Equipment Behavior Modelling for Fault Diagnosis and Deterioration Prognosis in Semiconductor Manufacturing

Hamideh Rostami

N° d'ordre NNT : 2018LYSEM028

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
MINES Saint-Etienne

Ecole Doctorale N° 488
Sciences, Ingénierie, Santé

Spécialité de doctorat: Génie Industriel

Soutenue publiquement/ à huis clos le 17/12/2018, par :
Hamideh Rostami

Equipment Behavior Modelling for Fault Diagnosis and Deterioration Prognosis in Semiconductor Manufacturing

Modélisation du Comportement des Equipements pour le Diagnostic des Pannes et le Pronostic des Détériorations des Machines en Fabrication de Semi-conducteurs

---

Devant le jury composé de :

Lambert-Lacroix Sophie, Professeur, Université Grenoble Alpes, Grenoble, France, Présidente
Castagliola Philippe, Professeur, Université de Nantes, Nantes, France, Rapporteur
S.Reis Marco, Professeur, Université de Coimbra, Coimbra, Portugal, Rapporteur
Chen Argon, Professeur, National Taiwan University, Taipei, Taiwan, Examinateur
Moyne James, Associate Research Scientist, University of Michigan, Michigan, US, Examinateur
Pinaton Jacques, ingnieur, STMicroelectronics, Rousset, France, Invité
Yugma Claude, Chargé de Recherche, EMSE, Gardanne, France, Directeur de thèse
Blue Jakey, Maître de Conférences, EMSE, Gardanne, France, Co-directeur de thèse

Spécialités doctorales	Responsables :	Spécialités doctorales	Responsables
SCIENCES ET GENIE DES MATERIAUX	K. Wolski Directeur de recherche	MATHEMATIQUES APPLIQUEES	O. Roustant, Maître-assistant
MECANIQUE ET INGENIERIE	S. Drapier, professeur	INFORMATIQUE	O. Boissier, Professeur
GENIE DES PROCEDES	F. Gruy, Maître de recherche	SCIENCES DES IMAGES ET DES FORMES	JC. Pinoli, Professeur
SCIENCES DE LA TERRE	B. Guy, Directeur de recherche	GENIE INDUSTRIEL	N. Absi, Maitre de recherche
SCIENCES ET GENIE DE L'ENVIRONNEMENT	D. Graillot, Directeur de recherche	MICROELECTRONIQUE	Ph. Lalevée, Professeur

EMSE : Enseignants-chercheurs et chercheurs autorisés à diriger des thèses de doctorat (titulaires d'un doctorat d'État ou d'une HDR)

ABSI	Nabil	MR	Génie industriel	CMP
AUGUSTO	Vincent	CR	Image, Vision, Signal	CIS
AVRIL	Stéphane	PR2	Mécanique et ingénierie	CIS
BADEL	Pierre	MA(MDC)	Mécanique et ingénierie	CIS
BALBO	Flavien	PR2	Informatique	FAYOL
BASSEREAU	Jean-François	PR	Sciences et génie des matériaux	SMS
BATTON-HUBERT	Mireille	PR2	Sciences et génie de l'environnement	FAYOL
BEIGBEDER	Michel	MA(MDC)	Informatique	FAYOL
BLAYAC	Sylvain	MA(MDC)	Microélectronique	CMP
BOISSIER	Olivier	PR1	Informatique	FAYOL
BONNEFOY	Olivier	MA(MDC)	Génie des Procédés	SPIN
BORBELY	Andras	MR(DR2)	Sciences et génie des matériaux	SMS
BOUCHER	Xavier	PR2	Génie Industriel	FAYOL
BRODHAG	Christian	DR	Sciences et génie de l'environnement	FAYOL
BRUCHON	Julien	MA(MDC)	Mécanique et ingénierie	SMS
CAMEIRAO	Ana	MA(MDC)	Génie des Procédés	SPIN
CHRISTIEN	Frédéric	PR	Science et génie des matériaux	SMS
DAUZERE-PERES	Stéphane	PR1	Génie Industriel	CMP
DEBAYLE	Johan	MR	Sciences des Images et des Formes	SPIN
DEGEORGE	Jean-Michel	MA(MDC)	Génie industriel	Fayol
DELAFOSSE	David	PR0	Sciences et génie des matériaux	SMS
DELORME	Xavier	MA(MDC)	Génie industriel	FAYOL
DESRAYAUD	Christophe	PR1	Mécanique et ingénierie	SMS
DJENIZIAN	Thierry	PR	Science et génie des matériaux	CMP
DOUCE	Sandrine	PR2	Sciences de gestion	FAYOL
DRAPIER	Sylvain	PR1	Mécanique et ingénierie	SMS
FAUCHEU	Jenny	MA(MDC)	Sciences et génie des matériaux	SMS
FAVERGEON	Loïc	CR	Génie des Procédés	SPIN
FEILLET	Dominique	PR1	Génie Industriel	CMP
FOREST	Valérie	MA(MDC)	Génie des Procédés	CIS
FRACZKIEWICZ	Anna	DR	Sciences et génie des matériaux	SMS
GARCIA	Daniel	MR(DR2)	Sciences de la Terre	SPIN
GAVET	Yann	MA(MDC)	Sciences des Images et des Formes	SPIN
GERINGER	Jean	MA(MDC)	Sciences et génie des matériaux	CIS
GOEURIOT	Dominique	DR	Sciences et génie des matériaux	SMS
GONDRAN	Natacha	MA(MDC)	Sciences et génie de l'environnement	FAYOL
GONZALEZ FELIU	Jesus	MA(MDC)	Sciences économiques	FAYOL
GRAILLOT	Didier	DR	Sciences et génie de l'environnement	SPIN
GROSSEAU	Philippe	DR	Génie des Procédés	SPIN
GRUY	Frédéric	PR1	Génie des Procédés	SPIN
GUY	Bernard	DR	Sciences de la Terre	SPIN
HAN	Woo-Suck	MR	Mécanique et ingénierie	SMS
HERRI	Jean Michel	PR1	Génie des Procédés	SPIN
KERMOUCHE	Guillaume	PR2	Mécanique et Ingénierie	SMS
KLOCKER	Helmut	DR	Sciences et génie des matériaux	SMS
LAFOREST	Valérie	MR(DR2)	Sciences et génie de l'environnement	FAYOL
LERICHE	Rodolphe	CR	Mécanique et ingénierie	FAYOL
MALLIARAS	Georges	PR1	Microélectronique	CMP
MOLIMARD	Jérôme	PR2	Mécanique et ingénierie	CIS
MOUTTE	Jacques	CR	Génie des Procédés	SPIN
NEUBERT	Gilles			FAYOL
NIKOLOVSKI	Jean-Pierre	Ingénieur de recherche	Mécanique et ingénierie	CMP
NORTIER	Patrice	PR1	Génie des Procédés	SPIN
O CONNOR	Rodney Philip	MA(MDC)	Microélectronique	CMP
OWENS	Rosin	MA(MDC)	Microélectronique	CMP
PERES	Véronique	MR	Génie des Procédés	SPIN
PICARD	Gauthier	MA(MDC)	Informatique	FAYOL
PIJOLAT	Christophe	PR0	Génie des Procédés	SPIN
PINOLI	Jean Charles	PR0	Sciences des Images et des Formes	SPIN
POURCHEZ	Jérémy	MR	Génie des Procédés	CIS
ROUSSY	Agnès	MA(MDC)	Microélectronique	CMP
ROUSTANT	Olivier	MA(MDC)	Mathématiques appliquées	FAYOL
SANAUR	Sébastien	MA(MDC)	Microélectronique	CMP
STOLARZ	Jacques	CR	Sciences et génie des matériaux	SMS
TRIA	Assia	Ingénieur de recherche	Microélectronique	CMP
VALDIVIESO	François	PR2	Sciences et génie des matériaux	SMS
VIRICELLE	Jean Paul	DR	Génie des Procédés	SPIN
WOLSKI	Krzystof	DR	Sciences et génie des matériaux	SMS
XIE	Xiaolan	PR0	Génie industriel	CIS
YUGMA	Gallian	CR	Génie industriel	CMP

Mise à jour : 03/02/2017

Acknowledgments

This thesis would not have been possible without the financial support of the ANR (Agence Nationale de la Recherche) in France under the international collaborating project IMAGINE (ANR-15-CE10-0003) with MoST (Ministry of Science and Technology) in Taiwan.

First and foremost, I am sincerely thankful to my supervisors Dr. Jakey Blue and Prof. Claude Yugma, who gave me confidence and freedom to complete this work. Through their extensive guidance, they taught me more than I can give them credit for here.

I am also extremely grateful for all of those with whom I have had the pleasure to work and discussion, among them Prof. Argon Chen and Prof. Marco Reis for their suggestions and valuable insight during my research.

I am indebted to my friends for their presence and their marvelous inspiration. Thanks for being by my side and always giving me reasons to be cheerful.

For my family, for whom all the words in the world cannot express my thanks. I offer my profound gratitude, for their unbounded thoughtfulness and support, mentally and spiritually. They have my deepest appreciation for being there when I needed it most.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter provides a general introduction of this thesis and starts by explaining the problem background and motivations in **Section 1.1**. The underlying problem is stated in **Section 1.2**. **Section 1.3** describes the research methodology and objectives following by the expected results.

1.1 Background and Motivation

Nowadays, high-tech production systems become more and more complicated, and data are generated and collected faster than data processing. Continuous data collection requires data analyzing with efficient novelty and effectiveness. Semiconductor industry with the improved Integrated Circuit (IC) design and manufacturing technology has collected varied data sources for mining operational knowledge. To obtain this knowledge, the manufacturer should analyze the received data quickly and efficiently to operate the equipment at high utilization and obtain high production yield.

Low tool utilization and inferior production yield are mainly due to production deficiencies such as process variations and unexpected tool breakdowns. In the semiconductor industry, unexpected failures have more consequences that are serious and tremendously increase investment and operation cost. The typical value of 300 mm fab for monthly production of 25,000 wafers exceeds US$ 2.0-3.0 billion. Such colossal investment urgently necessitates the improvement of operational effectiveness. Among various sorts of costs, equipment cost usually contributes the most significant part of the capital investment. This contribution is expected to take up to 75% of total investment in a typical 300 mm fab [1]. Accordingly, high tool utilization and top equipment effectiveness have become extremely necessary requisitions for semiconductor industries with intensive capital investment in equipment. However, the complex manufacturing environment has raised production deficiencies such as process variations and unexpected tool breakdowns and has made the evolution of machine conditions more and more difficult.

The earliest and most traditional strategy to prevent such violations into in-situ control at semiconductor production has been corrective strategies such as ”fix it when it

breaks." [2]. The problems with this strategy are various as the occurrence of unexpected breakdowns at inconvenient periods of production. This phenomenon leads to uncommitted friction and wears in production schedules that directly results in market profit loss and customer dissatisfaction. Instead, Condition-Based Monitoring (CBM) as a paradigm in Advanced Process Control (APC) theory infers equipment condition and alarms required action based on the runtime data analysis [3]. For this purpose, CBM requires a prognostic module that represents the healthy state of the equipments behavior, enables a manufacturer to avoid equipment breakdown and unnecessary maintenance, and a diagnostic module to identify the causes of the equipment failures.

These two prognostic and diagnostic modules help to develop an advanced equipment deterioration modeling and monitoring that not only results in efficient equipment condition monitoring but also helps to diagnose any failure cause. Such model in the shortest possible time is critical for minimizing scrap wafers, reducing unscheduled equipment breakdowns, minimizing unqualified periods of the equipment and consequently maintaining high process yields, which has motivated this research. Briefly speaking, the primary motivations of prognostic and diagnostic equipment deterioration modeling and monitoring are: (1) minimizing repair and maintenance costs and associated operational breakdowns, and (2) maximizing tool utilization and production yield. **Figure 1.1** shows the application of the prognostic and diagnostic results in the industry that takes place well in the current Advanced Process Control (APC) system in semiconductor manufacturing.

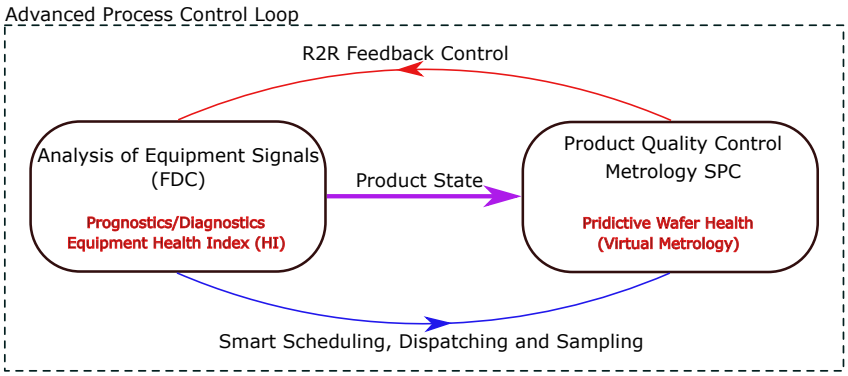


Figure 1.1: Application of the prognostic and diagnostic results in the industry.

Based on **Figure 1.1**, the prognostic and diagnostic findings of this thesis can be efficiently applied in the industry from different perspectives as 1) *Process* perspective, 2) *Equipment* perspective, and 3) *Product* perspective.

- From the *process* perspective, equipment deterioration modeling provides knowledge about the qualification level of equipment. This knowledge determines which set of equipment are qualified to undertake production activities and which set of equipment are less qualified to do operations. Accordingly, the manufacturer can develop an efficient and proactive production scheduling and dispatch. With the

aid of this model, such dynamic dispatching and scheduling will play an important role to minimize variability in almost all automated and sophisticated manufacturing factories, particularly semiconductor industry.

- From the *equipment* perspective, equipment deterioration modeling provides a diagnostic and prognostic maintenance plan to have a minimal maintenance cost and the minimum probability of unexpected equipment breakdown. Having fewer equipment downtime increases equipment utilization and production yield directly.
- From the *product* perspective, equipment deterioration modeling reports the health of equipment and recommends to readjust or re-set process settings though Run-to-Run system to avoid product defects. This feedback mechanism not only significantly minimizes the product waste and overall production cost, but also enhances the customer satisfaction and market share indirectly.

This research is a part of an international project named **IMAGINE** (**I**ntegrated **MA**nufacturing Decisions for Next **G**enerat**I**o**N** Factori**E**s) between *Ecole des Mines de Saint-Etienne* in France and *National Taiwan University* in Taiwan. The French part has been financed by the *French National Research Agency* (ANR-15-CE10-0003). It is worthy to mention that all the developed methodologies in this thesis have been tested and validated through the practical case studies with the local partners (STMicroelectronics in France).

1.2 Problem Statement

In this thesis, we model the equipment deterioration in Batch Manufacturing Processes that are pervasive modes in todays semiconductor fab. Semiconductor manufacturer has been taking the advantage of the rapid IT evolution, which enables more efficient analyses in equipment deterioration modeling. We benefit from such a booming growth of data as an important key for APC solutions to prognose equipment deterioration and diagnose failure causes. The batch manufacturing processes include an extensive data collection with three-dimensions (product $\times$ sensor $\times$ observation). In semiconductor manufacturing, the batch process data obtained from the equipment are usually referred to be the Fault Detection and Classification (FDC) data.

The first issue in equipment deterioration modeling of FDC data is the health index (HI) extraction from a large and heterogeneous data size, which has been focused unevenly in the literature. The second issue is multiple recipe contexts in semiconductor productions thereby particular process dynamics happen. For instance, in semiconductor manufacturing processes, almost all processes (e.g., etching and deposition) are carried out with different process recipes. Therefore, the normal changes in sensor readings between two different recipes must not be declared as equipment deterioration or fault.

Therefore, the underlying problem is to extract the knowledge (if any) regarding the FDCs behavioral variations and develop an efficient equipment deterioration model. This model is then used to prognose the equipment condition and remaining useful life as well as to diagnose the failure causes.

1.3 Research Methodology and Objectives

From a methodological point of view, there exist typically three approaches for equipment deterioration modeling and monitoring such as Physical model-based, Knowledge-based, and Data-driven prognostic approaches [4, 5, 6]. Physical model-based, accurately and analytically quantify the characteristics of a failure mode using physical rules regarding mathematical formulations. This interpretation requires a deep understanding of the equipment behavior and detail information of each failure mode [7]. Not surprisingly, physical models are mathematically complex particularly for multiplex manufacturing systems like semiconductor that contains hundreds or even thousand convoluted failure modes. Knowledge-based approaches assess the match between an observed condition and a databank of historical situations. This assessment is then used to monitor the equipment condition or even to predict and interpret the Remaining Useful Life (RUL) of the equipment [6]. Knowledge-based approaches such as expert systems contains an accumulated subject matter experience from experts and a set of precise IF-THEN rules. The accuracy and applicability of the knowledge-based approaches are limited due to the humane comprehension uncertainty about the current equipment behavior and its future evolution [8].

Thanks to the advancement of data collection technologies in semiconductor industry, data-driven approaches have become efficient methods to learn the system behavior through large volume of historical data of equipment conditions. Data-driven approaches are increasingly used for the equipment behavior modeling and monitoring [9, 10, 11] as well as the equipment failure diagnostic [12, 13].

Accordingly, this thesis aim at developing an efficient and novel data-driven approach that takes FDC data to prognose the equipment deterioration as well as diagnose failure causes. **Figure 1.2** illustrates the major steps of the proposed data-driven research methodology, wherein 1) *Data Pretreatment* is a set of activities to purify and prepare the data for further analysis. Some of these activities are eliminating/replacing the missing data, eliminating redundant variables etc., 2) *Data Processing* is to do a set of pre-processes such as integration, unfolding and decomposition of signals to make the data more informative and ready for building the model, 3) *Feature Selection and Extraction* is to transform the data into a set of features that efficiently represent the equipment condition, 4) *Equipment behavior Modeling and Controlling* is the main body of this methodology to model the equipment behavior based on the condition data, and finally 5) *Behavior Prognosis* and *Failure Diagnosis* are performed based on the equipment model and extracted features. The result of the Behavior Prognosis step is a trend that shows the condition of the equipment through the time. In addition, the result of the Failure Diagnosis is characterizing the equipment failures with corresponding fault roots.

Therefore, the objectives of solving the underlying problem using the proposed data-driven approach are:

- Extracting valuable knowledge from the FDC data regarding the equipment health condition and the equipment failure causes,

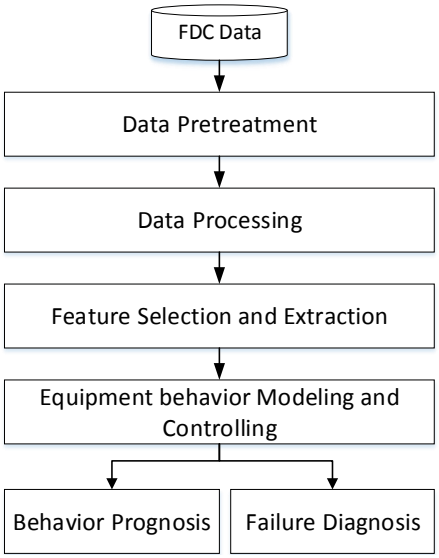


Figure 1.2: Research methodology of the proposed data-driven approach is illustrated.

- Developing an efficient and novel data-driven approach to model and monitor the equipment deterioration and
- Employing the proposed approach to diagnose the equipment failure causes in terms of fault signatures with corresponding roots.

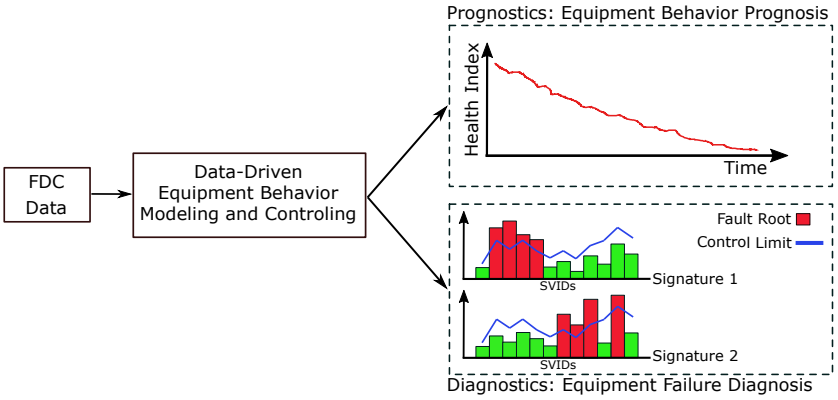


Figure 1.3: Equipment behavior modeling and controlling obtains expected prognostic and diagnostic results from FDC data.

In the following, **Figure 1.3** illustrates the transformation of the FDC data into prognostic and diagnostic results using the proposed data-driven approach. The prognostic results contain the equipment deterioration model that quantify the equipment health as an index (Health Index) over the time. This model can be used further to

monitor the equipment behavior. On the other hand, diagnostic results include finding different fault signatures and their corresponding roots that have created equipment's health degradation.

Chapter 2

Literature Review

In this chapter, the relevant scientific efforts are reviewed to position our work in the literature regarding its originality and contributions. Firstly, a comprehensive terminology of equipment behavior modeling, monitoring, and different prognostic and diagnostic methods are discussed in **Section 2.1** to **Section 2.3**. In **Section 2.4**, Advance Process Control (APC) in the semiconductor industry is briefly presented. Relevant papers studying equipment behavior prognosis and equipment failure diagnosis are reviewed in **Section 2.5**, respectively, followed by an analysis performed to find out the literature gap as well as to position this research in the literature. Finally, the research methodology in this thesis and its contributions are explained in **Section 2.6**.

The modern industries with advanced technologies increasingly require to work at high availability, low environmental risks, and to keep the machine at high utilization level. Technological development has resulted in unexpected violations that have more consequences which are severe and exponentially increase investment and operation cost. It becomes difficult or almost impossible to identify and predict condition variations promptly.

The economic consequences from an unexpected one-month stoppage in the industry may become as high as up to US$ 2.0-3.0 billion [1]. With such an enormous investment cost, the operational effectiveness of industrial equipment and production processes has a significant impact on the profitability and competitiveness of manufacturing industries. The efficient availability emphasizes the increasing importance of effective monitoring strategies of the equipment and production processes in industry.

Traditional strategies to cope with production variations and equipment breakdown have been Corrective Maintenance (CM) or Preventive Maintenance (PM) strategies [14], [2]. CM takes place when the equipment has failed, and PM takes place periodically to prevent unexpected breakdowns regardless of the equipment's health condition [14]. If equipment fails and PM is performed unnecessarily, these strategies result in uncommitted friction and wear in production schedules that impose enormous costs to the industry. Instead, Condition-Based Monitoring (CBM), as a type of predictive maintenance, involves using sensors to measure the condition of the equipment over time while they are operating and alarms required action based on the runtime data analysis [3]. With CBM, maintenance only takes place when the data reveals performance

deterioration or failure that is likely to occur.

CBM aims at avoiding unnecessary breakdowns by alarming the requirement of preventive actions (e.g., re-adjustment, repairs or replacements) only when there is enough evidence of deterioration in the equipment health condition. A properly established and effectively implemented CBM program can signicantly cut down maintenance costs by reducing the number of unnecessary scheduled PMs. Two wings enforcing CBM to reach such success are prognostic and diagnostic modules, whereby CBM monitors and assesses equipment health based on condition measurements that do not interrupt normal equipment operations [4]. By this assessment, CBM determines whether the equipment is necessitated doing maintenance or not; and if it is necessary, predicts when any action needs to be performed. Moreover, the prognosis aspect of CBM models the equipment health monitoring through the time and notifies engineers whenever equipment health deteriorates to an alarm level. This capability will provide adequate time for the engineers to take diagnostic actions to detect a fault(s), identify fault fingerprint(s) and extract fault root(s) [15].

2.1 Condition-Based Monitoring

Condition-Based Monitoring (CBM) is a strategy that estimates and monitors the condition of the equipment to decide whether any preventive action is required. Unlike scheduled preventive actions, maintenance and Out-of-Control Action Plan (OCAP) should only be conducted if enough symptoms appear, such the decreasing performance, and/or specific indicators signify the deterioration for imminent failure.

In CBM, the equipment health is measured based on various monitoring indicators, such as vibration, temperature, contaminants, and noise levels in mechanical equipment or projected gas volume, temperature and radio frequency in semiconductor industry [15]. The motivation of CBM is that 99% of equipment failures are preceded by specific signs or indications that alarm failure is going to occur [16]. Therefore, CBM is vital for managing equipment utilization, reducing life-cycle cost, avoiding unexpected equipment breakdown, etc.

The principal of CBM is, therefore, equipment behavior modeling built over signals which are continuously monitored by using a set of sensors installed inside or on the equipment [17]. Based on the equipment behavior model, the necessity of the preventive actions (e.g., re-adjustments, repairs or replacements) is alarmed only when required or when equipment health degrades to an emergency level. This emergency level is established by expert engineers and differs from one equipment to another. An essential capability for any condition modeling and monitoring process is that alarms should be triggered within an extended enough period before the failure. Required actions can be taken before the equipment breaks down or deteriorates to the unacceptable condition.

Two modules of condition modeling and monitoring to reach the above-mentioned advantageous are equipment prognosis and diagnosis. The prognostic module develops an equipment deterioration model whereby future condition can be prognosed. The diagnostic module implements an equipment behavior model for detecting the faults

and identifying the causes. These two modules are often employed integrally as the field of Prognostic and Health Management (PHM). As a prognostic discipline, PHM aims at providing an overview of the equipment health including early failure detection and Remaining Useful Life (RUL) prediction. Diagnostics is also integrated into PHM to identify and determine the relationship between cause and effect in the equipment and consequently isolate faults and identify failure causes. Conventionally, the modeling approaches for equipment prognosis and the failure diagnosis problem can be categorized into three main categories: Physical model-based, Knowledge-based, and Data-driven prognostic approaches [14, 4, 5, 18, 19].

2.1.1 Physical Model-based Approaches

Physical model-based approaches accurately and analytically quantify the behavioral characteristics of the equipment's physics through mathematical formulations. These approaches require a deep understanding of the equipment behavior in response to both internal and external abnormalities [7]. Once a physical model is derived, the actual condition of the equipment regarding sensor measurements are compared with outputs of the model. Substantial differences between actual condition and the model production are indicated as an abnormality (i.e., deterioration or failure) while small and acceptable differences occur under normal conditions [4]. The physical-model based approaches apply analytical models to evaluate the consistency of substantial differences. These analytical models could be physical specific or explicit mathematical model. These approaches become efficient if a correct and accurate model is built. However, they are not only mathematically complex but also they must be resolvable into individual failure modes. Accordingly, physical-model based approaches may not be feasible for complex systems (e.g., semiconductor industry) because of hundred or even thousand failure modes that are majorly convoluted.

2.1.2 Knowledge-based Approaches

Knowledge-based approaches are usually extracted by an engineer or some empirical knowledge which are later transferred to the rules and assess the similarity between an observed condition and a databank of previously dened situations [6].

An expert system as a sub-category of the knowledge-based approaches includes a set of programs that simulate the performance of human experts in a particular eld [6]. Such a knowledge-based system contains an accumulated subject matter experience from experts and a set of precise IF-THEN rules for using that knowledge to specific problems. One or more experts construct the rules over some years and, for being useful, the rules must be as complete and precise as possible [20]. Despite the physical model-based approaches, knowledge-based approaches are particularly efficient for real non-linear systems that are extremely difficult to be modeled and the linear approximation of the model results in significant errors. Also, knowledge-based diagnostic approaches are more flexible, easy to understand and follow.

Although the output of the knowledge-based approaches is understandable and reasonable, they are as good as the expert's knowledge. Also, they are disabled to predict continuous variable such as RUL since the outputs are determined by a discrete set of rules [6]. Furthermore, to perform equipment failure diagnosis, a deep understanding of the physical properties of the equipment is either unavailable or too costly to obtain.

2.1.3 Data-driven Approaches

The manufacturers have been taking advantage of the rapid IT evolution, which enables efficient analyses in equipment condition prognosis and diagnosis. In the high-tech manufacturing industries such as the semiconductor industry, the booming growth of data is an essential key for APC solutions to monitor, prognosis, and diagnose at better levels. While the move to big data solutions for APC systems is critical and necessary, manufacturers build data-driven approaches to learn the system behavior through a large volume of historical data of equipment conditions. Data-driven approaches are increasingly used for the equipment behavior modeling and monitoring [9, 10, 11] as well as the equipment failure diagnostic [12, 13]. They perform prognosis and diagnosis by training a set of run-to-failure units to build a model for the system deterioration, which in the deterioration features are extracted from raw signals [9]. It is worth mentioning that any non-perfect data-driven approach might be enhanced by the appropriate application of Subject Matter Expertise (SME) as a complementary capability. This expertise can be adapted from the knowledge-based approaches (e.g., to determine warning levels in a failure prognosis scheme, expert opinion can be elicited and employed in this regard).

Data-driven approaches can signicantly reduce the number of unexpected equipment downtime resulting in minimal maintenance costs if properly designed and effectively implemented for prognostic and diagnostic modeling. In this approach, any prognostic and diagnostic program consists of three essential steps [14]:

1. *Data collection* step, to obtain data relevant to the equipment condition;
2. *Data processing* step, to prepare, handle and manipulate the data for better understanding and interpretation of the equipment condition;
3. *Feature extraction* step, to summarize prepared data into features that well represent the equipment condition.

Data collection

Data collection is a process of collecting and storing useful data from targeted physical assets for prognostic and diagnostic purposes. This process is a vital step in implementing a CBM program for equipment behavior prognosis or equipment failure diagnosis. Effective CBM helps to establish the asset for parameter operating and to position the sensors correctly to collect the data. CBM data can be identified from different points of view such as the data value or the data application. Formerly, the data value is grouped into nominal (discrete or continuous), categorical or ordinal data. Nominal data have meaning as a measurement like a number of faulty products (discrete) or temperature

sensor reading (continuous). Categorical data represent characteristics such as the type of products (e.g., type "A", "B" etc.). Finally, the mix of both numerical and categorical data makes ordinal data such as quality of products that can belong to categories and have nominal value in each category. From the application point of view, two main types of collected data in a CBM program are event data and condition monitoring data [14]. Event data include the information about what has happened to the equipment (e.g., breakdowns, downtime, failure causes) and/or what has been done for these events (e.g., repair, preventive maintenance, piece exchange, parameter readjustment, etc.). Condition monitoring data are the measured information related to the equipment health condition/state and are very versatile based on the equipment mechanism (e.g., mechanical, electrical and chemical mechanisms).

Data processing

As the first sub-step of data processing, data cleaning is an essential step since data, particularly manually entered event data, always contains errors (noises). With data cleaning, it is assured as high as possible that further equipment behavior modeling is performed on clean (error-free) data. Without the data cleaning step, equipment behavior modeling may lead to confusing or even wrong results (i.e., increase in errors type I and type II).

The second sub-step of data processing is data manipulation. Data manipulation is the process of changing data to make it more informative to be analyzed. A variety of models, algorithms, and tools are available in the literature to manipulate data for a better understanding and interpretation of data. The models, algorithms, and tools used for data manipulation directly depend on the way, sampling frequency and dimension of the data. Condition monitoring data can be collected in three ways: 1) Data are obtained as a single value at a specific period for a condition monitoring variable such as temperature and pressure of the equipment. These data are also called as value type data; 2) Data are collected as time series at a specific time period for a condition monitoring variable such as vibration of mechanical equipment; 3) Data are collected in two dimensions (images) at a particular period of time for a condition monitoring variable such as the images of the product surface quality. These data are also called multi-dimensional data.

Accordingly, based on the type and the form of collection, data processing methods, and algorithms differ from multivariate analysis (e.g., Principle Component analysis PCA [15]) to decomposition technique (e.g., Wavelet Decomposition [21]).

Feature extraction

Feature extraction step is the transformation of input data into a set of features, which are distinctive properties of input patterns and help in differentiating between the categories of input patterns. This step has become one of the main principles of CBM to identify a set of equipment behavior indicative features within the equipment sensor readings and to utilize these features with an appropriate approach to prognose the

equipment health and diagnose the equipment failures. Various studies have focused on this problem and addressed the importance of extracting the underlying features by equipment behavior. The primary challenge in feature extraction is deciding which feature types represent the equipment behavior since the sensitivity of various features may differ from one working condition to another. The variety of features and the way of extracting them depends on the data type as well as the method of processing the data and building the equipment condition model [13, 22, 23, 24, 25].

2.2 Equipment Behavior Prognosis

Prognostic has been defined in different wordings as some of them quoted in **Table 2.1**. The primary foundation of these definitions is a) prognostic involves predicting the time progression of a specic failure mode from its incipience to the time of component failure, b) a prediction of future component status is required, and c) prognostic and diagnostic are interconnected, but they are not the same [6].

Based on the definition presented by ISO13381-1, prognostic modeling is an estimation of time to failure and risk for one or more existing and future failure modes [19]. This definition states that equipment behavior prognostic is interested in both predicting the effects of known failure modes on the equipment health as well as how these may initiate other failure modes.

Hence, an efficient prognostic model should be able to answer most of the following questions [5], [6]:

(a) How is the machine operating now?
(b) How severe is the deterioration?
(c) How quickly is the machine expected to deteriorate from its current state to functional failure?
(d) When will the machine fail with what likelihood? i.e., what is the RUL with what range?
(e) What will be the primary faults that cause the deterioration?
(f) Why does the fault occur?

Based on the denition proposed in **Table 2.1**, questions (e) and (f) can be considered as diagnostic questions, while the first four are thereby the realm of pure prognostic. Ongoing diagnostics through the prognosis process is also necessary to detect underlying fault modes that cause equipment deterioration so that required maintenance actions will be appropriately proposed.

Instead of building a sophisticated equipment behavioral model to include all possible factors, it is more practical to develop a prognostic model to take only the deterioration-related factors into account. The objectives of a prognostic model are twofold: 1) developing equipment health model (i.e., questions (a)-(c)) and 2) predicting RUL (i.e., question (d) with a certain likelihood and range).

Table 2.1: Prognostic's definition.

Prognostic is ... (direct quote)	**Ref.**
An estimation of time to failure and risk for one or more existing and future failure modes	[19]
The capability to provide early detecting of the precursor and/or incipient fault condition of a component, and to have the technology and means to manage and predict the progression of this fault condition to component failure	[26]
Predictive diagnostics, which includes determining the remaining life or time span of proper operation of a component	[27]
Failure prognosis involves forecasting of system degradation based on observed system condition	[28]
The ability to assess the current health of a part for a xed time horizon or predict the time to failure	[29]
Prognostic builds upon the diagnostic assessment and are dened as the capability to predict the progression of this fault condition to component failure and estimate the RUL	[30]
Prediction of when a failure may occur. To calculate the remaining useful life on an asset	[31]
The process of health assessment and prediction, which includes detecting the incipient failure and predicting RUL	[5]

As the essential prognostic task, equipment behavior modeling aims at analyzing real-time data to observe, evaluate and quantify the equipment health. For this task, all condition monitoring information are merged into a single value as the Health Index (HI). Equipment HI provides a real-time manifestation of the constant condition and can trigger an alarm before functional failure occurs. Conceptually, the equipment health cycle can be divided into three main zones: normal, deteriorating and unacceptable zones as shown in **Figure 2.1**, wherein a situation with only one component failure and failure mode is consider. It is noted that in more complex equipment with multiple components and failure modes, the equipment health cycle can be drawn by merging the health cycle of different components that requires more sophisticated approaches. In the normal zone, equipment is operating normally without any evidence of health deterioration. After normal operation for a specific period, the equipment enters the deterioration phase, and the equipment health starts degrading. From the entrance point of equipment to the deterioration zone until a point (i.e., prediction point) where there are enough evidence and enough information about health deterioration, equipment health model is built.

With a truthful equipment behavioral model, the RUL to the unacceptable zone can be predicted. The prediction should be made long enough before the functional failure to trigger a preventative maintenance action. RUL, also called remaining service life,

residual life or remnant life refers to the time left before observing a functional failure [14]. It is worth mentioning here that proper design of equipment health model is very crucial to predicting the RUL correctly.

As elaborated in **Section 2.1**, three main categories of modeling approach in condition monitoring are physical model-based, knowledge-based and data-driven approaches. In the following, the aim is to explain some of the most relevant methods of each category applied in equipment behavior modeling.

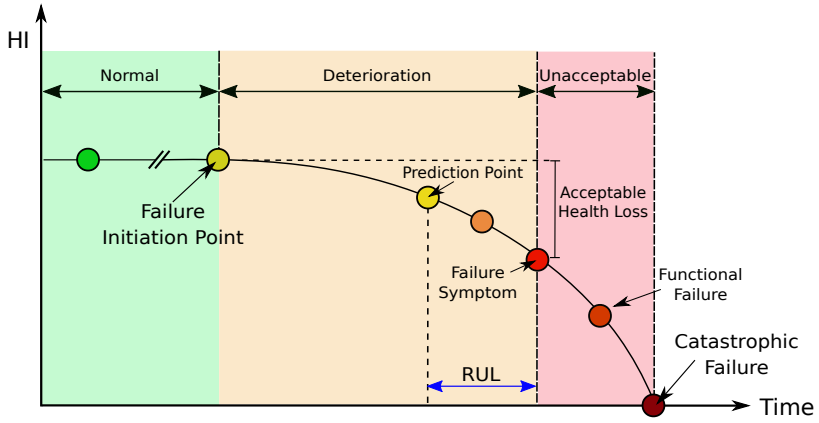


Figure 2.1: Equipment health cycle.

2.2.1 Physical Model-based Prognostic Approaches

Physical model-based prognostic approaches can be analytically designed to minimize the effect of unknown failures and perform the consistent sensitivity analysis. These approaches consider that an accurate mathematical model can be constructed from first principles. Physical models solve a deterministic equation or set of equations derived from empirical data to predict the equipment RUL. Accordingly, they require specific failure mechanism knowledge and theory relevant to the monitored equipment. Knowing the failure mechanism helps to derive the explicit relationship between the condition revealing features and the lifetimes [32].

These physical model-based methods often use residuals as features, where the residuals are the difference between the sensed measurements of the real system and the outputs of the mathematical model. The convention is that more and more the system deteriorates, the residuals become larger and larger and small/acceptable residuals may happen in the presence of normal variations, noise and modeling errors. The acceptance threshold is defined using statistical techniques to prognose the failures. In addition, the magnitude of the residuals can be modeled as the severe of the deterioration or failure. In some cases, detailed simulation is applied to model the physics of the equipment [28], [33],[34]. In [34], the authors developed a physical model-based approach by simulating the physics of an automotive component.

The main advantage of physical model-based approaches relies on their ability to incorporate a physical understanding of the monitored equipment. This understanding helps to explain accurately the behavior of the equipment and engineers can easily take appropriate actions. Moreover, if the understanding of the equipment deterioration improves, the model can be modified to increase its accuracy. On the other hand, this strong relationship with a mathematical model may become the Achilles heel of these approaches since it can be difficult, if not impossible, to catch or define the system's behavior, particularly in complex systems like semiconductor manufacturing. Furthermore, some authors believe that the condition monitoring approaches must evolve as the equipment does.

2.2.2 Knowledge-based Prognostic Approaches

It is usually difficult to build an accurate mathematical model based on the physics of the equipment. Accordingly, knowledge-based approaches which require no physical model are being utilized. Therefore, the knowledge-based prognostic approaches are used when there is a lot of experience but not enough details to develop accurate quantitative models. One well-known example of knowledge-based approaches is expert systems.

Expert systems are suitable for prognostic problems that can be solved by human experts. This system is usually considered as a computer system that is programmed to exhibit specialist knowledge [35]. The performance of these systems is based on the combination of computers power with the laws of reasoning. In an expert system, the expert's knowledge is stored into computers in the form of rules (rules are expressed in the form: IF condition, THEN consequence) and the system simulates the human's way of thinking and inferencing. Afterward, the created rules are used to generate solutions. Any expert system is built through four steps as knowledge acquisition, rules generation, rules verification and validation of models [35].

Among the prognostic tasks, expert systems have been mostly used for deterioration interpretation and maintenance action propositions and not for deterioration modeling. On the other hand, they are potent approaches in equipment failure diagnosis.

The main drawbacks of the expert systems are the difficulty in acquiring knowledge and generate the rules as well as handling new situations that are not covered explicitly in the current rules bank. Also, the behavior of the complicated and convoluted manufacturing systems requires a considerable number of rules to be described, and when the number of rules increases dramatically, the prognostic system becomes computationally inefficient.

As an example, the authors of [28] proposed an intelligent knowledge-based diagnostic and prognostic process. The process employs graph-based dependency models for fault diagnosis and analytical models for test design and fault detection. The proposed method contains four significant steps such as modeling, inferring, adaptive learning and predicting.

2.2.3 Data-driven Prognostic Approaches

Data-driven prognostic approaches are the most employed methods in the literature and cover a vast variety of methods varying from simple univariate regression models or multivariate statistical methods to sophisticated Artificial Neural Networks (ANNs). Data-driven prognostic approaches are usually developed from collected input/output data. These approaches can process a wide variety of data types and exploit the nuances in the data that cannot be described by physical models or be discovered by rule-based systems. Comparing to physical model- and knowledge-based approaches, data-driven approaches are considered as good compromises between the accuracy and the applicability.

Data-driven approaches to prognose the equipment RUL typically follow one of two strategies. The first strategy is a two-stage approach as 1) deterioration modeling then 2) RUL estimation. In the deterioration modeling, appropriate data processing (e.g., dimension reduction, data transformation, etc.), feature extraction, or pattern matching techniques are employed to map the equipment signals or features onto a single dimension degradation or health index [36]. At the end of this stage, usually, a (linear/nonlinear) deterioration trend is obtained. Technically, this first stage has the overlap with the realm of failure diagnostics since it is concerned with posterior event analysis (i.e., deterioration causes identification). In the RUL estimation stage, the deterioration model/trend is then extrapolated into the future until a predefined critical threshold limit is exceeded. The first-strategy approaches are normally built using multivariate statistical and regressive methods [34, 35, 36, 37]. The second strategy is a single-stage approach wherein the equipment RUL is directly estimated from the monitored signals or features. In this situation, the remaining life of the system is the output generated by the models. The second-strategy approach typically uses ANN methods [36]. ANNs can model complex non-linear systems and can generalize and adapt solutions from a limited data set [6]. Despite these advantages, ANNs are called as black box approaches because the mechanism that transforms the input into the output is obfuscated by a figurative box that is difficult if not impossible to interpret. Accordingly, it becomes often impossible to do cause-effect analysis in ANN-based equipment behavior prognosis approaches.

In the following, the regression method is explained as an example of the data-driven prognostic approaches.

Regression models estimate the failure initiation and progression based on historical inspection observations on similar equipment. Prognosis of future deterioration is performed by comparing these observations with models representing healthy behavior. Regression models use temporal data such as condition or process monitoring outputs.

Among the most well-known regression methods, trend extrapolation is one of the most straightforward and understandable forms of data-driven prognostic approach for engineers. This approach is based on simple trend analysis of a single monotonic parameter correlated with equipment health. This parameter may have originated from a single sensor or some sensors aggregated into a single variable. The aggregated variable is then plotted as a function of time representing the equipment health over the time.

Afterward, this trend is formulated using standard regression methods to predict the equipment RUL. An alarm threshold can be established, and the end of the equipment life is the time when trend reaches the threshold. More than one alarm level can also be determined for early warning and nal failure, as shown in **Figure 2.2**.

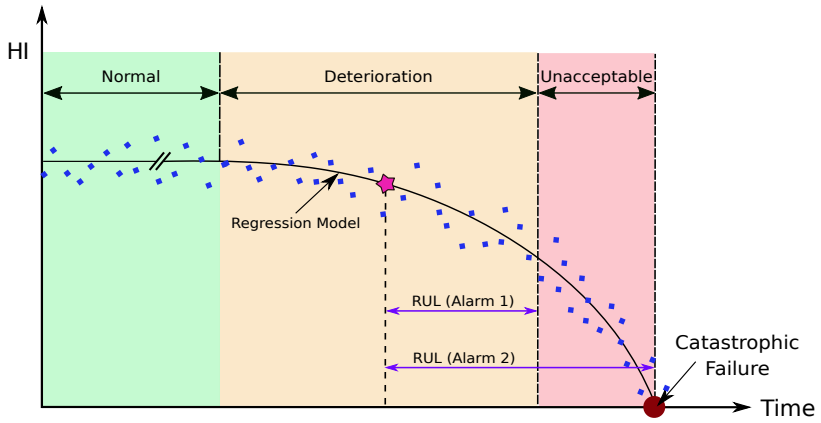


Figure 2.2: Statistical trend extrapolation with two alarms.

When new data arrives, it is assumed that behavior ts the known trend and so RUL merely is determined as the time between data arrival and the time at which the trend exceeds the alarm limit(s). Although representing equipment health and predicting equipment RUL using trend extrapolation is probably the most desired implemented approach within the industry, there are few published examples in refereed literature.

2.3 Equipment Failure Diagnosis

Equipment failure diagnosis, in contrast to the less well-understood prognosis, has been the subject of numerous investigations over the past decades. Equipment failure diagnosis approaches are to detect failures or anomaly conditions, isolate which component of the equipment is faulty, extract that failure modes exist and decide on the potential impact of a failing or failed component on the equipment health. The terms fault detection, isolation, and identication are typically used to convey the following meaning, [38, 39]:

- Fault detection: a task to indicate whether something is going wrong in the monitored system;
- Fault isolation: is a task to determines the location and the type of faults; and
- Fault identification: is a task to determine the nature (root) of the fault when it is detected.

In the following, three main equipment failure diagnosis approaches are explained in more details.

2.3.1 Physical Model-based Diagnostic Approaches

Physical model-based failure diagnostics can be structured as the detection, isolation, and identification of faults on the equipment by means of mathematical/analytical methods to extract the features from the equipment signals and diagnose the equipment faults. In these approaches, a fixed or variable threshold are set on residual signals calculated from the difference between the measurements of the real system and those of the physical model. Different residuals can be also generated where each is sensitive to a particular failure mode. Once the threshold is exceeded, the analysis of each residual leads to fault isolation and identification.

Any physical model-based failure diagnosis approach typically contains two blocks of residual generation and residual evaluation [39]. The residual generation block uses available input and output information from the monitored equipment to generate the residual signals based on analytical models. The analytical models could be physical specific or explicit mathematical model, e.g., parity equations and state observers, of the monitored equipment. Based on these models, residuals are generated using methods such as Kalman filter, system identification, and parity relations. The residual signals also called as fault symptoms indicate any fault occurrence where zero or close to zero values indicate no fault detection and significant magnitude of residuals indicates that fault has occurred. In the residual evaluation block, residuals are examined for the likelihood of faults to determine if any fault has occurred. This block may do a simple threshold test on the instantaneous or moving the average value of residuals; or more advanced threshold testing build by other statistical methods [40, 41].

2.3.2 Knowledge-based Diagnostic Approaches

Knowledge-based diagnostic approaches perform equipment failure diagnosis by evaluating on-line monitored data according to a rule set determined by expert knowledge [4]. This knowledge consists of the locations of input and output process variables, characteristics of abnormal conditions, failure modes, etc. On the other hand, the operators'and engineers'intelligence related to the equipment can be implemented into this approach. This knowledge aids to diagnose the equipment failures based on previous experiences.

Rule-based expert systems are the well-employed knowledge-based diagnostic approaches [18]. These systems are widely utilized for diagnostic purposes where experience and expertise are available, but a correct understanding of the equipment physics is either unavailable or too costly to obtain. The rule-based expert systems represent the knowledge in the form of production rules, wherein a rule explains the required action once a fault symptom appears. The main characteristic of these systems is the empirical association between causes and effects based on the expert's knowledge. This cause-effect association determines logical event chains that are used to describe the occurrence of the faults [42].

In addition to proper representation rules-based knowledge, an appropriate interface for the human-computer dialogue is also required. Afterward, the appeared fault's symptoms are presented to the user in a screen where the user can click the specific

symptom to start a searching process for the cause of the failure [23, 43]. Finally, a decision tree can be used to define the various logical paths that the expert must follow to reach conclusions.

There is the main difference between knowledge-based approaches in the equipment failure diagnosis and equipment behavior prognosis. In the behavior prognosis, no fault may have occurred, and the expert system is mostly used to interpret the deterioration causes and maintenance action proposition, while in failure diagnosis the expert system alarms the fault occurrence (fault has already occurred) as well as the fault causes.

2.3.3 Data-driven Diagnostic Approaches

Data-driven diagnostic approaches are popular methods of fault diagnostics to detect whether a specic fault exists or not based on the available condition monitoring information without intrusive inspection of the equipment [14]. The fault detection problem can be formulated as a hypothesis test problem with null hypothesis H_0: Fault A is present, against alternative hypothesis H_1: Fault A is not present. In actual fault diagnostic problems, hypotheses H_0 and H_1 are interpreted into an expression using specic distributions, or the parameters of a specic distribution. Test statistics are then constructed to summarize the condition monitoring information to decide whether to accept H_0 or reject it [44]. Interested readers are invited to study a recently proposed framework for fault diagnosis, called structured hypothesis tests, for handling complicated multiple faults of different types [45].

As a common data-driven diagnostic approach, Statistical Process Control (SPC) method has been well developed and widely used. The SPC principle is to quantify the drift of the current signal from a reference signal representing the normal condition to see whether the current signal is within the control limits or not [14]. Furthermore, clustering techniques (e.g., K-means clustering [46]), as multivariate statistical diagnostic approaches group signals into different fault classes based upon the similarity of the characteristics or features they possess. These approaches minimize intra-group and maximize inter-group variance [14].

Equipment failure diagnosis has been recently a very popular topic for industrial processes since they are free of the equipment physics and are capable of handling a large number of variables (hundreds). These approaches use techniques such as Principal Component Analysis (PCA) [47] or Partial Least Squares (PLS) [48]. PCA is widely used for dimension reduction, fault detection, and isolation [38] by constructing a model from the eigenvalues and eigenvectors of the covariance matrix of the data. Dimension reduction is performed by deleting the eigenvectors, or PCs, associated with sufficiently small eigenvalues. The equipment failure diagnosis comes from observing the interrelation between the variables and monitoring any drift in these relationships due to the fault occurrence in the equipment. Two fundamental statistical assumptions for the application of PCA are 1) variables are independent and identically follows the normal distribution, and 2) variables are uncorrelated in time (not auto-correlated) [38].

As other sets of data-driven diagnostic approaches, pattern recognition and deep learning approaches have been widely applied. In these approaches, the process is map-

ping the information obtained from the measurement and/or feature space to the equipment fault space [4, 49, 50]. Most of the pattern recognition approaches use ANN methods; however, it is not easy to apply ANN methods due to the lack of appropriate training data and specic knowledge to train the models. Accordingly, most of the existing researchers used experimental data for training the model.

Table 2.2 describes the advantages and disadvantages of different prognostic and diagnostic approaches in the literature. Also, **Table 2.3** explains where to use and where not to use particular prognostic and diagnostic approaches.

Table 2.2: Advantages and disadvantages of prognostic and diagnostic approaches.

Approach	**Method**	**Advantages**	**Disadvantages**
Physical model-based	Parity equations State observers	– Provide most accurate and precise models – Condence limits provided – Outputs can be easily understood – High accuracy in RUL prediction	– All failure modes should be precisely dened – The accuracy and robustness are subject to the experimental conditions under which models are developed – Complicated or impossible in non-linear systems – Impossible if having numerous failure modes
Knowledge-based	Expert systems	– Simple but time-consuming to develop – Easy to understand – Rules can be added or removed easily – Explanation of the reasoning process, induction and deduction process is easy	– Relies entirely on knowledge of subject matter experts – Lack of generality – Poor handling of novel situations – A signicant number of rules required – Precise inputs required – Inability to learn from their errors – No condence limits supplied – Not feasible to provide exact RUL prediction – Failure to represent time-varying phenomena – Development and maintenance is costly
Knowledge-based	Fuzzy logic systems	– Helpful in developing uncertain models of data – More compatible with human reasoning process than traditional symbolic approach – Fewer rules required comparing to expert systems – Inputs can be imprecise, noisy or incomplete – Failure severity can be reported – Condence limits can be provided on the output with some types of models	– Not feasible in the situation that membership functions are complicated to be determined – Lack of generality – Developing rules requires high expertise – Development and maintenance is costly – Linguistic terms may decrease the accuracy of the model
Data-Driven	Aggregate reliability functions	– Simple to be understood by engineers – Theoretically can be performed for all components of the equipment, if a small number of failure modes exist – Condence limits are available for RUL prediction of models	– Lack of generality for other equipment – It requires a statistically signicant sample size for each failure mode – Lack of accuracy in RUL prediction when lack of data – No alarm before the failure occurs, Requires high data quality (accuracy, archival length, context, no missing data, etc.)

Continued on next page

Table 2.2 – *Advantages and disadvantages of prognostic and diagnostic approaches (continued).*

Approach	**Method**	**Advantages**	**Disadvantages**
Data-Driven	Bayesian Networks	– Handle model multivariate, dynamic processes – Accommodate incomplete and noisy measurements – Manage incomplete data sets – Useful for learning the causal relationship – Expert knowledge can be integrated – Over tting of data can be edited by available algorithms – Variants available for non-linear processes – Other advantages depend on the underlying Bayesian technique	– They are useful only when a priori knowledge is available – Unpredictable failure roots cannot be modeled – Exploring unknown networks are computational difcult – Modeling experts are required – They require measurement data – Avoiding degeneracy requires a large number of samples
Data-Driven	Markov Model	– Able to model numerous equipment and failure modes – Specic knowledge of failure progression is not Required – They can manage incomplete data sets – They can model spatial and temporal data – Computationally efcient once developed	– They require a large volume of data for training – They assume only single monotonic and non-temporal failure deterioration – They are unable to model previously unanticipated failure roots
Data-Driven	Principal Component Analysis (PCA)	– It reduces multi-dimensional data sets to lower dimensional ones	– Its performance varies for different applications – It is unable to model non-linear systems (Linear transformation) – They are enabled to model the distribution of convoluted failure modes
Data-Driven	Trend Extrapolations	– They are simple to be applied – They are independent of process dynamic (normal ups and downs required by process recipes) – They are clear to be explain – They easily provide alarms – They do not require advanced software tools	– The monotonic and single-parameter trend can present few failures – The trend may be affected by measurement noises

Continued on next page

Table 2.2 – *Advantages and disadvantages of prognostic and diagnostic approaches (continued).*

Approach	**Method**	**Advantages**	**Disadvantages**
Data-Driven	Gaussian Mixture Models (GMM)	– They can model non-stationary data containing several distributions – They can model arbitrary distribution functions as a mixture of Gaussians – Following mixture models over the time provide good equipment condition trend – Distribution of different failure modes can be separately followed-up	– The optimization parameters are sensitive to the initialization methods – Determining the number of mixture models is not easy – They are enabled to model the distribution of convoluted failure modes
Data-Driven	Self-Organizing Maps (SOM)	– They do not need prior output (Unsupervised learning methods) – They provide excellent visualization – They can efficiently cluster different failure modes – They efficiently transfer high-dimensional data into a two-dimensional map	– Lack of standard algorithm to determine the structure and shape of the map – They suffer from block-box processing
Data-Driven	Artificial Neural Networks (ANNs)	– They can model complex, multi-dimensional and non-linear equipment – They do not require a physical understanding of the equipment and failure modes – They are available with a large variety for different purposes – They can be applied to different types of data	– They require a signicant amount of data for training – They do mostly trial and error for determining the most appropriate model – They are not timely-efficient – They need data enough clean – The high number of input variables makes them highly complex – They are black-box processing, and they do not provide cause-effect analyses – Outputs need to mapped to a physical representation
Data-Driven	Support Vector Machine (SVM)	– It achieves better decision accuracy in particular cases because of the maximized decision boundary – It is efficient for non-linear data thanks to kernel functions – It is competent enough for the large dataset and real-time analysis – It can cluster different failure modes efficiently	– Several parameters that need to be tuned accurately – They cannot provide deterioration trends – They need prior output (Supervised learning methods)

Table 2.3: Where (not) to use prognostic and diagnostic approaches.

Approach	**Method**	**WHEN TO BE APPLIED? IF ...**	**WHEN NOT TO BE APPLIED? IF ...**
Physical model-based	Parity equations State observers	– Failure modes are easy to be understood, and they are well dened through mathematical formulations – A physical model for each failure mode is available – Equipment conditions can be represented statistically – Appropriate data for modeling is available – High accuracy of equipment behavior modeling and monitoring is required – High accuracy of RUL prediction is required	– A physical model is not available – Failure modes are numerous and convoluted – Failure modes are convoluted – Equipment condition is complicated to be described
Knowledge-based	Expert systems	– There is equipment that is simple and well understood – Human experts are available to define the knowledge rules – Equipment conditions are explainable and predictable – Failure modes are well understood and can be characterized simply	– Equipment condition is complex to be described – Human experts are not available to dene knowledge rules – Failure modes are numerous and convoluted – Highly accuracy is required for equipment behavior modeling as well as RUL prediction
Data-driven	Fuzzy logic systems	– Equipment is complicated, and operating conditions are uncertain – Precise input is not available – Data is highly noisy and uncertain – One or more variables are continuous – There is no mathematical model available to be implemented	– No human experts are available to dene fuzzy rules – Input data is discrete and limited to a small number of options – Failure modes are numerous and convoluted – Highly accuracy is required for equipment behavior modeling as well as RUL prediction
Data-driven	Aggregate reliability functions	– Dataset is statistically large – Few numbers of failure modes exist – There is no condition monitoring data of the equipment – RUL is used for maintenance planning	– There is no possibility to differentiate the signicant number of possible failure modes – Past operating conditions of the equipment are not representative of its actual condition
Data-driven	Bayesian Networks	– There is incomplete data – There is multivariate data – Failure causes are known – Modeling experts are available – Accurate prediction of RUL is required – There are non-linear and non-Gaussian noises	– Failure causes are unknown – Modeling experts are not available – Lack of training data – There are linear and Gaussian noises – There are multiplicative noises

Continued on next page

Table 2.3 – *Where (not) to use prognostic and diagnostic approaches (continued).*

Approach	Method	WHEN TO BE APPLIED? IF ...	WHEN NOT TO BE APPLIED? IF ...
Data-driven	Markov Model	– There is incomplete data – There is multivariate data – Failure causes are known – Accurate prediction of RUL is required – No cause-effect analysis is required	– Sufcient data related to failure mode is not available for training – Suitable hardware for computation is not available – The cause-effect analysis is required
Data-driven	Principal Component Analysis (PCA)	– It is not possible to determine physical and statistical models – There is the need to reduce the dimensionality of the data – Failures impact equipment condition by changing the variance structure of the process – The cause-effect analysis is required	– Failures do not impact equipment condition by changing the variance structure of the process – Several complicated and convoluted failure modes exist
Data-driven	Trend Extrapolations	– It is possible to assign a single monitoring parameter to each failure mode – It is possible to summarize equipment condition into a single monitoring parameter – Operating conditions are stable or do not affect monitored parameter – Measurements are repeatable, reliable and not highly sensitive to measurement processes	– It is not possible to assign a single monitoring parameter to each failure mode – It is not possible to summarize equipment condition into a single monitoring parameter – There are varying operating conditions that affect the measured parameter but are not related to failure – The trend is not monotonic
Data-driven	Gaussian Mixture Models (GMM)	– There is non-stationary data with multiple distributions – Equipment processing data consists of a mixture of Gaussian models – Enough data is available to build Gaussian models – There is enough condition monitoring data for RUL prediction	– Data is stationary – Failure modes are highly convoluted – No statistical behavior exist for the equipment condition
Data-driven	Self-Organizing Maps (SOM)	– There is enough data for training – Input data is high-dimensional – It is not possible to determine physical and statistical models – No cause-effect analysis is required – There is enough condition monitoring data for RUL prediction	– The main aim is equipment behavior modeling – There is no information about outputs – No cause-effect analysis is needed – There is not enough data for training

Continued on next page

Table 2.3 – *Where (not) to use prognostic and diagnostic approaches (continued).*

Approach	**Method**	**WHEN TO BE APPLIED? IF ...**	**WHEN NOT TO BE APPLIED? IF ...**
Data-driven	Artificial Neural Networks (ANNs)	– There is a significant amount of noisy, numerical and temporal data – It is not possible to determine physical and statistical models – Precise estimation of RUL prediction is required – No cause-effect analysis is needed – Fault classification is more required – There is enough condition monitoring data for RUL prediction	– Data is complex or nominal – The data structure is simple – There are no temporal inputs – There is not enough data for training – The cause-effect analysis is required
Data-driven	Support Vector Machine (SVM)	– The main aim is fault detection and classification – There is enough information about outputs – It is not possible to determine physical and statistical models – There is enough data for training – Input data are not classifiable, and they can be projected into a higher dimensional space by kernel functions – No cause-effect analysis is required	– The main aim is equipment behavior modeling – There is no information about outputs – No cause-effect analysis is needed – There is not enough data for training

2.4 Advanced Process Control in Semiconductor

Semiconductor manufacturing also referred to as microelectronics manufacturing, processes semiconductor wafers in a fabrication facility or fab through hundreds of steps. These steps focus on features definition in repeated patterns on the wafer which called die [51]. Wafers move through the fab in lots, often of a constant size due to standard containers. The production of wafers is achieved in a multi-step process involving a variety of complex chemical processes such as deposition, photolithography, etching, ion implantation, and photoresist strip [52]. These front-end processes are revisited so that multiple layers are constructed. Once front-end processing is completed, back-end processing uses assembly, test and packaging capabilities to complete the process of converting individual die to chips [51]. The front-end equipment is highly sophisticated with each machine costing $\geq$ \$1M (USD) and containing hundreds of components and thousands of failure points at a minimum [53].

Advanced process control (APC) has become an essential framework in the semiconductor fabrication to keep the product quality and the yield (e.g., $\geq 90\%$) at high level and remain protable in a competitive global environment. Generally, APC enables cost reduction in manufacturing by accelerating the process development, reducing the monitor wafers, decreasing the process variation, increasing the yield, shortening the control loops in case of failures, ensuring product reliability and increasing the equipment utilization. Accordingly, the principal motivation for implementing APC is to improve equipment yield by controlling processes and equipment to reduce process variability and to increase equipment efficiency [54]. For improvement achievement, a wide variety of modeling approaches (i.e., physical model-based, knowledge-based and data-driven) have been developed and used. To maintain the processes at their specification levels and to monitor equipment for existing possible failures, APC makes it possible by preventing unexpected process downtime, increasing tool utilization, and reducing variability [54].

APC can encompass all kinds of equipment and process control systems in semiconductor manufacturing such as Run-to-Run (R2R) control, Virtual Metrology (VM), SPC, and Fault Detection and Classification (FDC) systems [51]. These APC solutions (please see **Figure 1.1** in **Chapter 1**) have become indispensable in all front-end semiconductor manufacturing. R2R is the technique of modifying recipe parameters between production runs to improve processing performance. A run can be a batch, lot, or an individual wafer. In serial processing, this method can just be applied between two measurements. On the other hand, VM is the technique of deriving wafer parameters or product parameters from existing manufacturing parameters or upstream metrology (e.g., process state, additional sensors, temperature, pressure, gas flow, etc.) by using physical models.

As the APC framework has been widely implemented in this industry, the impact of equipment health on the product quality is usually compensated by the R2R regulator. The product quality, namely the metrology, can be maintained at the acceptable level even the equipment gradually deteriorates. To associate equipment health with metrology, R2R regularization will be considered to recover the truthful process drift regarding

fine-tuning the recipes.

These challenges have existed in the industry for decades and are not specific to semiconductor industry; however, they are somewhat unique to semiconductor manufacturing and therefore key in implementing the APC solutions in the semiconductor industry [51], [54]. The first challenge is the equipment and process complexity. As noted earlier, each semiconductor front-end machine is costly and contains hundreds of components and thousands of possible failure points. This complexity makes it impossible to concisely define and model the component interactions as well as the detailed process events. Accordingly, vast data archives are required to cover and characterize all forms of events and interactions.

The second challenge is the process dynamics. The majority of the semiconductor processes are influenced by significant process dynamics that result in drift or shift of process operation [51]. Internal process/equipment factors creating the dynamics could be chamber seasoning over time; for example, a deposition process requires heating the chamber to a temperature high enough to ensure that the reaction can be triggered. If the temperature sensor readings are collected and reviewed after the process, a non-stationary profile can be seen in three steps: warming-up, main deposition, and cooling down. This non-stationarity shown in the temporal sensor readings does not come from the seasonality, periodicity or abnormal drifts. Instead, it is a normal phenomenon encoded in the recipe about the chemical and physical laws. There are also external factors that change the context of operation such as maintenance events and frequent changes in products. The third challenge is data quality. Data quality encompasses diverse issues such as accuracy, completeness, context richness, availability, and archival length [51] where the industry needs to improve data quality to support the APC solutions performance.

As mentioned earlier in **Section 2.1.3**, with the advancement of IT devices and the possibility of collecting a significant volume of data, data-driven approaches have received the highest attention. On the other hand, these challenges remark that the modeling approaches for the semiconductor manufacturing cannot be limited to data-driven; however data-driven approached are still the most efficient techniques to cope with APC challenges. Therefore, in the following sections, the three fundamental aspects of data-driven approaches: data collection, data processing, and feature extraction, will be linked to the aforementioned APC challenges. For example, data collection relates to data quality challenge and contains which data to be collected. In addition, data processing and feature extraction correspond to the complexity and process dynamics challenges, wherein the ways to process the data and to select correct features significantly impact the data-driven approaches performance.

2.4.1 Data Collection

As explained in **Section 2.1.3**, CBM data can be identified based on their statistical value (i.e., nominal, categorical and ordinal) or their application (i.e., event or condition monitoring data). In this section, we focus on condition monitoring data that are the core of data collection for APC solutions [19, 55]. Condition monitoring data can be

collected from the product (so-called metrology data such as wafer thickness, surface quality, etc.), from the process (e.g., pressure, temperature, radio frequency, etc.) or the equipment (e.g., vibration, acoustic, etc.). Increasingly in the industry and recently in semiconductor manufacturing, condition monitoring data are easily collected thanks to the rapid development of computer and advanced sensor technologies [51].

As mentioned in **Section 2.1.3**, two main types of collected data in a CBM program are event data and condition monitoring data. Also, two primary sources for collecting condition monitoring data are R2R and FDC systems [56]. Run-to-Run (R2R) system is a form of discrete process and equipment control in which the product recipe concerning a particular equipment process is modified at each run for minimizing process drift, shift, and variability [56]. Practically speaking, the inputs and outputs of each process run are compared by the R2R controllers at the end of each run. By this comparison, the controller adjusts the process variables (e.g., temperature, pressure, etc.) to avoid nonconformists in the product at the end of the next run. By repeating this process in between each run, one can minimize process drift immediately and prevent nonconformity propagation through the overall process.

On the other hand, the FDC system monitors the real-time performance of a tool to ensure that its performance does not result in mis-processing. Incorporating FDC into the production system improves overall equipment efficiency by monitoring equipment health and detecting emerging faults. Despite R2R control systems that actively control process variables at the end of each run, an FDC system monitors variables during the operation of a process to ensure that its performance is as expected. Such monitoring enables detection of operational faults in real time and facilitates interruption of the process before the equipment fails. In the semiconductor industry, the equipment determines the availability of the necessary status variables called the SVIDs (Status Variable IDentification). These SVIDs are required to collect the data related to the process such as the pressure, temperature, gas flow and radio frequency information during the processing of a wafer lot. Accordingly, the collected FDC data is expressed as a data cube in three dimensions: wafer, SVIDs and Process observations. **Figure 2.3** depicts the FDC collection system for an etching process in the semiconductor industry.

2.4.2 Data Pre-processing

As mentioned in **Section 2.1.3**, two sub-sets of data processing are data cleaning and data manipulation. Although data cleaning is critical and it is one of the significant challenges in APC (i.e., data quality challenge), it is beyond the scope of this thesis and will not be discussed in detail here. It was elaborated so that condition monitoring data can differ from value type data, time series data and multi-dimensional data. This section only deals with value type and time series'condition monitoring data. A major challenge in data processing is to select the best way of treating and manipulating that provide the data with adequate information. Afterward, manipulated data with an appropriate approach make it possible to extract features and build effective prognostic and diagnostic models correctly. One issue in value type data is the complex correlation structure when the number of variables is enormous. For coping with this issue and

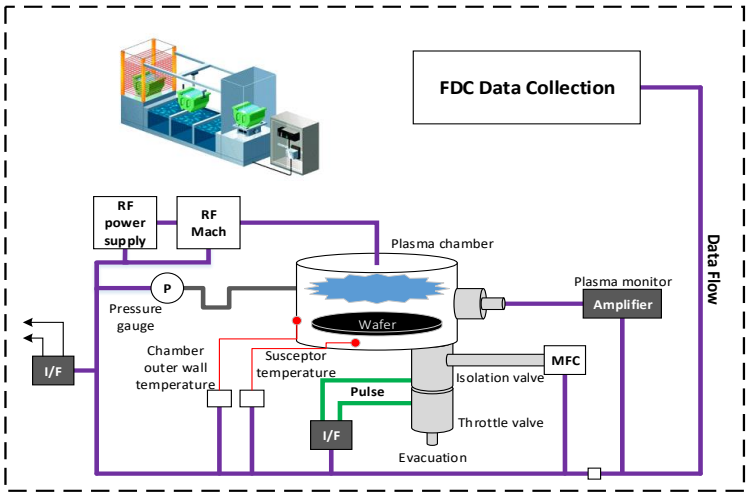


Figure 2.3: Research methodology of the proposed data-driven approach is illustrated.

preparing data for feature extraction, multivariate analysis techniques such as PCA and Independent Component Analysis (ICA) have been widely utilizing to handle data with complex correlation structure. These approaches are highly used to reduce the dimension of data when the number of variables is enormous. An example of applying dimension reduction techniques in equipment failure diagnostics is given in [15].

Data analyses can be directly conducted on the original temporal data (raw data) without any manipulation and transformation, or the data can be first decomposed based on different frequencies then the analyses are performed on the new components. The first analyses can be called as original-domain analysis, and the latter exists as time-frequency analysis.

In the original-domain analysis, the data value is known for all real numbers, for the case of continuous time, or at various separate instants in the case of discrete time. It is noteworthy that original-domain techniques such as multivariate analyses are appropriate only for stationary data and not for the data which have process dynamics and non-stationary profiles. Another way of analyzing the stationary data can be transforming the original temporal data into frequency domain then extracting required features from frequency components of the signal. Its advantage over original-domain analysis is the ability to identify and isolate specific frequency components of interest [39]. The most known technique for this transformation is Fast Fourier Transform (FFT). The main idea of FFT is to either look at the whole spectrum or look closely at specific frequency components of the signal and thus extract features from the frequency component [14]. The drawbacks of these analyses are their inability to handle non-stationary signals, which are very common when equipment failures occur, particularly in semiconductor manufacturing.

For processing the non-stationary temporal data, in addition to exploring the original domain, there also exist the time-frequency analyses, which are also known as signal processing. Signal processing techniques are employed to prepare the data for better analysis and interpretation. Thus, time-frequency analysis, which investigates waveform signals in both time and frequency domain, has been developed for non-stationary waveform signals. Two primary and well-known techniques for transferring the temporal signals into the time-frequency domain are Short-Time Fourier Transform (STFT) and wavelet transformation. The idea of STFT is to decompose the waveform signal into segments with short-time window and then apply Fourier transform to each segment.

Wavelet transformation has been rapidly developed in the past decade and has broad application. Unlike STFT, which provides a time-frequency representation of a signal, wavelet transform provides a time-scale representation of a signal [39]. Wavelet transformation decomposes the signal into a series of oscillatory functions with different frequencies at a different time. The wavelet transformation can produce a high time resolution at low frequencies and a high time resolution at high frequencies. It can also be utilized for de-noising of raw signals.

Semiconductor data are among the most complete and complicated data that contains product mixes, process dynamics, convoluted failure modes and highly correlated signals. Accordingly, processing semiconductor data requires appropriate and effective analyzing techniques.

2.4.3 Feature Extraction

The principles of condition monitoring in any area are 1) the identication of a set of features within the equipment sensor readings that are indicative of the equipment behavior and 2) the utilization of those features with an appropriate approach to prognose the equipment health and diagnose the equipment failures. Feature extraction methods vary based on the data type (e.g., continuous, discrete, nominal, and categorical) and generate different results according to the data mining techniques. In this section, we investigate the features extraction based on the type of data processing as original-domain features, frequency-domain features, and time-frequency domain features.

Different types of the original-domain feature can be directly extracted from the data such as statistical features, upper and lower bound of the signal, autoregressive coefcients, etc. These features can be extracted only for one-dimensional signals. In the following, only statistical features are explained since they are more common in the semiconductor industry. Other techniques are used more over vibration signals of mechanical equipment.

Statistical features from the original temporal data mostly include mean, root mean square (RMS), standard deviation and variance that have usually been used in past studies to identify the differences between signals from different working periods of the equipment. Such simple features are more suitable for stationary signals. For non-stationary signals, more advanced statistical features such as skewness and kurtosis can be applied. Skewness is used to measure whether the signal is negatively or positively skewed, while kurtosis measures the taillessness of the Probability Distribution Function

(PDF) of a real-valued random variable. In a similar way to the concept of skewness, kurtosis is a descriptor of the shape of a probability distribution and indicates if the signal is an impulse in nature.

Accordingly, for a signal with a normal distribution, signal has a skewness value of zero. Statistical features supervise the PDF of the signal [55]. It is a well-known fact that any change of the equipment behavior implicitly affects the PDF of the corresponding signals. Besides those mentioned statistical features on the original temporal data, shape factor is a non-dimensional feature that is affected by an object's shape but is independent of its dimensions [55]. This feature is affected by the change in the mean and RMS. **Table 2.4** summarizes the most common statistical features for the original temporal signals. In **Table 2.4**, variance, skewness, and kurtosis have become dimensionless features after being divided by a power of standard deviation (σ).

Frequency-domain feature extraction requires the transformation of the temporal signals initially into frequency-domain signals using FFT. Although FFT works well for stationary periodic signals, it is less useful for non-stationary signals that arise from time-dependent events. It is a fact that any fault and abnormal behavior in the equipment contribute to the normal signal as a component with the new frequency. Accordingly, distinguishing newly created frequencies in the signal and their progression through the time may help to prognose equipment behavior or diagnose equipment failures. Frequency-domain features can be extracted in different ways, and the most common features are statistical features e.g., Frequency Center (FC), Root Mean Square Frequency (RMSF) and Root Variance Frequency (RVF) [55]. The main fact in these features is when equipment behavior changes or when equipment fails, the frequency element changes, and the values of the FC, RMSF, and RVF also change.

Furthermore, Spectral Skewness (SS), Spectral Kurtosis (SK), spectral entropy and Shannon entropy are other advanced statistical measures applied to the magnitude spectrum [55]. The SS measures the symmetry of the distribution of the spectral magnitude values around its mean, while the SK measures the distribution of the spectral magnitude values and compare to a Gaussian distribution. Finally, FC, RVF, SS, and SK can be dened as the rst-order to fourth-order moments of the Fourier spectrum, respectively [57]. **Table 2.5** summarizes the statistical frequency-domain features.

The most well-known methods to transform original temporal signals into time-frequency domain are short-time Fourier transform (STFT) and wavelet transformation that implement a mapping of one-dimensional temporal signals to a two-dimensional function of time and frequency. These methods are commonly used for the non-stationary signal. The problem with STFT is that it provides constant resolution for all frequencies since it uses the same window for the analysis of the entire signal. Accordingly, STFT is not suitable for non-stationary signals. Despite STFT, the wavelet transformation method employs wavelet function as the basis function [25]. Accordingly, wavelet transformation can provide an excellent energy concentration properties and can decompose the signal with a certain number of coefcients. These coefficients, Approximation coefcient and Detail coefficient, are considered as new signals, and different features can be extracted for each of them. They are obtained by passing the original temporal signal

through the low-pass and high-pass filters, respectively. Finally, all the features of **Table 2.4** can also be extracted from both Approximation and Detail signals.

Table 2.4: Statistical original-domain features. $x_i : i = 1, ..., N$ is a point of the signal.

Feature	**Description**	**Formulation**
Peak value (P)	Peak value shows the maximum normal value of the signal	$\max_{i=1,...,N} x_i$
Mean (μ)	Mean is the expected value of a signal to measure the central tendency either of a probability distribution or the random variable characterized by that distribution	$\frac{\sum_i x_i}{N}$
Variance (Var)	Variance measures the dispersion of a signal around their reference mean value	$\frac{\sum_i (x_i - \mu)^2}{(N-1)\sigma^2}$
RMS	RMS starts to be increased as abnormality initiated in the signal and its value increases gradually as the fault is developing	$\sqrt{\frac{\sum_i x_i^2}{N}}$
Skewness (Sk)	Skewness is used to measure the asymmetry behavior of a signal through its PDF	$\frac{\sum_i (x_i - \mu)^3}{(N-1)\sigma^3}$
Kurtosis (Ku)	Kurtosis quanties the peak value of the PDF	$\frac{\sum_i (x_i - \mu)^4}{(N-1)\sigma^4}$
Crest indicator (CI)	It is defined as the ratio of the maximum positive peak value of the signal	$\frac{P}{RMS}$
Autocorrelation (r_k)	Measures the correlation (r_k) between x_i and x_{i+k}, where $k = 0, ..., N$ is the time lag	$\frac{\frac{\sum_{i=1}^{N-k}(x_i-\mu)(x_{i+k}-\mu)}{N}}{\sigma^2}$
Linear correlation ($Corr$)	Measures the strength of a linear association between two signals x and y.	$\frac{\sum_i (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_i (x_i - \mu_x)^2 \sum_i (y_i - \mu_y)^2}}$

Table 2.5: Statistical frequency-domain features. $x_i : i = 1, ..., N$ is a point of the signal.

Feature	**Description**	**Formulation**
Frequency Center (FC)	FC is the expected value of frequency elements to measure the position changes of main frequencies ($x_i^{'} = x_{i+1} - x_i$)	$\frac{\sum_{i=2}^{N} x_i^{'} x_i}{2\pi \sum_{i=1}^{N} x_i^2}$
Root Mean Square Frequency ($RMSF$)	RMSF starts to be increased as abnormality initiated in the signal. It is to measure the position changes of main frequencies	$\sqrt{\frac{\sum_{i=2}^{N} (x_i^{'})^2}{4\pi^2 \sum_{i=1}^{N} x_i^2}}$
Root Variance Frequency (RVF)	RVF shows the convergence of the power spectrum	$\sqrt{\frac{\sum_{i=2}^{N} (x_i^{'})^2}{4\pi^2 \sum_{i=1}^{N} x_i^2} - FC^2}$
Spectral Skewness ($SS(n)$)	SS measures the symmetry of the distribution of the nth spectral magnitude values around its mean	$\frac{2\sum_{i=0}^{N/2-1} (\lvert X(i,n)\rvert - \mu_{\lvert X\rvert})^3}{N\sigma_{\lvert X\rvert}^3}$
Spectral Kurtosis ($SK(n)$)	SK measures the distribution of the spectral magnitude values and compares to a Gaussian distribution. $X(i,n)$ is the ith point of the nth spectrum	$\frac{2\sum_{i=0}^{N/2-1} (\lvert X(i,n)\rvert - \mu_{\lvert X\rvert})^4}{N\sigma_{\lvert X\rvert}^4} - 3$
Spectral entropy ($S(E)$)	Spectral Entropy describes the complexity of a system. p_i is the probability of point i appearing in the spectrum	$-\sum_{i=1}^{N} p_i \ln p_i$
Shannon entropy ($H(A)$)	Spectral Entropy describes the complexity of a system	$-\sum_{i=1}^{N} p_i \log_2 p_i$

2.5 Literature Review

This section reviews the relevant literature by dividing the papers into two main categories as equipment behavior prognosis and equipment failure diagnosis. This section discusses the researches done in the semiconductor industry (since it is the case study of this thesis) alongside the chemical industry since they share the same FDC data structure and production features. A few papers might also be reviewed from other industry with different data if they provide relevant insight into equipment behavior prognosis and equipment failure diagnosis.

2.5.1 Literature Review: Equipment Behavior Prognosis

This section surveys the papers studied in the semiconductor and chemical industries that have researched on equipment behavior prognosis. Chen et al. [1] developed a data-driven approach by using the multivariate process capability index to integrate the equipment multiple parameters into an overall equipment health index. To do that, they first applied some multivariate statistical methods, e.g., PCA and time series analysis to do original-domain data preprocessing. They proposed then a Machine Capability Index (MCI) using preprocessed equipment data based on the idea of multivariate process capability Indices (PCIs), which was manipulated for equipment capability evaluation. This index considered the covariance and mean matrix of observation vectors as well as the incorporation of subject matter expertise to define machine condition. Afterward, they established a prognosis model based on an aging Markovian deterioration model to describe the wave-moving probability distribution of machine condition. The authors supposed that the machine likely enters into less healthy states as it grows older. Despite the semi-Markovian model in the literature, they introduced an aging factor that discounted the transitions'probabilities to more robust states while increasing the transitions'probabilities to less healthy states. Finally, dynamic PM policy under Markovian aging deterioration which was rooted in a stochastic health model was developed.

Chao et al. [58] proposed a data-driven model for batch process monitoring in the semiconductor industry. The initial idea came from Multi-way Principal Component Analysis (MPCA) knowing that it is often impossible to take benefit from pure MPCA due to the increase in the number of variables. They introduced a set of step-trend variables such that the mean shifts of each sensor variables from the reference profile are calculated in each step. The reason is in the semiconductor industry the processes often consist of ramp up-down steps that result in variables with sharp and significant changes. The considerable variation can be observed during the initial periods of the process. On the other hand, the steady difference states between batches impose another interpretation challenge. Afterward, they calculate the score of a first principal component obtained from MPCA to monitor equipment deterioration. Eventually, they estimated the Hotelling T^2 to detect the batch faults. Finally, they concluded that weighted PCA of step-trend variables could be used to monitor equipment behavior, and Hotelling T^2 and the cumulative probability of the residual variables to detect potential defective batch.

Tew et al. [46] proposed an original-domain based analysis to fault diagnostics and prognostics that aims to facilitate equipment health management in a manufacturing environment. The data set is vibration data collected for each bearing type and machine operating speed. Each dataset is sub-divided into the training set, containing 80% of the samples, and the testing set, which includes the remaining 20% of the samples. The extracted features were the peak value; root means square value, standard deviation and kurtosis value of the vibration signal. Afterward, they combined the extracted original-domain features into a feature set associated with each segment of the vibration data using the ColumnModifier module which is functions to obtain and merge columns from dierent files/databases. Therefore four extracted features combined into a single data table in which each row contained the four features computed from the corresponding segment of the vibration signal. Then, they applied three classifiers as Multi-Layer Perceptrons (MLP), k-Nearest Neighbors (kNN) and SVM, wherein the bearings with inner race fault, outer race fault, ball fault, and combination fault are grouped into one superset of faulty bearings. The Class Labeler modules assigned a "0" label onto each feature set for good bearing data, and a "1" label onto each feature set for the faulty bearing data. The authors finally captured the characteristics of the machine states with several data-driven approaches such as SVM and MLP which outperformed kNN classifier. It is worthy to mention that their data set was purely high qualified under controlled manufacturing environment, so it would be a challenge to apply this model on the data set contains noises, missing or even non-meaningful values.

In another paper [59], the central assumption and explanation are based on the various variations at different sensor profile which are mainly due to on-off recipe actions at specific points. The authors proposed a PCA-based data-driven method to address the long-term and short-term effects of tool aging and first-wafer in a lot cycle, respectively. To do that, they firstly determined a fixed-reference original-domain profile for each sensor variable to describe the on-off actions and a general similarity in the patterns due to recipe actions. They then established the level shifts of these patterns in each step to capture and to remove natural within-profile variations due to long-term aging trends and the short-term first-wafer effects. Despite the recipe dependency of this model, the detection of real faults is facilitated by these systematic normal phase separation. In the end, they formulated the residuals of this model to have a health index for each wafer by using standard multivariate statistical analysis.

Krueger et al. [60] proposed semi-physical model-based and knowledge-based methodologies about data collection, integration, and aggregation which is used in a proposed generalized linear model to predict semiconductor yield based on original-domain defect metrology data. They applied this technique to both die and wafer levels, and compared the prediction errors and significant process improvement factor, with existing models in the literature. They discussed the nested or hierarchical structure to identify the primary outlier sources, which means in multi-factor experiments, the levels of one factor are similar but not identical for different levels of another factor. They concluded that the nested structure could not be used at the wafer level due to equipment and process monitoring time consuming, while it can be used at die-level technique and provided

more detailed information about significant predictors, including specific wafers and die. For complementary, they proposed the nested die-level logistic regression models to show the predictive power at both the die and wafer levels. Thieullen et al. [61] surveyed several papers up to 2011 that study data-driven approaches for two issues of Prognostic and Health Management (PHM) methodologies with applications focused on semiconductor manufacturing process: the development of indicators for health assessment, and prognostic purposes.

As another data-driven approach using original temporal data, Yang and Lee [62] proposed Bayesian Belief Network (BBN) application on Chemical Vapor Deposition (CVD) tool to investigate the causal relationship among process variables on the semiconductor tools and to evaluate their influence on the wafer quality. The goal was fault diagnosis and prognosis in semiconductor manufacturing. By statistical inference on BBN models at different periods of the process, this relation and their influence on wafer indicated by the network structure and by the conditional probabilities in the model. They suggested by this model that one can diagnose causes when the bad wafer is produced or can predict the wafer quality when abnormal is observed during the process. In this study, the two main challenges which should be mentioned are first to have a normal process drift or shift model as a baseline and secondly to evaluate all the obtained interactions between the sensors. Accordingly, a process engineer expertise is required.

Bleakie et al. [63] presented a data-driven integrated feature extraction, equipment monitoring, and fault modeling approach applied to temporal sensor data from Plasma Enhanced Chemical Vapor Deposition (PECVD) tool. Firstly they extracted forty features from multiple sensor readings, including dynamic features like rise-time, overshoot, and steady-state values, along with statistical features, such as mean value, variance, and range. Afterward, they standardized features to eliminate the physical units and make them dimensionally homogeneous. To standardize, they subtracted feature mean and divided it by its standard deviation, where the mean and standard deviation were calculated from the normal data set. Then they analyzed the sensitivity of the feature to various equipment conditions from normal to faulty situations. For this aim, they applied Linear Discriminant Analysis (LDA) for finding the most sensitive features between two classes of data. They quantified the changes in the most sensitive features by tracking the overlap volume between probability density function (PDF) of the normal feature and current system behavior one. The feature PDFs were approximated using GMMs.

Nguyen et al. [64, 8] presented a data-driven approach to health index extraction for discrete manufacturing processes tools based on the deterioration reconstruction. They used the temporal FDC data and monitored the trend of significant SVIDs which carry the equipment deterioration information. Afterward, the authors identified the critical points of degraded SVIDs based on an optimization algorithm with variance maximization objective. The crucial aspect of a degraded sensor is the observation interval at which the variance is maximum. They used then Hybrid-wise Multiway PCA (E-HMPCA) and the index of Squared Prediction Error (SPE) to perform deterioration detection and diagnosis for the batch process machine. The process was considered

reliable if SPE is under its upper control limit. Then, they modeled the HI with a Gamma process. For the online supervision, they also estimated the probability density function of RUL for each inspection time which was supported by an off-line analysis. An application of the proposed method in a real industrial case showed a small error of RUL estimation for the online supervision. A further improvement of the proposed method is necessary to overcome the influences of local fluctuations of the HI in some particular situations.

Yu et al. [65] proposed a data-driven approach based on hierarchical indices for real-time equipment monitoring using the relative importance of the SVIDs, steps, and observation. They started by data preparation in original-domain and to tackle with missing data, observations were considered individually, and the results were then combined to develop the equipment indices. Instead of using p-values as equipment indices, they thought "1" and "0" to represent the changing/unchanging of an observation trend, respectively. For defined outlier zone, authors approximated it by the Gaussian distribution assumption. Afterward, they developed hierarchical indices using the weighted sums in which the weights, were provided by engineers, show the relative importance of each observation. They finally located the causes when equipment indices decreased.

In a similar work, Nguyen et al. [66] presented three data-driven methods for equipment behavior prognosis by extracting the HI of Plasma Enhanced Chemical Vapor Deposition (PECVD) equipment. The first method is Degradation Reconstruction Combined with PCA (DR-PCA), the second method is Degradation Reconstruction Combined with EWMA-Hybridwise Multiway PCA (DR-E-HMPCA), and the last one is Significant Points Combined PCA (SP-PCA). After applying on the simulated and real FDC data, they concluded the SP-PCA method gives the most reliable Health Index, though it is time-consuming. In the real time since the process condition is changing, one should consider updating this methodology to online prognosis.

As an extension of [66], Nguyen et al. [8] studied the equipment behavior prognosis in batch manufacturing processes (BMP) that plays a vital role in many production industries, such as in semiconductor, electronic and pharmaceutical industries. They exhibit some batch-to-batch or unit-to-unit variations due to many reasons such as variations in impurities and deviations of the process variables from their trajectories. The authors proposed a data-driven prognostic approach for the batch process on chemical vapor deposition machine. For this aim, they firstly reduced the data size to extract a raw health index which represents the operating state of the system. Next, they calculated the first principle component of the data set by PCA and later identified the sensor-observation points which mean the feature that carries the deterioration information based on the correlation value between the first principle component and each column of data set matrix. Finally, they modeled the deterioration by gamma process and then extracted remaining useful life estimation with using probability density function with a confidence interval. The proposed method was applied to semiconductor manufacturing equipment with two industrial data sets provided by STMicroelectronics.

Wang et al. [67] proposed a data-driven fault prognostic method on a simulated fabrication data set in semiconductor industry based on Bayesian networks to represent

the relationships between the current process status and the ongoing process fault for both continuous and discrete temporal process variables. In this modified Bayesian network for each fabrication process, continuous variables are for process monitoring, and the discrete variables illustrate where there is a fault in the process or not. This paper used the joint probability distribution property of Bayesian network which is represented by the product of all variables'conditional probability distributions, to predict whether there will be a fault in the next step. To specify which sensor caused the defect, they calculated the divergence of the sensor value from the normal mean value. In the end, they stacked up to the network of each process together, to have a complete model for whole fabrication processes.

Rostami et al. [37, 68] proposed a data-driven approach employing time-frequency analysis. Their approach develops a healthy state model for equipment deterioration monitoring using Wavelet Packet Decomposition (WPD) concept on FDC data in semiconductor manufacturing. They decomposed the SVID readings into approximation and detail components obtained from WPD and calculated the energies of each component at a given level decomposition. They introduced then these energies as features which show the equipment deterioration trend. To track the energy change over process run, they proposed to track their PDF. Accordingly, they used GMM to construct the PDFs and followed the evolution of feature distributions. It should be mentioned that they defined a normal GMM baseline which is constructed upon normal wafers or post-PM wafers. To quantify differences between normal GMM and current GMMs, they calculated the distance between their PDFs, iteratively, by Hellinger distance which was modified and relaxed by the authors. The challenge is how to set the size of normal GMM to compare fairly with current GMMs since the size of the normal baseline is the function of the data set size.

In a similar work, Rostami et al. [68] modeled the equipment deterioration with the complex FDC data in semiconductor manufacturing. They proposed applying WPD to decompose the temporal SVID readings into the Approximation and the Detail signals over the time-frequency domain. They explained that since faults contribute to the sensor readings with high-frequency signals and affect the energy of the signals, sensor decomposition help to extract the deterioration-related features. Then they calculated the energy values of the decomposed signals as deterioration-related features. They emphasized that the Approximation and Detail features carry information regarding the long-term patterns, e.g., process dynamics, and short-term changes, e.g., sudden ups and downs, respectively. To prognose equipment deterioration and to diagnose the potential failures, they tracked the PDF of the features over time. To reach this goal, they developed an iterative windowing approach to monitoring the evolution of the PDF over different sets of wafers for the energies of Approximation signals and the energies of Detail signals, separately. Afterward, they again used GMM to estimate the PDF of the energy features. The drift of GMM from normal basis window to the current window was characterized by calculating the Hellinger distance, which is an f-divergence-based function. Finally, they concluded that the Approximation and Detail signals could be used separately for failure diagnosis and deterioration prognosis, respectively.

Jia et al. [69] proposed a data-driven Diffusion Map based methodology for an online machine deterioration assessment, abnormality detection, and diagnosis. The deterioration assessment uses a baseline, which represents the healthy run, and the deviation from this baseline is measured as deterioration. They also proposed a difference based methodology for nonlinear dimension reduction. The authors selected diffusion map because of its better performance and more robust result on the real-world data set. They elaborated the proposed method on etching process fault detection. Finally, the comparison of the proposed method's result with PCA-based methods for data reduction and with a SOM-based method for equipment health monitoring illustrated the outperforming of the proposed approach. Although this methodology has been proposed for machine health monitoring, fault prognosis as a critical part of this domain is missed.

As a novel data-driven approach, Rostami et al. [70] utilized FDC data to develop an equipment deterioration model over the time and to identify the caused SVIDs. They depicted that deterioration touches the sensor readings in two ways: long-term level changes and short-term fluctuations which former appears in the signal as low-frequency components while later contributes to the signal as high-frequency components. They distinguished these impacts using WPD and decomposed signals into low frequency (Approximation) and high frequency (Detail) components. They calculated then the determinant of the correlation between Approximation signals in each wafer and repeated it for all wafers, to show long-term level changes causes into deterioration pattern. They afterward, identified contributing SVIDs on deterioration causes, because the long-term level change appeared only in some of SVIDs, and not between all SVIDs. For this aim, they proposed a greedy algorithm to pick out a group of SVIDs expected to be the deterioration causes. **Table 2.6** summarizes the literature review on equipment behavior prognostic papers.

Figure 2.4a to **Figure 2.4d** are depicted based on **Table 2.6** and visually show how equipment behavior prognosis problems have been studied in each of semiconductor and chemical industries. **Figure 2.4a** illustrates that the equipment behavior prognosis has received more attention in the semiconductor industry comparing to the chemical industry. **Figure 2.4b** shows that the majority of the studies have applied data-driven approaches to prognose the equipment behavior; while the knowledge-based approaches have the least contribution. Such a minimum percentage for knowledge-based techniques reveals that semiconductor/chemical industries are complicated and experts cannot provide enough knowledge to prognose the equipment behavior. The popularity of the data-driven approaches lies in their efficiency and applicability comparing to the other modeling approaches. **Figure 2.4c** illustrates that the majority of the reviewed papers have used FDC data for equipment behavior prognosis. Finally, it can be observed in **Figure 2.4d** that most of the existing approaches in the equipment behavior prognosis have processed the data in the original domain and a few papers also in the time-frequency domain.

Table 2.6: Literature: Equipment Behavior Prognosis

Reference No.	Data	Industry		Modeling			Data			Processing	
		Semiconductor	Chemical	Physical model-based	Knowledge-based	Data-driven	FDC	Metrology	R2R	Original Domain	Time-Frequency Domain
[71]	2004	✓			✓		✓	✓		✓	
[1]	2007	✓				✓	✓			✓	
[58]	2008	✓				✓	✓			✓	
[46]	2011	✓				✓	✓			✓	
[59]	2011	✓				✓	✓			✓	
[60]	2011	✓		✓		✓		✓		✓	
[6]	2011		✓	✓	✓		✓			✓	
[62]	2012	✓				✓	✓			✓	
[63]	2013	✓				✓	✓			✓	
[72]	2013	✓			✓	✓		✓		✓	
[65]	2014	✓				✓	✓			✓	
[66]	2015	✓				✓	✓			✓	
[8]	2016	✓				✓	✓			✓	
[67]	2017	✓				✓	✓			✓	
[37]	2017	✓				✓	✓				✓
[68]	2017	✓				✓	✓				✓
[70]	2018	✓				✓	✓				✓
[69]	2018	✓				✓	✓			✓	
Total Papers		**17**	**1**	**2**	**3**	**16**	**16**	**3**	**0**	**15**	**3**

2.5.2 Literature Review: Equipment Failure Diagnosis

This section surveys the relevant papers studied in the semiconductor and chemical industries that have researched on equipment failure diagnosis.

Chen and Liu [24] intended to develop a data-driven approach based on a neural network architecture named Adaptive Resonance Theory Network 1 (ART1), which recognizes spatial defect patterns to aid in the diagnosis of failure causes. The authors conducted the proposed methods on original temporal data obtained from a semiconductor manufacturing company in Taiwan and to evaluate the training performance of the ART1 network; it was compared with SOM. By the result, they showed that for the new patterns, ART1 converged much faster than SOM regarding data training. They also concluded that thought this approach provided the automated classification of known patterns and detection of new unknown patterns as well; it could not classify the patterns of defects. Furthermore, due to limited capability for die-level in ART1 network, it could organize the pattern of the defective die.

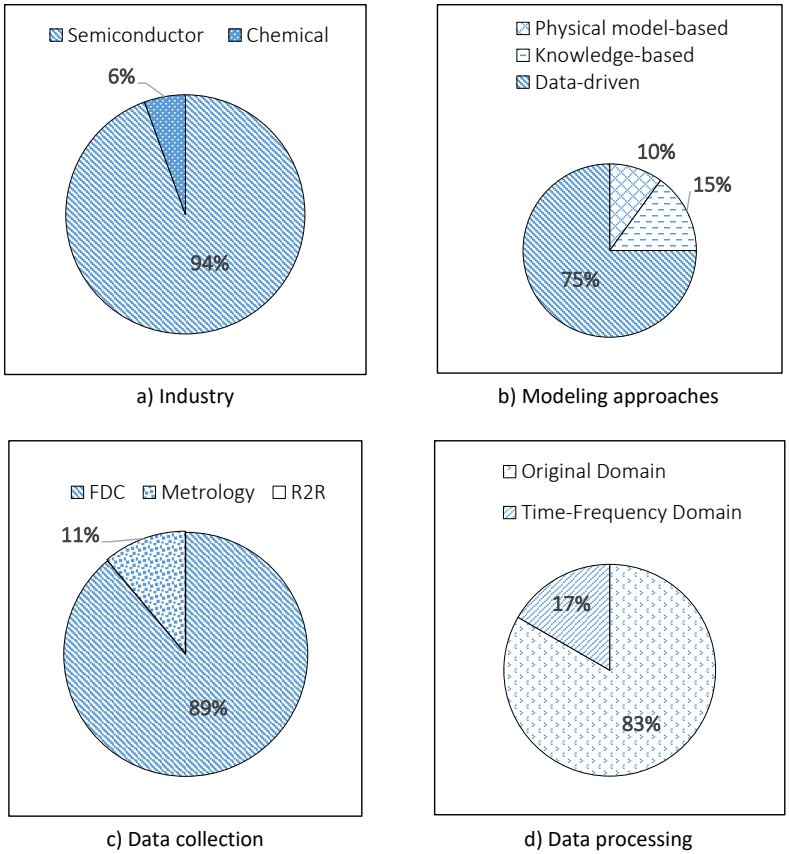


Figure 2.4: Literature analysis: Equipment Behavior Prognosis.

Yue et al. [73] introduced the batch process monitoring to semiconductor fabrication for plasma etchers by using emission spectra and the data-driven Multi-way PCA (MPCA) method to analyze multiple scan sensitivity within a wafer for several typical faults. They tested two MPCA schemes for fault detection and wavelength selection. The proposed method selected critical wavelengths to detect the failure. However, due to the chosen wavelength constraint and ignorance in the normal drifts, the proposed method is not guaranteed to be robust.

Su et al. [74] proposed a data-driven neural network approach for semiconductor wafer post-sawing inspection by introducing three types of neural networks as back-propagation, Radial Basis Function (RBF) network, and learning vector quantization. Regarding the final result, they compared the proposed approach with two other inspection methods, visual inspection, and feature extraction inspection. Finally, it was illustrated that the inspection time by the proposed approach is less than one second per die.

Spitzlsperger et al. [75] presented a data-driven adaptive Hoteling's T^2 control chart for the semiconductor etching process. In this paper, the authors emphasized the insufficiency of Hotelling process control charts for monitoring the status of the oxide etching

process due to the Hotelling statistic movement since it is unstable for the period after the maintenance. The proposed method updates only the univariate that is given to drift, but the problem has not been completely resolved.

Gondra [76] developed a variety of data-driven prognostic models based on machine learning techniques such as ANN and SVM to observe the correlation between process metrics and fault-proneness to estimate the process faults. The author performed a comparative experimental study of the effectiveness of ANN and SVM when applied to the problem of classifying modules as faulty or fault-free on a data set obtained from NASA Metrics Data Program data repository. The experimental results confirmed the superior performance of SVMs over ANNs when viewing fault-proneness prediction as a binary classification task.

Chen et al. [77] proposed a data-driven approach that determines a recipe-independent health indicator based on the Generalized Moving Variance (GMV) by consolidating the large number of SVIDs into a single tool health indicator. To do, they observed the distribution of the tool parameters readings. They proposed then using moving variance/covariance, which was calculated by a small number of consecutive observations in a moving time window. Generalized variance, the determinant of the covariance matrix, is in effect proportional to the volume of data distributed in the multidimensional variable space. They assumed that there should be a regular size of data distribution under a normal process run and therefore the distribution of SVIDs is a function of tool's health. The proposed estimation was model-freed. With this tool health indicator, they employed the exponentially weighted moving average (EWMA) control scheme to detect abnormal health. After that, they developed a two-step diagnosis method to find the causes of the health abnormalities. The first step included the decomposition of the generalized variance into the variance and the covariance. The second step involved the detection of the anomalies of the SVID variability and/or relationships. They tested the method with actual plasma-enhanced chemical vapor deposition (PECVD) and physical vapor deposition (PVD) production data.

Li et al. [48] developed a new data-driven fault diagnosis approach based on Total Projection to Latent Structures (T-PLS). They used four kinds of monitoring statistics in T-PLS and offered a new definition of variable contributions of the T^2 statistics in PLS. Finally, the authors derived the variable contributions of all statistics to identify the faults.

Yu [78] proposed Principal Components (PCs)-based GMM in semiconductor manufacturing to detect the faults. The author used PCs generated by PCA as inputs of GMM to estimate the PDF of semiconductor manufacturing process observations, which can handle complex data with nonlinearity or multi-modal features. The author also proposed two quantification indexes, i.e., negative Log likelihood probability and Mahalanobis distance for assessing process states and then used a Bayesian inference-based calculation method to provide the process failure probability.

Hung et Chen [79] proposed data-driven inference procedures to detect and classify the faults by monitoring the change of the covariance matrix of FDC temporal data in semiconductor manufacturing. They applied Bartlett and Cholesky decomposition

theories to test the evolution of the covariance matrix pattern. These theories aimed to overcome the type-I error challenges. They, therefore, offered two test methods with specific null distributions that overcome the substantial sample requirement difficulties in conventional covariance tests. They compared the proposed method with conventional univariate SPC method and proof its outperforming. It is noteworthy that their proposed classification rules can be combined with other fault detection methods like likelihood ratio test (LRT), in which once the faults are detected, the proposed classification can be applied to identify the sources of the failure.

Chien et al. [80] combined data-driven approaches such as MPCA and SOM to construct the model to detect faults and to derive the rules for fault classification. In their method, the 3-dimensional data was unfolded and projected onto the score and residual spaces to reduce the high data dimension to a few PCs. While D (Hotelling's T^2) and Q (residual) statistics were used for abnormal events detection, different types of process defects were detected based on the extracted control limits. It is noteworthy that using the statistical control charts on the historical data cannot grantee that the faults are distinguished correctly as the detected drifts might be from the normal variation, such as the process change. They then clustered the Out-Of-Control (OOC) observations using the SOM algorithm. However, clustering the observations based on their coordinates may lead to incorrect fault classification. Furthermore, their approach cannot distinguish the normal variation from the recipe changes, and these variations might be misunderstood as the fault when clustering the observations based on their coordinates.

Blue et al. [81, 82] developed a data-driven tool condition hierarchy monitoring approach based on general moving variance and multivariate methodologies to detect the tool faults in semiconductor manufacturing. Firstly, they classified the SVIDs into similar variations groups. In this way, they calculated the Coefficients of Variation (CV) for all the process steps through all SVIDs. Afterward, they stacked up to the CV matrices for each level and got an overall matrix. In the hierarchical agglomerative clustering method, they calculated the absolute value of the correlation coefficient between two SVIDs from whole CV matrix. This value resulted in the correlation-based similarity matrix. The agglomerative procedure was continued to generate potential SVID grouping schemes regarding a dendrogram. They then applied the idea of generalized moving variance on each SVID group to depict the group condition of the tool. To develop the tool condition monitoring, they took the first principal component of PCA applied on moving covariance matrix of each group and got the matrix of 1^{st} PCs. Then to generate an overall tool condition monitoring they again applied generalized moving variance on the matrix of 1^{st} PCs. Since they consolidated the massive amount of tool sensors data into one single indicator, they claimed that this overall indicator was recipe-independent. By this overall tool condition, they detected the abnormal tool conditions and explained them by the causal sensor groups.

By integrating the Reconstruction-Based Multivariate Contribution Analysis (RBMCA) with Fuzzy-Signed Directed Graph (SDG), He et al. [83] developed a hybrid data-driven and knowledge-based fault diagnosis method to identify the cause of the detected fault. First, the authors proposed an RBMCA based fuzzy logic to represent the signs of the

process variables. The fuzzy logic was then extended to examine the potential relationship between causes to effects regarding the Degree of Truth (DoT).

Rato et al. [84] proposed a data-driven approach for process monitoring in semiconductor manufacturing on the FDC temporal data collected from the etching process. Since in a chamber, mainly the gas flows, pressure, temperature, and power module are controlled to reach the desired quality targets on the wafer surface, the collected variables by sensors are temperature, gas flow, radio frequency power, and pressure readings. The proposed approach was based on monitoring of a particular type of features, i.e., the energies. These energies were computed from the application of a translation-invariant wavelet decomposition along the time series profile of each variable in one batch. They defined the vector of wavelet energies at each scale. Each entry of the vector of wavelet energies carried information about the dynamic activity of a given variable in the frequency-band corresponding at given scale. Then they monitored the energies vector by applying a latent variable process-monitoring framework, namely MSPC-PCA (Multivariate Statistical Process Control-PCA). Therefore their proposed methodology was composed two significant stages, firstly a signal decomposition stage, where a translation-invariant wavelet transform was applied to each variable and secondly a bilinear PCA modeling stage, performed over the wavelet coefficients'energy. To diagnose the faults, the authors calculated contribution plot of the proposed approach to investigate the origin of the failure.

Anand et al. [85] presented a statistical-data-driven approach using Bayesian analysis to diagnose and isolate the source(s) of yield loss of a given process workflow for the semiconductor manufacturing. The problem was a discrepancy observed between the Forecasted wafer sort yield and the Calculated wafer sort yield for a manufacturing process due to a mismatch between any subset of preceding process steps and their corresponding forecast models. By comparing these two sorts of yield, they proposed a cycle of graph-based or network-based representations of manufacturing processes to isolate the steps responsible for an observed discrepancy.

Li et Zhang [86] developed data-driven diffusion maps based on kNN rule technique for fault detection in the semiconductor industry. This proposed method used the correlation dimension estimator to assess the dimension of the original data set. Then, it reduced the size of a data set while persevering the information by using a diffusion map. Afterward, the adapted kNN rule has been applied to detect the faults.

Rostami et al. [87] applied several data-driven approaches for equipment failure diagnosis. The authors first proposed the SVM classifier to detect the abnormal observations in semiconductor FDC data. Then tey analyzed the normal process dynamics by SOM and kept them in different clusters. These clusters were further modeled by PCA and represented a specific dynamic behavior. Fault fingerprints were then extracted by projecting the abnormal data into the PCA models and illustrated a different type of fault that may exist in the data set.

As an extended study of [87], Rostami et al. [15] developed an equipment condition diagnosis model in semiconductor manufacturing. For this aim, they trained SVM on the FDC temporal data to distinguish normal data from abnormal ones. Then they

clustered the normal wafers into several consecutive clusters by K-means clustering, to differentiate normal process from faulty-related variations. They after applied PCA onto these clusters to modularize the process dynamics and construct the normal models. They projected abnormal data to the normal models and calculated D (Hotelings T^2) and Q (residual) statistics, representing the consistency in the score and residual spaces, respectively. They estimated the control limits for the process variables. In the next stage, they computed the contribution value of variables for each Out-Of-Control observation and applied again k-mean on the contributions to extract the highest possible number of fault fingerprints. They explained that since all faulty data must share the same faulty variables, the contribution value of a variable indicates that how much that variable creates abnormality deviation. Finally, they summarized the contribution value to provide unique contribution plot. They developed the control limits for contribution values to find the process variables that are different in the abnormal data compared to the normal models. They interpreted the variables as the fault roots if their summarized contribution values exceeded the corresponding control limits.

Table 2.7 summarizes the literature review on equipment behavior prognostic papers. **Figure 2.5a** to **Figure 2.5d** are depicted based on **Table 2.7** and visually show how equipment failure diagnosis problems have been studied in each of semiconductor and chemical industries. Despite the equipment behavior prognosis, **Figure 2.5a** illustrates that the equipment failure diagnosis has received equal attention in the semiconductor and chemical industries. **Figure 2.5b** shows that the majority of the studies have applied data-driven approaches to diagnose the equipment failure; while the knowledge-based approaches have the least contribution. Such a minimum percentage for knowledge-based techniques reveals that semiconductor/chemical industries are convoluted processes and experts cannot provide enough knowledge to characterize the equipment failures. Similar to prognostic approaches, the data-driven approaches have also received the highest popularity comparing to the other modeling approaches. **Figure 2.5c** illustrates that the majority of the reviewed papers have used FDC data for equipment failure diagnosis and few papers have employed metrology and R2R data. Finally, it can be observed in **Figure 2.5d** that most of the approaches in the equipment failure diagnosis have processed the data in the original domain and no paper in the time-frequency domain.

Table 2.7: Literature: Equipment Failure Diagnosis

Reference No.	Data	Industry		Modeling			Data			Processing	
		Semiconductor	Chemical	Physical model-based	Knowledge-based	Data-driven	FDC	Metrology	R2R	Original Domain	Time-Frequency Domain
[88]	1992	✓				✓	✓			✓	
[89]	1997		✓			✓	✓			✓	
[90]	1997	✓				✓	✓			✓	
[91]	1998	✓				✓			✓	✓	
[73]	2000	✓				✓	✓			✓	
[92]	2000		✓			✓	✓			✓	
[24]	2000	✓				✓	✓	✓		✓	
[93]	2000	✓				✓	✓			✓	
[94]	2000	✓				✓	✓			✓	
[95]	2001		✓			✓	✓			✓	
[96]	2001	✓				✓		✓		✓	
[97]	2002		✓			✓	✓			✓	
[98]	2002	✓				✓	✓			✓	
[99]	2003		✓			✓	✓			✓	
[100]	2004	✓				✓	✓			✓	
[101]	2004		✓			✓	✓			✓	
[102]	2004		✓			✓	✓			✓	
[103]	2004		✓			✓	✓			✓	
[104]	2004	✓				✓	✓			✓	
[105]	2004	✓				✓			✓	✓	
[106]	2004	✓				✓		✓		✓	
[75]	2005	✓		✓			✓			✓	
[107]	2005		✓			✓	✓			✓	
[108]	2005		✓			✓	✓			✓	
[109]	2005	✓				✓	✓			✓	
[110]	2006	✓				✓	✓	✓		✓	
[111]	2006	✓			✓	✓	✓			✓	
[112]	2006		✓			✓	✓			✓	
[113]	2006		✓			✓	✓			✓	
[114]	2007	✓				✓	✓			✓	
[115]	2007		✓			✓	✓			✓	
[116]	2008	✓				✓	✓			✓	
[117]	2008		✓			✓	✓			✓	
[118]	2008		✓			✓	✓			✓	
[119]	2008		✓	✓			✓			✓	
[48]	2009		✓			✓	✓			✓	
[120]	2009		✓			✓	✓			✓	
[121]	2009		✓			✓	✓			✓	
[122]	2009		✓			✓	✓			✓	
[77]	2009	✓				✓	✓			✓	
[123]	2010		✓	✓			✓			✓	
[124]	2012		✓			✓	✓			✓	

Continued on next page

Table 2.7 – *Literature: Equipment Failure Diagnosis (continued).*

Reference No.	Data	Industry		Modeling			Data			Processing	
		Semiconductor	Chemical	Physical model-based	Knowledge-based	Data-driven	FDC	Metrology	R2R	Original Domain	Time-Frequency Domain
[81]	2013	✓				✓	✓			✓	
[125]	2013		✓			✓	✓			✓	
[80]	2013	✓				✓	✓			✓	
[23]	2013	✓			✓	✓	✓			✓	
[83]	2014		✓		✓		✓			✓	
[126]	2014		✓			✓	✓			✓	
[82]	2014	✓				✓	✓		✓	✓	
[127]	2015		✓			✓	✓			✓	
[128]	2016		✓			✓	✓			✓	
[129]	2016	✓				✓	✓			✓	
[47]	2016	✓				✓			✓	✓	
[55]	2017	✓				✓	✓			✓	
Total Papers		**27**	**27**	**3**	**3**	**50**	**49**	**4**	**4**	**53**	**0**

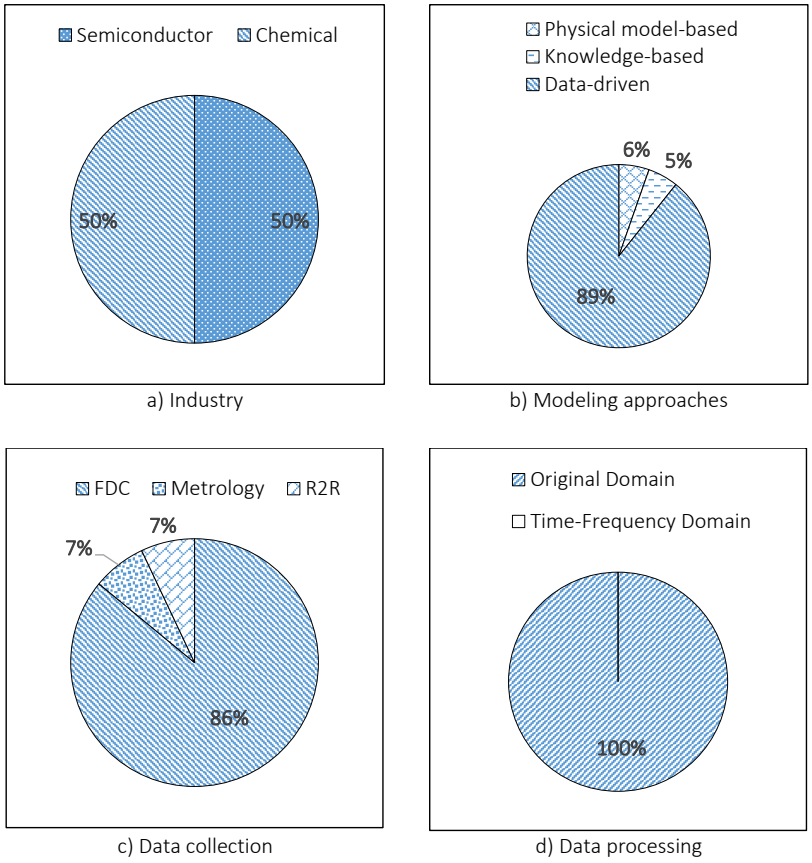


Figure 2.5: Literature analysis: Equipment Behavior Prognosis.

2.6 Thesis Framework

Based on what has been reviewed, this research develops efficient approaches to mine and analyze FDC data from etching equipment (**Figure 2.3**) in the semiconductor industry. Different methods are proposed for both prognostic and diagnostic purposes. From the prognostic point of view, a data-driven approach is developed to prognose the etching equipment behavior with corresponding root causes. Regarding the diagnostic aspect, another data-driven approach is proposed to diagnose the etching equipment failure with corresponding fault fingerprints and fault root causes. These purposes are explained more in **Sections 2.6.1** and **Sections 2.6.2**, respectively.

2.6.1 Methodology for Equipment Behavior Prognosis

As in the digital universe, semiconductor manufacturer has been taking advantage of big data evolution, which leveraging it into fault detection improvement and predictive maintenance support. In the semiconductor industry, the booming growth of data is

an essential key for APC solutions to monitor, prognose, and diagnose at better levels. While the move to big data solutions for APC systems is critical and immigrates the benefits [130], a semiconductor manufacturer can build prognostic approaches that learn the system behavior through large volumes of historical equipment conditions data. This definition is termed the data-driven approaches, also known as data mining or machine learning approaches [125]. With a variety of process types and tens or even hundreds of sensors to collect data in semiconductor manufacturing, data-driven approach represents a platform where big data solutions are introducing notable benefits [9, 10, 11].

In the following, we generally explain the equipment behavior prognosis and its details are provided in **Chapter 3**. Since the main prognostic purpose is to avoid upcoming failures, the equipment behavior prognosis is built over normal Batch Manufacturing Process (BMP) data called FDC; however, most of the current prognostic models have been applied to continuous processes and rotatory equipment rather than to BMP data. The FDC data are pervasive types of data in today fab which include an extensive data collection with three-dimensions (wafer $\times$ SVID $\times$ observation).

The first issue in equipment behavior prognosis of FDC data is the health index (HI) extraction from a large and heterogeneous data size, which has been focused unevenly in the literature. The second issue is multiple recipe contexts in semiconductor productions thereby particular process dynamics (normal ups and downs) happen. For instance, in semiconductor manufacturing processes, etching and deposition processes are carried out over time with different process recipes. Herein, the normal changes in sensor readings must not be declared as equipment deterioration.

To the best of our knowledge, two main types of variations exist among normal wafers in FDC data: 1) *recipe-related* variation and 2) *fault-related* variation. In another point of view, recipe-related variation can be replaced by a more general category as context-related variation that includes recipe-related variation as well as PMs and consumable changes that might be normal changes. But, this thesis focuses on in-between period of two PMs and mostly concentrates on the recipe changes which are more visible and challenging in the semiconductor industry. Therefore, we continue with the word recipe-related variation in the manuscript. Recipe-related variation included all normal changes in the signal that come from the recipe settings, and they are a part of production process. The recipe-related variation is also called as process dynamics. Fault-related variations imply all unexpected and abnormal changes that may propagate in the signal and result in the equipment failure. An indispensable capability of any prognostic approach should be to distinguish these two types of variations and not to produce false alarms in case of normal process dynamics. Therefore, the proposed equipment behavior prognosis must be recipe independent, and consequently, the approach is only sensitive to fault-related variation.

The third issue is that faults contribute to the signal with different frequencies in the timescale and most commonly used approaches based on the original temporal data are unable to discover these abnormalities correctly as well as to differentiate recipe-related variations from fault-related variations. The original-domain analysis is unable to capture all information from raw signals, while deterioration and abnormal events

contribute to the signal with new and unexpected frequencies. On the other hand, frequency-based approaches like Fast Fourier Transform (FFT) are suitable only for stationary signals; however, non-stationarity such as process dynamic is a standard feature of batch processes, particularly in the semiconductor industry. Accordingly, the time-frequency analysis is required to process and analyze the data in both time and frequency scales. Approximation and Detail components of a signal (generated by time-frequency analysis) provide us the data analysis at different time and frequency scales. The fourth issue is that almost all prognostic models only deal with quantifying and monitoring the equipment deterioration but not the reason(s) of this deterioration.

Consequently, we get a step toward introducing a predictive approach to overcome these four issues. Accordingly, this thesis develops a data-driven approach in the time-frequency domain for batch processes with two goals: 1) prognosis the equipment behavior to reveal and monitor the equipment deterioration and 2) diagnose the causes of this deterioration.

2.6.2 Methodology for Equipment Failure Diagnosis

Conventional equipment failure diagnosis approaches are usually performed in two steps [22] (i) fault detection to determine whether a fault has occurred and (ii) fault classification to categorize the cause of the observed out-of-control status. Automated and efficient equipment failure diagnosis approaches can overcome the wastage and barriers caused by the poorly maintained, degraded, and/or improperly controlled equipment [22].

Similar to equipment behavior prognosis, equipment failure diagnosis should be capable of distinguishing recipe-related from fault-related variations. Classical FDC methods are unable to differentiate the process dynamics from the fault symptoms. Accordingly, besides to fault detection and fault classification steps, a third step is proposed in this thesis called *process dynamic decomposition*. The *process dynamic decomposition* clusters the non-stationary sensor readings into stationary partitions.

A majority of equipment failure diagnosis approaches in the literature lack a mechanism to extract the fault root causes. The three steps mentioned above in the equipment failure diagnosis approach become more usable if the approach can extract the root cause for each class of faults. Therefore, a fourth step is also proposed in this thesis as *Fault root extraction* that comes to practice once the failure is detected and classified to extract the faulty variables to diagnose the causes of the fault.

For the equipment failure diagnosis approach, a novel data-driven method is proposed for combining these four steps to efficient and effective detection, classification and root extraction of specific faults in the semiconductor manufacturing process.

Chapter 3

Equipment Failure Diagnosis

In this chapter, a new approach to equipment failure diagnosis is proposed, especially, for the semiconductor industry. After providing a general introduction of the problem and corresponding motivation and challenges, **Section 3.1**, describes the proposed data-driven equipment failure diagnosis approach in detail.

Semiconductor manufacturing consists of highly complex and lengthy wafer fabrication processes with at least 300 process steps and a large number of interrelated variables. The complicated setting in the manufacturing environment further increases the difficulty of process stability maintenance and quality control. CBM and product quality analysis are vital for detecting critical abnormal events in wafer fabrication for optimal tool utilization and high production yield [1]. As a consequence, the equipment failure diagnosis approach shall be implemented and executed in the most precise manner because an oversight in fault detection or classification will then cause cascading product loss in the production line.

One of the most significant challenges of equipment failure diagnosis in semiconductor manufacturing is to cope with different types of variations in the FDC data. To the best of our knowledge, two main types of variations among wafers in the FDC data are *recipe-related variations* and *fault-related variations*. The *recipe-related* variations are normal changes in the variables and consists of *intra-recipe* and *inter-recipe* variations. The *intra-recipe* variation, considered as the normal process dynamics, is induced from the settings of a single recipe and can be observed as normal ups and downs, i.e., the recipe pattern, during one wafer process run. The *inter-recipe* variation comes from the product change, i.e., the shift from one recipe to another in the consecutive process runs.

As explained in **Section 2.4**, one of the challenges in the semiconductor data is the ability to distinguish the normal non-stationarity (i.e., recipe-related variations) appeared in the temporal sensor readings from the abnormal drifts (i.e., fault-related variations). The normal non-stationeries are encoded in the recipe and lie in the chemical and physical laws of the process. This type of non-stationarity is regarded as the process dynamics. On the other hand, the fault-related variation results from any abnormal changes and unexpected behavior that propagate in the signal, and results in the equipment failure. The fault-related variation can be further categorized in two types of events: *unexpected failure* and *gradual deterioration*. The *unexpected failure* occurred

to the production machines leads to disqualified products, i.e., abnormal wafers. The sensor readings typically contain observable drifts in comparison with the normal profiles. The *gradual deterioration* happens to the equipment from one wafer process run to another, and should be captured and analyzed as the equipment deterioration models. Among the fault-related variations, unexpected failure usually dominates the gradual deterioration. **Figure 3.1** illustrates the four types of variations in the FDC data as intra-recipe variation (up-left), inter-recipe variation (up-right), gradual deterioration (down-left) and unexpected failure (down-right).

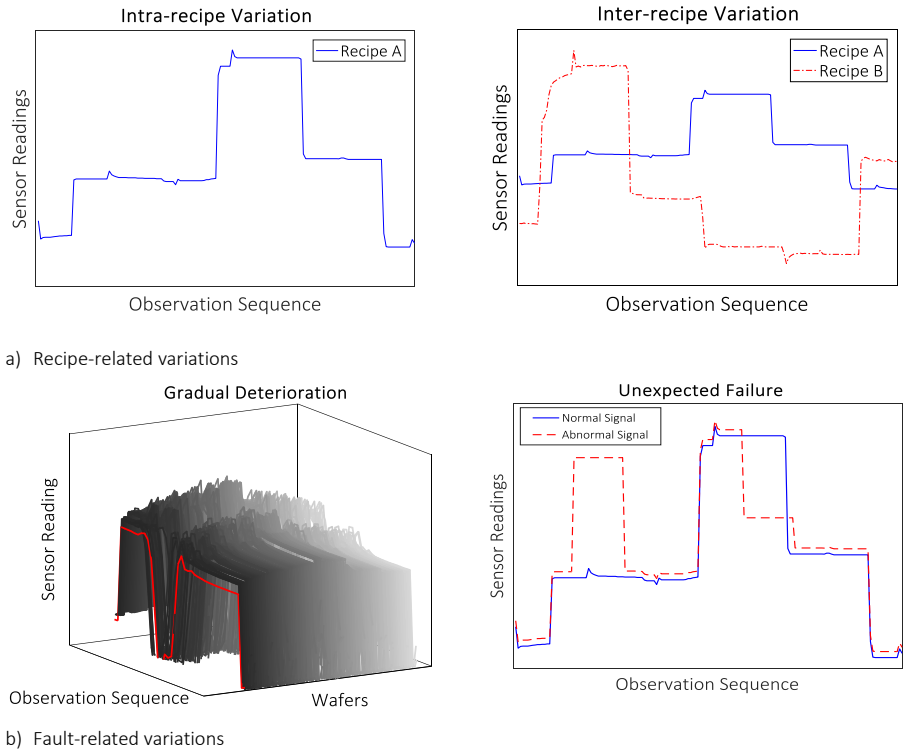


Figure 3.1: Illustration of the recipe-related variations and the events associated with fault-related variations.

It was highlighted in **Section 2.6.2** that two conventional steps in most of the classical failure diagnosis approaches are *fault detection* and *fault classification* [22]. These steps are typically to detect and classify any drift and abnormality occurred in the data comparing to the normal situation under a single recipe or a group of similar recipes. In this thesis, this step is called *process dynamic decomposition* to cluster the non-stationary sensor readings into stationary partitions. It is worth mentioning that the process dynamic decomposition is employed over the data of Normal Operation Condition (NOC) from a single recipe. Afterward, each stationary partition is utilized for further comparisons to detect faults/abnormalities.

After detecting the faults, it is even more critical to extract the fault roots of each

faulty class(s) in the semiconductor industry in order to moderate the loss. Therefore, a *Fault root extraction* step is proposed in this thesis. Once the faults classes are classified, this step looks for the faults root(s) for each class. These fault roots are expected to be different since different classes of faults have been identified. To integrate all these steps, we propose a data-driven equipment failure diagnosis approach for efficient and effective detection, classification and root extraction of specific faults in the semiconductor manufacturing process.

3.1 Equipment Failure Diagnosis Approach

Before entering to the detail, you are provided with the general framework of the proposed data-driven approach as **Figure 3.2**. This framework illustrates and explains what we are looking for and expect from the data-driven equipment diagnosis approach. Each block in the framework corresponds to a main step/purpose in the equipment failure diagnosis. Next to each block, the goal, tools to implement and the expected results of that step have been explained. It is noteworthy that depending on the data and the proposed method, this structure might be changed but still the most of the blocks should be incorporated in any advanced failure diagnosis approach. This framework is further explained and showed in detail and is decomposed into its sub-blocks/sub-steps in **Figure 3.3** and **Figure 3.4**.

The proposed data-driven approach, illustrated in the main framework of **Figure 3.2**, is adjusted into two main steps as *Equipment Anomaly Detection* and *Automatic Fault Fingerprint Extraction* to address the underlying diagnosis problem in this thesis. The first stage relates to the **Anomaly Detection** block and is to detect and classify the anomaly, and to retain the NOC data, which are then fed into the second stage that consists of four sub-stages: *Process Dynamic Decomposition* (relates to **Normal Model Creation** block), *Abnormal Data Categorization* (relates to **Normal Model Creation** and **Abnormal Data Analysis** blocks), *Fault Fingerprint extraction* (relates to **Fault Root Isolation** block), and *Fault Root Summarization* (relates to **Fault Root Isolation** block). These sub-stages are to decompose the process dynamic and to create normal models. The normal models are then utilized to categorize the abnormal data to find the fault fingerprints and their corresponding fault roots. Different and the most efficient machine learning techniques are employed in each stage such as Principal Component Analysis (PCA) [1], Support Vector Machine (SVM) [131], and K-means clustering algorithms [132]. **Figure 3.3** and **Figure 3.4** illustrate the flowchart of the proposed data-driven equipment failure diagnosis approach and the schematic view of the procedure, respectively. These stages are explained in detail in the further sections.

As explained in **Chapter 2**, the batch process data collected from the semiconductor equipment are usually called the FDC data that contain the sensor readings, also known as Status Variable IDentification (SVID). In the following sections, the structure of the FDC data is introduced, and the necessary notations throughout this chapter are provided. The proposed equipment failure diagnosis approach is then explained step by step.

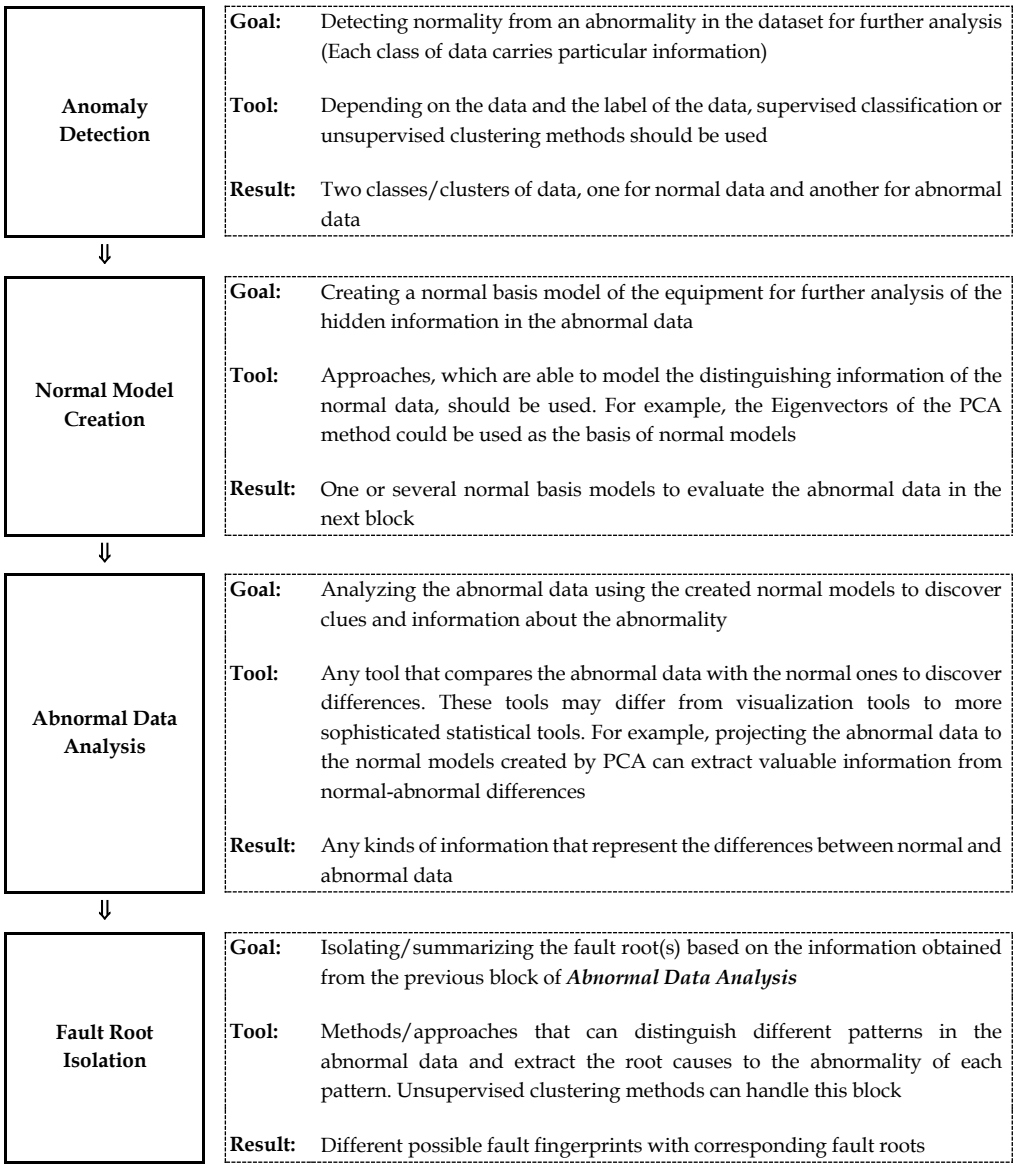


Figure 3.2: General flowchart of the proposed equipment failure diagnosis approach.

3.1.1 Fault Detection and Classification (FDC) data

To characterize the equipment in the semiconductor manufacturing environment, the FDC data are collected via the embedded sensors of the machine wherein a wafer is being processed.

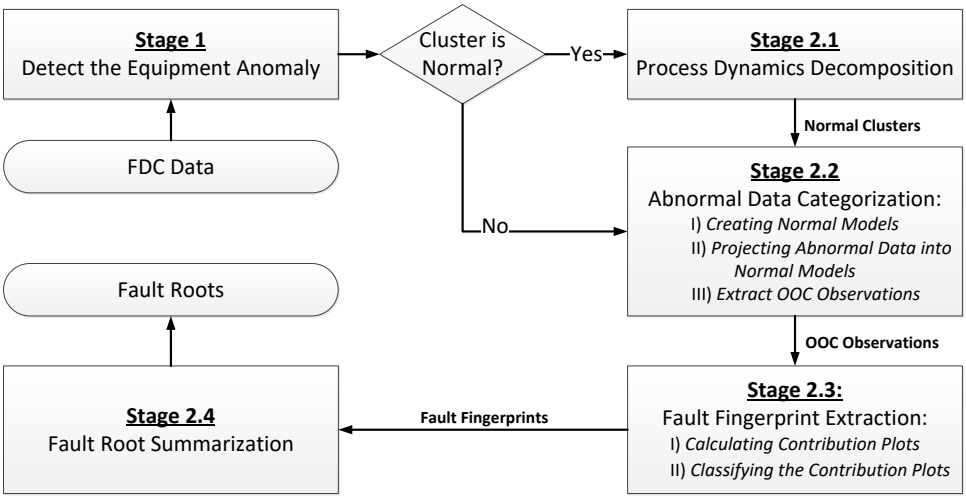


Figure 3.3: Detailed flowchart of the proposed data-driven failure diagnosis approach.

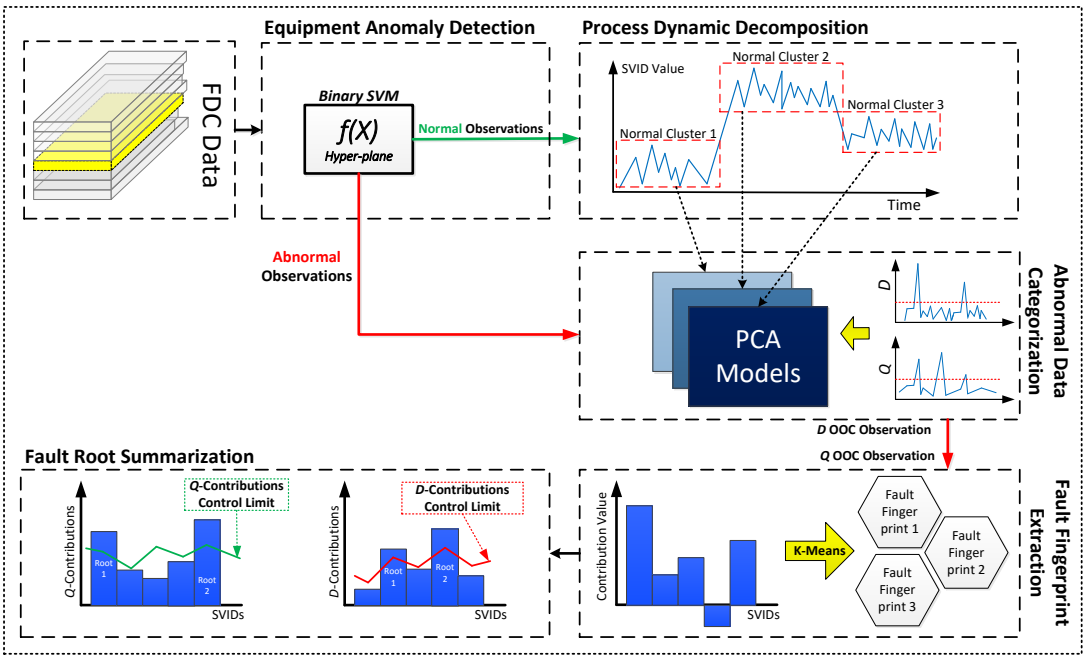


Figure 3.4: Schematic view of the proposed data-driven failure diagnosis approach.

Traditionally, fault detection and classification systems transform sensor data into summary statistics, which are then monitored to capture process excursions. The monitoring scheme is always intervened with domain experts and is nonsystematic. As shown in **Figure 3.5**, the FDC data is depicted as a data cube in three dimensions: wafer (k)

$\times$ SVID (j) $\times$ temporal observations (n_k) where the third dimension n_k is usually inconsistent from one wafer (batch) to another. It is because the process time varies from one wafer to another, given different process physics. For instance, the etching process stops only when the stop layer is reached and detected. This event causes the difference of the process time from one wafer to another and, consequently, alters the length of FDC data collected [47].

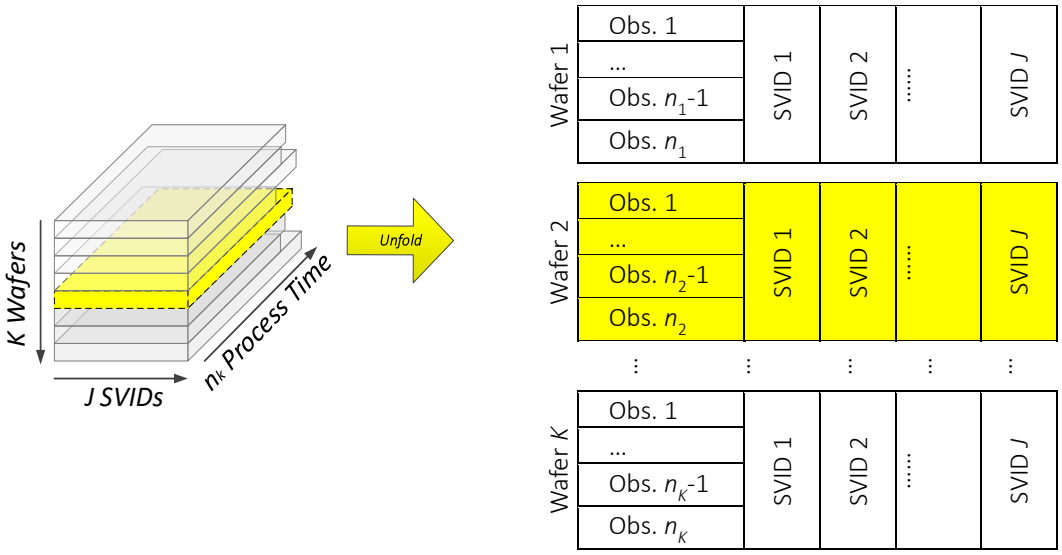


Figure 3.5: Irregular FDC data cube is illustrated and unfolded to become a data table.

Assuming there are J sensors installed in the equipment and thus J SVIDs, $v_1, v_2, ..., v_J$, will be collected when one wafer is being processed. Each SVID is collected as a temporal profile, i.e., time series, and the FDC data for wafer $k(k = 1, ..., K)$ are defined as $W(k)$ in (3.1)

$$W(k) = \begin{bmatrix} v_{1,1}^{(k)} & v_{1,2}^{(k)} & \cdots & v_{1,J}^{(k)} \\ v_{2,1}^{(k)} & v_{2,2}^{(k)} & \cdots & v_{2,J}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n_k,1}^{(k)} & v_{n_k,2}^{(k)} & \cdots & v_{n_k,J}^{(k)} \end{bmatrix}_{n_k \times J} = [V_1^{(k)}, V_2^{(k)}, ..., V_J^{(k)}], \tag{3.1}$$

where $v_{i,j}^k$ denotes the collected FDC observation of j^{th} SVID at i^{th} time stamp for wafer k; n_k is the number of observations of each SVID for wafer k and as explained is usually not the same from one wafer to another. $V_j^{(k)} = [v_{1,j}^{(k)}, v_{2,j}^{(k)}, ..., v_{n_k,J}^{(k)}]^T$ is the signal of the j^{th} SVID with a length of n_k of the k^{th} wafer. To deal with the irregular data cube, conventional unfolding techniques are usually employed [133]. In this research, the FDC

data are unfolded by stacking the wafers and the long data table is denoted as:

$$X = \begin{bmatrix} W(1) \\ W(2) \\ \vdots \\ W(K) \end{bmatrix}_{N \times J}, \tag{3.2}$$

where $N = \sum_{n=1}^{K} n_k$. The FDC system collects a significant amount of data in real-time and sometimes has problems assembling the data properly in the databases. This leads to out-of-range values, infeasible data combinations, missing values, non-informative variables, etc. [134]. Analyzing data that has not been carefully prepared for diagnostic approaches lead to misleading results. Thus, the pretreatment of the FDC data is of the first priority before running an analysis.

In the FDC data, common problems are wafers with exceptionally few or large observations compared to the normal runs. These wafers are eliminated from the FDC data-set in this thesis. Missing values are also interpolated as the mean values of two neighbor observations of the corresponding SVIDs. Finally, non-informative SVIDs, such as those with one consistent value all the time, are eliminated from the data set.

Without loss of generality, each SVID is normalized to zero mean and unit variance. The purpose of normalization is to avoid particular variables dominating the model results due to the value scales.

3.1.2 Data Pretreatment

Before starting to develop the equipment failure diagnosis model, the collected FDC data of **Section 3.1.1** should be prepared for further utilizations. The critical concern of the FDC data is the data collection quality, which is evaluated by the Data Collection Quality Value (DCQV) [133]. DCQV is the ratio of the collected number of observations over the theoretical number of observations. Due to the IT interfacing issues between the database servers and the production machine, there exist missing times tamps and the corresponding sensor readings from batch to batch, i.e., wafer. Therefore, the wafer with DCQV lower than 95% will be firstly filtered out. Furthermore, the FDC data of one wafer that contains zero variation for all the SVID, i.e., readings of each SVID are constant, will be deleted from the data set. Finally, non-informative SVIDs are eliminated from the data set based on the domain knowledge of engineers.

3.1.3 Notations

Necessary notations utilized in the proposed data-driven failure diagnosis approach are listed in **Table 3.1**.

3.2 Stage 1: Equipment Anomaly Detection

The first stage of the proposed method (see **Figure 3.3** and **Figure 3.4**) utilizes a binary Support Vector Machine (SVM) as a state-of-the-art classifier to detect if an observation is normal or abnormal in the FDC data [131]. The SVM classifier is trained with known labels, denoted as a $N \times 1$ vector with binary values: "+1" and "-1", indicating if the corresponding observation is normal or abnormal, respectively. Abnormal condition indicates there are faulty signals during the corresponding process periods.

Table 3.1: The notation definition.

Notation	**Definition**
$k = 1, ..., K$	Wafer k among K wafers
$j = 1, ..., J$	SVID j among J SVIDs
$i = 1, ..., n_k$	i^{th} observation of n_k time-stamps for wafer k
$c = 1, ..., M$	c^{th} cluster among the M clusters built only based on the normal observations
$r_c = 1, ..., R_c$	r_c^{th} principal component (PC) out of R_c PCs in the PCA model for cluster c. $R_c \leq J \quad \forall c = 1, ..., M$
N	Total number of observations in the unfolded FDC data
$N_{G,c}$	Total number of normal (Good) observations in the c^{th} cluster
N_B	Total number of abnormal (Bad) observations ($\sum_c N_{G,c} + N_B = N$)
$l = 1, ..., N_B$	l^{th} observation in the collection of all abnormal data
$X_{G,c}$	The $N_{G,c} \times J$ matrix of centered normal data in the c^{th} cluster
$T_{G,c}$	The $N_{G,c} \times R_c$ score matrix of normal observations in the PCA model of the c^{th} cluster
P_c	The $J \times R_c$ loading matrix in the PCA model for the c^{th} cluster
$E_{G,c}$	The $N_{G,c} \times J$ residual matrix of the normal data in the PCA model of the c^{th} cluster
X_B	The $N_B \times J$ matrix of total centered abnormal data
$T_{B,c}$	The $N_B \times R_c$ score matrix of abnormal data projected into the PCA model of the c^{th} cluster
$E_{B,c}$	The $N_B \times J$ residual matrix of the abnormal data projected into the PCA model of the c^{th} cluster
$D_{l,c}$	D statistics of l^{th} out-of-control (OOC) observation for the c^{th} cluster ($l \in OOC_c^D$)
$Q_{l,c}$	Q statistics of l^{th} OOC observation for the c^{th} cluster ($l \in OOC_c^Q$)
CL_c^D	Control limits of D statistics for the c^{th} cluster
CL_c^Q	Control limits of Q statistics for the c^{th} cluster
F_{jlc}^D	The contribution of SVID j in l^{th} OOC observation on D statistics for the c^{th} cluster ($l \in OOC_c^D$)
F_{jlc}^Q	The contribution of SVID j in l^{th} OOC observation on Q statistics for the c^{th} cluster ($l \in OOC_c^Q$)

It is noteworthy that the FDC data of the semiconductor industry are complex and not linearly separable, and the optimal classification is obtained only by curved (non-linear) hyper-planes. In this regard, the SVM can also be used in nonlinear classification via properly selected kernel functions [131]. By using the nonlinear mapping kernel functions, the original data are mapped into a high-dimensional feature space, where the linear classification is then possible. Among a long list of kernel functions [131], four mostly utilized functions are "Linear", "Polynomial", "RBF", and "Sigmoid" functions as shown in **Table 3.2**. The performance of the SVM classifier technique is improved once the kernel parameters are adjusted carefully in advance. These parameters consist of first regularization parameter (or misclassification penalty) C, that determines the tradeoff cost between minimizing the training error and minimizing the complexity of the model; and second parameters σ, a, and the polynomial degree, d, of the kernel functions which define the non-linear mapping from the input space to other spaces, and third the constant parameter b. Among the different kernel functions, the one with the highest accuracy will be found and set in the proposed method. The accuracy is simply calculated as dividing the number of corrected classified observation by the total number of observation.

The performance of the SVM algorithm highly depends on the level of its parameters. The optimum tuning of the parameters allows to have the lowest errors type I and II. Since the detected observations by the SVM will be fed to the following stages of the algorithm for creating the normal model and even characterizing the abnormal ones, it is necessary to find the optimum level of the parameters. In order to adjust the SVM parameters to achieve the highest accuracy, different evolutionary algorithms have been utilized to optimize the SVM settings, such as Genetic Algorithm (GA) [131, 135], particle swarm optimization (PSO) [136], artificial bee colony algorithm [137] and Grid algorithm [138]. Among them, GA has been widely and successfully employed in various optimization problems [135, 139, 140] and applied in modeling problems with varying resolutions and structures. The GA can thoroughly explore and exploit nonlinear solution spaces without requiring gradient information or a priori knowledge about the problem characteristics [141]. One of the main reasons that we selected GA among others is its power in escaping from local optima due to its exploitation and exploration in solution space using both crossover and mutation operators. While PSO and Bee algorithms easily fall into local optimum in high-dimensional space and have a low convergence rate in the iterative process. On the other hand, the Grid algorithm is time-consuming and does not perform well. Based on the Darwinian principle of survival of the fittest, the GA as a particular class of evolutionary algorithms is time-efficient and has well heuristic search method that can obtain the optimal solution. Accordingly, a GA similar to [131] is tailored and used to find the optimal value of the SVM parameters.

3.3 Stage 2: Automatic Fault Fingerprint Extraction

In **Figure 3.3**, and **Figure 3.4**, Stage 2 consists of four sub-stages: process dynamics decomposition (Stage 2-1), abnormal data categorization (Stage 2-2), fault fingerprint

extraction (Stage 2-3), and fault root summarization (Stage 2-4), which are explained in details in the following sub-sections.

Table 3.2: Kernel functions utilized in Stage 1: Equipment Anomaly Detection.

	Kernel function			
	Linear	**Polynomial**	**RBF**	**Sigmoid**
Formulation	$x^T y + b$	$(ax^T y + b)^d$	$e^{(-\frac{\|\|x-y\|\|^2}{2\sigma^2})}$	$tanh(ax^T y + b)$

3.3.1 Stage 2-1: Process Dynamics Decomposition

After distinguishing the normal observations from the abnormal ones in Stage 1, Stage 2-1 is to cluster the hundred or even millions of normal observations into a smaller number of clusters to decompose the process dynamics and not to detect them as a fault. As an example in **Figure 3.6**, the SVID temporal profile shows a non-stationary pattern induced from the recipe setting and jeopardizes all the algorithms with the stationarity assumption. For this aim, Stage 2-1 provides several normal process clusters, where intra-cluster observations can be treated as stationary. It must be noticed that Stage 2-1 is conducted only on the normal data classified from Stage 1.

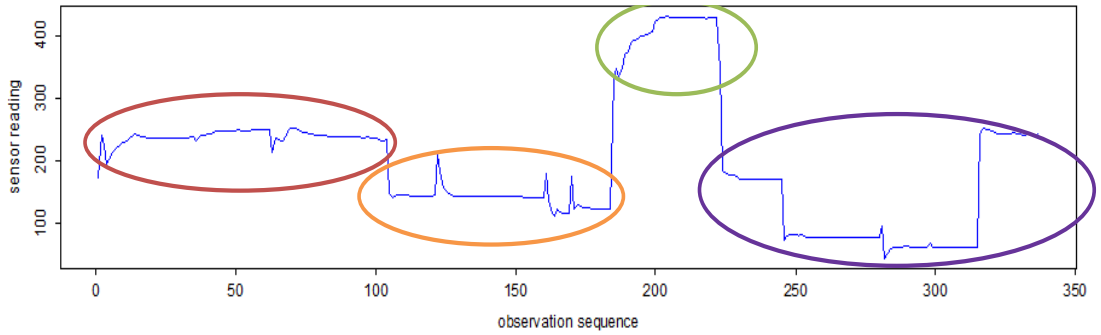


Figure 3.6: The non-stationary temporal profile of an SVID for one wafer is illustrated.

As the abnormal data shall contain faulty signals, differentiating the process dynamics is not meaningful. The K-means clustering algorithm is employed with the correlation between observations as the similarity measure. To better decompose the process dynamics, a dummy time-stamps variable is added to the data with the size equal to the number of observations in wafer k. Therefore, the FDC data $W(k)$ in (3.1) is augmented to become $W^*(k)$ of (3.3). By this new dummy variable, if two observations, say i_1 and i_2 have similar SVID values (at y axis) but are located at the very different time-stamps (at x axis), ex: $1 < i_2 \ll i_2 < n_k$, they shall not be grouped into the same cluster as

they belong to different part of the process dynamics.

$$W^*(k) = \begin{bmatrix} v_{1,1}^{(k)} & v_{1,2}^{(k)} & \cdots & v_{1,J}^{(k)} & 1 \\ v_{2,1}^{(k)} & v_{2,2}^{(k)} & \cdots & v_{2,J}^{(k)} & 2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ v_{n_k,1}^{(k)} & v_{n_k,2}^{(k)} & \cdots & v_{n_k,J}^{(k)} & n_k \end{bmatrix}_{n_k \times (J+1)} , \tag{3.3}$$

At the end of this stage, we have several cluster of normal observations (where each cluster corresponds to a part of the process dynamic) which are used to build normal basis models. The normal models are utilized to categorize the abnormal observations.

3.3.2 Stage 2-2: Abnormal Data Categorization

As the process dynamics are decomposed and the clusters of the normal observations are obtained, the abnormal data should be analyzed accordingly to find the fault patterns. In this regard, each cluster of the normal observations is used to create a normal model specific to that corresponding part of the process dynamic. Accordingly, PCA is firstly applied to summarize each cluster into a normal baseline model. PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.

Abnormal data are then projected into each model to measure the divergence of the abnormal data from the normal models. Hereafter, the notations are based on **Table 3.1**. For creating the normal baseline models, PCA decomposes the c^{th} matrix of centered normal data $X_{G,c}$ (obtained from Stage 2-1), into uncorrelated score matrix $T_{G,c}$ by an orthogonal loading matrix, P_c and the unexplained variation of $X_{G,c}$, that is, residual matrix $E_{G,c}$ [80]:

$$X_{G,c} = T_{G,c}P_c^{'} + E_{G,c}, \qquad \forall c \tag{3.4}$$

Most variations in the c^{th} cluster can be explained by the first R_c PCs of the matrix $T_{G,c}P_c^{'}$. Given the abnormal data X_B, abnormal score and residual matrices are calculated as (3.5) and (3.6), respectively.

$$T_{B,c} = X_B P_c, \qquad \forall c \tag{3.5}$$

$$E_{B,c} = X_B(I - P_c P_c^{'}), \qquad \forall c \tag{3.6}$$

where I is the identity matrix.

For fault detection, the normal PCA models developed based on normal observations are then used to check new measurement data. The differences between the new measurement data and their projections to the built normal model, the residuals, are then subjected to some sort of statistical test to determine if they are significant. Usually the Q statistic, also called squared prediction error (SPE), and the Hotelling's (T^2) statistic,

also called as D statistic, are used to represent the variability in the residual subspace and principal component subspace.

After building the normal baseline PCA models and projecting abnormal data into each model, faulty observations can be detected by the score space ($T_{B,c}$) and the residual space ($E_{B,c}$). Within the c^{th} cluster, the D_{lc} statistic and Q_{lc} statistic are calculated as (3.7) and (3.8), respectively, to measure the divergence of abnormal observations. The D_{lc} statistic monitors systematic variation in the score space, while the Q_{lc} statistic monitors variation which is not explained by the selected PCs [80].

$$D_{lc} = T_{B,lc} S_c^{-1} T_{B,lc}^{'} \sim \frac{R_c(df^2-1)}{df(df-R_c)} F_\alpha(R_c, df-R_c), \qquad \forall c,l \tag{3.7}$$

$$Q_{lc} = E_{B,lc} E_{B,lc}^{'} = X_{B,l}^{'}(I - P_c P_c^{'}) X_{B,l} \sim \chi^2_{(h)}, \qquad \forall c,l \tag{3.8}$$

where S_c^{-1} is the diagonal matrix containing the inverse eigenvalues associated with the R_c eigenvectors in the c^{th} normal cluster. In addition, df is the degree of freedom that can be the number of observation in the c^{th} normal cluster.

In formulation (3.7), the D_{lc} statistic provides an indication of unusual variability within the normal subspace. The value of D_{lc} for an abnormal observation is equal to the sum of squares of the adjusted (unit variance) scores on each of the principal components in the normal model. Furthermore in formulation (3.8), the Q_{lc} statistic shows how well an abnormal observation fits into the PCA model built on the normal data. It is a measure of the difference (residual) between the abnormal observation and its projection onto the principal components retained in the normal model.

The normal observations determine the control limits for the D_{lc} and the Q_{lc} statistics under normal conditions. The D_{lc} represents the squared length of the projection of the abnormal observation into the space spanned by the normal PCA models. It is an indication of how far the PCA estimate of the abnormal observation is from multivariate mean of the data (i.e., the intersection of the principal components). Therefore, if sample has an abnormal value of D_{lc} but Q_{lc} value below the limit, it is not necessarily a fault (it can also be a change of the operating region). The statistical control limit for the values of D_{lc}, i.e., CL_c^D, can be calculated according to statistical F-distribution as (3.9) [80].

The normal observations determine the control limits for the D_{lc} and Q_{lc} statistics under normal conditions. The control limit for the D_{lc} statistic, i.e., CL_c^D, is calculated by (3.9) [80].

$$CL_c^D = \frac{R_c(df^2-1)}{df(df-R_c)} F_\alpha(R_c, df-R_c), \qquad \forall c,l \tag{3.9}$$

where $F_\alpha(R_c, df-R_c)$ is the upper 100 α% critical point of the F-distribution with R_c and $df - R_c$ as the degree of freedom [142] and df is the number of observations in the data set used in the calculation of the PCA model in the corresponding normal cluster. The control limit CL_c^D defines an ellipse on the plane within the data are assumed to be normal.

To calculate the control limit for the Q_{lc} statistic, let $\theta_1^c = trace(V^c)$, $\theta_2^c = trace(V^c)^2$, $\theta_3^c = trace(V^c)^3$, where V^c is the covariance matrix of normal residual. Also let $h_{0,c} = 1 - \frac{(2\theta_1^c\theta_3^c)}{(3\theta_2^c)}$. Then,

$$(\frac{Q_{lc}}{\theta_1^c})^{h_{0,c}} \sim N[1 + \frac{\theta_2^c h_{0,c}(h_{0,c}-1)}{(\theta_1^c)^2}, \frac{2\theta_2^c h_{0,c}^2}{(\theta_1^c)^2}], \qquad \forall l, c \tag{3.10}$$

This means that the probability distribution for Q_{lc} can be approximated as (3.11). To calculate the control limit for the Q_{lc} statistic, firstly the probability distribution for Q_{lc} can be approximated as (3.11) [143]

$$\int_0^{Q_{lc}} f(q)\mathrm{d}q \approx \int_{-\infty}^{z} g(x)\mathrm{d}x, \tag{3.11}$$

where $g(x)$ is the normal density function and z is defined as (3.12):

$$z = \frac{\theta_1^c[(\frac{Q_{lc}}{\theta_1^c})^{h_{0,c}} - 1 - \theta_2^c h_{0,c}(h_{0,c}-1)/(\theta_1^c)^2]}{\sqrt{2\theta_2^c h_{0,c}^2}}, \tag{3.12}$$

Conversely, for a fixed type I error α, the upper control limit for the Q_{lc} statistic, CL_c^Q, can be approximated as 3.13, provided that all the eigenvalues of the covariance matrix are known.

$$CL_c^Q = \theta_1^c[\frac{z_\alpha\sqrt{2\theta_2^c h_{0,c}^2}}{\theta_1^c} + 1 + \frac{\theta_2^c h_{0,c}(h_{0,c}-1)}{(\theta_1^c)^2}]^{\frac{1}{h_{0,c}}}, \qquad \forall c \tag{3.13}$$

where z_α is the standardized normal variable with (1-α) confidence limit, having the same sign as $h_{0,c}$.

By comparing the D_{lc} and the Q_{lc} statistics with their corresponding control limits i.e., CL_c^D and CL_c^Q, the Out-Of-Control (OOC) observations can be identified. Hereafter, the OOC observations based on D and Q statistics for the c^{th} cluster are denoted as OOC_c^D and OOC_c^Q, respectively. Accordingly, observation l is considered as an OOC based on D_{lc} and Q_{lc} statistics if $D_{lc} > CL_c^D$ and $Q_{lc} > CL_c^Q$, respectively. In the c^{th} cluster, the sets of OOC observations due to D_{lc} violation (OOC_c^D) and due to Q_{lc} violation (OOC_c^Q) are archived. It is noteworthy that the archived OOC observations carry important information regarding fault fingerprints.

3.3.3 Stage 2-3: Fault Fingerprint Extraction

As the OOC observations are identified and collected, different fault fingerprints should be extracted to explain the root causes of the fault events better. Chien et al. [80] clustered the OOC observations using a (Self-Organizing Map) SOM network. However,

clustering the points based on their coordinates may lead to incorrect information. One way to cope with this issue is to calculate the contribution of the variables (SVIDs) for each OOC observation based on the D_{lc} and the Q_{lc} statistics, and then cluster the OOC observations in terms of their contribution values.

Another reason to examine the contribution values instead of the observations coordinates is that in the case of a process perturbation, one of the D_{lc} and the Q_{lc} statistics or even both might fall out of their control limits. If only the D_{lc} statistic exceeds the control limit, i.e., $D_{lc} > CL_c^D$, the model of the process is still valid, but the distance between the batch and the center of the model is too large. In this case, contributions of each SVID to the D_{lc} statistic should be examined. If the Q_{lc} statistic exceeds the control limit, i.e., $Q_{lc} > CL_c^Q$, a new event, that cannot be described by the process model, has happened in the data. In that case, the contributions of each process variable to the Q_{lc} statistic should be examined as well. To further consolidate the contribution values, clustering is applied to extract the common faulty patterns.

Nomikos [144] provided approximated formulations to calculate the contribution values which can be valid for principal component decompositions. In another approach, [92] generalized the formulations proposed by [144], while the new formulations can be also used for nonorthogonal scores and loadings as (3.14). For calculating the contribution of each variable to the D_{lc} statistic, (3.7) is decomposed into more details as (3.14).

$$\begin{aligned} D_{lc} &= T_{B,lc} S_c^{-1} T_{B,lc}^{'} \\ &= T_{B,lc} S_c^{-1} [X_{B,l}^{'} P (P^{'} P)^{-1}]^{'} \\ &= T_{B,lc} S_c^{-1} \sum_{j=1}^{J} [X_{B,l,j} P_j^{'} (P^{'} P)^{-1}]^{'} \qquad \forall c, l \in OOC_c^D \\ &= \sum_{j=1}^{J} T_{B,lc} S_c^{-1} [X_{B,l,j} P_j^{'} (P^{'} P)^{-1}]^{'} \\ &= \sum_{j=1}^{J} F_{jlc}^D, \end{aligned} \tag{3.14}$$

Thus, the contribution of the variable j to the D_{lc} statistic, F_{jlc}^D, is obtained as (3.15):

$$F_{jlc}^D = T_{B,lc} S_c^{-1} [X_{B,l,j} P_j^{'} (P^{'} P)^{-1}]^{'} \qquad \forall j, l \in OOC_c^D, c, \tag{3.15}$$

where $T_{B,lc}$ is the l^{th} row of the matrix $T_{B,lc}$, $X_{B,jl}$ is the array of the l^{th} row and j^{th} column of the matrix $X_{B,jl}$, and P_j is the corresponding loading element.

The summation of residuals of each variable j helps to figure out in which time-stamp the disturbance happened. Consequently, the contribution of the variable j to the Q_{lc} statistic, F_{jlc}^Q, can be obtained as:

$$Q_{lc} = \sum_{j=1}^{J} F_{jlc}^Q = \sum_{j=1}^{J} (E_{B,jlc})^2 \qquad \forall l \in OOC_c^Q, c, \text{and} \tag{3.16}$$

$$F^Q_{jlc} = (E_{B,jlc})^2 \qquad \forall j, l \in OOC^Q_c, c, \tag{3.17}$$

where $E_{B,jlc}$ is the array of the l^{th} row and j^{th} column of the matrix $E_{B,c}$. Hereafter, the vectors of contribution values of l^{th} OOC observation on D_{lc} ($VI^D_{c,l}$) and Q_{lc} ($VI^Q_{c,l}$) statistics are gathered into the unique matrices denoted as (3.18) and (3.19), respectively.

$$VI^D = \begin{bmatrix} VI^D_{c,1} \\ \vdots \\ VI^D_{c,l} \\ \vdots \\ VI^D_{c,N_B} \end{bmatrix}, \quad VI^D_{c,l} = [F^D_{1lc}, ..., F^D_{jlc}, ..., F^D_{Jlc}], \tag{3.18}$$

$$VI^Q = \begin{bmatrix} VI^Q_{c,1} \\ \vdots \\ VI^Q_{c,l} \\ \vdots \\ VI^Q_{c,N_B} \end{bmatrix}, \quad VI^Q_{c,l} = [F^Q_{1lc}, ..., F^Q_{jlc}, ..., F^Q_{Jlc}], \tag{3.19}$$

Finally, the K-means clustering algorithm with correlation similarity function is separately applied on VI^D and VI^Q in order to extract the potential fault fingerprints.

3.3.4 Stage 2-4: Fault Root Summarization

The last step of the proposed method is to find out the root causes in each created cluster, i.e., the fault fingerprints, from VI^D and VI^Q. For this aim, this section proposes a new way of calculating the control limits for the contribution values of D_{lc} and Q_{lc} statistics. These control limits help to find the process variables that are different in the abnormal data compared to the normal models. If one process variable has high contribution value in the normal model, it can also be expected to have high contribution value in the abnormal data. However, if a process variable has high contribution value in the abnormal data but low contribution value in the normal model, this probably is due to a special event in the abnormal data.

It must be noticed that the control limits of the contribution values are calculated from the normal observations. The control limit of the j^{th} variable in the c^{th} cluster for the D_{lc} statistic, CVL^D_{jc}, is calculated as (3.20). An underlying assumption for CVL^D_{jc} is that the set of $F^D_{jlc} : l \in IC_c$ follows normal distribution.

$$CVL^D_{jc} = \text{mean}\{F^D_{jlc}\}_{l \in IC_c} + 3\text{std}\{F^D_{jlc}\}_{l \in IC_c} \qquad \forall j, c, \tag{3.20}$$

where IC_c is the set of In-Control observations constructing the c^{th} cluster. The control limit of the j^{th} variable in the c^{th} cluster for the Q_{lc} statistic, CVL^Q_{jc}, is calculated as (3.21). To the best of our knowledge and despite of [80] for proposing a constant value

for all SVIDs, we propose the control limit (3.21) for each SVID since the contribution of variables are different and independent.

$$CVL_{jc}^{Q} = \theta_{1,j}^{c}[\frac{z_\alpha\sqrt{2\theta_{2,j}^{c}h_{0,jc}^{2}}}{\theta_{1,j}^{c}} + 1 + \frac{\theta_{2,j}^{c}h_{0,jc}(h_{0,jc}-1)}{(\theta_{1,jc})^{2}}]^{\frac{1}{h_{0,jc}}}, \qquad \forall j,c, \qquad (3.21)$$

where $h_{0,jc} = 1 - ((2\theta_{1,j}^{c}\theta_{3,j}^{c})/((3\theta_{2,j}^{c})^{2}))$, $\theta_{1,j}^{c} = (V^{c})_{jj}$, $\theta_{2,j}^{c} = (V^{c})_{jj}^{2}$, $\theta_{3,j}^{c} = (V^{c})_{jj}^{3}$, V^{c} is the covariance matrix of normal residual, and z is the standardized normal variable with $(1-\alpha)$ confidence limit, having the same sign as $h_{0,jc}$.

To compare the contribution values with the calculated control limits, we need to summarize the set of contribution vectors belonging to each fault fingerprint. Accordingly, the contribution vectors, related to the D_{lc} statistic (i.e., $VI_{c,l}^{D}, l \in OOC_{c}^{D}$), are summarized into an unique vector by calculating the $\mu + 3\sigma$ for each SVID. Thus, the contribution vectors, related to Q_{lc} statistic (i.e., $VI_{c,l}^{Q}, l \in OOC_{c}^{Q}$), are summarized into a unique vector by making average of the contribution values corresponding to each SVID. Finally, the summarized contribution values are compared with their corresponding control limits and the set of SVIDs that exceed the control limit is considered as the fault roots.

As a summary, this chapter developed a data-driven approach for the equipment failure diagnosis using the FDC data in the semiconductor industry. The proposed data-driven equipment failure diagnosis approach terminates by fault root summarization, and the extracted non-identical fault fingerprints and their corresponding roots become valuable information for practitioners. Engineers can then check the highlighted SVIDs, i.e. sensors, of the machine according to the fingerprint and decide if a corrective maintenance shall be performed.

It is worth mentioning that using fault fingerprints/roots to monitor the equipment condition and prognose the behavior is a difficult task, particularly in the semiconductor industry, wherein faults are numerous and highly confounded. Moreover, monitoring the fault fingerprints/roots do not necessarily provide a prognostic mechanism to model the equipment deterioration, and the fault roots should be different from the deterioration roots. The equipment deterioration might relate to particular process variables (SVIDs) while the fault symptoms appear in a different set of process variables. This issue necessitates developing a prognostic approach not only to find the deterioration cause roots but also to avoid the equipment failure.

Another characteristic of the proposed data-driven equipment failure diagnosis approach is that this approach treats each observation of the FDC data as an individual sample . Accordingly, the proposed approach can be considered as an observation-based approach. Most approaches in the literature summarize the observations of a wafer into a single data point and are considered as wafer-based approaches. Depending on the data collection system, each of observation-based or wafer-based mechanisms can be adapted to build diagnostic and prognostic approaches. In **Chapter 4**, an equipment behavior prognosis approach is developed by summarizing the observations of each wafer into individual features for deterioration modeling.

Chapter 4

Equipment Behavior Prognosis

Keeping the machine at a high utilization level is critical for manufacturing industries, in particular, the semiconductor fabrication, moving towards advanced technologies. Production deficiencies such as process variations and unexpected machine breakdowns have caused low-grade product yield and decreased the machine utilization significantly. Conventionally, regular machine maintenance would help to revive the machine condition for the better production efficiency. However, the maintenance schedule is usually settled by fixing either the time interval or the number of products processed between two maintenances CBM serves as a new control scheme to characterize the equipment behavior and triggers the corresponding control actions whenever required. In the prognostic perspective, equipment deterioration is modeled and monitored. An effective deterioration model, in the shortest reaction time, is critical for minimizing scrap wafers, reducing unscheduled machine breakdowns, increasing equipment utilization, and maintaining high production yields.

This chapter develops an efficient equipment behavior approach for modeling and monitoring the equipment deterioration using the batch process data of the semiconductor industry (please refer to **Section 3.1.1**). In this regard, several challenges make it difficult to employ classical techniques and methods in the literature to efficiently and effectively model and monitor the equipment deterioration.

The first challenge is that most of the current prognostic models have been applied to one-dimensional data, e.g., vibration signals, rather than to batch manufacturing processes, i.e., FDC data. Depending on the process characteristics, the length of observations of one product may differ from another. Therefore, it takes more preprocessing steps to extract the equipment Health Index (HI). Accordingly, a proper data treatment with tolerable information loss alongside an efficient analyzing approach is required for modeling and monitoring the equipment deterioration effectively. It is also critical to deal with the fact that the process recipes of a single machine are frequently changed in the semiconductor manufacturing environment. Different recipes lead to dissimilar sensors profiles from one batch to another and should be carefully grouped and analyzed according to their natures. For example, a normal change between two different recipes must not be detected as the equipment deterioration or a fault. The second challenge is that current prognostic models only deal with quantifying and monitoring the equipment

deterioration, and the root cause to the deterioration is not investigated [2, 145, 5, 14]. The third challenge comes from the presence of different types of variations in the FDC data of the semiconductor equipment. As explained in **Chapter 3**, two main types of variations among wafers in the FDC data are recipe-related variations and fault-related variation, which leads to unexpected failures and the gradual equipment deterioration. **Chapter 3** dealt with the fault diagnosis when an unexpected failure occurs and this chapter works on characterizing the deterioration.

In this thesis, the deterioration is assumed to be caused by the *macro-* and *micro-level* variations of the signals. *Macro-level* variation corresponds to all significant shifts of the mean level of the signal. **Figure 4.1** illustrates an example of the macro-level variation of an SVID as the temporal profiles of consecutive wafers progressively shift. *Micro-level* variation corresponds to the implicit changes over the main pattern of an SVID. The SVID variation can be amplified as the equipment deteriorates while the mean level remains the same. **Figure 4.2** illustrates the micro-level variation in a wafer.

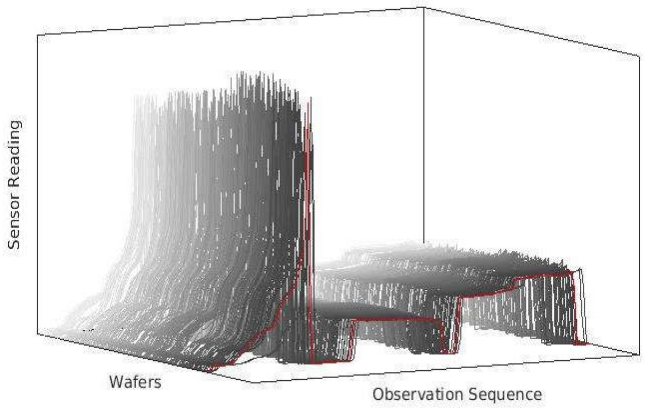


Figure 4.1: The macro-level variation of an SVID across different wafers is illustrative. The main pattern of the SVID is gradually shifted through the wafers. The red line indicates the profile of the final normal wafer.

Similar to the equipment failure diagnosis in **Chapter 3**, the proposed equipment behavior prognosis approach here should be able to distinguish different types of variations among the normal wafers and, in particular, be capable of differentiating the macro-level variations from the micro-level variations. To overcome the issues above, the data-driven equipment behavior prognosis approach for analyzing the batch process data is proposed with two goals: 1) exploiting the temporal data of batch processes to characterize the equipment behavior, and 2) identifying the deterioration trend with the most likely causes.

Like almost all data-driven approaches, this section proposes an equipment behavior prognosis approach through four main consecutive stages: 1) *data pretreatment*, 2) *data processing*, 3) *feature extraction*, and 4) *deterioration modeling*. **Figure 4.3** depicts the general framework of what we are doing and we expect in this thesis from the

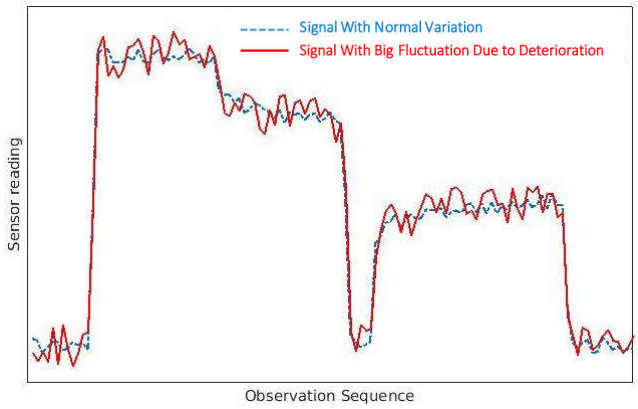


Figure 4.2: The micro-level variation of the signal is observed while the main pattern remains the same.

data-driven equipment behavior prognosis. Each block in the framework corresponds to a main step/purpose in the equipment behavior prognosis. Next to each block, the goal, tools to implement and the expected results of that step have been explained. A detailed schematic of the general framework will be provided in **Figure 4.8**. The Data Pretreatment block is performed as explained in **Section 3.1.2**.To construct the equipment behavior model, the batch process data, i.e., the FDC data, described in **Section 3.1.1** and **Figure 3.4**, should be pre-treated to extract significant features, which will be then used as the input of the model.

4.1 Data Processing: Discrete Wavelet Transformation

As elaborated above, a significant challenge is to select the most appropriate data analysis approach to provide and process the data with adequate and correct information. The active approach should manipulate and prepare the data for further feature extraction and equipment deterioration modeling and monitoring. It was previously revealed that different types of variations, in particular, the macro-level and micro-level variations are confounded in the FDC data and should be differentiated. For this aim, Discrete Wavelet Transformation (DWT), a signal processing technique to decompose a temporal signal into different frequency domains, is adopted in this section. DWT has been widely used in the past decade in different filed such as equipment condition monitoring [25], fault diagnosis [146], data and image compression [147], partial differential equation solving [148], texture analysis [149] and noise/trend reduction [150]. It is worth to mention that in semiconductor manufacturing data, the variation is exposed across time and space (i.e., temporal and spatial variation respectfully) with a number of different scales [151]. For example, the sensor data are collected along time dimension in a wafer processing and include the temporal variation. This variation is being tracked via some measures from one to the next wafer in the meantime of equipment deterioration.

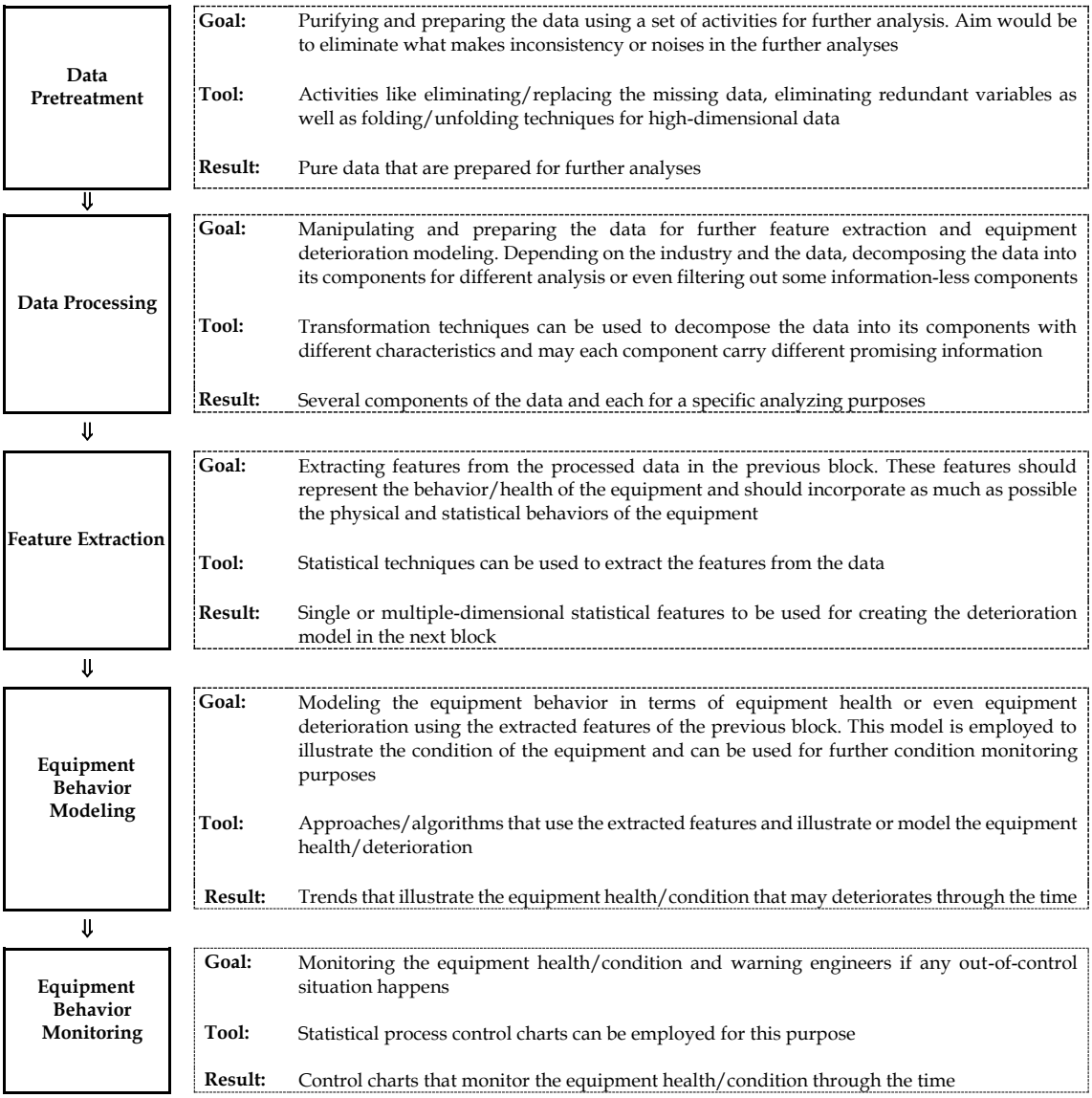


Figure 4.3: General flowchart of the proposed equipment behavior prognosis approach.

The temporal variation of a non-stationary profile in semiconductor data-set consists of systematic and random parts, which must be distinguished from one another before equipment deterioration studies. Accordingly, each signal is decomposed into a series of oscillatory functions with different frequencies at a different time. The idea of DWT is rooted in the traditional Fourier Transform (FT) [80] but has developed many unique

properties. It starts by iteratively sweeping the signal using a succession of different scale sizes. Smaller scale size is used to analyze the high-frequency components of the signal, and the larger one is used to catch the low-frequency components. As a result, each signal is then decomposed into the high-frequency and low-frequency components. From DWT, the systematic and random parts are captured by low-frequency components and high-frequency one, respectively.

The DWT comprises both wavelet decomposition and wavelet reconstruction, and these transformations are performed through two primary functions: the scale function (Φ, father wavelet) and the detail function (Ψ, mother wavelet).

$$\begin{aligned} \Phi_{l,p}(i) &= 2^{\frac{1}{2}}\Phi(2^l i - p), \int \Phi(i)\mathrm{d}i = 1, \\ \Psi_{l,p}(i) &= 2^{\frac{1}{2}}\Psi(2^l i - p), \int \Psi(i)\mathrm{d}i = 1, \end{aligned} \tag{4.1}$$

where $l \in \{0, ..., L\}$ is the decomposition-level (dilation) index, and p is the translation index. Once l increases, the width of the wavelet decreases, and the height of the wavelet increases (see **Figure 4.4**). Also, p has the effect of sliding the function along the x-axis. The raw signal $V_{n_k}^j(i)$ (i.e., i^{th} observation of j^{th} SVID with length of n_k of k^{th} wafer) is represented by the linear combination of essential functions in $L^2(R)$, as described in (**4.2**).

$$V_{n_k}^j(i) = \frac{1}{\sqrt{M}} \sum_p A_{L,p}^{j,k} \Phi_{L,p}(i) + \frac{1}{\sqrt{M}} \sum_{l=1}^{L} \sum_p D_{l,p}^{j,k} \Psi_{l,p}(i) \qquad \forall j, k, \tag{4.2}$$

where $A_{L,p}^{j,k}$ and $D_{l,p}^{j,k}$ are, respectively, approximation and detail coefficients corresponding to the j^{th} SVID with length of n_k of the k^{th} wafer. L and M correspond to the final decomposition level and the length of the components, respectively. Because the sets $\{\Phi_{L,p}(i)\}_{p\in Z}$ and $\{\Psi_{l,p}(i)\}_{(l,p)\in Z^2}$ are orthogonal to each other, we can simply take the inner product to obtain the wavelet coefficients as (**4.3**) and (**4.4**).

$$A_{L,p}^{j,k} = \frac{1}{\sqrt{M}} \sum_i V_{n_k}^j(i) \Phi_{L,p}(i) \qquad \forall p, \tag{4.3}$$

$$D_{l,p}^{j,k} = \frac{1}{\sqrt{M}} \sum_i V_{n_k}^j(i) \Psi_{l,p}(i) \qquad \forall l, p, \tag{4.4}$$

Figure 4.4 illustrates an example of DWT with three levels of decomposition ($L = 3$). The decomposed components of the signal with different frequency domains are selected and analyzed according to characteristics of the signal. The DWT shown in **Figure 4.4** uses a pair of the low pass (solid arrows) and high pass filters (dashed arrows) to decompose the signal into an approximated low-frequency component (solid circles) and a detailed high-frequency one (shaded circles). Conventionally, they are also referred to Approximation and Detail components, respectively.

Figure 4.5 depicts the decomposition of a raw signal into the approximation (left branch) and the detail (right branch) components. It can be seen that the approximation component carries the information of the systematic pattern while detail component reveals the evidence of random noise out of the raw signal. Via the wavelet decomposition, the macro-level and micro-level variations in the raw signal can be found and analyzed separately to model the equipment deterioration.

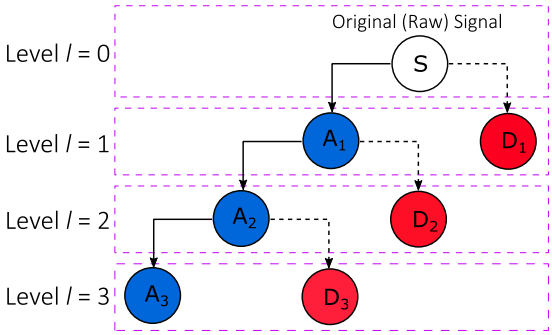


Figure 4.4: This figure shows an example of DWT in three levels ($L = 3$).

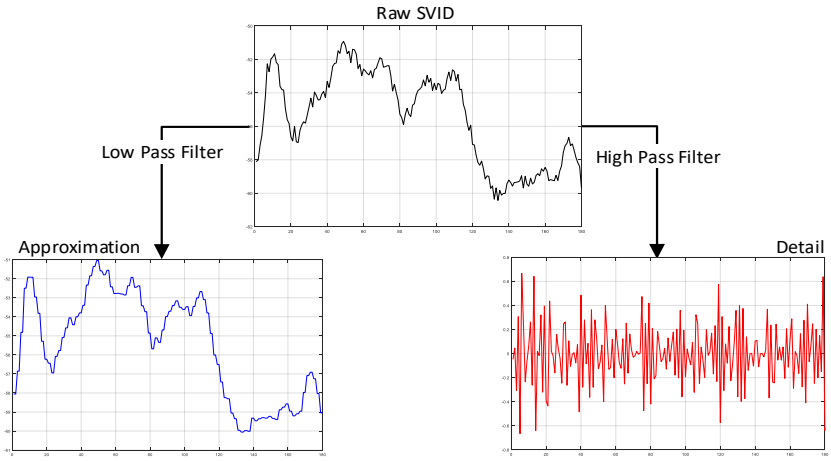


Figure 4.5: Approximation and Detail signals obtained by DWT are illustrated.

For starting a DWT analysis, the first step is to find a best-fit wavelet. The raw signal decomposition continues until a given number of levels. This iterative decomposition usually is stopped at a given number of levels or with acceptable reconstruction error. Since we run decomposition for one level, it is essential to feed a proper wavelet into the model. If the wavelet is a poor match for the signal, we may not get meaningful

information, at least not within the few calculation levels. It is therefore essential to use the most appropriate wavelet for the decomposition [37].

Several types of mother wavelets can be used for signal transformation. In fact, for the same signal, different mother wavelets leads to different decomposition results. Wavelets are characterized by a set of properties such as orthogonality, compact support, symmetry and vanishing moment [14]. This section aims at finding the best mother wavelet to capture as many information as possible from the raw signal. It happens if there exists the highest similarity between the selected mother wavelet and the raw signal. Several approaches have been employed to determine the similarity between the signal and the mother wavelet regarding qualitative and quantitative criteria [37, 152, 153].

Table 4.1: The wavelet library includes the most popular wavelet families, which are examined for selecting the best mother wavelet.

Family	**Wavelet Symbols**	**Properties**
Haar	haar	– Asymmetric – Orthogonal – Biorthogonal
Daubechies	db2, db3, db4, db5, db6, db7, db8	– Asymmetric – Orthogonal – Biorthogonal
Symlets	sym2, sym3, sym4, sym5, sym6, sym7, sym8	– Near symmetric – Orthogonal – Biorthogonal
Coiflets	coif1, coif2, coif3, coif4, coif5	– Near symmetric – Orthogonal – Biorthogonal
Biorthogonal	bior1.1, bior1.3, bior1.5, bior2.4, bior2.6, bior2.8, bior3.3, bior3.5, bior3.7, bior3.9, bior4.4, bior5.5, bior6.8	– Symmetric – Not orthogonal – Biorthogonal
Reverse biorthogonal	rbio1.1, rbio1.3, rbio1.5, rbio2.4, rbio2.6, rbio2.8, rbio3.3, rbio3.5, rbio3.7, rbio3.9, rbio4.4, rbio5.5, rbio6.8	– Symmetric – Not orthogonal – Biorthogonal
Meyer	dmey	– Symmetric – Orthogonal – Biorthogonal

Shape matching by visual inspection to select the best mother wavelet is one of the qualitative approaches. On the other hand, the core of quantitative approaches is quantifying the similarity between signal and mother wavelet methods. The example of these approaches are the Minimum Description Length (MDL) measurement [154], maximum cross-correlation coefficient criterion between the raw signal and the mother wavelet [153], and the maximum correlation coefficient between the raw signal and the reconstructed approximation components in different decomposition levels. In this thesis, the best mother wavelet is the one that provides the highest correlation coefficient between the raw signal and the reconstructed approximation components in different decomposition levels. The examined wavelets in this section come from different families as Haar, Daubechies, Symlets, Coiflets, Biorthogonal, Reverse biorthogonal and Meyer [155].

In **Table 4.1**, orthogonal filters lead to orthogonal wavelet basis functions, and the resulting wavelet transform is hence energy preserving. This property implies that the mean square error (MSE) introduced during the quantification of the DWT coefficients is equal to the MSE in the reconstructed signal.

However, in the case of biorthogonal wavelets, the basis functions are not orthogonal and thus not energy preserving. Also, symmetric wavelets are important when the aim is to build bases of regular wavelets over an interval, rather than the real axis [155]. Interested readers are referred to [155] for mathematical and statistical properties of wavelets. The detail list of wavelets examined in this section is in **Table 4.1**. **Figure 4.6** depicts a sample wavelet from each wavelet family.

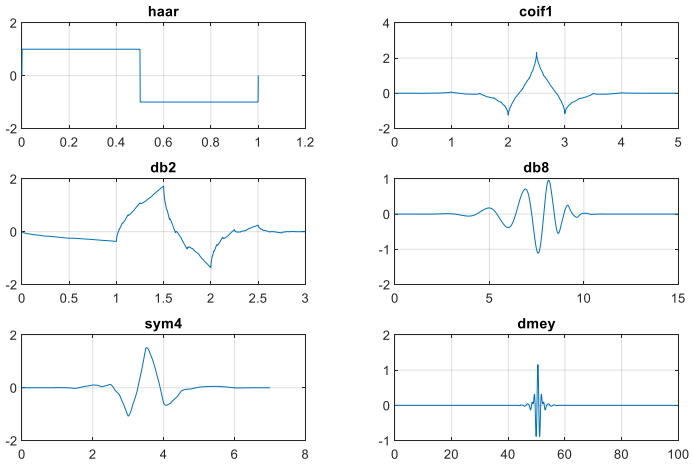


Figure 4.6: The sample wavelet functions which used for signals decomposition.

After finding the best mother wavelet for each SVID, the FDC data of each wafer k, W(k), presented as (3.1) in **Chapter 3**, is decomposed into approximation and detail components as $A(k)$ and $D(k)$, respectively.

4.2 Feature Extraction for Equipment Behavior

The principles of equipment behavior prognosis in any area are 1) to identify a set of features within the equipment sensor readings that indicate equipment behavior and 2) to prognose the equipment health deterioration and diagnose the causes of this deterioration.

Since we aim to study the macro- and micro-level variations, the variance/covariance matrix for the both of approximation and detail signals in each wafer is calculated to prevent from the recipe-related changes as well as to avoid the impact of different scales from one SVID to another. In this regard, we not only take the *variation* of each single signal into account, but we also consider the *co-variation* between signals.

In this section, we use the Pearson correlation coefficient between SVIDs as the feature indicating equipment behavior. From a physical point of view, when equipment is gradually degraded, the recipe set points are not achieved, and the controllers cannot realize this deterioration. Hence, the process compensates this deterioration by changing the level of specific SVIDs (either less or more than set value). Accordingly, the correlation between these SVIDs changes (either stronger or weaker). From a statistical point of view, the collected data from sensors are often cross-correlated. Consequently, one should consider not only the variation of every variable but also the co-variation among variables [2].

Based on this idea, the macro-level and micro-level variations are investigated by following the behavior of correlation matrix on approximation (i.e., $A(k)$) and detail (i.e., $D(k)$) components, respectively. Therefore, the correlation matrix of approximation and detail components of wafer k is represented as $\widehat{\rho}_A(k)$ and $\widehat{\rho}_D(k)$.

$$A(k) = \begin{bmatrix} a_{1,1}^{(k)} & a_{1,2}^{(k)} & \cdots & a_{1,J}^{(k)} \\ a_{2,1}^{(k)} & a_{2,2}^{(k)} & \cdots & a_{2,J}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_k,1}^{(k)} & a_{n_k,2}^{(k)} & \cdots & a_{n_k,J}^{(k)} \end{bmatrix}_{n_k \times J}, \text{ and}$$

$$D(k) = \begin{bmatrix} d_{1,1}^{(k)} & d_{1,2}^{(k)} & \cdots & d_{1,J}^{(k)} \\ d_{2,1}^{(k)} & d_{2,2}^{(k)} & \cdots & d_{2,J}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n_k,1}^{(k)} & d_{n_k,2}^{(k)} & \cdots & d_{n_k,J}^{(k)} \end{bmatrix}_{n_k \times J}, \quad \forall k \tag{4.5}$$

$$\widehat{\rho}_A(k) = \begin{bmatrix} ra_{1,1}^{(k)} & ra_{1,2}^{(k)} & \cdots & ra_{1,J}^{(k)} \\ ra_{2,1}^{(k)} & ra_{2,2}^{(k)} & \cdots & ra_{2,J}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ ra_{J,1}^{(k)} & ra_{J,2}^{(k)} & \cdots & ra_{J,J}^{(k)} \end{bmatrix}_{J \times J}, \text{ and}$$

$$\widehat{\rho}_D(k) = \begin{bmatrix} rd_{1,1}^{(k)} & rd_{1,2}^{(k)} & \cdots & rd_{1,J}^{(k)} \\ rd_{2,1}^{(k)} & rd_{2,2}^{(k)} & \cdots & rd_{2,J}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ rd_{J,1}^{(k)} & rd_{J,2}^{(k)} & \cdots & rd_{J,J}^{(k)} \end{bmatrix}_{J \times J}, \quad \forall k \tag{4.6}$$

where $ra_{j,j'}^{(k)}$ and $rd_{j,j'}^{(k)}$ are the correlation coefficient between SVIDs j and j' ($j, j' = 1, ..., J$), respectively in approximation and detail components.

The approximation and detail correlation matrices are calculated for each wafer through the FDC data-set. Finally, the determinant value is calculated to transform

the correlation matrices into single values. Consequently, the determinants of the approximation and detail correlation matrices are considered as the macro-level health index (i.e., $HI_A^{(k)}$) and micro-level health index (i.e., $HI_D^{(k)}$) per wafer k.

$$HI_A^{(k)} = \det(\widehat{\rho}_A(k)), \qquad \forall k \tag{4.7}$$

$$HI_D^{(k)} = \det(\widehat{\rho}_D(k)). \qquad \forall k \tag{4.8}$$

By drawing these determinant values, we expect to find the equipment behavior pattern, which could be decreasing or increasing. In case of decreasing (increasing) trend, the determinant values start from 1(0) when equipment usually is working, and they end 0(1) when equipment is completely deteriorated. It is worth mentioning that this index is mainly used to represent the behavior and the condition of the equipment. From an engineering point of view, when the equipment functions normally, the correlation between SVIDs should be stable without unexpected excursions. These indices effectively helps to model and monitor the behavioral changes in the relationship between SVIDs. Accordingly, a high index value does not imply the equipment is healthier and the lower index value does not necessarily indicate the equipment is complete deteriorated or ill-functioned. These index indicates when the equipment health might be jeopardized in case of unexpected faults or gradual deterioration.

4.3 Equipment Deterioration Identification

After extracting HI for each wafer, the next step is to model the deterioration over the time regarding HI with deterioration-related SVIDs. Since macro-level and micro-level variations happen only in certain SVIDs, following the determinant of correlation matrices using all the SVIDs would easily cancel out the deterioration effect. Therefore, a searching algorithm is developed to identify the contributing SVIDs to the expected deterioration trend. These SVIDs are finally classified as the deterioration causes. The progression of macro-level and micro-level variations through the wafers can be considered as the deterioration of the equipment. These progressions can be observed as the change of the correlation between contributing SVIDs (i.e., deterioration causes). Since correlation and determinant have inverse relation ($det = 1 - Corr^2$), when correlation becomes stronger (weaker), the determinant value must decrease (increase) gradually. Therefore, the equipment health curve indicates the deterioration if it has a significantly negative or positive slope. The proposed searching algorithm should look for those SVIDs that create a significant deterioration trend regarding two characteristics: I) slope significance and II) Goodness-of-Fit (GoF). For evaluating the two properties of a deterioration trend, the ordinary regression analysis is employed to estimate the deterioration over time. As shown in (4.9) and (4.10), the coefficients, α_A and α_D, are used to evaluate the slope significance of the deterioration trends, and β_A and β_D are the intercepts for the regression lines over the approximation and detail signals, respectively. The adjusted R^2s of the regression model are used to evaluate the GoF of a deterioration

trend and expressed as R_A^2 and R_D^2.

$$y_A = \alpha_A k + \beta_A, \tag{4.9}$$

$$y_D = \alpha_D k + \beta_D. \tag{4.10}$$

For optimizing the two characteristics at the same time, the integrated objectives for evaluating the deterioration trends modeled in (4.9) and (4.10) are defined as:

$$F(S_A) = |\log \alpha_A| \times (1 - R_A^2), \tag{4.11}$$

$$F(S_D) = |\log \alpha_D| \times (1 - R_D^2). \tag{4.12}$$

where $F(S_A)$ and $F(S_D)$ are the objective values for macro-level driven set of SVIDs, S_A, and micro-level driven set of SVIDs, S_D. $|\log \alpha_A|$ and $|\log \alpha_D|$ are used to indicate whether the slope is significant in terms of downward/upward deterioration trends. As explained earlier, when equipment is gradually deteriorated, the correlation between these SVIDs become stronger or weaker. Actually, becoming stronger or weaker are considered in terms of any increase or decrease in the value of the correlation and no matter how much these changes are. These two cases are separately defined as follows:

- *Stronger Correlation.* Any increase in the correlation matrices $\widehat{\rho}_A$ and $\widehat{\rho}_D$ leads to decrease in the value of both macro-level HI, HI_A, and micro-level HI, HI_D. Accordingly, the final deterioration trend becomes decreasing and consequently the slope values α_A and α_D become negative. In conclusion, the more the negative values of α_A and α_D, the minimum the values of $|\log \alpha_A|$ and $|\log \alpha_D|$.
- *Weaker Correlation.* Any decrease in the correlation matrices $\widehat{\rho}_A$ and $\widehat{\rho}_D$ leads to an increase in the value of both macro-level HI, HI_A, and micro-level HI, HI_D. Accordingly, the final deterioration trend becomes increasing and consequently the slope values α_A and α_D become positive. In conclusion, the more the positive values of α_A and α_D, the minimum the values of $|\log \alpha_A|$ and $|\log \alpha_D|$.

It is worth mentioning that the objective functions (4.11) and (4.12) work correctly for both decreasing and increasing trends. It means that minimizing the objective functions aims at finding the significant decreasing trend and the significant increasing ones simultaneously. To cope with this issue and not confounding the increasing and decreasing trends, the simplest way is penalizing the increasing trends (i.e., multiplying the objectives with a significant positive value if $\alpha_A, \alpha_D > 0$) when looking for the decreasing trends and vice versa, penalizing the decreasing trends (i.e., multiplying the objectives with a big positive value if $\alpha_A, \alpha_D < 0$) when looking for the increasing trends. Hence, the more the smaller values of $|\log \alpha_A|$ and $|\log \alpha_D|$, the more significant the deterioration trends. In the same direction as $|\log \alpha_A|$ and $|\log \alpha_D|$, minimizing the terms $(1 - R_A^2)$ and $(1 - R_D^2)$ leads to higher GoF at both increasing and decreasing trends. Therefore, deterioration trends with the minimal values of $F(S_A)$ and $F(S_D)$ are the ones we are looking for.

In summary, the sets of SVIDs (i.e., S_A and S_D) which minimizes (4.11) and (4.12) should be searched systematically. One way might be to enumerate all the combinations over the J SVIDs and identify the set of SVIDs with the minimum objective value. However, the complete enumeration is computationally expensive (if not impossible). To cope with this issue, a Stepwise Deterioration Searching Algorithm (SDSA) is developed as a constructive iterative algorithm that starts with an empty set of SVIDs. Each iteration of the SDSA contains two operators as forward-selection and backward-elimination. Forward-selection happens when a SVID is decided to be added to the set, and backward-elimination is for eliminating a SVID from the selection set. These operators are triggered if and only if adding or removing a SVID will make the deterioration trend more significant, i.e., make the objective values smaller. The SDSA terminates if none of forward-selection and backward-elimination can happen. In summary, we investigate three possibilities when a SVID candidate is being evaluated:

- Extension: the forward-selection happens and the SVID candidate is added to the selection list, and the list is extended,
- Substitution: both forward-selection and backward-elimination happen simultaneously, and the candidate SVID is substituted by one of the current SVIDs in the list,
- No-change: If none of extension and substitution happens. In this case, the current list of SVIDs does not change.

Let S_J denotes the set of all SVIDs and redefine S_A and S_D as the set of contributing SVIDs to the deterioration trend over the approximation signals, and the set of contributing SVIDs to the deterioration trend over the detail signals. These sets are defined as (4.13)-(4.15), respectively.

$$S_J = \{j | j = 1, ..., J\}, \tag{4.13}$$

$$S_A = \{j | j \in S_J, F(S_A) = \min_{S_J} F\}, \tag{4.14}$$

$$S_D = \{j | j \in S_J, F(S_D) = \min_{S_J} F\}, \tag{4.15}$$

The SDSA starts from the first SVID (i.e., SVID1) in S_j. For all the pairs of SVID1 and any other SVID in S_j (where $j \neq 1$), the HIs based on two SVIDs are calculated, the deterioration trend is fitted and the corresponding objective values are estimated. Then, the pair with the minimum objective value is formed and, the added SVID is removed from S_j. Now, the algorithm looks for the third SVID to add to the list. Whenever an SVID is tested, all extension and substitution operations are performed and the best combination is formed. This iterative mechanism continues until the current SVID list remains unchanged (i.e., *No-change* happens). Finally, the current list is archived as the list started from SVID1. Afterward, the algorithm restarts from second SVID and so on. This restart repeats J times (total number of SVIDs is equal to J) and the created lists are archived. Finally, the list with the minimum objective value is returned as the final contributing SVIDs. **Figure 4.7** and **Figure 4.8** depict the Pseudo code and

the detailed flowchart of the proposed statistical data-driven equipment deterioration modeling and monitoring, respectively. The notations in **Figure 4.7** are for finding the decreasing trend from the approximation signals. The settings can be easily changed for other types of trends (i.e., from detail signals or increasing trends).

Finally, after finding the contributing SVIDs to the deterioration trend over the approximation and the detail components of the signal, the final deterioration trends can be extracted and utilized for further monitoring. In this regard, once a new wafer is produced in the equipment, the FDC data of the contributing SVIDs are first decomposed into two components, which are then summarized into health indices. The deterioration trends are updated by the new calculated HIs.

As a summary, a data-driven approach for the equipment behavior prognosis using the FDC data in the semiconductor industry is developed. The proposed method aims at modeling the equipment deterioration. An essential result is finding and reporting the contributing SVIDs to the equipment deterioration, which are called as deterioration roots. It is worthy to mention that the deterioration causes are not necessarily equivalent to the extracted fault roots identified in **Chapter 3**. Accordingly, the deterioration symptoms might differ from the failure symptoms, and this difference necessitates the development of an independent prognostic approach. An important application of the proposed deterioration models is to utilize these trends to monitor the equipment condition and trigger an alarm whenever an unexpected excursion is detected. The monitoring mechanism based on properties of the extracted deterioration models will be discussed in **Chapter 6** as one of the future research directions.

input : FDC Data
output: Deterioration Trend

TrendType = DECREASING // Can be also INCREASING //
for $j \in S_J$ **do** // Algorithm starts from each SVID //
  $S_A = \{\}$;
  $F^C = G$; // F^C: Current objective value, $G \gg 0$ //
  $S_A = S_A \cup j$;
  for $j' \in S_J : j' \neq j$ **do**
    calculate HI_A and **fit** y_A
    calculate $F(j, j')$ for SVID pair (j, j')
    if $\alpha_A, \alpha_D > 0$ **then** // $\alpha_A, \alpha_D < 0$ if INCREASING //
      $F(j, j') = F(j, j') \times G$; // Penalty of INCREASING //
    end
    archive F
  end
  $S_A = \left\{ S_A \cup j'' | F(j, j'') = \min_{j'} F(j, j') \right\}$;
  save current objective function
  for $j' \in S_J : j' \notin S_A$ **do**
    do *Extension*:
      add j' to the list and **calculate** $F(S_A \cup j')$
      archive $F(S_A \cup j')$
    end *Extension*
    do *Substitution* N times $(N = |S_A|)$:
      replace j' with $j''(j'' \in S_A)$ and
      calculate HI_A and **fit** y_A for $S_A - j'' \cup j'$
      calculate $F(S_A - j'' \cup j')$
      archive $F(S_A - j'' \cup j')$
    end *Substitution*
  end
  if $F(Best\ Combination)|_{j'} < F^C$ **then**
    $F^C = F$(Best Combination);
    $S_A = S_A \cup j'$;
  else
    terminate the algorithm and **return** S_A for SVID j
      // *No-change* in the list //
  end
  archive S_A and F^C for starting SVID j
end
return an SVID list with the minimum objective value

Figure 4.7: This figure shows the pseudo code of the proposed equipment deterioration modeling approach from the approximation signal.

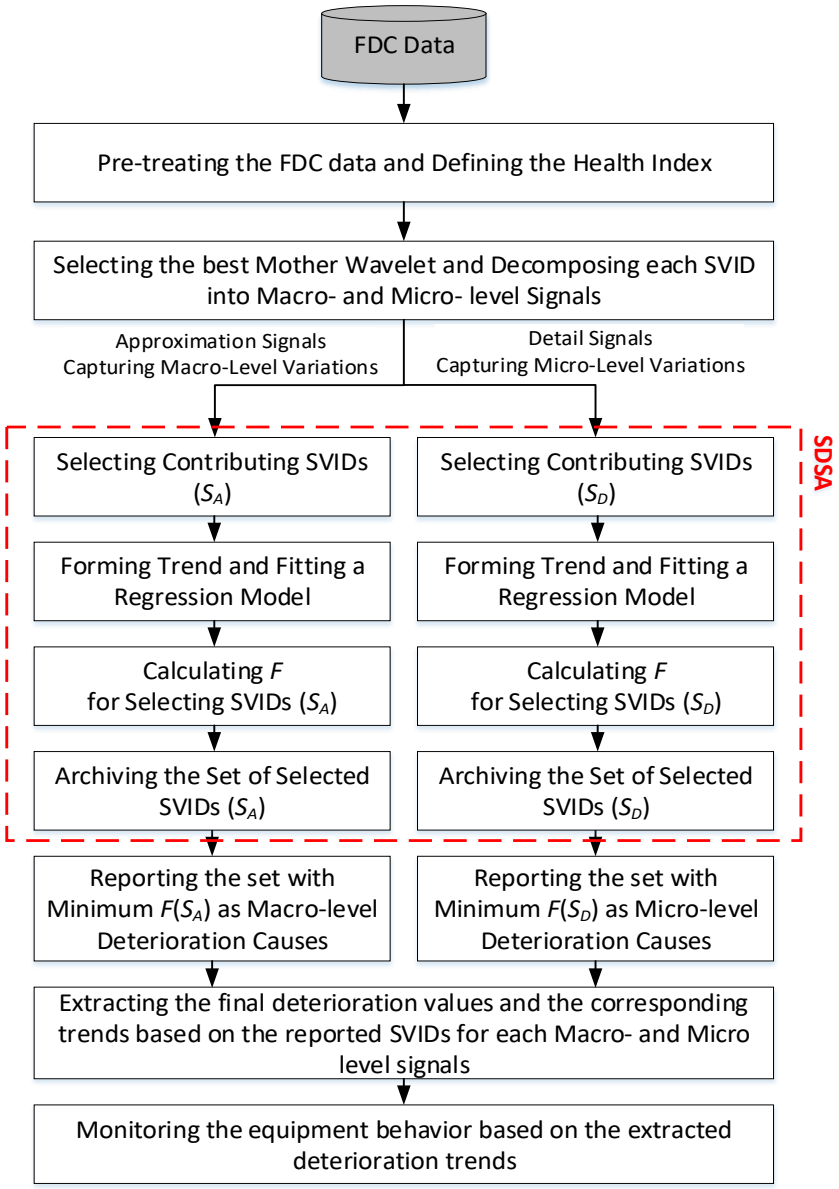


Figure 4.8: This detailed flowchart depicts the proposed equipment deterioration modeling and monitoring approach.

Chapter 5

Computational Results

In this chapter, the proposed equipment diagnosis and prognosis approaches are validated with the practical data-set from the local semiconductor manufacturer. The proposed algorithms are compiled in MATLAB software and the experiments are executed on a computer of 3.4 GHz quad core CPU with 4 GB memory. The FDC data from an etching tool are explained in **Section 5.1**. The results of the proposed *Equipment Failure Diagnosis* and *Equipment Deterioration Modeling* are discussed in **Section 5.2** and **Section 5.3**, respectively.

5.1 FDC Data Brief

Starting with a uniformly doped silicon wafer, it requires hundreds of sequential process operations to fabricate an IC chip. The most important process operations in the semiconductor fabrication are shown in **Figure 5.1** [156]. In deposition, a thin film layer that will form the wiring, transistors and other components is deposited on the wafer. The thin film is then coated with the photoresist. The circuit pattern of the photomask (reticle) is then projected onto the photoresist using photolithography technology. The developed photoresist is used as a mask and the etching process can remove the unwanted areas.

The FDC data used to validate the proposed approaches comes from an etching process in the semiconductor fabrication. **Figure 5.2** depicts the FDC system for an etching process [157]. The etch system shapes the thin film into the desired patterns using liquid chemicals, reaction gases or ion chemical reaction. It is frequently used in the IC production lines. Etching as a front-end process occurs after a photoresist film is patterned onto the wafer. To keep the pattern on the wafer surface, unwanted areas are removed by the etching process. Within an etch chamber, highly reactive plasma gasses react with the wafer to remove the film where the pattern leaves its expos. Once complete, the wafer has a dielectric film with a pattern ready to receive tungsten or copper, which serves as an interconnection to the next layer. It is vital that the etching process is correctly performed because it is difficult to restore the etched areas. As a result, improperly etched wafers are most likely scrapped, and a high waste cost is

induced. Accordingly, monitoring the behavior of the etch equipment and preventing it from any failure are required since they significantly affect the product quality and production yield.

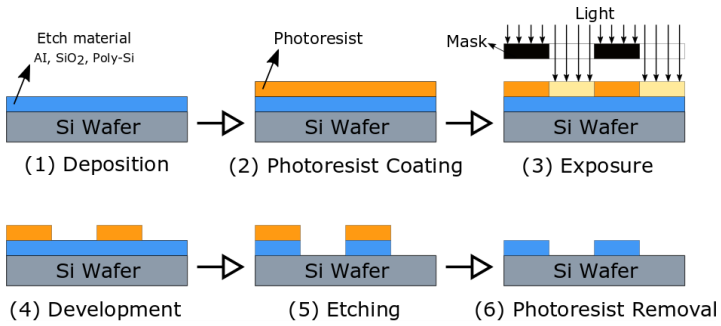


Figure 5.1: Main operations in the semiconductor manufacturing.

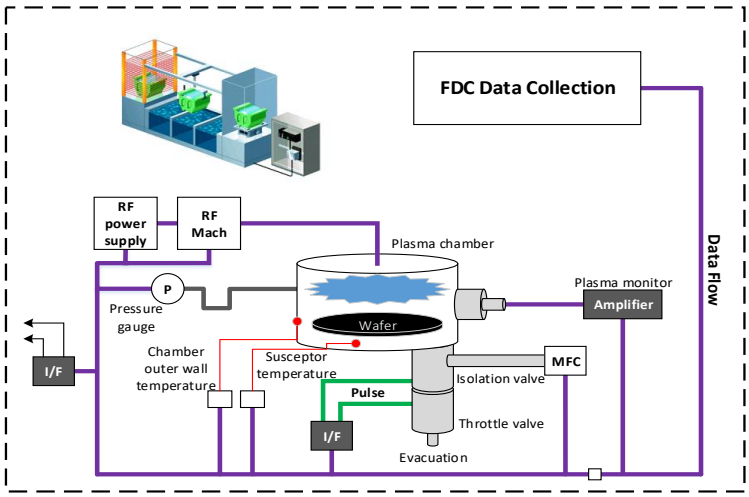


Figure 5.2: The FDC system for an etching process.

In the FDC data, there are 760 wafers (264,340 observations in total) processed under eight different recipes. Indeed, these eight recipes belong to a product family that utilize common equipment and do not differ too much in term of processing characteristics. There are some differences in the level of parameters that need to be considered. Thirty-one sensor readings, i.e., SVIDs, were collected simultaneously to keep track of the equipment signals and process states during the process. The SVIDs are listed in **Table 5.1** with basic descriptions. Due to the confidentiality, the full SVID names cannot be disclosed. It is known in advance that an unexpected Corrective Maintenance (CM) event was performed after the 480^{th} wafer to fix a process drift. Unfortunately, the fault

was not completely resolved, and the process remained unstable until the end of the data-set where a periodic Preventative Maintenance (PM) was executed. Accordingly, the first 479 wafers are normal wafers, and the rest 281 wafers are faulty (abnormal) wafers as the priori information. It is worth mentioning that the proposed equipment failure diagnosis approach works on the whole data-set while the equipment behavior prognosis approach is conducted only on normal wafers.

Table 5.1: The 31 SVIDs with basic descriptions.

SVID	Description	SVID	Description	SVID	Description
V_1	RF_01	V_{12}	RF_08	V_{22}	Gas flow 11
V_2	RF_02	V_{13}	RF_09	V_{23}	Gas flow 12
V_3	RF_03	V_{14}	Temperature 01	V_{24}	Gas flow 13
V_4	RF_04	V_{15}	Pressure 01	V_{25}	Pressure 02
V_5	Gas flow 01	V_{16}	Gas flow 05	V_{26}	Pressure 03
V_6	Gas flow 02	V_{17}	Gas flow 06	V_{27}	Pressure 04
V_7	Gas flow 03	V_{18}	Gas flow 07	V_{28}	RF_10
V_8	Gas flow 04	V_{19}	Gas flow 08	V_{29}	RF_11
V_9	RF_05	V_{20}	Gas flow 09	V_{30}	Temperature 02
V_{10}	RF_06	V_{21}	Gas flow 10	V_{31}	Temperature 03
V_{11}	RF_07				

5.2 Results of the Diagnosis Approach

The above FDC data collected are used to test of the proposed data-driven equipment failure diagnosis. The results are discussed in the following subsections. Without loss of generality, all data were firstly standardized with zero mean and unit variance for further implementation.

5.2.1 Stage 1: Equipment Anomaly Detection

This stage was developed as an early step to detect if there is any abnormality in the data using the supervised SVM method. To start training the SVM model, the first 479 wafers (167,522 observations) are labeled as normal while the rest are marked as abnormal.

It is obvious that higher performance for Stage 1 can be obtained by better tuning of its parameters. These parameters are categorized into SVM parameters and GA parameters. As mentioned in **Section 3.2**, four kernel functions Linear, Polynomial, RBF, and Sigmoid, have been tried and their parameters are listed in **Table 5.2**. The level of these parameters depends on the data-set and differs from a data-set to another. Accordingly, determining the ranges of parameters is a trial and error process. It is worth mentioning that the range of parameters must be neither very tight that may prevent us from optimality, nor very vast that makes GA very time-consuming to find the optimal values.

GA parameters include population size (NPop), maximum iterations (Itr_{max}), crossover rate (C_{Rate}), and mutation rate (M_{Rate}). These parameters must be tuned to reach the

highest performance of GA. These parameters are tuned using the full factorial approach [158]. Such an experiment allows us to study the effect of each parameter on the response variable (i.e., the fitness value in GA), as well as the effects of interactions between parameters on the response variable. Considering different levels for each GA parameters and applying a full factorial design, the levels 100, 100, 75%, and 30% are tuned for $NPop$, Itr_{max}, C_{Rate}, and M_{Rate} parameters, respectively.

Table 5.2: Kernel functions.

Kernel	**Formulation**	**Parameters & tuning range**
Linear	$x^T y + b$	– Penalty $C \in [1, 128]$ – Constant $b \in [-30, 30]$
Polynomial	$(ax^T y + b)^d$	– Penalty $C \in [1, 128]$ – Constant $a \in [0, 15]$ – Constant $b \in [-30, 30]$ – Constant $d \in [2, 10]$
RBF	$\exp(-\frac{\|x-y\|^2}{2\sigma^2})$	– Penalty $C \in [1, 128]$ – Constant $\sigma \in [1, 20]$
Sigmoid	$\tanh(ax^T y + b)$	– Penalty $C \in [1, 128]$ – Constant $a \in [0, 15]$ – Constant $b \in [-30, 30]$

Table 5.3: Kernels accuracy and parameters

Kernel	**Accuracy(%)**	**Error***(%)	**Parameters**
Linear	62.00%	38.00%	$C = 51, b = 1$
Polynomial	57.26%	42.74%	$C = 86, a = 4.74, b = 1, d = 7$
RBF	**97.31%**	02.69%	$C = 17, \sigma = 13.7$
Sigmoid	86.73%	15.27%	$C = 103, a = 1.5, b = -25$

*Error $= 1 - Accuracy$

Considering the different subsets of the data and applying 3-fold cross-validation [131] in the SVM algorithm (same results were obtained for 5, 7 and 10 number of folds), the optimal parameters obtained by GA and the accuracy of different kernels are summarized in **Table 5.3**. It can be seen that the RBF kernel provides the highest accuracy equal to 97.31%. The FDC data passed through the trained SVM model are then separated into two classes: *normal* and *abnormal* observations.

5.2.2 Stage 2: Automatic Fault Fingerprint Extraction

As the faults can be identified in the aforementioned SVM classifier, the root causes to the faults shall be extracted for further actions. The four steps to extract the fault fingerprints described in **Section 3.3** are performed over the data and the results are discussed in the following paragraphs.

(a) **Process Dynamic Decomposition**

This stage was to cluster the hundred or even thousands of normal observations (detected by Stage 1) into a smaller number of clusters to decompose the process dynamics and not to detect them as a fault. For this purpose, two well-known unsupervised clustering algorithms are employed as Self-Organizing Map (SOM) [159] and K-means algorithms.

Accordingly, the SOM and K-means algorithms are employed on the collected normal observations from the Stage 1 with different number of clusters to decompose the process dynamics into a number of clusters. In order to determine the number of clusters in the clustering algorithms (i.e., number of clusters), two most-utilized approaches in the literature are employed as *Elbow* method and *Internal Evaluation* [160]. Most generally, the *Elbow* method is more to identify the optimum number of clusters in a clustering algorithm while *Internal Evaluation* method can be applied also to compare the performance of different clustering algorithms.

- ***Elbow Method***

The *Elbow* method is to interpret and validate the consistency within the cluster analysis designed to find the appropriate number of clusters in a data-set. This method looks at the percentage of variance explained as a function of the number of clusters. Therefore, the optimum number is a number of clusters so that adding another cluster does not give much better modeling of the data. The *Elbow* method is performed for different numbers of clusters (i.e., $K = 2, ..., \Sigma_c N_{G,c}$) for both K-means and SOM clustering algorithms. For the K-means clustering, the Euclidian Distance and the Correlation between the data points have been considered as the criteria for clustering the data.

Figure 5.3 shows the scree plot of the percentage of variance explained as a function of the number of clusters for both K-means and SOM clustering algorithms. Since having more than 12 clusters does not explain more variance percentage in the both algorithms, **Figure 5.3** only shows the results for 2 to 12 number of clusters. It is evident that having 8 numbers of clusters does not add more explanation for the variance of the data in the both algorithms. Therefore, 7 numbers of clusters provide enough explanation of the variance. In addition, it can be condensed that K-means with correlation as the similarity function provides better variance explanations comparing to the K-means with Euclidian distance as the similarity function.

- ***Internal Evaluation***

In the *Internal Evaluation*, clustering algorithms are performed with different number of clusters and their performance is evaluated based on the input data that they have clustered. The best score is assigned to the algorithm that produces clusters with high similarity within a cluster and low similarity between clusters. The *DaviesBouldin* (DB) [161] and the *Dunn* (DN) [162] indices are the two main indices for internal evaluation.

Figure 5.3: Percentage of variance explained given different numbers of clusters.

The DB and DN indices are presented as Equations (5.1) and (5.2) to evaluate the performance of the clustering algorithms. These indices are called internal criteria since they are calculated based on the same input data that the clustering algorithms have been performed on it. From Equation (5.1), it can be seen that the DB index works with the average distance and there is not enough information about the variance of the data in each cluster. As a complementary approach, the DN index can evaluate the clustering result by taking the variance of the points into account [162]. In **Figure 5.3**, it was concluded that considering more than 11 clusters does not give much better modeling of the data. Therefore, the *Internal Evaluation* method is performed on the K-means and the SOM clustering algorithms with 5 to 11 number of clusters. In addition, K-means clustering algorithm is run by the correlation as the similarity function.

$$DB = \frac{1}{NC}\sum_{\omega=1}^{NC}\max_{\omega'\neq\omega}\left(\frac{\alpha_\omega + \alpha_{\omega'}}{d(C_\omega, C_{\omega'})}\right), \tag{5.1}$$

where NC is the number of clusters, C_ω is the centroid of cluster ω, α_ω is the average distance of all elements in cluster ω to centroid C_ω, and $d(C_\omega, C_{\omega'})$ is the distance between centroid C_ω and $C_{\omega'}$. The algorithm that produces clusters with low intra-cluster distances (high intra-cluster similarity) and high inter-cluster distances (low inter-cluster similarity) will have a low DB index. Therefore, the smaller DB index leads to be better clustering result.

The DN index of Equation (5.2) aims at identifying the dense and well-separated clusters. It is defined as the ratio of the minimum inter-cluster distance (highest intra-cluster similarity) to the maximum intra-cluster distance (lowest inter-cluster similarity).

$$DN = \frac{\min\limits_{1\leq\omega'\leq\omega'\leq K} d(\omega,\omega')}{\max\limits_{1\leq\omega''\leq K} d'(\omega'')}, \tag{5.2}$$

where $d(\omega, \omega')$ represents the distance between clusters ω and ω', and $d'(\omega'')$ measures the intra-cluster distance of cluster ω''. The inter-cluster distance $d(\omega, \omega')$ between two clusters may be any number of distance measures, such as the distance between the centroid of the clusters. Similarly, the intra-cluster distance $d'(\omega'')$ may be measured in a variety ways, such as the maximal distance between any pair of points in cluster ω''. Since the internal criterion seeks the clusters with high intra-cluster similarity and low inter-cluster similarity, the algorithms that produce clusters with high DN index are more desirable.

For the SOM clustering algorithm, the Euclidian distance is considered as the dissimilarity metric, and for the K-means clustering algorithm, both Euclidian distance and (1- Correlation) are considered as the dissimilarity measures. **Table 5.4** shows the *DB* and *DN* indices under different numbers of clusters. It can be seen that employing the K-means clustering algorithm with the correlation similarity function in all the numbers of clusters provides better clustering comparing to the K-means and the SOM clustering algorithm with Euclidian distance as the dissimilarity function. Based on **Table 5.4**, considering 7 numbers of clusters provides lower and higher values of *DB* and *DN* indices, respectively. This result supports the results of the *Elbow* method in **Figure 5.4**. On the other hand, by comparing the *DB* and *DN* values for the SOM and the K-means algorithms, it can be concluded that not only the K-means outperforms the SOM utilizing the Euclidian distance, but also K-means with correlation function provides better clustering comparing to the K-means with Euclidian distance function. Therefore, the superiority of the K-means algorithm is proved in comparison to the SOM algorithm. In addition, utilizing the correlation as the similarity function provides better process dynamic decomposition.

By decomposing the dynamics of a single wafer process run, i.e., a batch of 350 observations in average, into 10 sequential deciles, **Figure 5.4** shows that observations from the ramp-up and the ramp-down steps of the wafer process have been grouped into mainly two clusters (i.e., clusters 1 and 4). While the middle of the process containing the main etching steps (i.e., deciles 3 to 8), is distributed dispersedly in different clusters, it can be learned intuitively that the main etching steps are very non-stationary and compose of interchanging upward and downward trends with different distributions. Based on **Figure 5.4**, **Figure 5.5** further illustrates the inter-cluster borders across the wafer processing time.

(b) **Abnormal Data Categorization**

The process dynamic in each cluster, which contains only normal observations, is then summarized by a reduced PCA model. Each abnormal observation, filtered out from the first stage, is then projected into all the PCA models and classified as Out-Of-Control (OOC) by checking if its D and Q statistics exceed the corresponding control limits.

Table 5.4: DB and DN indices for evaluating the SOM and K-means.

Algorithm	**Index**	**Number of Clusters**						
		5	6	7	8	9	10	11
SOM	*DB*	0.64	0.77	0.62	0.79	0.81	0.86	0.88
	DN	0.46	0.62	0.76	0.79	0.67	0.70	0.65
K-means (Corr.)	*DB*	0.79	0.74	0.55	0.62	0.74	0.89	0.87
	DN	0.67	0.75	0.82	0.79	0.72	0.63	0.64
K-means (Euclidian)	*DB*	0.69	0.73	0.61	0.77	0.82	0.81	0.87
	DN	0.51	0.66	0.78	0.64	0.71	0.66	0.53

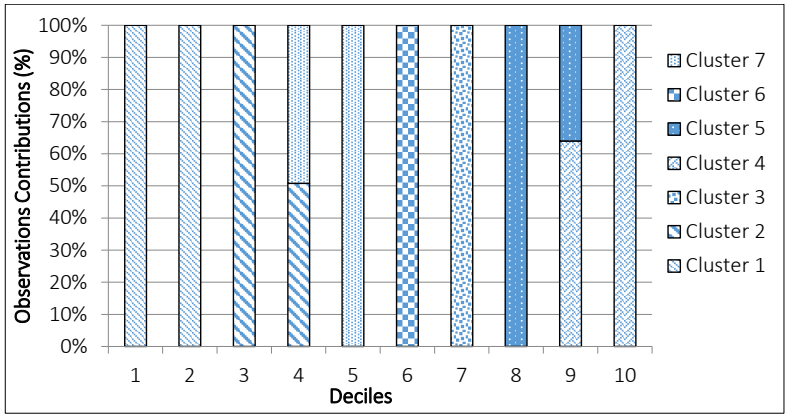


Figure 5.4: Division of the wafer process in 10 consecutive deciles and the distribution of the 7 clusters.

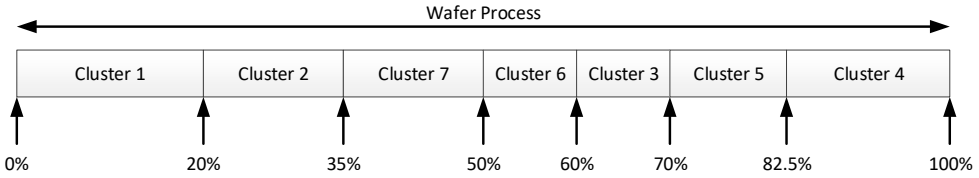


Figure 5.5: The cluster distribution within a wafer process run.

Figure 5.6 to **Figure 5.9** shows the control charts of the D and Q statistics after projecting all the abnormal observations into the PCA models built based on the clusters 1, 6, 3, and 4 (i.e., the clusters from the beginning, the middle and the end of the process to cover the duration of the whole process). The red line indicates the control limit in all control charts. From **Figure 5.6** to **Figure 5.9**, it can be interpreted that the middle of the wafer process (i.e., clusters 6 and 3) consists of a higher number of OOC observations in comparison with the beginning (i.e., cluster

1) and the end (i.e., cluster 4) of the wafer process. By the other words, the main etching steps produce more abnormality comparing to ramp-up and ramp-down steps.

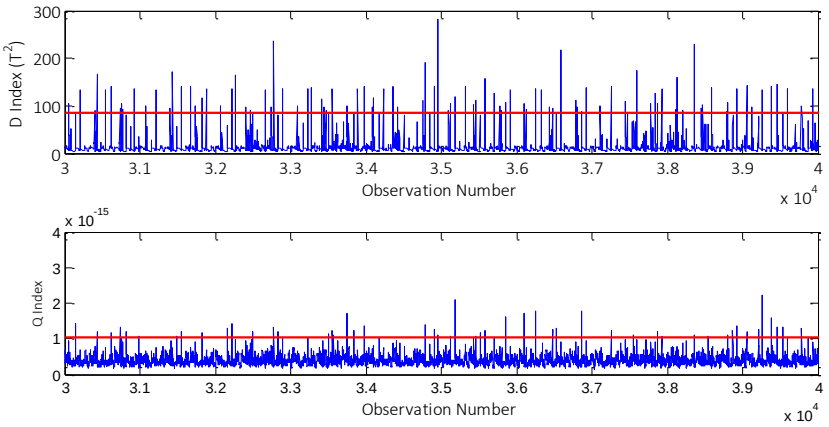


Figure 5.6: Control charts of the D and Q statistics for cluster 1.

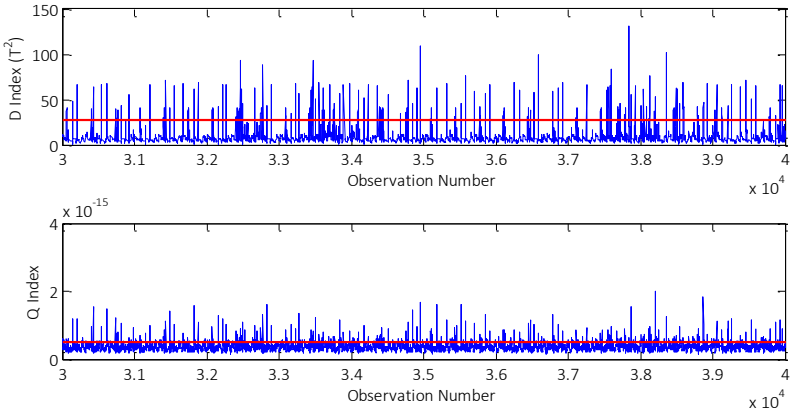


Figure 5.7: Control charts of the D and Q statistics for cluster 6.

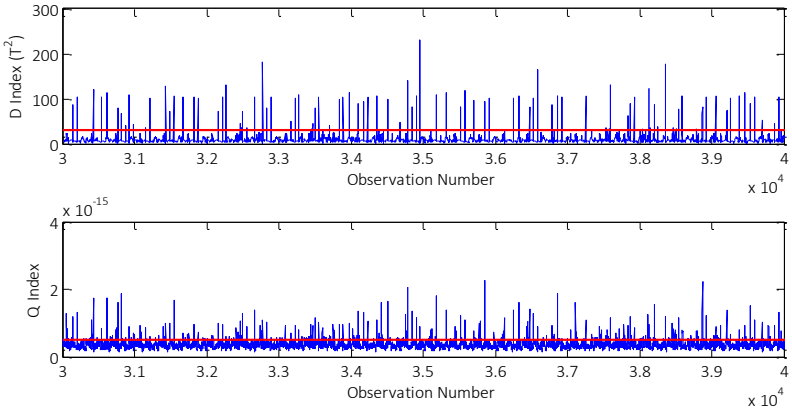


Figure 5.8: Control charts of the D and Q statistics for cluster 3.

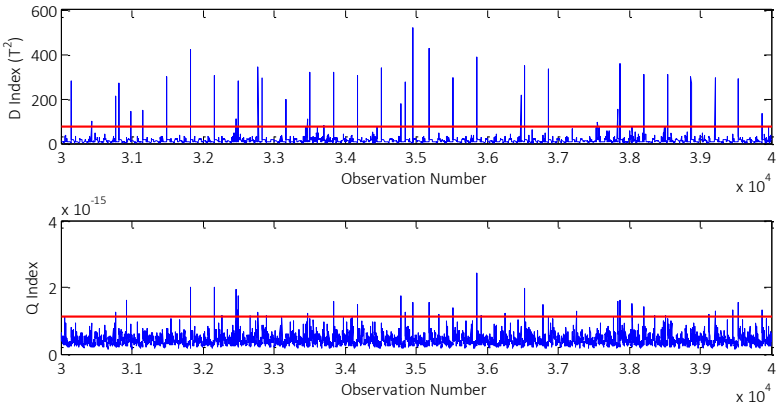


Figure 5.9: Control charts of the D and Q statistics for cluster 4.

(c) Fault Fingerprint Extraction

Given that observation is classified as OOC from one PCA model, the contribution values from all the SVIDs can be obtained by calculating Equations (3.15) and (3.17) for the D and Q statistics, respectively. **Figure 5.10** shows the contribution plots for two samples of OOC. The main idea of the Stage 2-3 is to cluster these contribution plots to find the specific patterns. These contributions thus form a specific pattern and can be interpreted as one fault type. All the OOC observations after projecting the abnormal data into all PCA models are collected, and their contribution values are calculated. The K-means clustering algorithm is then applied again to group these contribution values into different patterns.

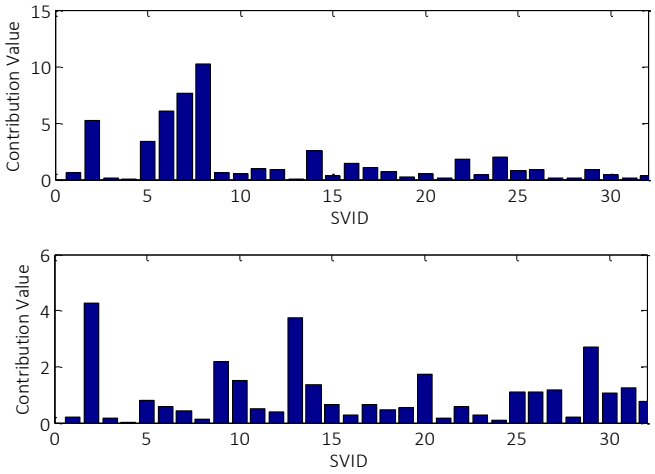


Figure 5.10: Contribution plots for two samples of OOCs.

To find the main fault patterns from the contribution values, the K-means algorithm with the correlation between points as the similarity function was performed with different number of clusters, $K = 1, ..., \text{No. of OOCs}$. By applying the *Elbow* method, it was shown that 3 numbers of clusters provide better variance explanation of the OOC contribution plots. Besides to *Elbow* method, the *DB* and *DN* indices were also calculated to identify the number of clusters ranging from 2 to 5. In addition, the K-means is separately conducted on the OOC observation sets from D and Q statistics. Similar to **Table 5.4**, **Table 5.5** presents the DB and DN indices for different numbers of clusters for the K-means algorithm, separately for D and Q statistics. It can be shown that, for each of D and Q statistics, three fault fingerprints can be found from the contribution values of the OOC observations.

Table 5.5: DB and DN indices for evaluating the K-means clustering on the extraction of fault fingerprints.

Index	**Control Chart**	**Number of Clusters**			
		2	3	4	5
BD	*D*	0.72	0.51	0.63	0.84
	Q	0.82	0.63	0.77	0.79
DN	*D*	0.75	0.92	0.82	0.64
	Q	0.64	0.86	0.71	0.73

(d) **Fault Root Summarization**

After clustering the OOC contribution values for each of D and Q control charts and creating fault fingerprints, the final step of the proposed method is to extract the fault roots of each fault fingerprints. As explained in **Section 3.3.4**, the proposed control limits Equations (3.20) and (3.21) are calculated for the

contribution values of each fault fingerprint extracted from D and Q statistics, respectively. Accordingly, we first need to provide a unique contribution plot for each fault fingerprint. For this aim, the contribution vectors, relating to D statistic (i.e., $VI^D_{c,l}, l \in OOC^D_c$), are summarized into a unique vector by calculating the $\mu + 3\alpha$ for each SVID. Thus, the contribution vectors, relating to Q statistic (i.e., $VI^Q_{c,l}, l \in OOC^Q_c$), are summarized into a unique vector by making average of the contribution values corresponding to each SVID. For this summarization, we need to first investigate that the contributions values of each SVID follow the normal distribution.

Different statistical tests such as Kolmogorov-Smirnov (KS) and Lilliefors (LF) tests [163] are conducted to investigate whether the contribution values of each SVID follow normal distributions. It is worth mentioning that the Lilliefors test is based on the Kolmogorov-Smirnov test by considering unknown values for the mean and the variance of the data. **Table 5.6** shows the results of these tests for a set of SVIDs in each fault fingerprint. The values of tests equal to 0 for KS and LF tests in **Table 5.6** show the normal distribution for the corresponding SVID. The normality for all SVIDs has been checked and almost all of them follow normal distribution. **Table 5.6** shows the result only for eight sample SVIDs.

Table 5.6: Normality tests for the contribution values of the D and Q statistics.

FF*	**Stat***	**SVID**			
1	D	SVID 18 $KS, LF = 0$	SVID 21 $KS, LF = 0$	SVID 22 $KS, LF = 0$	SVID 28 $KS, LF = 0$
	Q	SVID 8 $KS, LF = 0$	SVID 12 $KS, LF = 0$	SVID 16 $KS, LF = 0$	SVID 24 $KS, LF = 0$
2	D	SVID 4 $KS, LF = 0$	SVID 7 $KS, LF = 0$	SVID 14 $KS, LF = 0$	SVID 23 $KS, LF = 0$
	Q	SVID 6 $KS, LF = 0$	SVID 9 $KS, LF = 0$	SVID 20 $KS, LF = 0$	SVID 30 $KS, LF = 0$
3	D	SVID 10 $KS, LF = 0$	SVID 16 $KS, LF = 0$	SVID 19 $KS, LF = 0$	SVID 27 $KS, LF = 0$
	Q	SVID 2 $KS, LF = 0$	SVID 13 $KS, LF = 0$	SVID 15 $KS, LF = 0$	SVID 26 $KS, LF = 0$

*FF: Fault Fingerprint; Stat: Statistics

After testing the normality of all the contribution values corresponding to each SVID in each fault fingerprint, **Figure 5.11** to **Figure 5.13** show the fault fingerprint 1 to 3 for the D statistic and **Figure 5.14** to **Figure 5.16** show the fault fingerprint 1 to 3 for the Q statistic, respectively. The SVIDs are interpreted as the fault roots if their summarized contribution values exceed the corresponding control limits. Considering the D statistics, **Figure 5.11** to **Figure 5.13** illustrate that the fault roots are {*Gas flow 01, Gas flow 02, Gas flow 03, Gas flow 04*}, {*RF_03, RF_07, RF_08, RF_09, RF_10, Pressure 04*}, and {*RF_04, Gas flow 02,*

Gas flow 12, RF_09, Temperature 01, Pressure 01} for fault fingerprints 1 to 3, respectively. Considering the Q statistics, **Figure 5.14** to **Figure 5.16** illustrate that the fault roots are {*RF_02, RF_06, RF_09, Gas flow 02, Temperature 01, Pressure 02, Pressure 03, Pressure 04*}, {*RF_04, RF_07, RF_08, Pressure 01, Gas flow 12*}, and {*RF_06, Gas flow 07, Gas flow 10, Gas flow 13*} for fault fingerprints 1 to 3, respectively. It can be concluded from **Figure 5.11** to **Figure 5.16** that the main fault root is due to the malfunctioning Gas flows.

As a summary, the proposed data-driven equipment failure diagnosis approach works directly based on the observations of the FDC data-set. This approach detected the normal from abnormal observations with a 98% accuracy. The normal observations were used to decompose the process dynamics into seven groups. It was observed the middle of the wafers process contains higher dynamics and, consequently, is decomposed into several stationary clusters. The stationarity of these clusters were proved through different statistical tests. The variance structure of these clusters were captured in the normal models based on PCA. Projecting the abnormal observations to these models helped to find out their deviations from the normal basis. The divergence of each abnormal observation was transformed to the contribution values, which were further clustered into three meaningful fault fingerprints. In each fault fingerprint, the corresponding fault roots were identified by the control limits. It was observed that the gas flow SVIDs are classified as the fault roots mostly in the D statistics. From the perspective of the Q statistics, pressure SVIDs are also among the fault roots.

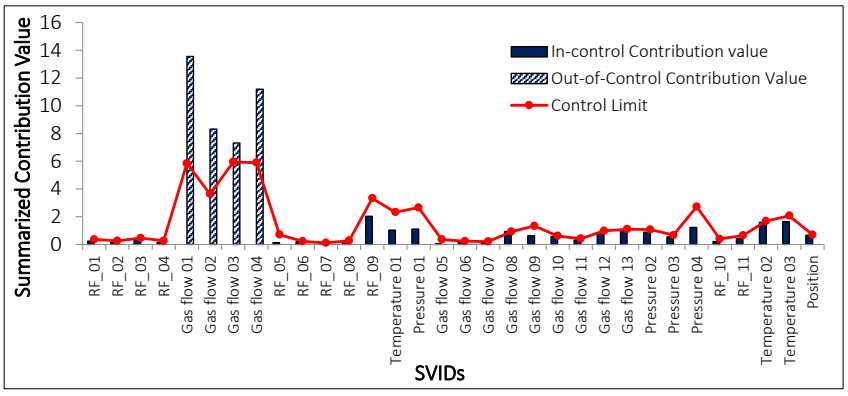


Figure 5.11: Fault roots of the fault fingerprint 1 coming from the D statistic.

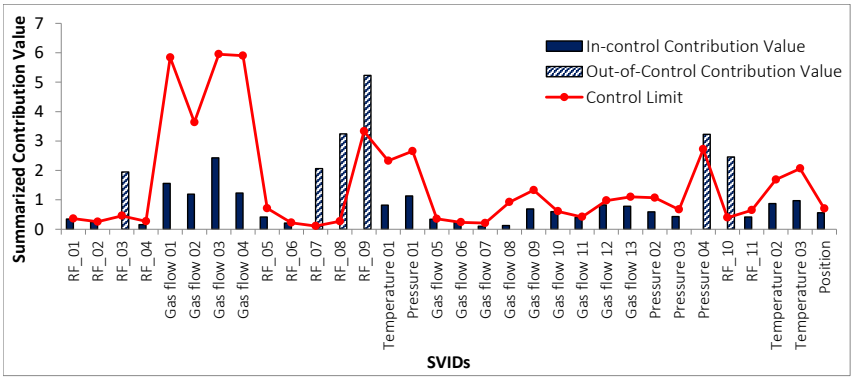


Figure 5.12: Fault roots of the fault fingerprint 2 coming from the D statistic.

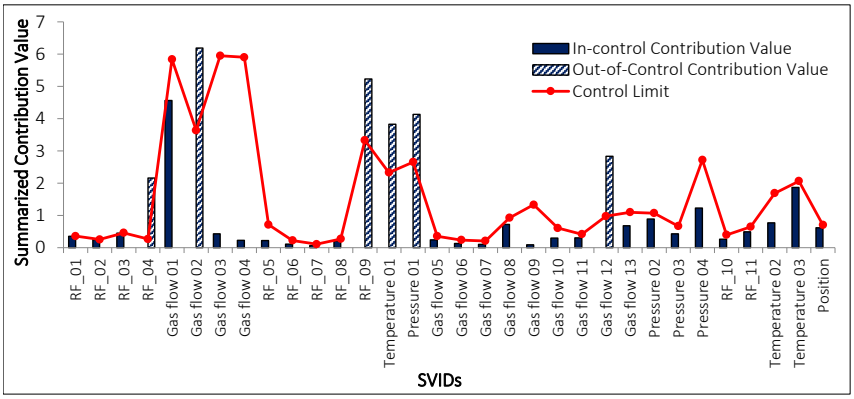


Figure 5.13: Fault roots of the fault fingerprint 3 coming from the D statistic.

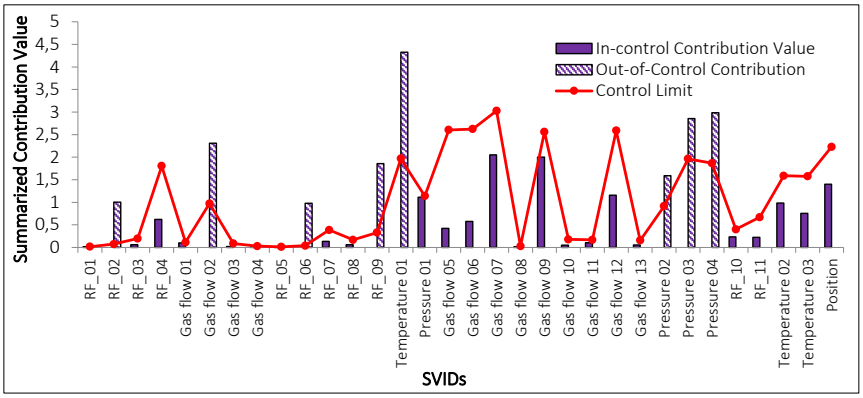


Figure 5.14: Fault roots of the fault fingerprint 1 coming from the Q statistic

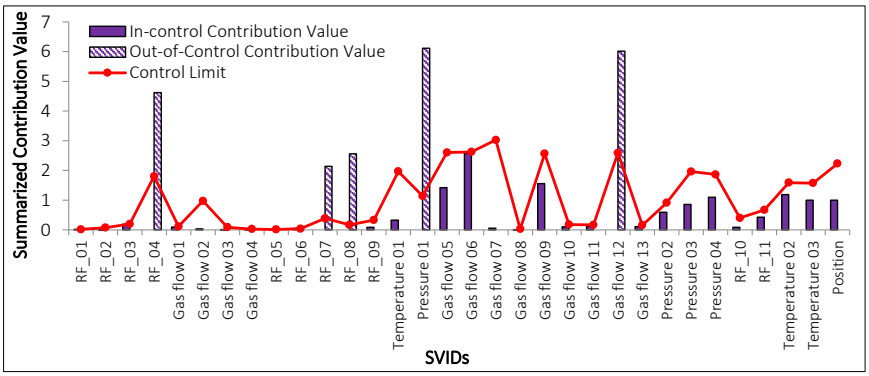


Figure 5.15: Fault roots of the fault fingerprint 2 coming from the Q statistic.

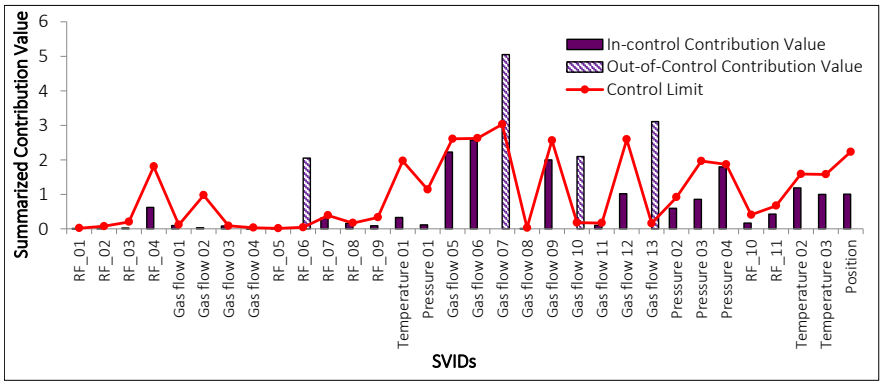


Figure 5.16: Fault roots of the fault fingerprint 3 coming from the Q statistic.

5.3 Results of the Prognosis Approach

The equipment deterioration modeling approach proposed in **Chapter 4** is validated in this section. The same FDC data-set is utilized knowing that seven out of eight recipes are labeled as normal process runs. Therefore the proposed prognosis approach is conducted on the first 479 normal wafers. The data-set is divided into two overlapped sets. The first set, labeled as 3_Recipes,consists of only three recipes out of the seven recipes and contains 392 wafers. The second set, labeled as 7_Recipes, is the whole data-set, i.e., 479 wafers. The set of 3_Recipes accounts for the 82% of the wafers in the FDC data. Although all the seven recipes are from the same family of products, these three recipes have the highest similarity comparing to the others. There are still visible differences between these three recipes and other four recipes. Consequently, the 3_Recipes set has the lowest inter-recipe variations when comparing to the 7_Recipes set.

The Discrete Wavelet Transformation decomposes the raw signals into approximation

and detail components. Consequently, each of the two sets can be viewed and analyzed in terms of three aspects: Raw, Approximation, and Detail. On the other hand, as explained in **Section 4.3**, two types of INCREASING and DECREASING deterioration trends are investigated on the signals of each data subset. In the DECREASING trend, the correlation between SVIDs become stronger, and the determinant of correlation is decreasing, while in the INCREASING trend it is vice versa. As a result, 12 possible scenarios (2 sets $\times$ 3 aspects $\times$ 2 trends) should be investigated. The 12 scenarios are summarized in **Table 5.7**, and each scenario is identified with a particular symbol. For instance, 3A$\uparrow$ represents a scenario that takes the approximation component of the 3_Recipes set and looks for an increasing deterioration trend while 7R$\downarrow$ indicates a scenario that takes the raw signals of the 7_Recipes set and looks for a decreasing deterioration trend.

Table 5.7: Different twelve possible scenarios with particular symbols.

Dataset	**Deterioration Trend**	**Signal**		
		Raw	Approximation	Detail
3_Recipes	INCREASING	3R$\uparrow$	3A$\uparrow$	3D$\uparrow$
	DECREASING	3R$\downarrow$	3A$\downarrow$	3D$\downarrow$
7_Recipes	INCREASING	7R$\uparrow$	7A$\uparrow$	7D$\uparrow$
	DECREASING	7R$\downarrow$	7A$\downarrow$	7D$\downarrow$

5.3.1 Best Mother Wavelet Selection

Before starting to decompose the SVID signals and model the equipment behavior based on the approximation and the detail components, the first step was to find the most suitable mother wavelet for each SVID. Hence, a library of 49 orthogonal wavelets was used based on **Table 4.1**, and for each SVID a wavelet is selected that provides the highest correlation coefficient between raw signal and approximated signal in the first level of decomposition. **Table 5.8** shows the best mother wavelet chosen for each SVID, and the selected mother wavelet differs from one SVID to another. Accordingly, different mother wavelets for different SVIDs discover the necessity of selecting the best mother wavelet for each SVID. The results of **Table 5.8** is valid for all scenarios.

One may argue that why not considering a unique common mother wavelet such as Haar for decomposing all SVIDs. Even though unique mother wavelet leads to satisfactory results in particular cases, but it is evident that the form and characteristics might be significantly different from one SVID to another. In a more general view, when new data-set arrives or even the behavior of new and different equipment is going to be modeled, the unique mother wavelet disables to catch the highest possible information from the signals. Accordingly, an automatic procedure is required as what is explained in the best mother wavelet selection. Furthermore, adding this step does not impose any computational complexity on the approach, and it is performed in some seconds.

Table 5.8: Best mother wavelet was selected for each SVID.

SVID j	Wavelet	SVID j	Wavelet	SVID j	Wavelet
1	bior2.4	12	sym5	22	haar
2	rbio3.3	13	rbio5.5	23	dmey
3	rbio3.3	14	sym7	24	rbio3.3
4	db3	15	rbio1.3	25	sym5
5	db3	16	bior3.9	26	rbio1.3
6	coif1	17	haar	27	haar
7	db3	18	rbio3.5	28	coif3
8	db2	19	haar	29	rbio1.5
9	sym5	20	haar	30	sym5
10	bior5.5	21	rbio3.3	31	sym5
11	sym5				

5.3.2 Micro-level Variation Imposition for Wavelet Decomposition

In the proposed data-driven approach, only the first-level decomposition was considered because it was observed that the approximation and detail signals at the first-level are very representative. Drilling down to higher levels does not exhibit more insights.

Each SVID is decomposed into approximation and detail components using the best mother wavelets shown in **Table 5.8**. The goal of this step is to distinguish the macro-level variation from the micro-level one. Despite the hidden existence of macro-level variation in the FDC data, very significant micro-level variation in the signals was not observed. Accordingly, these variations are imposed into the data-set, and then the distinguishing capability of the proposed approach can be validated regarding variation differentiation and finding corresponding causes. As a matter of fact, the introduced variations represent an adjustment not destroying the data-set, as the profile behavior stays unchanged. For this aim, SVIDs of pressures and gas flows are selected that are more likely to get micro-level variation. In reality, when equipment is deteriorating, the variation magnitude is likely to enhance through the gas- or pressure-related sensors. Therefore a source of correlated noises is introduced to these SVIDs, and when the equipment deteriorates gradually, these noises become more correlated. Therefore, the only decreasing trend will be extracted from the detail signals. Hence, two 3D↑ and 7D↑ will not be studied in this chapter since no significant trend neither exists nor introduced to their corresponding data-sets. Consequently, the scenarios are reduced to totally ten cases (i.e., 3R↑, 3R↓, 7R↑, 7R↓, 3A↑, 3A↓, 7A↑, 7A↓, 3D↓, and 7D↓).

Two ways of introducing micro-level variation are adding zero-mean (i.e., $\mu = 0$) noises to the signals then either increasing the standard deviation of the noises or increasing the correlation between signals through the wafers. As mentioned in **Chapter 4**, as the equipment deterioration occurs, the correlation between signals is increased. Therefore, increasing the standard variation of the noises through the wafers does not necessarily leads to the correlation increase between detail signals. Accordingly, the second way is adapted and the correlation between the random noises is increased directly. For this aim, the zero-mean (i.e., $\mu = 0$) noises with small standard deviations (i.e., σ) are introduced to SVIDs of pressures and gas flows, wherein the correlation between

noises increases gradually through the wafers.

Minimum and maximum correlation values are considered as $\rho_{min} = 0$ and $\rho_{max} = 1$, respectively. Afterward, the correlation of introduced noises into wafer k is calculated as $\rho_k = \rho_{min} + d(k-1)$, where $d = (\rho_{max} - \rho_{min})/K$ is the increasing step of correlation from wafer to wafer. Finally, the noises are generated randomly following normal multivariate distribution $N(\mu, \Sigma_k)$ for each wafer k, where Σ_k is the covariance matrix calculated from σ and ρ_k. SVID numbers 16 (Gas flow 05), 21 (Gas flow 10), 26 (Pressure 03) and 27 (Pressure 04) have been selected to introduce micro-level variation. The same procedure is employed for both 3_Recipes and 7_Recipes sets.

It is expected that wavelet decomposition not only keeps macro-level variation in the approximation signals but also transfers the micro-level variation into the detail signals. The ideal case is that the total introduced micro-level variation go to the detail component. It guarantees that macro-level and micro-level variations will be distinguished and they can be analyzed independently.

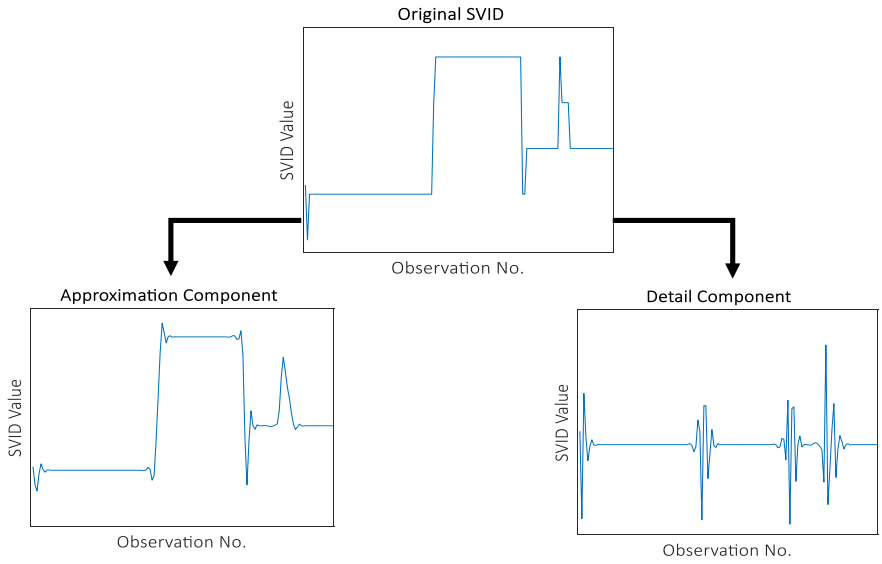
Table 5.9 shows that wavelet decomposition can transfer the majority of the micro-level variation to the detail component signal. As a result, the performance of wavelet decomposition in distinguishing macro-level and micro-level variations is validated.

Table 5.9: Percentage of total introduced micro-level variation transferred to each component.

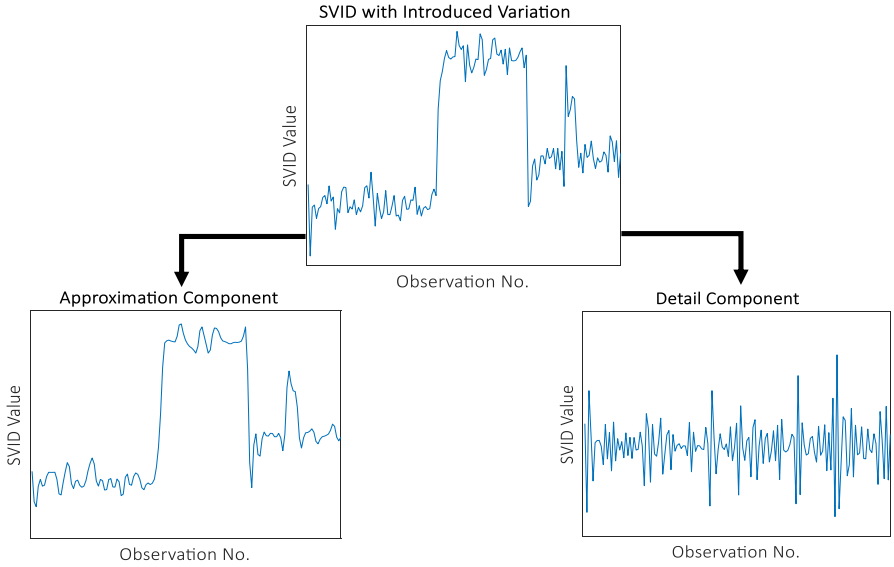
Component	**SVIDs**			
	16	21	26	27
Approximation	25%	07%	14%	07%
Detail	**75%**	**93%**	**86%**	**93%**

5.3.3 Equipment Deterioration Model Considering All SVIDs

This section examines if any trend can be extracted in the approximation, in the detail or directly by the raw signals of the 3_Recipes and 7_Recipes data-sets without applying the proposed SDSA algorithm. For this aim, the HI_A, HI_D, and HI_R trends are depicted directly in **Figures 5.18** to **5.23**. The HI_R stands for the determinant of correlation (health index) over the raw signals. **Figures 5.18** to **5.20** illustrate respectively the equipment HI based on the approximation, the detail and the raw signals in the 3_Recipes set when all SVIDs are incorporated. **Figures 5.21** to **5.23** show the same results for the 7_Recipes set. As can be seen, the equipment HI through the wafers based on approximation, detail and raw signals, respectively, demonstrate no gradual decreasing or increasing trend. Actually, when all SVIDs are considered in calculating the determinant of correlation, some redundant SVIDs destroy the hidden underlying deterioration trends. These results validate the necessity of the proposed SDSA.



(a) Signal decomposition before introducing variations



(b) Signal decomposition after introducing variations

Figure 5.17: Variation is introduced to the raw signal, and the majority of this variation is captured by the Detail component. Here, 75% of the added variance is transferred to the Detail component, and 25% stays in the Approximation component.

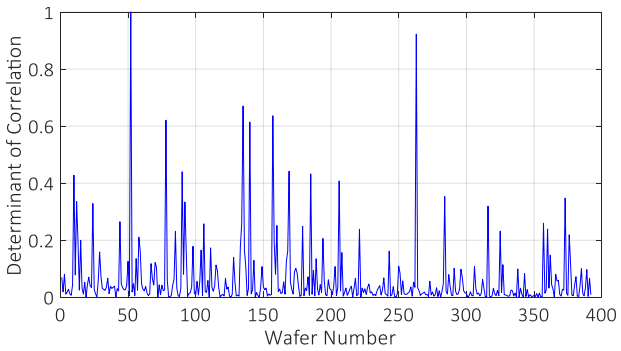


Figure 5.18: Equipment HI based on approximation with all SVIDs in 3_Recipes data-set.

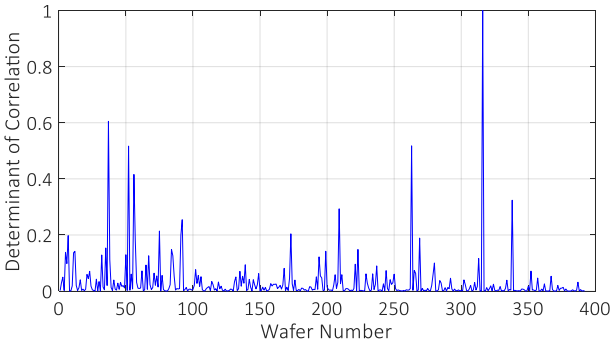


Figure 5.19: Equipment HI based on detail with all SVIDs in 3_Recipes data-set.

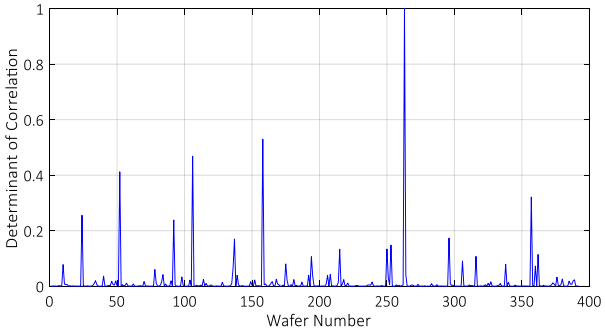


Figure 5.20: Equipment HI based on raw signals with all SVIDs in 3_Recipes data-set.

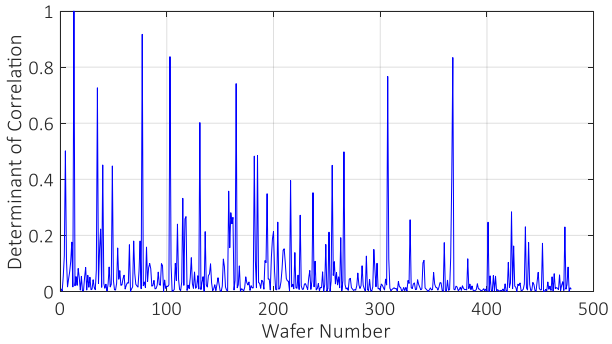


Figure 5.21: Equipment HI based on approximation with all SVIDs in 7_Recipes data-set.

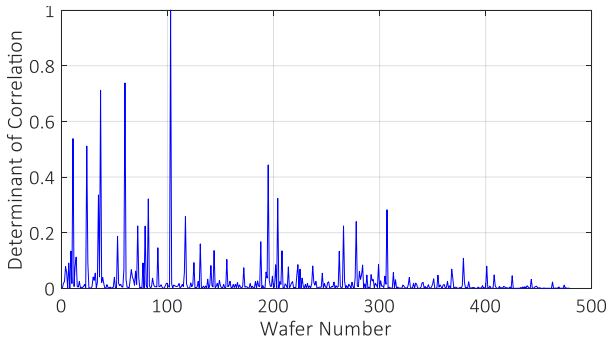


Figure 5.22: Equipment HI based on detail with all SVIDs in 7_Recipes data-set.

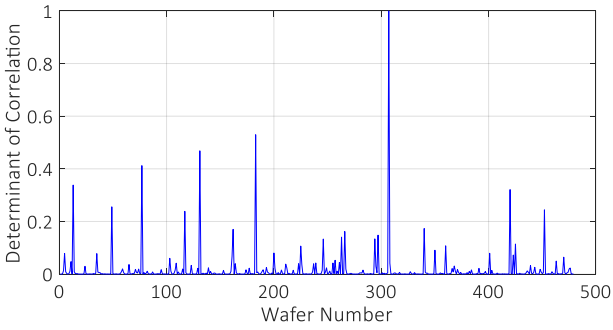


Figure 5.23: Equipment HI based on raw signals with all SVIDs in 7_Recipes data-set.

5.3.4 Equipment Deterioration Model using SDSA

For finding the most significant decreasing/increasing deterioration trends, the proposed SDSA algorithm is applied to all ten scenarios. The performance of the proposed SDSA is compared with a simple greedy algorithm and Genetic Algorithm (GA) [131].

A greedy algorithm is a constructive iterative algorithm that starts with an empty set of SVIDs. Unlike SDSA, each iteration of the greedy algorithm contains only forward-selection, and once an SVID enters the list, it never exists the list. Therefore, whenever a redundant SVID enters, that list has a lower chance to reach to the optimum list. In GA, each solution is represented as a $1 \times J$ matrix of binary values, wherein value 1 in the j^{th} bit shows that j^{th} SVID is in the list and value 0 indicates that j^{th} SVID is not in the list. A fixed number of paired individuals (i.e., crossover rate, C_{Rate}) are crossed using one-point or two-point crossover operators. In addition, a fixed number of individuals (i.e., mutation rate, M_{Rate}) are mutated in order to escape from local optima. GA parameters are fixed as $NPop = 100$, $C_{Rate} = 80\%$, $M_{Rate} = 20\%$. A maximum computational time is considered as the stopping criterion for the GA. The GA is performed for the approximation and the detail signals separately.

Tables 5.10 to **5.12** compare the performance of all three algorithms (i.e., SDSA, greedy algorithm and GA) and report the contributing SVIDs in each scenario. In these tables, S_R stands for the set of contributing SVIDs to the most significant trends based on the raw signals. In addition, α_R and R_R^2 are the slope and the GoF of the deterioration trend obtained from S_R. **Table 5.10** compares the performance of the three algorithms in obtaining the most significant DECREASING trend for the approximation and the raw signals (i.e., scenarios 3A↓, 3R↓, 7A↓ and 7R↓). **Table 5.11** shows the results for obtaining the most significant INCREASING trend for the approximation and the raw signals (i.e., scenarios 3A↑, 3R↑, 7A↑ and 7R↑). Finally, **Table 5.12** reports the results for obtaining the most significant DECREASING trend for the detail signals (i.e., scenarios 3D↓ and 7D↓). In **Tables 5.10** to **5.12**, third to fifth columns (in the heading) provide the characteristics of the deterioration trend created by corresponding contributing SVIDs in the sixth column. The reported time in the last column measures the amount of time the CPU spent running the algorithms. This does not count any other programs that might be running, and also does not count CPU time spent in the kernel (such as for file I/O). The stopping criterion for GA is fixed on 3600 seconds (more than SDSA and Greedy algorithm). This value is big enough and gives enough chance to GA for finding better solutions than SDSA and Greedy algorithm.

After having a quick look on the results, it can be seen that both SDSA and greedy algorithms provide the same trend significance and they obtain a remarkably better solution with lower values of the objective function comparing to the GA in almost all cases except 7R↓. The only drawback of the SDSA compared to the greedy algorithm is its computational time. However, this small difference can be neglected from an industrial point of view. The superiority of the SDSA is further elaborated in comparison with the greedy algorithm.

It is also observed from **Table 5.10** that the SVIDs contributing to the decreasing deterioration trends are mostly different from those of contributing to the increasing trends of **Table 5.11**. It reveals that the equipment deterioration with macro-level variations might appear with different symptoms in the signals and each symptom impacts a particular set of SVIDs. A common finding in **Tables 5.10** and **5.11** is that the gas flow SVIDs contribute to both decreasing and increasing trends with macro-level

variations. Accordingly, the gas flows number 01, 02, 03 and 04 are the ones which are representatives of the equipment health in case of macro-level variations.

Another results learning from **Tables 5.10** to **5.12** is that the GA reports somehow different contributing SVIDs comparing to the SDSA and the greedy algorithm. It may be because the SDSA and the greedy algorithm directly construct the contributing SVID set based on the correlation between SVIDs while the GA evolves the SVIDs set based on random operators (i.e., crossover and mutation). Accordingly, it seems that the GA will lead to less correct (if not incorrect) results and set of contributing SVIDs. Furthermore, the contributing SVIDs based on the raw signals are different from those of obtained from the approximation signals, S_A. Regarding S_R sets, some SVIDs do not lead to a significant deterioration trend. This issue will be further illustrated by comparing the trends obtained from S_A and S_R. Finally, knowing the GA as one of the most potent searching algorithms, **Tables 5.10** to **5.12** reveal that not only both SDSA and greedy algorithm provide solutions as good as (even better than) the GA, but also they are significantly more time-efficient.

Looking at **Table 5.12**, it is observed that the proposed SDSA and the greedy algorithm both completely find out those SVIDs (i.e., 16, 21, 26 and 27) that received noises in **Section 5.3.2**. It validates the performance of the proposed equipment behavior prognosis approach in extracting the deterioration trend from the micro-level variations in the signals.

Figures 5.24 to **5.27** depict the decreasing deterioration trends of **Table 5.10**. The first observation is decreasing trend of scenarios 3A↓, and 7A↓ in **Figures 5.24** and **5.26** wherein the equipment behavior contain two phases as downward and upward and contain the fitted linear regression (dashed line).

Table 5.10: The most significant DECREASING trend obtained from the raw and the approximation signals.

Sc.	Alg.	Obj.	Value	Slope ($\times 10^{-4}$)		GoF		Contributing SVIDs	Time (s)
	SDSA		4.8960	-8.71		0.3668		$\{6, 18, 8, 12, 15\}$	1200
3A↓	Greedy	$F(S_A)$	4.8960	α_A -8.71	R_A^2	0.3668	S_A	$\{18, 6, 8, 12, 15\}$	500
	GA		4.9492	-8.52		0.3578		$\{5, 7, 11, 15, 30\}$	3600
	SDSA		5.0761	-7.94		0.3491		$\{28, 5, 30\}$	1400
3R↓	Greedy	$F(S_R)$	5.0761	α_R -7.94	R_R^2	0.3491	S_R	$\{28, 5, 30\}$	800
	GA		5.2906	-8.02		0.3208		$\{5, 7, 11, 13, 14, 15, 30\}$	3600
	SDSA		5.9891	-4.84		0.2745		$\{6, 7, 5, 19, 10, 29\}$	1426
7A↓	Greedy	$F(S_A)$	5.9891	α_A -4.84	R_A^2	0.2745	S_A	$\{7, 6, 5, 19, 10, 29\}$	500
	GA		6.0204	-4.76		0.2720		$\{5, 6, 7, 10, 14, 19, 29\}$	3600
	SDSA		6.7480	-3.03		0.2242		$\{6, 16, 8, 27, 12, 13\}$	1500
7R↓	Greedy	$F(S_R)$	6.3107	α_R -5.60	R_R^2	0.2227	S_R	$\{7, 8, 23, 27, 16, 20, 26, 9, 19, 13\}$	520
	GA		5.6088	-6.19		0.3012		$\{7, 8, 15, 16, 18, 19, 20, 22, 23, 25, 26, 30\}$	3600

Sc.: Scenario; Alg.: Algorithm; Obj.: Objective

Table 5.11: The most significant INCREASING trend obtained from the raw and the approximation signals.

Sc.	Alg.	**Obj. Value**		**Slope** ($\times 10^{-4}$)		**GoF**		**Contributing SVIDs**		**Time (s)**
	SDSA		5.1993		8.70		0.2613		$\{1, 5, 21\}$	800
3A↑	Greedy	$F(S_A)$	5.1993	α_A	8.70	R_A^2	0.2613	S_A	$\{1, 5, 21\}$	800
	GA		5.2374		8.60		0.2571		$\{1, 6, 21, 25\}$	3600
	SDSA		5.2125		8.91		0.2574		$\{1, 5, 21\}$	880
3R↑	Greedy	$F(S_R)$	5.2125	α_R	8.91	R_R^2	0.2574	S_R	$\{1, 5, 21\}$	390
	GA		6.3416		6.30		0.1392		$\{1, 6, 13, 21, 25\}$	3600
	SDSA		6.1468		5.10		0.1890		$\{1, 5\}$	1800
7A↑	Greedy	$F(S_A)$	6.1468	α_A	5.10	R_A^2	0.1890	S_A	$\{1, 5\}$	470
	GA		6.2797		4.90		0.1744		$\{1, 6, 10, 13, 14, 16, 21, 25, 31\}$	3600
	SDSA		6.1453		5.09		0.1895		$\{1, 5\}$	1600
7R↑	Greedy	$F(S_R)$	6.1453	α_R	5.09	R_R^2	0.1895	S_R	$\{1, 5\}$	470
	GA		6.3041		5.06		0.1691		$\{1, 6, 10, 16, 21, 25, 31\}$	3600

Sc.: Scenario; Alg.: Algorithm; Obj.: Objective

Table 5.12: The most significant DECREASING trend obtained from the detail signals.

Sc.	Alg.	**Obj. Value**		**Slope** ($\times 10^{-4}$)		**GoF**		**Contributing SVIDs**		**Time (s)**
	SDSA		1.8946		-16.1		0.7353		$\{27, 16, 26, 21, 5, 3\}$	1500
3D↓	Greedy	$F(S_D)$	1.8946	α_D	-16.1	R_D^2	0.7353	S_A	$\{16, 26, 27, 21, 5, 3\}$	1000
	GA		1.9217		-16.0		0.7317		$\{5, 16, 17, 21, 26, 27\}$	3600
	SDSA		1.0489		-1.40		0.8557		$\{21, 11, 16, 26\}$	2300
7D↓	Greedy	$F(S_D)$	1.0489	α_D	-1.40	R_D^2	0.8557	S_R	$\{16, 21, 26, 11\}$	510
	GA		1.0513		-1.40		0.8554		$\{11, 16, 21, 26, 31\}$	3600

Sc.: Scenario; Alg.: Algorithm; Obj.: Objective

However, the deterioration trend is expected to be decreasing in the long-term to show the equipment health is going worse and worse over time. For instance, the wafers number 300 and 350 are respectively the turning points in **Figures 5.24** and **5.26** from a downward phase to the upward phase. The upward trend signifies an improvement in the equipment health. This issue and this improvement could be due to the automatic Run-to-Run regulation of the etching system, but further investigation with the engineers is required.

Figures 5.25 and **5.27** depict the deterioration trends of the scenarios 3R↓ and 7R↓ wherein the wavelet decomposition step has been skipped, and the proposed feature extraction approach and SDSA are performed directly on the raw signals to find the most significant equipment deterioration trend. Comparing to **Figures 5.24** and **5.26**, it can be seen that no significant trend is observed but the downward spike. This result validates the necessity of the wavelet decomposition for distinguishing the macro-level variation from the micro-level one. Actually, without signal decomposition, these two variations are convoluted, and no significant deterioration trend can be in the 3R↓ and 7R↓ scenarios.

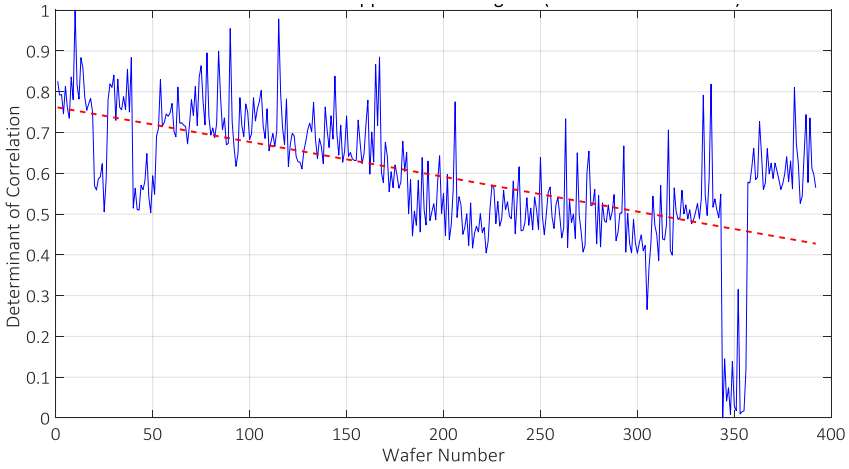


Figure 5.24: Deterioration trend of the scenario 3A↓ obtained by SDSA.

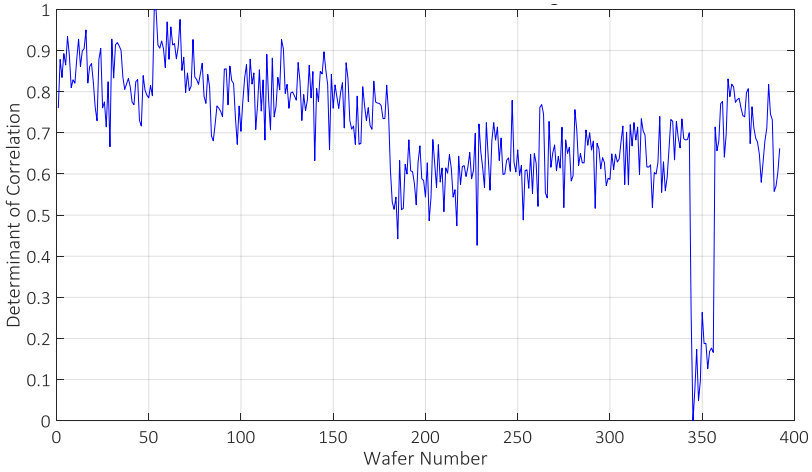


Figure 5.25: Deterioration trend of the scenario 3R↓ obtained by SDSA.

The deterioration trends of **Figures 5.24** and **5.26** contain two spikes, one at the beginning and one in the upward phases. Precisely in **Figure 5.24**, two sudden shifts were observed in the original data-set as **Figure 5.28** which result in upward and downward spikes. It is worth mentioning that the proposed approach alarms not only sudden macro-level changes in the contributing SVIDs but also prognoses the long-term deterioration trend over the whole wafers. Although the proposed prognostic approach is sensitive to sudden variations, these changes do not impact the ongoing deterioration trend of the equipment. As can be seen, the downward shift is more severe, and consequently, its spike (around wafer 350) is more significant.

Figures 5.29 and **5.30** illustrate the deterioration trend of the scenarios 3D↓ and

7D↓ of **Table 5.11** It is observed that the proposed equipment behavior prognosis approach can remarkably distinguish the micro-level variation and keep it in the detail signals. Considering the 3_Recipes data-set, both trends of **Figures 5.24** and **5.29** can be adapted to monitor the equipment deterioration. In addition, the trends of **Figures 5.26** and **5.30** can be monitored when using the 7_Recipes data-set.

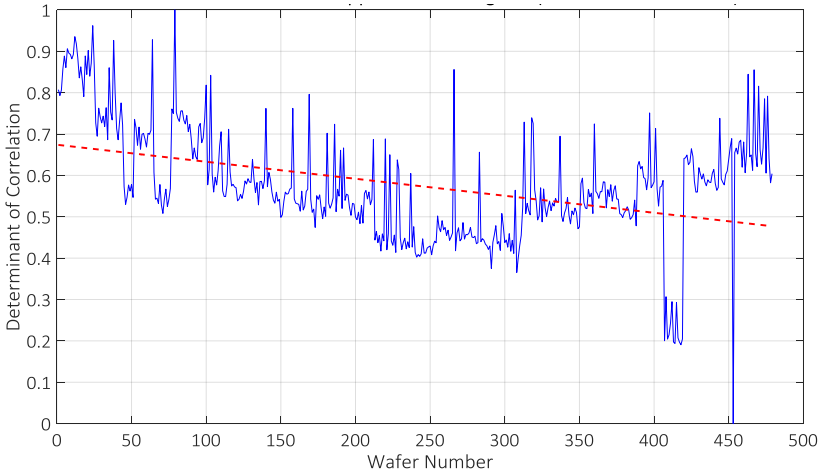


Figure 5.26: Deterioration trend of the scenario 7A↓ obtained by SDSA.

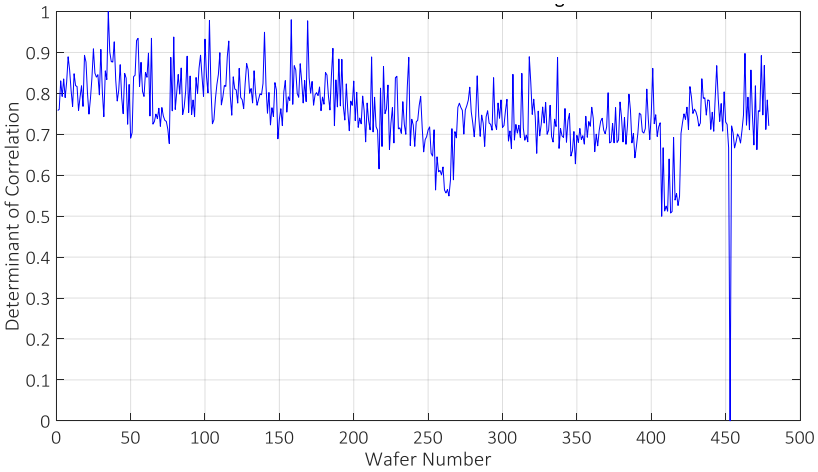


Figure 5.27: Deterioration trend of the scenario 7R↓ obtained by SDSA.

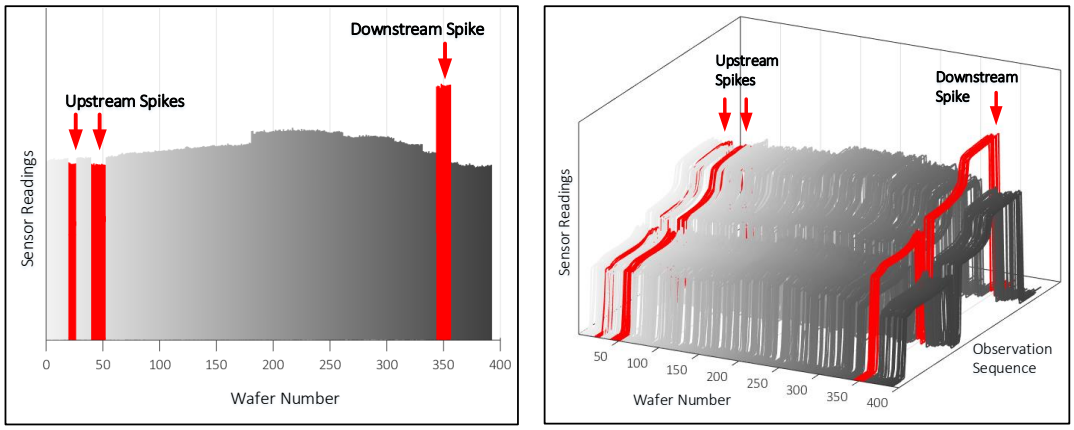


Figure 5.28: Upward and downward spikes in the profiles of one of the contributing SVID that have been observed in the trend of the 3A↓ scenario.

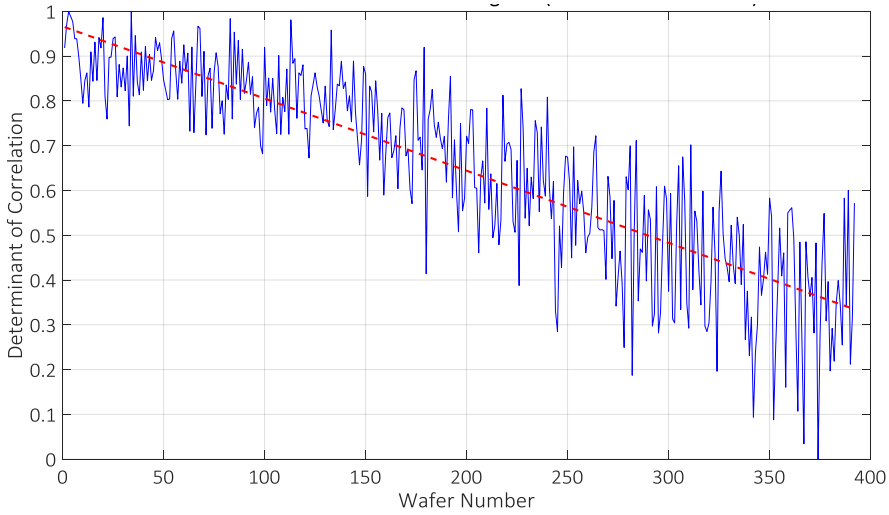


Figure 5.29: Deterioration trend of the scenario 3D↓ obtained by SDSA shows a significant deterioration tendency.

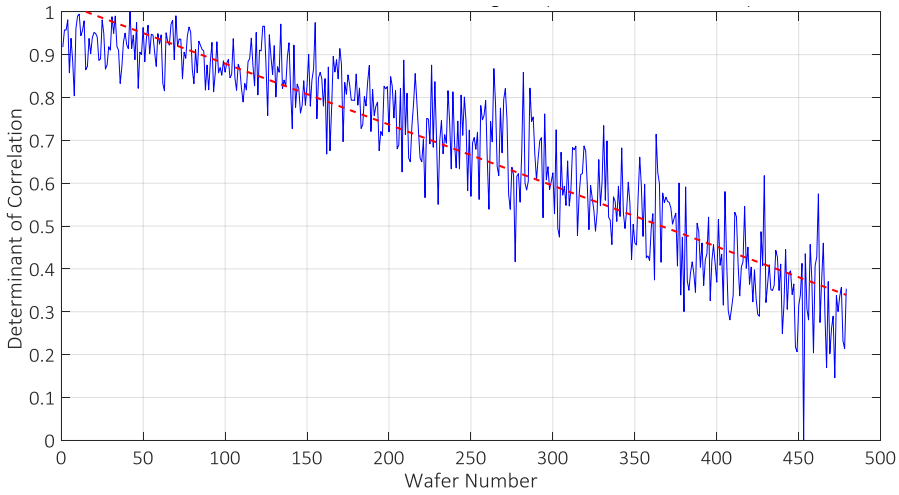


Figure 5.30: Deterioration trend of the scenario 7D↓ obtained by SDSA shows a significant deterioration tendency.

In the following, the increasing trends in the scenarios 3A↑, 3R↑, 7A↑, and 7R↑ are illustrated and evaluated. These trends are respectively depicted in **Figures 5.31** to **5.34**. Since the increasing trends are not as significant and clear as the decreasing trends, it is concluded that the deterioration of the process/equipment can happen only when the correlations between SVIDs are gradually increased, which implies the process is gradually regulated to push the physical/chemical reactions.

Another insight learning from **Figures 5.31** to **5.34** is that despite the results of **Table 5.10** the raw signals (i.e., cases 3R↑ and 7R↑) provide the increasing trends as significant as the approximation signals (i.e., cases 3A↑ and 7A↑). This event is because no decrease of correlation was expected to appear among micro-level variation and consequently no increasing trend from the detail signals. Hence, there is no convolution between macro-level and micro-level variations regarding increasing trend. Consequently, the wavelet decomposition step might be skipped for obtaining the increasing trend in this particular FDC data but the proposed prognostic approach is an automotive approach that guarantees no convolution of variations happens in the signals, and the most significant trend (if any) will be surely found.

5.3.5 Performance Comparison between SDSA and Greedy Algorithm

Although it was observed (in **Section 5.3.4**) that the most significant deterioration trend obtained by the greedy algorithm is the same as the SDSA, this section provides more analyses on the performance of the proposed SDSA. The analyses are about testing its ability to escape from the local optimum and if it can find out other deterioration trends with different sets of contributing SVIDs apart of the most significant one.

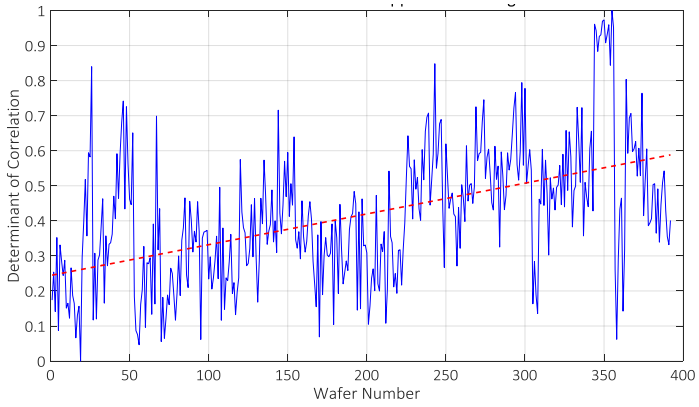


Figure 5.31: Deterioration trend of the scenario 3A↑ obtained by SDSA.

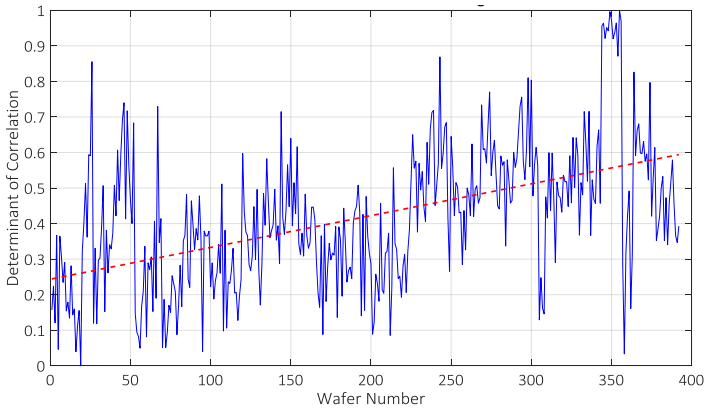


Figure 5.32: Deterioration trend of the scenario 3R↑ obtained by SDSA.

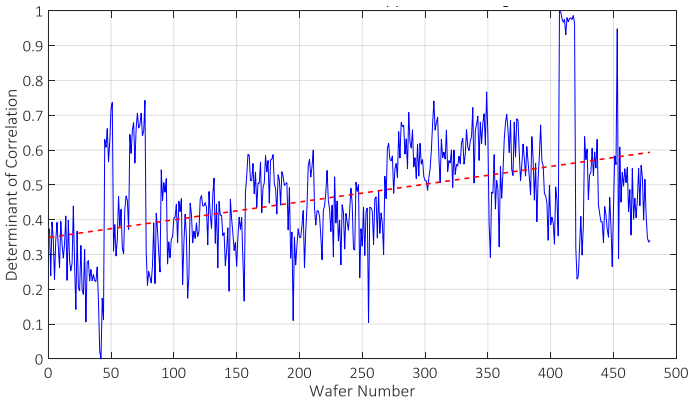


Figure 5.33: Deterioration trend of the scenario 7A↑ obtained by SDSA.

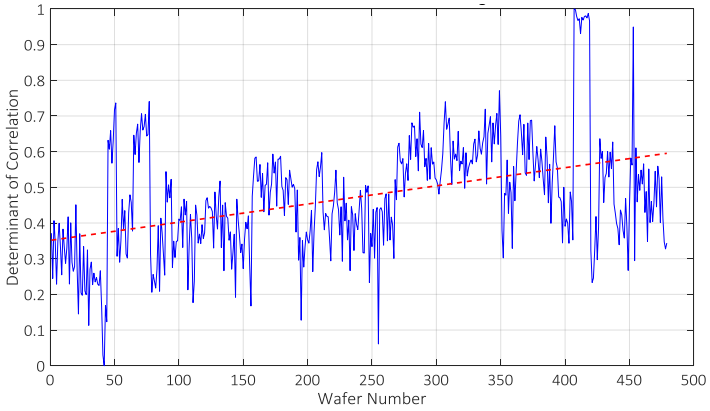


Figure 5.34: Deterioration trend of the scenario 7R↑ obtained by SDSA.

As mentioned in **Chapter 4**, the proposed SDSA and the greedy algorithm both are iterative algorithms, and they create the list of contributing SVIDs by starting from each SVID. Therefore, each algorithm leads to J number of SVID sets, where J is the number of all SVIDs in the FDC data-set. This section does a detailed analysis of these sets of SVIDs. **Tables 5.13** to **5.24**, report the results of the scenarios 3A↓, 7A↓, 3A↑, 7A↑, 3D↓ and, 7D↓ for the SDSA and the greedy algorithm. In **Tables 5.13** to **5.24**, the first column represents the starting SVID; the second to the forth columns provide the characteristics of the most significant trend corresponding by starting from the particular SVID of the first column. The fifth column shows the final list of the contributing SVIDs when starting from each SVID. In the last column, different groups of contributing SVIDs obtained by the corresponding algorithm are numerated. In this column, a lower number of different groups shows the higher performance of the algorithm to converge to the same solutions.

Table 5.13 and **Table 5.14** show the results of the case 3A↓ for the SDSA and the greedy algorithm, respectively. By looking at the second column of **Table 5.13** and **Table 5.14**, it is observed that by starting from the same SVID, in the majority of cases, SDSA provides a better solution in term of lower objective value. It can be easily counted that the SDSA reaches to better solutions in 28 cases (out of total 31) comparing to the greedy algorithm in the 3A↓ case.

In addition, the last column of **Table 5.13** indicates that totally 4 different sets of contributing SVIDs that create significant trend have been found while this value for the greedy algorithm is 25, and the most of the contributing sets of **Table 5.14** contain irrelevant SVIDs. This issue is because the greedy algorithm gets trapped in local optimum and starting from an SVID often leads to a different final set of contributing SVIDs. Hence, the SDSA algorithm is well converged to the most significant trends. Furthermore, we know that the best solution found by both algorithms is the same as $\{6, 8, 12, 15, 18\}$. This solution has been bolded in **Table 5.13** and **Table 5.14**, wherein it can be seen that the proposed SDSA obtains the set $\{6, 8, 12, 15, 18\}$ three times while the greedy algorithm reaches to this solution only once. On the other hand, the

SDSA obtains another solution $\{12, 6, 8, 31, 29\}$ nine times that is a good alternative for the optimal solution $\{6, 8, 12, 15, 18\}$ (i.e., both slope and R^2 values are quite close). Therefore, the SDSA is highly effective in producing optimal and near-optimal solutions.

Another observation is that in the contributing sets of SVIDs obtained by the greedy algorithm (in the fifth column of **Table 5.14**), the starting SVID always belong to the set and this is because no backward-elimination operator is employed in the greedy algorithm. So, the redundant entered SVIDs will no exit the list and make the local optimum issues. Furthermore, the final set of contributing SVIDs strongly depends on the starting SVID. This issue leads to a higher variety in the contributing sets of SVIDs and less reliability for determining the main cause roots of the equipment deterioration.

Table 5.15 and **Table 5.16** show the detail results of the case 7A↓ for the SDSA and the greedy algorithm, respectively. Comparing the objective function in these tables discovers that independent from the starting SVIDs, the SDSA leads to better solutions with lower values of objective functions comparing to the greedy algorithm. Furthermore, the SDSA algorithm often converges to the same solutions, and totally seven different sets (out of 31) of contributing SVIDs are found. These sets can be good alternatives since their objective function values are close. On the other hand, the greedy algorithm rarely converges to the same solution and leads to different 26 (out of 31) sets of contributing SVIDs.

Table 5.17 and **Table 5.18** show the detail results of the case 3A↑ for the SDSA and the greedy algorithm, respectively. In the 3A↑ case, in a majority of starting SVIDs (i.e., 30 cases out of total 31), the proposed SDSA finds more significant trends comparing to the greedy algorithm. Based on the last column of **Table 5.17**, the proposed SDSA remarkably leads to less different sets of contributing SVIDs (totally 4 different sets) in comparison with the greedy algorithm (totally 23 different sets) and the most of the contributing sets of **Table 5.18** contain unjustifiable SVIDs. Furthermore, the SDSA and the greedy algorithm have reached out ten times and twice the best solution (i.e., contributing set $\{1, 5, 21\}$), respectively. On the other hand, the best alternative contributing set $\{6, 15\}$ has been found ten times by the proposed SDSA; however, the alternative solution can be identified twice in the results of the greedy algorithm.

Table 5.19 and **Table 5.20** show the detail results of the case 7A↑ for the SDSA and the greedy algorithm, respectively. Similar to the case 3A↑, the SDSA results in better solutions regarding lower objective function values and entirely converge to only seven (out of 31) different sets of contributing SVIDs. The greedy algorithm cannot converge to the same solutions and produces twenty-four (out of 31) different and non-alternative sets of contributing SVIDs. Also, the SDSA finds the combination {1,5} as the contributing SVIDs with the minimum objective value (best solution) eleven times (out of 31) while the greedy algorithm finds it only twice (out of 31). Therefore, the SDSA is more powerful and reliable to find the best solution.

Table 5.21 and **Table 5.22** show the detail results of the case 3D↓ for the SDSA and the greedy algorithm, respectively. Similar to the 3A↓ and 3A↑ cases, it is observed that by starting from the same SVID in the 3D↓ case, the SDSA reaches to better solutions in all 31 cases comparing to the greedy algorithm. In addition, the last column of **Table**

5.21 indicates that totally 3 different sets of contributing SVIDs that create significant trend have been found while this value for the greedy algorithm is 26, and even the most of the contributing sets of **Table 5.22** have (received) no micro-level variations.

The best solution found by both algorithms (**Table 5.21** and **Table 5.22**) is the same as $\{16, 26, 27, 21, 5, 3\}$ and this solution has been obtained sixteen and only four times by the SDSA and the greedy algorithm, respectively. The best alternative obtained by the SDSA is the contributing set $\{5, 16, 27, 26, 17, 3, 21\}$ which has been found twelve times by this algorithm. The best solution and the alternative one are completely similar except in the SVID number 17. Hence, it can be considered that the SDSA has been converged to the same solution in 28 times (i.e., sixteen times for the best solution and twelve times for the alternative solution). Finally, **Table 5.23** and **Table 5.24** show the detail results of the case 7D↓ for the SDSA and the greedy algorithm, respectively. Almost the same interpretations can be provided for this case, and it is evident that the SDSA correctly pick out the same SVIDs that receive intentional micro-level variation. Finally, these explanations demonstrate the outperformance of the proposed SDSA in extracting the both decreasing and increasing deterioration trend from the approximation and the detail signals.

5.3.6 Discussion on the Deterioration Trends

By looking at **Table 5.13**, it is observed that Gas flow 02, 03, 04, and 07 alongside pressure 01 have been diagnosed as the main representatives of the etching equipment deterioration. There exist also another set of SVIDs (i.e., $\{12, 6, 8, 31, 29\}$) which is a good alternative for the best set and generates the most significant deterioration trend. All of these sets share the common identification of equipment deterioration over the Gas flow SVIDs. It is worth mentioning that reporting these SVIDs and explaining the deterioration symptoms (i.e., macro- or micro-level variations) give enough clues to the engineers to distinguish the deterioration and take necessary preventive actions. Some hints brought to mind from these results are discussed hereafter.

In current practice, clean room enhancement and the increasing utilization of the automated equipment enable reducing the contamination caused by particles or films. Although, contaminants from external sources have been diminished but contaminating particles are still generated inside the process chamber during the processing of semiconductor wafers. Some identified sources of contamination include the process gases and liquids, the interior walls of the process chambers and the mechanical wear of the wafer handling equipment [164]. Silicon-based ceramics have been widely used in semiconductor plasma processing equipment as a shield to protect the ceramic parts inside etchers or chemical vapor deposition reactor chambers from corrosion caused by fluorocarbon corrosive gases [165]. These materials interact with plasma and are eroded, resulting in the generation of contaminant particles on the wafer. These particles are combined and charged negatively which may ultimately contaminate a substrate that is being processed in the chamber.

From the physical point of view, two possible events may happen to incur the equipment deterioration in an etching process. The first event relates to contamination coated

on the chamber and the second event represent the interfering of the contamination particles in the plasma bombarding process.

In the first case, when etching equipment is gradually deteriorated, e.g., when the etching tool is contaminated, particles are coated on the chamber walls during the plasma process. Consequently, recipe set points (i.e., temperature targets, gas flow levels, etc.) are not achieved. R2R controller intends to regulate more gas into the chamber or to increase the temperature of the chuck to reach to the targeting set point. With the contamination of coated chamber wall, the chamber space becomes smaller, and the level of specific SVIDs increases in the long term. Therefore, the correlation between these SVIDs become stronger, and these SVIDs are those who signify the contamination.

In the second case, during a plasma etching, the high-energy reactive plasma ions can be easily combined with any available chemical elements in the chamber to generate contaminating particles. Accordingly, when the radio frequency (RF) power is turned on, the generated contaminating particles float or suspend in the chamber due to the interaction with high energy plasma ion particles [164]. Hence, floating particles interact with plasma ions and interfere with the plasma bombarding to the wafer surface [166]. This malfunctioning results in the reaction of controllers in putting more gas into the chamber. Therefore, the level of certain SVIDs increases. After the etching process of each wafer is conducted, the RF power and the heating lamps are switched off. It leads to the sudden loss of energy in the suspended contaminating particles and causing them to be quickly deposited on the chamber walls, an upper electrode, and the wafer. This event happens to each wafer, and the process regulator correspondingly increases the level of certain gas flows at each etching operation. The more the contamination stuck to the chamber walls, the more difficult it can be removed. This is why the level of gas flow SVIDs gradually increases, and the correlation between these SVIDs become strong.

Table 5.13: Results of SDSA on the scenario 3A↓.

Starting SVID	$F(S_A)$	$\alpha_A \times 10^{-4}$	R_A^2	S_A	**Group**
1	7.568	-4.7	0.0855	$\{1, 19, 11\}$	1
2	5.016	-7.8	0.3577	$\{5, 30, 3\}$	2
3	5.016	-7.8	0.3577	$\{3, 5, 30\}$	2
4	5.016	-7.8	0.3577	$\{30, 5, 3\}$	2
5	5.016	-7.8	0.3577	$\{5, 30, 3\}$	2
6	4.915	-8.7	0.3623	$\{6, 8, 12, 31, 29\}$	3
7	5.016	-7.8	0.3577	$\{30, 5, 3\}$	2
8	4.915	-8.7	0.3623	$\{8, 6, 12, 31, 29\}$	3
9	5.016	-7.8	0.3577	$\{30, 5, 3\}$	2
10	5.016	-7.8	0.3577	$\{30, 5, 3\}$	2
11	4.915	-8.7	0.3623	$\{8, 6, 12, 31, 29\}$	3
12	4.915	-8.7	0.3623	$\{12, 6, 8, 31, 29\}$	3
13	7.568	-4.7	0.0855	$\{1, 19, 11\}$	1
14	4.915	-8.7	0.3623	$\{6, 8, 12, 31, 29\}$	3
15	**4.896**	**-8.5**	**0.3668**	**{15,6,12,8,18}**	**4**
16	5.016	-7.8	0.3577	$\{30, 5, 3\}$	2
17	**4.896**	**-8.5**	**0.3668**	**{6,15,12,8,18}**	**4**
18	**4.896**	**-8.5**	**0.3668**	**{18,6,8,12,15}**	**4**
19	7.568	-4.7	0.0855	$\{19, 1, 11\}$	1
20	5.016	-7.8	0.3577	$\{5, 30, 3\}$	2
21	5.016	-7.8	0.3577	$\{3, 5, 30\}$	2
22	7.568	-4.7	0.0855	$\{11, 19, 1\}$	1
23	5.016	-7.8	0.3577	$\{5, 30, 3\}$	2
24	4.915	-8.7	0.3623	$\{6, 8, 12, 31, 29\}$	3
25	4.915	-8.7	0.3623	$\{6, 8, 12, 31, 29\}$	3
26	4.915	-8.7	0.3623	$\{6, 8, 12, 31, 29\}$	3
27	5.016	-7.8	0.3577	$\{30, 5, 3\}$	2
28	5.016	-7.8	0.3577	$\{3, 5, 30\}$	2
29	4.915	-8.7	0.3623	$\{29, 6, 8, 12, 31\}$	3
30	5.016	-7.8	0.3577	$\{30, 5, 3\}$	2
31	5.016	-7.8	0.3577	$\{30, 5, 3\}$	2

Table 5.14: Results of greedy algorithm on the scenario 3A↓.

Starting SVID	$F(S_A)$	$\alpha_A \times 10^{-4}$	R_A^2	S_A	**Group**
1	7.568	-4.7	0.0855	$\{1, 19, 11\}$	1
2	6.871	-6.3	0.1425	$\{2, 4, 15, 16, 27, 30, 18, 12, 8, 7, 5, 29\}$	2
3	5.016	-7.8	0.3577	$\{3, 5, 30\}$	3
4	6.871	-6.3	0.1425	$\{4, 2, 15, 16, 27, 30, 18, 12, 8, 7, 5, 29\}$	2
5	5.304	-7.8	0.3214	$\{5, 9, 28\}$	4
6	4.916	-8.7	0.3623	$\{6, 8, 12, 31, 29\}$	5
7	4.949	-8.6	0.3578	$\{7, 5, 15, 11, 30\}$	6
8	4.916	-8.7	0.3623	$\{8, 6, 12, 31, 29\}$	5
9	5.304	-7.8	0.3214	$\{9, 5, 28\}$	4
10	5.432	-7.7	0.3058	$\{10, 5, 28\}$	7
11	6.176	-7.2	0.2174	$\{11, 24, 8, 6, 10, 5, 7, 9, 14, 13\}$	8
12	6.331	-6.5	0.2075	$\{12, 24, 8, 6, 10, 13, 15, 18\}$	9
13	5.255	-7.8	0.3274	$\{13, 19, 11, 25, 9, 8, 6\}$	10
14	5.032	-8.1	0.3536	$\{14, 8, 6, 12, 31\}$	11
15	5.256	-8.2	0.3240	$\{15, 18, 5, 7, 11\}$	12
16	5.176	-9.1	0.3255	$\{16, 5, 7, 12, 26, 10, 14\}$	13
17	10.32	-5.3	0.0004	$\{17, 15\}$	14
18	**4.896**	**-8.5**	**0.3668**	**{18,6,8,12,15}**	**15**
19	7.568	-4.7	0.0855	$\{19, 1, 11\}$	1
20	5.381	-7.4	0.3153	$\{20, 7, 5, 6, 12, 29, 19\}$	16
21	5.931	-7.6	0.2431	$\{21, 11, 31, 8, 7, 5, 19, 13\}$	17
22	6.057	-7.0	0.2341	$\{22, 19, 1, 11, 8, 6, 5, 7, 30, 29\}$	18
23	5.706	-8.5	0.2624	$\{23, 7, 8, 5, 6, 12, 10, 13, 29\}$	19
24	5.710	-7.4	0.2736	$\{24, 8, 7, 5, 10, 15\}$	20
25	5.000	-8.4	0.3547	$\{25, 8, 6, 10, 31\}$	21
26	5.196	-8.1	0.3326	$\{26, 8, 6, 12, 18\}$	22
27	5.016	-8.7	0.3495	$\{27, 5, 7, 15, 11, 14\}$	23
28	5.031	-7.8	0.3563	$\{28, 5, 30\}$	24
29	4.916	-8.7	0.3623	$\{29, 6, 8, 12, 31\}$	5
30	5.016	-7.8	0.3577	$\{30, 5, 3\}$	3
31	5.532	-7.1	0.3002	$\{31, 5, 28\}$	25

Table 5.15: Results of SDSA on the scenario 7A↓.

Starting SVID	$F(S_A)$	$\alpha_A \times 10^{-4}$	R^2_A	S_A	**Group**
1	**5.99**	**-4.84**	**0.27**	**{ 6, 7, 5, 19, 10, 29}**	**1**
2	9.49	-1.20	0.01	$\{2, 4, 18, 17, 14\}$	2
3	6.37	-3.26	0.26	$\{5, 10, 6\}$	3
4	9.69	-1.03	0.00	$\{4, 29\}$	4
5	6.37	-3.26	0.26	$\{5, 6, 10\}$	3
6	6.37	-3.26	0.26	$\{6, 5, 10\}$	3
7	**5.99**	**-4.84**	**0.27**	**{7, 6, 5, 19, 10, 29}**	**1**
8	6.41	-2.94	0.26	$\{8, 6, 12, 31\}$	5
9	6.37	-3.26	0.26	$\{10, 5, 6\}$	3
10	7.98	-3.50	0.07	$\{10, 24, 22\}$	6
11	8.94	-1.87	0.02	$\{12, 19\}$	7
12	8.94	-1.87	0.02	$\{12, 19\}$	7
13	8.94	-1.87	0.02	$\{12, 19\}$	7
14	7.98	-3.50	0.07	$\{22, 10, 24\}$	6
15	8.47	-2.60	0.04	$\{15, 21, 19, 13\}$	8
16	6.37	-3.26	0.26	$\{6, 5, 10\}$	3
17	6.41	-2.94	0.26	$\{8, 6, 12, 31\}$	5
18	6.41	-2.94	0.26	$\{8, 6, 12, 31\}$	5
19	8.94	-1.87	0.02	$\{19, 12\}$	7
20	8.94	-1.87	0.02	$\{19, 12\}$	7
21	6.37	-3.26	0.26	$\{5, 6, 10\}$	3
22	6.41	-2.94	0.26	$\{6, 8, 12, 31\}$	5
23	6.41	-2.94	0.26	$\{12, 6, 8, 31\}$	5
24	7.98	-3.50	0.07	$\{24, 10, 22\}$	6
25	6.41	-2.94	0.26	$\{6, 8, 12, 31\}$	5
26	6.37	-3.26	0.26	$\{10, 6, 5\}$	3
27	6.41	-2.94	0.26	$\{6, 8, 12, 31\}$	5
28	6.37	-3.26	0.26	$\{6, 5, 10\}$	3
29	9.69	-1.03	0.00	$\{29, 4\}$	4
30	6.37	-3.26	0.26	$\{10, 5, 6\}$	3
31	6.37	-3.26	0.26	$\{10, 5, 6\}$	3

Table 5.16: Results of greedy algorithm on the scenario 7A↓.

Starting SVID	$F(S_A)$	$\alpha_A \times 10^{-4}$	R^2_A	S_A	**Group**
1	6.80	-4.66	0.18	$\{1, 7, 6, 8, 5, 15, 14, 29\}$	1
2	9.49	-1.20	0.01	$\{2, 4, 18, 17, 14\}$	2
3	6.93	-4.00	0.18	$\{3, 31, 5, 6, 8, 29\}$	3
4	9.69	-1.03	0.00	$\{4, 29\}$	4
5	6.37	-3.26	0.26	$\{5, 6, 10\}$	5
6	6.37	-3.26	0.26	$\{6, 5, 10\}$	5
7	**5.99**	**-4.84**	**0.27**	**{7,6,5,19,10,29}**	**6**
8	6.41	-2.94	0.26	$\{8, 6, 12, 31\}$	7
9	6.34	-4.21	0.24	$\{9, 5, 6, 14, 31\}$	8
10	7.98	-3.50	0.07	$\{10, 24, 22\}$	9
11	9.04	-1.69	0.02	$\{11, 19\}$	10
12	8.94	-1.87	0.02	$\{12, 19\}$	11
13	6.00	-4.93	0.27	$\{13, 19, 10, 5, 6, 7, 29\}$	12
14	8.10	-3.13	0.06	$\{14, 10, 24, 16\}$	13
15	8.47	-2.60	0.04	$\{15, 21, 19, 13\}$	14
16	6.03	-4.52	0.28	$\{16, 5, 6, 7, 10, 29\}$	15
17	9.56	-1.13	0.01	$\{17, 22, 13\}$	16
18	7.37	-3.62	0.13	$\{18, 27, 3, 5, 8, 6, 10\}$	17
19	8.94	-1.87	0.02	$\{19, 12\}$	11
20	9.64	-1.04	0.01	$\{20, 11\}$	18
21	8.76	-2.00	0.03	$\{21, 16, 26, 13\}$	19
22	6.59	-4.76	0.20	$\{22, 8, 6, 5, 15, 7, 14, 30\}$	20
23	7.04	-3.78	0.17	$\{23, 7, 8, 6, 10, 16, 29\}$	21
24	7.98	-3.50	0.07	$\{24, 10, 22\}$	9
25	6.06	-3.22	0.30	$\{25, 8, 6, 12, 9, 18, 29\}$	22
26	6.16	-4.36	0.26	$\{26, 16, 5, 6, 7, 29\}$	23
27	6.56	-3.06	0.24	$\{27, 8, 6, 12, 16\}$	24
28	6.77	-3.93	0.20	$\{28, 5, 6, 8, 20, 25\}$	25
29	9.69	-1.03	0.00	$\{29, 4\}$	4
30	6.48	-3.71	0.24	$\{30, 5, 6, 14\}$	26
31	6.34	-4.21	0.24	$\{31, 5, 6, 9, 14\}$	8

Table 5.17: Results of SDSA on the scenario 3A↑.

Starting SVID	$F(S_A)$	$\alpha_A \times 10^{-4}$	R_A^2	S_A	**Group**
1	**5.199**	**8.77**	**0.2613**	**{1, 5, 21}**	**1**
2	6.052	5.62	0.2613	$\{5, 19, 20\}$	2
3	5.240	7.61	0.2613	$\{6, 15\}$	3
4	5.240	7.61	0.2613	$\{6, 15\}$	3
5	**5.199**	**8.77**	**0.2613**	**{5, 1, 21}**	**1**
6	5.240	7.61	0.2613	$\{6, 15\}$	2
7	**5.199**	**8.77**	**0.2613**	**{5, 1, 21}**	**1**
8	6.196	6.57	0.2613	$\{8, 23\}$	4
9	**5.199**	**8.77**	**0.2613**	**{1, 21, 5}**	**1**
10	6.196	6.57	0.2613	$\{8, 23\}$	4
11	**5.199**	**8.77**	**0.2613**	**{1, 5, 21}**	**1**
12	**5.199**	**8.77**	**0.2613**	**{1, 5, 21}**	**1**
13	**5.199**	**8.77**	**0.2613**	**{5, 1, 21}**	**1**
14	**5.199**	**8.77**	**0.2613**	**{1, 5, 21}**	**1**
15	5.240	7.61	0.2613	$\{15, 6\}$	3
16	**5.199**	**8.77**	**0.2613**	**{1, 5, 21}**	**1**
17	6.052	5.62	0.2613	$\{5, 20, 19\}$	2
18	6.196	6.57	0.2613	$\{23, 8\}$	4
19	6.052	5.62	0.2613	$\{19, 5, 20\}$	2
20	6.052	5.62	0.2613	$\{20, 5, 19\}$	2
21	5.240	7.61	0.2613	$\{15, 6\}$	3
22	5.240	7.61	0.2613	$\{6, 15\}$	3
23	6.196	6.57	0.2613	$\{23, 8\}$	4
24	5.240	7.61	0.2613	$\{15, 6\}$	3
25	5.240	7.61	0.2613	$\{15, 6\}$	3
26	5.240	7.61	0.2613	$\{15, 6\}$	3
27	5.240	7.61	0.2613	$\{6, 15\}$	3
28	6.052	5.62	0.2613	$\{5, 19, 20\}$	2
29	**5.199**	**8.77**	**0.2613**	**{5, 1, 21}**	**1**
30	5.240	7.61	0.2613	$\{15, 6\}$	3
31	6.196	6.57	0.2613	$\{23, 8\}$	4

Table 5.18: Results of greedy algorithm on the scenario 3A↑.

Starting SVID	$F(S_A)$	$\alpha_A \times 10^{-4}$	R^2_A	S_A	**Group**
1	**5.1993**	**8.77**	**0.2613**	**{1, 5, 21}**	**1**
2	7.0419	4.61	0.0834	$\{2, 19, 5, 21\}$	2
3	6.4787	5.71	0.1326	$\{3, 30, 24, 6, 9, 25, 29\}$	3
4	8.0013	2.62	0.0296	$\{4, 30, 10\}$	4
5	**5.1993**	**8.77**	**0.2613**	**{5, 1, 21}**	**1**
6	5.2400	7.61	0.2703	$\{6, 15\}$	5
7	7.1412	4.39	0.0763	$\{7, 14, 15\}$	6
8	6.1960	6.57	0.1544	$\{8, 23\}$	7
9	5.7385	6.74	0.2142	$\{9, 21, 6, 1\}$	8
10	6.7059	5.07	0.1162	$\{10, 23\}$	9
11	6.2932	5.61	0.1593	$\{11, 5, 22, 20, 24\}$	10
12	6.4285	5.41	0.1454	$\{12, 5, 22, 20, 24\}$	10
13	5.3501	7.70	0.2538	$\{13, 1, 21, 6, 25\}$	11
14	7.1412	4.39	0.0763	$\{14, 7, 15\}$	6
15	5.2400	7.61	0.2703	$\{15, 6\}$	5
16	6.7486	5.13	0.1092	$\{16, 7, 11, 22\}$	12
17	6.9520	3.92	0.1137	$\{17, 20, 13, 21, 6, 3, 1, 11\}$	13
18	7.1670	4.69	0.0650	$\{18, 10\}$	14
19	6.0521	5.62	0.1913	$\{19, 5, 20\}$	15
20	6.0521	5.62	0.1913	$\{20, 5, 19\}$	15
21	5.2319	8.58	0.2590	$\{21, 6, 1, 25\}$	16
22	6.6568	5.20	0.1197	$\{22, 24, 1, 6, 25\}$	17
23	6.1960	6.57	0.1544	$\{23, 8\}$	7
24	5.7799	7.17	0.2017	$\{24, 6, 1, 25, 30\}$	18
25	5.6727	7.22	0.2157	$\{25, 6, 30, 1, 21\}$	19
26	5.3814	8.18	0.2430	$\{26, 6, 21, 1\}$	20
27	6.1612	6.30	0.1640	$\{27, 21, 1, 6, 25\}$	21
28	6.3301	5.82	0.1503	$\{28, 19, 6, 21, 26, 1\}$	22
29	6.2191	5.49	0.1716	$\{29, 1, 5, 21, 19\}$	23
30	5.6727	7.22	0.2157	$\{30, 6, 25, 1, 21\}$	20
31	8.1786	2.41	0.0182	$\{31, 10, 2, 19\}$	23

Table 5.19: Results of SDSA on the scenario 7A↑.

Starting SVID	$F(S_A)$	$\alpha_A \times 10^{-4}$	R_A^2	S_A	**Group**
1	**6.15**	**5.11**	**0.19**	**{1, 5}**	**1**
2	8.64	1.44	0.02	$\{2, 16\}$	2
3	6.60	5.55	0.12	$\{6, 30\}$	3
4	6.60	5.55	0.12	$\{30, 6\}$	3
5	**6.15**	**5.11**	**0.19**	**{5, 1}**	**1**
6	6.60	5.55	0.12	$\{6, 30\}$	3
7	6.84	3.11	0.15	$\{15, 14, 23, 10\}$	4
8	**6.15**	**5.11**	**0.19**	**{5, 1}**	**1**
9	6.66	3.80	0.15	$\{23, 10\}$	5
10	6.66	3.80	0.15	$\{10, 23\}$	5
11	6.86	3.81	0.13	$\{14, 23, 6, 21, 26, 29, 13, 1\}$	6
12	6.86	3.81	0.13	$\{23, 14, 6, 21, 26, 29, 13, 1\}$	6
13	**6.15**	**5.11**	**0.19**	**{5, 1}**	**1**
14	6.84	3.11	0.15	$\{14, 15, 23, 10\}$	4
15	6.66	3.80	0.15	$\{10, 23\}$	5
16	8.64	1.44	0.02	$\{16, 2\}$	7
17	**6.15**	**5.11**	**0.19**	**{5, 1}**	**1**
18	6.66	3.80	0.15	$\{23, 10\}$	5
19	6.60	5.55	0.12	$\{30, 6\}$	3
20	**6.15**	**5.11**	**0.19**	**{1, 5}**	**1**
21	**6.15**	**5.11**	**0.19**	**{5, 1}**	**1**
22	6.66	3.80	0.15	$\{23, 10\}$	5
23	6.66	3.80	0.15	$\{23, 10\}$	5
24	**6.15**	**5.11**	**0.19**	**{5, 1}**	**1**
25	6.60	5.55	0.12	$\{30, 6\}$	3
26	6.60	5.55	0.12	$\{30, 6\}$	3
27	**6.15**	**5.11**	**0.19**	**{1, 5}**	**1**
28	**6.15**	**5.11**	**0.19**	**{5, 1}**	**1**
29	**6.15**	**5.11**	**0.19**	**{5, 1}**	**1**
30	6.60	5.55	0.12	$\{30, 6\}$	3
31	8.64	1.44	0.02	$\{16, 2\}$	4

Table 5.20: Results of greedy algorithm on the scenario 7A↑.

Starting SVID	$F(S_A)$	$\alpha_A \times 10^{-4}$	R^2_A	S_A	**Group**
1	**6.15**	**5.11**	**0.19**	**{1, 5}**	**1**
2	8.64	1.44	0.02	{2, 16}	2
3	8.62	1.66	0.01	{3, 30, 2, 26}	3
4	7.85	2.51	0.05	{4, 6, 24, 3, 13, 26, 10, 31, 16}	4
5	**6.15**	**5.11**	**0.19**	**{5, 1}**	**1**
6	6.60	5.55	0.12	{6, 30}	5
7	9.03	0.94	0.03	{7, 14}	6
8	7.87	2.41	0.06	{8, 1}	7
9	7.58	3.40	0.05	{9, 10118}	8
10	6.66	3.80	0.15	{10, 23}	9
11	8.60	1.68	0.01	{11, 12, 5, 22, 20, 4, 13, 1}	10
12	8.60	1.68	0.01	{12, 11, 5, 22, 20, 4, 13, 1}	10
13	7.99	2.24	0.05	{13, 1, 5}	11
14	6.84	3.11	0.15	{14, 15, 23, 10}	12
15	6.84	3.11	0.15	{15, 23, 10, 14}	12
16	8.64	1.44	0.02	{16, 2}	2
17	8.69	1.57	0.01	{17, 20, 2, 16}	13
18	8.34	2.02	0.02	{18, 10}	14
19	7.11	3.72	0.10	{19, 6, 23, 30, 25, 10}	15
20	6.65	4.57	0.14	{20, 5}	16
21	6.34	5.15	0.16	{21, 1, 6, 26, 23, 13}	17
22	8.35	1.39	0.06	{22, 10, 23}	18
23	6.66	3.80	0.15	{23, 10}	9
24	7.72	3.26	0.04	{24, 1}	19
25	6.25	4.56	0.19	{25, 6, 30, 1}	20
26	6.60	4.16	0.15	{26, 6, 30, 1}	20
27	7.91	2.50	0.05	{27, 24, 6, 23, 3, 26}	21
28	8.17	2.12	0.03	{28, 1, 25, 13}	22
29	8.02	2.41	0.04	{29, 1}	23
30	6.60	5.55	0.12	{30, 6}	5
31	8.53	1.78	0.01	{31, 2, 28, 26}	24

Table 5.21: Results of SDSA on the scenario 3D↓.

Starting SVID	$F(S_D)$	$\alpha_D \times 10^{-4}$	R_D^2	S_D	**Group**
1	**1.895**	**-16.1**	**0.7353**	**{27, 16, 26, 21, 5, 3}**	**1**
2	1.926	-15.7	0.7317	$\{5, 16, 27, 26, 17, 3, 21\}$	2
3	**1.895**	**-16.1**	**0.7353**	**{16, 27, 26, 21, 5, 3}**	**1**
4	**1.895**	**-16.1**	**0.7353**	**{16, 26, 21, 27, 5, 3}**	**1**
5	1.926	-15.7	0.7317	$\{16, 5, 27, 26, 17, 3, 21\}$	2
6	**1.895**	**-16.1**	**0.7353**	**{26, 16, 21, 27, 5, 3}**	**1**
7	1.926	-15.7	0.7317	$\{16, 27, 5, 26, 17, 3, 21\}$	2
8	1.926	-15.7	0.7317	$\{5, 16, 27, 26, 17, 3, 21\}$	2
9	1.926	-15.7	0.7317	$\{16, 5, 27, 26, 17, 3, 21\}$	2
10	1.926	-15.7	0.7317	$\{16, 5, 27, 26, 17, 3, 21\}$	2
11	1.926	-15.7	0.7317	$\{16, 27, 5, 26, 17, 3, 21\}$	2
12	1.926	-15.7	0.7317	$\{16, 27, 5, 26, 17, 3, 21\}$	2
13	1.926	-15.7	0.7317	$\{16, 26, 27, 5, 17, 3, 21\}$	2
14	1.926	-15.7	0.7317	$\{16, 5, 27, 26, 17, 3, 21\}$	2
15	8.835	-02.2	0.0161	$\{15, 18, 17\}$	3
16	**1.895**	**-16.1**	**0.7353**	**{16, 26, 27, 21, 5, 3}**	**1**
17	8.835	-02.2	0.0161	$\{17, 15, 18\}$	3
18	8.835	-02.2	0.0161	$\{18, 15, 17\}$	3
19	**1.895**	**-16.1**	**0.7353**	**{16, 26, 27, 21, 5, 3}**	**1**
20	**1.895**	**-16.1**	**0.7353**	**{26, 21, 16, 27, 5, 3}**	**1**
21	**1.895**	**-16.1**	**0.7353**	**{21, 26, 16, 27, 5, 3}**	**1**
22	**1.895**	**-16.1**	**0.7353**	**{16, 26, 21, 27, 5, 3}**	**1**
23	**1.895**	**-16.1**	**0.7353**	**{26, 21, 16, 27, 5, 3}**	**1**
24	**1.895**	**-16.1**	**0.7353**	**{26, 21, 16, 27, 5, 3}**	**1**
25	**1.895**	**-16.1**	**0.7353**	**{16, 27, 26, 21, 5, 3}**	**1**
26	**1.895**	**-16.1**	**0.7353**	**{26, 16, 27, 21, 5, 3}**	**1**
27	**1.895**	**-16.1**	**0.7353**	**{27, 16, 26, 21, 5, 3}**	**1**
28	**1.895**	**-16.1**	**0.7353**	**{16, 26, 21, 27, 5, 3}**	**1**
29	1.926	-15.7	0.7317	$\{27, 16, 5, 26, 17, 3, 21\}$	2
30	**1.895**	**-16.1**	**0.7353**	**{26, 21, 16, 27, 5, 3}**	**1**
31	1.926	-15.7	0.7317	$\{26, 16, 27, 5, 17, 3, 21\}$	2

Table 5.22: Results of greedy algorithm on the scenario 3D↓.

Starting SVID	$F(S_D)$	$\alpha_D \times 10^{-4}$	R^2_D	S_D	**Group**
1	2.175	-15.5	0.6977	$\{1, 3, 26, 16, 21, 27, 5\}$	1
2	3.382	-13.1	0.5396	$\{2, 9, 8, 29, 27, 16, 26, 21, 19\}$	2
3	2.175	-15.5	0.6977	$\{3, 1, 26, 16, 21, 27, 5\}$	1
4	2.281	-16.1	0.6814	$\{4, 23, 21, 16, 26, 27, 5, 30\}$	3
5	4.368	-12.1	0.4110	$\{5, 8, 15, 27, 26, 16, 21, 23, 14, 19, 4, 31\}$	4
6	9.073	-01.8	0.0125	$\{6, 4, 29\}$	5
7	8.580	-02.3	0.0428	$\{7, 29\}$	6
8	3.283	-14.1	0.5488	$\{8, 9, 29, 28, 30, 20, 21, 16, 26, 27, 19\}$	7
9	3.283	-14.1	0.5488	$\{8, 9, 29, 28, 30, 20, 21, 16, 26, 27, 19\}$	7
10	4.829	-10.9	0.3571	$\{10, 8, 5, 15, 27, 26, 16, 21, 6, 23, 14, 17\}$	8
11	8.981	-02.0	0.0118	$\{11, 29\}$	9
12	9.206	-01.5	0.0174	$\{12, 29\}$	10
13	2.585	-14.1	0.6447	$\{13, 31, 29, 20, 28, 21, 16, 26, 27, 17\}$	11
14	3.394	-12.8	0.5391	$\{14, 8, 9, 28, 22, 30, 21, 16, 26, 27\}$	12
15	8.835	-02.2	0.0161	$\{15, 18, 17\}$	13
16	**1.895**	**-16.1**	**0.7353**	**{16, 26, 27, 21, 5, 3}**	**14**
17	8.835	-02.2	0.0161	$\{17, 15, 18\}$	13
18	8.835	-02.2	0.0161	$\{18, 15, 17\}$	13
19	3.663	-13.3	0.5002	$\{19, 25, 27, 21, 16, 7, 17, 20, 14, 31\}$	15
20	1.956	-16.1	0.7268	$\{20, 21, 16, 26, 27, 7\}$	20
21	**1.895**	**-16.1**	**0.7353**	**{21, 26, 16, 27, 5, 3}**	**14**
22	2.890	-15.7	0.5975	$\{22, 19, 21, 16, 26, 27, 17, 7, 3\}$	21
23	2.195	-16.4	0.6927	$\{23, 21, 16, 26, 27, 5, 30\}$	22
24	4.857	-12.7	0.3409	$\{24, 21, 16, 26, 27, 19, 5, 28, 1, 31\}$	23
25	3.663	-13.3	0.5002	$\{25, 27, 16, 21, 19, 7, 17, 20, 14, 31\}$	15
26	**1.895**	**-16.1**	**0.7353**	**{26, 16, 27, 21, 5, 3}**	**14**
27	**1.895**	**-16.1**	**0.7353**	**{27, 16, 26, 21, 5, 3}**	**14**
28	2.286	-16.0	0.6809	$\{28, 20, 21, 16, 26, 27, 7\}$	24
29	3.639	-12.3	0.5081	$\{29, 5, 8, 28, 20, 21, 16, 26, 27, 19, 17\}$	25
30	1.925	-16.1	0.7310	$\{30, 21, 26, 16, 27, 5\}$	26
31	2.585	-14.1	0.6447	$\{31, 13, 29, 20, 28, 21, 16, 26, 27, 17\}$	12

Table 5.23: Results of SDSA on the scenario 7D↓.

Starting SVID	$F(S_D)$	$\alpha_D \times 10^{-4}$	R_D^2	S_D	**Group**
1	1.05	-14.2	0.86	$\{16, 11, 21, 31, 26\}$	1
2	1.09	-15.8	0.85	$\{17, 26, 21, 16, 27\}$	2
3	**1.05**	**-14.2**	**0.86**	**{21, 11, 16, 26}**	**3**
4	1.09	-15.8	0.85	$\{26, 17, 21, 16, 27\}$	2
5	8.68	-2.12	0.04	$\{7, 29\}$	4
6	8.92	-2.04	0.02	$\{6, 15, 17, 29, 31, 14, 4, 22\}$	5
7	8.68	-2.12	0.04	$\{7, 29\}$	4
8	**1.05**	**-14.2**	**0.86**	**{16, 21, 11, 26}**	**3**
9	**1.05**	**-14.2**	**0.86**	**{21, 16, 11, 26}**	**3**
10	**1.05**	**-14.2**	**0.86**	**{26, 11, 16, 21}**	**3**
11	8.68	-2.12	0.04	$\{7, 29\}$	4
12	9.14	-1.68	0.01	$\{12, 13, 29\}$	5
13	9.14	-1.68	0.01	$\{13, 12, 29\}$	5
14	8.68	-2.12	0.04	$\{7, 29\}$	4
15	**1.05**	**-14.2**	**0.86**	**{11, 26, 16, 21}**	**3**
16	**1.05**	**-14.2**	**0.86**	**{16, 21, 26, 11}**	**3**
17	**1.05**	**-14.2**	**0.86**	**{16, 21, 26, 11}**	**3**
18	**1.05**	**-14.2**	**0.86**	**{26, 11, 16, 21}**	**3**
19	8.82	-2.24	0.02	$\{19, 23, 22, 25, 13, 30, 14\}$	6
20	8.68	-2.12	0.04	$\{7, 29\}$	4
21	**1.05**	**-14.2**	**0.86**	**{21, 16, 26, 11}**	**3**
22	8.82	-2.24	0.02	$\{22, 19, 23, 25, 13, 30, 14\}$	6
23	8.82	-2.24	0.02	$\{23, 19, 22, 25, 13, 30, 14\}$	6
24	8.82	-2.24	0.02	$\{23, 19, 22, 25, 13, 30, 14\}$	6
25	8.68	-2.12	0.04	$\{29, 7\}$	4
26	**1.05**	**-14.2**	**0.86**	**{26, 21, 16, 11}**	**3**
27	**1.05**	**-14.2**	**0.86**	**{21, 16, 26, 11}**	**3**
28	**1.05**	**-14.2**	**0.86**	**{11, 16, 21, 26}**	**3**
29	8.68	-2.12	0.04	$\{29, 7\}$	4
30	8.82	-2.24	0.02	$\{30, 23, 19, 22, 25, 13, 14\}$	6
31	**1.05**	**-14.2**	**0.86**	**{16, 21, 26, 11}**	**3**

Table 5.24: Results of greedy algorithm on the scenario 7D↓.

Starting SVID	$F(S_D)$	$\alpha_D \times 10^{-4}$	R^2_D	S_D	**Group**
1	2.54	-12.4	0.66	$\{1, 28, 29, 14, 26, 21, 16, 27, 17, 5\}$	1
2	8.75	-2.27	0.02	$\{2, 4, 7, 8, 9, 28, 29, 31, 14\}$	2
3	8.80	-2.28	0.02	$\{3, 25, 8, 10, 29, 31\}$	3
4	8.75	-2.27	0.02	$\{4, 2, 7, 8, 9, 28, 29, 31, 14\}$	2
5	8.94	-1.76	0.03	$\{5, 29\}$	4
6	9.00	-1.92	0.01	$\{6, 12, 17, 29, 31, 14, 4, 15\}$	5
7	8.68	-2.12	0.04	$\{7, 29\}$	6
8	1.78	-13.0	0.76	$\{8, 9, 27, 16, 21, 26, 17\}$	7
9	1.78	-13.0	0.76	$\{9, 8, 27, 16, 21, 26, 17\}$	7
10	1.11	-15.5	0.85	$\{10, 27, 16, 21, 26, 17\}$	8
11	9.32	-1.36	0.01	$\{11, 29, 31\}$	9
12	9.14	-1.68	0.01	$\{12, 13, 29\}$	10
13	9.14	-1.68	0.01	$\{13, 12, 29\}$	10
14	2.41	-13.0	0.67	$\{14, 29, 28, 26, 21, 16, 27, 5\}$	11
15	2.41	-12.4	0.67	$\{15, 10, 8, 27, 16, 21, 26, 17\}$	12
16	**1.05**	**-14.2**	**0.86**	**{16, 21, 26, 11}**	**13**
17	1.09	-15.8	0.85	$\{17, 27, 16, 21, 26\}$	14
18	8.79	-2.27	0.02	$\{18, 15, 25, 29, 4, 14, 22, 23, 31\}$	15
19	8.72	-2.29	0.03	$\{19, 24, 22, 25, 31, 4, 29, 14\}$	16
20	2.47	-12.2	0.67	$\{20, 29, 31, 14, 4, 26, 21, 16, 27\}$	17
21	**1.05**	**-14.2**	**0.86**	**{21, 16, 26, 11}**	**13**
22	8.82	-2.24	0.02	$\{22, 19, 23, 25, 13, 30, 14\}$	18
23	8.82	-2.24	0.02	$\{23, 19, 22, 25, 13, 30, 14\}$	18
24	8.72	-2.29	0.03	$\{24, 19, 22, 25, 31, 4, 29, 14\}$	16
25	8.74	-2.47	0.02	$\{25, 7, 8, 15, 3, 22, 17, 30\}$	19
26	**1.05**	**-14.2**	**0.86**	**{26, 21, 16, 11}**	**13**
27	1.09	-15.8	0.85	$\{27, 16, 21, 26, 17\}$	14
28	2.41	-13.0	0.67	$\{28, 14, 29, 26, 21, 16, 27, 5\}$	11
29	8.68	-2.12	0.04	$\{29, 7\}$	6
30	8.71	-2.32	0.03	$\{30, 24, 19, 22, 25, 14, 29, 4\}$	20
31	1.05	-14.2	0.86	$\{31, 21, 16, 26, 11\}$	21

Chapter 6

Conclusion and Future Research Direction

The thesis and future research directions are concluded in this chapter. In **Section 6.1**, each chapter is summarized. The proposed equipment failure diagnosis and the equipment behavior prognosis are concluded in **Sections 6.2** and **Sections 6.3**, respectively. Especially, initial developments and experiments regarding the future research directions are demonstrated in **Section 6.3**.

6.1 Chapters Summary

In **Chapter 1**, a general introduction to this thesis starts by explaining the problem background and motivations is provided. The underlying problem is stated, and a general research methodology is described in this chapter. In **Chapter 2**, the relevant scientific studies to position this work in the literature regarding originality and contributions are reviewed. A comprehensive terminology of the equipment behavior modeling, monitoring, and different prognostic and diagnostic methods is provided. The Advanced Process Control (APC) in the semiconductor industry, and corresponding challenges are elaborated. The relevant papers studying equipment behavior prognosis and equipment failure diagnosis are reviewed. Based on the conducted review, the gaps between the literature and this research are analyzed and expected to close by the proposed approaches. Finally, the detail research methodology in this thesis and its contributions are explained. In **Chapter 3**, a data-driven equipment failure diagnosis approach that uses the Fault Detection and Classification (FDC) data from the semiconductor industry is developed to find out fault fingerprints with their corresponding fault roots. The well-known supervised and unsupervised techniques, such as principal component analysis (PCA), support vector machine (SVM) and K-means algorithm, are employed to build a novel equipment failure diagnosis approach. In **Chapter 4**, an efficient equipment behavior prognosis approach for modeling and monitoring the equipment deterioration using the same FDC data is proposed. The approach is explained step by step from FDC data decomposition to deterioration trends extraction. A new searching mechanism is developed in this chapter to find out the roots of equipment deterioration. In **Chapter 5**, comprehensive experimental results for both of the proposed equipment failure di-

agnosis and the equipment behavior prognosis are presented. The results are analyzed, and numerous engineering insights are extracted and validated.

6.2 Proposed Equipment Failure Diagnosis Approach

The first proposed approach in this thesis was a data-driven equipment failure diagnosis to analyze the semiconductor equipment data that may contain different types of recipe-related or fault-related variations. The proposed approach distinguished the normal recipe-related variations (e.g., process dynamics) from the fault-related variations. The proposed data-driven approach contained two main stages as *Equipment Anomaly Detection* and *Automatic Fault Fingerprint Extraction*.

These stages were performed by employing the conventional soft computing algorithms such as Support Vector Machine (SVM), K-Means clustering algorithm, and Principal Component Analysis (PCA). In the proposed two-stage methodology, Stage 1 detects the normal from the abnormal conditions using SVM, while Stage 2 projects the abnormal data into different PCA models which are built to preserve the process dynamics. By summarizing the variable contributions to the Out-of-Control (OOC) observations, fault fingerprints and their corresponding cause roots were automatically extracted using the K-means algorithm.

Different statistical tests were conducted to justify the use of the soft computing algorithms in the proposed approach. These tests showed the superiority of the proposed algorithms with specific settings in comparison to other well-known algorithms. To validate the proposed approach, the FDC data of 760 wafers (264,340 observations in total) provided by a local IC maker were analyzed. The data cover eight different recipes in an etching process, wherein 31 sensor readings were collected simultaneously during a process run, i.e., a wafer.

As the prior information, it was known that there exists a severe process drift followed by a Corrective Maintenance (CM) within the data duration. Through the case study, different fault fingerprints were extracted with their corresponding root causes. The results of the case study show convincing practicability of the proposed equipment failure diagnosis approach. In particular, the three fault fingerprints, which were extracted automatically, presented consistent causes with the known faults to the process drift in the FDC data.

In the proposed methodology, the FDC data are labeled as normal and abnormal, which can be further simplified to be only one normal class for training a one-class SVM model.

Nevertheless, the normal wafers and their FDC data must be carefully filtered and selected to ensure that the SVM model can learn and remember the normal process dynamics, including different standard recipe bodies and the tolerable natural noises of the sensor readings. Since the data-driven SVM model is subject to the data complexity and resolution, as long as the normal condition may be changed, e.g., performing the corrective/preventive maintenances (CM/PM), recovering from machine breakdown, or processing a new recipe/product, the SVM model shall be refreshed or rebuilt to cover

as much (normal) information as possible. By this, the normal condition and the characteristics of normal wafer are well defined and readily available based on the recipes. Modeling the faulty situations depends on whether enough information about the fault(s) characteristics are available or not. In case of having enough information about the failures, they can be easily modeled by two (multiple)-class SVM. If not, a One-class SVM is employed wherein any record that does not meet the normal condition is considered as abnormal record. Therefore, we emphasize that knowing the normal condition well is sufficient to employ SVM.

6.2.1 Contributions

The main contributions of the proposed equipment failure diagnosis approach have been:

- *Differentiating two kinds of variations in the FDC data.* The proposed approach deals with two types of variations in the data as normal recipe-related variation (i.e., process dynamics) and fault-related variations. These two variations must not be confounded. Regarding this issue, the dynamics of a single wafer process run from the data-set were decomposed into 10 sequential deciles and it was observed that observations from the ramp-up and the ramp-down steps of the wafer process were grouped into mainly two clusters while the middle of the process containing the main etching steps is distributed dispersedly in different clusters. It was then learned intuitively that the main etching steps are decidedly non-stationary and compose of interchanging upward and downward trends with different distributions. Accordingly, coping with process dynamic is an essential task in most of fault diagnosis approaches, particularly in the semiconductor industry.
- *Extracting the fault fingerprints with their corresponding cause roots.* The proposed approach clusters the contribution plots of the out-of-control observations and obtains the highest possible number of fault fingerprints and finally identified their corresponding cause roots. The contribution values and their control limits presented in this approach are capable of capturing the fault fingerprints corresponding to the abnormal process events. By looking into the fault fingerprints, the process variables to be responsible for the anomaly can be easily identified. The proposed approach extracts the unknown fault fingerprints despite the literature that in most of the studies, the faulty profiles are usually given.
- *Proposing a control limit for identifying the fault cause roots.* Most of the existing studies in the literature consider the SVIDs with the higher contribution value as the fault roots, while a higher value of contribution values does not necessarily mean the fault causes. Hence, although the contribution values of some SVIDs are high, they must not be considered as the fault roots because they do not pass their control limit. This approach proposes a control limit for the contribution value of each SVID based on the normal observations and those SVIDs are identified as the fault roots if they pass the proposed control limits.

6.2.2 Future Research Direction

One of the major future research directions for the equipment failure diagnosis approach is dealing with multiple recipes which are remarkably different in profiles. Actually, the recipes in this study belonged to a family of product; therefore, there was no need to build different normal baselines for different recipes. Accordingly, coping with different recipes from different product families could be a challenging issue that requires to be investigated.

Another possible research direction could be summarized the observation cubes into wafer-based features, whereby the proposed methodology might become simpler. For example, process dynamic decomposition can be eliminated since process dynamic is incorporated in wafer-based features.

6.3 Proposed Equipment Behavior Prognosis Approach

The second proposed approach in this thesis was a data-driven equipment behavior prognosis for modeling and monitoring the equipment deterioration in a semiconductor industry using the batch process data. The proposed approach has two aims as 1) exploiting the temporal data of batch processes to characterize the equipment behavior, and 2) identifying the deterioration trend with the most likely causes. In this regard, several challenges appeared in this study that have made it difficult in employing classical techniques and methods in the literature to efficiently and effectively model and monitor the equipment deterioration.

This proposed prognostic approach included four main consecutive stages as 1) Data pretreatment, 2) Data processing, 3) Feature extraction, and 4) Deterioration modeling. The data pretreatment stage prepared the FDC data by deleting the constant and non-informative SVIDs based on the domain knowledge of engineers. Similar to the equipment failure diagnosis problem, one of the challenges in front of the proposed prognostic approach was to distinguish different types of variations in the FDC data-set. The proposed approach encountered two macro-level and micro-level variations with a different domain of frequencies; the former were low-frequency variations, and the latter one was high-frequency variations. Accordingly, in the data processing stage, the discrete wavelet transformation (DWT) technique was employed to distinguish the macro-level variation from the micro-level ones in each SVID using the most appropriate selected mother wavelets. After decomposing the SVIDs and capturing the macro-level and micro-level variation, respectively in the approximation and the detail signals, equipment deterioration features were calculated in the feature extraction stage. By knowing some physical meanings behind the semiconductor etching process, the determinant of correlation between decomposed SVIDs was considered as deterioration features. Finally, a stepwise searching mechanism was developed to find out the most significant deterioration trends from the approximation the detail signals with the contributing SVIDs to these trends.

The normal wafers of the same FDC data in the first proposed approach (i.e., equipment failure diagnosis) was used to validate the proposed equipment deterioration mod-

eling and monitoring approach. The most increasing and decreasing deterioration trends were illustrated with corresponding contributing SVIDs, and it was observed that the gas flow SVIDs are mostly the representatives of the equipment deterioration.

The proposed stepwise searching algorithm was benchmarked with a greedy algorithm and the well-known genetic algorithm and its performance regarding solution quality, convergence rate and computational time was verified through a comprehensive set of experiments. The proposed approach can automatically adapt to the new data-set and can extract increasing/decreasing deterioration trends from different macro-level and micro-level variations.

6.3.1 Contributions

The main contributions of the proposed equipment behavior prognosis approach have been:

- *Differentiating different types of variations in the FDC data.* An important challenge in the proposed approach was the presence of different kinds of variations in the FDC data of the semiconductor equipment. The existing prognostic approaches only consider one type of variation while equipment deterioration may appear in both macro-level and micro-level variations in the data. Accordingly, the proposed approach decomposed the signals into approximation (high-frequency) and detail (low-frequency) signals and extracted independent deterioration trend from these signals. The experimental results were also demonstrated that confounded variations in the signals might lead to less accurate trends if any trend can be extracted. Each extracted trend can be independently followed to monitor the equipment behavior.
- *Combining the discrete wavelet transformation with correlation determinant.* This combination provides an efficient approach to account for both variation and co-variation of the signals in two macro and micro levels of variations. The wavelet transformation avoids the convolution of faulty-related variations and the correlation determinant attempts to the co-variation between the signals. Each of the deterioration models obtained from the approximation signals (macro level) and the detail signals (micro level) can be utilized to monitor and prognose the equipment behavior since they independently alarm particular cause roots. Based on the provided cause roots, engineers can select and adapt further precautions.
- *Proposing an efficient and effective stepwise searching algorithm.* It has been the fact that deterioration usually appears in only certain SVIDs and involving all SVIDs to build the deterioration model may not lead to any significant trend. This issue was also demonstrated through the experimental results. Therefore, the proposed stepwise algorithm looks for the contributing SVIDs in a greedy manner where two forward-selection and backward-elimination operators are combined to escape from the local optimum.
- *Proposing a simple and practical deterioration modeling and monitoring approach.* Despite the most existing researches in the literature, the proposed approach is not

only novel regarding academic aspects, but also effective and easy to understand from the practical point of view.

6.3.2 Future Research Directions

In this section, two main future research directions are proposed as 1) Monitoring the equipment health/condition using the proposed equipment behavior prognosis approach and 2) Overcoming the overfitting problem (if any) of the proposed SDSA searching algorithm.

- ***Equipment health/condition monitoring***

To make the proposed equipment behavior prognosis approach practical for the industry, the deterioration models should be integrated into the monitoring scheme of the equipment condition. The monitoring mechanism can be utilized to optimize the production control, such as the job scheduling and maintenance planning.

A conventional way of monitoring the process/equipment condition is to use the statistical process control (SPC) charts to determine if a quality/behavior related feature is no longer under control. In the aspect of equipment failure diagnosis, the SPC chart could be used to judge the equipment stability. Furthermore, the SPC chart is expected to detect the unexpected anomalies (i.e., sudden drift or deterioration) in the equipment behavior prognosis. Control limits are usually determined based on the 3-sigma criterion given the assumption that the monitored statistics are following a normal distribution, wherein 99.73% of the data should fall into the range of $\mu \pm 3\sigma$ accompanied with the type I error of 0.27%.

To apply the common monitoring scheme of $\bar{x}$ control chart, the underlying distribution for the proposed equipment health index (HI) in equation (4.7) should be studied. Per our investigation, the determinant of a correlation matrix can approximately follow a normal distribution. Due to the lack of theoretical proof, a simulation study is conducted to learn the approximating behavior of the determinants, i.e., the HI.

In this regard, a large number of realistic correlation matrices are generated for each wafer and then these correlations matrices are transformed into determinant values. Having enough determinant values (e.g., 10000 samples) for each wafer makes it possible to test if the determinant follows a normal distribution. For generating numerous realistic correlation matrices, firstly the contributing SVIDs S_A in each wafer are normalized. By this normalization, the covariance and correlation matrices become identical. By putting the correlation matrix in the Wishart distribution [167] with a degree of freedom equal to the number of observations in wafer k, it is possible to generate the random realistic correlation matrices. Afterward, the generated correlation matrices are transformed to determinant values. As an initial evaluation, it was observed that the determinant values follow the Gaussian distribution. **Figure 6.1** shows the histogram of determinant values of thousand generated correlation matrices from a sample wafer.

For supporting the results of the simulation, different statistical tests such as Kolmogorov-Smirnov and Chi-square goodness of fit tests [163] were conducted to investigate whether

the simulated determinant values of each wafer follow a normal distribution. These tests were performed for numerous wafers and the results for almost all wafers showed that the determinant values follow a normal distribution at 5% significance level. Now, the control charts $\mu \pm 3\sigma$ can be directly applied on the deterioration trends obtained from approximation (i.e., HI_A) and detail (i.e., HI_D) signals. First, we draw these control limits over HI_D. For this aim, we calculate HI_D without imposing the micro-level variations into the SVIDs. Since the detail signals are not originally correlated, the HI_D trend becomes quiet stable. **Figure 6.2** shows the control charts over the HI_D values.

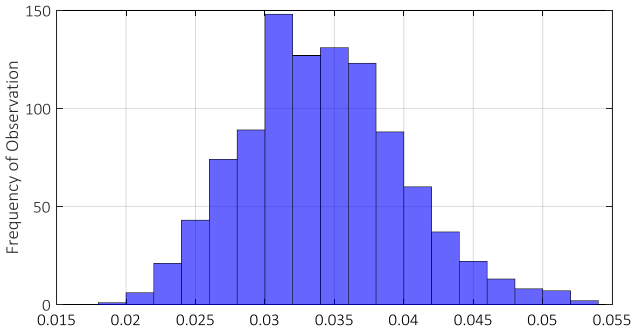


Figure 6.1: Histogram chart for a sample wafer shows the Gaussian distribution of determinant values of the generated correlation matrices.

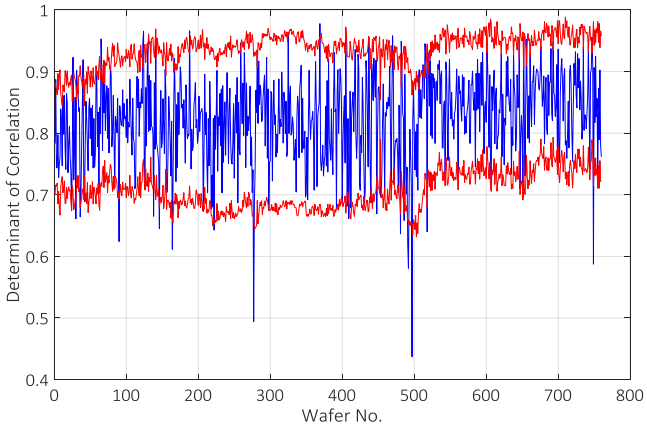


Figure 6.2: Control charts are built over the determinant values of the detail signals.

The $\mu \pm 3\sigma$ control charts cannot be directly implemented over the determinant values of the approximation signals since the HI_A are non-stationary. For this aim, the values of μ should be estimated for the HI_A trend. In the following, the idea of the Exponentially Weighted Moving Average (EWMA) method is adapted to estimated μ for each wafers.

Let $HI_A(k)$ be the health indicator based on the approximation signals, i.e., the determinant of correlation between the contributing SVIDs S_A for wafer k. The EWMA statistic Z_k of the health indicator for wafer is calculated as follows:

$$Z_k = \lambda HI_A(k) + (1 - \lambda)Z_{k-1} \tag{6.1}$$

where λ is the smoothing constant. A starting value z_0 also should be defined before the first sample is taken. If a target value μ is specified, then $z_0{=}\mu$. Otherwise, it is typical to use the average of some preliminary data points (i.e., $z_0{=}\bar{x}$). In this section, the average of the first five samples are considered. The weighting constant λ indicates how much previous observations influence the current EWMA Z_k. Values of λ near to 1 put almost all weight on the current observation. On the other hand, for values of λ near 0, a small weight is applied to almost all of the past observations, and consequently the, EWMA statistics become smoother. It is worth mentioning that since the EWMA is a weighted average of the current and all past observations, it is generally insensitive to the normality assumption. Therefore, it can be a useful controlling chart procedure to use with individual observations.

Based on equation (6.1), the proposed upper and lower control limits of wafer k, $UCL_{(k)}$ and $LCL_{(k)}$, are provided as equations (6.2) and (6.3), respectively.

$$UCL_{(k)} = Z_{(k)} + 3\sigma_{(k)} \quad \forall k \tag{6.2}$$

$$LCL_{(k)} = Z_{(k)} - 3\sigma_{(k)} \quad \forall k \tag{6.3}$$

where $Z_{(k)}$ is the EWMA statistic for the determinant value of wafer k and $\sigma_{(k)}$ is the standard deviation of the determinant values of the generated correlation matrices. For validating the proposed control limits (6.2) and (6.3), the FDC data containing 760 wafers (479 normal wafers + 281 abnormal wafers) is utilized. As mentioned in **Section 5.1**, an unexpected Corrective Maintenance (CM) event was performed after the 480^{th} wafer to fix a process drift. Unfortunately, the fault was not completely resolved, and the process remained unstable until the end of the data-set where a periodic Preventative Maintenance (PM) was executed. **Figure 6.3** shows the HI_A for all 760 wafers and the control limits for each wafer. The λ is considered equal to 0.01 for calculating the EWMA statistics. As it can be seen, a sudden and harsh drift has happened after 480^{th} wafer, and this jump continues until 505^{th} wafer once the CM is performed and the determinant values go down but still not into the expected range as wafers 1 to 479. In addition, the equipment behavior between wafers 400 to 480 has been zoomed in and it can be seen that the determinant value of several wafers has been out of the control limits. If this behavior is repeated unexpectedly, require actions should be taken. One way of utilizing this control chart is raising an alarm once the determinant value jumps out of the control charts like those of wafers 480 to 505. This alarm is to notify the engineers for taking required actions to stop producing abnormal wafers.

An important issue is resetting the EWMA statistics after the CM performed, i.e., wafer 505. This resetting is necessary to avoid the accumulation of the faults (of wafers 480 to 505) into the EWMA statistics of wafers 506 to 760. Therefore, the new control limits are calculated for post-CM.

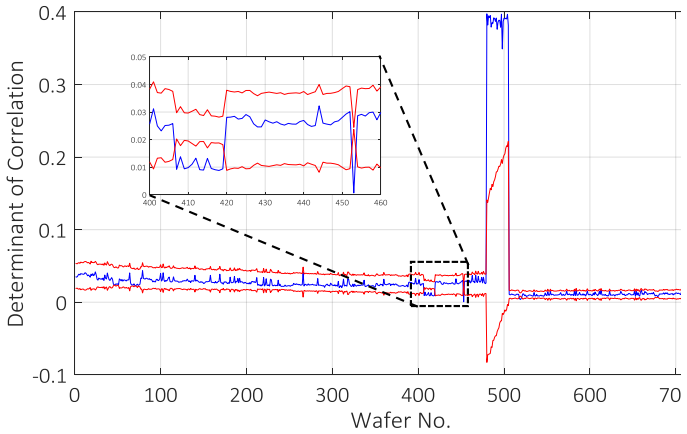


Figure 6.3: Determinant values are monitored using the control charts to alarm the unexpected and harsh behavioral changes.

Using the proposed control limits, the determinant of correlation between the contributing SVIDs should stay in-control to be sure that the equipment is working normally. Approaching the lower or upper bounds indicates underlying behavioral changes in the SVIDs that could be either due to equipment deterioration or failure propagation in the signals. Using this control limits, it is possible also to detect the sudden drifts due to unexpected severe failures. Finally, the EWMA statistics alongside the control limits can be utilized to predict the future condition of the equipment. If this condition can be well associated with the health of the equipment, the proposed deterioration model of **Chapter 4** with the control limits (6.2) and (6.3) can be employed to estimate the remaining useful life (RUL) of the equipment.

- ***Overfitting of the SDSA***

In this part of the future research direction, we briefly study the problem of overfitting in the proposed SDSA algorithm. As explained in **Chapter 4** (**Section 4.3**), the SDSA is a constructive iterative stepwise searching algorithm that employs two forward-selection and backward-elimination operators. Forward and backward stepwise selection are not guaranteed to give us the best set of contributing SVIDs but that is the price to pay in order to avoid overfitting. Even if the number of all SVIDs is small (i.e., less than 30), looking at all possible combination of SVIDs may not be the best research to do. It is not convincing that if the best set of SVIDs leads to the most significant deterioration trend on the training data, this set is really going to be the best overall set in the context of test data or new coming data, which is what we definitely care about. Also, forward-selection and backward-elimination somehow overcome overfitting issue comparing to the complete search, but still there are some issues to solve in this

regard. Overfitting occurs when a model begins to memorize training data rather than learning to generalize from the trend. In the other word, overfitting generally occurs when a model is excessively complex, such as having too many parameters to be selected (i.e., 31 SVIDs in this thesis) relative to the number of observations (i.e., 479 normal wafers in this thesis). In order to avoid overfitting in an algorithm, it is necessary to use additional techniques such as the early stopping of the algorithm [168, 169] or restricted forward-selection [170]. Early stopping rules provide guidance as to how many iterations can be run before the learner begins to over-fit [169]. These techniques can be applied to the proposed SDSA to observe how the performance of the SDSA changes.

In order to apply early stopping in the proposed SDSA, the data-set of the normal wafers should be divided into training and validation sets. For this aim, a significant volume of data should be provided; however, this thesis suffered from lack of big data. The idea of dividing the data-set is to obtain the contributing SVIDs from the training set then to verify if this set gives again the most significant trend for the validation test. The idea of implementing early stopping in the proposed SDSA is easy to implement. Accordingly, the SDSA iterative training process is stopped once the generalization accuracy starts to drop. This generalization performance is obtained by withholding a sample of the data (the validation set) [170]. Actually, the iterative process starts once the objective function on the validation set starts to worsen. At this point, the training process can be stopped and the contributing set of SVIDs is reported. Another way of overcoming the overfitting issue in the proposed SDSA is using the restricted forward-selection technique [171]. In this technique, the idea is to ignore extra searches of SVIDs which are not really relevant and do not significantly improve the objective function. This technique can be implemented as follows:

1. Starting from each SVID, create all pairs of SVIDs and calculate the objective function of each pair (we have totally J-1 number of pairs at each start of the algorithm, where J is the number of all SVIDs). Then, sort the pairs of SVIDs in terms of objective function from the most important pair P_1 to the least important pairs P_{J-1} (i.e., $P_1 < P_2 < \cdots < P_{J-1}$ in terms of objective function);
2. Select the SVID to enter that make the pair P_1 with the starting SVID (the best pair in terms of objective function).
3. Add the next SVID from either P_2 or P_3, or any other one until $P_{(J-1)/2}$. There are $((J-1)/2)-1$ of these pairs. The winner of this round will be the third SVID to be added to the list;
4. Add the next SVID from either P_2 or P_3, or any other one until $P_{(J-1)/3}$. There are $((J-1)/3)-1$ of these pairs. The winner of this round will be the forth SVID to be added to the list;
5. Continue this procedure, until no improvement in the objective function is found by adding a new SVID to the list.

Bibliography

[1] Argon Chen and G. S. Wu. Real-time health prognosis and dynamic preventive maintenance policy for equipment under aging Markovian deterioration. *International Journal of Production Research*, 45(15):3351–3379, August 2007.

[2] Ranganath Kothamasu, Samuel H. Huang, and William H. VerDuin. System health monitoring and prognostics a review of current paradigms and practices. *The International Journal of Advanced Manufacturing Technology*, 28(9-10):1012–1024, July 2006.

[3] Rosmaini Ahmad and Shahrul Kamaruddin. An overview of time-based and condition-based maintenance in industrial application. *Computers & Industrial Engineering*, 63(1):135–149, August 2012.

[4] Tran Van Tung and Bo-Suk Yang. Machine Fault Diagnosis and Prognosis: The State of The Art. *International Journal of Fluid Machinery and Systems*, 2(1):61–71, 2009.

[5] Jay Lee, Fangji Wu, Wenyu Zhao, Masoud Ghaffari, Linxia Liao, and David Siegel. Prognostics and health management design for rotary machinery systemsReviews, methodology and applications. *Mechanical Systems and Signal Processing*, 42(1):314–334, January 2014.

[6] J. Z. Sikorska, M. Hodkiewicz, and L. Ma. Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing*, 25(5):1803–1836, July 2011.

[7] George Vachtsevanos, Frank Lewis, Michael Roemer, Andrew Hess, and Biqing Wu. Fault Prognosis. *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*, March 2007.

[8] T. B. Lien Nguyen, Mohand Djeziri, Bouchra Ananou, Mustapha Ouladsine, and Jacques Pinaton. Fault prognosis for batch production based on percentile measure and gamma process: Application to semiconductor manufacturing. *Journal of Process Control*, 48:72–80, December 2016.

[9] Z. Xi, R. Jing, P. Wang, and C. Hu. A Copula-based sampling method for data-driven prognostics and health management. In *2013 IEEE Conference on Prognostics and Health Management (PHM)*, pages 1–10, June 2013.

[10] Khanh Le Son, Mitra Fouladirad, Anne Barros, Eric Levrat, and Benot Iung. Remaining useful life estimation based on stochastic deterioration models: A comparative study. *Reliability Engineering & System Safety*, 112:165–175, April 2013.

[11] Gang Li, S. Joe Qin, Yindong Ji, and Donghua Zhou. Reconstruction based fault prognosis for continuous processes. *Control Engineering Practice*, 18(10):1211–1219, October 2010.

[12] Dongyang Dou and Shishuai Zhou. Comparison of four direct classification methods for intelligent fault diagnosis of rotating machinery. *Applied Soft Computing*, 46:459–468, September 2016.

[13] Hongyang Yu, Faisal Khan, and Vikram Garaniya. Nonlinear Gaussian Belief Network based fault diagnosis for industrial processes. *Journal of Process Control*, 35:178–200, November 2015.

[14] Andrew K. S. Jardine, Daming Lin, and Dragan Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483–1510, October 2006.

[15] Hamideh Rostami, Jakey Blue, and Claude Yugma. Automatic equipment fault fingerprint extraction for the fault diagnostic on the batch process data. *Applied Soft Computing*, 68:972–989, July 2018.

[16] Heinz P. Bloch and Fred K. Geitner. *Machinery Failure Analysis and Troubleshooting: Practical Machinery Management for Process Plants*. Butterworth-Heinemann, December 2012. Google-Books-ID: NdDmayNBAskC.

[17] Jaime Campos. Development in the application of ICT in condition monitoring and maintenance. *Computers in Industry*, 60(1):1–20, January 2009.

[18] D. Miljkovi. Fault detection methods: A literature survey. In *2011 Proceedings of the 34th International Convention MIPRO*, pages 750–755, May 2011.

[19] ISO 13381-1:2015 - Condition monitoring and diagnostics of machines – Prognostics – Part 1: General guidelines.

[20] Tatiana Biagetti and Enrico Sciubba. Automatic diagnostics and prognostics of energy conversion processes via knowledge-based systems. *Energy*, 29(12):2553–2572, October 2004.

[21] H. Rostami, J. Blue, C. Yugma, and J. Pinaton. Equipment health modeling for deterioration prognosis and fault signatures diagnosis. In *2017 28th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, pages 357–362, Saratoga Springs, NY, May 2017.

[22] Manjeevan Seera, Chee Peng Lim, Dahaman Ishak, and Harapajan Singh. Offline and online fault detection and diagnosis of induction motors using a hybrid soft computing model. *Applied Soft Computing*, 13(12):4493–4507, December 2013.

[23] Bo Li, Ting Han, and Fuyong Kang. Fault diagnosis expert system of semiconductor manufacturing equipment using a Bayesian network. *International Journal of Computer Integrated Manufacturing*, 26(12):1161–1171, December 2013.

[24] Fei-Long Chen and Shu-Fan Liu. A neural-network approach to recognize defect spatial pattern in semiconductor fabrication. *IEEE Transactions on Semiconductor Manufacturing*, 13(3):366–373, August 2000.

[25] Z. K. Peng and F. L. Chu. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. *Mechanical Systems and Signal Processing*, 18(2):199–221, March 2004.

[26] S. J. Engel, B. J. Gilmartin, K. Bongort, and A. Hess. Prognostics, the real issues involved with predicting life remaining. In *2000 IEEE Aerospace Conference. Proceedings (Cat. No.00TH8484)*, volume 6, pages 457–469 vol.6, 2000.

[27] A. Hess, G. Calvello, and P. Frith. Challenges, issues, and lessons learned chasing the "Big P". Real predictive prognostics. Part 1. In *2005 IEEE Aerospace Conference*, pages 3610–3619, March 2005.

[28] Jianhui Luo, M. Namburu, K. Pattipati, Liu Qiao, M. Kawamoto, and S. Chigusa. Model-based prognostic techniques [maintenance applications]. In *Proceedings AUTOTESTCON 2003. IEEE Systems Readiness Technology Conference.*, pages 330–340, September 2003.

[29] T. Brotherton, P. Grabill, D. Wroblewski, R. Friend, B. Sotomayer, and J. Berry. A testbed for data fusion for engine diagnostics and prognostics. In *Proceedings, IEEE Aerospace Conference*, volume 6, pages 6–3029–6–3042 vol.6, 2002.

[30] P. Baruah and R. B. Chinnam *. HMMs for diagnostics and prognostics in machining processes. *International Journal of Production Research*, 43(6):1275–1293, March 2005.

[31] S. A. Lewis and T. G. Edwards. Smart sensors and system health management tools for avionics and mechanical systems. In *16th DASC. AIAA/IEEE Digital Avionics Systems Conference. Reflections to the Future. Proceedings*, volume 2, pages 8.5–1–8.5–7 vol.2, October 1997.

[32] O. E. Dragomir, R. Gouriveau, F. Dragomir, E. Minca, and N. Zerhouni. Review of prognostic problem in condition-based maintenance. In *2009 European Control Conference (ECC)*, pages 1587–1592, August 2009.

[33] M. J. Roemer and G. J. Kacprzynski. Advanced diagnostics and prognostics for gas turbine engine risk assessment. In *2000 IEEE Aerospace Conference. Proceedings (Cat. No.00TH8484)*, volume 6, pages 345–353 vol.6, 2000.

[34] C. Sankavaram, B. Pattipati, A. Kodali, K. Pattipati, M. Azam, S. Kumar, and M. Pecht. Model-based and data-driven prognosis of automotive and electronic systems. In *2009 IEEE International Conference on Automation Science and Engineering*, pages 96–101, August 2009.

[35] Ying Peng, Ming Dong, and Ming Jian Zuo. Current status of machine prognostics in condition-based maintenance: a review. *The International Journal of Advanced Manufacturing Technology*, 50(1-4):297–313, September 2010.

[36] Shane Butler. *Prognostic Algorithms for Condition Monitoring and Remaining Useful Life Estimation.* phd, National University of Ireland Maynooth, September 2012.

[37] Hamideh Rostami, Jakey Blue, and Claude Yugma. Equipment Deterioration Prognosis and Fault Diagnosis in Semiconductor Manufacturing. Dublin, Ireland, April 2017.

[38] Gary R. Halligan and S. Jagannathan. PCA-based fault isolation and prognosis with application to pump. *The International Journal of Advanced Manufacturing Technology*, 55(5-8):699–707, July 2011.

[39] Silvio Simani, Cesare Fantuzzi, and Ron J. Patton. *Model-based Fault Diagnosis in Dynamic Systems Using Identification Techniques.* Advances in Industrial Control. Springer-Verlag, London, 2003.

[40] Michle Basseville. Detecting changes in signals and systemsA survey. *Automatica*, 24(3):309–326, May 1988.

[41] R. Isermann. Supervision, fault-detection and fault-diagnosis methods An introduction. *Control Engineering Practice*, 5(5):639–652, May 1997.

[42] Murali Krishnamurthi and Don T. Phillips. An expert system framework for machine fault diagnosis. *Computers & Industrial Engineering*, 22(1):67–84, January 1992.

[43] Venkat Venkatasubramanian, Raghunathan Rengaswamy, and Surya N Kavuri. A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies. *Computers & Chemical Engineering*, 27(3):313–326, March 2003.

[44] Hoon Sohn, Keith Worden, and Charles R. Farrar. Statistical Damage Classification Under Changing Environmental and Operational Conditions. *Journal of Intelligent Material Systems and Structures*, 13(9):561–574, September 2002.

[45] Mattias Nyberg. A General Framework for Fault Diagnosis Based on Statistical Hypothesis Testing. In *Proceedings of Twelfth International Workshop on Principles of Diagnosis Dx01, March 7-9, 2001, Via Lattea, Italian Alps*, pages 135–142, 2001.

[46] K. L. Tew, S. D. Teddy, P. Y. Watt, and Xiaoli Li. A systematic approach towards manufacturing health management. In *2011 Prognostics and System Health Managment Confernece*, pages 1–6, May 2011.

[47] Jakey Blue, Agns Roussy, and Jacques Pinaton. Run-to-run Sensor Variation Monitoring for Process Fault Diagnosis in Semiconductor Manufacturing. In *Proceedings of the 2016 Winter Simulation Conference*, WSC '16, pages 2523–2534, Piscataway, NJ, USA, 2016. IEEE Press.

[48] Gang Li, Si-Zhao Qin, Yin-Dong Ji, and Dong-Hua Zhou. Total PLS Based Contribution Plots for Fault Diagnosis. *Acta Automatica Sinica*, 35(6):759–765, June 2009.

[49] Rui Zhao, Ruqiang Yan, Zhenghua Chen, Kezhi Mao, Peng Wang, and Robert X. Gao. Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, 115:213–237, January 2019.

[50] H. Lee, Y. Kim, and C. O. Kim. A Deep Learning Model for Robust Wafer Fault Monitoring With Sensor Measurement Noise. *IEEE Transactions on Semiconductor Manufacturing*, 30(1):23–31, February 2017.

[51] James Moyne and Jimmy Iskandar. Big Data Analytics for Smart Manufacturing: Case Studies in Semiconductor Manufacturing. *Processes*, 5(3):39, July 2017.

[52] Anne-Sylvie Charles, Ioana-Ruxandra Floru, Catherine Azzaro-Pantel, Luc Pibouleau, and Serge Domenech. Optimization of preventive maintenance strategies in a multipurpose batch plant: application to semiconductor manufacturing. *Computers & Chemical Engineering*, 27(4):449–467, April 2003.

[53] Gary S. May and Costas J. Spanos. Introduction to Semiconductor Manufacturing. *Fundamentals of Semiconductor Manufacturing and Process Control*, May 2006.

[54] Claude Yugma, Jakey Blue, Stphane Dauzre-Prs, and Ali Obeid. Integration of scheduling and advanced process control in semiconductor manufacturing: review and outlook. *Journal of Scheduling*, 18(2):195–205, April 2015.

[55] Wahyu Caesarendra and Tegoeh Tjahjowidodo. A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. *Faculty of Engineering and Information Sciences - Papers: Part B*, January 2017.

[56] Rajesh Vijayaraghavan. *Fault detection and classification in etch tools*. PhD thesis, Texas Tech University, 2003.

[57] Rangaraj M. Rangayyan. *Biomedical Signal Analysis: A Case-Study Approach*. Wiley-IEEE Press, New York, NY, 1 edition edition, December 2001.

[58] A. G. Chao, S. T. Tseng, D. S. H. Wong, S. S. Jang, and S. P. Lee. Systematic applications of multivariate analysis to monitoring of equipment health in semiconductor manufacturing. In *2008 Winter Simulation Conference*, pages 2330–2334, December 2008.

[59] Shui-Pin Lee, An-Kuo Chao, Fugee Tsung, David Shan Hill Wong, Sheng-Tsiang Tseng, and Shi-Shang Jang. Monitoring Batch Processes with Multiple OnOff Steps in Semiconductor Manufacturing. *Journal of Quality Technology*, 43(2):142–157, April 2011.

[60] D. C. Krueger, D. C. Montgomery, and C. M. Mastrangelo. Application of Generalized Linear Models to Predict Semiconductor Yield Using Defect Metrology Data. *IEEE Transactions on Semiconductor Manufacturing*, 24(1):44–58, February 2011.

[61] A. Thieullen, M. Ouladsine, and J. Pinaton. A Survey of Health Indicators and Data-Driven Prognosis in Semiconductor Manufacturing Process. *IFAC Proceedings Volumes*, 45(20):19–24, January 2012.

[62] Lei Yang and Jay Lee. Bayesian Belief Network-based approach for diagnostics and prognostics of semiconductor manufacturing systems. *Robotics and Computer-Integrated Manufacturing*, 28(1):66–74, February 2012.

[63] Alexander Bleakie and Dragan Djurdjanovic. Feature extraction, condition monitoring, and fault modeling in semiconductor manufacturing systems. *Computers in Industry*, 64(3):203–213, April 2013.

[64] Thi-Bich-Lien Nguyen, Mohand Djeziri, Bouchra Ananou, Mustapha Ouladsine, and Jacques Pinaton. Fault Prognosis with Stochastic Modelling on Critical Points of Discrete Processes | PHM Society. *Second European Conference of the Prognostics and Health Management Society*, 5:1–8, 2014.

[65] Hui-Chun Yu, Kuo-Yi Lin, and Chen-Fu Chien. Hierarchical indices to detect equipment condition changes with high dimensional data for semiconductor manufacturing. *Journal of Intelligent Manufacturing*, 25(5):933–943, October 2014.

[66] T. B. L. Nguyen, M. A. Djeziri, B. Ananou, M. Ouladsine, and J. Pinaton. Health Index Extraction Methods for Batch Processes in Semiconductor Manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, 28(3):306–317, August 2015.

[67] G. Wang, R. M. Hasani, Y. Zhu, and R. Grosu. A novel Bayesian network-based fault prognostic method for semiconductor manufacturing process. In *2017 IEEE International Conference on Industrial Technology (ICIT)*, pages 1450–1454, March 2017.

[68] Hamideh Rostami, Jakey Blue, and Claude Yugma. Equipment Health Diagnosis and Prognosis using a Wavelet-based Windowing Approach in Semiconductor Manufacturing. Austin, Texas, US, October 2017.

[69] Xiaodong Jia, Chao Jin, Matt Buzza, Yuan Di, David Siegel, and Jay Lee. A deviation based assessment methodology for multiple machine health patterns classification and fault detection. *Mechanical Systems and Signal Processing*, 99:244–261, January 2018.

[70] Hamideh Rostami, Jakey Blue, and Claude Yugma. FDC-based Equipment Deterioration Modeling and the Root Cause Identification in Semiconductor Industry. *European advanced process control and manufacturing (apc|m) Conference,*, April 2018.

[71] Yu-Chuan Su, Fan-Tien Cheng, Min-Hsiung Hung, and Hsien-Cheng Huang. Intelligent prognostics system design and implementation. *IEEE Transactions on Semiconductor Manufacturing*, 19(2):195–207, May 2006.

[72] Mohammed Farouk Bouaziz, Eric Zama, and Frdric Duvivier. Towards Bayesian network methodology for predicting the equipment health factor of complex semiconductor systems. *International Journal of Production Research*, 51(15):4597–4617, August 2013.

[73] H. H. Yue, S. J. Qin, R. J. Markle, C. Nauert, and M. Gatto. Fault detection of plasma etchers using optical emission spectra. *IEEE Transactions on Semiconductor Manufacturing*, 13(3):374–385, August 2000.

[74] Chao-Ton Su, Taho Yang, and Chir-Mour Ke. A neural-network approach for semiconductor wafer post-sawing inspection. *IEEE Transactions on Semiconductor Manufacturing*, 15(2):260–266, May 2002.

[75] G. Spitzlsperger, C. Schmidt, G. Ernst, H. Strasser, and M. Speil. Fault detection for a via etch process using adaptive multivariate methods. *IEEE Transactions on Semiconductor Manufacturing*, 18(4):528–533, November 2005.

[76] Iker Gondra. Applying machine learning to software fault-proneness prediction. *Journal of Systems and Software*, 81(2):186–195, February 2008.

[77] A. Chen and J. Blue. Recipe-Independent Indicator for Tool Health Diagnosis and Predictive Maintenance. *IEEE Transactions on Semiconductor Manufacturing*, 22(4):522–535, November 2009.

[78] J. Yu. Fault Detection Using Principal Components-Based Gaussian Mixture Model for Semiconductor Manufacturing Processes. *IEEE Transactions on Semiconductor Manufacturing*, 24(3):432–444, August 2011.

[79] Hung Hung and Argon Chen. Test of covariance changes without a large sample and its application to fault detection and classification. *Journal of Process Control*, 22(6):1113–1121, July 2012.

[80] Chen-Fu Chien, Chia-Yu Hsu, and Pei-Nong Chen. Semiconductor fault detection and classification for yield enhancement and manufacturing intelligence. *Flexible Services and Manufacturing Journal*, 25(3):367–388, September 2013.

[81] J. Blue, D. Gleispach, A. Roussy, and P. Scheibelhofer. Tool Condition Diagnosis With a Recipe-Independent Hierarchical Monitoring Scheme. *IEEE Transactions on Semiconductor Manufacturing*, 26(1):82–91, February 2013.

[82] J. Blue, A. Roussy, and J. Pinaton. FDC R2r variation monitoring for sensor level diagnosis in tool condition hierarchy. In *25th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC 2014)*, pages 92–97, May 2014.

[83] Bo He, Tao Chen, and Xianhui Yang. Root cause analysis in multivariate statistical process monitoring: Integrating reconstruction-based multivariate contribution analysis with fuzzy-signed directed graphs. *Computers & Chemical Engineering*, 64:167–177, May 2014.

[84] T. J. Rato, J. Blue, J. Pinaton, and M. S. Reis. Translation-Invariant Multiscale Energy-Based PCA for Monitoring Batch Processes in Semiconductor Manufacturing. *IEEE Transactions on Automation Science and Engineering*, 14(2):894–904, April 2017.

[85] D. Anand, J. Moyne, and D. M. Tilbury. A Method for Reducing Noise and Complexity in Yield Analysis for Manufacturing Process Workflows. *IEEE Transactions on Semiconductor Manufacturing*, 27(4):501–514, November 2014.

[86] Yuan Li and Xinmin Zhang. Diffusion maps based k-nearest-neighbor rule technique for semiconductor manufacturing process fault detection. *Chemometrics and Intelligent Laboratory Systems*, 136:47–57, August 2014.

[87] Hamideh Rostami, Jakey Blue, and Claude Yugma. Equipment anomaly detection and automatic fault fingerprint extraction in semiconductor manufacturing. In *Procedding of International Symposium on Semiconductor Manufacturing Intelligence*, Hsinchu, Taiwan, August 2016.

[88] C. J. Spanos, H. F. Guo, A. Miller, and J. Levine-Parrill. Real-time statistical process control using tool data [semiconductor manufacturing]. *IEEE Transactions on Semiconductor Manufacturing*, 5(4):308–318, November 1992.

[89] S. Joe Qin, Hongyu Yue, and Ricardo Dunia. Self-Validating Inferential Sensors with Application to Air Emission Monitoring. *Industrial & Engineering Chemistry Research*, 36(5):1675–1685, May 1997.

[90] B. Kim and G. S. May. Real-time diagnosis of semiconductor manufacturing equipment using a hybrid neural network expert system. *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part C*, 20(1):39–47, January 1997.

[91] Jill P. Card, Mark Naimo, and William Ziminsky. Run-to-run process control of a plasma etch process with neural network modelling. *Quality and Reliability Engineering International*, 14(4):247–260.

[92] Johan A. Westerhuis, Stephen P. Gurden, and Age K. Smilde. Generalized contribution plots in multivariate statistical process monitoring. *Chemometrics and Intelligent Laboratory Systems*, 51(1):95–114, May 2000.

[93] A. Chen, R. S. Guo, and G. S. Wu. Real-time equipment health evaluation and dynamic preventive maintenance. In *Proceedings of ISSM2000. Ninth International Symposium on Semiconductor Manufacturing (IEEE Cat. No.00CH37130)*, pages 375–378, 2000.

[94] Weihua Li, H. Henry Yue, Sergio Valle-Cervantes, and S. Joe Qin. Recursive PCA for adaptive process monitoring. *Journal of Process Control*, 10(5):471–486, October 2000.

[95] H. Henry Yue and S. Joe Qin. Reconstruction-Based Fault Identification Using a Combined Index. *Industrial & Engineering Chemistry Research*, 40(20):4403–4414, October 2001.

[96] Chih-Min Fan, Ruey-Shan Guo, A. Chen, Kuo-Ching Hsu, and Chih-Shih Wei. Data mining and fault diagnosis based on wafer acceptance test data and in-line manufacturing data. In *2001 IEEE International Symposium on Semiconductor Manufacturing. ISSM 2001. Conference Proceedings (Cat. No.01CH37203)*, pages 171–174, 2001.

[97] Junghui Chen and Kun-Chih Liu. On-line batch process monitoring using dynamic PCA and dynamic PLS models. *Chemical Engineering Science*, 57(1):63–75, January 2002.

[98] Q. P. He and J. Wang. Large-Scale Semiconductor Process Fault Detection Using a Fast Pattern Recognition-Based Method. *IEEE Transactions on Semiconductor Manufacturing*, 23(2):194–200, May 2010.

[99] Ningyun Lu, Fuli Wang, and Furong Gao. Combination Method of Principal Component and Wavelet Analysis for Multivariate Process Monitoring and Fault Diagnosis. *Industrial & Engineering Chemistry Research*, 42(18):4198–4207, September 2003.

[100] H. H. Yue and M. Tomoyasu. Weighted principal component analysis and its applications to improve FDC performance. In *2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601)*, volume 4, pages 4262–4267 Vol.4, December 2004.

[101] Nomikos Paul and MacGregor John F. Monitoring batch processes using multiway principal component analysis. *AIChE Journal*, 40(8):1361–1375, June 2004.

[102] Bakshi Bhavik R. Multiscale PCA with application to multivariate statistical process monitoring. *AIChE Journal*, 44(7):1596–1610, April 2004.

[103] Dunia Ricardo and Joe Qin S. Subspace approach to multidimensional fault identification and reconstruction. *AIChE Journal*, 44(8):1813–1831, April 2004.

[104] S. A. Velichko. Multi-parameter model based advanced process control. In *2004 IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop (IEEE Cat. No.04CH37530)*, pages 443–447, May 2004.

[105] Sang Jeen Hong, G. S. May, Sungwon Park, and Dong-Chul Park. A modular neural network for R2r diagnosis of semiconductor fabrication equipment: a reactive ion etching application. In *2004 IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop (IEEE Cat. No.04CH37530)*, pages 43–47, May 2004.

[106] Sang Jeen Hong and G. S. May. Neural network-based real-time malfunction diagnosis of reactive ion etching using in situ metrology data. *IEEE Transactions on Semiconductor Manufacturing*, 17(3):408–421, August 2004.

[107] Kyungpil Kim, Jong-Min Lee, and In-Beum Lee. A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction. *Chemometrics and Intelligent Laboratory Systems*, 79(1):22–30, October 2005.

[108] Ji-Hoon Cho, Jong-Min Lee, Sang Wook Choi, Dongkwon Lee, and In-Beum Lee. Fault identification for process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 60(1):279–288, January 2005.

[109] K. Mai and M. Tuckermann. SPC based in-line reticle monitoring on product wafers. In *IEEE/SEMI Conference and Workshop on Advanced Semiconductor Manufacturing 2005.*, pages 184–188, April 2005.

[110] G. A. Cherry and S. J. Qin. Multiblock principal component analysis based on a combined index for semiconductor fault detection and diagnosis. *IEEE Transactions on Semiconductor Manufacturing*, 19(2):159–172, May 2006.

[111] Chin Sun, S. G. Bisland, K. Nguyen, and Long Vu. Prognostic/Diagnostic Health Management System (PHM) for Fab Efficiency. In *The 17th Annual SEMI/IEEE ASMC 2006 Conference*, pages 433–438, May 2006.

[112] Yale Zhang and Michael S. Dudzic. Online monitoring of steel casting processes using multivariate statistical technologies: From continuous to transitional operations. *Journal of Process Control*, 16(8):819–829, September 2006.

[113] Yale Zhang and Michael S. Dudzic. Industrial application of multivariate SPC to continuous caster start-up operations for breakout prevention. *Control Engineering Practice*, 14(11):1357–1375, November 2006.

[114] Q. P. He and J. Wang. Fault Detection Using the k-Nearest Neighbor Rule for Semiconductor Manufacturing Processes. *IEEE Transactions on Semiconductor Manufacturing*, 20(4):345–354, November 2007.

[115] Zhiqiang Ge and Zhihuan Song. Process Monitoring Based on Independent Component AnalysisPrincipal Component Analysis (ICAPCA) and Similarity Factors. *Industrial & Engineering Chemistry Research*, 46(7):2054–2063, March 2007.

[116] Q. P. He and J. Wang. Principal component based k-nearest-neighbor rule for semiconductor process fault detection. In *2008 American Control Conference*, pages 1606–1611, June 2008.

[117] Ashraf AlGhazzawi and Barry Lennox. Monitoring a complex refining process using multivariate statistics. *Control Engineering Practice*, 16(3):294–307, March 2008.

[118] Stella Bezergianni and Aggeliki Kalogianni. Application of Principal Component Analysis for Monitoring and Disturbance Detection of a Hydrotreating Process. *Industrial & Engineering Chemistry Research*, 47(18):6972–6982, September 2008.

[119] Sang Wook Choi, Julian Morris, and In-Beum Lee. Dynamic model-based batch process monitoring. *Chemical Engineering Science*, 63(3):622–636, February 2008.

[120] Carlos Alcala and S. Joe Qin. Reconstruction-based Contribution for Process Monitoring. *IFAC Proceedings Volumes*, 41(2):7889–7894, January 2008.

[121] Berber R., Atasoy I., Yuceer M., and Deniz G. Online Statistical Process Monitoring and Fault Diagnosis in Batch Baker's Yeast Fermentation. *Chemical Engineering & Technology*, 32(4):650–658, February 2009.

[122] Zhou Donghua, Li Gang, and Qin S. Joe. Total projection to latent structures for process monitoring. *AIChE Journal*, 56(1):168–178, August 2009.

[123] Vinay Kariwala, Pabara-Ebiere Odiowei, Yi Cao, and Tao Chen. A branch and bound method for isolation of faulty variables through missing variable analysis. *Journal of Process Control*, 20(10):1198–1206, December 2010.

[124] Jialin Liu. Fault diagnosis using contribution plots without smearing effect on non-faulty variables. *Journal of Process Control*, 22(9):1609–1623, October 2012.

[125] Pieter Van den Kerkhof, Jef Vanlaer, Geert Gins, and Jan F. M. Van Impe. Analysis of smearing-out in contribution plot based fault isolation for Statistical Process Control. *Chemical Engineering Science*, 104:285–293, December 2013.

[126] Jialin Liu, David Shan Hill Wong, and Ding-Sou Chen. Bayesian filtering of the smearing effect: Fault isolation in chemical process monitoring. *Journal of Process Control*, 24(3):1–21, March 2014.

[127] Te-Hui Kuang, Zhengbing Yan, and Yuan Yao. Multivariate fault isolation via variable selection in discriminant analysis. *Journal of Process Control*, 35:30–40, November 2015.

[128] Hongquan Ji, Xiao He, and Donghua Zhou. On the use of reconstruction-based contribution for fault diagnosis. *Journal of Process Control*, 40:24–34, April 2016.

[129] H. Rostami, J. Blue, and C. Yugma. Equipment Condition Diagnosis and Fault Fingerprint Extraction in Semiconductor Manufacturing. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 534–539, December 2016.

[130] J. Moyne, J. Samantaray, and M. Armacost. Big Data Capabilities Applied to Semiconductor Manufacturing Advanced Process Control. *IEEE Transactions on Semiconductor Manufacturing*, 29(4):283–291, November 2016.

[131] Hamidey Rostami, Jean-Yves Dantan, and Lazhar Homri. Review of data mining applications for quality assessment in manufacturing industry: support vector machines. *International Journal of Metrology and Quality Engineering*, 6(4):401, 2015.

[132] Adriane B. S. Serapio, Guilherme S. Corra, Felipe B. Gonalves, and Veronica O. Carvalho. Combining K-Means and K-Harmonic with Fish School Search Algorithm for data clustering task on graphics processing units. *Applied Soft Computing*, 41:290–304, April 2016.

[133] Q. Peter He and Jin Wang. Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes. *AIChE Journal*, 57(1):107–121.

[134] R. Pearson. *Mining Imperfect Data*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics, January 2005.

[135] Pao-Hua Chou, Menq-Jiun Wu, and Kuang-Ku Chen. Integrating support vector machine and genetic algorithm to implement dynamic wafer quality prediction system. *Expert Systems with Applications*, 37(6):4413–4424, June 2010.

[136] Shih-Wei Lin, Kuo-Ching Ying, Shih-Chieh Chen, and Zne-Jung Lee. Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Systems with Applications*, 35(4):1817–1824, November 2008.

[137] Dalian Yang, Yilun Liu, Songbai Li, Xuejun Li, and Liyong Ma. Gear fault diagnosis based on support vector machine optimized by artificial bee colony algorithm. *Mechanism and Machine Theory*, 90:219–229, August 2015.

[138] Engin Avci. Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithmsupport vector machines: HGASVM. *Expert Systems with Applications*, 36(2, Part 1):1391–1402, March 2009.

[139] David E. Goldberg and John H. Holland. Genetic Algorithms and Machine Learning. *Machine Learning*, 3(2-3):95–99, October 1988.

[140] Mehrdad Mohammadi, Ali Siadat, Jean-Yves Dantan, and Reza Tavakkoli-Moghaddam. Mathematical modelling of a robust inspection process plan: Taguchi and Monte Carlo methods. *International Journal of Production Research*, 53(7):2202–2224, April 2015.

[141] Chih-Hung Wu, Gwo-Hshiung Tzeng, Yeong-Jia Goo, and Wen-Chang Fang. A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. *Expert Systems with Applications*, 32(2):397–408, February 2007.

[142] J. F. MacGregor and T. Kourti. Statistical process control of multivariate processes. *Control Engineering Practice*, 3(3):403–414, March 1995.

[143] J. Edward Jackson and Govind S. Mudholkar. Control Procedures for Residuals Associated With Principal Component Analysis. *Technometrics*, 21(3):341–349, August 1979.

[144] Paul Nomikos. Detection and diagnosis of abnormal batch operations based on multi-way principal component analysis World Batch Forum, Toronto, May 1996. *ISA Transactions*, 35(3):259–266, January 1996.

[145] Tran Van Tung and Bo-Suk Yang. Machine Fault Diagnosis and Prognosis: The State of The Art. *International Journal of Fluid Machinery and Systems*, 2(1):61–71, 2009.

[146] Jinglong Chen, Zipeng Li, Jun Pan, Gaige Chen, Yanyang Zi, Jing Yuan, Binqiang Chen, and Zhengjia He. Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 70-71:1–35, March 2016.

[147] Macarena Boix and Begoa Cant. Wavelet Transform application to the compression of images. *Mathematical and Computer Modelling*, 52(7):1265–1270, October 2010.

[148] Oleg V Vasilyev and Christopher Bowman. Second-Generation Wavelet Collocation Method for the Solution of Partial Differential Equations. *Journal of Computational Physics*, 165(2):660–693, December 2000.

[149] S Arivazhagan and L Ganesan. Texture classification using wavelet transform. *Pattern Recognition Letters*, 24(9):1513–1521, June 2003.

[150] Donghui Zhang. Wavelet Approach for ECG Baseline Wander Correction and Noise Reduction. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 1212–1215, January 2005.

[151] Duane S. Boning and James E. Chung. Statistical metrology: Understanding spatial variation in semiconductor manufacturing. In *Microelectronic Manufacturing Yield, Reliability, and Failure Analysis II*, volume 2874, pages 16–27. International Society for Optics and Photonics, 1996.

[152] M. A. Kramer. Autoassociative neural networks. *Computers & Chemical Engineering*, 16(4):313–328, April 1992.

[153] Yan Wang, F. Benito, G. A. Vera, and M. Jamshidi. Control design for diagnostic and prognostic of hardware systems. In *2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No.04CH37542)*, volume 1, pages 457–462 vol.1, July 2004.

[154] E. Y. Hamid and Z. I. Kawasaki. Wavelet-based data compression of power system disturbances using the minimum description length criterion. *IEEE Transactions on Power Delivery*, 17(2):460–466, April 2002.

[155] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41(7):909–996, October 1988.

[156] James Moyne, Enrique del Castillo, and Arnon M. Hurwitz, editors. *Run-to-Run Control in Semiconductor Manufacturing*. CRC Press, Boca Raton, 1 edition edition, November 2000.

[157] Data Collection in Semiconductor Manufacturing Equipment and Facilities | India | Omron IA.

[158] Douglas C. Montgomery. *Design and Analysis of Experiments*. John Wiley & Sons, 2017. Google-Books-ID: Py7bDgAAQBAJ.

[159] Teuvo Kohonen and Panu Somervuo. Self-organizing maps of symbol strings. *Neurocomputing*, 21(1):19–30, November 1998.

[160] Gobinda G. Chowdhury. *Introduction to Modern Information Retrieval*. Facet Publishing, 2010. Google-Books-ID: cN4qDgAAQBAJ.

[161] D. L. Davies and D. W. Bouldin. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227, April 1979.

[162] J. C. Dunn. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 3(3):32–57, January 1973.

[163] Hubert W. Lilliefors. On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown. *Journal of the American Statistical Association*, 64(325):387–389, March 1969.

[164] Kuo-Liang Lu and Yung-Chih Yao. Method for preventing contamination in a plasma process chamber, April 2001.

[165] Tzu-Ken Lin, Wei-Kai Wang, Shih-Yung Huang, Chi-Tsung Tasi, and Dong-Sing Wuu. Comparison of Erosion Behavior and Particle Contamination in Mass-Production CF4/O2 Plasma Chambers Using Y2o3 and YF3 Protective Coatings. *Nanomaterials*, 7(7), July 2017.

[166] Katsumi Ukai. Dry Etching (Part 1): Particulate Contamination Due to Dry Etching. In *Ultraclean Surface Processing of Silicon Wafers*, pages 361–370. Springer, Berlin, Heidelberg, 1998.

[167] John Wishart. The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population. *Biometrika*, 20A(1/2):32–52, 1928.

[168] Federico Girosi, Michael Jones, and Tomaso Poggio. Regularization Theory and Neural Networks Architectures. *Neural Computation*, 7(2):219–269, March 1995.

[169] Yuan Yao, Lorenzo Rosasco, and Andrea Caponnetto. On Early Stopping in Gradient Descent Learning. *Constructive Approximation*, 26(2):289–315, August 2007.

[170] John Loughrey and Pdraig Cunningham. Overfitting in Wrapper-Based Feature Subset Selection: The Harder You Try the Worse it Gets. In Max Bramer, Frans Coenen, and Tony Allen, editors, *Research and Development in Intelligent Systems XXI*, pages 33–43. Springer London, 2005.

[171] Kan Deng and Andrew W. Moore. *On Greediness of Feature Selection Algorithms*. Carnegie Mellon University, The Robotics Institute, 1998.

Ecole Nationale Supérieure des Mines
de Saint-Etienne

NNT : 2018LYSEM028

Hamideh Rostami

Equipment Behavior Modelling for Fault Diagnosis and Deterioration Prognosis in Semiconductor Manufacturing

Speciality: Industrial Engineering

Keywords: Equipment behavior prognosis, Equipment failure diagnosis, Deterioration modeling and monitoring, Fault detection and classification, Semiconductor industry

Abstract: Moving toward advanced technologies requires the modern industries, in particular, the semiconductor, to keep their equipment at a high utilization level and low environmental risk. Production deficiencies such as process variations and unexpected equipment breakdowns have made it difficult (if not impossible) to stay at high-grade product yield and significant equipment utilization. In this thesis, the aim is to propose efficient equipment behavior prognosis, and equipment failure diagnosis approaches in Batch Manufacturing Processes that are pervasive modes in today's semiconductor fab.

With the advancement of sensor information technology, efficient data-driven approaches are proposed for both prognostic and diagnostic purposes. In the fault diagnosis, this research firstly applies the Support Vector Machine (SVM) classifier to detect the abnormal observations. The normal process dynamics are then decomposed into different clusters by K-means clustering. Principal Component Analysis (PCA) is used to model each part of the process dynamics. Fault fingerprints can be extracted finally by consolidating the out of control scenarios after projecting the abnormal observations into the PCA models.

In prognostics, an equipment deterioration modeling and monitoring approach for batch processes is developed with two aims: exploiting the temporal FDC (Fault Detection and Classification) data to characterize the equipment behavior and modeling the deterioration trend with the potential causes. By using the Discrete Wavelet Transformation (DWT), the temporal data are decomposed into approximation and detail components to detect two types of deterioration caused by macro- and micro-level variations. Several scenarios of case studies are conducted based on the practical dataset provided by a local IC maker. The results show that the proposed approaches can effectively prognose the equipment behavior and diagnose the equipment failure with the correct causes.

Ecole Nationale Supérieure des Mines
de Saint-Etienne

NNT : 2018LYSEM028

Hamideh Rostami

Modélisation du Comportement des Equipements pour le Diagnostic des Pannes et le Pronostic des Détériorations des Machines en Fabrication de Semi-conducteurs

Spécialité: Génie Industriel

Mots clefs: Pronostic du comportement des équipements; Diagnostic de pannes équipement; Modélisation et surveillance de la détérioration; Détection et classification des fautes; Industrie de semi-conducteurs

Résumé: Les défauts de production dus aux variations dans le processus de fabrication et aux pannes inattendues des équipements rendent difficile la conservation d'un rendement élevé de production dans l'industrie de fabrication de semi-conducteurs. L'objectif de ce travail de thése est de proposer un pronostic efficace du comportement des équipements ainsi qu'un diagnostic des pannes dans le processus de fabrication de semi-conducteurs. Avec des capteurs plus performants, des approches efficaces basées sur des données sont proposées pour le pronostic et le diagnostic. Pour le diagnostic, cette recherche applique d'abord la méthode de Machines à Vecteurs de Support (Support Vector Machine) pour détecter les anomalies constatées dans les observations. La dynamique du processus normale est ensuite décomposée en différentes groupes par la méthode de partitionnement à K-moyennes. L'Analyse en Composantes Principales (ACP) est utilisée pour modéliser chaque partie de la dynamique du processus. Les empreintes de défaut peuvent enfin être extraites en consolidant les scénarios hors contrôle après avoir projeté les anomalies dans les modèles ACP. En pronostic, une approche de modélisation et de surveillance de la dégradation des équipements pour le processus de fabrication par lots est développée avec deux objectifs : exploiter les données temporelles de Détection et Classification des Fautes (FDC) pour caractériser le comportement des équipements et modéliser la tendance de détérioration avec les causes potentielles. La transformation en ondelettes discrète (DWT) décompose les données temporelles en composantes d'approximation détaillées afin de détecter deux types de détérioration provoquée par des variations au niveau macro et micro. Les résultats montrent que les approches proposées permettent de prédire efficacement le comportement des équipements et de diagnostiquer la défaillance avec les causes premières.