UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Enginyeria Electrònica

"Contributions to industrial process condition forecasting applied to copper rod manufacturing process"

Thesis submitted in partial fulfillment of the requirement for the PhD Degree issued by the Universitat Politècnica de Catalunya, in its Electronic Engineering Program.

Daniel Zurita Millán

Supervisors:

Dr. Juan Antonio Ortega Redondo

Dr. Miguel Delgado Prieto

September, 2017

El futuro tiene muchos nombres.
Para los débiles, lo inalcanzable.
Para los temerosos, lo desconocido.
Para los valientes, la oportunidad.

Víctor Hugo

Abstract

Ensuring reliability and robustness of operation is one of the main concerns in industrial manufacturing processes, due to the ever-increasing demand for improvements over the cost and quality of the processes outcome. In this regard, a deviation from the nominal operating behaviours implies a divergence from the optimal condition specification, and a misalignment from the nominal product quality, causing a critical loss of potential earnings. Indeed, since a decade ago, the industrial sector has been carried out a significant effort towards process condition monitoring approaches, that is to extract knowledge regarding the condition of the different machines and process involved in the manufacturing processes.

However, information about the future condition of an industrial process represents a significant tool in order to gain response time against undesired deviations of the process condition. Thus, the combination of the future knowledge of the process status and the consequent assessment of the future condition, is a required objective towards the next generation of industrial monitoring strategies.

In this regard, the proposed thesis consists on the investigation of a novel industrial process condition forecasting methodology able to assess the current and the future condition of an industrial process with high accuracy and generalization capabilities. This is faced by the modelling and the forecasting of the target signals representative of the global behaviour of the industrial process. For then, fuse the future behaviours of the target signals to provide information regarding the future condition of the whole industrial process. The suitability and the performances of the proposed methodologies have been validated by means of experimental data of a real industrial process, a copper rod manufacturing process.

Keywords

Artificial Intelligence
Condition Monitoring
Feature Extraction
Forecasting
Fuzzy Neural Networks
Industrial Plants
Machine Learning
Predictive models
Prognosis
Time Series Analysis

Acknowledgement

Cuando decidí embarcarme en este proyecto lo hice sabiendo que no lo haría solo. De hecho, lo he conseguido gracias a todas las personas que a lo largo del recorrido de la tesis me han ofrecido de una forma u otra un su apoyo, su consejo y su cariño. En este sentido, soy consciente que unas pocas líneas no son suficientes para describir todo lo que habéis hecho por mí, pero aún y así lo voy a intentar.

Primero, agradecer a mi familia, a mi madre por su cariño, sus mimos, por preocuparse por mí y no dejar que me perdiera a lo largo de estos años por el camino de la vida. A mi padre, por su cariño, sus consejos, su apoyo y por ser para mí un modelo a seguir. A mi hermano, por su cariño, aunque sea a veces a regañadientes, por sus charlas y su apoyo indispensable. A mi abuela, por su cariño y su amor a lo largo de toda una vida. A mi tía, mis tíos y mis primos por estar siempre ahí y regalarme esos momentos inolvidables. Gracias.

Segundo, a mis directores y mis compañeros de MCIA, por hacer que cada día sea una aventura diferente y conseguir que me levante con una sonrisa para ir a trabajar. En especial a Miguel, por ser para mí todo un ejemplo de constancia, trabajo y superación. Sin su amistad y consejo no habría llegado donde estoy ahora, ni estaríais leyendo estas líneas. Gracias de corazón. A Juan Antonio, por su apoyo y supervisión a lo largo de la tesis, por aportar siempre esa visión que me ha hecho mejorar como persona. A Luis, por convencerme de entrar a formar parte de MCIA, por tu apoyo y tu consejo. A Vicente, por intentar llevarme siempre por el buen camino, por preocuparte por mí y hacerme crecer, gracias. A Jesús, por compartir conmigo todos estos años, codo con codo y artículo a artículo. A Carlos y Sandra, por todo el cariño dado, por estar pendientes de mí y darme un empujón siempre que lo necesito, gracias por estar siempre ahí. A María, por su apoyo y confianza, por contagiarme su positividad y hacerme reír cuando tanta falta me ha hecho, gracias. A Enric y Carles, por esas conversaciones dignas de aparecer en los libros de historia. A Joan, por las visitas al despacho cuando aún no sabía ni que era un JCR. A Anderson, porque nunca me reiré tanto testeando un micro en el laboratorio. Y a todos los compañeros que me dejo en el tintero pero que igualmente me habéis aportado mucho, sin vosotros no hubiera sido lo mismo, gracias.

Tercero, pero no menos importante, al resto de mi familia, a Cristina, por su amistad y cariño a lo largo de todos estos años, por estar siempre ahí cuando la necesitaba, por ser como eres, gracias. A Sara, por contagiarme su alegría y su entusiasmo durante todo este tiempo. A Anna, por su amistad y por los buenos consejos que siempre me has dado. A Alex, por sus no tan buenos consejos que más de una vez me han arrancado carcajadas. A José, por estar siempre ahí y ser como un hermano postizo. A Mari Carmen, por su cariño y comprensión los primeros años de la tesis, gracias.

A toda la gente del laboratorio Ampere en Lyon, por acogerme, cuidarme y hacerme sentir como si estuviera en mi casa. Gracias.

En estas líneas intento describir todo lo que me habéis ayudado y me habéis hecho crecer a lo largo de todos estos años. Prometo juntar todos estos recuerdos y llevarlos en la mochila para seguir recorriendo el nuevo camino que se me aparece. Gracias a todos.

Contents of the document

Index of figures

Index of tables

Acronyms and their definitions

ACO	Ant Colony Optimization
AI	Artificial Intelligence
ANFIS	Adaptive Network-based Fuzzy Inference System
ANN	Artificial Neural Networks
ARMA	Autoregressive moving average
BMC	Bayesian Monte Carlo
BMU	Best Matching Unit
CA	Correspondence Analysis
CPS	Cyber-Physical Systems
CRMP	Copper Rod Manufacturing Process
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
EMD	Empirical Mode Decomposition
FEA	Finite Element Analysis
FTPC	Fault-Tolerant Process Control
GA	Genetic Algorithms
HHT	Hilbert-Huang Transform
HMM	Hidden Markov Model
kNN	k- Nearest Neighbours
LDA	Linear Discriminant Analysis
MEA	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MF	Membership Function
MU	Matching Units
NFN	Neo Fuzzy Neuron
NN	Neural Network
PCA	Principal Component Analysis
PDF	Probability Density Function
PHM	Prognostics and Health Management
RMS	Root Mean Square
RMSE	Root Mean Squared Error
RUL	Remaining Useful Life
SOM	Self-Organizing Maps
SMCM	Sequential Monte Carlo Methods
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
TTF	Time to Failure
WT	Wavelet Transform
ZDM	Zero-Defect Manufacturing

1.

Introduction

This chapter outlines the basis on which this thesis research is engaged. It starts from the introduction to the research topic to the objectives and the hypotheses of this thesis research. This chapter includes also the state of the art in regard with the most recent research and trends that can be found in the scientific literature related with the thesis's field.

CONTENTS:

1. Introduction

1.1 Research topic

The reliability of manufacturing process operations and safety of electromechanical assets, are becoming critical aspects in the modern industry [1]. In this regard, the industrial sector has being carried out a significant effort towards the integration of process condition monitoring approaches since a decade ago [2]. However, the condition forecasting of industrial processes, although important to allow preventative actions, is still a challenge in the field. That is, information about the future condition of an industrial process represents a significant feature in order to gain response time against undesired deviations of the process condition. Thus, the combination of the future knowledge of the process's behaviour, and the consequent assessment of the future condition, is a required objective towards the next generation of monitoring strategies to be applied to the modern industry [3].

In order to stablish the framework of this thesis, it is mandatory to define three critical concepts behind the **industrial process condition forecasting** research topic.

First, some definitions can be found in the literature about **industrial process**, as [4]:

> **"The sequence of operations that is necessary to complete the manufacturing of a part or product".**
>
> **"The series of activities performed upon material to convert it from the raw or semi-finished state to a state of further completion and a greater value".**
>
> **"Sequence of interdependent and linked procedures which, at every stage, consume one or more resources (employee time, energy, machines) to convert inputs (data, material, parts, etc.) into outputs. These outputs then serve as inputs for the next stage until a known goal or end result is reached".**

Therefore, an industrial process is considered as a sequential group of processes and machineries that operate with the aim of transforming an input material in a final manufactured asset. Then, it implies that the manufactured product, during the procedure, varies measurable and traceable characteristics.

Second, the **condition of an industrial process** is considered as the ratio between the current condition of the different machines and processes involved in the manufacturing process, and the optimal condition, given when the machines are working properly without failures, and the processes are well adjusted to their nominal configurations. Indeed, the operating condition includes not just classical fault conditions related with electromechanical parts, but also, as more important, functioning deviations from the expected nominal operation, always considering that such deviations affect in some manner the final characteristics of the manufactured product, that is, the quality of the manufactured product.

Finally, third, the **condition forecasting** implies that information regarding the condition of the industrial process is given in future time over a defined time horizon. Thus, the characteristics of the future manufacturing assets can be inferred by the analysis of the future condition of the process, since deviations from the nominal behaviour of the manufacturing process will affect the characteristics of the final product, for example, in terms of quality aspects of the manufactured unit.

Indeed, the industrial process condition forecasting would provide the opportunity to an earlier corrective action over undesired manufacturing process performances, before major affectation into the characteristics of the final product.

The block diagram of the considered industrial process condition monitoring framework is shown in **Fig. 1.1.1**. That is, a manufacturing process is fed by an input material, which is transformed through different processes formed by industrial machinery to provide the final product. Thus, in this research framework, the industrial monitoring system is intended to retrieve data from the manufacturing process in order to assess the current and the future condition, providing with it a way to infer the characteristics of the future manufactured product.

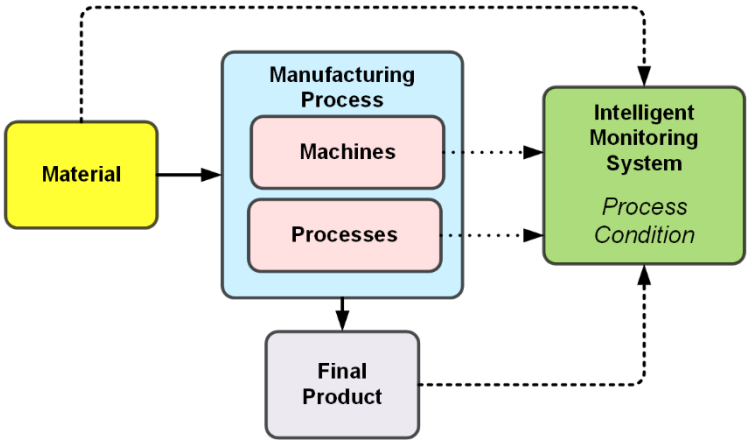

Fig. 1.1.1 Industrial process condition monitoring approach

In fact, the process condition monitoring and its forecasting capabilities represent the basis of the new industrial manufacturing paradigm, the so-called **Zero Defects Manufacturing** (ZDM) [5]. This is an ideal scenario for companies since it involves the minimization of production costs and increase of economic performances. In this regard, the **zero defects manufacturing means that the final product presents characteristics in the expected range, thus, maintaining the desired quality level and, in consequence, keeping the manufacturing process operating in the nominal conditions.**

Independently of the quality criteria used in a specific manufacturing plant, the deviation of the considered quantitative parameters during the final product inspection represents a loss of benefits by reducing the market price or, directly, by product's rejection. Nowadays, high manufacturing ratios requirements motivate the use of complex machinery and processes. This fact, causes a loss in the process understanding by the operators and users, since the intrinsic relations of what is happening during the different manufacturing processes and how are affecting such behaviours to the features of the final product is getting lost. That is, a huge amount of information is usually available in industrial manufacturing plants for process monitoring approaches, but there is not the implementation of corresponding methodologies to obtain proper condition monitoring about *How is operating the manufacturing line in terms of final product's characteristics?* and, *What is the expected quality assessment of the manufactured asset?*

According to this scenario, the condition monitoring and forecasting of the final product's characteristics in manufacturing processes is not an easy issue, since quality is affected by multiple factors such as the company manufacturing strategy, the chosen suppliers and, specially, the condition of the industrial equipment and the

associated manufacturing processes. Thus, in order to face such questions, the product's quality is approached by the process condition monitoring point of view. That is, the condition of the different processes is monitored in order to characterize and identify manufacturing process deviations [6]**:**

> **"ZDM assumes that the quality of a product is totally associated to the proper functioning of the machinery and different processes involved in its fabrication. If all of them are in perfect condition, the final product's quality must be optimal".**

It is stated, then, that **there is a relation between the final product's quality assessment and the condition of the related manufacturing process** [7]. Then, the problem of product quality forecasting turns into industrial process condition forecasting.

The manufacturing plant condition forecasting needs to be supported by **industrial monitoring systems** able to assess the condition of different processes. The general framework of industrial monitoring systems focused on condition forecasting is shown in **Fig. 1.1.2**, where the main blocks and the corresponding objectives are represented [8].

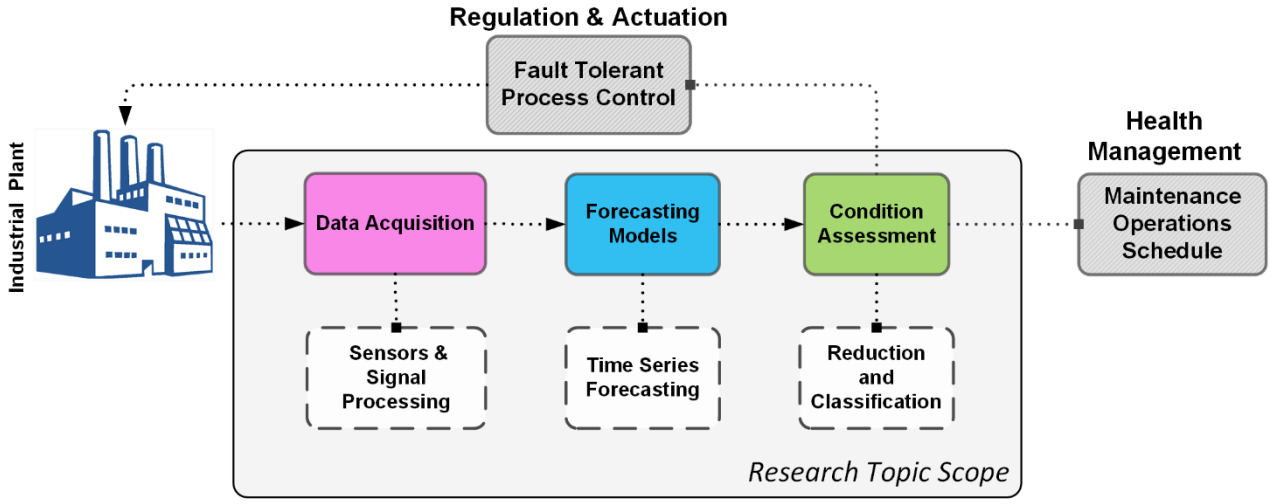

Fig. 1.1.2. General scheme of an Industrial monitoring system for condition forecasting.

- **Data acquisition:** The manufacturing plant must be properly monitored; this means that, at least, basic information must be acquired. Dealing with zero-defect manufacturing, the considered information covers three main sections: (i) information of the material to be processed, (ii) information of the process operation and, (iii) information of the final product characteristics.
- **Forecasting Models:** This block aims to forecast, based on historical data and the current status, the evolution of the expected process signals or calculated features among a defined prediction horizon, *p*.
- **Condition assessment:** Aims to assess the current or future condition of the process regarding the information provided by the forecasting models. It is classically known as diagnosis block, and performs tasks as process behaviour identification and process condition classification.

Although out of the scope of this thesis research topic, the availability of reliable condition forecasting information allows the development of two aspects also considered in the ZDM framework:

- Optimal maintenance scheduling. Future process condition is evaluated in order to plan the optimal maintenance tasks taking into account the current and future conditions of the plant, the resources, the maintenance schedule, the costs and the logistics [9]. This kind of approaches are based on optimization algorithms, fused together with artificial intelligence techniques to automatically generate the optimal maintenance plans for improving the availability and performance of the associated processes.

- Fault tolerant process control. It is focused on production KPIs and self-regulation, and uses the estimated future condition in order to, automatically, correct operating deviations by performing predictive control over the manufacturing processes [10]. This solution implies a self-regulation control of the industrial plant to always work inside the nominal conditions.

1.1.1 Case study: copper rod manufacturing process

In order to validate the effectiveness and performances of the contributions considered in this thesis, a real industrial process has been considered. Thus, the research framework of this thesis, is focused on a specific industrial process line related with a copper rod manufacturing process, however, this industrial process presents the main characteristics that allows the applicability and generalization of the thesis's contributions to multiple industrial scenarios. Moreover, considering real data the developed methodology must overcome the variability and the casuistic of a real process in order to assure the applicability of the proposed approaches.

- The manufactured product quality can be quantified by a certain metric.

- The manufacturing process can be divided in sub-processes in which it can be identified an input and output status of the manufactured product.

- The condition of the considered sub-processes can be represented by the analysis of the signals involved.

Indeed, the variability consideration of industrial plants, in terms of different kinds of manufacturing procedures, is out of the scope of this thesis research. That is, the industrial process condition forecasting is focused in this thesis from an information technology point of view, in which performing information treatment procedures and methodologies represent the challenging aspects considered.

However, the applicability of the proposed thesis requires the understanding and considerations of the specific aspects that take place in the selected manufacturing plant. In this regard, the Copper Rod Manufacturing Procedure (CRMP) is summarized in **Fig. 1.1.4**. The CRMP process is divided in five main elements:

i. The shaft furnace, is a vertical natural gas fired furnace in charge of melting the input high purity copper cathodes.
ii. The holding furnace, acts as a lung for the copper melting process which aim is to provide a constant flow of copper to the rest of the process.
iii. The tundish is a ceramic valve that controls the melted copper flow to the rest of the process.
iv. The casting wheel is in charge of solidifying by a heat extraction process the melted copper.
v. The roughing mill, reduced the diameter of the raw copper rod to meet the specified diameter conditions fixed by the plant operators.

The manufacturing process proceed as follows: first, high purity copper cathodes are melted by natural gas fired burners arranged in rows around the shaft furnace. Second, the copper flows from the shaft furnace via a gas-fired launder to the holding furnace, which acts as a lung. The holding furnace, which is also fired with natural gas, serves as a buffer to provide a constant flow to the rest of the process and, if required, can be used to increase the temperature. Third, the molten copper flows from the holding furnace via another gas-fired launder to a tundish with a ceramic valve, which feeds the casting wheel. A water cooled steel band encloses half of the casting wheel, forming the casting cavity in which the molten copper solidifies to form a raw rod by means of a heat extraction process. In this regard, acetylene, burnt with air, produces a soot dressing for the casting wheel and steel band facilitating heat transfer between the cooper and the steel band. Both casting wheel and the steel enclosure are refrigerated by means of a water cooling open circuit. Fourth, after being bevelled and shaved, the cast bar is moved to a rolling mill consisting of a roughing section and one finishing section, which reduces the bar to its final diameter. Finally, the copper bar is cooled to proceed with the coiling and packaging, where the copper rod is strapped and covered with a polyethylene film, giving with it the final manufactured product, the copper rod.

Indeed, the manufacturing process implies the transformation of the melted input copper in a solid copper rod. Such transformation is based on a controlled solidification and copper roughing process. However, the solidification of the copper is a critical aspect within the manufacturing process, in which the excessive heat should be correctly extracted from the copper bar. A correct heat extraction procedure avoids having quality issues with the final manufactured product. Indeed, quality issues are related with the apparition of a certain degree of porosity in the final manufactured unit. Porosity is caused by the apparition of oxygen bubbles in the internal structure of the copper during the solidification. Porosity in copper causes the apparition of cup-cone failures during the future wire-drawing process made by the final consumer, and therefore causes the rejection of the produced copper rod unit. An example of the cup-cone failure produced during the wire drawing of the final copper rod is shown in **Fig.1.1.3**.

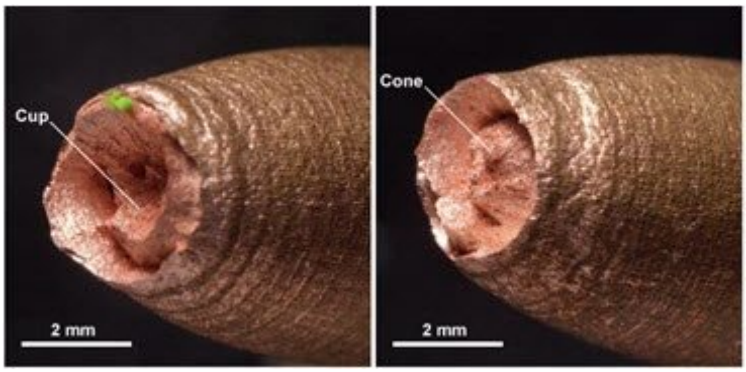

Fig. 1.1.3. Detail of a cup-cone failure within a manufactured copper rod jumbo.

The probability of having porosity in a certain manufactured rod is defined by the behaviour of the critical process signals during the solidification procedure. Deviations from the nominal operation ranges of the target signals define a certain probability of having porosity in the final manufacturing unit.

Therefore, avoiding the inherent complexity of the process, the condition of the process, and in consequence the quality of the manufactured copper rod, is summarized in regard with the probability of porosity apparition (cup-cone failures) in the final copper rod manufactured unit, called jumbo, that corresponds to 100m

of manufactured copper rod. Such probability is estimated by a manual quality inspection procedure made with samples from final manufactured copper rod jumbos. Additionally, the factory presents a Defectomat®, located at the end of the manufacturing process. In this regard, the Defectomat® is a contact-free non-destructive testing unit based on the eddy current principle that is used to detect short flaws and cracks in metallic materials. Its function is to continuously scans the manufactured copper rod in order to find signs of cavities caused by porosity in the internal structure of the copper. Then, it indicates for each jumbo manufactured the number of cavities detected.

Therefore, the manual inspection, combined with the Defectomat® information, fixes a range of porosity probability that is summarized in a discrete quality label assigned to each jumbo unit. This range is defined as the probability to find cup-cone failures in the totality of a jumbo unit. In this regard, the factory defines copper rod with the highest quality, Q1, corresponding to a probability of 0 % to 25 % of having porosity probability, medium quality, Q2, from 25 & to 75%, and low quality, Q3, from 75 % to 100%, which are the quality information provided by the company and represent the condition of the industrial process.

Indeed, two critical parts of the copper rod manufacturing process are considered for the development of the thesis. The first one, the tundish, is considered for the study since is the last reference of the input copper before starting the solidification procedure. In this regard, since the measure of the input copper flow is not given, it is critical to know the input temperature for monitoring possible deviations that cause an excess of the refrigeration in the posterior casting process. The particularities of such process is the high non-linear characteristics of the tundish temperature, which is crucial to approach industrial time series methodologies. The second one, the casting process in charge of extracting the heat from the melted copper. It is critical since any deviations in such procedure, summarized in the refrigeration index of the casting wheel, will cause an increase on the porosity probability of the final manufactured copper. The particularities of this process is the high affectation of the auxiliary signals to the refrigeration index, that should be considered in the condition forecasting approaches proposed in this thesis. Both processes will be used to test the suitability of the methodologies developed in this thesis. It should be noticed that a full list of the signals considered in this thesis is given in Annex I.

Therefore, the proposed thesis aims to propose new methodologies for the assessment of the future condition of an industrial process. It should be considered that the proposed methodologies are not generalized to any industrial process available. Even though, they should be considered and validated over an experimental environment by setting boundaries to the problem in order to assure the impact of the contributions. In this regard, the contributions are based on statistical approaches from the signals, and the method used as the validation of the hypothesis is by the application of such approaches to the presented industrial process. However, with the objective to analyse the impact and the performance of the proposed methodologies, are also considered in this thesis the application of experimental test benches available in the scientific community, such as electro-mechanical platforms, synthetic industrial process models, etc. Therefore, the methodologies proposed in this thesis are first validated in the aforesaid test benches, for then, applying them to the proposed industrial process, the Copper Rod Manufacturing Process.

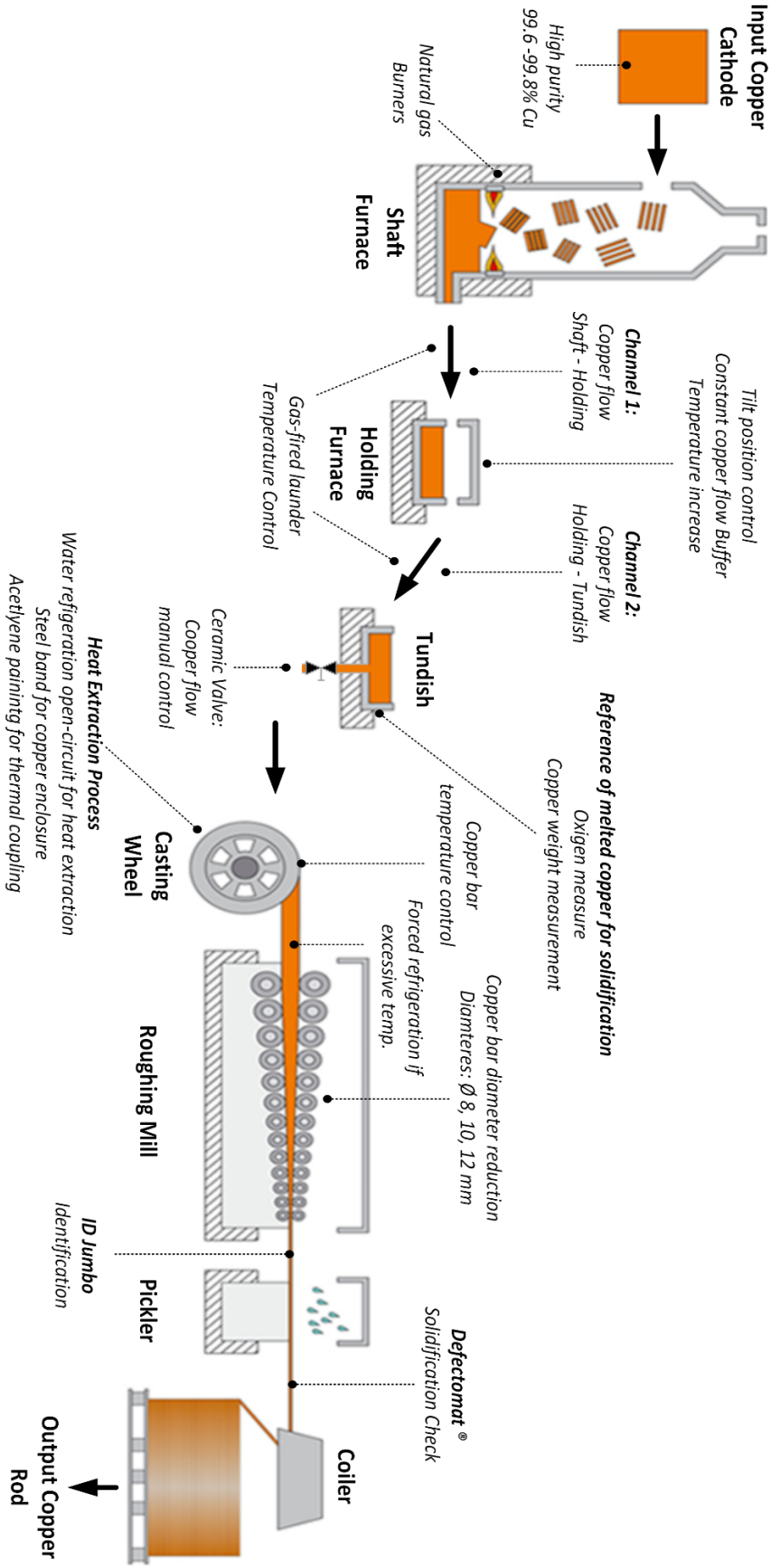

Fig. 1.1.4 Diagram of the main elements of a CRMP. Partially supported by [11].

1.2 Research problem

The most significant works in the industrial process condition forecasting topic can be found in recent literature. Regarding the fundamentals for developing the methodologies, W. Y. Wang K. [12], presented a framework called intelligent fault diagnosis and prognosis systems, which integrates data acquisition, data mining, fault diagnosis and prognosis, operation/plan optimization and feedback control. O. Myklebust [8], complements the previous work and proposes a ZDM method with focuses on maintenance, in which different types of data-driven prognostic models were able to detect early indication of abnormal behaviour or wear of manufacturing equipment. However, problems such as the definition of a coherent methodology able to deal with the variability of the data, the complexity of real industrial processes, and the adaptation of the data-driven methods to the characteristics of the application, causes a lack of coherent industrial process condition forecasting methodologies.

Dealing with proposed methodologies in the literature, J. Lee *et al.* [13], proposed a watchdog agent that bases its degradation assessment on the readings from multiple sensors that measure critical properties of the process or machinery being considered. Autoregressive moving average modelling and match matrix methods were used to forecast the condition of roller bearing based machine. Despite having good performance, **the information for assessing the condition are not properly combined**, which causes the method to achieve a poor performance with a dataset different that the used for the training procedure of the algorithms, which represents a lack of variability consideration that is critical for real industrial processes, since it **implies a loss on assessing reliability**.

Several authors, [14]–[16], propose a methodology for condition forecasting based on cyber-physical systems (CPS). CPS are models or “virtual twins” of the real machines and processes that focuses on forecasting the expected behaviour of these elements to assess their condition. This approach most based in the concept of target signal modelling which gives the final condition in regard with predicted deviations. The problem here relies on the fact that the **forecasting models are designed by a blind optimization process**. In such approach, most of the parameters of the model are configured by using an optimization algorithm, normally genetic algorithm based, that searches for the best combination of inputs in regard with a model error defined criteria [17].

The problem of these approaches is that **modelling is abstracted from the physical meaning of the process, and the intrinsic properties of the target signals to model are not considered**. That is, the resulting **models are overfitted to the training set**, since they have been configured and optimized to achieve the lowest error **without considering generalization capabilities**, which are crucial to deal with the natural variability of industrial processes.

In this regard, for developing industrial condition forecasting methodology the blocks of condition assessment and model forecasting needs to be addressed. Indeed, if coherent methodologies for assessing the future condition of industrial process wants to be defined, there are still important questions to be asked in this topic.

A relation of these questions non-properly addressed in the literature are listed next:

In regard with the future condition forecasting

(i) *Given the future behaviour of a set of target signals of an industrial process, How to fuse such information to assess the future condition of the process? That is, How to deal with the forecasting and diagnosis errors to reach a reliable outcome?*

In regard with the signal forecasting modelling

(ii) *Given a target signal from an industrial process, How to deal with model performance decrease in front of high signal's contents in terms of dynamics' information?*

(iii) *How to handle with reduced datasets, in which the representation of the different behaviours of the process is limited?, consequently, How to optimize the convergence of modelling without incurring overfitting?*

In regard with the pre-processing of available signals for forecasting modelling

(iv) *What forecasting horizon must be considered in regard with the modelling performance?*

(v) *Which past values of the available signals are optimum to be fed in the forecasting model?*

(vi) *Which auxiliary signals available in the process can be used to improve forecasting modelling performance?, also, How to combine significant auxiliary signals to allow proper forecasting modelling generalization and convergence?*

1.3 Hypotheses

In order to address the current limitations in regard with the research topic, the following hypotheses have been formulated:

H1 The future condition of an industrial process can be estimated with higher performance and reliability by fusion procedures of forecasting models that assess the future behaviour of critical process signals.

H2 The decomposition of signals' dynamics and collaborative modelling schemes will improve the performance, mainly, dealing with non-linear and non-periodic signals.

H3 The consideration of significant and non-redundant input signals in direct modelling schemes will allow higher convergence performances without incurring overfitting.

H4 The forecasting horizon, and the selection of the past values, can be optimized in terms of modelling performances by the analysis of the target signal's characteristics.

H5 The adaption of non-linear mapping strategies to auxiliary signals pre-processing will allow the improvement of the model generalization and convergence capabilities.

In consequence,

H6 The implementation of coherent industrial process condition forecasting approaches will allow higher monitoring performances, including enhanced modelling accuracy and condition assessment reliability.

These exposed assumptions represent the basis of the resulting thesis research. The hypotheses are investigated by means of the research work reflected in this thesis document.

1.4 Aim and objectives

In order to solve the research problem and test the research hypotheses, **the aim of this thesis consists of the investigation of a novel industrial process condition forecasting methodology able to assess the current and the future condition of an industrial process with high accuracy and generalization capabilities**.

Such methodology must allow the identification and isolation of process deviations before they affect condition of the process condition and, in consequence, final product's quality. The methodology is intended to be applied in a real industrial application, a copper rod manufacturing process, that serves as an industrial test bench for the development of the proposed thesis.

To successfully accomplish the thesis purpose, the following specific objectives are identified:

- **The proposal of a methodology that considers the future behaviours given by signals' models, and fuses such information to assess the future condition of the industrial process.**

- **The proposal of industrial time series modelling schemes considering signal dynamics decomposition in order to increase modelling performance.**

- **The proposal of suitable signal forecasting modelling schemes to allow higher convergence performances without incurring overfitting.**

- **The proposal of pre-processing approaches to identify optimum forecasting horizon and optimize auxiliary signals information.**

- **The validation of the proposed contributions by means of the industrial process considered in this thesis.**

Chapters' description

In order to cover the exposed objectives, this thesis has been divided into different stages, which are reflected in the chapters described below.

Chapter 2 consists on the description of the proposed contributions to enhance the pre-processing of industrial signals with the objective to solve the problematic of design the structure of forecasting models when applied to industrial time series.

In **Chapter 3**, contributions to the analysis and the consideration of the target signal dynamics to improve the forecasting accuracy and robustness versus industrial time series are proposed. In this regard, a novel signal dynamics aggrupation and a new modelling method to be applied in industrial time series is also proposed.

A methodology for the combination of the information given by the critical process signals to give the future condition of the industrial process is approached in **Chapter 4**.

Although each chapter concludes with a partial conclusion focused on its respective topic, in **Chapter 5** the thesis work is analysed from a general point of view, and the conclusions and contributions are clearly exposed.

Finally, the publications and collaborations resulting from the research work development are presented in **Chapter 6**.

1.5 State of the art in industrial process condition forecasting

The condition forecasting has been a widely studied topic on multiple application fields, such as wind turbines or heating, ventilation and air conditioning systems [18]. Industrial processes refer to manufacturing procedures composed, mainly, by electrical, thermal, mechanical and/or chemical stages, in which an asset is sequentially and repetitively produced. In this sense, the physical signals involved in a monitoring system present specific characteristics, as a complex non-periodic behaviour that only presents dependability in a certain time horizon, and non-linear relations among the inputs to define the condition of the process. In this regard, dealing with the condition forecasting of industrial processes, different approaches have been reported in the literature, however, three main schemes can be identified as common strategies.

- **Direct condition forecasting**

In this approach, represented in **Fig. 1.5.1**, the critical *n* time-based signals identified from the available historical database, *X (t,n)*, are introduced as inputs into a forecasting model. The model performs both, the determination of the relations among the inputs, and the assessment of the future condition of the system at a defined forecasting horizon, *p*, *C (t+p)*, usually, as a continuous parameter related with the remaining useful life of the component under inspection or the expected time to failure of the process. Note that in this type of approaches, the input data and the ability of the model to combine the inputs critically affects the final result of the prognostic.

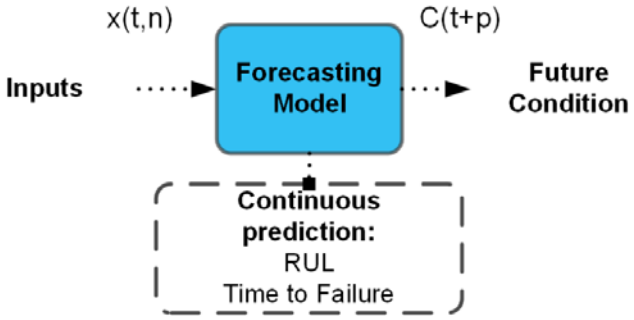

Fig. 1.5.1. Main block diagram of a direct forecasting scheme

A significant work following this approach is the one presented by A. K. Mahamad, *et al.* [19], in which an artificial neural network model is used. The proposal considers time and fitted measurements of root mean square (RMS) and kurtosis from current and previous instants of time in order to estimate the remaining useful life of an electromechanical machine. Also, T. Benkedjouh *et al.* [20], presents a method for condition forecasting of critical process components based on Principal Component Analysis (PCA), and support vector data description. In this study, PCA is used to reduce the dimensionality of the input signals, while support vector data description is used to fit the trained data into a hyper sphere, which radius models the condition of the system as the increase of data dispersion with the degradation of the nominal condition.

The direct condition forecasting is the simplest approach, and provides a good response dealing with a reduced number of input variables, or in those applications in which the number of predefined conditions (classes), is reduced. That is, the main limitation dealing with diagnosis and forecasting models lies on the overfitting of the resulting mathematical description. In this sense, dealing with direct condition forecasting, the

increased demand of defining both diagnosis relations and forecasting estimations, lies to an over-fitted response of the resulting scheme, in which the generalization capabilities in front of new data are compromised.

- **Discrete condition forecasting**

This approach, represented in **Fig. 1.5.2**, corresponds to the classical diagnosis scheme, in which the available time based signals are processed to be characterized by a set of numerical features. Then, the set of features is used as input to a classification algorithm, which relates the pattern with one of the predefined classes following the relations defined during the previous training stage. Thus, the classifier determines the current condition of the process, mostly in a discrete mode, that is, one condition among all the available, although severity degree is often delivered. Later, a forecasting model analyses information regarding previous conditions to forecast the next expected condition.

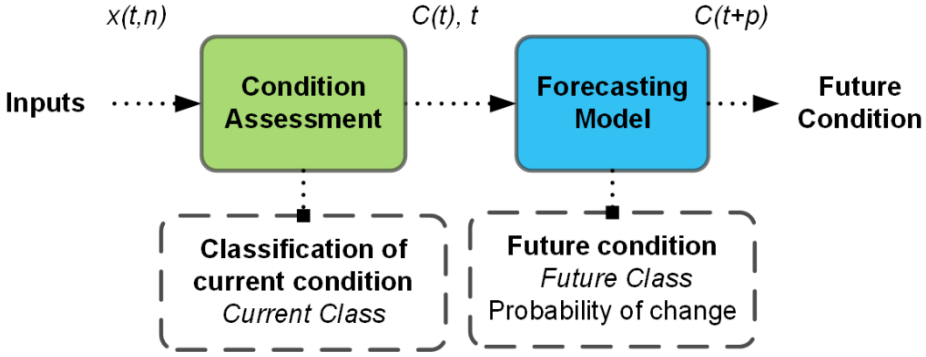

Fig. 1.5.2. Block diagram of the discrete condition forecasting approach.

A clear example of this approach is presented by C. Chen *et al.* in [21], where a classical classification scheme is applied to assess the condition of the system, but, later, a dynamic wavelet neural network is applied as predictor to forecast the expected failure progression. Also, A. Soualhi *et al.*, in [22], proposes a bearings condition forecasting approach where, first, a set of time domain features is estimated from vibration signals. Next, the features are used as inputs to an unsupervised classification algorithm, an artificial ant clustering. Finally, the imminence of the next degradation status is given by hidden Markov models, and the estimation of the remaining time before the next degradation state is given by an ANFIS algorithm.

In general terms, the discrete condition forecasting approach entail enhanced accuracy and efficiency in condition assessment. However, the main disadvantage of this approach lies on the low resolution of information used for the generation of the forecasting model. Indeed, only current and previous values of the discrete condition are used to forecast the future behaviour. That is, diagnosis output corresponds to class's assignment along time, this information is useful dealing with forecasting under run to failure scenarios, where the evolution of a specific fault is tracked in time. However, such monotonic approaches do not perform properly dealing with multiple causes of deviation in industrial processes defined by multiple non-linear relations among signals.

- **Condition assessment based on signal forecasting**

In this approach, represented in **Fig. 1.5.3**, the available signals or the corresponding characteristic features are modelled by dedicated forecasting models. Later, the forecasted values are combined to retrieve the future condition of the system in a predefined time instant. In this regard, the assessment block is in charge of extracting the condition information in regard with the future behaviour of the inputs modelled.

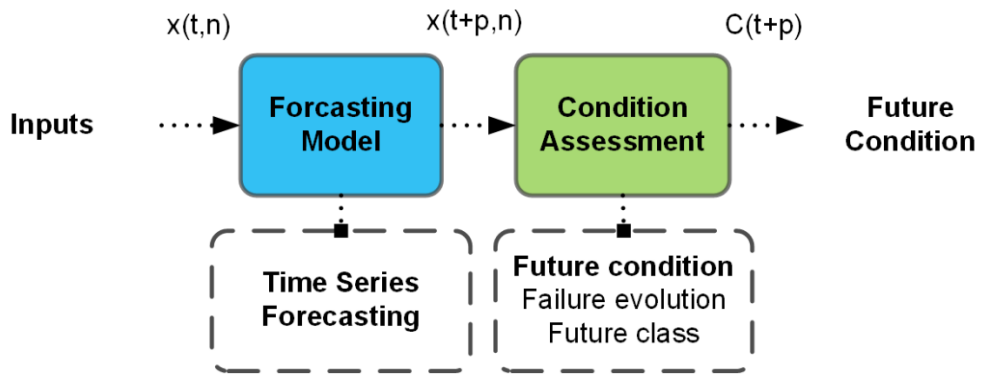

Fig. 1.5.3. Block diagram of the future condition assessment based on signal forecasting.

Significant studies based on this approach have been presented. For instance, C. Chaochao *et al.* in [21], use an Adaptive Neuro Fuzzy Inference System (ANFIS), trained via historical process data. The ANFIS output is introduced into a hidden Markov model to describe the failure condition propagation among time. Also, A. Widodo and B-S. Yang, in [23], present an intelligent machine prognostics system based on survival analysis and Support Vector Machine (SVM). Survival analysis utilizes censored data to forecast the survival probability of failure time of machine components. SVM is trained from historical data that corresponds to target vectors of estimated survival probability. After validation process, SVM is employed to assess the failure time of individual unit of a machine component.

Discussion and conclusions

The condition assessment based on signal forecasting provides good performances since the modelling of the future condition is approached as a classical time series problem preserving as much information as possible from the original information. However, this approach drags two different sources of error, the inherent forecasting error of the input modelling block, and the class assessment error introduced by the classifier. Therefore, to overcome the limitations of this approach, efforts should be made in the improvement of both forecasting model and condition assessment stage. **The main contribution of this thesis lies on the proposal of a condition forecasting scheme that overcomes the accuracy and performance limitations of classical methods. In this regard, the proposal combines the modelling of the critical process's signals to obtain their future behaviours and, later, the condition of the process at a future time by means the non-linear combination of such forecasted time-series.**

In this regard, the presented state of the art of this thesis is focused on covering the two main blocks used in any industrial condition forecasting approach: the forecasting model, and the condition assessment. Therefore, state of the art will cover an **introduction to the data formatting** in industrial process monitoring. This is made in order to illustrate the problematic of dealing with real industrial process data, such as the filtering and adequation of the input data. Then, an overview of the **main families of methods to perform forecasting models for industrial time series** is given, giving special attention to ANFIS based time series modelling. Finally**, the main methods behind the condition assessment block** in charge of both the combination of the information, and the assessment or classification of the condition are explained

1.5.1 Data formatting in industrial process monitoring

The classical data acquisition stage represents the extraction of raw data from sensed physical magnitudes present in the industrial manufacturing process. Indeed, the most common acquisition approaches are based on physical magnitudes such as currents, vibrations, ultrasonic measurements, acoustic emissions, temperature, pressure, stress, and others.

Industrial data acquisitions systems are evolving due to the information complexity in most of industrial processes. This evolution is based on include certain signal processing capabilities in the sensors in order to reduce the quantity of information to transit and process, and the ability to create networks of sensors in order to improve, contrast and complement the quality of the information acquired from the plant [17]. These kind of sensors are being called smart sensor and defined as sensors that provide extra functions beyond those necessary for generating a correct representation of the sensed quantity [18]. Smart sensor can also monitor the physical magnitudes, but they can project this information to a higher level of abstraction for instance, giving information about the degradation level of the gears of a machine [19].

In the literature, the considered information extracted from a manufacturing system can be structured as shown in **Fig. 1.5.4**. Data from the machines and equipment will be captured using hard/smart/soft sensors in order to evaluate the machine condition and detect any kind of failure [21]. Process and control information can also be used to evaluate the condition of the machine or either to assets the performance of a specific process. Furthermore, it is also interesting to record and use for condition assessment purposes the alarms triggered in the system due to high/low values, abnormal operation condition, and any other non-expected event [13, 20]. K.-S. Wang [16], in 2013, reported that the integration of product information is useful in order to handle the defect propagation within the processes and readjust the prognostic and condition assessments models versus new real performance data.

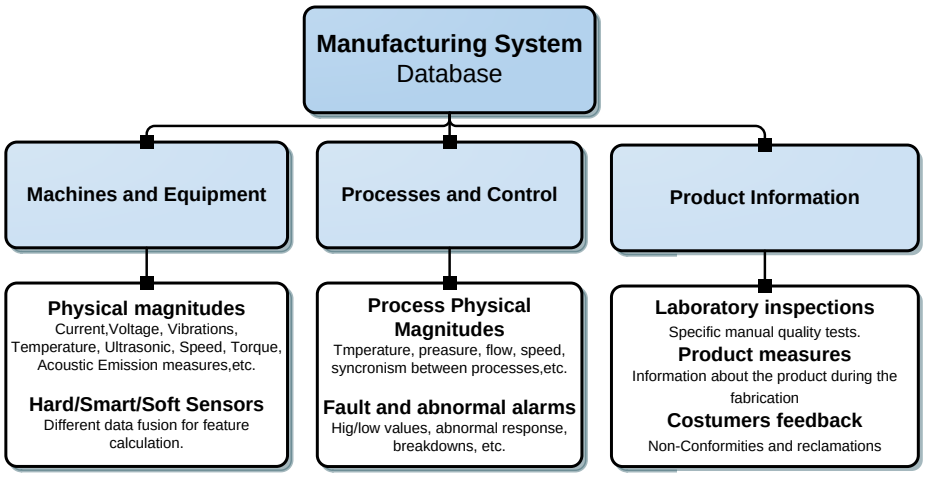

Fig. 1.5.4. Considered information acquired from an industrial process.

Once data is acquired, it is necessary to assure both coherence and significance of the data. In this regard, as all data is automatically acquired and stored, either by the PLCs of the factory or the SCADA systems, it may present incoherencies that should be fixed. For instance, it is usual that during the normal operation of the processes the acquisition devices produce a failure in the storage of a single observation, that is marked as an

incoherence in the database, known as Not a Number (NaN). Therefore, such values should be identified and removed from the analysis [24].

An important issue is related with the coherence of the time vector. Indeed, it is common that due to external problems or interruptions, the acquisition systems miss some points during the storage of the values. This should be detected by analysing the increments of the time vector to detect those areas where the stored samples do not follow the specified sampling frequency. Therefore, the number of missing points should be estimated, and the missing samples interpolated in order to assure the coherence of the signals [25]. Other observations that should be detected and removed and those related with non-desired behaviours of the process, high periods of inactivity, time periods where the process is under some kind of maintenance action, or where new provisional configurations are being tested should be identified and removed from the posterior analysis, since its behaviour is considered to be an exceptional situation that cannot be considered in the modelling procedure.

After the proper consolidation of the database, the condition assessment or the modelling approaches can be started. In this regard, if the data is not properly conditioned, several errors will be induced to the posteriors steps of the condition forecasting methodology. Next section covers the state of the art in the generation of forecasting models for industrial time series.

1.5.2 Modelling methods for time series forecasting

This section covers the state of the art in the modelling methods for time series forecasting. Even though, an initial explanation should be made in order to contextualize the term forecasting and explain its relation with condition assessment.

Condition assessment is able to determine if the system is working properly or not by calculating, analysing and classifying a set of features extracted from the system. But answering the question of: *"What is the remaining operational lifetime of a system once a failure condition is detected?" or What is the expected behaviour of this process among time?* with condition assessment methods is not possible. This is due to the fact that in condition assessment the time factor is not involved, since it is a specific analysis to determine the condition of the system in a specific moment of time. However, if the actual and previous condition states of the system are known, a forecast of the estimated condition evolution among time can be performed, this origins the concept prognosis which is to predict a failure or a behaviour before it occurs [26]. It should be noticed, that for this thesis, the terms forecasting and prognosis represents synonyms in regard with predicting future behaviours or conditions.

Formally, **Prognosis** is literally the ability to acquire **knowledge** (gnosis) **about events before they actually occur** (pro). Accordingly, the term prognosis has been used first in the medical field to imply the forecasting of the probable course or disease. However, in the state of the art of condition monitoring, several definitions can be found in: M. Lebold and M. Thurston [27], define prognosis as the ability to perform a reliable and sufficiently accurate prediction of the **Remaining Useful Life (RUL)** of equipment in service; and specify that the primary function of prognostic is to project the current condition state of equipment into the future taking into account estimates of future usage profiles. A. Hess in [28] defines that prognosis, in its simplest way, is to monitor and detect the initial indications of degradation in a component, and be able to consistently make accurate predictions. For the International Organization for Standardization: "Prognosis is the estimation of time to failure and risk for one or more existing and future failure modes" [29]. G. Vachtsevanos, et al. in [30], define "prognosis" as the ability to predict accurately and precisely the RUL of a failing component or subsystem. A. Heng, et al. in [31], defined Prognosis as the forecast of the RUL, future condition, or probability of reliable operation of system based on the acquired condition monitoring data (current and previous).

Forecasting is approached by different ways, and a wide range of techniques and methodologies can be found in the actual state of the art. Despite this fact, compared to condition assessment, the literature of forecasting is much smaller, which indicates is a quite new area of interest [32]. Forecasting and prognosis methods can be classified in different ways regarding the way they are implemented (AI, statistics, experience), their objective (RUL estimation, time to next state estimation, etc.), for this research work, the classification of forecasting methods will be made according to their nature. Therefore, three big families can be found, as shown in **Fig. 1.5.5**, which are: (i) Model-Based prognosis; (ii) Probability based Prognosis, and (iii) Data-Driven Prognosis. [16]-[17].

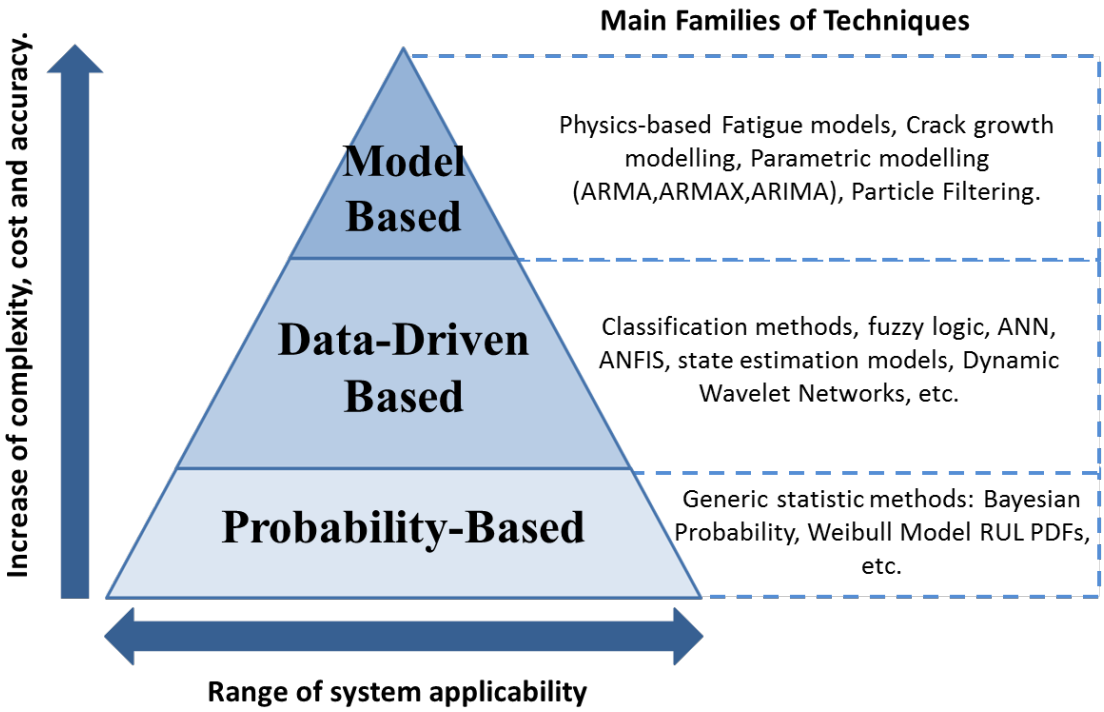

Fig. 1.5.5. Main classification and basic characteristics of the prognosis methods

Probability-based modelling

For simple non-critical systems, there is not the necessity to create a complex model that tries to simulate the real physics involved in the system, or the dynamical response in terms of stochastic modelling. If a good data historic from previous failure conditions is accessible, for example the shown in **Fig.1.5.6,** Probability-Based prognosis methods can be used.

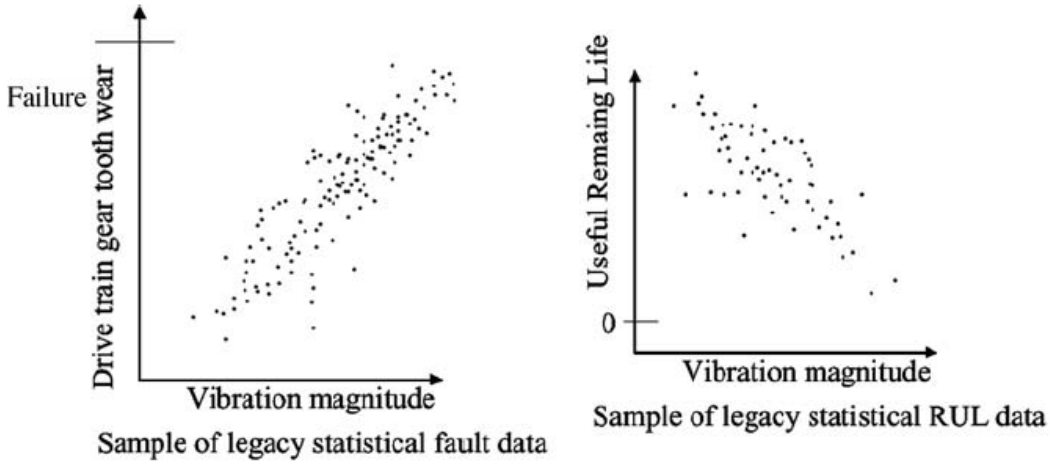

Fig. 1.5.6. Historical data from previous failures used to estimate the RUL of the component [30]

These methods require less computational cost and complexity, due to the fact that all the prognostic issue resides in estimating different Probability Density Function (PDF). These PDFs can be estimated in base of the observed statistical data, in spite of solving the nonlinear dynamic equations, and can give sufficient information

from the prognosis point of view, such as the RUL, the TTF, etc. Due to the fact that these methods are based on statistical analyses methods, they also can give confidence limits for their prediction, giving with this an indicator of accuracy and precision of the prediction.

The most known probability of failure curve , also known as hazard function, was proposed by H. Chestnut [34], and it is known as bathtub curve , which is shown in **Fig. 1.5.7**. This curve identifies the different probabilities of failure of a machine or component among time and it is divided in three phases: (i) At the early phase, all the elements are new and the rate of failure is relatively high, due problems such as defective components, misalignments, etc., the failures during this initial phase are known as infant mortality or wearing failures. (ii) Once the components settle, the failure rate is almost constant and low. Finally, after specific time of use, the failure probability increases till the total failure of the system; this is known as wear-out failure.

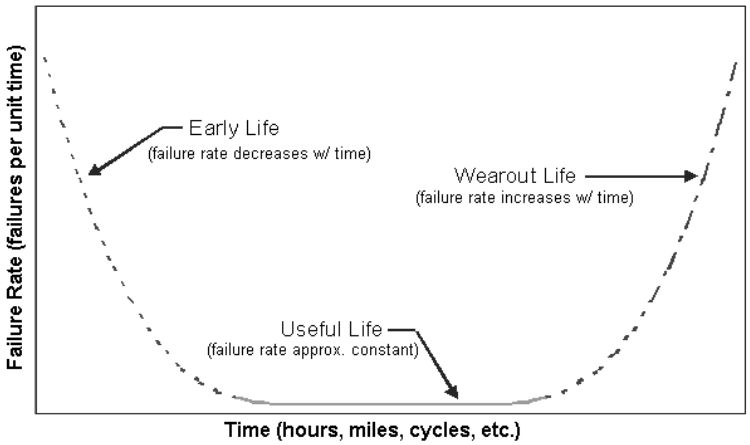

Fig. 1.5.7. The Hazard function known as the bathtub curve, probability of failure versus time.

In this regard, the main probabilistic methods are the Bayesian probabilistic analysis, the Dempster-Shafer's theory and the time to failure analysis using the Weibull distribution. It should be noticed that other probabilistic techniques such as Hidden Markov Models and Fuzzy logic can also be used to estimate the PDF of a historical data, but as they are more related to complex models and AI techniques, considered data-driven approaches. In this regard, S. Zhang, et al. [35], presents a method using recursive Bayesian analysis to estimate asset condition reliability using condition monitoring data. It enables reliability evaluation using observations from individual assets. The method can be performed in an on-line mode, and functions in both normal and abnormal zones. Remaining useful life can also be evaluated based on the derived reliability curve; given an acceptable reliability (hazard). An experiment on bearing life testing was employed to validate the method. B. Saha, et al. in [36], proposes to use Bayesian statistical approach to measure the RUL PDF under operational conditions in complex systems whose internal state variables are either inaccessible to sensors or hard to measure under operational conditions. Accordingly, models of electrochemical processes in the form of equivalent electric circuit parameters were combined with statistical models of state transitions, aging processes, and measurement fidelity in a formal framework.

G. Niu and B-S. Yang in [37], develop a data-driven machinery prognosis strategy for industry application. They use a regression technique to perform the task of time-series prediction. The proposed scheme was validated using condition monitoring data of a methane compressor to predict the degradation trend. The author concludes that the proposed methods have a low error rate. W. He, et al. in [38], proposed a method to forecast

the RUL for lithium-ion batteries using regression and the Bayesian Monte Carlo methods. In this work, an empirical model based on the physical degradation behaviour of lithium-ion batteries is developed. Model parameters are initialized by combining sets of training data based on DST. Bayesian monte carlo is then used to update the model parameters and predict the RUL based on available data through battery capacity monitoring. The author concluded that, as more data becomes available, more improves the accuracy of the model in predicting RUL.

A. K. Schömig and O. Roser in [39], investigate how the Weibull distribution is suited as an alternative to common approaches in the stochastic modelling of machine downtimes and times between failures, and present an algorithm that automatically calculates Weibull parameters from machine data based on the SEMI E10 standard. Q. Zhang, et al. in [40], construct a mixture Weibull proportional hazard model to predict the failure time of a mechanical system with multiple failure modes. The historical lifetime and monitoring data of multiple failure modes are combined to estimate the system failure probability density and reliability. Acquired data is the input of the model that estimates the system reliability and predict the system failure time.

Discussion and conclusions

Therefore, probabilistic approaches are mainly used when the problem to be approached is simple in terms of signal complexity, or the metric that wants to be extracted at a future horizon is a basic probability of future operation. Indeed, the literature also identifies the inherent difficulty of forecasting the future behaviours by using statistical methods, due to the fact that several problems such as manufacturing variability, mission history variations, life degradation, and false-alarm probability are present in most of the systems. For this reason, the performances of such metrics are fare away from the requirements of an industrial condition forecasting scheme and are not suitable for the development of the methodologies from the proposed thesis.

Model-based modelling

Model–based prognosis methods try to find a mathematical model that formally describes the behaviour of the monitored system. By integrating physical and stochastic modelling techniques, the model can be used to evaluate the distribution of RUL as a function of uncertainties in component strength/ stress properties, loading, or lubrication conditions for a particular fault. These techniques have the greatest accuracy among the forecasting techniques, due to the fact that ideally, the generated model is a virtual twin of the monitored system. However, they require high computational cost and a complexity, sometimes unapproachable for most of complex industrial systems. The model should be defined specifically for the concrete application. In these methods, three different approaches can be found: physics-based models and stochastic-based models [41].

Some characteristics of physical models are: (i) They require an immense computational cost due to the complexity of the physical formulation behind its behaviour, most of the time these models are implemented by means of Finite Element Analysis (FEA). (ii) High complexity for nonlinear or multi-stage systems, all the behaviour of the system should be modelled, this includes the interactions between different materials, components and subsystems, sometimes unapproachable for most of industrial processes; for this reason, these kind of models are often applied to characterize a single component such a bearing or a gear. (iii) High

accuracy and precision, the physical formulation describes exactly the behaviour of the component, to the point that these models do not need any measurable event to forecast the evolution of the damage (iv) Low generalization capabilities, these models are totally specific for the application.

There are mainly, two main families of model based forecasting approaches: (i) Physics based modelling. Typically involve building mathematical models to describe the physics of the system and failure modes. These models attempt to combine system-specific mechanistic knowledge and data to provide ‘‘knowledge-rich’’ prognosis output [42]. (ii) Parametric Modelling. An identification methodology is applied in order to find a polynomial based model that establishes a relation between the input signals and the measured outputs [43].

In these kind of methodologies, the structure and the coefficients of the model is previously fixed by the user, the most commonly used model structures are ARMA (auto-regressive moving average model) [44], ARMAX (auto-regressive moving average model with exogenous inputs) [45], and ARIMA (auto-regressive integrated moving average model) [46]. Once the structure is fixed, the problem turns into the identification of those coefficients that minimize the error between the output of the model and the expected behaviour of the system represented by the acquired data. These techniques are less computational expensive than the physics-based models due to the fact that most of the work is done offline, and once the model has been created, it can be used to predict the output of the system.

Most of the different structures proposed in the state of the art in parametric models for forecasting are based in these model architectures and identification errors. However, it is difficult to find forecasting methodologies based only in the utilization of parametric models, several modifications and combinations of model identification with probabilistic or data-driven methods can be found. For instance, there is a great synergy between AI methods such as Neural Networks and parametric models. In this topic, A. A. El Desouky and M.M. Elkatab present in [47] a modelling approach where different neural-network configurations with an adaptive learning algorithm are pursued using a hybrid ARIMA-ANN to arrive at a suitable prognosis model for forecasting the load demand.

M. Orchard, et al. in [48], suggest that a possible solution to forecasting problems (such as large-grain uncertainty, critical states consideration, and randomness in the model generated), is based on the application of Particle Filtering (PF). Particle filter is based on the recursive Bayesian estimation techniques; they combine information from physical models, and on-line data obtained from the process to readjust the predictions. It should be noticed that the used models are assumed to be nonlinear to accurately represent the underlying dynamics of the physical system, and also non-Gaussian to represent complex uncertainties and disturbances. This is of particular benefit in diagnosis and prognosis of complex dynamic systems, such as engines, gas turbines, gearboxes, etc., because of the nonlinear nature and ambiguity of the rotating machinery world when operating under fault conditions. Moreover, particle filtering allows information from multiple measurement sources to be fused in a principled manner. Under this approach, prognosis is based on both an adaptive failure or deviation progression model and an accurate Particle Filter estimate of the current state. Both pieces of information allow predicting the trend of fault evolution or process deviations in time and therefore the probability of failure if load conditions are maintained. As new data becomes available, long-term predictions are updated and a learning correction-based procedure is used to improve both their accuracy and precision.

Several authors also work in the combination of high order particle filtering with AI methods for conducting forecasting of industrial processes; in this field, such as the work made by C. Chen et. al in [21], where

prognostic method is developed using ANFIS and high-order PF. Khan, et al. in [49], present a combination PF and stochastic modelling based prognosis study for predicting remaining useful life of steam generator tubing in which a PF is used to predict the RUL of the steam generator tubes. The uncertainty involved in prediction of remaining useful life of tubes using particle filter algorithm is controlled by an RUL correction loop based on an ARMA. W. Jinjiang and R. X. Gao in [50], proposes also an interaction scheme for the combination of stochastic models and PF in order to improve the accuracy and reduce the complexity presented by physics-based models in the problem of bearing remaining life prediction. Additionally, B. E. Olivares, et al. in [51] presents the implementation of a particle-filtering-based prognostic framework that allows estimating the state of health (SOH) and predicting the remaining useful life (RUL) of energy storage devices, and compares the performance of a PF model with the ideally physics-based prognosis model in terms of accuracy, precision, complexity and computational cost.

Discussion and conclusions

Model based approaches presents many limitations to be applied in the modelling of industrial process behaviours. In this regard, physical modelling requires the mathematical definition of all the relations between the input material, the processes and the machines involved. Such approaches can be applied to monitor one specific critical part of the plant, but is unavailable for most industrial plants for modelling the whole process.

Otherwise, parametric modelling presents more affordable solution, since only the architecture of the model needs to be fixed. Despite this fact, it has been demonstrated in literature that current AI methods outperform the classical model parameter estimation approaches by recursive estimation. In this regard, the adaptation and generalization capabilities of data-driven methods outperform classical parametric modelling.

Beyond all, particle filtering is an interesting approach that presents good contributions to be applied in the modelling of industrial processes. Since it provides a link to re-estimate the output of a physical model in regard with the observed data, which is translated into an improvement of the generalization and adaptation to the process behaviour, classical limitations of model based approaches. However, a physical equation that fixes the behaviour of the industrial process monitored is required, even though, this can be difficult to be defined for complex processes in which the relations among the different signals involved is not clear.

As a conclusion, the consideration of physical based modelling for industrial condition forecasting results in high computational cost strategies that requires a deep knowledge of the monitored process, for this reason the thesis will be focused on data-driven models, which are more aligned with the modelling necessities of industrial processes.

Data-driven modelling

Due to the complex nature of industrial processes both physical and probabilistic methods present many limitations in terms of: computational cost and generalization, for the physical approach, and a lack of precision and adaptability to the probabilistic ones [52]. On the other hand, technology improvements in the field of sensing and instrumentation is leading industries into rich data environments, in which data regarding different parts of the process is acquired and systematically stored in a database.

Data-Driven approaches aims at using real data acquired from the process in order to generate a predictive model able to approximate and track the behaviour of the process, and then forecast the global behaviour of the system; instead of building models based on the physical comprehension of the system or the human expertise [32]. These kinds of methods usually use AI paradigms to learn from “examples” and try to capture the subtle functional relation among the data, for then reproducing the expected behaviour among time in base to new inputs. Data-Driven methods present minor accuracy and lower complexity compared with model-based methods, although they present great generalization and data fusion capabilities, and are more suitable for real industrial scenarios, which means that the user does not need a deep knowledge of the interactions between the process being monitored. Therefore, these methods are more suitable for dealing with forecasting in complex nonlinear systems, such as industrial processes where high amounts of data are involved, which is precisely the scenario where this research work is focused.

Data-driven techniques, especially those based on AI algorithms, have been increasingly applied to machine and processes prognosis, and the state of the art is nowadays a trending research topic around the world. However, it should be remarked that industrial time series forecasting does not has a clear methodology, and a high variety of theoretical methods and different combinations of techniques can be found in the state of the art. It is well known that the utilization of Neural Networks or any of their derivatives is wide, P. J. Werbos in [53], reported, in one of the earliest works which such modelling approaches, that ANN trained with the backpropagation algorithm outperform traditional statistical methods such as regression and stochastic modelling such as Box-Jenkins. F. S. Wong in [54], collects the previous work, and describes a backpropagation neural network approach for time series forecasting, the structure is applied to forecasting of electricity load, stock market and interbank interest rate forecasting with promising results. In the prognosis field, T. Khawaja, et al. in [55], introduces a novel confidence prediction neural network which aims to predict the time to failure as close as possible to the real one, which is constructed with a Q-learning routine that improves online prognostics accuracy. J. Abdi, et al. in [56], compares different IA approaches to forecast traffic flow, and concludes that the forecasting algorithm which uses the Q-learning algorithm is capable of finding the optimal forecasting approach. Which remarks the suitability of unsupervised methods for performing forecasting models.

Z. Xiaodong, et al. in [57], proposed scheme consists of three main components including PCA, and Hidden Markov Models (HMM) and an adaptive stochastic fault prediction model. The principal signal features extracted by PCA are utilized by HMM to generate a component health/degradation index, which is the input to an adaptive prognostics component for on-line remaining useful life prediction. M. Dong and D. He in [58], present a statistical modelling methodology for performing both diagnosis and prognosis in a unified framework. The methodology is developed based on segmental hidden semi-Markov models (HSMMs). An HSMM is a HMM with temporal structures. As a conclusion, they expose the performance increase of HSMMs in regard with the classical HMM approach is about in 30 %, which is the result of including the temporal information of the signal in the modelling.

The specific literature published to date shows that forecasting of industrial manufacturing plants, in terms of target signals evolution for supervision purposes, is still a novel field for research, while similar fields like wind turbines and electric load consumptions are being dealing with since a decade ago [59], [60]. In comparison with wind speed and load consumption application fields, forecasting of target signals in industrial processes implies a combined scenario where the model needs to predict non-periodic and chaotic signals with complex

dynamics, as is the case of wind forecasting, that are influenced by operating conditions defined by non-linear relations among auxiliary process signals, such as in load forecasting. Indeed, an important subject dealing with forecasting performance is the dynamics contents of the forecasted signal.

In this regard, unlike model-based approaches by means of physical and mathematical knowledge of the process, data-driven schemes are becoming in an efficient alternative considering the usual availability of industrial historical records in terms of time plots of various signals [52]. The two main limitations related to industrial process forecasting found in literature are: First, how to design a model from a target signal when this signal presents a chaotic and non-periodic behaviour, as is the case of industrial processes, where the correlation of the target signal with the associated process quickly decreases within a short period of time [61]. Second, how to take advantage and exploit of the information regarding auxiliary signals of the process by introducing them in the model without increasing the complexity [62], [63].

Furthermore, artificial neural networks have been developed to solve the nonlinear time series forecasting problem in the industrial field [64]. Thus, in [65], an autoregressive integrate moving average model is presented for wind speed forecasting supported by particle swarm optimization. Also, in [66], a genetic algorithm is applied to select the optimum time signal delays to be applied over the inputs of different forecasting algorithms for performance comparison. Advanced nonlinear methods from the field of artificial intelligence have also been widely employed for electric load forecasting, in which due to the periodic nature of loads, most of the effort is focused on identifying trends of the operating point. Therefore, in [67], three advanced machine learning feature selection algorithms, mutual information, RRelief and correlation-based selection, are applied in a two-step approach previous to the neural network algorithm for electricity load prediction. Similarly, in [68] the performance of various feature selection algorithms is analysed in terms of electric utility load forecasting based on support vector regression machines. In [69], the optimum set of energy drivers is estimated by GA in order to increase the performance of the energy forecasting modelled by means of an Adaptive Neuro-Fuzzy Inference System (ANFIS). ANFIS models fuses the parametric adaptability of ANN, and the generalization capabilities of the fuzzy logic. Indeed, ANFIS based forecasting models offers a very reliable and robust condition predictor, since it can capture the system dynamic behaviour quickly and accurately [70]–[72]. The integration of both knowledge fuzzification and non-linear functions estimation provide smoothness and adaptability to the modelling process. Therefore, due to the high applicability, ANFIS conforms the main state of the art of time series forecasting. For this reason, the explanation of the ANFIS and its different structures when dealing with time series forecasting is given in a dedicated section.

1.5.3 Adaptive neuro fuzzy inference systems

ANFIS Modelling architecture

There is a great deal of combinations between AI techniques in order to take advantages from each of the AI techniques involve. However, the most important hybrid system is the Adaptive Network-based Fuzzy Inference System (ANFIS). The ANFIS method was proposed by J. S. R. Jang in 1993 [73]. This method is a universal approximation algorithm, which takes into account both Fuzzy Logic System and Artificial Neural Network technologies: It is fuses the parametric adaptability of NN, and linguistic functions of the fuzzy logic methods. Its structure is totally fixed, and it follows a feedforward connection schema [74]. The ANFIS architecture can be divided in five different layers as shown in **Fig. 1.5.8**.

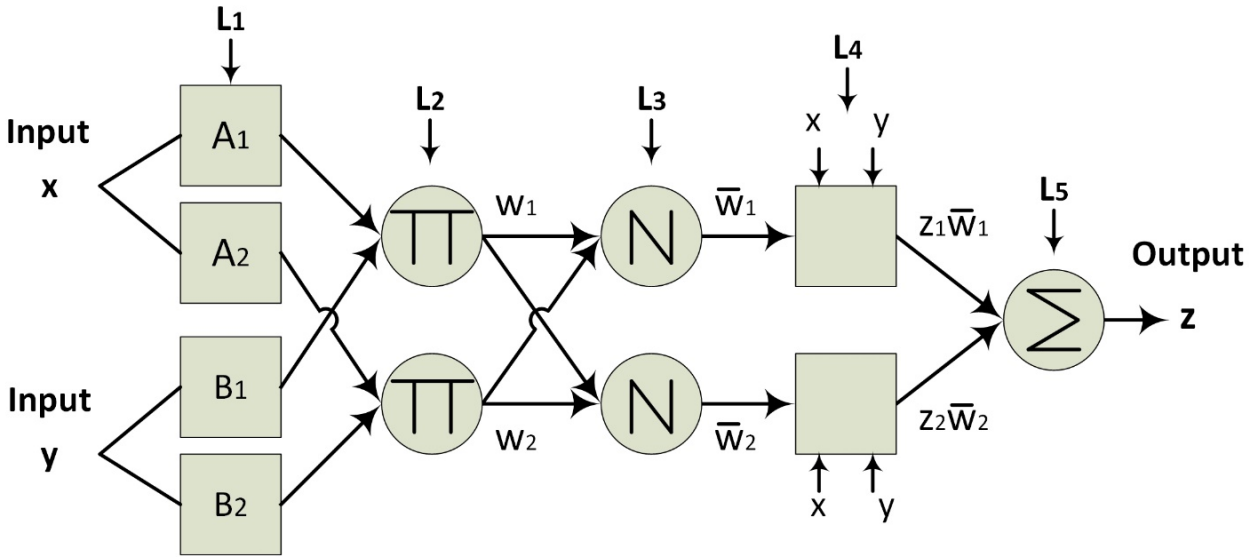

Fig. 1.5.8. ANFIS scheme considering two inputs and two membership functions to fuzzify each input.

Several nodes can exist in each layer, in order to explain the behaviour; the notation used is O_i^j as the *i-th* node of the layer *j*.

- **L1: Input Layer**

This layer is in charge of addressing the input signal to a node. Each node i, is an adaptive node with an output function defined as:

$$O_i^1 = \mu \cdot A_i(x)\,, i = 1{,}2 \qquad \textbf{Eq. 1.5.1}$$

$$O_i^1 = \mu \cdot B_{i-2}(y), i = 3{,}4 \qquad \textbf{Eq. 1.5.2}$$

where x and y are the inputs to the nodes; A_i and B_{i-2} are the linguistic labels associated with the nodes, this membership functions are generally defined as Gaussians functions. Fuzzy ruled-based system, can be also instead. However, the fuzzy membership functions do not need to be known since ANFIS will adaptively determine these membership functions. However, a good initialization of these sub-networks will aid in training of the Neural-Fuzzy system by giving it a consistent seed value, the final convergence of the learning algorithm will be increased.

- **Layer 2: Fuzzification**

Each node of the layer multiplies the input signals, as shown by equation shown in Eq. 1.5.3. Then, each output represents the activation variables of a fuzzy rule.

$$O_i^2 = W_i = \mu A_i(x) \cdot \mu B_i(y), \quad i = 1,2 \qquad \textbf{Eq. 1.5.3}$$

- **Layer 3: Fuzzy Rules Definition**

The fuzzy rule module provides the antecedent consequence statements of fuzzy logic. These statements provide the condition of the fault being monitored given the linguistic operating range of the inputs. Each node calculates the reason between the *i-th* activation variable, and the sum of all the activation variables Eq. 1.5.4. The outputs of this layer are called normalized activation variables.

$$O_i^3 = \overline{W_i} = \frac{w_i}{w_1 + w_2}, \quad i = 1,2 \qquad \textbf{Eq. 1.5.4}$$

- **Layer 4: Defuzzification**

In this layer, each node computes the contribution of the i-th fuzzy rule over all the outputs by applying Eq. 1.5.5. The outputs of each node are named consequents.

$$O_i^4 = \overline{W_i} \cdot Z_i = \overline{W_i} \cdot (p_i x + q_i y + r_i), \quad i = 1,2 \qquad \textbf{Eq. 1.5.5}$$

- **Layer 5: Output Layer**

The nodes of the last layer calculate the output signals as the summation of all the input signals as shown in equation Eq. 1.5.6.

$$O_i^5 = \sum_i \overline{W_i} \cdot Z_i = \frac{\sum_i \overline{W_i} \cdot Z_i}{\sum_i W_i} \qquad \textbf{Eq. 1.5.6}$$

Wang, et al. in [75], compare the performance of recurrent neural networks and ANFIS systems, which were adopted to develop an on-line machine fault prognostic system. In order The performance of the developed prognostic system is evaluated by using three test cases from mechanical components. The author concludes that the ANFIS based prognostic system offers a very reliable and robust machine health condition predictor since it can capture the system dynamic behaviour quickly and accurately.

Forecasting performance evaluation

Classically, in order to evaluate the performance and the accuracy of a model, statistical metrics have been used, which are: Root Mean Squared Error (RMSE), MAE and MAPE, which are defined below [76]. The RMSE, shown in Eq. 1.5.7, is one of the most used performance metrics for the development of forecasting models, is a measure of the standard deviation of the differences between predicted values and observed values. It is useful in order to analyse the global behaviour of the model, but is very sensitive with the amplitude.

In this regard the MAE, Eq. 1.5.8, is used to evaluate the forecast since it is less sensitive to outliers. Finally, the MAPE, Eq. 1.5.9, helps to determine the mean deviation of each sample normalized by the amplitude, so it helps to unify the scale and can be used to compare the errors of signals with different levels of amplitudes.

$$RMSE = \sqrt{\frac{\sum_{l=1}^{L}(y(\mathrm{t}) - \hat{y}(t))^2}{L}} \qquad \textbf{Eq. 1.5.7}$$

$$MAE = \frac{\sum_{l=1}^{L}|(y(\mathrm{t}) - \hat{y}(\mathrm{t}))|}{L} \qquad \textbf{Eq. 1.5.8}$$

$$MAPE = \frac{\sum_{l=1}^{L}\left|\frac{(y(t) - \hat{y}(\mathrm{t}))}{y(\mathrm{t})}\right|}{L} \cdot 100\% \qquad \textbf{Eq. 1.5.9}$$

where *y(t)* corresponds to a target signal, $\hat{y}\ (t)$, to the estimated signal output of the model, and *L* corresponds to the total length of the signal in terms of number of samples.

Discussion and conclusions

Among the multiple available techniques, ANFIS, represents the most successful approach due to its ability to determine non-linear relationships between the forecasted values and the data that affect them. However, although the fuzzification and defuzzification procedures allow a better training than a classic neural network and, in consequence, a better performance, large amounts of training data limit its modelling capabilities.

However, an important consideration dealing with the forecasting of a time-series is the dynamics of the forecasted signal. In order to consider most of the dynamics contained in the target signal, a great deal of training data is used. As a consequence, the ANFIS based scheme suffers a loss of performance, generally leading the system to an overfitted response. Therefore, the major drawbacks of ANFIS is a relatively easy over adaptation of the structure to the training set, and slow convergence speed due to the input adaptation procedure.

Industrial time series modelling with ANFIS

The literature published to date shows that forecasting of manufacturing processes, in terms of industrial time series forecasting for supervision purposes, is still a novel field for research, in which performing methodologies are expected. Indeed, as has been previously aforesaid, it is necessary to remark that there are two main challenges related to industrial process forecasting. First, the consideration of suitable procedures to deal with highly non-linear signal behaviours. Second, the assessment and exploitation of such auxiliary information related with the target signal, which is required to enhance the forecasting performance avoiding computational complexity and model overfitting.

Auxiliary signal management with ANFIS

Several ANFIS approaches are proposed in the literature to deal with such problematics. In regard with the exploitation of auxiliary information, C. H. Su *et al.*, in [77], propose the use of an ANFIS, to predict the evolution of a non-linear time series. In such wok, a non-linear input selection method based on an adaptive expectation method is implemented to select the best suitable inputs for the ANFIS model. K. Kampouropoulos *et al.* in [17] propose the combination of ANFIS and Genetic Algorithm (GA) in order to find the best subset of inputs from a pool of available auxiliary signals, inputs are treated as chromosomes for the GA algorithm, best combination is given by optimizing the Root Mean Squared Error (RMSE) of the modelling. In this regard, is this combination of ANFIS modelling and GA algorithms the most common approach in literature to solve the problem of auxiliary signals management.

- **Genetic algorithm based optimization**

A Genetic Algorithm (GA) is a derivative-free and stochastic optimization method. Its orientation comes from ideas borrowed from the natural selection as well as evolutionary process. GA is used as a general purpose solution to demanding problems. It has the unique features of parallel search and global optimization. In addition, GA needs less prior information about the problems to be solved than the conventional optimization solutions. Note that generally, the optimization methods or search algorithms can be classified in two classes: deterministic methods including classical derivative-based algorithms, in which the derivative is numerically computed by finite differences, and more recently derivative free optimization algorithms. This second group based on stochastic methods, rely on random sampling to better explore the search space, and include recently introduced bio-inspired algorithms.

An important part of the initial GA implementation was carried out by Holland in [78], GA is inspired by the mechanism of natural selection, a biological process in which stronger individuals are likely be the winners in a competing environment. It presumes that the potential solution of a problem is an individual and can be represented by a set of parameters. These parameters are related to the genes of a chromosome and can be structured by a string of values in binary form. A positive value, generally known as fitness value, is used to reflect the degree of “goodness” of the chromosome for solving the problem, and this value is closely related to its objective value.

In a practical application of GA, a population pool of chromosomes is needed and they can be randomly set initially. The size of this population varies from one problem to the other. In each cycle of genetic operation, a subsequent generation is created from the chromosomes in the current population. This can only be successful if a group of those chromosomes, generally called “parents” or a collection term “mating pool,” are selected via a specific selection routine. The genes of the parents are to be mixed and recombined for the production of offspring in the next generation. It is expected that from this process of evolution (manipulation of genes), the “better” chromosome will create a larger number of offspring, and thus has a higher chance of surviving in the subsequent generation, emulating the survival-of-the-fittest mechanism in nature. A scheme called *roulette wheel selection* [79] is one of the most commonly used. Two fundamental functions are required to facilitate the GA evolution, *crossover* and *mutation*.

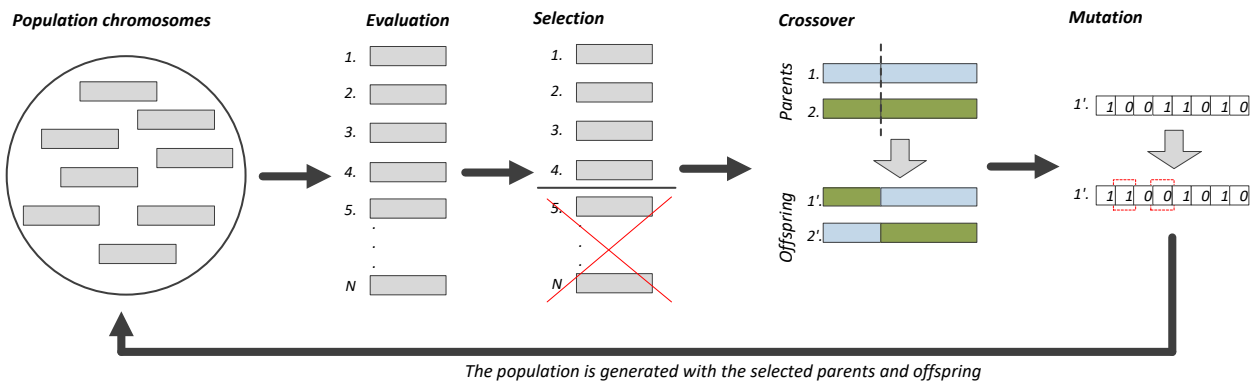

Fig. 1.5.9. Example of Genetic Algorithm operation: Evaluation, Selection, Crossover and Mutation stages.

As it is shown in the previous figure, during the crossover operation, a threshold is needed, this crossover point is probabilistically set. The portion of the two chromosomes beyond this cut-off point to the right is to be exchanged to form the offspring. For mutation, the process is applied to each offspring individually after the crossover exercise. It alters each bit with a small probability. The cycle evolution is repeated until a desired termination criterion is reached. This criterion can also be set by the number of evolution cycles (computational runs), the amount of variation of individuals between different generations, or a predefined value of fitness.

Since GA is generally an auxiliary optimization method, it is not applied independently in practice. The GA limitations rely on high computational cost and problems with search blocks in local maximums or minimums.

- **GA based ANFIS auxiliary inputs management**

A representation of the GA for ANFIS optimal input selection is shown in **Fig. 1.5.9.** This methodology performs as follows, a set of available auxiliary signals considered as potential inputs of the forecasting model is defined as *x(t,D),* where *D* indicates the dimensionality of the set, that is the number of available input. The target signal that wants to be modelled is defined as y(t). Then, the GA is initialized with a seed value, that corresponds normally to a random combination of inputs, this forms the provisional subset of inputs, *opt**, to be used in the design and test of the ANFIS model, *x(t,opt*)* together with the current time value of the target signal, *y(t)*. Then, a generic ANFIS model is developed using the specified combination of inputs to forecast the target signal among a defined forecasting horizon, *p,* obtaining *y(t+p)* as the output of the model.

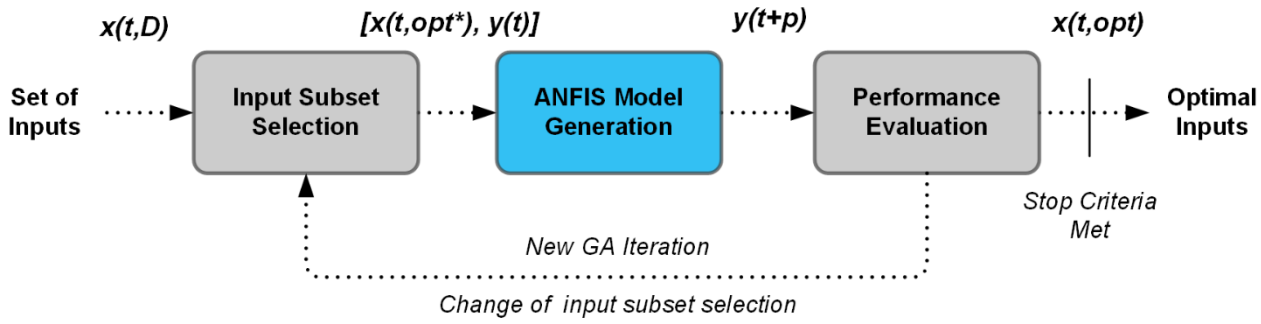

Fig. 1.5.10. Main Approach for ANFIS with GA for auxiliary signals input selection.

Then, the performance of the model is evaluated by using a standard error metric, normally the RMSE shown in Eq. 1.5.7 is used in most of approaches. This error corresponds to the objective function to be minimized by the GA optimization. Finally, the achieved error is compared towards a stop criterion, normally defined in terms of absolute achieved error or by algorithm configuration in terms of generations, if the stop criteria is met, the optimization algorithm stops giving with it the optimal combination of inputs. If the stop criteria are not met, the algorithm iterates and searches for another subset of inputs and repeats the procedure.

Discussion and conclusions

This approach gives good results for forecasting applications in which only some inputs are strong related with the behaviour of the target signal that wants to be modelled. However, it is based on the overfitting of the ANFIS model to the best combination of inputs found in the database, for this reason, the model is totally adapted to the best combination that causes the lowest error, but no other issues such as overfitting or generalization capabilities are considered in such approach. Furthermore, it corresponds to a blind optimization process that does not analyse the relation of the auxiliary signals with the target one. This aspect is critical when dealing with industrial time series, since the relation of the auxiliary signals can be correlated in different ways among the behaviours presented in the process. For this reason, an approach for managing the auxiliary signals as a fusion approach that consider relations among signals either than a selection one needs to be performed. Such approach is not properly defined in the state of the art, and will be developed in chapter 2 of this thesis.

Signal dynamics consideration in ANFIS

For the handling of highly non-linear signal behaviours the literature non-focused on a classical single model approach, propose to split the target signals in different details in regard with its dynamics content. Then, a dedicated ANFIS model for each set of frequencies is considered. Several authors state that decomposing the signal outperforms classical single-model approaches [80], [81]. Most of these approaches are based on Empirical Mode Decomposition (EMD) for extracting the dynamic modes. This is due to the fact that EMD offers an adaptive capability during the signal decomposition process, resulting in a collection of Intrinsic Mode Functions (IMF), and a residue. L. Y. Wei, in [82], propose the combination of EMD and ANFIS modelling. The author states that the combination of EMD and ANFIS improves the resulting performance of a time series forecasting in signals with high variability, that is, multiple oscillation modes. Dynamics based modelling will be further approached in Chapter 3 of this thesis. Indeed, the main objective of this section is to present the most used approach in the state of the art for considering dynamics in the modelling.

Therefore, previous to the explanation of the ANFIS and EMD approach, it is mandatory to define the theoretical concepts behind the empirical mode decomposition.

- **Empirical mode decomposition**

Empirical mode decomposition (EMD), was proposed by Huang et al. in 1998 [83], as a new method for analysing nonlinear and non-stationary signals. The EMD allows decomposing the full signal into a finite number of Intrinsic Mode Functions (IMFs), giving information concerning the intrinsic oscillation modes. The properties

of IMFs allow using them for forecasting purposes when dealing with complex signals with high dynamical content [84]. The procedure to extract the IMFs is illustrated in **Fig. 1.5.11**.

The IMFs must satisfy two conditions: First, the number of extrema (sum of maxima and minima) and the number of zero crossing differs only by one, and second, at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. If both these properties hold, the decomposition is complete, local, and adaptive. The IMFs represent oscillatory modes, but are much more general than harmonic functions. In fact, it is worth noting that the IMFs are modulated both in amplitude and in frequency and, consequently, are not restricted to be stationary.

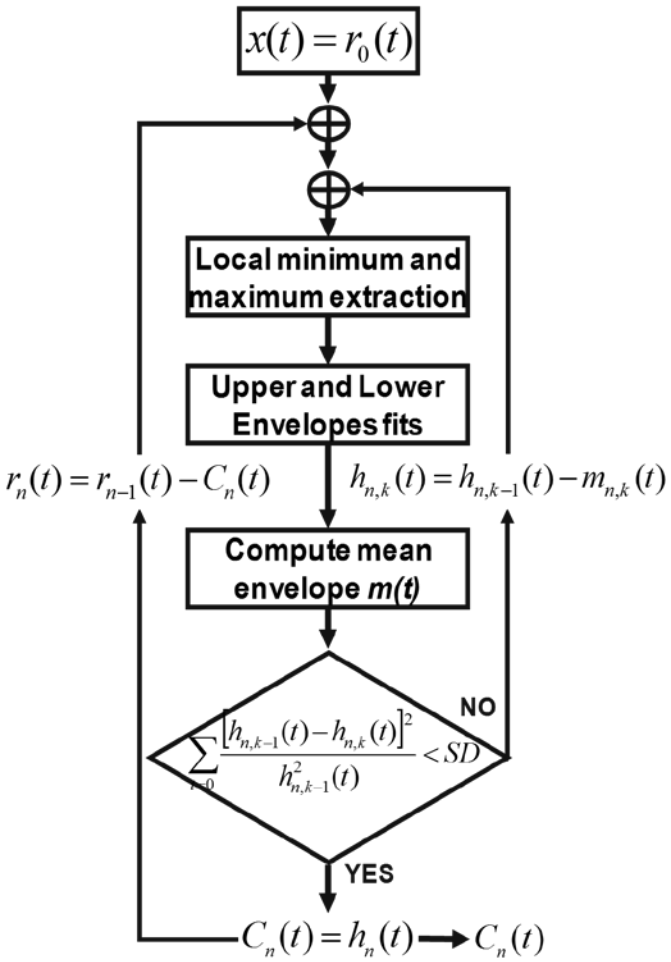

Fig. 1.5.11. Procedure to decompose a target signal in a set of IMFs.

The EMD method extracts, from a given signal, *x(t)*, the *M* number of IMFs by the so called sifting process described following: (i) For the first IMF (*k=1*), Find the upper envelope of *x(t)* as the cubic spline interpolated of its local maxima, and the lower envelope, as the cubic spline interpolated of its local minimum. (ii) Compute $\mu_k(t)$ as the average of the upper and lower envelopes. (iii) Compute the detail, $h_k(t)$ by subtracting $\mu_k(t)$ from the original signal $h_k(t) = x(t) - \mu_k(t)$.

Usually $h_k(t)$ is not an IMF and then the process has to be reiterated taking $h_k(t)$ in place of the original signal. In practice the process is usually reiterated up to *i* times, until the IMF properties are satisfied as indicated in Eq. 1.5.10, where $h_{k,i}(t)$ is the sifting result in the i_{th} iteration for calculus of the *k-th* IMF.

$$h_{k,i}(t) = h_{k,i-1}(t) - \mu_{k,i}(t) \qquad \textbf{Eq. 1.5.10}$$

Then, the stopping criterion for the sifting process can be defined by limiting the size of the standard deviation, SD, computed from the two consecutive sifting results, as indicated in Eq. 1.5.11. The value of SD is typically set between 0.2 and 0.3. Once the stopping criterion is achieved after *i* iterations, a IMF, $IMF_k(t)$, is obtained as $h_{k,i}(t)$. The next IMFs, corresponding to larger scales, will be computed in a similar way with the first one but taking as the starting signal the residue obtained by subtracting $IMF_k(t)$ from the original signal, $r_{k+1}(t) = x(t) - IMF_k(t)$.

$$SD = \sum_{t=0} \frac{\left[h_{k,i-1}(t) - h_{k,i}(t)\right]^2}{h_{k,i-1}^2(t)} \qquad \textbf{Eq. 1.5.11}$$

The obtained IMFs can be re-assembled in order to obtain the original data, as shown in Eq. (3), where $r_M(t)$, the last residue, is a constant or a monotonic function which represents the general trend of the decomposed time series.

$$x(t) = \sum_{k=1}^{M} IMF_k(t) + r_M(t) \qquad \textbf{Eq. 1.5.12}$$

- **ANFIS modelling based on signal decomposition**

In this regard, the main approach used to combine signal decomposition scheme and ANFIS modelling is shown in **Fig. 1.5.10**.

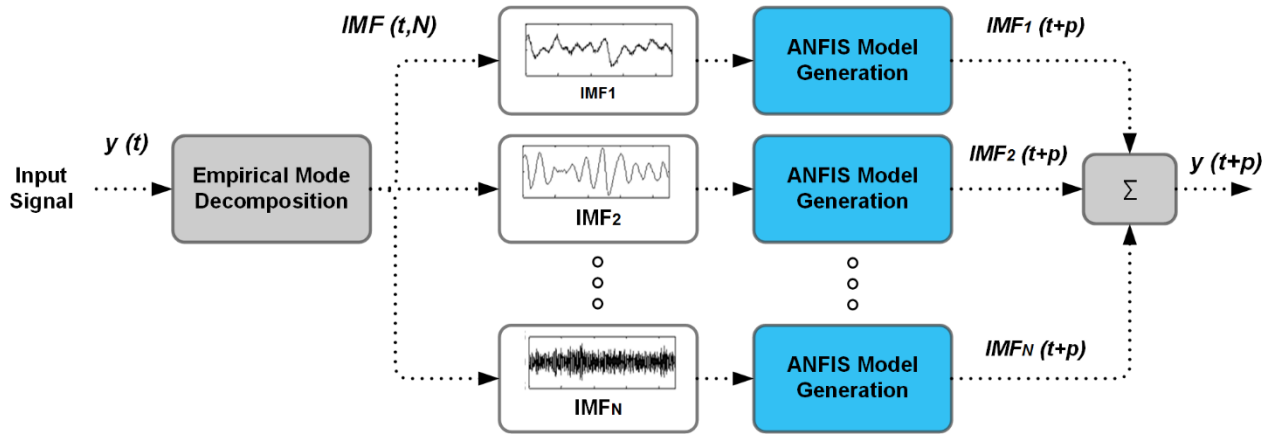

Fig. 1.5.12. Main approach for ANFIS based dynamics decomposition and consideration

This schema work as follows, first, a target signal y(t) is separated into a set of *N* IMFs by means of the classical EMD. Then, the resulting *N* IMFs, *IMF (t,N)*, are modelled with a dedicated ANFIS model in order to estimate the future behaviour of the IMF at a forecasting horizon p, IMF_N (t+p). These individual models use generally as inputs the current value of the IMF, and past values to get a measure of the signal tendencies. Finally, the different models are trained with their specifics datasets to provide the forecasted behaviour of all

IMFs. Then, due to the superposition properties of the IMFs, all model outputs are directly combined to give the final forecasting outcome, *y(t+p).*

Discussion and conclusions

This approach outperforms classical single model methods in terms of accuracy and adaptation over the used datasets. However, as the IMF are directly decomposed and modelled, that is, one forecasting model is required for each signal partitioning, which represents a high computational-burden strategy and, moreover, the simplicity of modelling some of the IMF can lead the corresponding model and, then, the global forecasting performance, to an intense overfitting. Therefore, the proposed thesis provides solid methodologies for overcoming such difficulties when dealing with rich dynamics signals, such approaches are discussed in Chapter 3 of this thesis.

1.5.4 Condition assessment

This section covers the state of the art in the condition assessment. In this regard, **condition or health assessment** in industrial process monitoring can be defined as the application of diagnosis methods to a system, in order to evaluate or classify its condition and detect any sign of failure or abnormal operation in it. Accordingly, diagnosis can be defined as the detection (an abnormal operating condition is detected and reported), isolation (determining which component, subsystem or system is failing or has failed), and identification (estimating the nature and extent of the fault) of an impending or incipient failure condition or a deviation, when the affected component (subsystem, system) is still operational even though at a degraded mode. Condition assessment is able to determine if the process is working properly or not by calculating, combining (reducing) and classifying a set of information extracted from the process. In this regard, the classical diagnosis schema is shown in **Fig. 1.5.13.**

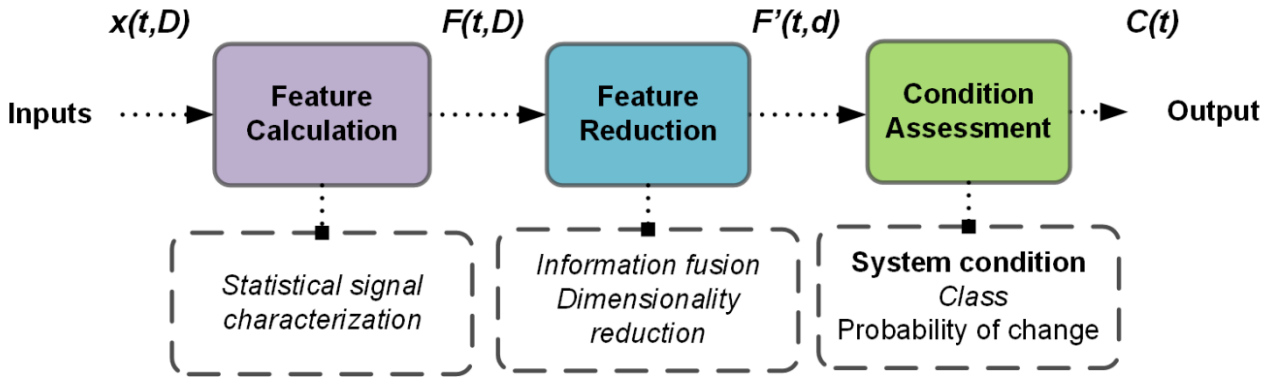

Fig. 1.5.13. Classical diagnosis schema that starts from a set of inputs, normally physical magnitudes acquired from the system, for then calculate statistical features to characterize those signals, combine those features in a proper reduction stage, and apply a classifier in order to assess the condition of the system.

In this regard, there are three critical blocks in condition assessment: (i) **Feature calculation**, that deals with the statistical characterization of the input data acquired. In this regard, regular metrics such as the RMS value, the max. value, min. value, crest factor, skewness or kurtosis have been used during the last decades to provide solid information from the analysed signals. (ii) **Feature reduction,** also known as dimensionality reduction, is the transformation of high-dimensional data, *D*, into a meaningful representation of reduced dimensionality, *d*, where *d*<<*D*. (iii) **Condition Assessment**, deals with the classification of the condition of the system assessed by the combination of the features involved.

Feature reduction approaches

The feature reduction approach attempts to extract some form of combination of all or a subset of the initial features set, in order to develop a reduced-size feature space representative of the initial data. Ideally, the reduced representation should have a dimensionality that corresponds to the intrinsic dimensionality of the data. **The intrinsic dimensionality of data is the minimum number of parameters needed to account for the observed properties of the data**. Therefore, **the aim of feature reduction is to find a reduced set of features containing the main descriptive relations between the considered condition system scenarios (classes) and the information in the database (features)**.

One of the drawbacks of this method is that the generated features have not understandable interpretation; the physical magnitude sense is missed during the reduction process. However, this kind of techniques is often referred as high-level feature extraction, feature fusion or feature compression due to its significant reduction capabilities.

The basic idea lies on the linear projection of high-dimensional data onto a lower dimensional space. Linear techniques result in each of the $d \leq D$ components of the new variable Y being a linear combination of the original variables X, as shown in equation Eq. 1.5.13.

$$Y_{dxN} = W_{dxD}\, X_{DxN} \qquad \textbf{Eq. 1.5.13}$$

where W_{dxD} is the linear transformation weighted matrix. The new variable Y is called the hidden or latent variables. Linear combinations are particularly attractive because they are simple to compute and are analytically tractable. There are different approaches to develop a linear feature extraction procedure, meanly, depending on the training objective such as Principal Component Analysis (minimizing reconstruction error) or Linear Discriminant Analysis (maximizing class reparability).

- **Principal component analysis**

Principal Component Analysis (PCA) is an unsupervised technique which represents the most used lineal feature extraction techniques [85]. The output due to the application of PCA over a feature set is a new set of features called principal components. Every principal component is a linear combination of the original features. There is not redundant information within *principal components* because all of them are orthogonal to each other in the feature space. In this sense, it retains most of the data-set information, that is, the variation present in the data-set, given by the correlation between the original variables. PCA identifies orthogonal directions of maximum covariance in the observations (maximum data dispersion direction), and projects these observations into a new dimensional space formed by the same number of components that features. However, this component can be ordered from the most significant to the less, which offer the possibility to establish a threshold value to obtain a final reduced set of only the highest covariance components.

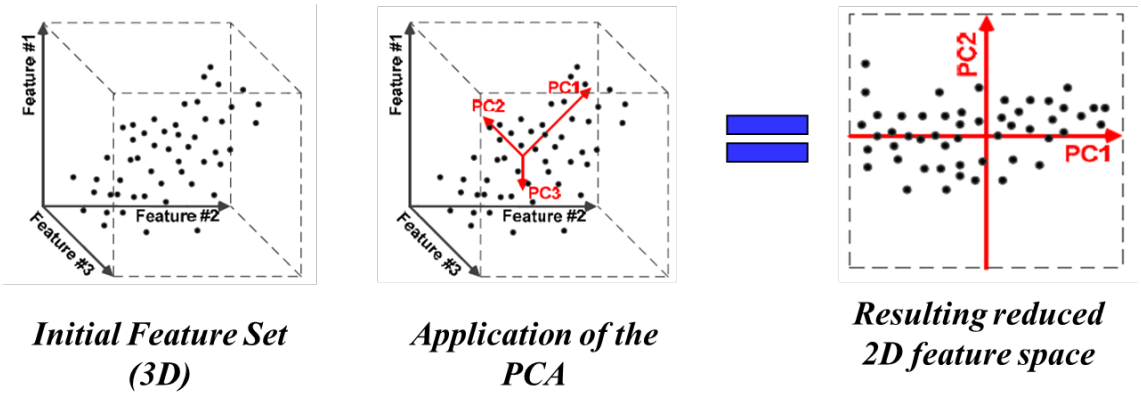

Fig. 1.5.14. Representation of a PCA process a) Initial 3-dimensional feature space b) Principal Components calculation (Application of the PCA) c) Representation of the observations in a reduced new feature space composed by two principal components.

In order to illustrate the PCA process, let x_w be a vector which minimize the sum of the squared distances between x_w and the rest of vectors in the data. Therefore, for each vector, the squared-error-criterion is calculated, as shown in Eq. 1.5.14. Following this idea, this first vector in **Fig. 1.5.14** corresponds to the arrow labelled as PC1, since it presents the lowest error in regard with the dataset.

$$J_0(x_0) = \sum_{i=1}^{N} \|x_w - x_i\|^2 \qquad \textbf{Eq. 1.5.14}$$

Therefore, the PCA method proceeds as follows: (i) First, the covariance matrix for the full set of observations is calculated in order to evaluate the relation between features. (ii) Second, diagonalization of covariance matrix and eigenvectors $x_{\lambda\text{-}i}$ and eigenvalues λ_i calculation is performed. (iii) Third, the eigenvectors are sorted in function of their eigenvalues. The eigenvectors which contributes less than a specified percentage in the data variation are removed from the final principal components set. (iii) The most significant eigenvectors are used as a projection destiny for the initial samples, in that sense, the projection of the initial samples over the *i*-th eigenvector is called *scores* $Y_i = x_{\lambda\text{-}i} * X_i$ and can be used as a reduced features set for a next classification stage.

Often there will be just a few large eigenvalues, which contain the information, and the remaining dimensions generally contain noise. Note that the intrinsic dimensionality is not the same as the linear dimensionality which is related to the number of significant eigenvalues of the scatter matrix of the data. This kind of techniques offer an important reduction effect, but the reduction process is based on the samples representation and, in this sense, there is a lack of attention between the reduction and the considered classes.

- **Linear discriminant analysis**

Linear Discriminant Analysis (LDA) is a supervised feature extraction technique that seeks to reduce dimensionality while preserving as much of the class discriminatory information as possible. Where PCA seeks directions that are efficient for representation, LDA seeks directions that are efficient for discrimination. In this sense, the PCA can be interpreted as a rotation of the original axes maintaining the orthogonally between them, however, in the application of discriminant analysis the resulting axes may not be orthogonal between them [86].

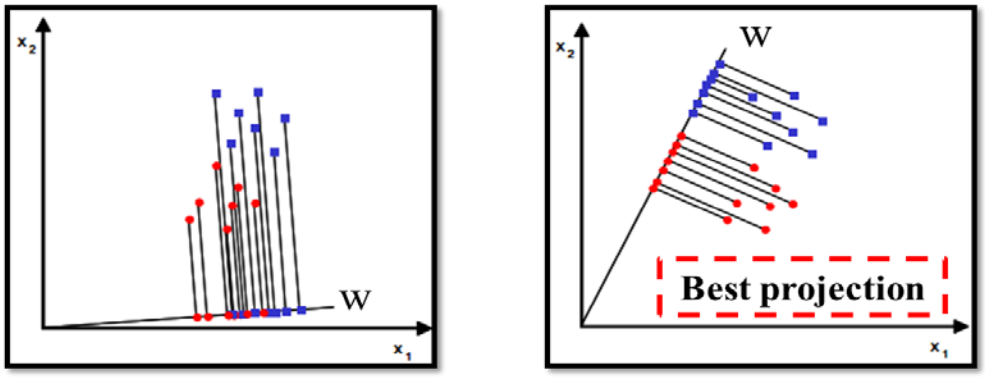

Fig. 1.5.15. Projection of the same set of samples onto two different lines in the directions marked as *w*. The right part of the figure shows a projection with greater separation between *blue* and *red* points.

The most extended example to explain the LDA operation is base considering the problem of projecting data from *D* dimensions in 1D as shown in **Fig. 1.5.15**. Of course, even if the samples formed well-separated, compact clusters in *D*-space, projection onto an arbitrary line will usually produce a confused mixture of samples from all of the classes and thus produce poor recognition performance. However, by moving the line around, we might be able to find an orientation for which the projected samples are well separated. This is exactly the goal of classical Linear Discriminant Analysis.

The criterion function for the best separation can be defined as shown in Eq. 1.5.15.

$$J(w) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{S}_1^2 + \tilde{S}_2^2} \qquad \textbf{Eq. 1.5.15}$$

Where $\tilde{m}_{c-i} = \sum_{y \in c_i} y \,/ n_{c-i}$ is the classes mean of the classes samples projection *(Y=w^T X),* taking *Y* as the projection of original points *X* into the defined line, and $\tilde{s}_i^2 = \sum_{y \in c_i} (y - \tilde{m}_{c-i})$ is the variance evaluation for the projected samples of each class *c*. This is called the **Fisher's linear discriminant** with the geometric interpretation that the best projection makes the difference between the means as large as possible relative to the variance. In other words, it searches the projection where the samples from the same class are very close to each other and, at the same time, very far from the other classes, this concept is graphically shown in **Fig. 1.5.16.**

Then, the evaluation of the discriminant coefficient is a useful tool to evaluate the disposition of the classes' samples in the feature space by the features calculation methods. Therefore, large values of discriminant coefficients imply good disposition of the samples in the feature space. The main drawback of this technique is the initial assumption, which takes into account that all the considered classes have the same variance, and this is usually not a real premise. Other important limitation during the application of this technique is in regard with the number of features in the original feature set, because this cannot be higher than the number of observations.

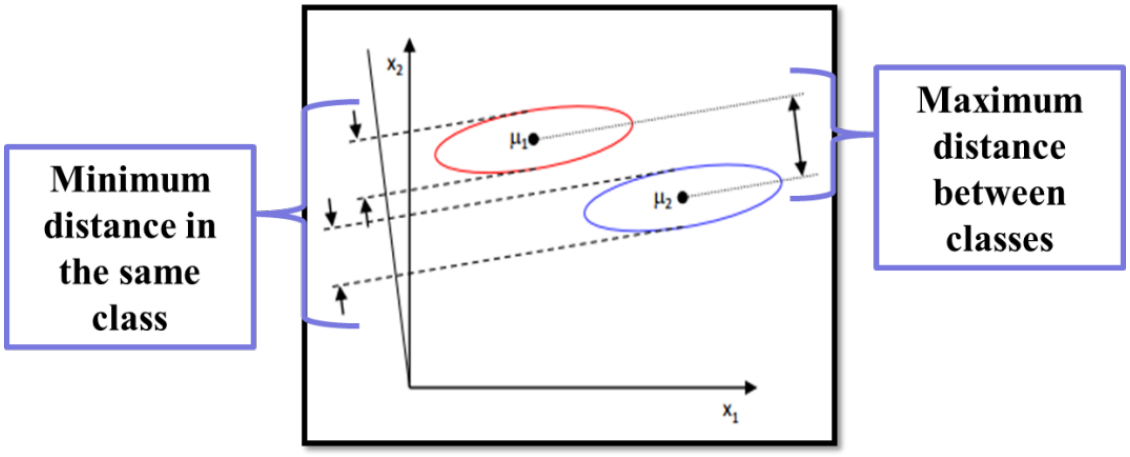

Fig. 1.5.16. Illustrated concept of the Fisher Linear Discriminant and its application in the LDA. µ1 correspond to samples of the class 1, and µ2 to class 2.

- **Self-organizing maps**

The Self-Organizing Map (SOM) learning rule, also called as Kohonen map, was proposed by Kohonen in 1990 to build topology preserving mappings [87]. The main important aspect is that SOM is based on neural networks. In this sense, SOM corresponds to a neural network grid trying to preserve the topological properties of the input space. The output space, is predefined as a regular d-dimensional grid. Generally, a two or three-dimensional grid is predefined and suitable for most of applications [88], [89].

In this regard, each point of the SOM grid represents a neuron or a Matching Unit (MU). For each neuron, MU_s, a D-dimensional weight vector $w_{v\text{-}j}$ is defined. The weights represent the coordinates of neurons position in the input space. Mapping is performed by assigning each data point, $x_{v\text{-}i}$, in the input space to one of these neurons, namely the one whose weight vector is closest to the point, which is called the Best Matching Unit (BMU). The position vector, $y_{v\text{-}i}$, in the output space is then given by the grid position of this neuron. The used error function corresponds as Eq. 1.5.16.

$$E_{SOM} = \sum_{j} \sum_{i \in s_{y-i}} \left(w_{v-j} - y_{v-i}\right)^2 \qquad \textbf{Eq. 1.5.16}$$

with $s_{y\text{-}i}$ being the set of data points which have neuron j as closest neuron. E_{SOM} expresses the average squared distance from a point to its representative. The minimization of E_{SOM} represents the objective of the operation, and is performed with respect to the weight vectors $w_{v\text{-}j}$. The classical gradient descent approach leads to the updating rule defined in Eq. 1.5.17.

$$w_{v-j}^{(t+1)} = w_{v-j}^{(t)} - \alpha^{(t)} \left(\nabla E_{SOM}^{(t)}\right)_{v-j} \qquad \textbf{Eq. 1.5.17}$$

However, this algorithm clusters the input space, but is not useful for mapping since the learning rate, α does not depend on the output space. For this reason, the learning rate is replaced by a neighbourhood function Nhf_{wn} which explicitly depends on the mapped space, as indicated in Eq. 1.5.18.

$$Nhf_{wn}^{(t)} = \begin{cases} \alpha^{(t)} \; if \; i \in N_{wn}^{(t)} \\ 0 \quad if \; i \notin N_{wn}^{(t)} \end{cases} \qquad \textbf{Eq. 1.5.18}$$

where $N_{wn}^{(t)}$ is the set of all neurons within a certain range of the BMU (the nearest element to the presented data point) in the output space. During training, this range and α(t) are decreased monotonically. Since neighbouring neurons in the output space will be neighbours in the input space, the mapping preserves the local topology.

Typically, the performance of SOM map is evaluated by average quantization error and topographic error in terms of the training data. The average quantization error means the average distance from each data vector to its BMU, and the topographic error means the percentage of data vectors for which the BMU and the second-BMU are not adjacent units.

In this regard, SOM does not follows the classical structure of the previous feature reduction method presented, that aims to extract a new feature in regard with some function applied to the input data. In this regard, **if the MU of the SOM are considered, it can be seen as a feature reduction method that aims to preserve the topological properties of the input space formed by the inputs or the signals into the latent space**. SOM will be used during the development of the proposed thesis to fuse the information regarding different auxiliary process signals and combine them in the BMU of the grid. Such approaches will be explained in section 2.3 of the proposed thesis.

Condition assessment

The aim of condition assessment is to recognize the hidden patterns in the data analysed in order to determine its belonging condition or class. In this regard, pattern recognition is the scientific discipline whose goal is the classification of objects into a number of categories or classes.

- **K-nearest neighbour**

In this regard, one of the most simply and used classification approaches is the *k*-Nearest Neighbour (*k*NN). This method leads to classify input data according to the most similar class. That is, the *k*-NN takes as a reference a set of points *X* in a *D*-dimensional space, where *D* implies the number of features describing each point in *X*.

An input data *y* is then evaluated by a distance function in order to analyse the nearest neighbours in *X*. The method always assigns a final class to the input, in order to include some control parameter, a limit distance can be configured, to ignore those samples extremely distant to the input. The class which collects a specified *k* number of nearest neighbours will be the assigned class for input *y* as it is represented in **Fig. 1.5.17**. The *k* value is completely up to user and it will affect directly the algorithm behaviour.

The use of large values of *k* has two main advantages: first, produce smoother decision regions, and second, provides probabilistic information (the ratio of samples for each class give information about the ambiguity of the decision). However, too large values of *k* are detrimental because, it destroys the locality of the estimation since farther examples are taken into account, and in addition, it increases the computational burden. Generally, *k* value between 2 and 10 gives proper results in the classification process, an example of this method is shown in **Fig. 1.5.17.**

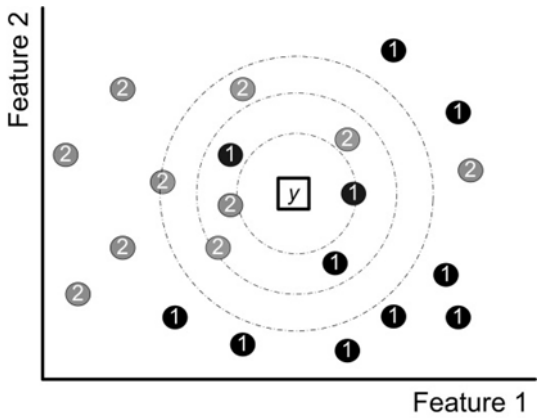

Fig. 1.5.17. k-NN representation between features vectors corresponding to two classes and an input features vector y. With a selected k value equal to 5, the y input is assigned to class 2.

- **Neural network**

Artificial neural network or commonly called Neural Network (NN) are data-driven self-adaptive information processing method inspired in biological systems, and represents the most commonly found data-driven technique in the diagnosis literature [31]. It is composed by a number of interconnected processing elements (neurons) working at the same time to solve a specific problem. The systems based in NN for pattern recognition are configured for a specific application, such as diagnosis, through a learning process. NN represents a no-lineal, multivariate and no-parametric algorithm.

The biological model of the neuron is shown in **Fig. 1.5.18.** The neural scheme can be described using the initial terms from the biological field: the neurons have a great deal of branches called dendrites which acts as a conduct to the electrical stimulation received from other neural cells to the cell nucleus or core. Is in this core, that different stimulus are treated and the neuron generates an electrical excitation which is sent through the axon to the synapsis, which are the link with other neurons and where the intensity of the transmitted excitation can be modified. Following this procedure, artificial neural networks were designed simulating the biological behaviour [90].

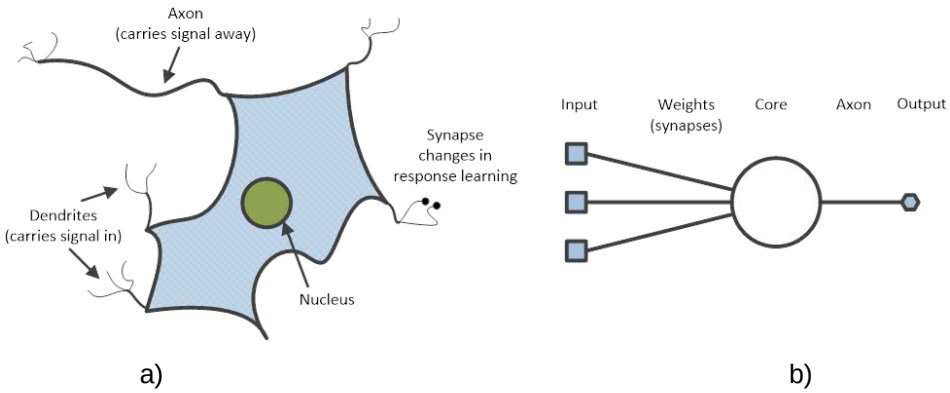

Fig. 1.5.18. Neural schemes. a) Biological neuron and b) Artificial neuron

Neural networks have been the most widely used artificial intelligence technique in condition monitoring and forecasting, due to their excellent pattern recognition capability and their ability to recognize fuzzy and imprecise signals, as it is reflected in a great deal of studies carried out in the last years [80-84]. The following characteristics make NN able to be used in many applications related to information fusion and forecasting:

- A neural network can be a Multi-Input and Multi-Output system. This structure demonstrates that neural network can manage complicated multiple object problems, like multiple failure conditions in a machine.
- Neural network processes the information in a parallel way which is similar to the way humans process complicated information. This special feature indicates that neural network can fuse information from different sources simultaneously.
- The knowledge in a trained neural network is stored in a distributed way, by means of a set of weights.
- Neural network has good fault tolerance performance. This property mainly originates from its parallel structure and distributed information storage system.

- Neural network has the ability to learn new knowledge. The learning process is implemented by the continuous adjusting of the weight values among the neurons. Hence, neural networks are well suited for practical problems where is easier to find CM than specific knowledge.

The basic processing unit in a neural network, the neuron, produces an output by forming a weighted sum of a number of inputs. The basic operation is that, if the summed value is higher than a threshold, then the neuron generates a pulse which is sent to neighbouring neurons. The McCulloch-Pitts neuron model published in 1943 [85], are the basic neuron presented, *perceptron*, and is assumed that incoming and outgoing signals may have only binary values 0 and 1. If incoming signals summed through positive or negative weights have a value larger than threshold *T*, then the neuron output is set to 1, otherwise it is set to 0. The neural model proposed by *McCulloch-Pitts* Neuron models have evolved in the use of different transfer function, $f(S_{ne-j})$, in the neuron core. In a general way: the output *y* of the neuron *j* is a function, generally non-linear, of the weighted sum of the inputs $x_v = \{fe1, fe2, \dots, feD\}$. The characteristic behaviour of the neuron can be described as shown in equation Eq. 1.5.17 and Eq. 1.5.18, where w_{ij} is the *i*-th weight of the *j*-th neuron.

$$y_j = f(S_{ne-j}) \qquad \textbf{Eq. 1.5.17}$$

$$S_{ne-j} = \sum_{i=1}^{D} w_{ji} * f_{e-i} \qquad \textbf{Eq. 1.5.18}$$

The diagram that represents the mathematical behaviour exposed in the formulation can be seen in **Fig. 1.5.19**.

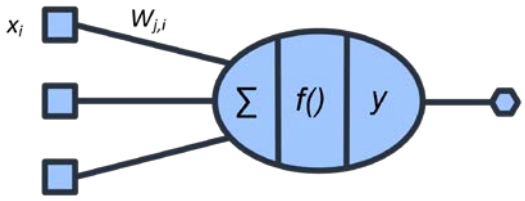

Fig. 1.5.19. Representation of the Neuron operators

The applied functions $f(S_{ne-j})$ depend of the NN configuration. A great deal of *transfer functions* can be found in the literature, however sigmoid functions are the most used, since they allow to give a fine probability in the output of the network.

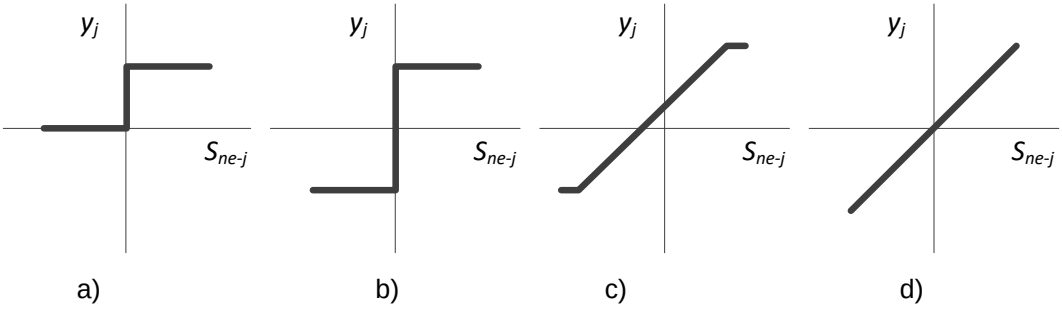

Fig. 1.5.20. Typical Neural transfer function. a) Hard threshold unipolar. b) Hard threshold bipolar. c) Saturated threshold bipolar. d) Pure threshold.

ANN are constructed by several neurons and these neurons are structured in different layers (as shown in **Fig. 1.5.21**), conceptually columns of neurons, generating a structure in which the neurons of consecutive layers are connected between them. Accordingly, all the neurons in a single layer have the same behaviour (activation function) and the same computational function inside the network. The neurons of a layer among them can be totally connected or non-connected. It should be noticed that the input layer does not perform any computation; it just redirects the input signals. The number of layers, and the way the neurons are connected conforms the architecture of the neural network.

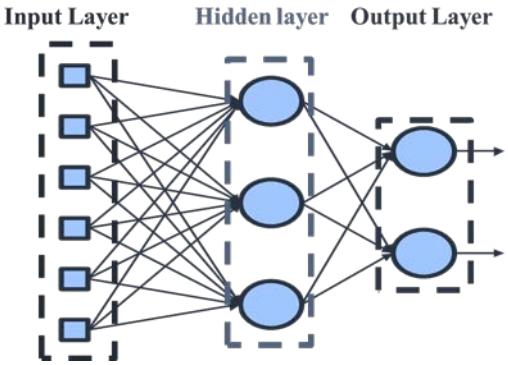

Fig. 1.5.21. A typical scheme of a three-layer perceptron which contains an Input-layer, a hidden layer and an output layer with (6 inputs, 2 outputs)

As a conclusion from the theoretical definition, the performance of a NN depends, basically, on the next decisions: (i) the transfer functions used in the hidden layers. (ii) The networks structure parameters such as number of layers, number of neurons and connection between neurons. (iii) The used learning process. The main drawback of the NN is that the relations between the inputs and the outputs are masked under the hidden layers of the network, so the decisions are not always evident.

Indeed, there exist many approaches in the literature in regard with the application of neural network based classifier for the assessment of the condition of a system, in which also SOM can be included. In this regard, SOM adjust the labels of the classes in regard with data density. That is, the samples with their respective labels are evaluated towards the MU of the SOM grid, then a label is assigned to each MU in regard with the majority of samples evaluated towards the specific neuron [91].

Discussion and conclusions

Feature reduction and classification methods have been studied during the last decades; however, there is a weak point with such methods regarding reliability in the classification outcome. That is, there is a necessity to include the estimation of the belonging degree of an observation to the assessed class. In this regard, the majority of the condition assessment approaches in the literature give a binary outcome in regard with the estimated class of a concrete observation. This scenario is acceptable for discrete condition detection, however is not the same case for applications in which the condition of the system needs to be extracted from a signal which presents a certain variation among time. Therefore, it is necessary that the classification approaches proposed by this thesis give an estimation of the reliability of the assessed class in regard with the knowledge of the observation.

Indeed, most of the classification approaches dedicate their modelling approaches in defining the best boundaries to separate the classes fixed by the labels of the samples. However, it is desirable or interesting to dedicate some part of the modelling efforts in the learning of the distribution of the data, in order to directly assess the associated boundaries from the supervised approach of the training. Such issues will be discussed in Chapter 4 of this thesis.

2.

Contributions to

Pre-processing for industrial forecasting models

The pre-processing is a critical stage previous to the development of a forecasting model. Dealing with industrial time series, the pre-processing is mandatory in order to properly design the modelling in terms of forecasting horizon, past values of signals and auxiliary signal management.

CONTENTS:

2. Pre-Processing for industrial forecasting models

2.1 Introduction

Prior to the generation of a forecasting model for predicting future behaviours of critical process's signals, it is mandatory to define the structure of the models in regard with the application. The procedure to define the structure of the models in regard with the target signal and the available information of the process is known as pre-processing stage.

In this regard, one of the most critical aspects to consider in this stage is the **forecasting horizon**. This parameter fixes the time window that the model should infer. The forecasting horizon is critical as directly conditions the model performance and feasibility. The forecasting horizon can be either fixed by the application, due to specific future step requirements [92], or optimized to reach the furthest horizon the model could give in regard with available information [93].

Most modelling approaches include information regarding **past inputs of the target** as auxiliary inputs. However, the best past inputs for a defined forecasting horizon must be selected, this procedure is faced primarily by using optimization methods (GA) [69]. In this respect, industrial time series present a characteristic non-periodic and complex behaviour that causes a quick drop of performance as horizon is extended. This is due to the fact that they are affected by the surrounding auxiliary signals and their own dynamics content. Hence, the pre-processing of such signals a combined scenario with two critical aspects: (i) the analysis of feasibility of the forecasting horizons and the associated past inputs, and (ii) auxiliary signals management, that is, the exploitations of the relations between auxiliary signals and target/s to increase modelling performance and generalization capabilities.

Regarding the management of auxiliary signals, further research has been made in the load forecasting field. In these applications, the signal to model presents lower dynamics and a more periodic approach than the industrial ones, but are also very influenced by the auxiliary signals of the process. Therefore, in this field, the efforts are concentrated on the selection of the best auxiliary inputs for a specific model to improve the performance. Z. Hu *et al.* [94], propose a hybrid filter-wrapper to select the best inputs for a Support Vector Machine (SVM). F. Keynia [95], presents a mutual information based feature selection method in which the most relevant features are ranked regarding their information addition to the forecasting process, the final modelling is faced by means of a NN composition. K. Kampouropoulos *et al.* [17], use a genetic algorithm based optimization in order to select the best inputs for an ANFIS based modelling. The inputs are treated as chromosomes for the GA algorithm, where the best combination is given by optimizing the Root Mean Squared Error (RMSE) of the modelling. The main issue behind using input selection methods for auxiliary data management is the difficulty from many modelling methods to handle a considerable number of inputs. Normally, model complexity and overfitting increases with the number of inputs, and also, the number of data required for a proper training increases [96]. So the model is optimized by limiting the number of inputs to the best set filtering signals that does not give additional information for a process, or auxiliary information that is not somehow related with the target [63].

By contrast, other data managing approaches aims to combine, rather than select, the information given by the auxiliary signals in order to give summarized information regarding the process. F. M. Bianchi *et al.* [97],

uses Principal Component Analysis (PCA) to compress auxiliary information and uses an Echo-State Network to model the evolution of the signal. C. Brighenti *et al.* [98], uses the topology preservation properties of Self-Organizing Maps (SOM) in order to analyse industrial process variations in the SOM clusters and its affectation to the target signal to be modelled. C.W. Frey [88], presents a SOM based schema with watershed transformations to detect different operating areas of an industrial automation network to monitor process changes and deviations. M. Dominguez *et al.* [99], uses also SOM based compression to monitor operating areas of an industrial process with the aim to detect abnormal behaviours. However, this information is used primarily as process monitoring, but is rarely exploited by the model to improve generalization in the forecasting.

As a conclusion, the literature in pre-processing for industrial forecasting models is limited to the direct configuration of the models by the utilization of an optimization algorithm that aims to select the best set of inputs in regard with a defined error criterion. Most of the time, this blind approaches causes the model to reach a local minimum during the convergence of the algorithm, since the response of the targets signal/s and the inputs is not analysed.

The contributions of this thesis research in the pre-processing for industrial forecasting models can be divided in two parts: (i) the first contribution consists on the proposal of input pre-processing methods to analyse the viability of the forecasting horizon, and to define the search intervals for the genetic algorithm for the past inputs. (ii) The second one, consists on the proposal of correlation analysis of auxiliary signals, together with SOM for the non-linear codification of the auxiliary information and their posterior consideration in the modelling method.

2.2 Contributions to optimal forecasting model design

Classically, a forecasting model for an industrial time series application presents a multi-input single output structure. In this regard, for a target signal, *y(t)*, the output of the model is directly the same signal predicted over a defined forecasting horizon, *p, y (t+p)*. However, the complexity of the modelling relies on the considered inputs for generating the model. Accordingly, the model may contain different inputs providing information regarding several aspects of the target signal itself, process information or either other relevant information to infer the future behaviour. In this regard, the main inputs used in industrial time series modelling are:

- The current value of the target signals, *y (t)*, to provide a solid seed value to construct the prediction.
- A *n* number of past values of the target signal delayed z_k samples as independent inputs, $y(t-z_1), \ldots y(t-z_k)$.
- Other information in regard with the target signal, such as tendencies, derivatives, mean values, etc. Used to extract simple dynamic information from the target signal that helps the model infer future behaviours.
- Auxiliary signals available in the process that affects in some measure the target signal that wants to be modelled.

Indeed, the selection of the forecasting horizon is a crucial step which affectation to the modelling must be studied. For this reason, it is necessary to emphasize that the result of a forecasting method is critically associated to the forecasting horizon used to train and validate the models. In many applications, the forecasting horizon is usually fixed by application requirements, or modelling limitations [100]. Yet, the procedure to select the best horizon for a given application is not well established in the literature, since even with a defined horizon, an analysis of the viability of such horizon must be performed, which is not the case in the literature.

Therefore, the behaviour of the signal to be forecasted must be analysed to select the most suitable horizon. **It is proposed as a first contribution to select the optimal horizon by analysing the performances decrease of the modelling method as the forecasting horizon is extended**.

2.2.1 Forecasting horizon selection based on auto-correlation analysis

The contribution approach includes a systematic analysis of the forecasting horizon affectation to the modelling performance, in order to establish a valid method for selecting the best forecasting horizon in regard with a proposed application. The configuration of the model is optimized one step beyond the current literature by analysing cross-correlations between a signal and past instants, which are selected by using GA optimization together with auto-correlation analysis, to find the best input configuration, shown in **Fig. 2.2.1.**

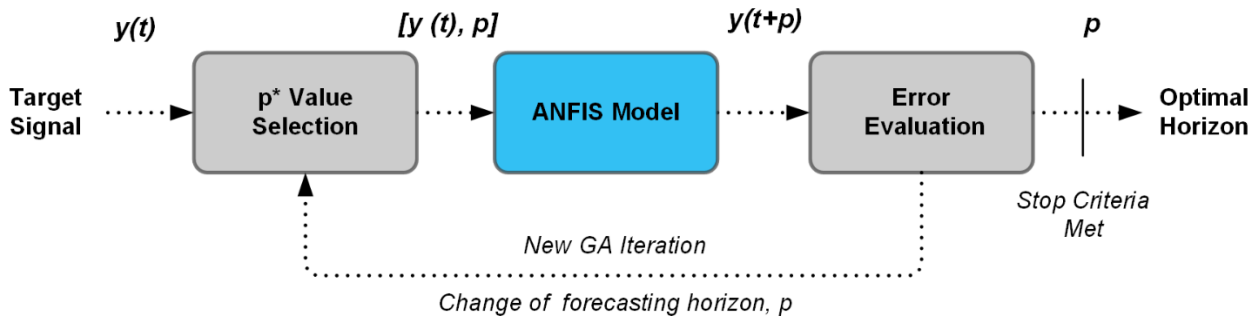

Fig. 2.2.1. Proposed method for selecting the optimal forecasting horizon p, of a given target signal.

In this regard, the analysis of the proposed methodology to find the most suitable forecasting horizon has been experimentally validated by means of the modelling of the vibrations in an electromechanical test bench. Such experimental set-up allows the analysis of the performances in comparison with similar approaches that can be found in the literature.

Vibration signals usually present a faster or at least equal dynamics in comparison with other magnitudes inherent to an electromechanical system, such as voltage, electric current and temperature. Due to this fact, the modelling problem should be approached as a time-series forecasting in which the forecasting horizon critically affects the performance of the models [101]. The global aim of the experimental study is to model and forecast the evolution curve of the vibration's RMS signal during the start-up and the thermal stabilization of an electromechanical actuator. The main point is to obtain a forecasting model of the RMS capable of giving the future value with enough resolution taking into consideration the dynamics of different failures occurring to the system. To do so, the target vibration signal is decomposed in three fixed frequency bands, Z_1, Z_2 and Z_3. The resulting RMS is calculated by a sliding window applied to the output of each filter. Classic ANFIS models are used to forecast the behaviour of the RMS in each filter output. Final forecasting is given by the direct combination of the three model outputs. After the output of the method, the presence of the failure can be anticipated by knowing the future value of the RMS and a simply statistical characterization as a diagnosis method.

The forecasting horizon analysis starts with the RMS signals calculated in the output of each filter as shown in **Fig. 2.2.2.** Further information regarding the calculation of the RMS and the raw data extracted from the test bench can be found in the Annex II of this thesis.

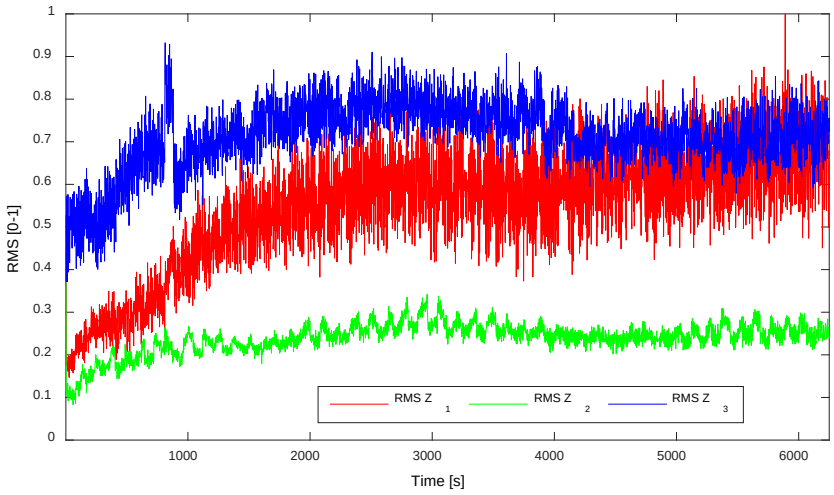

Fig. 2.2.2 RMS of the three frequency band that is intended to be modelled with a dedicated ANFIS model.

Due to the temporal characteristics of the application, the proposed forecasting horizon analysis interval comprises from *p=1* samples, 1 sec., to *p=600* samples, 600 seconds with an increase ratio of P_{inc}=5. Once the evaluation range is established, the RMS of each filter is evaluated towards a classic 1-input 1- output ANFIS model. The input of the model corresponds to the current value of the vibration RMS, and the output to the forecasted value of the RMS at a time horizon *p*. This simple model approach is used to isolate the error caused by the increase of the forecasting horizon, and for this reason no pas inputs are introduced.

The ANFIS models of each filter decomposition are trained and evaluated iteratively for all the specified horizons. The forecasted outcome of the three models is evaluated by classical performance metric, that is, RMSE. As a result, the performance of all operating conditions in regard with the metric used and the evaluated model is shown in **Fig. 2.2.3**. Nevertheless, the error of the model depends on the performance of the optimization algorithm used to select the best past inputs, for this reason, a smoothed response of the error has been generated in order to estimate the average expected error with the selected horizon.

The general results show that, as it was expected in time-series forecasting applications in which the forecasted signal does not present a periodic pattern, the lowest error can be found in the initial delays, in this case comprises p = [1-50]. Then, the error increases for all the models, especially the model of the first frequency band; that exhibits a linear behaviour by increasing gradually with the forecasting horizon. However, for the model of the second and third frequency bands, there is a second decay of error growth around *p=200*. This area also represents a low error region that can be utilized to develop the forecasting models, and thus, by defining the horizon *p=200,* the final forecasting can be optimized.

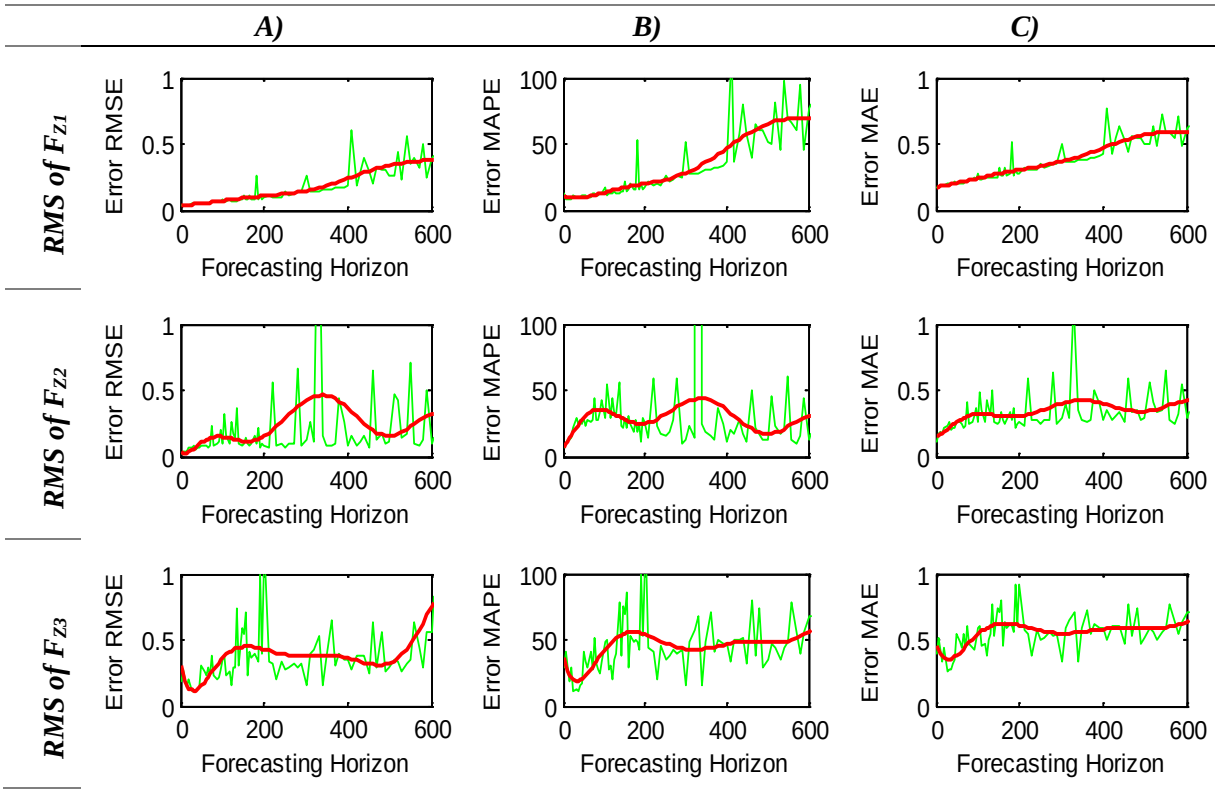

Fig. 2.2.3 Study of the affectation of the forecasting horizon to the forecasting modelling performance in a vibration monitoring based electromechanical test bench.

As specific conclusions, it should be noticed that low frequency dynamics are easier to be modelled independently of the selected forecasting horizon, since them represent the tendency of the vibration, and thus they suffer slower changes among time. It should be observed how error increases in the higher frequency bands, since fast dynamics are more punctual and closely related to adjacent samples of the current vibration value.

Indeed, the proposed horizon selection method is valid when the specific value should be selected. However, in many industrial applications, the forecasting horizon is fixes by the application requirements. Even though, due to the chaotic behaviour of industrial signals, it is required to analyse the feasibility of such horizon from the modelling point of view. In this regard, it is proposed to add the autocorrelation analysis to obtain a quantitative measure of the significance of the forecasting horizon regarding the target signal.

The resulting autocorrelation is obtained by **Eq. 2.2.1**, and represents a measure of statistical similitude between *y(t)* and its successive delays represented by *n*. The evolution of this value as *n* increases shows signal oscillation modes and periodicities in which the target signal and their delays are significant correlate. It should be noticed that in practice the forecasting horizon is only an abstract concept, since the model only sees the temporal relation between the current sample and the most advanced one. Due to this reason, the autocorrelation of a target signal *y(t)* with its successive delayed signals *y (t-n)*, can also be seen as the autocorrelation between the target signal *y(t)* the forecasted one among the horizon defined by *n*, *y(t+n)*. Therefore, the autocorrelation curve also gives the relation of the target signal similitude as long as the forecasting horizon, represented by the delays, is extended.

$$\mathrm{Crr}\ (\mathrm{n}) = \frac{cov\left(y(n), y\ (t+n)\right)}{\sigma\left(y\ (t)\right) \cdot \sigma\left(y(t+n)\right)}\ for\ n = 0 \ldots 10p \qquad \textbf{Eq. 2.2.1}$$

Therefore, an affordable horizon is selected by setting a correlation threshold, C_{th}, and checking if the value of the proposed horizon is below its value. It should be noticed that this coefficient corresponds to the average correlation which is a restrictive metric regarding the relation of both signals. Normally, the correlation threshold is set to be 40% for most of the industrial time series modelling applications.

Once the optimal forecasting horizon is selected or its suitability evaluated, the problem of model configuration derives to the selection of the best past values for the specific model. To do so, a generic GA based optimization algorithm has been used as in the literature. However, in order to increase the effectiveness of the optimization and reduce the associated computational cost in terms of search iterations, the contribution is to introduce a prior past index interval finding step.

Discussion and conclusions

The forecasting horizon is a critical aspect in modelling applications since it is a parameter with direct affectations in the performance of the model. In this regard, is mandatory to analyse the viability of such horizon if it is fixed by the application to detect unaffordable horizons. Furthermore, if an optimal horizon needs to be found, the thesis proposes a methodology to evaluate the optimal forecasting horizon of the model in regard with classical defined performance metrics.

2.2.2 Optimal range for past inputs selection

The autocorrelation analysis presented is useful also for the identification of the best past intervals to enhance the forecasting performance. Indeed, the optimization procedure to get the indexes for the past inputs gets simply as the search space is being bounded. The interval selection is made by calculating the correlation of the target signal by following Eq. 2.2.1 , *y (t)*, with successive delays of the same signal, *y (t-n),* where is fixed in the interval from 0 to 10 times the forecasting horizon considered $n \in [0, 10p]$. Then, the beast past input search interval is selected by selecting those intervals with the highest accumulated correlation over a defined threshold, C_{th}. It is common to find modelling situations; in which it is necessary to deal with different training sets or operating conditions. In these situations, the search of the optimal intervals turns into an iterative process that tries to localize those intervals with a consistent correlation, and once identified, select the maximum common range that accumulates the highest correlation.

Going back to the experimental study performed in the electro-mechanical test bench, **Fig. 2.2.4** shows the average correlation coefficient for the experimental set-up calculated with the healthy dataset with the RMS of each filter output. Due to the fast dynamics of the acquired vibration signal, the delayed signals of more than 20 acquisitions are considered to be non-significant for the application; therefore, the total number of delayed signals is set to 600. The amount of necessary correlation to consider a strong relation between the signal and a self-delayed is not defined. Thus, the experimental results show that useful intervals are considered with a correlation threshold above 40% (C_{th} >0.4). Additionally, in order to avoid finding empty intervals in which only few index are found, intervals with a length lower than 5 indexes are discarded. The interval that accumulates the highest C_{th} value is selected as the upper and lower values constrains of the optimization algorithm.

The correlation of the target signal in Z_1 shows a quick drop in the correlation coefficient in the first 20 delays, this behaviour is characteristic of non-linear forecasting applications, such as the study since the dependence of the target signal with the accumulative delays is high. The coefficient decrease continues until delay 200, in which there is a zone between till 400 seconds that the correlation increases moderately. This coefficient increase is due to a certain similarity that can be found in the signal. This interval can be exploded by the GA algorithm in order to find the best index for the model of the first band, Z_1. This fact confirms the previous selection of the forecasting horizon for Z_1 since the correlation increases in the same area that the selected horizon does.

The correlation of Z_2 presents a smoother decrease of the coefficient as the delay time increases. Moreover, it presents some oscillations as the delay increases; this resonance corresponds to signal periodicities that are characteristic of the dynamics of the signal. These periodicities can be used to enhance the forecasting capabilities of the model, since the past signals introduced are more correlated to the target signal. In this regard, the selected in the model of band Z_2 to exploit these periodicities is set I_{Z2} = [200 350] seconds. Note that another time, the periodicity also matches with the selected horizon extracted from the proposed method error analysis. Additionally, notice that the correlation in the second band is the highest in regard with the others. This happens because the fundamental harmonic of the operating speed is allocated primarily in the second band, and therefore, it presents a soft dynamic easier to be modelled.

Finally, Z_3 exhibits a lower correlation coefficient for the entire analysed interval. This behaviour is expected since it corresponds to the highest frequency band. In vibration forecasting, the high frequency band is closely related with noise while the system is working in proper condition, but it is modified with the apparition of incipient

failures. This behaviour near to the noise makes the signal chaotic and only dependent to adjacent past values; this causes the coefficient to drop quickly as the delay increases. Even tough, there is an upswing of the value in the interval of I_{Z2} = [200 350] that could be taken advantage of for the model of the band Z_3.

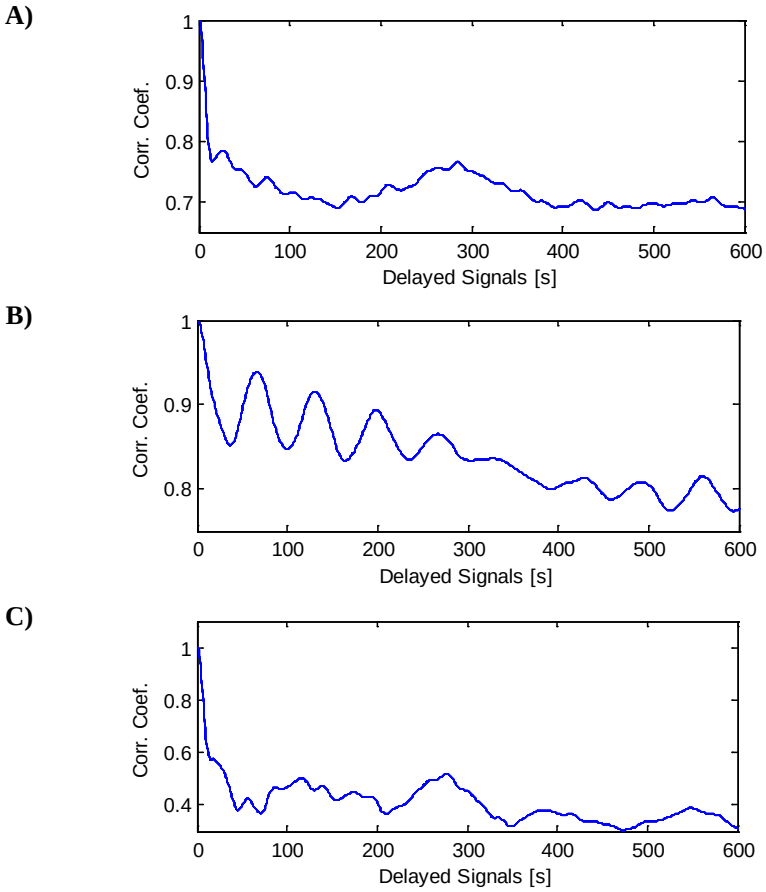

Fig. 2.2.4 Correlation coefficient for each target resulting RMS calculated at the output of each filter. A) Corresponds to the correlation in the first band, RMS_{B1}, B) to RMS_{Z2}, and C) to RMS_{Z3}.

Although each model in this experimental study presents the same quantity of inputs, the current value y(t) and two past inputs of the current signal $y\,(t\text{-}z_1)$ and $y\,(t\text{-}z_2)$, the delayed samples z_1 and z_2 could be different for each filter depending on the optimization results. The identified correlation intervals are introduced as boundaries for the Genetic Algorithm (GA). The cost function of the optimization procedure is defined as the Root Mean Squared Error (RMSE) of the forecasting model that can be seen in Eq. 1.5.7. The resulting past value indexes after 20 generations of the GA are shown in **Table 2.2.1**. Note that the indexes are referenced to the absolute delay; to get the relative indexes p value must be subtracted.

Table 2.2.1 Selected past index as a result from the constrained GA based optimization.

	z_1	z_2
RMS_{Z1}	217	389
RMS_{Z2}	285	397
RMS_{Z3}	247	348

Discussion and Conclusions

Both approaches cover the contributions to the optimal model configuration by means of the analysis of the target signal and the performance of the model in regard with the selected forecasting horizon. Therefore, this proposed analysis leads to an optimal configuration of the forecasting horizon, and thus the simplification of the configuration process during the selection of the past inputs while assuring the convergence to a reliable and robust forecasting.

2.3 Contributions to auxiliary process signals management

Due to the characteristics of industrial processes, there are auxiliary signals that are highly correlated with the target signal that wants to be modelled. Those signals have a great potential to be included in the modelling in order to increase adaptability and generalization. On the other hand, the classical feature selection stage used in the literature to identify the best set of auxiliary signals represents an overfitting approach usually neglected. The cost function related with feature selection techniques is commonly analysed in terms of forecasting performance, thus, the generalization capabilities of the original set of data is lost.

2.3.1 Auxiliary signals selection based on correlation analysis

Indeed, all the available auxiliary signal are introduced to the genetic algorithm without any analysis of the contribution or the correlation of such signal to the target that wants to be modelled. In this regard, the first contribution consists on proposing a correlation base auxiliary signal selection to overcome this limitation and assure that any signal introduced to the model presents consistent information.

This selection is made in regard with the diagram shown in **Fig. 2.3.1**, and is approached in two steps, first, the N_G auxiliary signals of each *G* processes, $x_G(t, NG)$ are analysed in terms of auto-correlation in order to remove from the analysis those signals that present redundant information within the process (high correlation). This is marked as Step 1 In the figure, and at the end, a subset of signals from the process are obtained. Second, the signals are analysed in terms of correlation with the target signal to be modelled, y(t). This second step is made in order to exclude those signals that are not related somehow with the target signal (present low correlation).

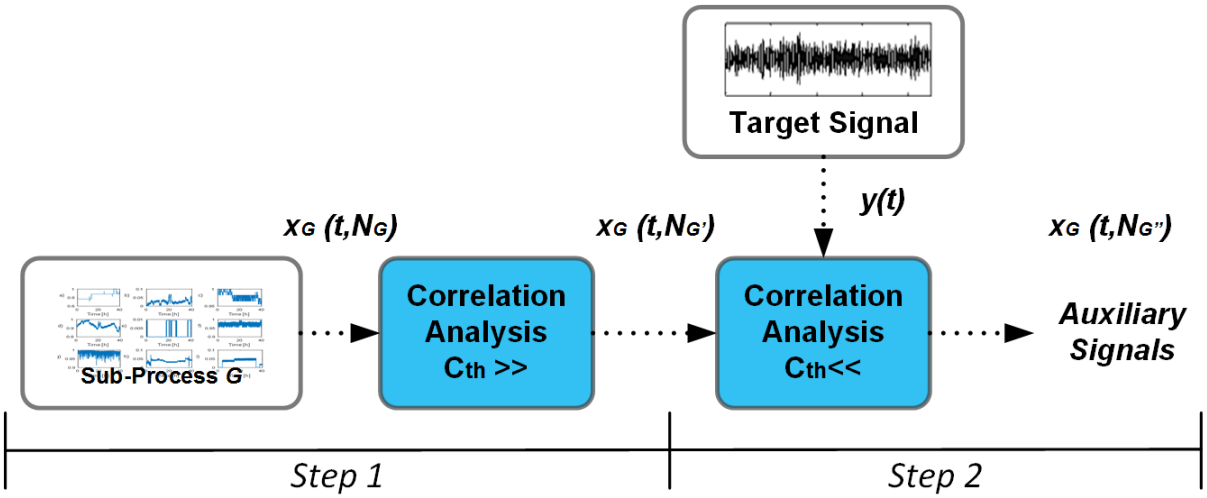

Fig. 2.3.1. Proposed approach for auxiliary signal selection based on correlation analysis.

In this regard, the correlation coefficient, C_{coef}, is proposed as a measure of the similarity between two signals x(t) and y(t). C_{coef} is calculated by means of **Eq. 2.3.1**, as a relation between the covariance matrix of two signals, *x(t)* and *y(t)*, divided by the product of their respective standard deviations. The point here is to exploit the useful information, excluding redundant information affects the performance of modelling in a negative way, since it unnecessarily increases the number of required inputs, slows convergence of training and causes a more overfitted approach.

$$C_{coef}(n) = \frac{cov\left(x(t), y(t)\right)}{\sigma\left(x(t)\right) \cdot \sigma\left(y(t)\right)} \qquad \textbf{Eq. 2.3.1}$$

The proposed auxiliary signals selection based on the correlation coefficient analysis is validated experimentally with the data from the industrial process in which this thesis is located, the copper Rod Manufacturing Process (CRMP). In this regard, the target signal and with it the objective of the proposed application corresponds to the refrigeration index of the casting wheel, $R_{ind}(t)$. This index is an indirect measure of the effectiveness of the heat extraction procedure among time. This magnitude is critical in the manufacturing process because represents the temporal heat extraction from the melted cooper during the casting procedure. Deviations in the refrigeration imply imperfections in the manufactured copper rod due to non-uniformities in the copper density. Therefore, this application is located in the heat extraction process of the CRMP, the location of the target signal and the auxiliary variables is shown in **Fig. 2.3.2**. The refrigeration index that wants to be modelled is shown in **Fig. 2.3.3**.

Auxiliary information in regard with the heat extraction process can be divided in two sub-processes, the water refrigeration, W_{sP1}, and the acetylene painting, A_{sP2}, both containing 9 auxiliary signals. In one hand, the matrix of the auxiliary signals from W_{sP1}, includes information regarding the casting wheel temperature, and measures of flow and pressure from the water cooling process in different parts of the wheel. The temporal waveform of W_{sP1} signals and their description can be seen in **Fig. 2.3.4**. On the other hand, the signals from the acetylene, the matrix A_{sP2} includes the temperature of the steel band, and measures of pressures and flows from the painting elements from different parts of the casting wheel. The temporal waveform of A_{sP2} signals and their description can be seen in **Fig. 2.3.5**.

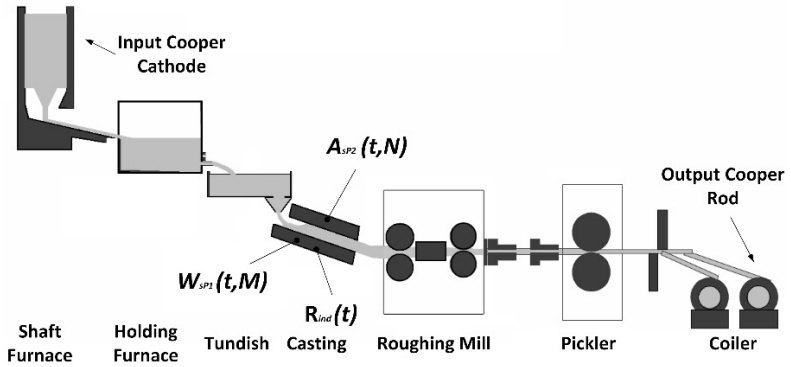

Fig. 2.3.2 Diagram of the main elements of the copper rod manufacturing process. Target signal: $R_{ind}(t)$ – Refrigeration index of the casting wheel measured in ºC. Sub-process: $W_{sP1}(t,M)$ – Water refrigeration process of the casting wheel. $A_{sP2}(t,N)$ – Acetylene panting process of the casting wheel.

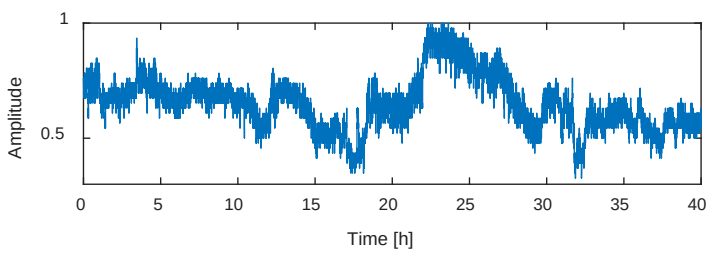

Fig. 2.3.3 Acquired refrigeration index, $R_{ind}(t)$, from the manufacturing plant in a periods of 40 hours of operation.

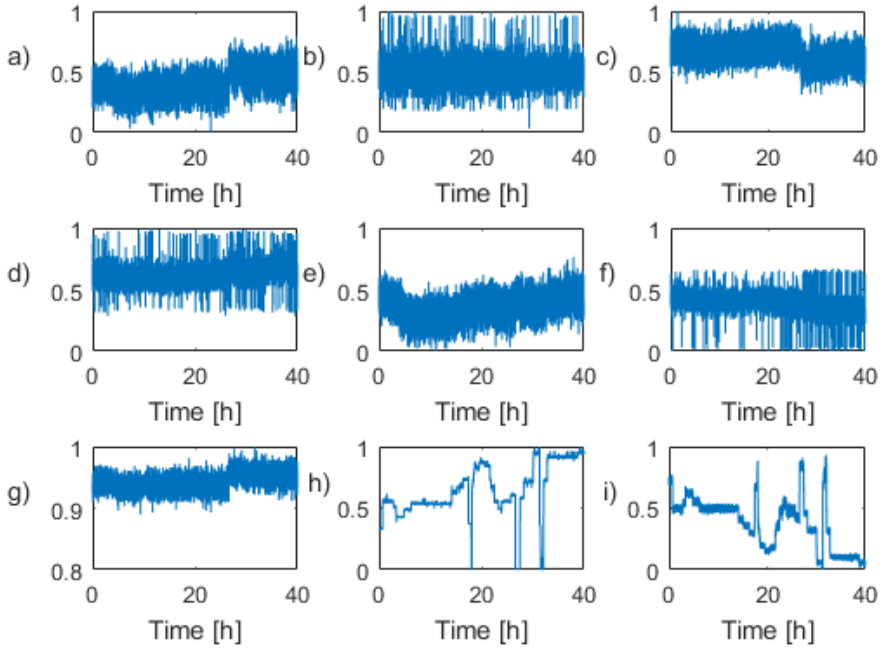

Fig. 2.3.4. Temporal form of signals from W_{sP1}. a) Temperature of the casting wheel [W_{sP1} (t, 1)], b) Water flow in the interior of the wheel [W_{sP1} (t, 2)], c) Water pressure in the interior of the wheel [W_{sP1} (t, 3)], d) Water flow in the exterior of the wheel [W_{sP1} (t, 4)], e) Water pressure in the exterior of the wheel [W_{sP1} (t, 5)], f) Water flow in the lateral of the wheel [W_{sP1} (t, 6)], g) Water pressure in the lateral of the wheel [W_{sP1} (t, 7)], h) Water flow in the center of the wheel [W_{sP1} (t, 8)], i) Water pressure in the center of the wheel [W_{sP1} (t, 9)].

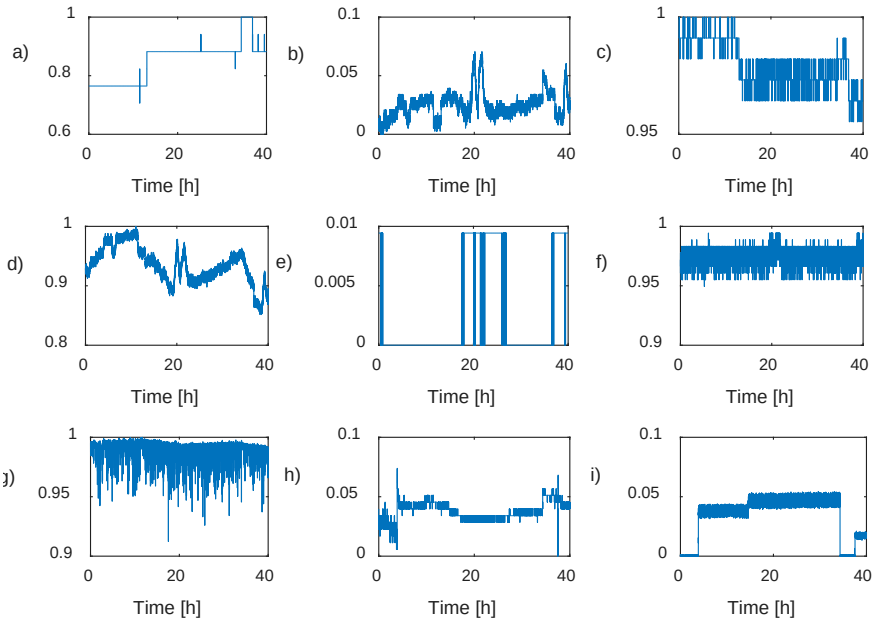

Fig. 2.3.5. Temporal form of signals from A_{sP2}. a) Temperature of the steel band [A_{sP2} (t, 1)]. Measures in the casting wheel: b) Acetylene flow in the exterior of the wheel [AsP2 (t, 2)], c) Acetylene pressure in the exterior [A_{sP2} (t, 3)], d) Acetylene flow in the interior [A_{sP2} (t, 4)], e) Acetylene pressure in the interior [A_{sP2} (t, 5)], f) Acetylene flow in the lateral of the wheel [A_{sP2} (t, 6)], g) Acetylene pressure in the lateral of the wheel [A_{sP2} (t, 7)], h) Acetylene flow in the center of the wheel [A_{sP2} (t, 8)]. Acetylene pressure in the center of the wheel [A_{sP2} (t, 9)].

The proposed input selection methods start with the calculation of the C_{coef} from all the auxiliary signals in each sub-process considered, W_{sP1} and A_{sP2} and all their combinations. Considering all the relations, the analysis is illustrated with a graphical correlation that represents cross relations among the auxiliary signals in the intensity of the colour that unifies the different signals. In this regard, **Fig. 2.3.6** shows the graphical

correlation of all signals from W_{sP1} and A_{sP2} respectively. For this graph, the stronger the colour of the line is, the higher positive (green) or negative (blue) correlation the signals present.

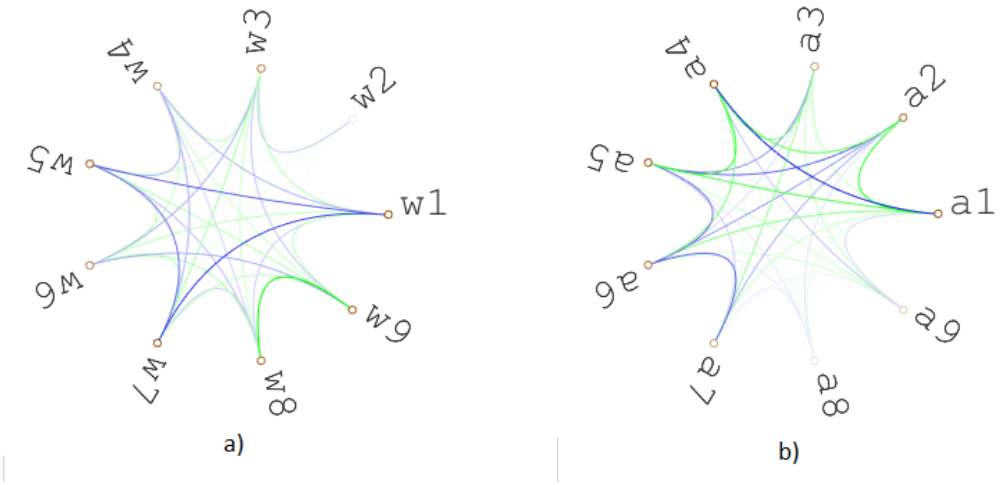

Fig. 2.3.6. Graphical representation of the correlation coefficient for a) Water refrigeration process, and b) Acetylene painting process.

Accordingly, if a correlation among the 50 % is considered that both signals give similar information, the signals W_{sP1} (t, 1), W_{sP1} (t, 5) W_{sP1} (t, 9), and A_{sP2} (t, 1), A_{sP2} (t, 5), A_{sP2} (t, 7), are removed for the pool of auxiliary signals to consider in the model. For the second step of the proposed analysis, previous selected signals are analysed towards to check their correlation with the target R_{ind} *(t)*. The resulting graphical correlation is shown in **Fig. 2.3.7**. Signals W_{sP1} (t, 2), W_{sP1} (t, 4), A_{sP2} (t, 3), A_{sP2} (t, 8), A_{sP2} (t, 9) do not show significant correlation with the target signal and are thus removed from the auxiliary input selection process.

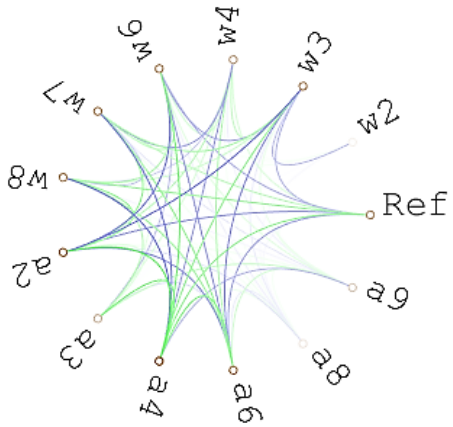

Fig. 2.3.7. Correlation representation of the selected signals versus the target signal Rind (t).

If an auxiliary signal from any sub-process achieves a C_{coef} lower than a predefined threshold of significant correlation, fixed at 10%, the signal is discarded from the analysis. the final pool of auxiliary signals that are correlated with the target and will be used as inputs of the model are listed below: W_{sP1} (t, 3), W_{sP1} (t, 6), W_{sP1} (t, 7), W_{sP1} (t, 8), A_{sP2} (t, 2), A_{sP2} (t, 4), A_{sP2} (t, 6). The scored C_{coef} of the signals is shown in **Table 2.3.1**.

Table 2.3.1 Ccoef of selected process signals versus the target Rind (t). Correlation is given as unitary value.

	$R_{ind}(t)$	W_{sP1} (t, 3)	W_{sP1} (t, 6)	W_{sP1} (t, 7)	W_{sP1} (t, 8)	A_{sP2} (t, 2)	A_{sP2} (t, 4)	A_{sP2} (t, 6)
$R_{ind}(t)$	**1**	0,23	0,24	-0,24	-0,36	0,34	-0,39	0,29
W_{sP1} (t, 3)	0,23	**1**	0,29	-0,31	-0,17	0,45	-0,39	0,36
W_{sP1} (t, 6)	0,24	0,29	**1**	-0,25	-0,29	0,43	-0,49	0,33
W_{sP1} (t, 7)	-0,24	-0,31	-0,25	**1**	0,18	-0,40	0,45	-0,39
W_{sP1} (t, 8)	-0,36	-0,17	-0,29	0,18	**1**	-0,41	0,48	-0,35
A_{sP2} (t, 2)	0,34	0,45	0,43	-0,40	-0,41	**1**	-0,61	0,38
A_{sP2} (t, 4)	-0,39	-0,39	-0,49	0,45	0,49	-0,61	**1**	-0,47
A_{sP2} (t, 6)	0,29	0,36	0,33	-0,39	-0,35	0,38	-0,47	**1**

Discussion and conclusions

The high correlation shown by these signals indicates the importance of considering such auxiliary information in industrial time series modelling, since the environmental variables present a strong influence in the target signal that wants to be modelled, and a change in these variables might help to anticipate a deviation of the process and thus improve modelling response and generalization. It should be remarked that auxiliary signal correlation analysis is important since it indicates the relation of process variables with the target signals to be modelled, if this approach is not controlled, classical input selection method could select two correlated signals that does not add any additional information of the process, and increase computational complexity of adding non-significant inputs to a model.

However, another conclusion extracted from this analysis is that industrial target signals are highly influenced by the surrounding auxiliary signals. As is the case of this experimental application, in which the target signal is influenced by seven auxiliary signals considered. Then, such strong influence should be considered during the modelling stage, and if all the signals should be included as inputs together with the classical information of the target signal, such as past values, tendencies, etc., the model will suffer a loss of performance by incurring in overfitting.

Summarizing, dealing with an industrial process time series forecasting, a set of auxiliary signals are complementary available with the target signal to be modelled, that all together define the process condition. These signals present a great potential to be informative enough to the modelling procedure. However, the direct introduction of such signals in the model algorithm may cause an unnecessary growth of the number of model inputs, increasing model complexity and the amount of data required to achieve a proper convergence during the learning phase. Indeed, forecasting models complexity increases with the number of inputs introduced. This is due to the fact that more weights and coefficients need to be estimated while the length of

training dataset is maintained. This problem is critical for example when dealing with ANFIS models, since modelling efforts are focused to adapt the membership functions to the input distributions [102]. For this reason, auxiliary inputs must be pre-processed to remove non-significant and redundant information and combined to summarize.

Consequently, it is mandatory the proposal of coherent auxiliary signal reduction methods in order to summarize the information regarding the surrounding processes behaviour to simplify as much as possible the structure of the model. This information will allow the model to improve generalization in order to overcome one of the limitations of modelling complexity, while assuring high ratios of accuracy and precision characteristic of such modelling approaches.

2.3.2 Auxiliary signals reduction by SOM mapping

In this way, the reduction process has been typically implemented with linear techniques such as Principal Component Analysis (PCA). However, PCA techniques has been discussed by many authors emphasizing its limitation dealing with large data sets, because it seeks for a global structure of the data. The information contained in a *D*-dimensional space mostly has a nonlinear structure. Concerning with this problems, manifold learning methods have been applied in the last years to preserve this information. Among them, Self-Organizing Maps (SOM) is the most used, which is based on developing a neural network grid to preserve most of the original distances between feature vectors representations in the input space [87]. The theoretical fundaments of SOM has been explained in Section 1.5.4.

Such space is initially predefined as a regular D-dimensional grid [89] then, SOM adapts this grid to data distribution defined by the auxiliary signals. In **Fig. 2.3.8**, the topology preservation mapping is illustrated. Prior to the training, the grid is defined, **Fig. 2.3.8 (a)**. During the training, **Fig. 2.3.8 (b)**, the grid successively adapts itself in order to preserve as much as possible the topology described by the original data. Finally, the resulting grid is evaluated over a new data, **Fig. 2.3.8 (c)**. Thus, for a new data point, the Euclidean distance, to each neuron in the *D*-space is calculated. The neuron with the shortest distance is considered to be the Best Matching Unit, BMU. All the coordinates of the point are mapped in the number of BMU, providing to the SOM the capability of input data codification. Indeed, SOM can be seen as a neural network that non-linearly discretizes the input data space and codifies such partition into the BMU number.

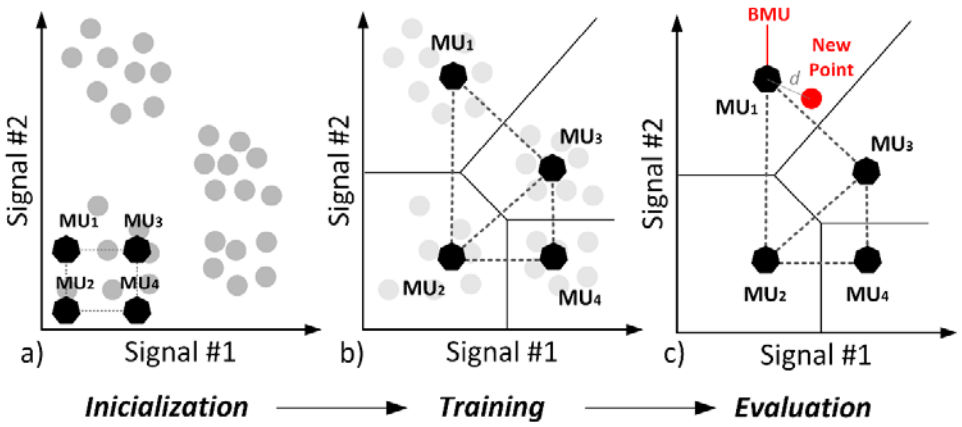

Fig. 2.3.8. SOM procedure to codify the input space.

Then, from the proposed method point of view, the available auxiliary signals are presented to the trained SOM, and a corresponding BMU is obtained. As seen in section 1.5.4, the BMU represents a discretized signal that summarizes process condition by the auxiliary signals. Thus, the SOM mapping codification is a two-step procedure. The first phase deals with the compression of the D-dimensional data space (usually D>2), into a d-dimensional grid formed by the SOM's neurons. Usually, for visualization purposes, d≤2. This step is achieved after the initialization and the training of the network with a training set. The training is the procedure to adapt the input SOM grid to the data space in order to maintain as much as possible the data topology, the information, by an adaptive analysis of the data distribution in the input space. The second step compresses the grid information into the BMU. After the training, the SOM network can be evaluated regarding an input in order to detect which is the associated BMU to this input.

Therefore, the proposed contribution relies on the use of the SOM as a topology preservation based auxiliary inputs mapping. The contribution is based on the premise that each BMU represents the mapping of a non-uniform region of the data space, and the behaviour of the process can be assessed by evaluating the corresponding BMU of a set of auxiliary inputs.

This contribution on the use of SOM is validated experimentally with the data from the heat extraction process as in the input signal selection commented before. Indeed, the following results start from the correlation based auxiliary signal selection explained before. Therefore, a dedicated SOM will be generated for each sub-process identified, the water cooling sub-process, W_{sP1}, and the acetylene painting process, A_{sP2}, that affects the target signal, the refrigeration index of the casting wheel. Then, SOM model is fed with a selection of the available auxiliary signals that are correlated with the target signal and the BMU is introduced as input of the model, that correspond to W_{sP1} (t, 3), W_{sP1} (t, 6), W_{sP1} (t, 7), W_{sP1} (t, 8) for the water process, and A_{sP2} (t, 2), A_{sP2} (t, 4), A_{sP2} (t, 6) for the acetylene one.

After useful information has been selected, the resulting reduced auxiliary signal inputs from each sub-process will be introduced in a SOM for mapping the operating point of each sub-process. This dedicated SOM will adapt the internal grid to the topology of this input data space. Therefore, the SOM grid has been configured with a 2D hexagonal connection grid with a size of [15x15], which fixes a total amount of 225 neurons to cover the whole space. Prior to the training, the input signals have been normalized in order to have mean zero and standard deviation equal to 1. Then, the SOM is initialized and trained with a batch algorithm and a total amount of 100 epochs were performed.

Typically, the performance of SOM map is evaluated by average quantization error, E_{qe}, and topographic error, E_{top}, in terms of the training data. The average quantization error means the average distance from each data vector to its BMU, and the topographic error means the percentage of data vectors for which the BMU and the second-BMU are not adjacent units. For this training set, the average quantization error is equal 1.081 to and the topographic error is 0.045 for W_{sP1}, and 0.977 for quantization error and 0.086 of topographic error for A_{sP2}. Theoretically, any process variation should be reflected in a certain region of the input data space covered by the SOM neurons. The resulting SOM mapping can be seen by means of the U- Matrix in **Fig. 2.3.9a** for W_{sP1} and **Fig. 2.3.9c** for A_{sP2}. The U-Matrix represents the distances between the neurons of the trained SOM. The dark areas represent regions of the map in which data is concentrated (low distance between neurons), and the light areas represent the frontiers of the clusters (high distances between neurons). In the figure, the dark area in the center correspond to the normal operation point of the process. Indeed, data in the light area

correspond to those points that are far from the center and are identified as punctual deviations from the nominal conditions. It can be appreciated in **Fig. 2.3.9a** how the water cooling process present a more disperse behaviour since there are not high density areas presented among the data, however, it can be appreciated a consistent operating point in the middle of the U-matrix. This fact is also reflected in the BMU, **Fig. 2.3.9b**, achieved in the number of jumps over different number of neurons. In the opposite, the acetylene painting process presents a more stable behaviour since five big operating points can be identified in the u-matrix of **Fig. 2.3.9c**, this results in a more stable behaviour of the obtained BMU as can be seen in **Fig. 2.3.9d**.

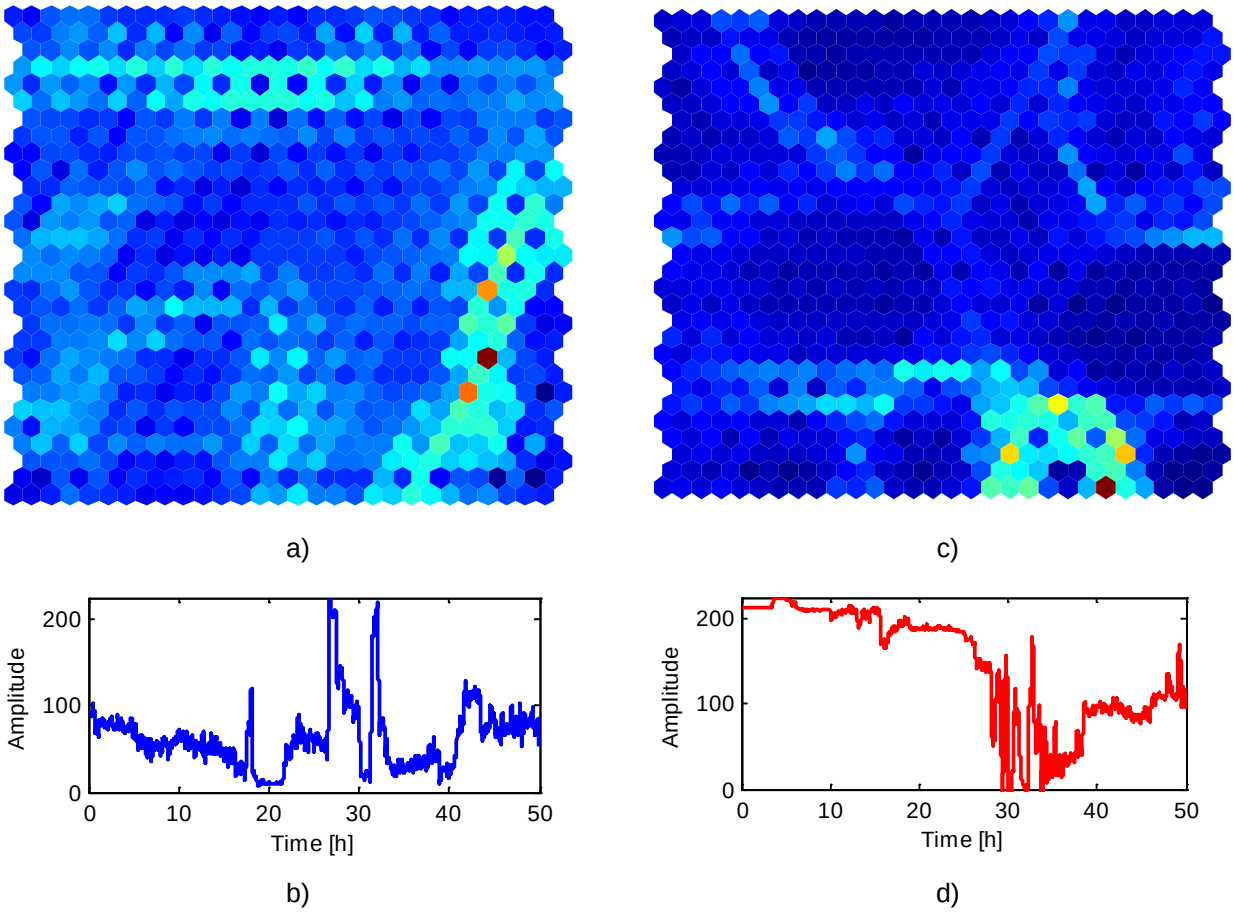

Fig. 2.3.9. Results from the SOM of applied to the water cooling sub-process W_{sP1}. a) U-matrix representing the inter-distances between neurons, b) BMU extracted with the temporal signal, BMU_{WsP1} (t). Results from the SOM to acetylene painting sub-process A_{sP2}. c) U-matrix representing the inter-distances between neurons, d) BMU extracted with the temporal signal, BMU_{AsP2} (t).

It has been seen how the SOM is able to capture each sub-process behaviour and compress this information in the BMU indicator. Furthermore, it has been seen how each sub-process may present a certain dynamic and might be more or less correlated with the target signal to be forecasted. Therefore, from the cooper manufacturing point of view, that the water cooling procedure presents a less stable behaviour due to the variations in the cooper temperature and the response of the control loops of the factory.

Discussion and conclusions

As a conclusion, the operating of each sub-process is reflected in the BMU provided by the SOM. In this regard, the auxiliary signals at every time step are summarized into the BMU and all the variability in the process behaviour is reflected as changes in the BMUs. Indeed, it should be noticed that the evaluated BMU corresponds to a temporal signal that summarizes the operating point of the process, and not to a static measure of the process. This means that that the BMU evolves as the auxiliary signals do, and therefore, presents a characteristic dynamic related to the process behaviour.

The methodology presented the suitability of using the SOM for compressing the information while maintaining the topographical properties of the auxiliary signals. In this regard, the contributions in this topic of using SOM based compression to extract the operating conditions from a set of auxiliary signals that conforms a sub-process helps the model to improve generalization in order to overcome one of the limitations of the modelling, while assuring high ratios of accuracy and precision characteristic of such modelling approaches.

2.3.3 Impact of auxiliary signal mapping in forecasting

A second study has been performed in order to assess the impact of the proposed SOM based mapping of the auxiliary signals in a classic approach of ANFIS based modelling. In this regard, the originality of this experimental study is to propose a complete industrial target signal forecasting methodology applied to the cooper manufacturing plant. The method uses the explained SOM based auxiliary signals mapping, to compress the information given by the most significant signals related with the temperature of the liquid copper in the tundish. Next, the implementation of the data-driven model is done by means an ANFIS based approach, which takes advantage BMUs of the SOM. In this work, the synergy between ANFIS models and self-organizing maps is exploited in order to increase the forecasting performance while assuring generalization with the validation set. This work represents an important step to the introduction of manifold learning techniques to the development of industrial time series forecasting.

The aim of the proposed forecasting scheme is to improve the auxiliary signals management in a data forecasting scheme, in order to increase the performance and generalization capabilities of the resulting forecasting models. This point is critical in those applications in which the time series to be forecasted are highly correlated by multiple measured information and, the pre-analysis of the information is not obvious, which lead often to non-proper data reduction strategies. The procedure of the proposed method is shown in **Fig. 2.3.10**.

First, the method splits the variables in the database in two different groups, the first one concern with the target signal that wants to be forecasted, the second group is formed by the auxiliary signals that are available in the process and are correlated somehow with the target signal. The signals contained in the second group will be considered as auxiliary inputs. Then, the signals are pre-processed and filtered in order to remove outliers, reduce electrical noise and detect abnormal behaviours.

Once the data is processed, the signal that wants to be predicted should be excluded from the set of inputs to the SOM network. This is important hence the function of the SOM is to map the input data space of the

auxiliary inputs in order to detect operation points that help to describe or anticipate the changes of the process, which are reflected in the forecasted signal and not to over-fit the model with information regarding the predicted signal.

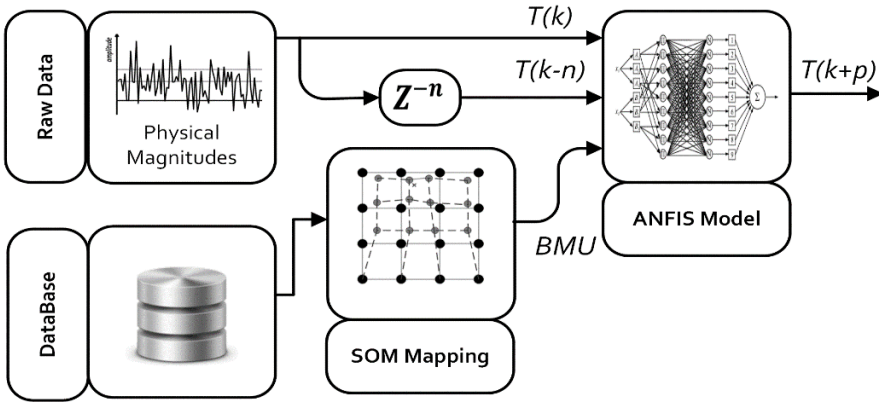

Fig. 2.3.10. Critical industrial signal forecasting method by means of ANFIS modelling and SOM Mapping

The ANFIS model for a target signal, *y*, has a single output, which is the predicted signal at a time horizon *p*, *y (k+p).* This time horizon should be configured depending on the application, the dynamics of the signal and the forecasting requirements.

The considered inputs of the model are of three types, *i.e. (i) T*he current value of the signal to be forecasted, *y (k). (ii)* A past values of the signal in order to have a reference of its tendency, *y (k-n)*. It should be noticed that the selection of the delay *n*, could affect the performance of the final model. For this reason, it is recommended to select this parameter by means of optimization algorithms like GA. (iii) The BMU of the SOM grid for the current value of the signal *y(k).* As a result, an ANFIS model will be generated, trained and validated.

Therefore, the aim of the experimental validation is to test the suitability of the method when dealing with the prediction of a variable, which affects the correct functioning of the process and the final manufactured product's quality. However, there are several factors and physical magnitudes that affects the cooper during the melting process and furthermore, the relations among these variables and its affectation is difficult to be known. For this reason, the scope of this scientific work is limited to the prediction of one single variable which is considered for the factory staff as a critical for the process and which prediction allows the quick correction of any deviations before it affects the quality of the final manufactured product.

The objective of this application is to forecast the tundish temperature, T_{tu} *(t)*, which is related to another process considered that presents less auxiliary signals than the previous exposed process of heat extraction. This is done in order to isolate the effects of two different sub-process and assess which is the gain of performance and generalization of including the mapped auxiliary signals in a classic modelling scheme. This variable is important in the manufacturing process hence is the last reference of the melted cooper before it enters to the casting wheel. Any deviation in this variable could cause imperfections in the output cooper, and should be corrected before it affects the final product.

The considered auxiliary signals that directly affects the tundish temperature are: (ii) Strainer's oxygen, O_{tu}. (ii) Ratio Gas/Air in the strainer's burner, R_{tu}. (iv) Cooper weight in the strainer, W_{tu}. Two data sets with 50 h of plant operation are used in this work, the first will be used to train both ANFIS model and SOM algorithm; the second set to test and validate the performance of the method against new data with a similar behaviour.

Both signals can be seen in **Fig. 2.3.11**. It should be noticed that these kind of industrial processes are influenced by several external factors such as the company production ratios, the temperature, the quality of the input material, etc. For this reason, the two data sets correspond to an operation time in two consecutive weeks. Despite this, it can be appreciated in the figure how the validation set presents a faster dynamic and a disturbance in the 20th hour of operation. These differences are important hence they will allow to test the generalization capabilities of the propose method.

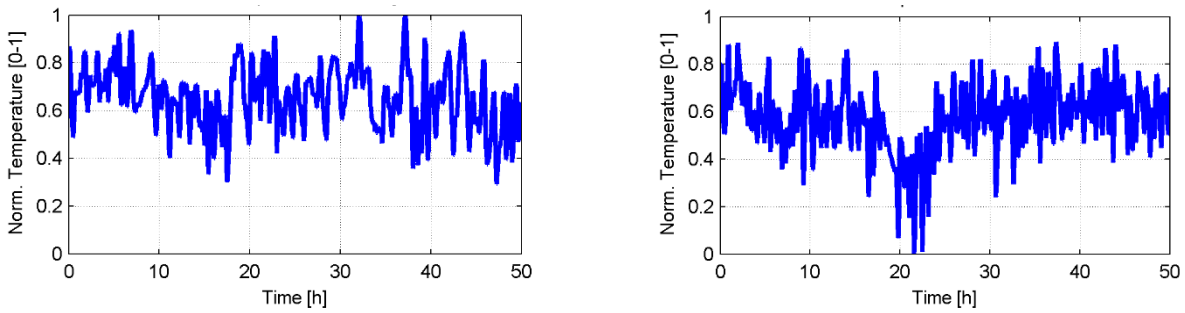

Fig. 2.3.11. Normalized waveform of the strainer temperature from [0 to 1] in the considered a) training set, and b) validation set.

The input data space, formed by the three considered auxiliary signals, to be mapped by the SOM is shown in **Fig. 2.3.12 (a).** Taking into consideration this data, the technique will adapt the initial grid to the topology of this input data space. Therefore, the SOM grid has been configured with a hexagonal connection and size of [15x15], which fixes a total amount of 225 neurons to cover the whole space. Prior to the training, the input signals have been normalized in order to have mean zero and standard deviation equal to 1. Then, the SOM is initialized and trained with a batch algorithm and a total amount of 100 epochs were performed. After the training, the updated SOM grid in the input space could be seen in **Fig. 2.3.12 (b).**

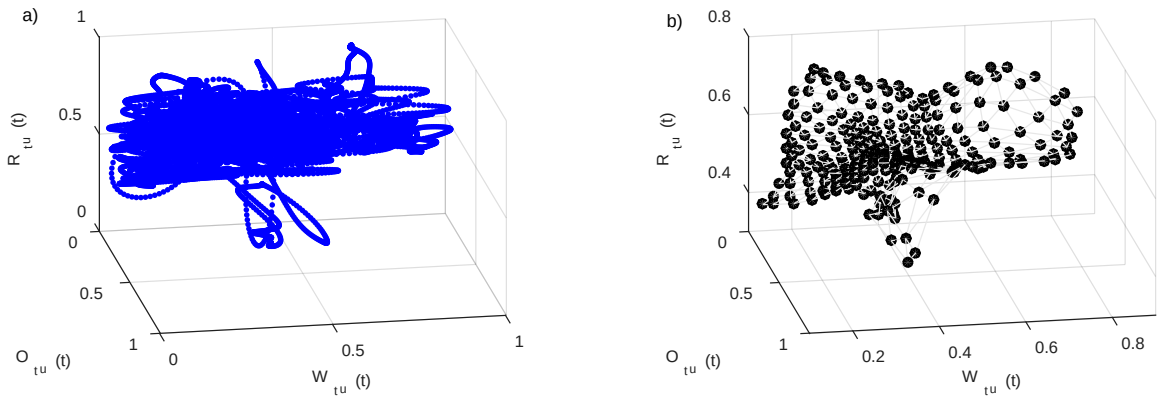

Fig. 2.3.12 Data and SOM grin in the input space. a) Input data space made by the variables: x) Strainer Oxygen, y) Strainer burner's ratio and z) Cooper weight in the strainer. b) Detail of how the SOM grid is covering the input data space after the training procedure of 100 epochs. x) Strainer Oxygen, y) Strainer burner's ratio and z) Cooper weight in the strainer.

As it can be seen, the data distribution process presents a central area with high density of data which corresponds to the main operating condition. However, there is a dispersion in each axis around the center, which means deviations from the nominal values reflected in the auxiliary signals.

Performance of SOM map is also evaluated by average quantization error and topographic error in terms of the training data. For this training set, the average quantization error is equal to 0.496 and the topographic error is 0.067. The achieved E_{qe} value means that most of the neurons are concentrated covering the central cluster and the distance between the outlier points and the other neurons outside the cluster is big. In the other hand, the low value of the E_{top} indicates that the initial topography is well conserved by the trained SOM, indicating with it the suitability of the projection.

The resulting SOM mapping can be seen by means of the U- Matrix in **Fig. 2.3.13**. Indeed, data in the light area correspond to those points that are far from the center and are identified as punctual deviations from the nominal conditions. Furthermore, a second cluster can be identified in the right corner of the U-Matrix. This high data density area corresponds to a second operation point also reflected in the input data space in **Fig. 2.3.12 (a)**. Note that the interpretation of the SOM information by the model differs from the simply understanding of the U-matrix, since the model uses this map in order to relate behaviours detected in the target signal with different operating areas codified in this map.

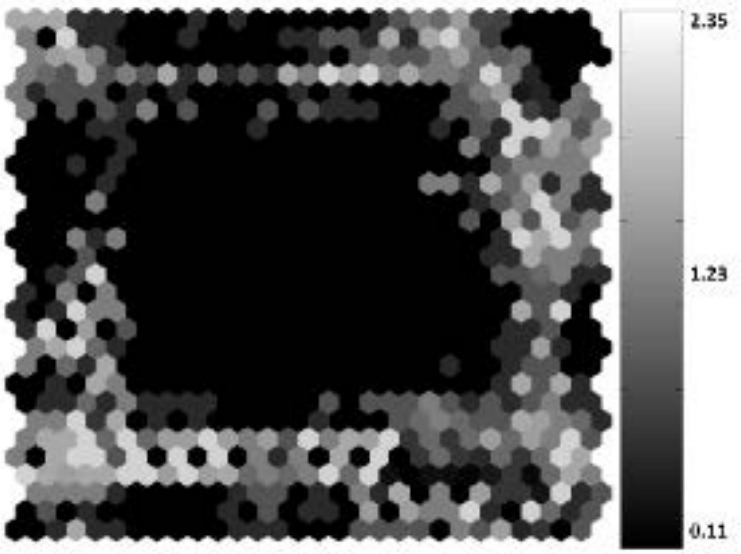

Fig. 2.3.13. U-Matrix of the trained SOM containing the distances between the neurons of the grid. A central cluster corresponding to the highest data density can be appreciated, furthermore, the data is scattered in all directions among the central cluster. This behaviour matches with the initial data topology.

The BMU for both test and validation sets is shown in **Fig. 2.3.14**. Any process variation should be reflected in a certain region of the input data space covered by the SOM's neurons. As can be seen in **Fig. 2.3.14 (b)** for the validation set, most of the data presents a BMU value from 50 to 120. This interval corresponds to the central cluster identified in the U-Matrix as the most common operation point. Values under and over this interval can be considered as changes in the operation point. Indeed, the training set exhibits a variation of the temperature around the 22^{th}-28^{th} hour of operation (**Fig. 2.3.11**). This behaviour is also reflected in the BMUS of the SOM. As can be seen in **Fig. 2.3.14 (b)**, the BMU is moved to one frontier of the SOM grid (around the BMU 225, that

corresponds to the upper right corner of the u-matrix), which confirms the relation between the tundish temperature and the process operation condition codified by means of the auxiliary signals.

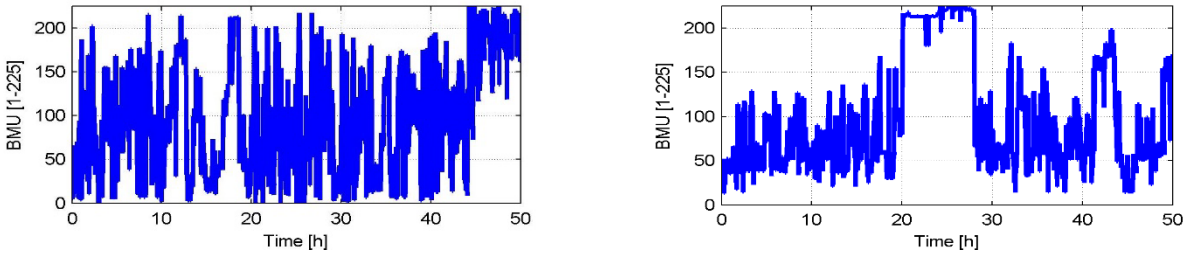

Fig. 2.3.14. Best matching units for both a) training and b) validation data sets among the corresponding 50h of plant operation.

Once the auxiliary inputs are summarized in to the process operating point by means of the SOM assessed BMU, the next step in the methodology is to generate and train, the ANFIS models with the training set. First of all, the prediction horizon *p* of the models should be selected. For this concrete experiment, the horizon has been configured to detect deviations in the actual cooper rod under production, before it affects the next manufactured unit. Considering the manufacturing time of one cooper rod, the prediction horizon is configured to be 900 seconds (90 samples).

According to section 2.2.1, for a given forecasting horizon, the viability of such horizon needs to be assessed. In this regard, the auto-correlation of the tundish temperature is shown in **Fig. 2.3.15**. As shown in the figure, the training set presents a quickly drop of correlation during the first iterations caused by the presence of high frequency oscillation modes in the signal. In comparison, the validation set presents a slower decay since the presence of higher frequencies is lower in this dataset. However, after the initial decay, the training set stabilizes and maintains a certain level of correlation, $C_{coef} \approx 50\%$, as time is increased. This is not the case for the validation set, that presents a sustained decrease for all the analysed range, less than 50% from 0.5 hours. This fact indicates that the complexity in terms of signal dynamics of the validation set is higher than the training set. Regarding the forecasted horizon required by the application, 15 minutes, it can be check, in **Fig. 2.3.15**, the viability of such horizon, since C_{coef} is around 45% for both datasets.

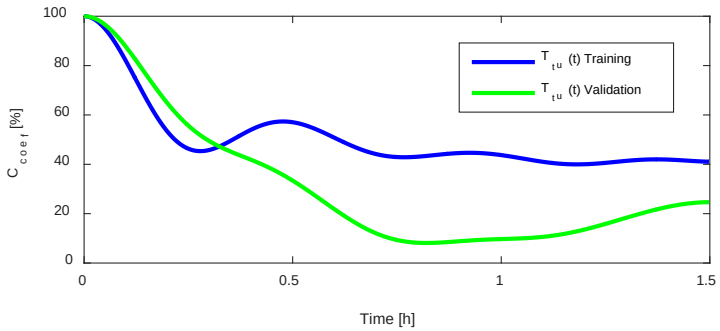

Fig. 2.3.15. Auto-correlation analysis of the tundish temperature. At 15 minutes, the correlation is 46.32% for the training set.

The ANFIS model has been configured according to a 3-input structure, that consists on: (i) the current value of the strainer's temperature, T_{tu} *(t)*; (ii) past values of the signal delayed z1 samples, T_{tu} *(t-z1)* (iii) process operating point of the auxiliary signal summarized by the BMU, *BMU (t).* Each input is normalized with the min-

max method in order to have a range from 0 to 1. Then, the inputs are fuzzified by means of three generalized bell-shaped membership functions. The temporal values that show an autocorrelation value over 0.4 are selected to define the pool of possible *z1* delay values, that is a range of [100 - 150] delayed samples for the training set. A GA is proposed to find such delays using the forecasting performance of the two ANFIS that forms the collaborative structure fed by the delayed target signal and the BMU.

Once the optimal delay is met, the model is trained by means of the classical hybrid learning algorithm. After 15 epochs, the prediction with the training set can be seen in **Fig. 2.3.16 (a-b)**, and in **Fig. 2.3.16 (c-d)** for the validation set. The resulting performance of the trained model is validated through the achieved RMSE, MSE, and MAPE.

Table 2.3.2. Performance evaluation of the ANFIS model with the training and the validation sets.

	Training	Validation
RMSE	0.0927	0.1483
MAE	0.2669	0.3389
MAPE	0.1194	0.8668

In order to compare the proposed approach with the most classical one proposed in the literature, the following comparative with a classical ANFIS with a GA input selection method, and an auxiliary signals reduction based on OCA compression.

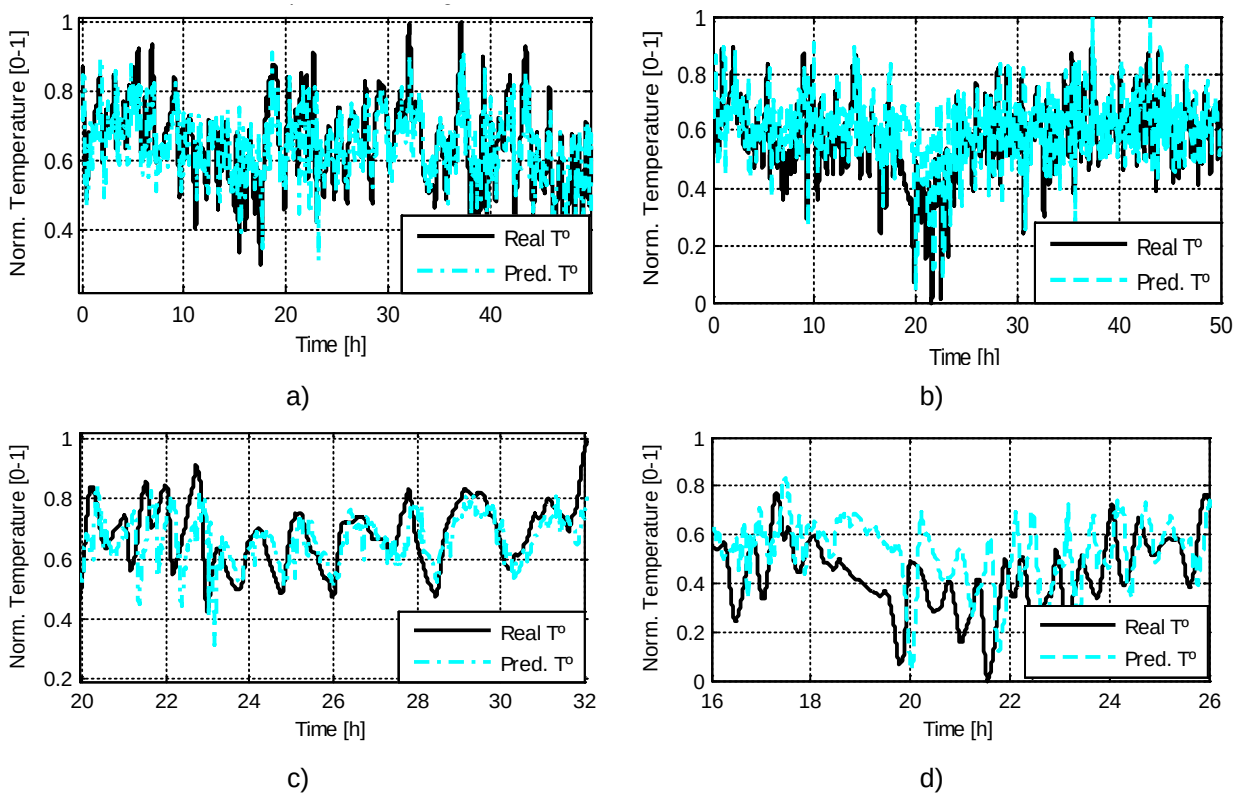

Fig. 2.3.16. Training and validation results of the ANFIS model with prediction horizon, p=90. a) Training result of the model, b) Detail of the training result in the highest error area. b) Validation result of the model, d) Detail of the validation result during the disturbance in the 20th hour of plant operation.

In this regard, the considered approaches are listed below, and the comparative of the different methods is shown in **Table 2.3.3**.

- ***Single ANFIS:*** For this model, no special processing has been applied to the signals, the auxiliary inputs are directly introduced to the ANFIS model, all together with the actual and the delayed sample of the strainer temperature. As a result, the model has five inputs which considerably increases the computational cost of the training procedure with big amounts of data.
- ***G-ANFIS:*** This model uses an input selection method based on a genetic algorithm in order to select the most suitable inputs that helps the model to reduce the fixed cost function. According to the literature, the selected cost function is the RMSE of the model with the validation data set [17].
- ***PCA-ANFIS:*** For this method, a PCA is used to compress the information of the process with the objective of maintaining the maximum variance among the samples. Then, the principal component is introduced in the ANFIS model in order to extract information from the process.

Accordingly, it can be seen how the basic ANFIS with all the different signals performs better in the training than other methods. This happens because the ANFIS is able to capture the relations among the data and adjust the model output accordingly. However, the model becomes too overfitted for the training and totally misses the validation process not being able to generalize the output of the model in regard with new data from the same process.

The G-ANFIS sacrifices performance in the training process (has a higher RMSE), since is seeking the optimal validation error. However, it shows an interesting fact, which is that the best individual combination of input signals cannot outperform the performance of combining the inputs to extract the information as SOM-ANFIS or PCA-ANFIS.

Finally, the PCA-ANFIS is able to compress the information regarding the process and helps the model to reach a good performance which is similar to the one proposed by the SOM-ANFIS. However, it presents some instabilities in the validation due to sudden changes in the temperature as can be seen in **Fig. 2.3.17**. Therefore, the PCA-ANFIS has a similar performance in absolute terms, but presents less stability among the validation than the proposed method, which an important factor when dealing with real industrial applications in order to avoid generating fake alarms.

Table 2.3.3. Performance evaluation of the ANFIS model with the training (Trn.) and the validation (Val.) data sets.

	SOM-ANFIS		**ANFIS**		**G-ANFIS**		**PCA-ANFIS**	
Data	*Trn.*	*Val.*	*Trn.*	*Val.*	*Trn.*	*Val.*	*Trn.*	*Val.*
RMSE	0.093	0.148	0.068	6.330	0.141	0.179	0.092	0.156
MAE	0.267	0.339	0.222	1.539	0.331	0.375	0.267	0.340
MAPE	11.94	86.70	8.081	544	93.67	75.83	11.941	104.7

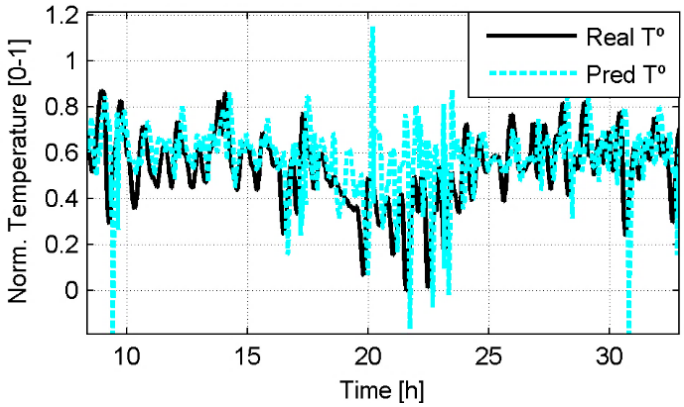

Fig. 2.3.17. Detail of the performance of the PCA-ANFIS model (p=90) with the validation dataset. Some disturbances peaks can be observed in the response of the model in points when the dynamic is faster.

Discussion and Conclusions

This study presents a complete industrial sub-process data forecasting scheme by means of industrial data from a cooper manufacturing plant. The results of the method show the impact of proposing a coherent auxiliary signal management in forecasting industrial time series. In this regard the synergy between ANFIS models and SOM helps to improve model accuracy, combining the inputs using the SOM results in a 17% less error than the best individual combination of inputs. Additionally, compressing the auxiliary signals with the SOM adds simplicity and generalization in front of new measurements, which are important features to be considered in real industrial applications.

Moreover, the information compression capabilities of the SOM for industrial applications have been presented. The SOM based mapping is able to adapt the grid to the input data space of the process and summarize this information into the BMU. Also, for this experimental application, the inherent correlation between the process variables and the target signal have been demonstrated; a sudden drop in the target signal is also reflected in the BMU, since there is a change of the operation point of the plant.

Furthermore, a comparative between the most common approaches in the literature has been made; the conclusion extracted from this comparison is that the combination of the information of the process outperforms the individual selection of inputs by means of optimization methods, which confirms the basis of this work.

It should be noted that this methodology can be applied to monitor and forecast time series of interest in different industrial applications. Future work can be concentrated on controlling the gas throttle/s and oxygen level using the predicted values which is a matter of interest for most cooper manufacturing plants and similar industries.

2.4 Conclusions

The presented chapter deals with the proposed contributions in regard with the data pre-processing in industrial target signals forecasting. In this regard, the contributions on this topic are related to:

i. How to analyse the target signal in order to find an optimal forecasting horizon, or given a defined forecasting horizon? How to evaluate its viability from a signal forecasting point of view?
ii. How to select a search range for finding the most suitable past inputs?
iii. How to select the auxiliary signals that are more related to the target that wants to be modelled?
iv. How to map the auxiliary signals in to the process operating point in industrial process forecasting by the use of Self Organizing Maps as non-linear topology preservation mappings?

In regard with the first contribution, the experimental results with the electromechanical actuator shown in this way highlight the necessity to analyse the target signal in order to find correlations or hidden periodicities that helps the model not to get stuck in a local minimum during the training phase. Furthermore, the autocorrelation analysis is presented itself as a low-computational way to extract the behaviour of the target signal in regard with its successive delays. This coefficient presents a reliable and fast approach that presents high benefits during the pre-processing stage that are reflected in a higher convergence and accuracy results during the final training of the model.

Furthermore, regarding (ii), the selection of the best past inputs for a given forecasting horizon, it should be concluded that, finding the most suitable range of possible past values by analysing the periodicities in the signal, corresponds to a low computational cost approach that presents an impact on the performance of the optimization algorithm in charge of finding the vest delayed samples. In this regard, this simply analysis fixes the real range of search and avoids the convergence towards a local minimum solution.

One of the most important contributions in this chapter is the solid affectation that present auxiliary signals in regard with target signals in industrial processes. Indeed, it have been demonstrated in regard with (iii), that approaching the selection of the target signals to consider in a two steps in regard with the cross correlation of such signals, allows to consider only the signals with a strong influence to the target signal while assuring a certain degree of information independence among the considered signals.

Finally, the codification of the auxiliary signal by means of the self-organizing maps present several benefits in the development of the industrial time series forecasting models. In this regard, it allows to compress all the information given by the significant auxiliary inputs in a single indicator, the BMU. Such magnitude has been demonstrated to present a dynamic that is strong related with the industrial process monitored, and hence, it is a representative measure about any undesired deviation occurred to that process. Furthermore, SOM based mapping improves generalization while assures high ratios of accuracy and precision in the modelling of such target signals.

Summarizing, the research work presented in this chapter address the proposal of an optimization procedure of the forecasting horizon and the past inputs selection, and the establishment of non-linear mapping of the auxiliary signals as model generalization improvement while increasing the convergence speed of the algorithm for the same dataset.

3.

Contributions to

Dynamics based industrial time series modelling

Signal's dynamics consideration is proposed as a critical factor in industrial time series modelling, since it conditions model performance and accuracy. In this regard, this chapter presents the contributions to enhance modelling results by specific methodologies based on the analysis of the signal's dynamics.

CONTENTS:

3. Industrial Time-Series Modelling

3.1 Introduction

The literature published to date shows that forecasting of industrial process condition, in terms of target signals evolution for supervision purposes, is still a novel field for research, in which coherent methodologies are expected. Indeed, there are two main challenges related to industrial target signal forecasting. First, the assessment and exploitation of auxiliary information related with the target signal, which is required to enhance the forecasting performance avoiding computational complexity and model overfitting [62], as it has been presented and studied in the previous chapter of this thesis document. Second, the consideration of suitable procedures to deal with highly non-linear signal's dynamics behaviours, as is the case of most industrial processes, where the auto-correlation of the target signal quickly decreases within a short period of time [61].

In this regard, H. A. Zamani *et al.*, in [103], propose the use of an ANFIS modelling for complex non-linear time series forecasting. In such method, a single ANFIS model to forecast a complex non-periodic gas concentration signal is used. Indeed, the single-model scheme corresponds to the classic approach, since is the easiest procedure used in literature to handle complex signal's dynamics. It corresponds to the use of the whole raw data inputs during the training of the algorithm. Therefore, dealing with industrial time series forecasting, such industrial signals response with a non-linear dynamic content. Such scenario uses to exhibit a short-term correlation that is difficult to be modelled with classical approaches. Thus, in such approaches, the modelling scheme is not able to learn the variability of the signal, and the forecasting scheme undergoes a loss of performance leading the system to an over-fitted response.

In order to deal with such problems, recent studies pointed out that splitting the target signal in order to be modelled in different dynamics is a suitable approach dealing with non-linear time series. Indeed, different multi-scale signal decomposition approaches are being proposed [104]. One of the classical approaches corresponds to the wavelet package decomposition, which allows splitting a time-signal into subspaces containing high and low frequency components by computing iterative filtering of each resulting subspace. R. Hooshmand *et al.*, in [105], proposes the combination of Wavelet Package Decomposition (WPD), and ANFIS, to perform an electric load forecasting. In such approach, WPD is used to extract high and low frequency modes of the signals. Then, a dedicated ANFIS model for each set of frequencies is considered. The authors state that decomposing the signal outperforms classical single-model approaches. However, although the number of resulting subspaces is controlled by the number of iteration levels, the filtering cut frequencies must be defined. This fact requires a deep knowledge of the process that, usually, corresponds to a non-linear and multi-dimensional response.

Other classical approach for non-linear and non-stationary signal analysis is the Empirical Mode Decomposition (EMD). The EMD decomposes a complex signal into a collection of Intrinsic Mode Functions (IMF), and a residue. Although the resulting number of IMF is not controlled, the adaptive capabilities to the signal under analysis represents a suitable approach for a great deal of signal decomposition purposes [84]. Recently, additional studies, have presented significant variants of the EMD, as the Ensemble Empirical Mode Decomposition (EEMD), that allows the extraction of low-correlated IMF by avoiding the classic mode-mixing problem of the EMD.

However, currently, there exist just a few works in the literature dealing with signal decomposition for modelling and, even less, considering EMD based approaches for signal forecasting. Moreover, in this last case, the available studies are based on dedicated models for each one of the extracted IMF. For instance, the authors, in [104], propose a EMD based approach to extract the main modes of a wind speed signal, and each resulting IMF is modelled by a dedicated artificial neural network; then, all model outputs are combined to give the prediction outcome. The results are convincing, however, the performance of the method presents an over-adaptation of the training set. L. Y. Wei, in [82], proposes the combination of EMD and ANFIS modelling. The author states that the combination of EMD and ANFIS improves the resulting performance of a time series forecasting in signals with high variability, that is, multiple oscillation modes.

As a conclusion, the literature in target signal dynamics analysis for industrial time series modelling is related to decomposing the signal either by a multi-scale decomposition method, or by an adaptive approach in order to generate dedicated models for such dynamic details extracted. However, there is still an open problem to optimize such forecasting schemes in terms of computational burden and resulting accuracy, as well as their validation of dynamics analysis methodologies over industrial time series modelling.

Therefore, the contributions of this chapter aims to cover the dynamics analysis necessities identified in the literature with its corresponding application to the modelling of the industrial process considered by this thesis, the copper rod manufacturing process. In this regard, the contributions on this topic can be divided in three main points: (i) the proposal of methodologies for the coherent aggrupation of signal dynamics in critical industrial signals forecasting. (ii) The proposal of the Neo Fuzzy Neuron as a modelling method able to overcome ANFIS limitations when dealing with industrial time series forecasting. (iii) The proposal of methodologies for modelling industrial time series considering both signal dynamics and auxiliary signals information.

3.2 Forecasting modelling based on signal dynamics aggrupation

The dynamics of a target time-domain signal, is referred in the literature to the different frequency modes that conforms the totality of the signal [106]. Therefore, dynamics can also be seen as the intrinsic oscillation modes of the target signal. Generally, in terms of signal modelling, more dynamic content implies more complexity to model such signal, due, meanly, to the presence of non-linear relations among the rest of the available auxiliary signals of the process to be modelled, and the increased decay ratio of periodicities on its response, at least, in regard with the high-frequency components.

Therefore, time series forecasting requires the identification of the most significant target signal's dynamics and their learning, to provide, later, a reliable prediction in a predefined forecasting horizon. Classically, overloaded forecasting approaches are applied, where the complete signal is used as input of the model. However, generally, only some modes of the original signal's dynamics can be properly learned by the model. Indeed, such classical modelling approaches deal with a loss of performance due to convergence difficulties. To face such problem, other methods in literature decompose the signal and generate dedicated models for each IMF detail, as presented in [107]–[109]. However, such approaches represents a high computational-burden strategy and, moreover, the simplicity of modelling some of the IMF can lead the corresponding model and, then, the global forecasting performance, to an intense overfitting, due to the highly adaptation of each IMF to the training signal [110].

Indeed, the main concern and the proposed contribution, is to find a high-performance decomposition based strategy by analysing the significance of each IMF in regard with the forecasting model. In this regard, considering a target signal decomposition in a set of IMF, it has been reported that the performance of a single model approach non-linearly decreases with the number of IMFs or dynamic modes considered [104]. This concept is illustrated in **Fig. 3.2.1**, where the loss of performance that suffers a classic single ANFIS model when additional IMFs are added to the training signal can be seen. In this regard, the figure illustrates the conceptual relation between the modelling error of an ANFIS model where successive IMFs are added to the training set of the model. Indeed, the model response increasing the error exponentially until a stability point is reached. This increase of the model error is caused by the limitation of the ANFIS model to capture wide band signals which contains a rich spectrum formed by different frequencies at the same time.

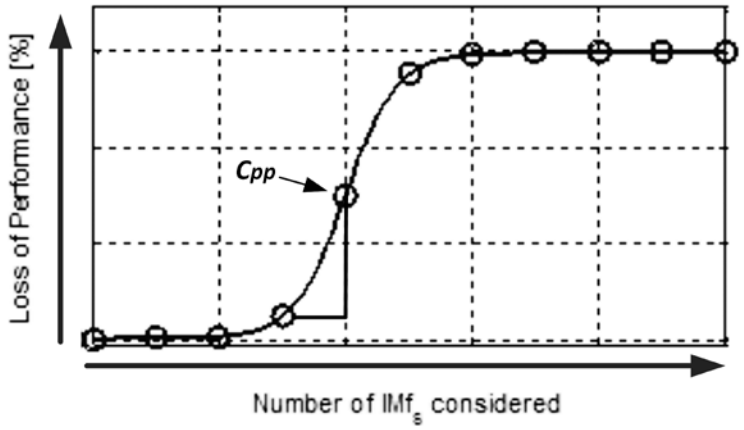

Fig. 3.2.1. Representation of the effect over the resulting accumulated error of the number of IMF considered for a single model. The critical performance point, C_{pp}, represents an allowable down limit in regard with the modelling performance.

3.2.1 Definition of the dynamics aggrupation methodology

The aim of the proposed method is the analysis of a signal decomposition result, by considering the significance of each IMF to package them in optimized groups for modelling performance enhancement. Indeed, such coherent packaging of the resulting IMF represents the multimodal modelling approach proposed to increase performance and generalization, while optimizing the number of required models.

Furthermore, as seen in **Fig. 3.2.1**, there is a point in which the error suddenly increases when including an additional IMF. This point is considered to be the Critical Performance Point, $C_{pp,}$ and defines the end of an IMF aggrupation, or pack, that can be processed by a single model. Hence, the aim of the method is the identification of the successive C_{pp} to form the N packs that consider all available IMF. Such identification is based on the evaluation of the accumulated dynamics in an error's graph. Therefore, an error threshold function must be defined to identify automatically such points. Then, if the error of a dynamics' pack surpasses the error function in some point, this represent a C_{pp}, and the pack is closed removing the last IMF added.

As a result, the problem turns into the identification of the optimum number of IMF that should form each package. The C_{pp} is mathematically described by the proposal of an error threshold function, $E_{TH.}$ That is, the E_{TH} defines the allowable model error curve in regard with the significance, in terms of energy contents, of the considered IMF compared with the original signal. The E_{TH} is defined as a decreasing second order function, as shown in Eq. (3.2.1).

$$\mathrm{E_{TH}}(Re_i, Er) \leq A \cdot Re_i^2 + B \cdot Re + C \qquad \textbf{Eq. 3.2.1}$$

The industrial applicability of this thesis has been taken into consideration for the mathematical definition of the error function, where, generally, low frequency modes represent long-term process behaviours. Therefore, this function has been designed to be permissive in terms of error with low-significant modes, which in most cases correspond to the higher frequencies contents of the signal under analysis, and restrictive with the high-significant modes, that usually represent the low frequencies and main trends of the signal. Note that the significance is quantified by the amount of energy that accumulates a certain IMF in regard with the original signal. Thus, the parameters *A, B* and *C*, used to define the error threshold function, would be identified by means of interpolation, as it is shown in **Fig. 3.2.2**. Three points are needed to allow the regression, that are: (i) the maximum allowed error for the low-significant modes, LW_{MAX}, (ii) the maximum allowed error for the high-significant modes, HG_{MAX}, and (iii) the smoothing factor, S_m, which fixes the decay of the curve.

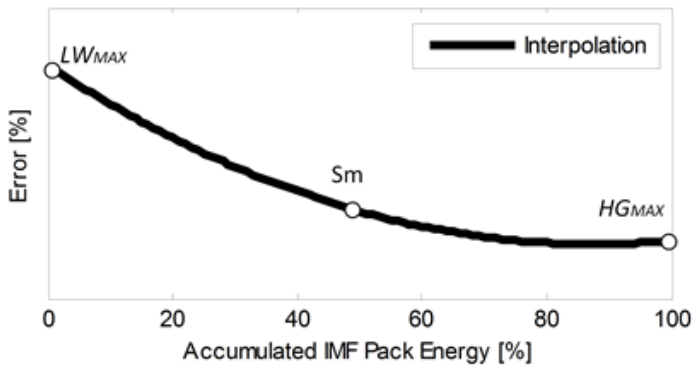

Fig. 3.2.2. Example of a resulting error threshold curve in regard with the accumulated energy, $R_{e,}$ of the IMF package. The curve is described by the interpolation of a second order function within the three points defined, LW_{MAX}, S_m and HG_{MAX}].

Thus, in order to carry out the IMF packaging procedure, the modelling errors must be estimated and, then, compared with the predefined error threshold curve. For this procedure, a simple model is proposed, that is, a classical time series one-input ANFIS structure, using current value of the target signal as an input of the model, and the predicted value as the output. This approach represents a trade-off between simplicity and performance analysis, and corresponds to a usual wrapper approach applied in multiple data-driven approaches. Then, when the model error surpasses the corresponding threshold, the last intrinsic mode function is removed from the model, and the fist IMF package is closed. Then, the procedure starts again with a new model consideration, and the iterative addition of the rest of the IMF till the *N* packs are formed. The objective is to obtain *N* IMF packages of which dynamic combination is affordable by a simple ANFIS model.

In order to carry out the IMF packaging procedure an iterative process is defined as follows: the EMD is applied to extract the *k* different IMF. Let $Pk_j(t)$ be the *j-th* package of IMF, where $j \epsilon [1..N]$ and contains IMFs from ki_j to kf_j, then,

1. In each iteration *j*, $Pk_j(t)$ is formed by Eq. (3.2.2). Note that in first iteration the method initializes by $ki_1 = kf_1 = 1$.

$$Pk_j(t) = \sum_{k=ki_j}^{kf_j} IMF_k(t) \qquad \textbf{Eq. 3.2.2}$$

2. $Pk_j(t)$, is modelled by a 1 input ANFIS model in order to obtain the predicted output, $\hat{y}_j(t)$, at a time horizon p, Eq. (3.2.3). The model uses as inputs the current value of the signal, $Pk_j(t)$.

$$\hat{y}_j(t) = ANFIS\left(Pk_j(t-p)\right) \qquad \textbf{Eq. 3.2.3}$$

3. The model performance is evaluated and the prediction error, E_r, is calculated by Eq. (3.2.4).

$$E_r = \frac{1}{L}\sum_{t=1}^{L}\frac{\left|y_j(t) - \hat{y}_j(t)\right|}{\left|y_j(t)\right|} \qquad \textbf{Eq. 3.2.4}$$

4. The significance in terms of relative energy, Re_j, of $Pk_j(t)$ versus the complete signal $x(t)$ is obtained by Eq. (3.2.5).

$$Re_j = \frac{\sum\left|Pk_j(t)\right|}{\sum|x(t)|}\cdot 100\ \% \qquad \textbf{Eq. 3.2.5}$$

5. The error threshold function evaluated in regard with both calculated points, *Re* and *Er*.

If the performance is under the allowable error defined in Eq. (3.2.1), another intrinsic mode function should be added to the model, the pack index is incremented by increasing $kf_j + 1$. Then, steps from 1 to 5 are repeated till the consideration of the *k* IMF. If the point is above the curve means that $Pk_j(t)$ should be closed with the IMF added in the previous iteration. Therefore, the number of IMF is decreased by $kf_j - 1$; and the initial intrinsic mode function of the next package, $Pk_{j+1}(t)$, is prepared by $ki_{j+1} = kf_{j+1} = kf_j + 1$. Then, $j+1$ and the algorithm returns to step 1. At the end of this procedure, all the IMF will be distributed in *N* packages.

3.2.2 Experimental validation of the signal's dynamics aggrupation method

An experimental application of the proposed dynamics analysis and packing method has been made in regard with the industrial process used in this thesis as validation platform, the CRMP. The objective of this experimental application is to forecast the tundish temperature, $T_{tu}(t)$. This magnitude is critical in the manufacturing process because represents the last reference of the melted copper before entering to the casting wheel. Deviations in the temperature of the tundish imply imperfections in the manufactured copper rod due to non-uniformities in the copper density. The forecasting of such temperature allows the corresponding actions to avoid the affectation to the next manufacturing batch.

In this regard, the aim of the proposed method is to exploit the forecasting capabilities by analysing the associated intrinsic mode functions of the signal to be predicted. Thus, taking advantage of the EMD performance, a two-step forecasting approach is proposed. Different packs of IMFs from the tundish temperature are analysed by the IMF packing procedure explained before. Once the different N sets of IMFs are consolidated, each set of IMFs is modelled separately with a dedicated ANFIS model to boost the dynamics forecasting capabilities. It should be noticed that in order to isolate the affectation of the dynamics packaging from the modelling method, a classic ANFIS approach with a regular configuration have been used in the method. Finally, considering the superposition properties of the resulting IMF of the EMD analysis, the outputs of the models are combined to obtain the forecasting estimation. In this regard, the proposed method is summarized in **Fig. 3.2.3**.

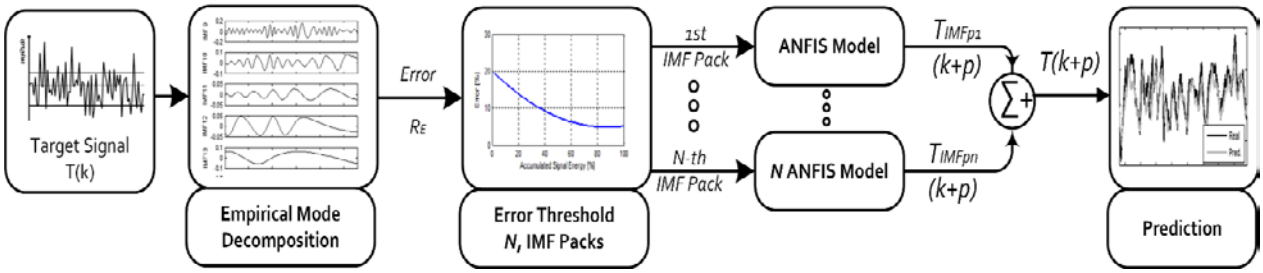

Fig. 3.2.3. Enhanced time series forecasting scheme by means of EMD analysis of particular dynamics and ANFIS modelling.

After the dynamics analysis and packing procedure, the resulting *N* ANFIS models are considered, that corresponds to the *N* packs of IMF found. For each ANFIS model, the proposed forecasting follows Eq. (3.2.5). The proposed structure for the *N* resulting models corresponds to a three inputs, one output scheme. The considered model output is the resulting signal at *p*, $P_k(t+p)$, *t*he considered inputs are:

(i) the current value of the modelled IMF package, $P_k(t)$,

(ii)- (iii) *two* past values of the signal in order to have a reference of its tendency, $P_k(t-z_1)$, and $P_k(t-z_2)$. It should be noticed that the delays are selected by means of optimization methods combined with the past values range selection exposed in Section 2.3.1 of this thesis.

$$P_k(t+p) = ANFIS(\,P_k(t), P_k(t-z_1), P_k(t-z_2)\,) \qquad \textbf{Eq. 3.2.5}$$

In order to compare the proposed forecasting method performance with classical approaches, a set of standard metrics have been considered. The Root Mean Squared Error (RMSE), reflects the standard deviation of the prediction error. The Mean Absolute Error (MAE), measures the average error. The Mean Absolute

Percentage Error (MAPE), reflects the average deviation of each observation divided by the signal amplitude, it exhibits the percentage precision of the modelling.

Next, the experimental results on the application of the proposed method to the tundish temperature of the copper rod manufacturing, T_{tu} *(t),* are shown. Such industrial system represents a challenging scenario for process forecasting due to the non-stationary operating conditions and the non-linear relation among the process signals. In this regard, two data sets corresponding to 48 hours of plant operation are used in this work, the first set, shown in **Fig. 3.2.4**, is used for training purposes, while the second test is used for testing the suitability and the generalization capabilities of the method.

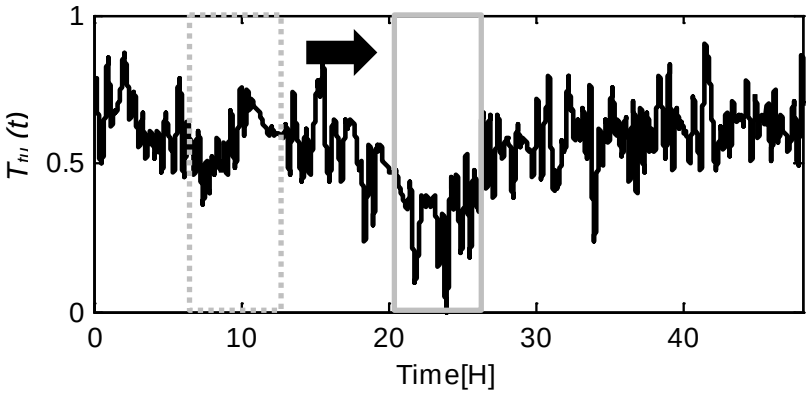

Fig. 3.2.4. Acquired tundish temperature, T_{tu} *(t)*, in the manufacturing plant in 48 hours of operation. Corresponds to the training signal of the method. The sliding window used for the computation of the EMD decomposition is also illustrated in the figure.

The forecasting horizon, *p*, represents one of the most critical parameters in regard with the forecasting performance. The application defines *p* to *p*=90 samples, that is, one copper rod batch manufacturing period (15 min). Even though, as mentioned in section 2.2.1 of the thesis, due to the chaotic behaviour of industrial signals, it is required to analyse the feasibility of such horizon from the modelling point of view. In this regard, it is proposed in this thesis the autocorrelation analysis in order to obtain a quantitative measure of the significance of the forecasting horizon regarding the target signal. The correlation curve of the training signal, extracted by applying n Eq. (2.2.1), is depicted in **Fig. 3.2.5**.

It shows a characteristic behaviour of non-periodic signals with a certain time constant. Indeed, the slope and the shape of the curve is characteristic of numerous industrial applications, and represents a qualitative indicator about the decrease of performance when the forecasting horizon is increased. The achieved correlation performance of T_{tu} *(t)* is 43.12 % which is considered to be significant in regard with the defined industrial application and the dynamics of the signal, therefore it is feasible to use p=90 as the forecasting horizon.

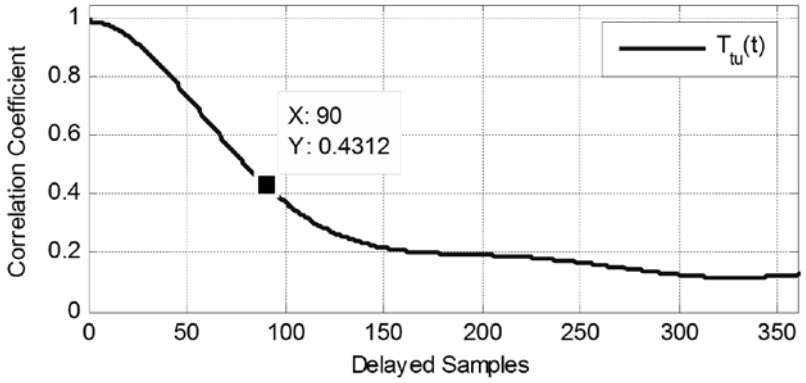

Fig. 3.2.5. Correlation coefficient between the tundish's temperature and its successive delayed signals. It shows the decrease of similitude when the delay (directly related with the forecasting horizon) is increased. For the selected p, the correlation is 43.12 %.

Prior to all calculations, considering the continuous operation requirements of the proposed method, it is required to use a temporal time buffer for the update of the EMD calculation every time a new observation is available. Then, it is necessary to define the length of the time buffer to compute the EMD. This buffer should be configured in regard with the application since it should be representative of the slowest dynamics presented in the tundish's temperature. Consequently, the slowest dynamic can be defined as the inverse of the first harmonic after the continuous content. Accordingly, DFT is used to analyse the spectrum of $T_{tu}(t)$. **Fig. 3.2.6**, shows the resulting DFT for the training set within 48 hours of plant operation. It can be appreciated that the first harmonic is located at $61.04 \cdot 10^{-6}$ Hz that corresponds to a temporal window of 4.5 hours (1620 samples). Furthermore, by adjusting the buffer to the lowest dynamic the stability of the IMF decomposition is assured, that means to obtain in each window a similar number of signal details.

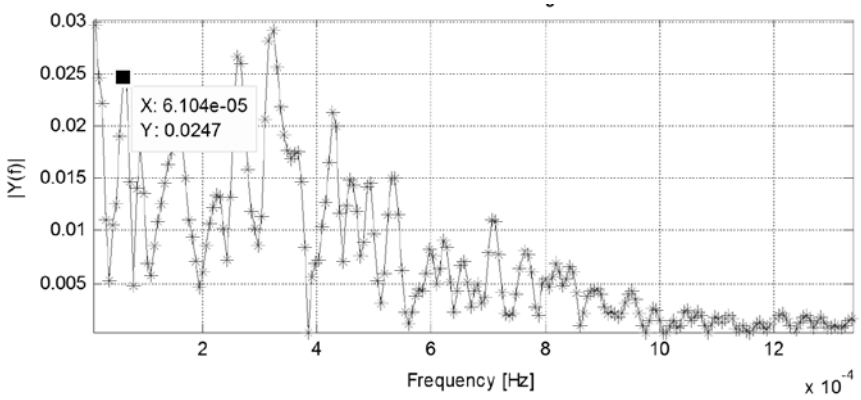

Fig. 3.2.6. FFT of the tundish temperature for 24h of operation. The temporal period associated to the first appreciated harmonic after the offset content is selected as the window width of the method. For this application, the slowest dynamics is at 6.104E-5 Hz, that corresponds to a temporal frequency of 4.5 hours of plant operation.

Considering the continuous operation requirements of the proposed method, the EMD is approached by means of the utilization of a time based buffer. The corresponding time window is selected in regard with the lowest dynamic modes contained in the target signal, that is a temporal window of 4.5 hours (1620 samples). Thus, the EMD is applied to $T_{tu}(t)$. A total number of k = 9 IMF are obtained. An example of a resulting IMF decomposition during the training process is shown in **Fig. 3.2.7**.

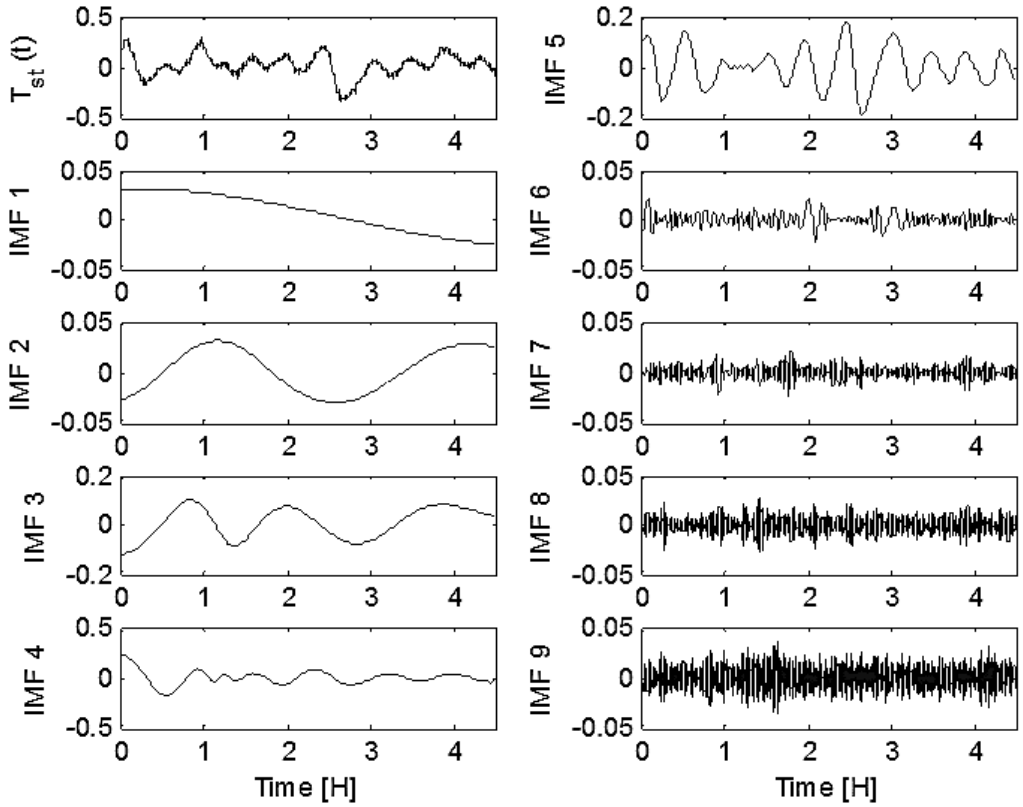

Fig. 3.2.7. Intrinsic mode functions extracted from the tundish's temperature of the training set during a time window processing.

Following the method, the IMF are successively evaluated over an error threshold curve to find the optimum number of *N* IMF packages. The E_{TH} has been defined by fixing the low significance error, LW_{MAX}, at *20%,* the smooth parameter, S_m , at *7.5%,* and the high significance dynamics error, HG_{MAX}, at *5%*, that results in *A= 0.002*, *B= 0.35* and *C= 20* following Eq. (3.2.1).

The procedure to form the first package of IMFs, $Pk_{j=1,}$ starts by evaluating IMF_1 in terms of relative energy and error, Eq. (3.2.5), versus E_{TH}. As the threshold is not met, consecutive IMF are successively added to $Pk_{j=1}$. The addition of IMF_4 in $Pk_{j=1}$, shows an error value over the admissible threshold. Thus, $Pk_{j=1}$ is formed by the three first IMF, from IMF_1 to IMF_3. Then, the second package, $Pk_{j=2}$, starts with IMF_4 and successively evaluates all modes till IMF_9. Since the error threshold is not reach anymore, *N=2* packages. This analysis is shown in **Fig. 3.2.8**. The resulting packages, $Pk_j(t)$, are shown in **Fig. 3.2.9**.

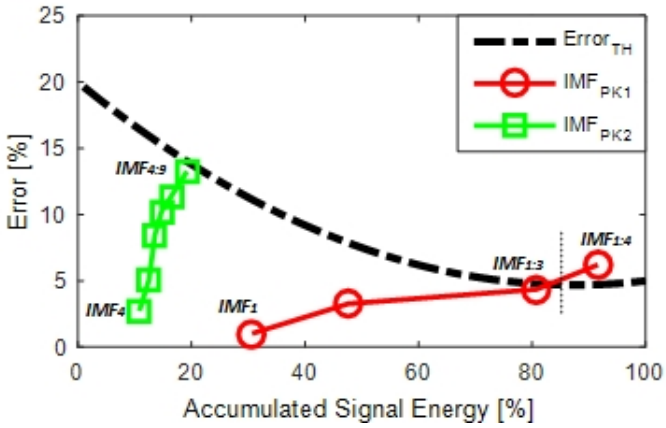

Fig. 3.2.8. Evaluation of the training set versus the E_{TH} to find the number *N* of IMF packages.

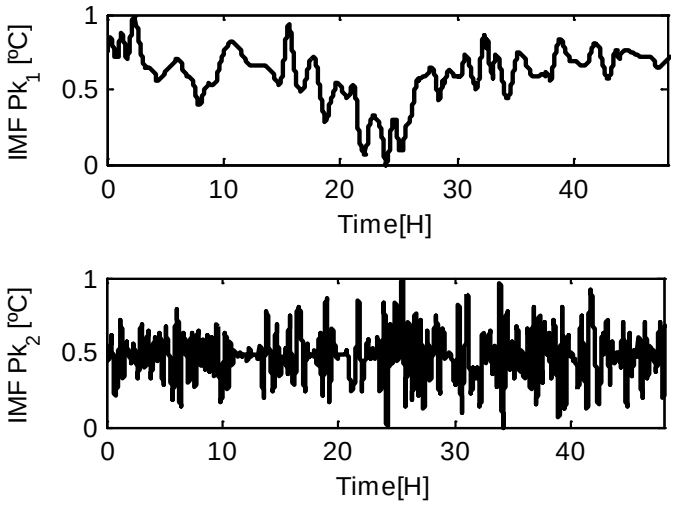

Fig. 3.2.9. Resulting IMF packages to be modelled. (a) Package *j*=1. (b) Package *j*=2.

In order to train the $N = 2$ models, the delays z_1 and z_2 value for the second and third input parameter must be selected. For this aim, the proposed combination of correlation range selection together with GA optimization explained in section 2.2.2 is proposed. As shown in **Fig. 3.2.10**, the resulting signal from Pk_1 shows a smoother autocorrelation decrease indicating that the loss of performance is smoother in regard with the forecasting horizon. Otherwise, the resulting signal from Pk_2 exhibits a sudden drop of autocorrelation till delay 60th.

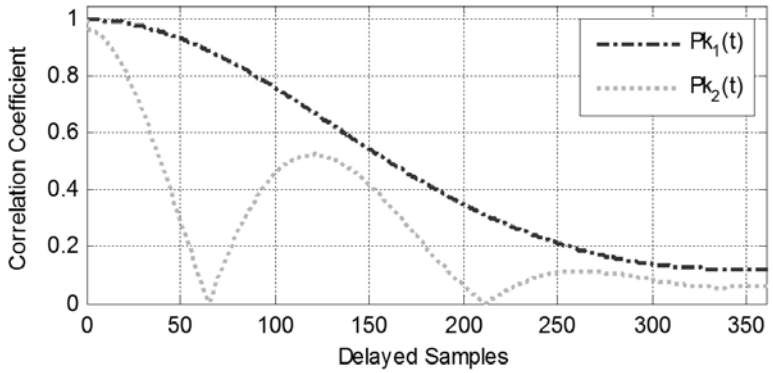

Fig. 3.2.10. Correlation coefficient between the IMF packages (Pk_1 and Pk_2), and their successive delayed signals.

This response is characteristic of high frequency modes, that are included in the Pk_2. That is, such difference in the autocorrelation response between Pk_1 and Pk_2 relies on the predominant low and high frequency contents, respectively. The temporal values that show an autocorrelation value over 0.4 have been considered as proper *range of* delays. A GA has been carried out to find such delays for each package, using the forecasting performance of the ANFIS as cost function. The resulting delays are z_1=11 and z_2=27 for $Pk_{j=1}$ and z_1=5 and z_1=11 for $Pk_{j=2}$, that corresponds to a delay of 110 and 270 seconds for $Pk_{j=1}$ and 50 and 110 seconds for $Pk_{j=2}$.

Finally, in order to complete the ANFIS design, each input is normalized with the *min-max* method, obtaining input signals in the range 0 to 1. The inputs are fuzzified by means of three generalized bell-shaped membership functions. The model is trained for 15 epochs by means of the classical hybrid learning algorithm, which is the combination of the least-squares method and the backpropagation gradient descent method. The

forecasting performance during the training can be seen in **Fig. 3.2.11 (a)**, and the corresponding validation in **Fig. 3.2.11 (b)**. The performance' metrics of the model are shown in **Table 3.2.1**.

Table 3.2.1. Performance of the IMF models

	$Pk_1(t)$		$Pk_2(t)$	
	Trn.	**Val.**	**Trn.**	**Val.**
RMSE	0.0357	0.0886	0.125	0.118
MAE	0.153	0.229	0.307	0.285
MAPE (%)	7.31	11.97	17.12	28.51

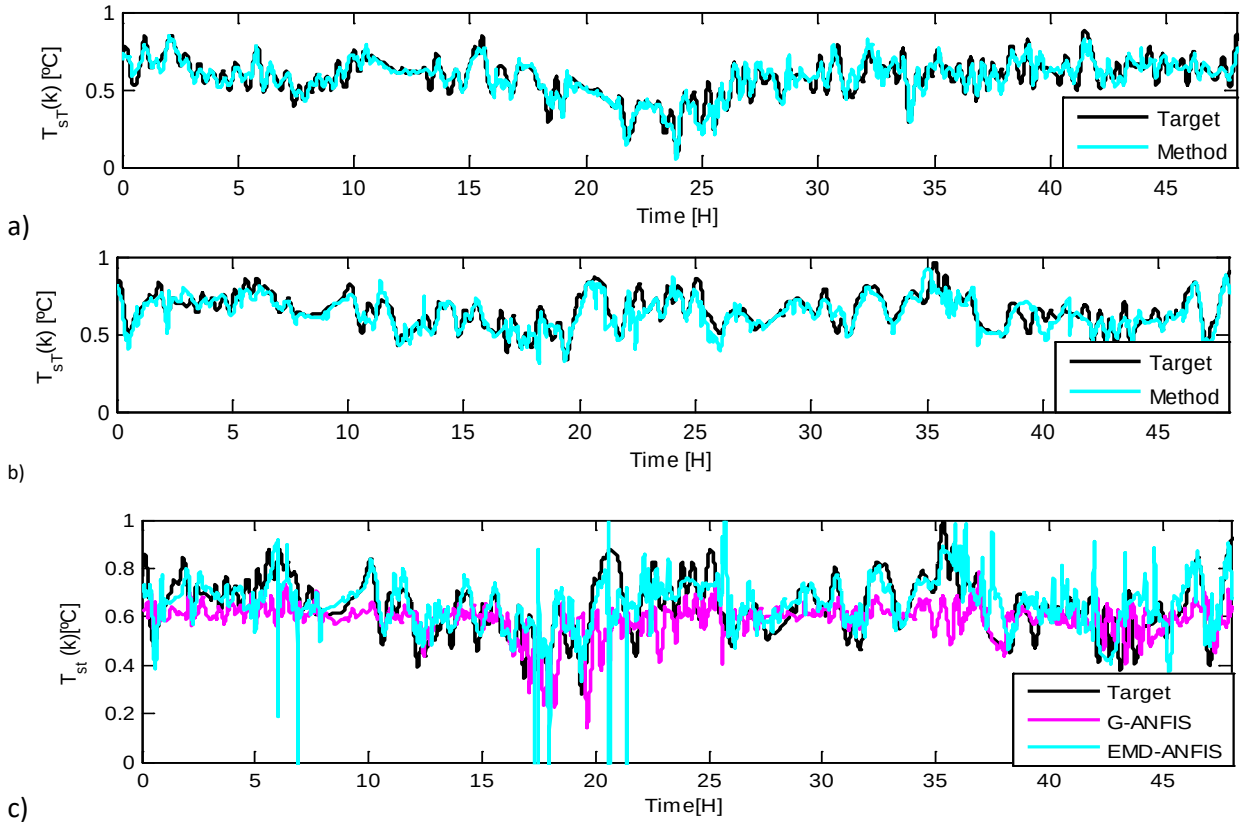

Fig. 3.2.11. Results of the forecasting models: a) Output of the EMD-SOMFIS method for the training set. b) Output of the EMD-SOMFIS method versus the validation set. c) Comparative of the G-ANFIS and EMD-ANFIS after the training process using the validation set when z1=5.

The results show that the proposed modelling methodology fits significantly both training and validation data sets with a MAPE error lower than 15%. However, although this performance represents competitive forecasting capabilities, it should be noted that most of the error is located at higher frequency components contained in the $Pk_2(t)$. Also, the low values shown by the RMSE implies significant generalization capabilities, that is, a smoothed response versus the outliers.

The performance of the proposed method has been compared with the two classical approaches corresponding to the state of the art: first, a single model approach and, second, one model for each IMF approach.

- **Description of the classical approach based on ANFIS modelling with input selection method**

The first classical modelling approach, a single ANFIS model, uses an input selection method based on a GA in order to select the most suitable inputs. According to the literature, the cost function is usually based on the MAPE estimation of the model against the validation data set [69], [94]. In this study the GA has been configured to select the best delayed inputs from [$T_{tu}(t)$, $T_{tu}(t- z_1)$, $T_{tu}(t- z_2)$]. The chromosomes of the GA are configured in regard with the kind of input. In this regard, limits of the GA have been configured to vary between 1 and 180 samples. After the application of the GA, the best selected inputs and the structure of the final model is defined in Eq. (3.2.6).

$$\widehat{T_{tu}}(t) = GANFIS(T_{tu}(t), T_{tu}(t-8), T_{tu}(t-25)) \qquad \textbf{Eq. 3.2.6}$$

- **Description of the classical approach based on EMD and ANFIS modelling for each IMF**

In order to implement the second method, a multi-model approach, the tundish's temperature is decomposed in IMF. A dedicated ANFIS model is generated for each IMF, which means a total number of 14 models [108], [111]. The model for each IMF uses as inputs the current value of the *k-th* IMF, $IMF_k(t)$, and the best delay found in the correlation of the target signal, $IMF_k(t-z1)$. Accordingly, the model is defined as in Eq. (3.2.7).

$$\widehat{T_{tu}}(t) = \sum_{k_i=1}^{k} ANFIS(IMF_k(t), IMF_k(t-z_1)) \qquad \textbf{Eq. 3.2.7}$$

- **Results of the comparison between proposed methodology and classical approaches**

The first method, single ANFIS modelling, exhibits the worst results for both training and validation datasets with a MAPE greater than 15% in both datasets. As expected, by means of a single model, it is difficult to forecast the behaviour of such a non-linear time series. Furthermore, as the cost function is made in regard with the validation set, the model shows an overfitted response for the validation. This fact, causes the model to approximate better the validation set than the training set, that is, MAPE in validation is lower than in the training.

In regard with the second method, it performs slightly better during the training procedure. As it can be seen in the table, it only presents a MAPE of 7.18%. However, the model exhibits a loss of performance when dealing with the forecasting of the validation set, as can be appreciated in **Fig. 3.2.11 (c)**. This means that the model suffers a lack of generalization by predicting each IMF individually compared with the proposed method.

Table 3.2.2. Performance of the Statistical Error Metrics

	EMD - SOMFIS		**EMD - ANFIS**		**G- ANFIS**	
Merit Factor	*Trn.*	***Val.***	***Trn.***	*Val.*	*Trn.*	*Val.*
RMSE	0.055	**0.076**	**0.049**	0.800	0.103	0.127
MAE	0.307	**0.229**	**0.194**	0.302	0.279	0.314
MAPE (%)	7.95	**8.12**	**7.18**	16.75	15.32	26.48

Regarding the computational cost of the on-line functioning of the algorithm, a regular computer with a CPU model Intel i7 3820 and 8 GB of RAM has been used to run the experiments. The average CPU time has been measured inside the algorithm. The results can be seen in **Table 3.2.3**. The specific time associated to

the different processes that are common to each algorithm are: IMF decomposition 12.29 ms and ANFIS evaluation 2.16 ms. The total computational time obtained in all the cases is inferior to the acquisition time of 10 seconds, validating with it the on-line implementation of the proposed algorithm in the industrial manufacturing plant.

As it was expected, the G-ANFIS requires the less computational cost since its execution is primarily limited to the evaluation of a single ANFIS model. However, the performance of the single model is unaffordable for a critical magnitude such the tundish's temperature. Furthermore, the proposed method requires over 34% less computational cost than the classic approach of a single model for each IMF and achieves better performance with other sets than the training one.

Table 3.2.3. Computational cost of each evaluated algorithm.

	Total CPU Time	
EMD-SOMFIS	*261.48*	*ms*
G-ANFIS	*10*	*ms*
EMD-ANFIS	*397.08*	*ms*

Discussion and conclusions

This experimental validation exhibits the benefits of the proposed dynamics based industrial time series forecasting method applied to complex industrial processes monitoring. The proposed method takes advantage of the EMD to adaptively pack signal's dynamics in regard with their significance and their modelling error. Finally, the forecasting is faced by means of a collaborative ANFIS strategy. The method has been experimentally validated by means of industrial data from a copper rod manufacturing process.

The proposed method has been successfully applied to forecast the evolution of the tundish's temperature in terms of low error and high generalization capabilities for a predefined prediction's horizon of 15 minutes. The results indicate that the method is suitable to be applied in industrial processes where fastest dynamics in the range of seconds, which represents a wide range of industrial applications.

Furthermore, a comparative with the classical methods in the state of the art, the input selection based G-ANFIS, and the IMF-ANFIS, has been made. The results show that signal decomposition based methods outperforms input selection methods for industrial process forecasting. This fact is due to the characteristic non-periodic behaviour of industrial manufacturing processes. Thus, the most significant information for forecasting purposes is the signal itself (decomposition), and the operation point of the process (process codification). The proposed method adapts better to the available training dataset increasing generalization capabilities and reducing forecasting complexity.

It should be noted that this methodology could be applied in multiple complex industrial processes. The potential of the dynamic's adaptation and codification stages has the processing capability to extract the relevant information while maintaining generalization capabilities.

This work results in a high performance and advanced prognosis structure to the development of industrial plants system prognosis procedures. This work represents an important study to the introduction of a signal analyses stage previous to any forecasting based methodology for industrial process forecasting.

3.3 Multi-dynamics based time series modelling

Classically, data-driven modelling approaches were faced by the application of the Neural Networks (NN) [112], or the Adaptive Neuro Fuzzy Inference Systems (ANFIS) [113]. In this regard, the high acceptance of the ANFIS based approaches is due to the trade-off between its performance and simplicity. That is, the ANFIS exhibits a significant capability to adapt the Membership Functions (MF) of the input layer by considering the interrelation between the inputs and the target. This crucial step can be understood as a signal selection layer that aims to weight the most significant information, and discard those inputs that do not affect the target [21]. However, this particularization of the model requires high amount of data, slows the convergence of the solution and masks the relations between the inputs and the target. This selection procedure turns into a limitation when the model needs to address a considerable amount of inputs, the signals presents significant non-linear behaviours, and the available data is limited, as is the case of industrial time series.

In this context, the Neo-Fuzzy Neuron (NFN), is a modelling method that offers interesting capabilities to be exploded in the industrial field. It presents an open and modular architecture that generates a dedicated modelling unit for each input that is introduced to the model [114]. Then, the inputs are connected in a collaborative way to give the final outcome. In this structure, the relation of the inputs with the target can be easily identified, this leads to higher solution convergence ratio, which implies lower data requirements and fewer training iterations, this results in a less overfitted approach [26]-[27]. Therefore, the NFN together with a simple correlation study in order to assure the utility of the selected model inputs is proposed as a reliable and accurate solution for industrial process modelling.

Indeed, ANFIS based approaches are the most used dealing with time series modelling due to the good trade-off between tuning simplicity and performance [21]. In this regard, ANFIS architecture is based on a second-order optimization method, which analyses cross-relations between model inputs and target signal. This analysis, although simplifies the design procedure, causes an increase of convergence complexity during the training stage of the model. As a consequence, the amount of data required to assure a proper solution increases exponentially with the number of inputs. Therefore, if not enough data is available, the training iterations will increase, and the model will incur in an overfitting response. This fact might be a limitation when dealing with real industrial problems, in which coherent data from the process is limited to the acquired behaviours.

In order to overcome such limitations, the neo-fuzzy neuron represents a novel and underexploited modelling architecture that allows a modular structure assigning one modelling unit for each input. Then, inputs are connected in a network of non-linear collaborative synapsis to estimate the outcome. This structure presents higher convergence ratios with less over-fitted results, which implies lower amounts of training samples required [117]. The NFN does not presents any input adaptation method, as the ANFIS approach; for this reason, it must be assured that significant and non-correlated inputs are being considered.

Previous research works have been made with NFN. Accordingly, K.T. Chaturvedi et al. [118], proposes the application of NFN for modelling the evolution of economic and power time series. Y. Bodyanskiy et al. [117], uses the NFN for modelling time series properties focusing on chaotic non-linear signals. A. Soualhi et al. [119], proposes the NFN a bearing condition monitoring, which Is the closest application to the industrial field. However, this last application, authors propose to use NFN modelling to track the degradation evolution of the bearings, in such applications, the performance of the model is closely related to the time vector, since it

uses the increase of time to propagate the degradation among time, which application is far from a classical time series forecasting problem. However, in the thesis application, NFN is intended to handle high dynamics signals characteristic of industrial processes, such problematic has not been properly approached with the NFN architecture in the literature.

Therefore, this thesis's section is organized as follows, first, the theoretical considerations of the Neo Fuzzy Neuron are explained in this section 3.3.1. The theoretical basis of the NFN is included in this section since the use of NFN for industrial forecasting modelling is not significantly considered in the current literature. Then, the suitability of NFN for modelling industrial time series are explained in section 3.3.2 and validated with an experimental application to the CRMP. As NFN represents a modelling structure able to deal with multiple inputs, the last part, section 3.3.3, is related to the consideration of the dynamics modes of a target signal into a NFN based modelling structure.

3.3.1 The Neo-fuzzy neuron

The Neo-Fuzzy Neuron (NFN) is a nonlinear multi-input single output modelling architecture that was introduced by T. Yamakawa *et. al.* in [120]–[122]. The structure of a single output, *n* inputs NFN can be seen in **Fig. 3.3.1**.

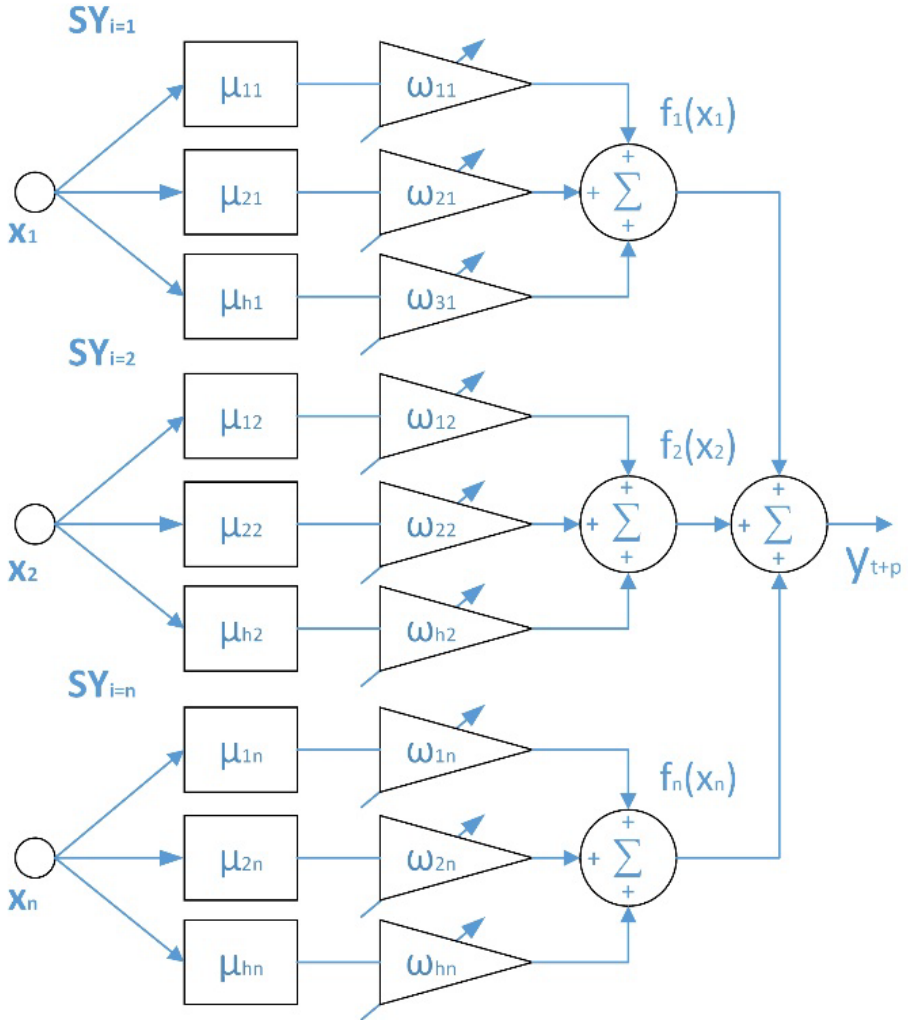

Fig. 3.3.1. Neo-Fuzzy Neuron structure. It represents a network with n=3 inputs, and h=3 MF to cover each input.

Let x_i be the *i-th* input, and $\hat{y}_{t+p}$ the predicted output at a predefined forecasting horizon *p*. In this regard, the output of the NFN is calculated by applying Eq. (3.3.1).

$$\hat{y}_{t+p} = \sum_{i=1}^{n} f_i(x_i) \qquad \textbf{Eq. 3.3.1}$$

The structural modelling units of the NFN are *n* non-linear synapses, SY_i, that connect the inputs of the model with the output by means of $f_i(x_i)$, as shown in Eq. (3.3.2).

$$f_i(x_i) = \sum_{j=1}^{h} \omega_{j,i} \cdot \mu_{j,i}(x_i) \qquad \textbf{Eq. 3.3.2}$$

where $\omega_{j,i}$ represents the interconnecting weights, and $\mu_{j,i}(x_i)$ represents the membership degree of the *i-th* input over *j-th* MF calculated by Eq. (3.3.3), where *h* denotes the number of MF to fuzzify each input.

$$\mu_{ji}(x_{t-i}) = \begin{bmatrix} \dfrac{x_i - c_{j-1,i}}{c_{j,i} - c_{j-1,i}} & x_i \in [c_{j-1,i}, c_{j,i}] \\ \dfrac{c_{j+1,i} - x_i}{c_{j+1,i} - c_{j,i}} & x_i \in [c_{j,i}, c_{j+1,i}] \\ 0 & otherwise \end{bmatrix} \qquad \textbf{Eq. 3.3.3}$$

In this regard, all MF of the NFN are triangular functions obtained after the normalization of the input variables as $x_i (0 \leq x_i \leq 1)$. Then, each input is fuzzified by a set of *h* equidistant triangular MF as shown in **Fig. 3.3.2.** Finally, the output value of the *i-th* synapse is given by Eq. (3.3.4).

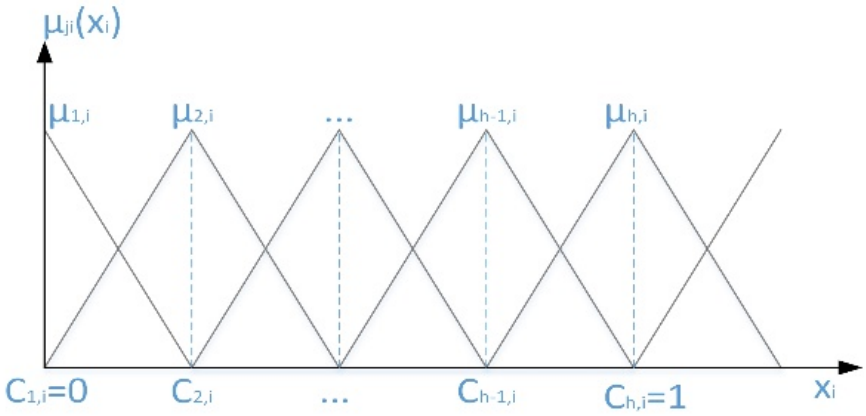

Fig. 3.3.2. Triangular MF that are used for input fuzzification. In the figure, h=5 MF are shown.

Finally, the output value of the synapse is given by Eq. (3.3.4). It should be noticed that *h* should be set in regard with the application, *h* must be increased in regard with target signal complexity.

$$f_i(x_{t-i}) = \omega_{j,i} \cdot \mu_{j,i}(x_{t-i}) + \omega_{j+1,i} \cdot \mu_{j+1,i}(x_{t-i}) \qquad \textbf{Eq. 3.3.4}$$

After fixing the structure, model training is the procedure to adapt, for each synapsis, the *h* x *n* weights of the model. In this regard, NFN uses the Stepwise Training as the learning algorithm to update the weights in regard with the achieved error. Thus, for a time instant, *t*, the training function is defined as the classic quadratic error of the output, shown in Eq. (3.3.5). Indeed, despite ANFIS approach, the weighting matrix of the NFN modelling represents a more direct relation between the modelling outcome and the individual affectation of each input.

$$E(t) = \frac{1}{2}\left(y_{t+p}(t) - \hat{y}_{t+p}(t)\right)^2 = \frac{1}{2}\left(y_{t+p}(t) - \sum_{i=0}^{n-1}\sum_{j=1}^{h}\omega_{i,j}(t-1)\cdot\mu_{j,i}(x_i(t))\right)^2 \qquad \textbf{Eq. 3.3.5}$$

Thus, learning phase can be seen as a optimization problem that aims to find the optimal weight configuration by means of the gradient descendent method, usually a back-propagation approach [123]. The weights are iteratively updated through the configured training iterations by following Eq. (3.3.6), where α is the learning rate that determines convergence speed. Notice that *α* can be adjusted to the model error in order to adapt the convergence speed to the amount of error achieved, and *q* denotes the *q*-th iteration of the learning algorithm.

$$\omega_{j,i}(t+1) = \omega_{j,i}(t) + \alpha\cdot\left(y_{t+p}(t) - \hat{y}_{t+p}(t)\right)\cdot\mu_{j,i}\left(x_i(t)\right) \qquad \textbf{Eq. 3.3.6}$$

Once the model is properly trained, standard statistic metrics can be used in order to evaluate the performance and the accuracy of the NFN model. As it has been aforementioned, the most used metrics in the literature are the root mean squared error, the mean absolute error, and the mean absolute percentage error [76].

Thus, dealing with NFN modelling, the degrees of freedom in regard with its configurations are:

- **The number of inputs:** inputs must be selected by the user with the inputs considered to be significant for the modelling problem. In this regard, it should be considered that NFN does not have any input adaptation method as classical approaches (i.e. ANFIS), and for this reason, it must be assured that each input presents consistent information.
- **Number of membership functions:** the number of membership functions to cover each input, *h*. The affectation of *h* to the modelling procedure is that when *h* is increased, the resolution over the description of the corresponding input is improved. However, having a high number of membership functions implies a large set of weights to be defined during the learning stage; thus, complexity is increased. The typical range is h=[5-20], as the recommended interval for most applications [124].
- **Learning convergence speed:** the convergence speed of the training algorithm, *α*. The learning rate, *α*, determines the convergence speed of the back-propagation training method. High values of *α* means faster convergence and a better adaptation of the datasets with fewer iterations. Yet, increasing convergence speed can drive the model to an unstable training. Consequently, when dealing with industrial signals, which are usually rich in terms of dynamics, a learning rate between *0.001< α <0.05* is recommended to assure the model stability [125].

- **Learning iterations:** The training iterations, *iter*. In this regard, the weights of the synapsis will be adapted by introducing the training data a defined number of iterations, as the number of iteration increases, the same data is passed through the model that results in a more overfitting approach and a loss of generalization in the output.

3.3.2 Industrial time series modelling with NFN

The main advantages of NFN method are a high learning rate with a limited database, computational simplicity, easy input handling, and the avoiding of overfitting in the output [119]. In this regard, the NFN together with a simply correlation study is proposed in order to in order to assure the significance of the selected model inputs.

In this experimental validation of the NFN, first, the available auxiliary signals of the database are analysed by means of the correlation analysis proposed in section 2.3.1. This correlation analysis aims to detect the appropriate set of inputs to the model. Then, the NFN is configured and trained to learn the behaviour of the target signal. Finally, the proposed method is compared in terms of performance with a classic ANFIS approach.

Hence, the objective of this experimental validation is to model and forecast the refrigeration index of the casting wheel, $R_{ind}(t)$. This index is an indirect measure of the effectiveness of the refrigeration among time. This magnitude is critical in the manufacturing process because represents the temporal heat extraction from the melted cooper during the casting procedure. Deviations in the refrigeration imply imperfections in the manufactured copper rod due to non-uniformities in the copper density. The modelling and forecasting of such index must allow the corresponding actions to avoid the affectation to the next manufacturing batch. This constrain fixes the forecasting horizon, p, of the model to 15 minutes (90 samples).

Available information in regard with the heat extraction process can be divided in two sub-processes, the water refrigeration process, W_{sP1}, and the acetylene painting process, A_{sP2}, both containing 9 auxiliary signals. The temporal waveform of W_{sP1} signals and their description can be seen in **Fig. 2.3.3**. Regarding the signals from the acetylene, temporal waveform and their description can be seen in **Fig. 2.3.4**. After the proposed correlation based signal selection process exposed in section 2.3.1 with the same dataset, the final pool of auxiliary signals that are correlated with the target and will be used as inputs of the NFN are: W_{sP1} (t, 3), W_{sP1} (t, 6), W_{sP1} (t, 7), W_{sP1} (t, 8), A_{sP2} (t, 2), A_{sP2} (t, 4), A_{sP2} (t, 6). The description of the signals of the CRMP is shown in Annex I of this thesis.

- **NFN model design and application**

According to the experimental application, the first important aspect to define are the inputs of the model. Considering the auxiliary information from the previous step, and information regarding the dynamics of the signal, the next inputs of the model are proposed:

- Input 1 - The current value of the refrigeration index, $R_{ind}(t)$.
- Input 2 and 3 - Two past values of the refrigeration index delayed z_1 and z_2 samples respectively, $R_{ind}(t\text{-}z_1)$ and $R_{ind}(t\text{-}z_2)$. z_1 and z_2 values are set by means of a GA optimization algorithm, the chromosomes of the algorithm are programmed to vary in the 1-180 samples range to find the best inputs. After the GA, the found values are z_1= 46, z_2=75 delayed samples respectively.

- Input 4 - Mean value of the signal in the last 60 min, $\bar{R}$ *(t).* This input is calculated to provide information regarding the low dynamics of the signal.
- Input 5 - Linear slope of the signal in the last 60 min, *M (t),* this input gives information regarding the variation of the signal in the last time interval.
- Input 6 and 7 - Auxiliary inputs found during the correlation analysis, W_{sP1} (t, 3), W_{sP1} (t, 6), W_{sP1} (t, 7), W_{sP1} (t, 8), A_{sP2} (t, 2), A_{sP2} (t, 4), A_{sP2} (t, 6). The output of the model is the refrigeration index among the forecasting horizon *p=90* samples*,* R_{ind} *(t+90).* In this regard, due to the characteristics of NFN, the number of synaptic nodes of the NFN is equal to the number of inputs of the model. Therefore, there are *n=12* synaptic nodes to model the signal.

The output of the model is the refrigeration index among the forecasting horizon *p=90* samples*,* R_{ind} *(t+90).* In this regard, due to the characteristics of NFN, the number of synaptic nodes of the NFN is equal to the number of inputs of the model. Therefore, there are *n=12* synaptic nodes to model the signal.

For this application, *h* is set to 15 MF per input, which is a usual value within the *h=[5-20]* recommended interval [124]. With the inputs and *h* parameters, the structure of the model is fixed. In order to face learning, the parameters of the training algorithm must be set. In this regard, as the NFN is trained by means of the backpropagation method, the NFN learning rate, α, that adjust the learning convergence of the algorithm must be set. For this application, α = 0.003. In this sense, the number of training iterations is set to 10 which offers a good trade-off between generalization and adaptation to the training set.

After the training phase of the NFN algorithm, the output of the model versus the training set can be seen in **Fig. 3.3.4**, and in **Fig. 3.3.5**, versus the validation set. Error performance metrics of the NFN algorithm are shown in **Table 3.3.1**. As can be seen in both figures, the NFN model is able to fit correctly the datasets achieving a MAPE error lower than 10% in both cases. However, most of the error in the validation set is concentrated on the two sudden changes that presents the process by 19h and 36h that can be considered as disturbances in the process.

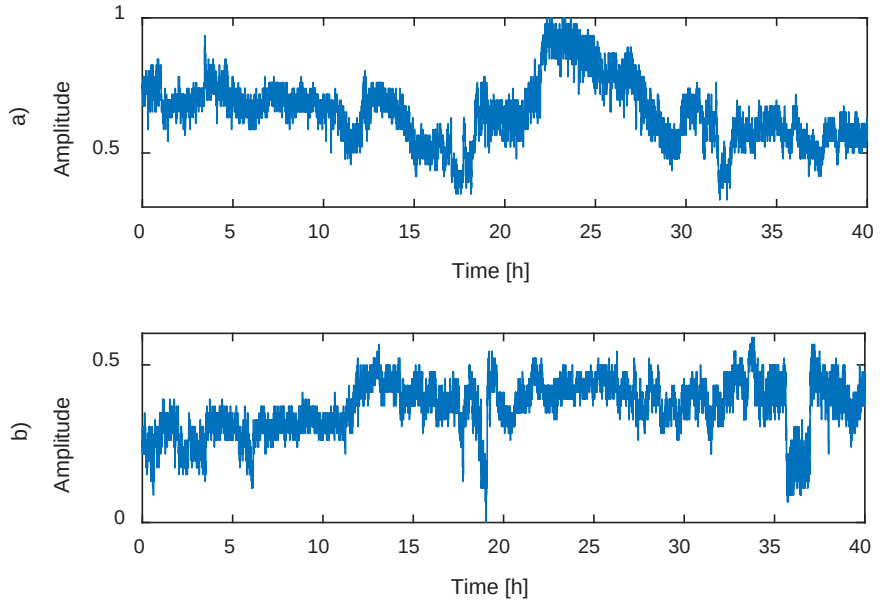

Fig. 3.3.3. Acquired refrigeration index, R_{ind} *(t)*, from the manufacturing plant in two periods of 40 hours of operation. a) Corresponds to the training set, while b) to the validation set.

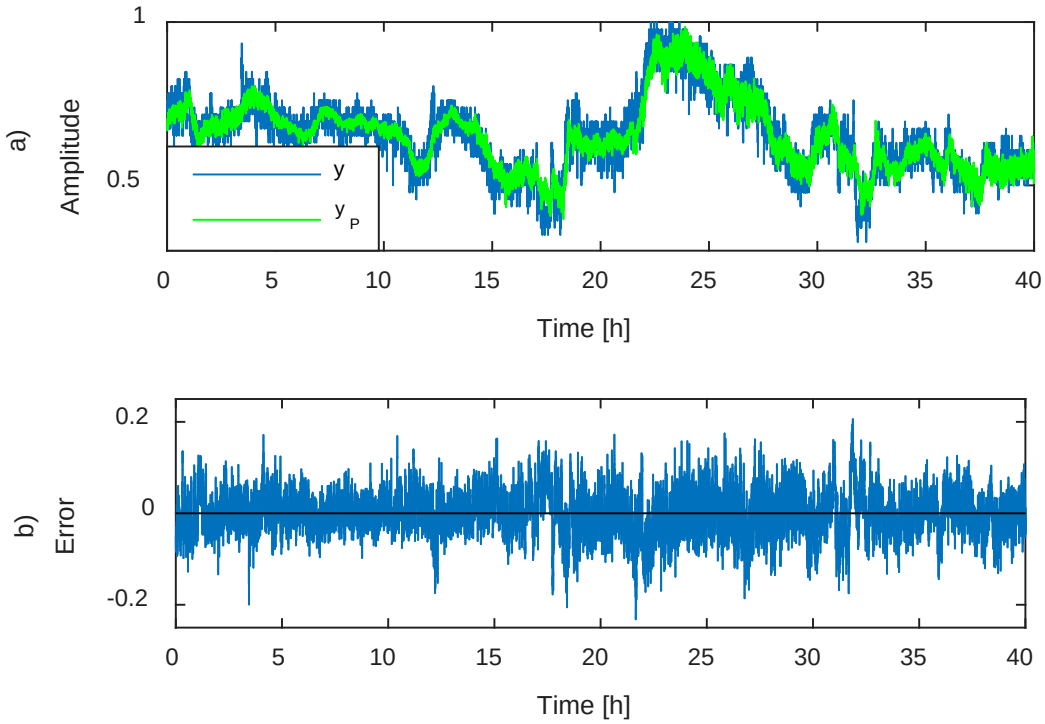

Fig. 3.3.4. Result of the NFN model versus the training set: a) temporal waveform of the signal, *y* is the target and y_P the model output b) achieved error.

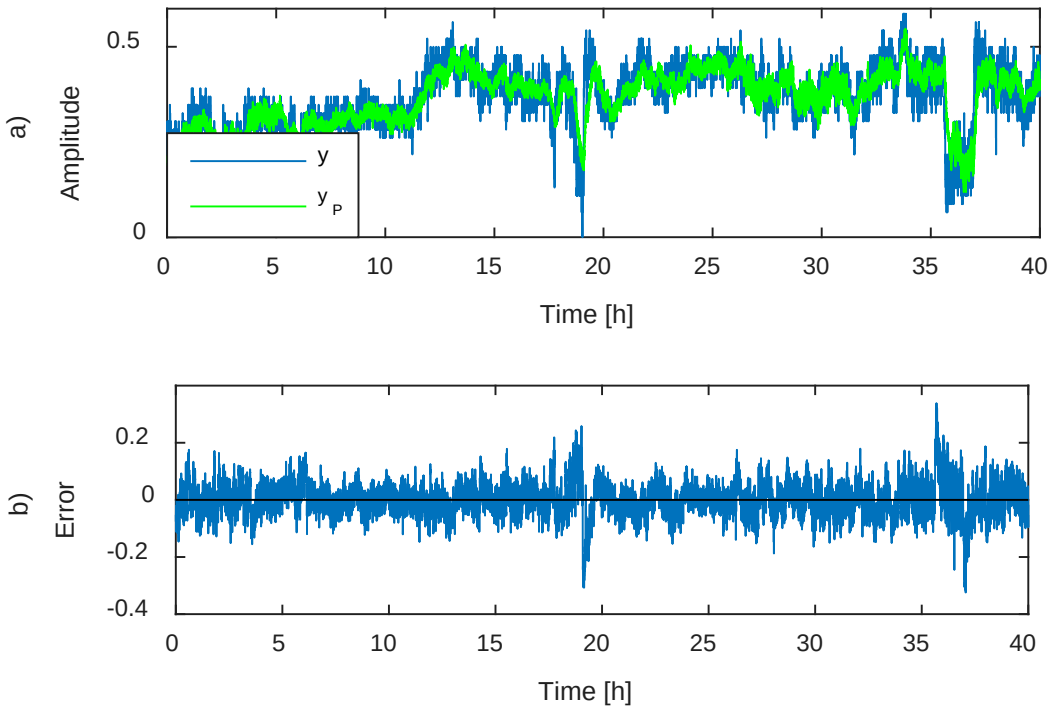

Fig. 3.3.5. Result of the NFN model versus the validation set: a) temporal waveform of the signal, *y* is the target and y_P the model output b) achieved error.

- **Comparison of NFN based scheme with classical ANFIS approach**

Same data has been used to compare the competency of the proposed NFN modelling approach in front of a classical ANFIS based approach. The G-ANFIS modelling strategy uses an input selection method based on a GA in order to select the most suitable inputs for a predefined cost function. According to the literature, the cost function is usually based on the MAPE estimation of the model against the validation set [69], [94].

In this study the GA has been configured to select the best inputs from the same set as the NFN model. The chromosomes of the GA are configured in regard with the kind of input. For the first, the past value, the limits of the GA have been configured to vary between 1 and 180 samples, for the rest of signals, binary inputs are used in order to incorporate (1) or discard the signal as an input of the model. After the application of the GA, the best selected inputs are: (i) current value, $R_{ind}(t)$. (ii) A past value of the index, $R_{ind}(t\text{-}56)$. (iii) Mean value of the signal in the last 60 min, $\bar{R}(t)$. (iv-v) auxiliary inputs W_{SP1} (t, 8), A_{SP2} (t, 2). Inputs are fuzzified by means of three generalized bell-shaped membership functions. The model is trained for 15 epochs by means of the classical hybrid learning algorithm, which is the combination of the least-squares method and the backpropagation gradient descent method. Results can be seen in **Fig. 3.3.6** and **Fig. 3.3.7** for both training and validation sets, performance metrics are shown in **Table 3.3.1.**

Table 3.3.1. Performance metrics of both NFN and GANFIS algorithms

	NFN		**G- ANFIS**	
Merit Factor	*Trn.*	*Val.*	*Trn.*	*Val.*
RMSE	0.047	0.063	0.053	0.095
MAE	0.191	0.228	0.204	0.316
MAPE (%)	5.53	9.51	6.58	16.14

As can be seen in both training and validation responses, the NFN presents a smoother response which is concentrated in following the mean signal value without incurring in the generation of outliers. This fact is reflexed in the low achieved RMSE value of the algorithm, which is more sensitive to samples displaced from the mean value.

Furthermore, the error in the training set presents a Gaussian distribution for both NFN and G-ANFIS modelling methods, which indicates the goodness of both methods (MAPE<7% in training). In this regard, the G-ANFIS method also achieves a good response in the training set, but it presents a more overfitting response in the validation set. This fact can be appreciated in **Fig. 3.3.7 (b)** as an increase of the error when a sudden change in the process is appreciated (around 19h and 36h of plant operation). This error increase is due to the over-adjustment of the ANFIS model to the signals. This results indicate and the suitability to apply NFN to the modelling of industrial time series.

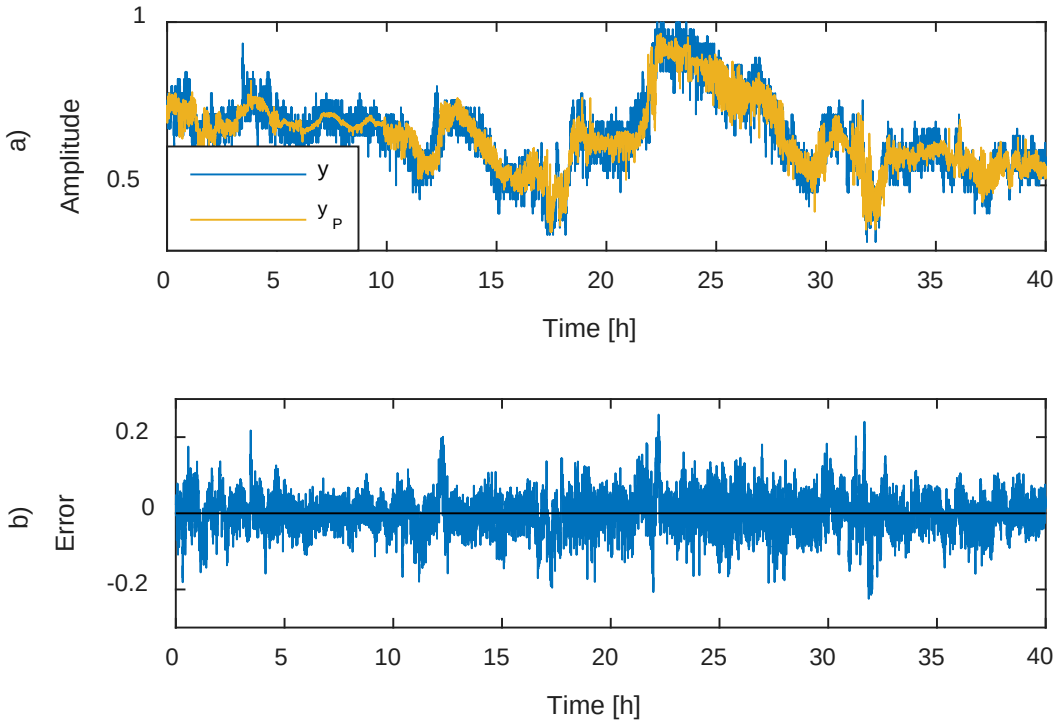

Fig. 3.3.6. Result of the G-ANFIS model versus the training set: a) temporal waveform of the signal, *y* is the target and y_P the model output b) achieved error.

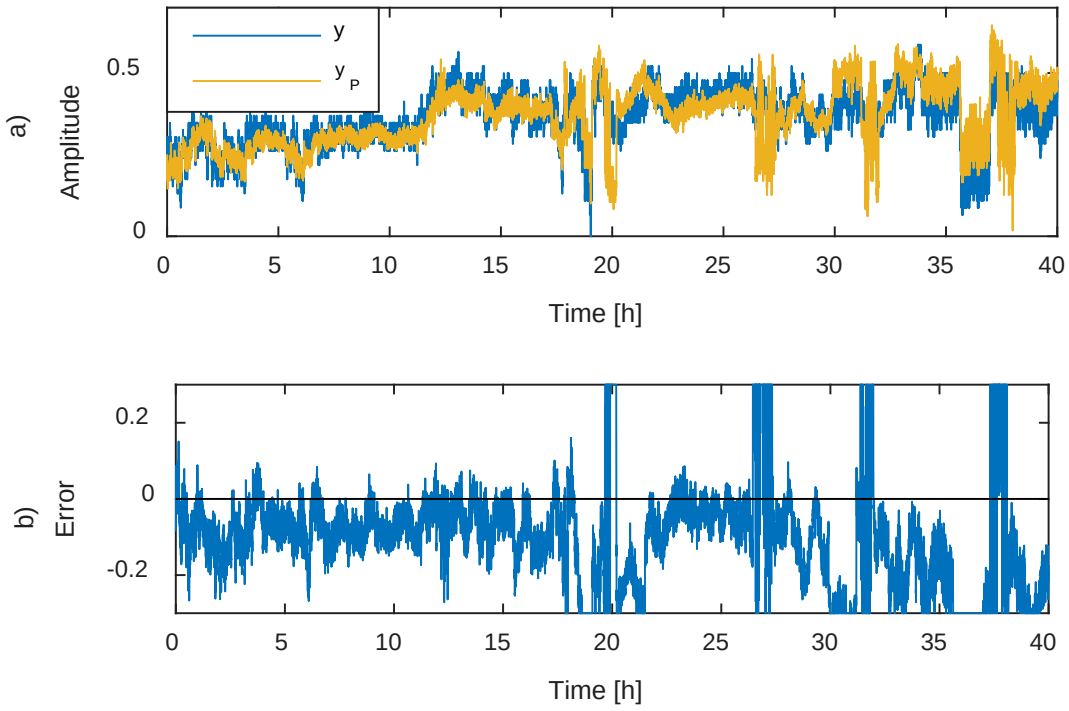

Fig. 3.3.7. Result of the NFN model versus the validation set: a) temporal waveform of the signal, *y* is the target and y_P the model output b) achieved error.

Discussion and conclusions

This experimental validation presents the Neo-Fuzzy Neuron (NFN) as a solid and robust method for modelling industrial time series. The proposed approach takes advantage of the process correlations in order to select the most suitable set of inputs to enhance performance of a NFN method. The proposed method has been successfully applied to forecast the evolution of the refrigeration index from a copper manufacturing plant in terms of low error and high generalization capabilities for a prediction horizon of 15 minutes. The results indicate that the method is suitable to be applied in industrial processes which fastest dynamics is measured in seconds, which is an affordable range of industrial applications. Note that is the first time that NFN has been applied to the modelling of industrial processes.

Furthermore, a comparative with the classical ANFIS method has been made. The results show that the NFN method adapts better to the available training dataset increasing generalization capabilities towards the validation set while reducing modelling complexity.

However, further modifications in the NFN algorithm can be made in order to enhance the results when dealing with industrial time series, such modifications include the particularization of the MF to the data density ranges of the signal, or the dynamic learning rate adaptation. In this regard, the MF of the neo-fuzzy neuron method are placed equidistantly among the 0-1 range considered. However, if a signal with high dynamic content presents small and quick variations among a certain sub-range of values within the 0-1 interval, it will require more resolution in the sub-range of values. For this reason, it could be interesting to adapt the MF of the NFN of the input data distribution, in order to gain resolution in those specified areas. However, the adaptation of the MF should not be too restrictive, since the model will be over-adapted to areas that concentrates the high data density ranges, and the maximums and minimums of the signal will not be properly modelled.

3.3.3 Multi-dynamics consideration in NFN modelling

The multi-input handling capabilities of the NFN can be exploited in order to consider the extracted dynamics of the signal in the modelling to overpass classical ANFIS performance in industrial time series modelling. Thus, this section is focused on this approach, in order to provide a solid forecasting methodology for industrial time series that decompose the dynamics of the signal in a set of independent modes, and consider them as inputs of a single NFN model.

The proposed method sets out a dynamic-based modelling approach to improve the corresponding forecasting performance metrics. Indeed, the method consists on taking advantage of advanced signal dynamics decomposition approaches, as the EEMD, in order to extract low-correlated IMF to be used later as inputs of a multi-input and non-linear modelling structure of the Neo-Fuzzy Neuron. Accordingly, the block diagram of the method is presented in **Fig. 3.3.8**. Thus, for a target signal *y(t),* a NFN-based model considering the resulting *k* IMF is proposed to estimate the corresponding value at a defined horizon *p*, that is, *y(t+p)*.

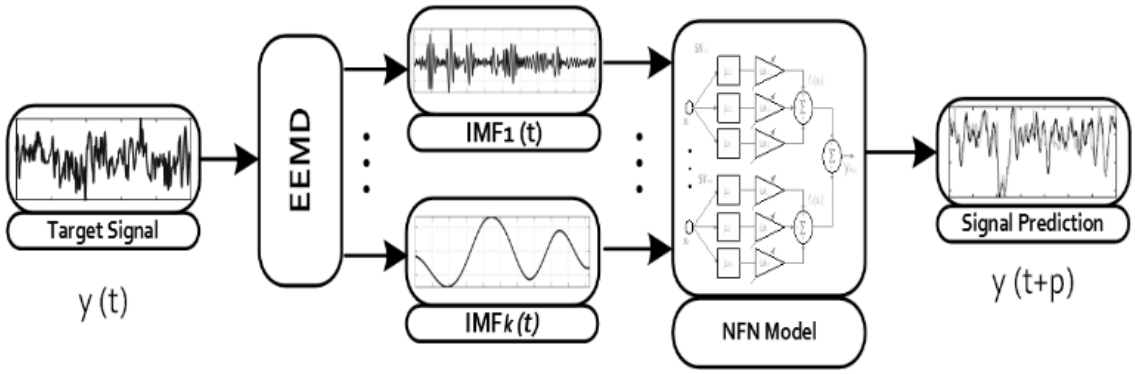

Fig. 3.3.8. Diagram of the proposed methodology, considering the target signal, *y(t)*, its decomposition in *k* IMF by the EEMD, and the NFN based model.

The ensemble empirical mode decomposition is an adaptive noise-assisted method [126], to overcome the limitations of the classic EMD. In this regard, IMFs are obtained by applying the classic EMD to the target signal with successive additions of independent equally-distributed white noise of the same standard deviation [127]. The added white noise is intended to cancel each other together with the intrinsic noise of the signal, significantly reducing the chance of mode mixing and assuring the dyadic property of the decomposition.

The configuration parameters of the EEMD are the number of trials, *nT*, the number of iterations of the EMD algorithm, *nIt*, and the common standard deviation of the white noise, *sD*. Thus, considering the target signal, *y(t)*, after the application of the EEMD, *k* independent IMFs are obtained as specified in Eq. (3.3.7). Note that *k* depends on the analysed signal, and thus the number of IMF varies with the application. In fact, such adaption capability of the decomposition is a desired feature in order to estimate the most significant sub-signals.

$$y(t) = \sum_{k=1}^{K} IMF_k(t) \qquad \textbf{Eq. 3.3.7}$$

Considering a set of IMF containing the most significant dynamics of the target signal, the structure of the NFN model is defined as shown in Eq. (3.3.8), since IMF are the only inputs considered, the synaptic nodes of the network are directly the *k* extracted IMF, so for any application, SY_n=*k*.

$$\hat{y}(t+p) = NFN\left(\sum_{k=1}^{K} IMF_k(t)\right) \qquad \textbf{Eq. 3.3.8}$$

Therefore, the proposed method, that corresponds to a novel contribution to the multi-dynamics based modelling in industrial processes, presents a high-performance industrial time series forecasting scheme supported by two main novelties. First, a frequency based modelling approach by means of a noise-assisted decomposition of the target signal. Second, the consideration of a suitable modelling architecture dealing with non-correlated and significant modelling inputs, the Neo-Fuzzy Neuron. The extraction of the dynamic modes from the target signal is faced by means of the Ensemble Empirical Mode Decomposition. The resulting Intrinsic Mode Functions represent inputs in a Neo-Fuzzy Neuron based model.

Considering that the proposed method includes a novel dynamics consideration in a NFN architecture, it is necessary to extend the validation to an additional case study, a reference benchmark of an industrial process of the literature. Therefore, in order to evaluate the competency of the method and its benefits in the modelling

of industrial time series, the proposed method is validated over a reference case study, the Tennessee Eastman Process (TEP) and, also, over real data from a Copper Rod Manufacturing Process. In this regard, the performance of the proposed method is tested, first, over the Tennessee Eastman Process (TEP), a detailed model of an industrial process created by the Eastman Chemical Company for evaluating process control and monitoring methods [128]. Such process allows the assessment of the proposed method in a reference benchmark widely used by the scientific community. Furthermore, in order to demonstrate the impact and applicability of the proposed method, the tests have been extended to a second data set, the Copper Rod Manufacturing Process, corresponding to the real industrial process considered as a benchmark for the development of condition forecasting methodologies of this thesis.

- **First case study: Tennessee Eastman process**

The TEP consists of five main units: the reactor, the product condenser, the vapour-liquid separator, the recycle compressor and the product stripper. It contains eight components named from A to H. The gaseous reactants, A, C, D and E, are introduced in the reactor in order to form the liquid products G and H by an irreversible exothermic reaction. The diagram of the TEP can be seen in **Fig. 3.3.9**, it contains two blocks of variables: 12 manipulated variables and 41 measured variables. Detailed information about the model can be obtained from [128]. Different versions have been made of the mathematical model of the TEP process, in this thesis, the version proposed by [129] is used.

From the process monitoring point of view, it is mandatory to know the internal status of the reactor unit to assure the proper functioning of the chemical process. In this regard, the objective of this application is to forecast the reactor pressure signal, $P_{re}(t)$, in order to identify deviations that might cause malfunctioning in the system.

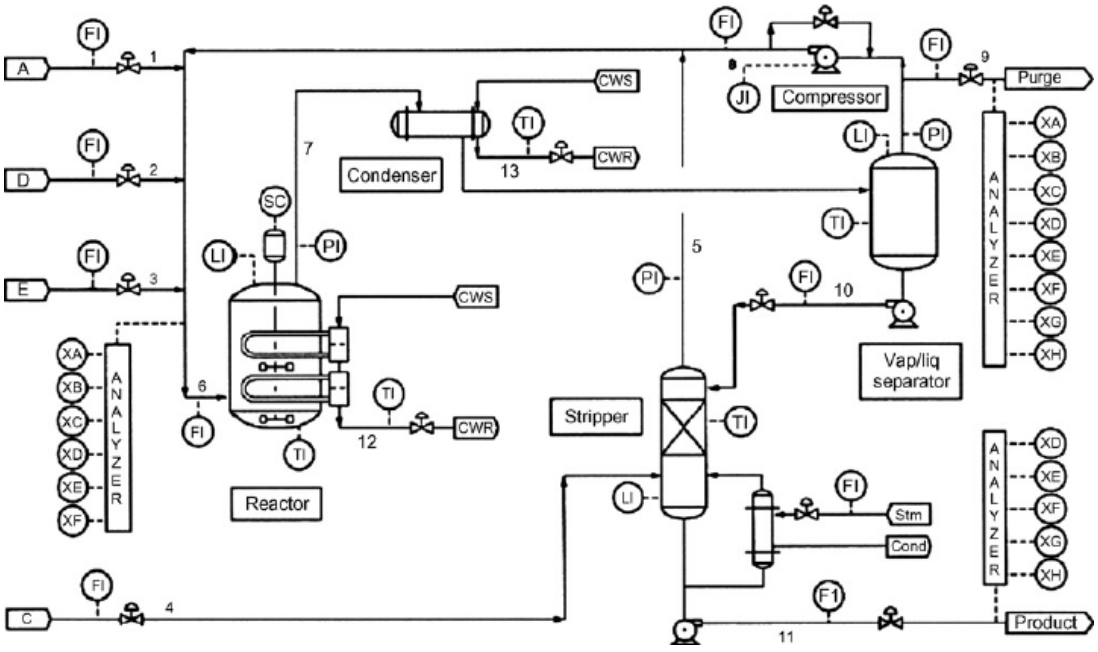

Fig. 3.3.9. Diagram of the Tennessee Eastman Process

The temporal waveform of the reactor pressure, $P_{re}(t)$, can be seen in **Fig. 3.3.10**. The sampling frequency is 1/36 Hz, and the simulation time has been configured to be 70 hours, where the first 35 hours will be used for training purposes, and the last 35 hours for validation tasks.

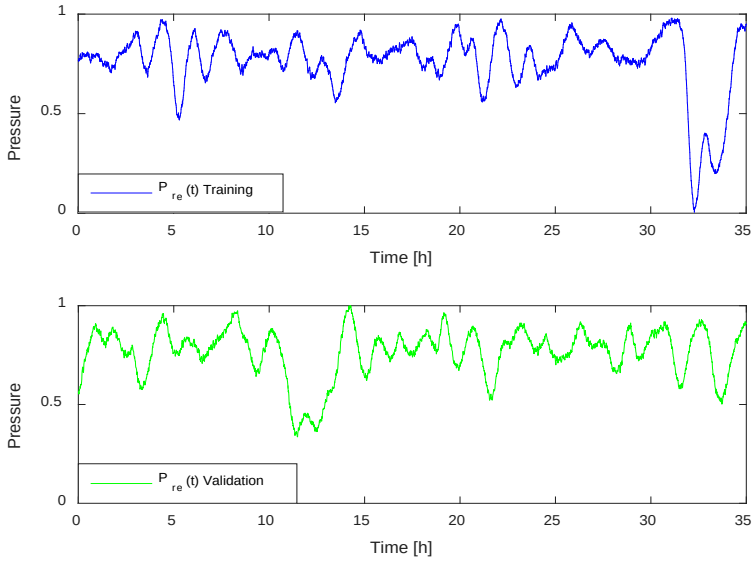

Fig. 3.3.10. Scaled pressure of the reactor between [0,1], $P_{re}(t)$, among the 70 hours of operation time.

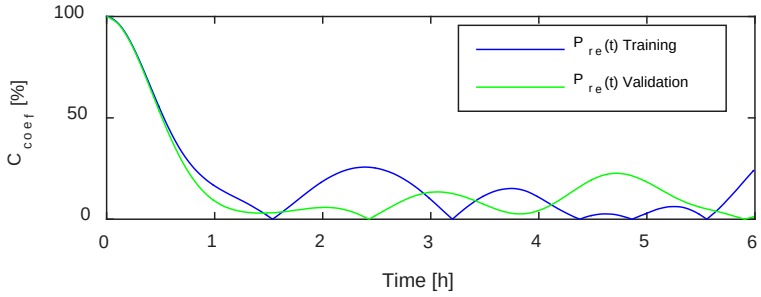

Fig. 3.3.11. Correlation analysis of the Pre (t) for both dataset partitions. At 35.4 minutes, the correlation is 40.39% for the training set.

In order to select the optimal forecasting horizon, the proposed methodology explained in section 2.2.1 based on the autocorrelation analysis has been used. In this regard, the drop of correlation caused by the extension of the forecasting horizon is shown in **Fig. 3.3.11**. As expected, the dynamic contents of the signal are highly related with short-term dynamics, above 75% of correlation until previous 20 minutes, and low related with long-term dynamics, below 10% from 1 hour. Thus, the forecasting horizon has been set to 60 samples of reactor operation, which corresponds to a good trade-off between the forecasting utility, 36 minutes of forecasting, and correlation viability, around 40%.

Considering the defined NFN model structure of the method, the outcome to forecast is reactor pressure at the predefined forecasting horizon of p equal to 60 samples or 36 minutes, $P_{re}(t+p)$, the resulting expression of the model is shown in Eq. (3.3.9).

$$\widehat{P_{re}}(t+p) = NFN\left(\sum_{k_i=1}^{k}\left(IMF_k(t)\right)\right) \qquad \textbf{Eq. 3.3.9}$$

The inputs of the model are the resulting IMF from the decomposition process. Indeed, as it has been aforementioned, the dynamics of the reactor pressure signal are extracted in intrinsic mode functions by means

of the EEMD algorithm following [126]. The decomposition has been set with a noise standard deviation of 0.1, 1000 realizations and 5000 maximum iterations during the shifting process.

The resulting thirteen IMF for both dataset partitions are shown in **Fig 3.3.12**. Therefore, the number of synaptic nodes is *n=13* independent modelling units. The number of memberships, *h*, has been set to 15 MF per input. The classical backpropagation algorithm has been selected for the training procedure. Learning rate was set to α = 0.005, and the number of training iterations to adjust the weights of the network has been set to 25.

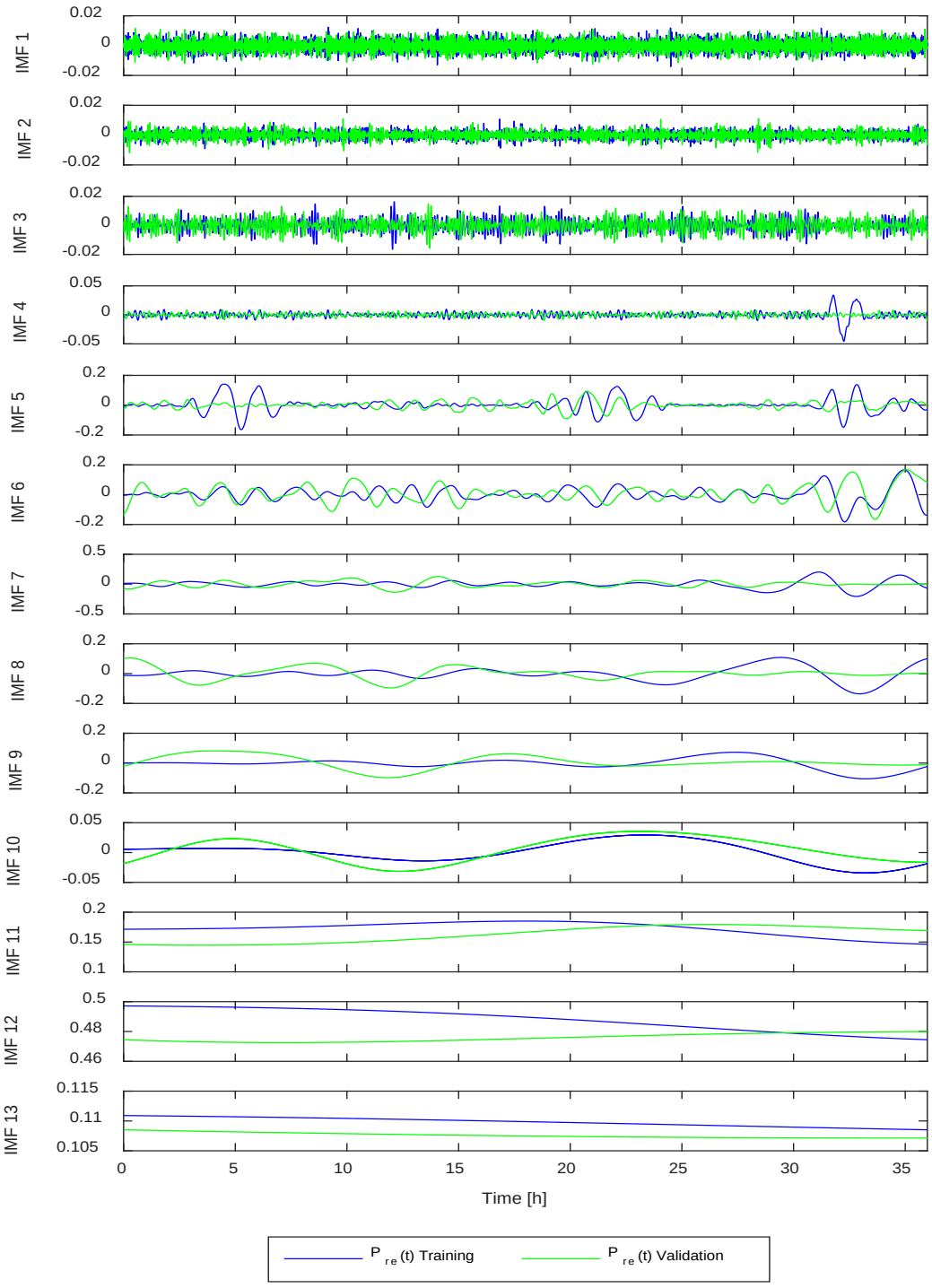

Fig. 3.3.12. Resulting EEMD decomposition of $P_{re}(t)$ for both training and validation datasets.

The output of the model, $P_{re}(t+p)$, during the training stage is shown in **Fig. 3.3.13 (a)**. The result dealing with the validation data partition is shown in **Fig. 3.3.13 (b)**. Performance metrics are summarized in **Table 3.3.2**. As can be seen, the NFN based $P_{re}(t+p)$ model fits correctly both datasets. Indeed, the low achieved RMSE dealing with the validation set, 0.050, together with the MAPE error, 4.87%, indicates that the model response does not generate predictions far from the mean value of the target signal. This is a critical requirement in industrial environments where an alarm will be associated to high/low values of the predicted signal, avoiding with it the generation of false positives. In addition, it can be appreciated that, most of the resulting error is due to two disturbances that causes a sudden drop of the reactor pressure, one around hour 10, and other around hour 12, in the validation set.

However, this fact reflects qualitatively, the significant generalization capabilities of the resulting model, since such behaviours are poorly represented in the training dataset, and, therefore, hardly considered during the model training. It is therefore mandatory to compare the performance of the proposed method with the classical approach in literature. In this regard, current approaches dealing with the consideration of the dynamics of the target signal for forecasting purposes are supported mainly by ANFIS modelling.

In this sense, the resulting sub-signals from a decomposition or filtering stage are, each one, modelled by a specific ANFIS algorithm and, later, the direct combination of all model outputs represents the target signal forecasting [108], [111]. This approach allows improving modelling performance by isolating each dynamic mode of the signal and generating a dedicated ANFIS based model for each IMF

In order to represent such approach, the EMD over the signal has been applied, and the corresponding ANFIS models defined. The ANFIS model for the *k*-th IMF presents the same configuration, that is, two inputs: *(i)* The current value of the *k*-th IMF, $IMF_k(t)$ and, *(ii)* a past value of the $IMF_k(t\text{-}n_k)$. Since the selection of the past value affects the performance of the model, a Genetic Algorithm (GA), has been used to select an optimal delay, n_k, for each ANFIS model. In this regard, the limits of the searching region of the GA have been set between 1 and 60 samples, that is, a maximum delay of 36 minutes. Thus, the mathematical model description of the EMD+ANFIS method is given in Eq. (3.3.10).

$$\widehat{P_{re}}(t+p) = \sum_{k_i=1}^{k=9} ANFIS(IMF_k(t), IMF_k(t-n_k)) \qquad \textbf{Eq. 3.3.10}$$

The EMD application over the reactor pressure signal results in eleven IMFs. Results of the application of this method are shown in **Fig. 3.3.13 (c)** and **Fig. 3.3.13 (d)** for training and validations sets respectively. Model error metrics are summarized in **Table 3.3.2**.

Table 3.3.2. Performance metrics of the proposed method and the comparative method (emd+anfis).

	Method		ANFIS + EMD	
	Trn.	*Val.*	*Trn.*	*Val.*
RMSE	0.009	0.050	0.052	0.076
MAE	0.076	0.186	0.200	0.240
MAPE	1.74%	4.87%	7.39%	9.68%

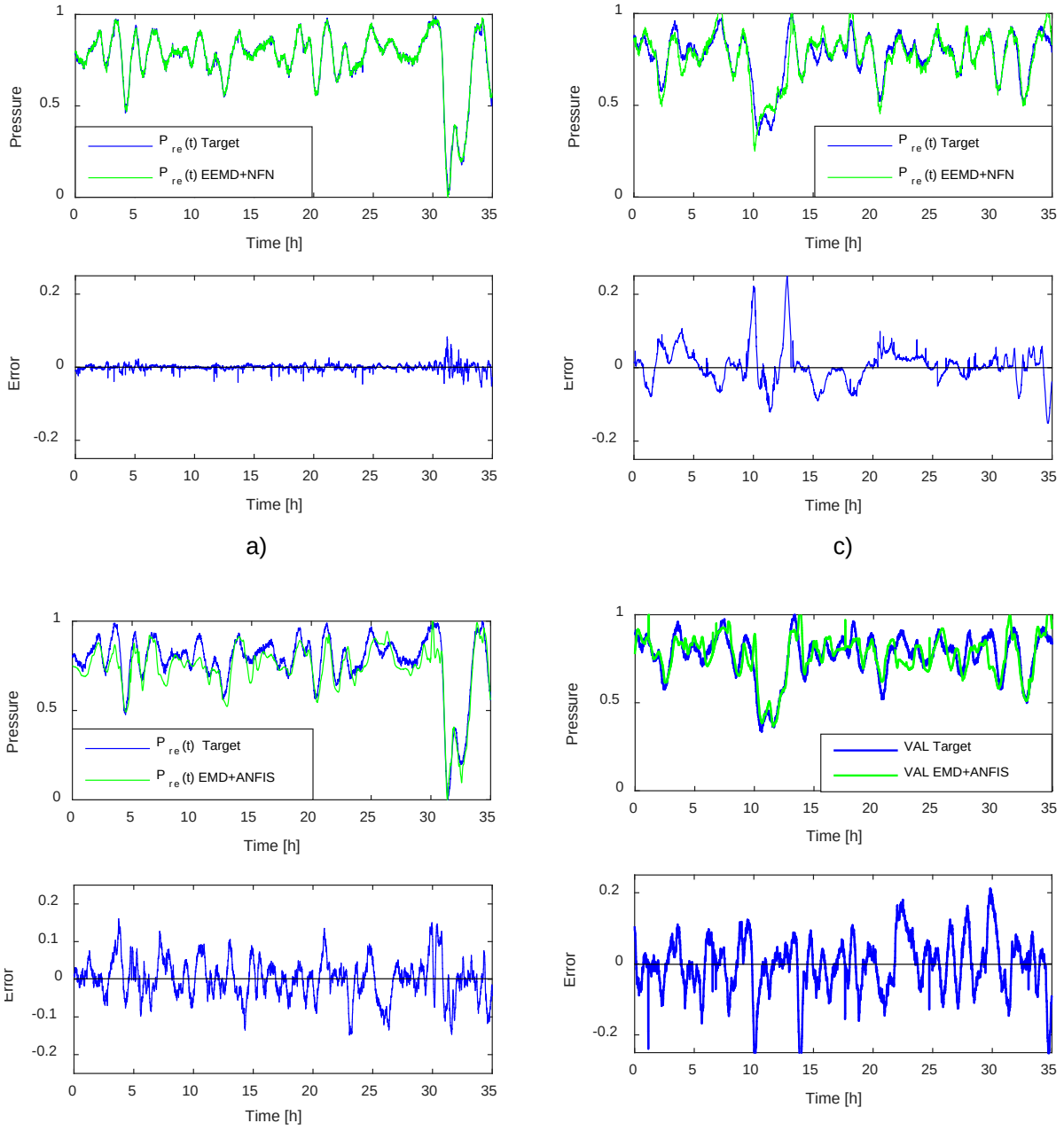

Fig. 3.3.13. Results on the modelling of $P_{re}(t)$ in terms of signal waveform and error. a) Training results of EEMD+NFN. b) Validation results of EEMD+NFN. c) Training results of EMD+ANFIS. d) Validation results of EMD+ANFIS.

It can be appreciated that the ANFIS based method presents acceptable modelling results with a MAPE error lower than 8% in the training set. Furthermore, most of the modelling error is located around the aforementioned disturbances in the reactor pressure, which is expected due to ANFIS convergence limitations with limited datasets. However, one of the main differences in regard with the proposed method is the difficulty for the EMD+ANFIS to follow the validation set. This is due to a common overfitting phenomenon, which is relatively easy to occur when using an ANFIS model. Since modelling efforts are concentrated in finding cross-relations in the input layer, the model convergence towards the optimal solution is slowed.

For this reason, the ANFIS model requires more data than the NFN for achieving a proper non-overfitting training result. This fact can be appreciated in the obtained results, the ANFIS based model is concentrated on tracking the mean value of the validation set, but is not able to define correctly the upper and lower values of the reactor pressure far from the mean. In consequence, the performance achieved in terms of RMSE is correct, and similar to the achieved by the NFN algorithm, but the MAPE error is higher since the instant value of the signal is lost. This problem is magnified if the complexity of the signal to model increases, as it is shown in the next section dealing with CRMP data.

- **Second case study: The Copper Rod Manufacturing Process**

In order to validate the suitability of the proposed method in front of real industrial process, with the considered industrial process in this thesis, the proposed method has been validated by modelling and forecasting the tundish temperature of the CRMP. In this regard, as previous experimental results, the objective of the experimental validation is to forecast the tundish temperature, $T_{tu}(t)$. The modelling and forecasting of such temperature allows reducing the affectation of such deviations to the next manufacturing batch. $T_{tu}(t)$, is acquired continuously at a period of 10 seconds, 0.1 Hz. Two data sets of $T_{tu}(t)$, that corresponds to 50 hours of plant operation are used. Training set is shown in **Fig. 3.3.14.**

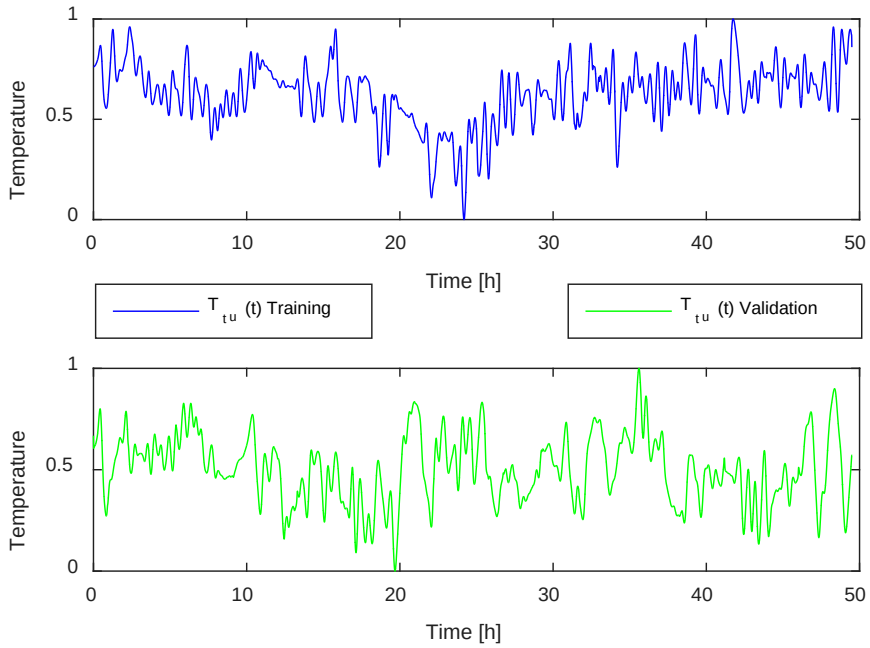

Fig. 3.3.14. Acquired tundish temperature, $T_{tu}(t)$, from the manufacturing plant in a period of 50 hours of operation for both training and validation sets. Signals have been scaled to preserve manufacturing information.

As it has been mentioned, the manufacturing time constrain fixes the forecasting horizon, p, of the model to 90 samples, that corresponds to 15 minutes. The auto-correlation analysis of the tundish temperature shows also the expected behaviour of industrial systems indicated before. It starts with the maximum correlation and the coefficient decays as the forecasting horizon is extended. However, the relation between both training and validation signals is substantially different that the observed in the TEP. As shown in **Fig. 3.3.15**, the training set presents a quickly drop of correlation during the first iterations caused by the presence of high frequency

oscillation modes in the signal. In comparison, the validation set presents a slower decay since the presence of higher frequencies is lower in this dataset. However, after the initial decay, the training set stabilizes and maintains a certain level of correlation, $C_{coef} \approx$ 50%, as time is increased. This is not the case for the validation set, that presents a sustained decrease for all the analysed range, less than 50% from 0.5 hours. This fact indicates that the complexity in terms of signal dynamics of the validation set is higher than the training set. Regarding the forecasted horizon required by the application, 15 minutes, it can be check, in **Fig. 3.3.15**, the viability of such horizon, since C_{coef} is around 45% for both datasets.

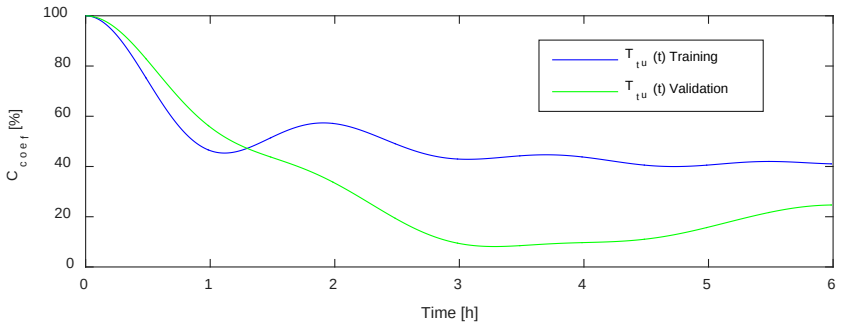

Fig. 3.3.15. Auto-correlation analysis of the tundish temperature. At 15 minutes, the correlation is 46.32% for the training set.

Considering that the output of the model is the tundish temperature at the defined forecasting horizon of p equal to 90 samples or 15 minutes, $T_{tu}(t+p)$, the resulting expression of the model is shown in Eq. (3.3.11).

$$\widehat{T_{tu}}(t+p) = NFN\left(\sum_{k_i=1}^{k}\left(IMF_k(t)\right)\right) \qquad \textbf{Eq. 3.3.11}$$

As shown in **Fig. 3.3.16**, thirteen IMFs result from the EEMD application over the tundish temperature signal. Similar to the configuration used in the previous TEP case study, fifteen membership functions have been considered for each of the IMFs used as inputs. After 25 training iterations, the proposed method results are shown in **Fig. 3.3.17 (a)** and **Fig. 3.3.17 (b)** for both training and validation sets respectively. Performance metrics of the method are shown in **Table 3.3.3**.

The performance of the model in terms of error should be highlighted for the training set, since the model is able to represent the signal dynamics achieving a MAPE error lower than 2%. The highest error is located in the lowest value of the temperature during the disturbance that affects the copper process at twenty-four hours. However, the affectation of such error caused by this disturbance is negligible. Regarding the validation set, it can be seen that the increase of signal complexity causes the error to be biased to the positive part, that means that the model's output is lower than the target signal.

This is caused by the characteristic waveform of the training set, hence the presence of the disturbance during time 20-30h is significant and is learned by the model. Despite this fact, the performance is acceptable considering the differences exposed between both training and validation datasets, and a MAPE error during validation lower than 4%.

Table 3.3.3. Performance metrics of the proposed method and the state of the art (emd+anfis) applied to the copper manufacturing process.

	Method		**ANFIS + EMD**	
	Trn.	*Val.*	*Trn.*	*Val.*
RMSE	0.007	0.0274	0.044	0.0751
MAE	0.066	0.146	0.183	0.228
MAPE	1.26%	3.53%	7.95%	15.08%

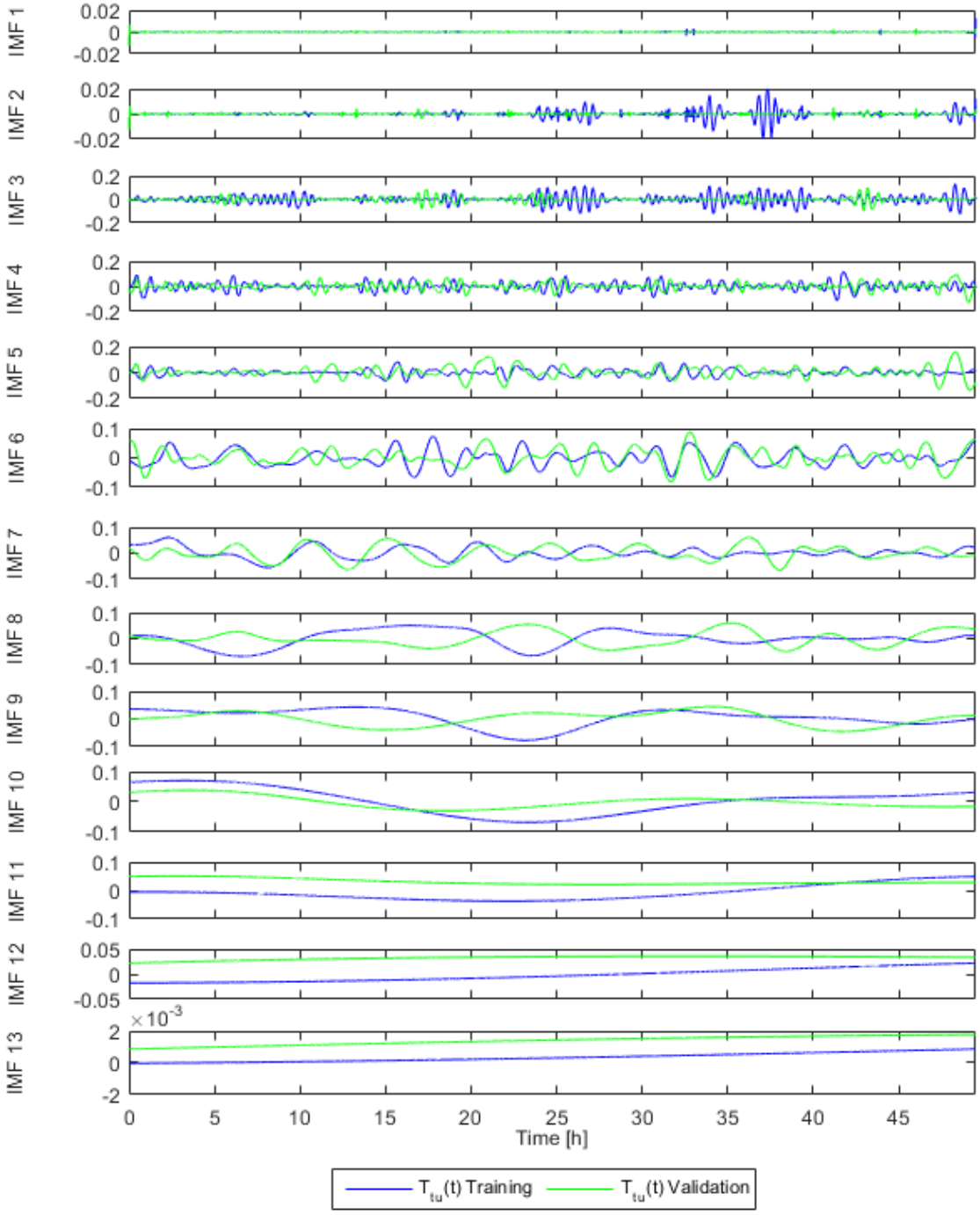

Fig. 3.3.16. EEMD decomposition of the tundish temperature

- **Comparison of multi-modal NFN based scheme with classical ANFIS Approach**

The performance of the proposed method is compared in front of the classical EMD+ANFIS representing the state of the art. The configuration follows previous section description. Thus, the Eq. (3.3.12) shows the mathematical description of the model, after finding 10 IMFs in the decomposition. The results are shown in **Fig. 3.3.17 (c)** and **Fig. 3.3.17 (d)** for both training and validation sets, respectively. The metrics of the algorithm assessment are summarized in **Table 3.3.3**.

$$\widehat{T_{tu}}(t+90) = \sum_{k_i=1}^{10} ANFIS(IMF_k(t), IMF_k(t - n_k)) \qquad \textbf{Eq. 3.3.12}$$

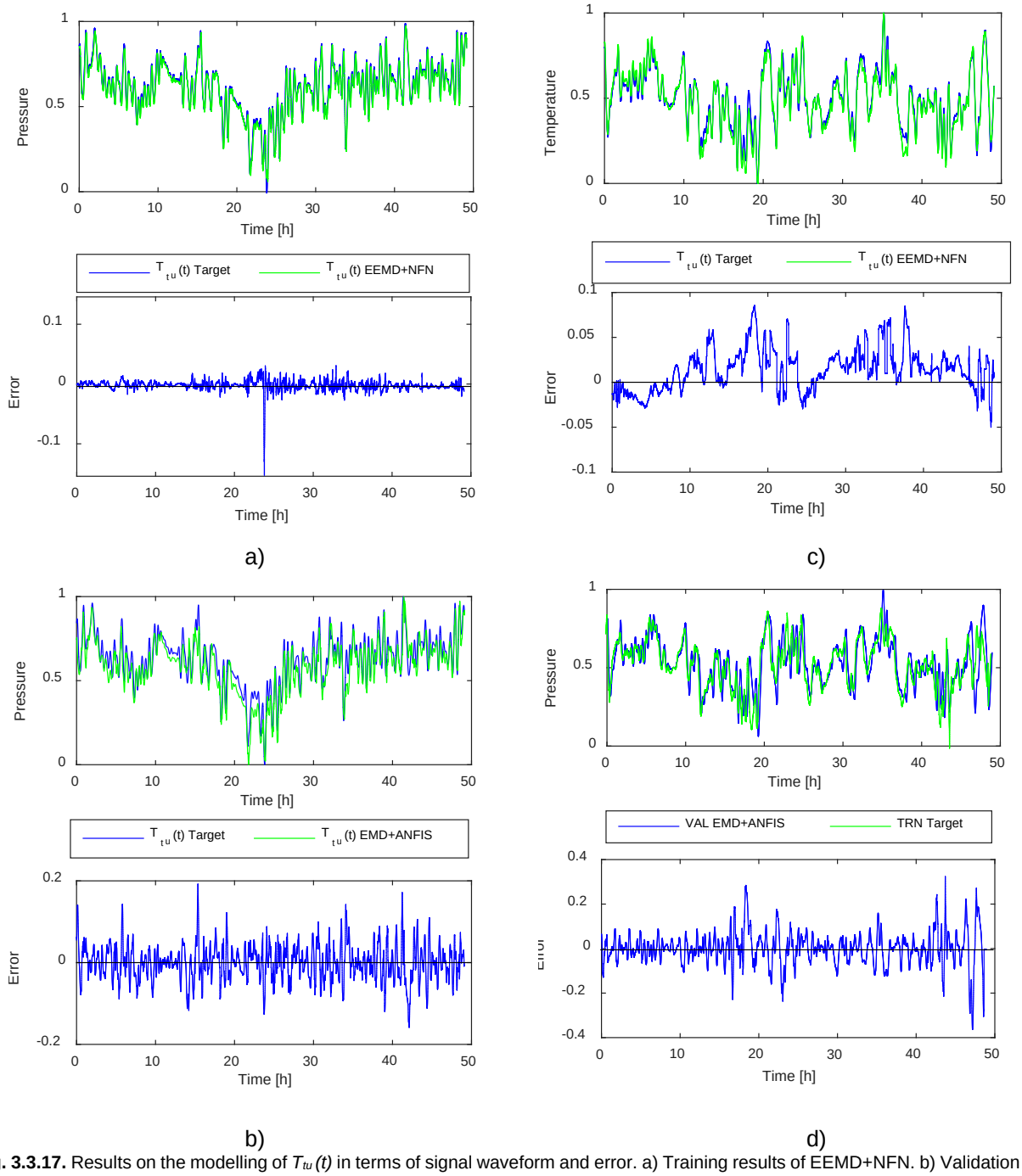

Fig. 3.3.17. Results on the modelling of $T_{tu}(t)$ in terms of signal waveform and error. a) Training results of EEMD+NFN. b) Validation results of EEMD+NFN. c) Training results of EMD+ANFIS. d) Validation results of EMD+ANFIS.

The performance of the ANFIS model in front of this case study is reduced in terms of MAPE error, since the training set presents an error of 7.95%. Indeed, the model also presents signs of overfitting as can be appreciated with the results of the validation set. The model is adapted to the variation and amplitude of the training set, thus only the mean value of the validation set is considered meanwhile the instant value is not well estimated. This fact is reflected in a MAPE error higher than 15%, which is far from the MAPE of 3.53% achieved with the proposed method.

Discussion and conclusions

This section presents a novel methodology for industrial time series forecasting based on multi-dynamics considerations by means of the Neo-Fuzzy Neuron modelling. There are two important aspects in this contribution method. The first one is the application of a decomposition scheme supported by EEMD. That is a noise-assisted signal adaptive decomposition method able to extract un-correlated dynamic modes of the target signal to forecast. The second, is the validation of the NFN as a modelling structure for industrial time series forecasting. The modular architecture of the NFN presents a capability to handle the dynamic modes extracted while assuring a proper convergence. Two relevant case studies have been considered, which represent an important validation set: The Tennessee Eastman Process, and a Copper Rod Manufacturing Process. Under both scenarios, the proposed modelling scheme exhibits a reliable forecasting outcome.

Furthermore, an exhaustive comparison with classical ANFIS approach has been considered. The results show that the proposed method better adapts itself to the available training dataset, increasing generalization capabilities towards the validation set, while reducing modelling complexity. This fact is validated quantitatively by the decrease of 11.55% over the MAPE generated by the ANFIS approach. It is worth noticing that the combination of signal decomposition and NFN avoids the need of additional inputs as past values of the target signal in order to reach the resulting performance. ANFIS approach, requires multiple model training, for which the past inputs are mandatory to infer the outcome, while NFN approach considers all the IMF in a unique model. As a result, the training procedure of the proposed method is simplified by avoiding the utilization of, classically, a genetic algorithm to identify the best-delayed samples for each of the IMF.

The proposed method represents a contribution to the advance towards the definition of modelling methodologies to be applied in real industrial processes.

3.4 Conclusions

The presented chapter deals with the proposed contributions to the field of signal's dynamics consideration for industrial time series modelling. In this regard, the main contributions of this chapter are related to:

- The proposal of a procedure for optimize the packing of significant dynamics modes in terms of relative energy and modelling error for improving forecasting accuracy and generalization, while reducing the number of required models.
- The proposal of the Neo Fuzzy Neuron as a modelling method to be applied with complex non-linear industrial signals where the model needs to handle multiple inputs, while assuring the convergence.
- The proposal of a forecasting methodology for combining noise-assisted signal decomposition approaches with the Neo-Fuzzy Neuron to improve the forecasting accuracy and generalization when dealing with complex signals avoiding the use of a multi-model approach.

The obtained results indicate the importance of the dynamics consideration. Indeed, it has been seen how classical single model approaches that not considers the dynamics of the signals are totally outperformed by signal decomposition modelling approaches. In this regard, the proposed coherent aggrupation of such extracted dynamics in terms of significance and modelling difficulty, represents an improvement for the increase of modelling performance if compared with single model approaches, and, at the same time, increases generalization while reduced the number of required models of literature decomposition and modelling approaches.

The proposal of the Neo Fuzzy Neuron for the modelling of industrial time series overpasses classical ANFIS approaches in terms of number of input management and convergence response with reduced datasets. However, as it does not provide any input adaptation method, it must be assured that consistent and non-correlated information is introduced to the model. Furthermore, the suitability of such modelling approach has been tested with a common benchmark in the literature, the achieved results indicate in a standard platform the suitability and the synergy between signal decomposition approaches and the Neo Fuzzy Neuron, since it is able to handle all the signal dynamics in a single model approach.

Thus, the extracted conclusions from this chapter demonstrate how the analysis and the coherent consideration of the signal's dynamics can contribute to the improvement of modelling performance dealing with complex industrial time series. It must be considered that during the analysis and validation of the proposals, the contributions in regard with the signal pre-processing presented in the previous chapter has been also applied.

Indeed, by means of a coherent methodology for developing the forecasting models of critical industrial series, considering the dynamics content of the target signal, the auxiliary signal management and optimal model configuration approaches discussed in this thesis, the industrial signal forecasting modelling can be enhanced and, then, the process condition forecasting can be properly approached as presented in the next chapter.

4.

Contributions to

Industrial process condition forecasting

This chapters presents the contributions of this thesis to industrial process condition forecasting. In this regard, the methodology to estimate the future condition in regard with the forecasted evolution of the process critical signals is exposed. This chapter concludes the contributions of the proposed thesis and closes the research work with the experimental validation of the proposed methodology in the copper rod manufacturing process.

CONTENTS:

4. Industrial plant condition forecasting

4.1 Introduction

Reliability and safety are becoming critical aspects in the modern industry. In this regard, the industrial sector has being carried out a significant effort towards process condition monitoring approaches since a decade ago [130], [131]. However, the condition forecasting of the industrial processes, although critical to allow preventative actions, is still an open-problem in the field. That is, information about the future condition of an industrial process represents a significant tool in order to gain response time against undesired deviations of the process condition. Thus, the combination of the future knowledge of the process status and the consequent assessment of the future condition, is a required objective towards the next generation of industrial monitoring strategies [3].

In this regard, in order to define the operating status of an industrial process, the physical modelling approach represents a non-viable method, since describing the multiple non-linear relations among different components in a complex process requires an unaffordable amount of resources and considerations. The data-driven approach, however, allows the extraction of such non-linear relations among process's signals from historic data analyses. Indeed, the data-driven approach provides a mathematical description of the process operation based on the identification of the main relations among the signals considered in existing data bases. A complete register of all the elements involved in a specific industrial process is not usually available, however, the target signals for the correct monitoring and interpretation of the process status are usually considered. Thus, the modelling of such signals represents a valuable source of information, since allows process condition assessment and the application of forecasting analyses [132]. Indeed, the combination of forecasted target signals by means of information fusion approaches will allow to obtain future condition of the monitored process. However, the condition forecasting of an industrial process deal with two sources of inaccuracies that must be overcome. First, the diagnosis error, which is the difference between the estimated condition and the real one. Second, the forecasting error, which is the difference between the estimated value of the target signal and the real value into a predefined time horizon.

The current state of the art shows three main approaches regarding condition forecasting approaches. First, a direct condition forecasting, which is based on a single model structure fed by information of the process, and trained to estimate a future condition, usually, related with the remaining useful life or time to failure [107]. In this scheme, the model carries out both diagnosis and forecasting tasks. Second, a discrete condition forecasting, which relies on using a classification algorithm to estimate the condition of the process from a set of predefined classes and, then, a sequence predictor is applied to forecast the future condition [98]. In this scheme, the condition of the process is the information to be forecasted by a specific historical analyser algorithm. The third one, a condition assessment based on signal forecasting, is based on the modelling and forecasting of the extracted signals and features from the process and their posterior evaluation to assess the future condition [92].

4.2 Contributions on future condition assessment

The condition assessment based on signal forecasting provides good performances since the modelling of the future condition is approached as a classical time series problem preserving as much information as possible from the original information. However, this approach drags two different sources of error, the inherent forecasting error of the input modelling block, and the class assessment error introduced by the classifier. Therefore, to overcome the limitations of this approach, efforts should be made in the improvement of both forecasting model and condition assessment stage.

The proposed methodology for industrial process condition forecasting is intended to overcome the limitations of the main approaches found in the literature. It combines high-performance time series modelling, together with non-linear process information. Thus, the objective is to forecast industrial process condition by the modelling, forecasting and posterior non-linear combination of process target signals. Novelties of the proposed methodology include, first, the time series modelling of target signals by means of a Neo-Fuzzy Neuron (NFN), which provides model design simplicity while increasing performance, as it has been analysed in the previous chapter. Second, a non-linear mapping of the process condition by means of a Self-Organizing Map (SOM), and the use of this map to assess the future condition in regard with the historical data and condition stages of the process.

The diagram of the proposed methodology for future condition assessment is shown in **Fig. 4.2.1**. The method proceeds as follows: *(i)* the *n* identified target signals are modelled by a set of *n* NFN based models in order to forecast its expected behaviour in a predefined time horizon, *(ii)* the *n* forecasted signals are codified by means of a SOM based topology preservation mapping, that summarizes the *n*-dimensional map into the Matching Unit (MU) of the grid, that is, the operating point of the process, and, finally, *(iii)* the information regarding future condition and forecasting robustness is given by the assessment of the operating points into the resulting supervised mapping from the SOM training.

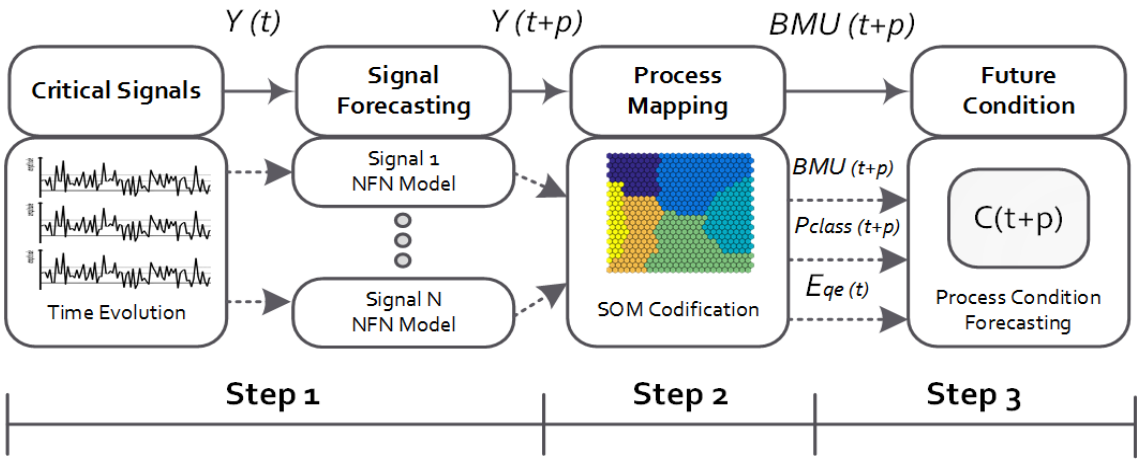

Fig. 4.2.1. Block diagram of the proposed condition forecasting method. Three steps are identified in the diagram: (i) Step 1: Target signals forecasting, (ii), Step 2: Process behaviour codification, and (iii) Step 3: future process condition assessment.

- **NFN based signal modelling**

The objective of the models is to drive the *N* target signals into a future time horizon, *p*. In this regard, the models present a multi-input single-output structure as shown in Eq. (4.2.1). where considering a target signal, *y(t)*, the model presents a single output that corresponds to a future value of the target signal, *y (t+p)*. The inputs of the model correspond to: *(i)* the current value of the target signal, *y (t)*, *(ii)* a past value of the signal delayed *z* samples, *y (t-z₁)*. It must be noted that z_1 is set by means of an optimization algorithm in regard with the achieved error. Finally, *(iii)* the mean value of the signal in the last 60 min, $\bar{y}$ *(t)*. This input is calculated to provide information regarding low dynamics of the signal.

$$y(t+p) = NFN\left(y(t), y(t-z_1), \bar{y}\,(t)\right) \qquad \textbf{Eq. 4.2.1}$$

By the intrinsic architecture of the NFN, the number of synaptic nodes is equal to the number of inputs of the model, so there are *Sy=3* synaptic nodes for each NFN model. The other configuration parameter of the NFN is related with the number of MF associated to the description of each input, *h*, that should be selected in regard with the complexity of the data to model. The defined structure of the NFN based model is the same for all *n* signals, however, the selection of the optimum *z* value is selected individually for each model. Performance of modelling is evaluated by classical error metrics. In this regard, the Root Mean Squared Error (RMSE), and the Mean Absolute Percentage Error (MAPE), are used to evaluate the goodness of the modelling.

- **SOM based process behaviour codification**

The operating point of the process, BMU *(t)*, is estimated by the SOM based combination of the *n* target signals considered into the process operating point. Such combination is carried out by the projection of the *n*-dimensional input into the SOM grid, and corresponding best matching unit assignment. It should be noticed that a new observation of the *n* target signals is considered an observation of the process.

During the training procedure, the SOM adapts the coordinates of the grid units, that is, the weights of the neurons, to the topology described by the *n*-dimensional input space. Therefore, considering such information, SOM can be evaluated with the output of the forecasting models in order to retrieve the future process operating point of the industrial process, as represented in Eq. (4.2.2).

$$BMU(t+p) = SOM\left(y_1(t+p), \ldots, y_N\,(t+p)\right) \qquad \textbf{Eq. 4.2.2}$$

Typically, the performance of SOM map is evaluated in terms topographic error, E_{top}, and average quantization error, E_{qe}. during the training stage. The topographic error represents the percentage of data vectors for which the first-BMU and the second-BMU are not adjacent units, and, average quantization error represents the average distance from each data vector to its BMU. High values of quantization error implies that the evaluated observation does not belong to the characteristic data density distribution modelled during the training of the grid. In this regard, the quantization can be understood as a measure of similarity between the initial knowledge used to train the method and the evaluated observation, and then, it can be associated to a probability value related with the forecasting outcome reliability.

- **Future condition assessment**

After the training of the SOM grid, a label is assigned to each MU in regard with the targets of the training dataset. This assignation is made following a majority voting approach. Thus, for a new observation to be evaluated, a BMU will result from the SOM, and its associated label will reveal the class outcome.

In order to provide a class membership degree, p_c, to the future condition, the center of each class in the input space is estimated, and the corresponding BMU, in terms of proximity to such classes' center are identified. The proposed method associates the highest membership degree, $p_c = 1$, to the points that belongs to the BMUs representing the center of the class. Therefore, the membership degree is reduced, following a sigmoid function, as distance from the center is increased. Hence, for each new observation to be evaluated, the related p_c is estimated.

Afterwards, the future process condition, *C (t+p)*, is estimated by the method applying Eq. (4.2.3). That is, future condition is given in regard with three parameters:

(i) The estimated future process operating point, *BMU (t+p)* and corresponding class,

(ii) *T*he distance to the center of the corresponding class, and

(iii) The quantization error, E_{qa}. It must be noted that the proposed quantization error included in the condition estimation is related with the current time instant, *t*, in order to gain robustness and reliability. Indeed, the quantization error, E_{qa}, and the distance to the class, p_c, are used as weighting factors over the condition forecasting outcome.

$$C(t+p) = \frac{Class\big(BMU(t+p)\big) - p_c(t+p)}{1 - E_{qa}(t)} \qquad \textbf{Eq. 4.2.3}$$

4.3 Experimental validation of the proposed condition forecasting methodology

The proposed methodology is validated in the CRMP industrial process considered in this thesis, that is, the Copper Rod Manufacturing Process, in which high purity copper cathode is transformed into high performance copper rod. Indeed, as it has been aforesaid, the aim of the CRMP is to manufacture copper rod from a continuous casting process. A detail of the monitored part of the process considered for the validation of the proposed condition forecasting methodology is shown in **Fig. 4.3.1**. Mainly, the manufacturing process implies the transformation of the melted input copper in a solid copper rod, and such transformation is based on a controlled solidification process.

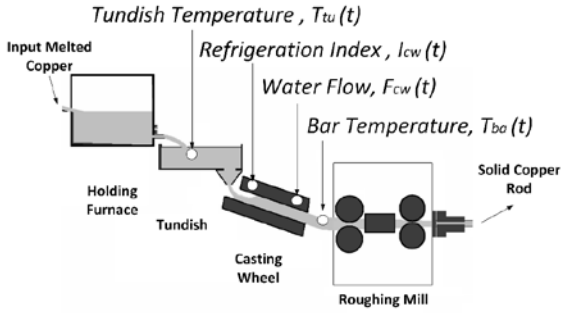

Fig. 4.3.1. Copper Rod Manufacturing Process plant diagram. Target signals considered are marked in the figure.

Indeed, the solidification of the copper is a critical aspect within the manufacturing process, in which the heat must be properly extracted from the copper bar. The problematic of the heat extraction in the CRMP are explained in section 1.1.2 of this thesis document. As a summary, the condition of the process is related to the probability of having porosity in a certain manufactured rod asset. That is defined in regard with the behaviour of the critical process signals during the solidification procedure. The probability associated with the produced unit are estimated *a posteriori* by a manual quality inspection procedure. This inspection fixes a range of porosity probability that is summarized in a discrete quality label assigned to each unit. This range is defined as the probability to find cup-cone failures in the manufactured unit. In this regard, the predefined quality ratios correspond to: high quality, *Q1*, probability between 0 % and 25 % of porosity affectation, medium quality, *Q2*, between 25 % and 75% and, low quality, *Q3*, between 75 % and 100%.

4.3.1 Modelling of the process's target signals

The available target signals related with the copper solidification process are listed in **Table 4.3.1**. In this regard:

- T_{tu} *(t)* corresponds to the temperature of the melted copper in the tundish. It is the last reference of the melted copper before entering the casting wheel, it is understood as the input temperature to the solidification process.
- T_{ba} *(t)* is the temperature of the solidified copper rod bar before entering to the roughing process. It represents the output temperature of the solidification process.

- The I_{cw} *(t)* is the heat extraction index of the casting wheel. It is measured as the difference between the input water temperature and the resulting water after the refrigeration of the wheel. It is a measure of the efficiency of the copper heat extraction procedure, low values of I_{cw} *(t)* indicate a proper heat extraction procedure.
- The F_{cw} *(t)* is the total water flow introduced to the solidification process. It shows deviations from the nominal configuration if the heat extraction is not properly done.

Deviations in the specified variables negatively affects the solidification procedure causing with it an increase of porosity apparition, and consequently, a decrease in the manufactured product quality level.

Table 4.3.1. Target signals of the CRMP and their associated description

Avb.	Description
T_{tu} **(t)**	Tundish Temperature [ºC]
T_{ba} **(t)**	Copper Rod Bar Temperature [ºC]
I_{cw} **(t)**	Casting Wheel Refrigearation Index [ºC]
F_{cw} **(t)**	Total Water Flow Casting Wheel [l/min]

All signals are acquired synchronously, and are automatically stored in a standard SQL database at a period of 10 seconds, that corresponds to a sampling frequency, f_s, of 0.1 Hz. A data sets shown in **Fig. 4.3.2**, corresponding to 250 hours of plant operation are used in this work. In this regard, 70% of data is used for training, and 30% for testing the suitability and the generalization capabilities of the method. It should be pointed that all signals have been normalized in the [0-1] interval regarding the max and min values found in the database.

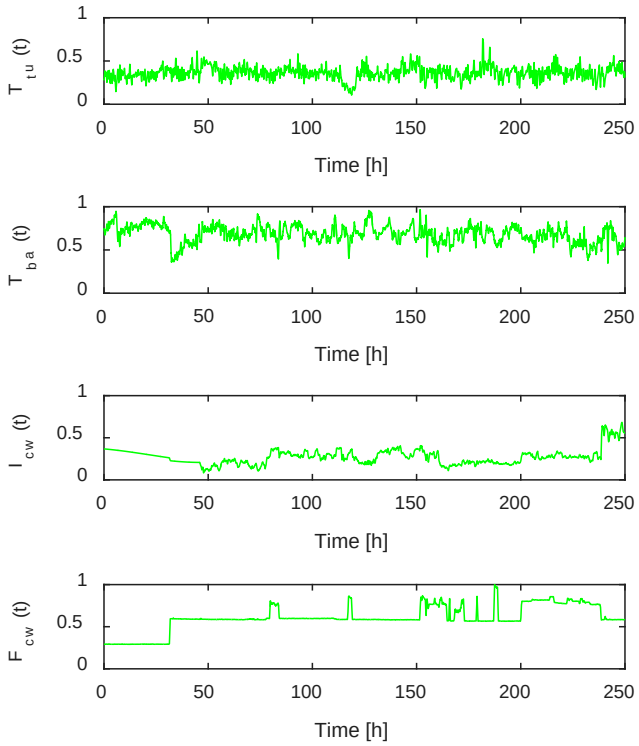

Fig. 4.3.2. Target signals of the CRMP database. A total amount of 250 h of plant operation time have been considered.

The target signals of the CRMP, T_{tu} (t), T_{ba} (t), I_{cw} (t), and F_{cw} (t), are modelled by a set of NFN based models. Thus, the output of the models corresponds to the target signal forecasted over a defined forecasting horizon that is common to all models. In this regard, the manufacturing time of one copper rod unit fixes the forecasting horizon of the model to p = 90 samples, that corresponds to 15 minutes of continuous copper casting. The main objective behind this horizon is to get information of future process condition before any deviation affects the next manufactured unit. The number of memberships, h, has been set to 15 MF per input. The classical backpropagation algorithm has been selected for the training procedure. Learning rate was set to α = 0.01, and the number of training iterations to adjust the weights of the network has been set to 20. After the training procedure, the resulting NFN models for the target signals are shown in **Fig. 4.3.3**, and the corresponding performance metrics summarized on **Table 4.3.2.**

Table 4.3.2. Error performance metrics of the forecasting model.

	T_{tu} (t)		T_{ba} (t)		I_{cw} (t)		F_{cw} (t)	
	Trn.	***Val.***	***Trn.***	***Val.***	***Trn.***	***Val.***	***Trn.***	***Val.***
RMSE	0.04	0.05	0.04	0.06	0.01	0.02	0.01	0.03
MAPE	5.22	7.81	4.98	9.68	8.67	12.95	9.83	16.98

The results show the capabilities of the NFN based modelling to fit properly all signals during both training and validation stages. Indeed, the results of the modelling show that, in terms of performance, the NFN achieves good results for the target signals modelled. In this regard, the temperatures T_{tu} (t) and T_{ba} (t) are the signals that achieve the lowest MAPE and RMSE errors, 7.81 % and 9.68%, respectively. This is due to a smoother dynamics content in both signals in comparison with I_{cw} (t) and F_{cw} (t), which dynamics present sudden changes to response versus undesired deviations in the process.

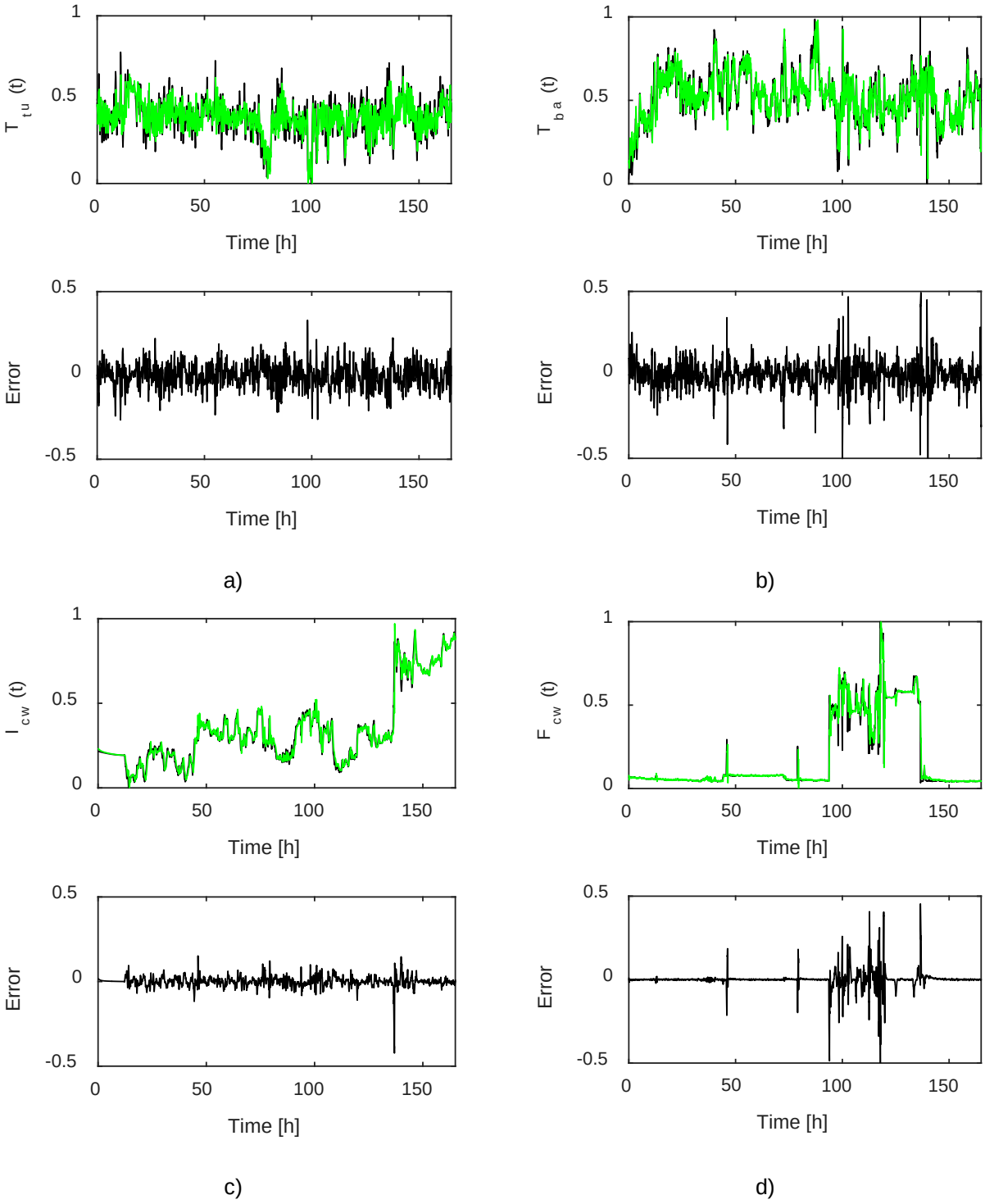

Fig. 4.3.3. Forecasting models of the critical process signals at *p*=90 samples, comparison with real data and resulting error. a) NFN modelling of the tundish temperature, *Ttu* (t). b) NFN modelling of the copper bar temperature, *Tba (t).* c) NFN modelling of the heat extraction index *Icw (t)*. d) NFN modelling of the total water flow, *Fcw (t)*.

4.3.2 Process operating point codification

After the modelling, all the forecasted signals are considered by means of a SOM based mapping to define the operating point of the process, *BMU (t)*. Indeed, each BMU can be seen as a non-linear region of the space defined by the target signals of the plant during the training stage. Such BMU summarizes all the variability of the signals in the corresponding region, and is representative of the process operating condition at the evaluated time instant.

The SOM grid has been configured by a hexagonal distribution 20 x 20, that is, a total of 400 units. The SOM is initialized and trained by a batch algorithm and a total amount of 150 epochs are performed. For this training set, the resulting errors are E_{qe} = 0.054 and E_{top} = 0.064. The obtained E_{qe} value corresponds to an average distance measurement among MUs. In the other hand, the low value of the E_{top} indicates that the initial topography is well preserved by the trained SOM, indicating with it the suitability of the projection.

The resulting process operating regions formed by the mapping of the target signals can be analysed by means of the U-Matrix shown in **Fig. 4.3.4**. The U-Matrix represents the distances between the neurons of the trained SOM. Thus, it is used to see if there are discrete operation areas in an industrial process.

It should be noticed that SOM works under an unsupervised approach. However, as the corresponding condition label of each observation is known, the behaviour of the process's operation can be compared with the identified clusters shown in the U-matrix. Indeed, three clusters can be identified in **Fig.4.3.4**, named clusters A, B and C, which can be related with the three known operating conditions of the process, that is Q1, Q2 and Q3.

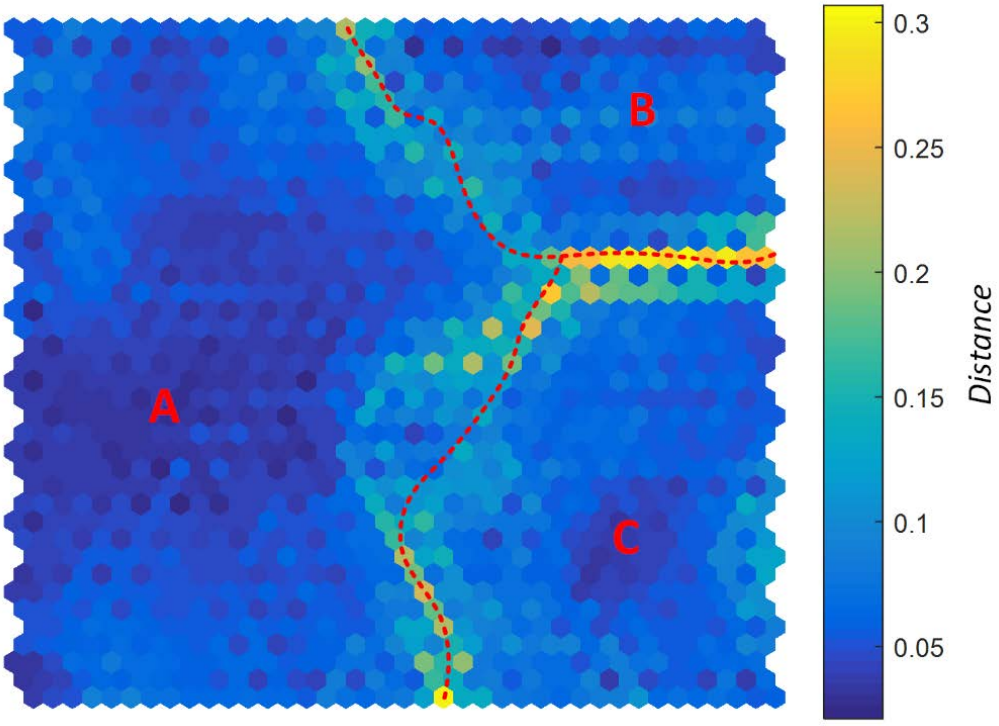

Fig. 4.3.4. U-Matrix of the trained SOM containing the distances between the neurons of the grid. The dark areas represent regions of the map in which data is concentrated (low distance between neurons), and the light areas represent the frontiers of the clusters (high distances between neurons). Red line defines the boundaries between low-distance neighbour MU and high-distance neighbour MU. The three clusters identified has been marked with A, B and C labels to facilitate the explanation.

A prior interpretation of the significance of the process operating regions, in regard with the manufactured copper rod quality and the physical affectation of the target signals can be made by looking at the individual affectation of each SOM input to the definition of the U-matrix. Such affectation is given in **Fig. 4.3.5**.

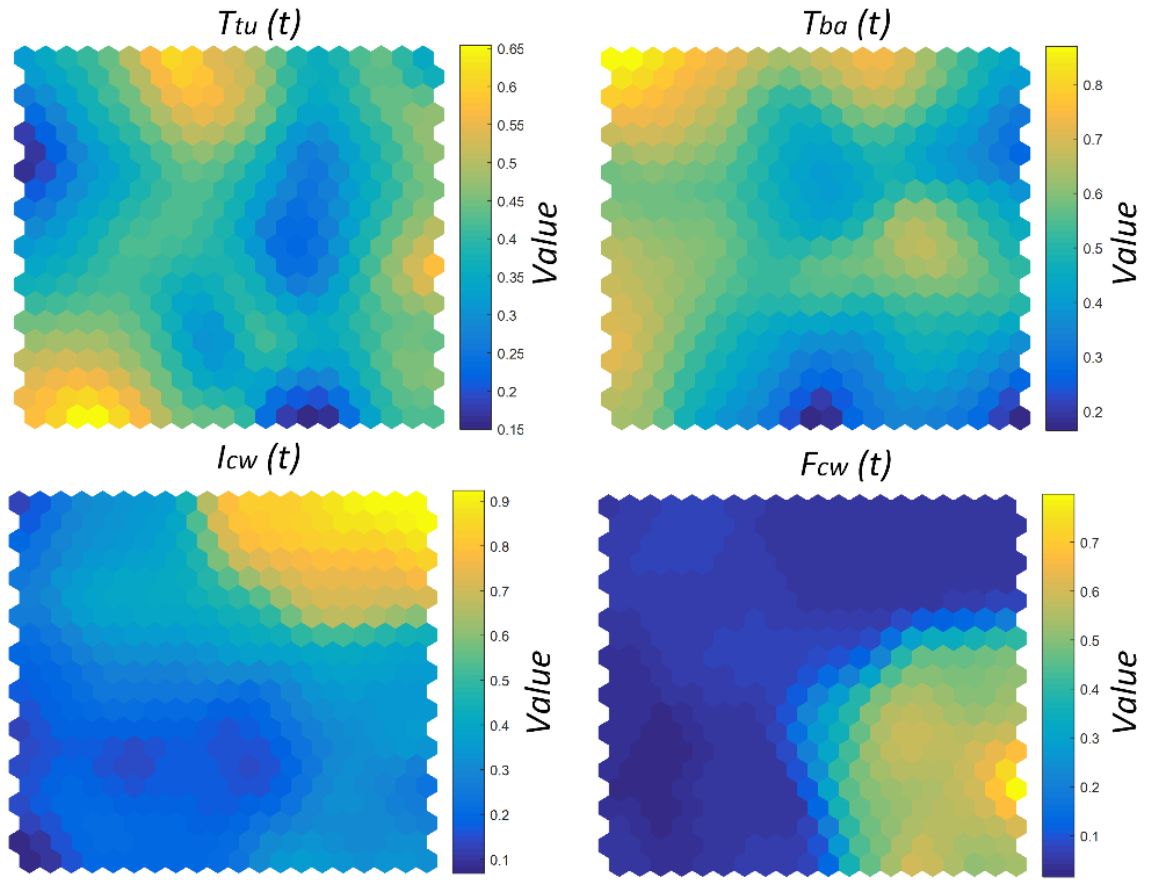

Fig. 4.3.5. Contributions of the four target signals, in terms of signal value, to the definition of the process operating regions seen in the U- Matrix.

The clearest affectation to the definition of the operating areas of the U-matrix is the signal F_{cw} *(t)*, since it presents a constant nominal values for the clusters A and B, with a value around 0.6, but defines the C cluster with high values of water caudal, F_{cw} > 0.8. This is an expected behaviour highly related to the copper solidification issue. That is, when the heat extraction process is not able to extract enough heat from the melted copper, the process requires to increment the total water flow in order to reduce T_{ba} *(t)* to the security operational ranges fixed by the process. This increase of the refrigeration water introduced to the process causes a sudden drop of the copper bar temperature, also visible in the affectation of T_{ba} *(t)* into the C cluster, causing high probability of porosity apparition in the copper manufactured product. Therefore, there exists a match between cluster C and a non-proper quality process condition, Q3 in this case taking into consideration the corresponding labels of the samples.

The T_{tu} *(t)* presents an almost uniform affectation to the definition of the U-matrix. This is an expected behaviour, since corresponds to the input temperature of the melted copper and presents a greater stability. However, two interesting behaviours of such signal can be appreciated also related with the cluster C. Such behaviours are also related with low quality condition, since it presents two areas with abnormal high and low

values in the corresponding C cluster. Indeed, such behaviour indicates that sudden drops or increases of the input temperature may also cause deviations from the optimal condition.

Similar behaviour is appreciated in $I_{cw}(t)$, where high heat extraction index causes small deviations towards the nominal condition of the process, causing a moderate probability of porosity in the final unit, associated to the cluster B, that corresponds to a medium quality condition, Q2. In the other hand, $T_{ba}(t)$, presents nominal values in the left part of the U-matrix, for this reason the left operating region, cluster A, corresponds to the nominal and desired operating condition of the process, Q1.

For each neuron considered in the U-Matrix, the SOM computes a majority-voting based procedure in order to assign one of the class labels, that is, Q1, Q2 or Q3. Thus, the SOM operates as a classifier assessing the corresponding condition from the combination of the forecasted input signals. For the considered experimental case, the final class partition given by the BMUs of the SOM is shown in **Fig. 4.3.6**. It can be seen how the final distribution of the BMUs targets matches with the previous considerations on the affectation areas of the target signals. Furthermore, the centres of each class are also defined and marked in the figure. These centres are used to evaluate the distance of a new point towards the center of the assessed class, and get a measure of the membership degree of such class, p_c.

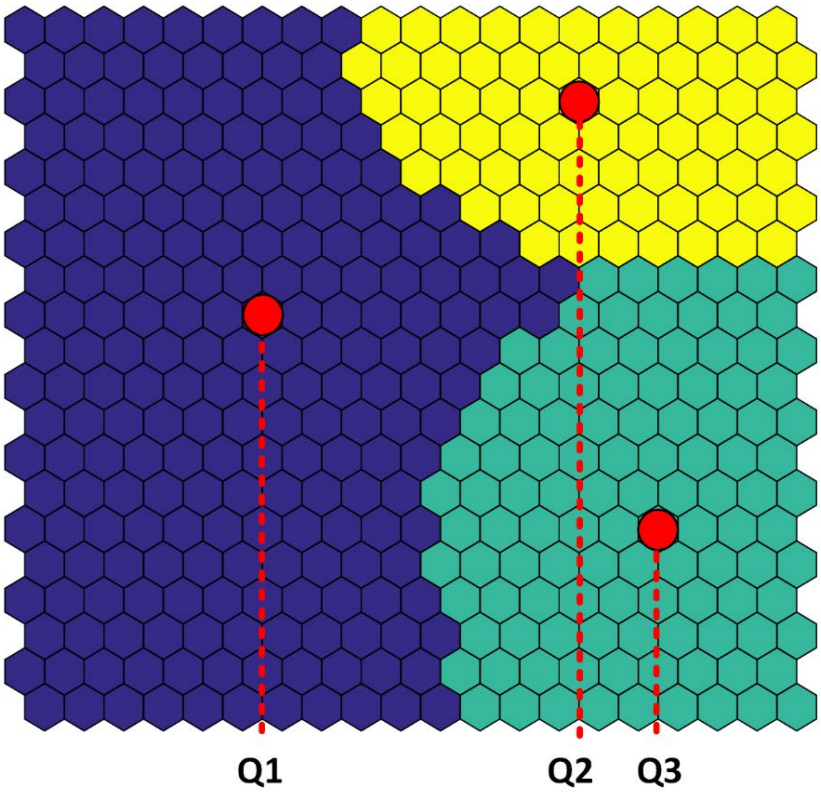

Fig. 4.3.6. Class partition of the SOM grid. (i) Q1 Center in BMU=131. (ii) Q2 Center in BMU=102. (iii) Q3 Center in BMU=323.

4.3.3 Condition forecasting

Once the SOM is trained, and the class labels assigned, the grid can be evaluated with the input signals in the forecasted time horizon *p*. The BMU signal corresponding to the future operating point of the process, and the quantization error can be seen in **Fig. 4.3.7** for the considered dataset.

In this regard, it can be appreciated that the method presents low signal forecasting error, since the topology of the input space formed by the target signals is conserved, this fact is indicated by the similitude of the target BMU and the forecasted one seen in the figure. Also, the corresponding E_{qe} *(t)* is low for most of the evaluated samples, fact that indicates that the majority of the forecasted points have been seen by the grid during the training. However, **Fig. 4.3.7** shows, around 100 to 150 hours of plant operation, some areas in which the achieved quantization error is high, which indicates that the algorithm is suffering a loss of reliability due to a non-known patterns in the input data.

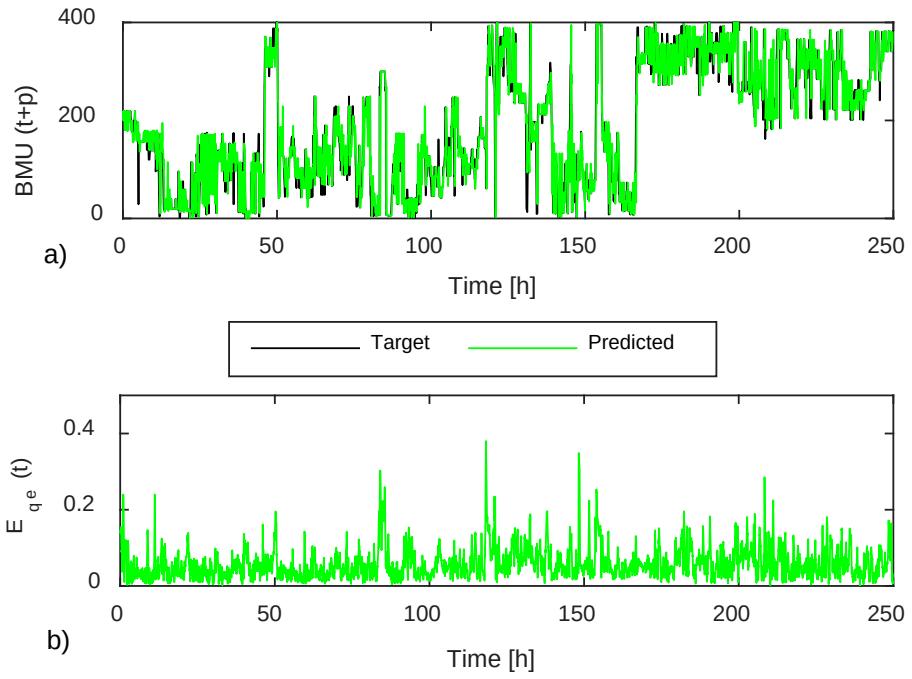

Fig. 4.3.7. Results on the evaluation of the SOM in terms of a) BMUs of the future condition estimated by the models, BMU (t+p), versus the expected BMU, and b) E_{qe} *(t)* with the current value of the target signals.

Furthermore, **Fig. 4.3.8** shows a time detail of the condition calculated from the assessed BMU and the probabilities extracted from the distances to the center of the class. In this regard, it is an example of the performance of the proposed method for estimating the future conditions. In this regard, as it is expected, the process of copper rod production operates most of the time under high quality conditions, Q1, having in this regard some deviations in this conditions to Q2 and Q3 during short periods of time. These changes of conditions are due to deviations from the nominal operation of the processes.

Also, it can be appreciated how p_c is high for almost all the samples forecasted, this indicates the suitability of the positioning of the SOM neurons during training of the grid, and the input space covering by such neurons. However, it can be seen in **Fig. 4.3.8 (a)** how there are some values with lower class membership probabilities. Nevertheless, such p_c values never decrease form 0.5, which is the minimum probability considered for accept an observation as belonging to that class.

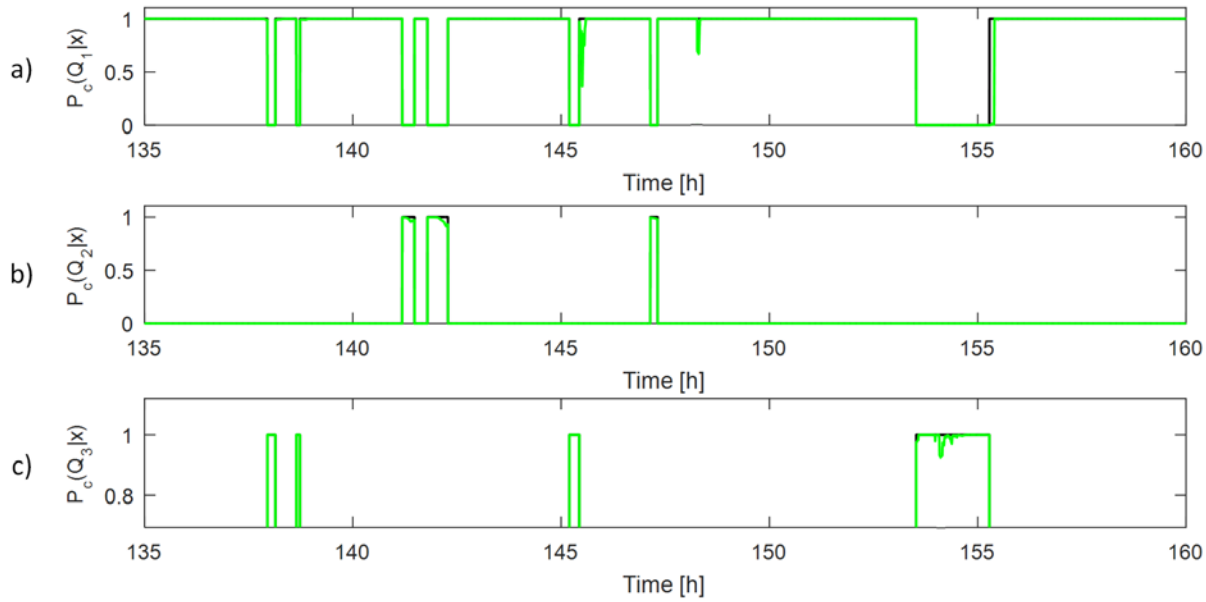

Fig. 4.3.8. Results from the condition assessment for each considered class in regard with the temporal evolution of the operating point. *a)* p_c value for the high quality versus expected target, $p_c(Q_1|x)$. *b)* p_c value for the medium quality condition versus expected target, $p_c(Q_2|x)$. c*)* p_c value for the low quality versus expected target, $p_c(Q_2|x)$.

Finally, the information is combined by following Eq. (4.2.3) to obtain the future condition of the industrial process. In this regard, the performance on the assessment of the future condition is summarized in the confusion matrix exhibit in **Table 4.3.3**. The overall system achieves a performance of 91.30%, which corresponds to a notable performance result when dealing with real industrial data. It can be seen that the class that achieves worst results is Q2. This is due to the low variability that presents this condition in regard with the nominal operating class Q1. However, Q3, that corresponds to the lowest quality and should be detected primarily, presents good classification performances, which hat implies that it can be assessed with a low error.

Table 4.3.3. Confusion matrix resulting on the assessment of the C(t+p) with the proposed method.

T \ P	1	2	3	Accuracy
1	43871 *95.57%*	1012 *2.21%*	1019 *2.22%*	95.37% *4.63%*
2	3197 *13.02%*	20624 *84.00%*	730 *2.97%*	84.00% *16.00%*
3	1151 *5.91%*	714 *3.67%*	17616 *90.43%*	90.43% *9.57%*
Accuracy	90.53% *9.47%*	92.97% *7.03%*	92.88% *7.12%*	**91.30%** ***8.70%***

In order to compare the performances of the proposed method with classical approaches, an ANFIS based modelling for predicting the evolution of the target signals, and a neural network for assessing the future condition have been implemented. In this regard, the ANFIS models have been configured with three inputs and a single output structure, as shown in Eq. (15). In this regard, the output is the target signal among the horizon *p*, *y (t+p)*. While the inputs correspond to the current value of the signal, *y(t)*, and two past values of the target signal delayed z_1 and z_2 samples accordingly. Note that z_1 and z_2 have been set by means of a genetic

algorithm. The obtained values are $z_1=44$ and $z_2=67$ samples. The resulting error metrics of the ANFIS models are summarized in **Table 4.3.4**.

$$y(t+p) = ANFIS\left(y(t), y(t-n_1), y\,(t-n_2)\right) \qquad \textbf{Eq. 4.2.4}$$

Table 4.3.4. Error performance metrics of the forecasting models with the classic ANFIS and Neural Network approach

	T_{tu} (t)		T_{ba} (t)		I_{cw} (t)		F_{cw} (t)	
	Trn.	***Val.***	***Trn.***	***Val.***	***Trn.***	***Val.***	***Trn.***	***Val.***
RMSE	0.07	0.012	0.078	0.099	0.029	0.215	0.045	0.083
MAPE	8.91	13.2	10.1	14.2	15.3	17.6	16.21	22.3

The NN is trained by using the training data and the class information of each observation. In this regard, a multi-layer NN has been configured with one hidden layer which is composed of 10 hidden neurons. The neurons have been configured with a sigmoid activation function, which is recommended for classification problems involving multiple independent attributes. The network is trained by means of classical backpropagation algorithm. The current information of the target signals is used for the training of the network, while the future behaviour is used for testing the performance of the future condition assessment. Indeed, after the modelling of the target signals, the NN is fed directly by the future value of the target signals given by the ANFIS models to assess the condition. Therefore, the resulting confusion matrix is shown in **Table 4.3.5.**

The combination of ANFIS and NN achieves a total performance of 80.9% of classification score. Therefore, the proposed method outperforms in more than 10% the classical approach. Mainly, this increase of performance is due to the improvement over both sources of error, the modelling and the class assessment. In this regard, the NFN based models achieve better performance metric than the ANFIS. This is mainly due to the convergence capabilities of the NFN dealing with multiple behaviours of industrial time series, which is clearly shown comparing the corresponding MAPE errors on **Table 4.3.2** and **Table 4.3.4**.

Furthermore, the classic NN is able to classify properly the Q1 condition, which is the class with the highest number of samples. However, the other conditions, Q2 and Q3, are not properly characterized. This is due to its less representative number of samples. In fact, under such non-balanced sets of samples, the SOM provides enhanced generalization capabilities at the time of delimiting the boundaries of the classes in the input space.

Table 4.3.5. Confusion matrix resulting on the assessment of the *C(t+p) with the classic ANFIS and Neural network method*.

T \ P	1	2	3	Accuracy
1	44167 *96.27%*	251 *0.55%*	1460 *3.18%*	96.3% *3.7%*
2	8310 *33.85%*	16225 *66.09%*	16 *0.06%*	66.1% *33.9%*
3	4501 *23.10%*	2630 *13.50%*	12350 *63.40%*	63.4% *36.6%*
Accuracy	77.5% *22.5%*	84.9% *15.1%*	89.3% *10.7%*	**80.9%** ***19.1%***

Discussion and conclusions

The proposed method indicates the synergy between the neo-fuzzy neuron for high-performance industrial time series modelling, together with the self-organizing maps for combining the values of the signals for assessing the future behaviour of the industrial process. In this regard, the proposed methodology also exhibits the significance of SOM for classifying the future condition of the process by considering information regarding input data distribution, the distances of the centres of the classes modelled by the associated MU and the reliability of the prediction given by the quantization error of the SOM grid evaluated towards the new observation. Furthermore, the proposed methodology has been validated over the copper rod manufacturing process achieving good performances in the assessment of the future condition of the industrial process.

However, it should be remarked that condition forecasting approaches need to deal with the combination of two different sources of error, the forecasting modelling error, and the classification error. In this regard, the synergy between the NFN based models and the SOM present a solid proposal for the minimization of such errors. In this regard, the NFN is able to capture the dynamics of the signal providing a low-error forecasting with enough robustness to consider sudden behaviours characteristic of industrial process signals.

However, is the application of SOM which presents more benefits for the reliable assessment of the future condition. In this regard, SOM is able to perform both feature reduction and condition assessment functions by the segmentation of the data input space in non-uniform regions summarized in the MUs. In this regard, it is also capable of giving a measure in regard with how far is a new observation from the topographical distribution of the MU, that is the quantization error. Such value presents a strong influence in the goodness of the assessment, since it is able to discriminate if a current condition has been seen by the grid, and consequently learned by the forecasting models during its training or not.

The consideration of such information for an on-line modification of the continuous learning and the reconfiguration of the models open a new field of possibilities to integrate new learning paradigms in condition forecasting methodologies for industrial processes.

4.4 Conclusions

This chapter presents a novel methodology for industrial condition forecasting applied to a real case study, a copper rod manufacturing process. The main contributions of this chapter is the proposal of a methodology for the assessment of the future condition of industrial process. In this regard, the novelties of the proposed methodology are the adaption of neo fuzzy neuron based modelling to forecast the critical process signals into a future time horizon, and the future condition assessment approach by a non-linear mapping of those signals into the operating areas of the process with self-organizing maps.

Regarding target signal forecasting, it is demonstrated that neo fuzzy neuron based modelling results in a suitable approach to capture and forecast the behaviour of several industrial process signals with different natures and a similar range of dynamics information. Furthermore, self-organizing maps represents an enhanced approach to model non-linear behaviours defined by the available process signals, and estimate the future process condition. Indeed, the resulting grid modelling emphasizes the expected underlying physical effects defined by the target signals of the plant, since the resulting data clusters matches with the copper porosity classes considered.

In order to provide robustness to the proposed method, two additional measures are considered during the assessment of the future condition, the class probability, and the quantization error. Thus, a membership degree in regard with the estimated class is computed taking into account the distance of the new sample under evaluation, and the class distribution characterization. Also, the quantization error allows the similarity estimation between the sample under evaluation and the initial available knowledge, and, consequently, the reliability of the forecasted condition outcome.

Furthermore, the assessment made by the SOM provides better generalization capabilities, since the frontiers of the classes formed by the different conditions are less overfitted to the training set. It must be considered that the proposed method represents an advance towards the definition of condition forecasting methodologies to be applied in real industrial processes.

Summarizing, the contributions exposed in this chapter presents a complete industrial process condition methodology. The proposed methodology uses a robust and reliable forecast given by enhanced forecasting modelling approaches that, combined with non-linear topology preservation fusion procedures, is used to estimate the future condition while assuring the reliability of the prediction.

Next section summarizes all the contributions and conclusions reached by the proposed thesis, together with the future work to continue with the new research topics resulting from this thesis.

5.

General conclusions and future work

The main contributions of this thesis research, as well as the conclusions and future work are presented in this chapter.

CONTENTS:

5. Conclusions and future work

5.1 Conclusions

This chapter presents the conclusions in relation with the stated hypothesis and the obtained results during the development of the thesis. The conclusions are divided in three main blocks corresponding to the three thesis' chapters: Chapter2 - Pre-processing for industrial time series modelling, Chapter 3 - Dynamics based industrial time series modelling, and Chapter 4 - Industrial process condition forecasting. In order to complete this conclusions' section, the global conclusions in regard with the research impact of this thesis is given.

Indeed, as it was initially stated, the three main parts considered during the development of this thesis work presented, in the state of the art, a lack in the definition of coherent methodologies and improvements approaches for their application to industrial condition forecasting.

In regard with the pre-processing, the selection of the optimal forecasting horizon and its associate past values has been, classically, an unattended issue in the majority of the studies found in the literature. An exhaustive analysis of the target signal with the optimal forecasting horizon and its associated optimal past values, considering all the available signals and the modelling method, is not a viable approach, since the affectation of a concrete modelling method to the forecasting horizon is difficult to be assessed. For this reason, it is logical to limit the search interval of the optimization problem by considering only the target signal, since it presents the highest affectation to the performance of modelling method. Accordingly, the method for selecting the optimal horizon proposed in this thesis does not searches a concrete value, it tries, instead, to find those range of values that lead to a major performance of the modelling. To do so, first, the method performs an auto-correlation analysis of the target signal, second, the selection of the range of forecasting horizon suitable for such signal and, third, the identification of the ranges of optimal past values to be considered for such horizon. Then, the application of the genetic algorithm for finding the specific values is reduced in terms of computational burden, since the solution space is significantly reduced and, also, in terms of suitability of the final solution, since the possibility to get stuck in a local minima is reduced.

One of the open points for discussion in regard with the pre-processing, is how to interpret the auto-correlation graph of a signal from an industrial process. In this regard, dealing with industrial processes, the autocorrelation analysis of the target signal usually shows an exponential decay as long as the delays are increased. An interesting parameter of such responses is the width of the high-correlation region. That is, the number of delays that keep a correlation coefficient above 40%. In fact, such width is related with the dynamics of the signal under analysis, in such way that the higher the frequencies of the signal dynamics are, the quicker is the exponential decay of the auto-correlation, since the periodicity affectation of the higher frequency modes is lost as the delays are extended.

In regard with the management of the available auxiliary signals, this thesis demonstrates that the target signal should be understood as the resulting behaviour of the process and, the available auxiliary signals are considered to conform the operation point in which the process is working. Thus, in this regard the contribution relies on analysing the auxiliary signals in an independent manner a part from the model. This approach implies a direct increase of modelling performance, since the model is less stressed in terms of computational requirements due to the number of inputs.

Considering the auxiliary signals themselves, the objective of the proposed methodology is to codify this operating point in a certain number of non-linear *N*-dimensional regions, where *N* is the number of auxiliary signals considered for the modelling. In order to assure the codification of such signals into the operating point, it must be guaranteed non-redundancy among the considered signals and significance of auxiliary signals in regard with the target. In this regard, the proposed methodology performs a two-step correlation analysis among the auxiliary signals, which results in a good trade-off between simplicity and performance results. In fact, it has been demonstrated that a relation exists between the target signal and the process condition, and such relation is fixed by the set of auxiliary signals. Thus, the proposed methodology aims for a modelling architecture that allows to preserve as much as possible such intrinsic relations, despite other approaches as PCA that extract only the relations that maximize data variance. The use of a SOM based topology preservation modelling approach, to combine the set of auxiliary signals, has been demonstrated as performing approach to reach such objective. In this regard, topology is considered as the preservation of the underlying physical behaviours described by the auxiliary signals. Then, in front of a new observation, the method outcome represents a scalar value that implies a certain range of combinations from the auxiliary signals, that is, an identifier of a specific non-uniform region of the codified input space.

All these contributions result in a significant impact on the pre-processing for industrial time series modelling. Actually, the modelling algorithm has the objective of maximize the performance in regard with a given target signal and a forecasting horizon. Thus, the model seeks, during the training procedure, to define a mathematical function that allows to know the outcome of the model, the forecasted value, in regard with the relations of the input values. With the proposed methodology, these relations are drastically reduced, since each output value of the SOM corresponds to a combination of the *N*-available signals. As a result, in regard with the auxiliary signals, the proposed method avoids training efforts in finding such relations among the signals, that implies an increase of the model generalization, and a direct increase of modelling convergence capabilities.

Furthermore, an interesting point of discussion is focused on the resolution of the SOM grid. In this regard, if the number of neurons of the grid is too high, the model has the risk of oversizing process dynamics. This happens due to high variations in the BMU that are generated for small changes in auxiliary signal values, hence, a fake dynamic is induced in the model due to an overfitted partition of the input data space. Therefore, in order to fix the resolution of the SOM grid in terms of number of neurons, the quantization error should be considered. In this regard, considering an input set scaled between 0 and 1, the resolution of the SOM should minimize the quantization error under 0.3 for, at least, the 80% of the samples. This would guarantee a good trade-off between simplicity of the neuron grid and performance in the mapping of the auxiliary signals.

Dealing with industrial time series forecasting, it has been demonstrated that single modelling of an industrial signal with all its dynamic modes results in a worst performance that considering multiple model for the dynamic modes extracted. Furthermore, it has been also demonstrated that the approaches that require a dedicated model for each signal dynamics extracted imply a high-computational burden strategy that may cause the model to incur in overfitting.

Thus, the contribution of the proposed method consists on relating each one of the dynamics of the signal with its relative significance, in terms of energy, in regard with the totality of the target signal, and, also with the

contribution of each dynamic to the modelling error. Therefore, it is proposed a relation among three factors, the dynamics, their significance and their associated modelling error.

This approach allows to propose a method for adaptively packaging the signal dynamics where, first, the target signal is decomposed in terms of intrinsic mode functions. Then, such IMFs are iteratively added to a package of dynamics that is evaluated in terms of significance and modelling error. These two parameters are evaluated with a proposed error threshold function, that indicates if more dynamics can be added to the current evaluated aggrupation, or a new model needs to be considered. It has been demonstrated that this approach allows to increase modelling performance in almost 20% in terms of MAPE error compared with classical single modelling approach and, also, reduces the number of required models in an 80% compared to the dedicated model per signal dynamic approach.

In regard with the proposed method, the significance is an intrinsic property of each IMF. However, it is critical to define the error threshold function. Indeed, what is proposed in the method is to quantify how much importance should be given to high and low frequency dynamics. Accordingly, the error threshold function is defined in regard with the maximum admissible error for both high and low dynamics. Thus, the definition of the function is simple, since setting high restrictive values of error corresponds to the worst case scenario in which a single model is generated for each IMF, that implies going back to the classical approach proposed in the literature. In this regard, the proposed error threshold function can be related with the forecasting horizon. Indeed, in electromechanical systems and industrial processes, typically, the low frequency dynamics are related with high signal amplitudes, since they represent tendencies of long-term behaviours of the system. Otherwise, high frequency modes are related with low signal amplitudes. Hence, if the forecasting horizon is extended, the admissible error for the high frequency modes can be extended for two reasons: first, significance of high frequency modes to the amplitude of the target signal is low (around 1%) and, second, the correlation between the target signal and the high frequency modes is drastically reduced as forecasting horizon is extended. If the forecasting horizon is short, the admissible error in regard with low frequencies should be also restrictive, since its affectation to the final performance is high in terms of signal amplitude. Also, the admissible error for high frequencies should be reduced, since shorter horizons usually seek a more demanding modelling scenario where higher performances are sought.

The affectation of the dynamics to the signal modelling cannot be solved by the direct filtering of the high frequency modes, since the performance of the modelling depends, predominately, on the quantity of signals dynamics that want to be modelled, and not to the bandwidth of those modes. In this regard, dynamic modes to model imply a higher number of sources of information that need to be considered for the modelling, for this reason, as more dynamic modes are considered, more data is required for the proper training of the modelling, that is, an increase of overfitting risk. Therefore, the decomposition of the signal in its dynamic modes rather than filtering high frequency dynamics is required to increase the performance of the modelling.

In regard with the main limitations of the ANFIS approach, it has been demonstrated that ANFIS present performance problems in terms of model accuracy if signal dynamics are not considered during the model design. Furthermore, its convergence capabilities are also compromised if a considered number of inputs are required to be introduced in the ANFIS model. However, if it is assured that the input signals of the model present significant relations with the target signal, and do not present redundant information among them (inputs signals are mutually low-correlated), it is feasible to use modelling methods that consider signal dynamics as a

set of direct inputs to the model, maximizing with it the convergence to the optimal training solution with a reduced dataset. This is the case of Neo-fuzzy neuron modelling approach, that has been classically used in other research fields, where input signals are clearly independent. Nevertheless, it was a novel underexploited method when dealing with industrial time series forecasting.

Thus, the forecasting modelling proposed includes the decomposition of the target signal by the utilization of the EMD, which adaptive nature provides a set of IMF as an output of the method. Despite those IMF are generally uncorrelated, the use of advances variations of the EMD, as the EEMD, allows to maximize those characteristics, and avoid mode mixing in the extraction of the set of IMFs. It should be considered that the proposed approach has been stated as a pure time series problem in order to allow direct comparison between the performance of the proposed method and the approaches available in the current state of the art.

In this regard, the experimental validation has been performed in two case studies, a scientific platform commonly used in the literature, and the copper rod manufacturing process, object of this thesis research. In such scenarios, the proposed method reduces in more than 50% the forecasting error in terms of MAPE, by comparison with the approach proposed in the literature. It should be remarked that the validation procedure of the model training has been performed with two different datasets, providing different operational periods of the target signal to model. This validation approach allows testing the generalization capabilities of the model in front of new data from the same process but with a different dynamic distribution.

The current state of the art in regard with future condition forecasting is based on the classical diagnosis scheme, and is approached by adding a diagnosis evolution analysis model. That is to track the class information given by the classifier in time, in order to retrieve metrics such as the time spent in the specific class, the remaining time to change to another class, etc. This specific approach, despite its simplicity, presents severe limitations dealing with the assessment of the future condition, mainly due to low performance of the forecasting motivated by a low resolution in the input information. In this regard, this specific approach is useful dealing with applications in which the condition of the system is characterized among its operational life. However, the proposed approach in which the assessment of the future condition is given in regard with the forecasted value of the target signals considered, represents a more adequate method for industrial processes, where non-monotonic behaviours are observed. That is, the condition of the process itself is reversible, and despite degradation approaches, it is able to vary from one condition to another in regard with the behaviour of its associated signals.

Indeed, for enhancing the performance of the future condition assessment, the defined methodology proposes the use of NFN in a first stage in order to forecast the behaviour of the target signals. In this sense, it has been previously demonstrated that it offers better performance that the classical ANFIS approach in terms of model accuracy and generalization capabilities. In regard with the reliability of the condition forecasting assessment, it has been considered to include during the diagnosis procedure a quantization of the similitude degree between the observation to evaluate and the initial knowledge used for the training of the models. To do so, it has been proposed the use of non-linear topology preservation mapping for modelling target signal information. That, as has been previously mentioned, it allows characterizing the underlying physical behaviours of the inputs by discretizing their associated space by a neural grid. Therefore, the SOM allows fusing the behaviours of the process by integrating the forecasting given by the NFN to assess the future condition. Unlike

classical diagnosis approach based on a neural network, that label the totality of the input space, SOM offers an adaptation of the neurons of the grid to the intrinsic input data topology.

In this regard, the proposed method, for the validation cases considered, outperforms the approach of the literature based on ANFIS an NN in 10% in terms of classification score. Furthermore, in regard with the misclassifications, 40% of misclassified data presents a similitude value under 50%, which indicates the outlier nature of the missed samples. Indeed, condition forecasting implies to deal with two sources of errors that are considered to be affected in cascade, the classification and the modelling errors. Therefore, forecasting error implies a high change of misclassifications due to a sudden modification of the topology of the input data space in regard with the current behaviour of the process.

As global conclusions from the proposed thesis work, it has been demonstrated the opportunity to enhance the classical future condition approaches present on the literature. Indeed, the analysis and management of the available signals to perform condition forecasting represent a key factor in the proposed methodology that, classically, has been unattended. The proposal of the codification of the auxiliary signals, and the decomposition of the target signal, allows the proposed method to use significant and non-correlated set of inputs in the modelling. The impact of the proposed method resides on the way that auxiliary signals are treated, since they are processed in an independent way in regard with the target signal modelling procedure. Indeed, the proposed method allows to compress the behaviour defined by the auxiliary signals in a single feature. Such feature is independent from the modelling method used and, therefore, can be used and exported to any modelling approach. This has also a high potential in the scalability of the modelling, auxiliary signals can be updated or new auxiliary signals can be incorporated to the mapping approach, since the structure of the final forecasting model will not be altered, only retraining is required.

The modular capabilities of the proposed industrial time series forecasting approach has been prioritized. In this regard, the potentiality of the neo-fuzzy neuron relies on the capability of incorporating and managing a high number of inputs without losing convergence performance. Therefore, the proposed approach for managing auxiliary signals and forecasting is the most appropriate for industrial environments, in which the integrity of the available database is limited. This implies that the number of observations available to learn the different behaviours reflexed in the database is relatively low in order to represent all the characteristic variability of the industrial processes.

Classically, diagnosis approaches are based on previous knowledge about the different conditions of the monitored processes, typically, based on electromechanical actuators. In this regard, each class implies a maintenance action necessary to bring the system back to the nominal condition. Thus, there is a direct relation between the cause (condition), and the effect (maintenance action required). However, this relation dealing with industrial manufacturing processes is complex and difficult to be assessed, since different signal behaviours may lead to the same condition (asset quality level), and, therefore, each combination of inputs may require a specific maintenance action. This is the case of the experimental plant used in this thesis, the copper rod manufacturing process, in which the traceability of the class associated to the underlying physical effect is required. This is due to the fact that different combinations of the signals from the heat extraction and the solidification procedures may lead to reach the same output copper rod quality level.

Therefore, classical classification approaches based on the mathematical characterization of the input data space in order to define the optimal boundaries, imply a loss of traceability between the physical effects defined by the considered signals and the condition of the system. Thus, for industrial manufacturing processes, as seen in this thesis, it is proposed the use of methodologies that allow to know at every moment the existent relation between the value of the signals and the condition of the process. Although this may require additional post-processing efforts, the affectation maps resulting of defining the topographical grid of the SOM map allow the possibility of offering such information to the technical experts of the process.

5.2 Future work

The main line of future work in relation with the proposed thesis results consist on the study and analysis of the capacity of hidden pattern extraction procedures that, in an automatic way, identify extract and give information in regard with which auxiliary signals and corresponding dynamics are more significant for the forecasting task. This work can be approached by means of advanced artificial intelligence methods such as recurrent neural networks combined with deep learning methodologies, that, considering current state of the art, would allow giving support information in the interpretation of the signal relations of the industrial process.

Another interesting future work to be considered refers to the integration and implementation procedures of the proposed condition forecasting schemes into industrial platforms, which represents a techno-scientific challenge in which the proposed algorithm should be adequate and optimized to the requirements and characteristics of digital platforms.

Furthermore, although in this thesis a procedure to quantify the diagnosis outcome reliability has been proposed in regard with the quantization error of the SOM grid, the management of those outliers, non-previously considered during the training phase of the model, represents not-known data which has great potentiality to be included in the knowledge of the models. The concept of integrating new knowledge in the models as soon as it becomes available in the industrial process belongs to a scientific field named novelty detection. In this regard, novelty detection could complement the proposed condition forecasting approaches with the capability to identify and assimilate new knowledge in an on-line manner. For this reason, the identification of novel knowledge, the management and the incorporation of such novelties in the condition forecasting approach is also considered as a future field of study following the contributions of the proposed thesis.

In general terms, the performed research of this thesis work has been located on data acquisition and pre-processing, the forecasting model generation and the future condition assessment. In this regard, with the achieved results, the proposed methodologies can be used for fault tolerant and optimal control of the manufacturing process, or for optimal maintenance actions scheduling. That is, the resulting condition information can be analyzed in terms of suitability to develop model predictive control approaches, and the generation of advanced expert systems for real time optimization of the maintenance procedures in regard with the expected future condition of the industrial process. Thus, the identification and incorporation of the requirements defined by such applications to the proposed condition forecasting approach represents also a challenging ways of continuation.

6.

Thesis results dissemination

The direct contributions resulting from this Thesis work, in international journals as wells as in specialized conferences, are collected in this Chapter. Additionally, the contributions in research projects related with the thesis topics are also briefly exposed.

CONTENTS:

6. Thesis results dissemination

6.1 Publications: Thesis contributions

Dissemination results directly related with the thesis contributions

Journals

D. Zurita-Millán, M. Delgado-Prieto, J.J. Saucedo-Dorantes, J. A. Cariño-Corrales, R. A. Osornio-Rios, J. A. Ortega-Redondo, R.J. Romero-Troncoso. "Vibration Signal Forecasting on Rotating Machinery by means of Signal Decomposition and Neuro-fuzzy Modelling," *Shock Vib.,* vol. 2016, pp. 1–13, 2016.

D. Zurita-Millán, M. Delgado, J. A. Carino, J. A. Ortega, and G. Clerc, "Industrial Time Series Modelling by means of the Neo-Fuzzy Neuron," *IEEE Access*, vol. 4, pp. 6151 - 6160, 2016.

D. Zurita-Millán, M. Delgado, J. A. Carino, J. A. Ortega, R.J. Romero-Troncoso, H. Razik, G. Clerc, "Ensemble Empirical Mode Decomposition and Neo-Fuzzy Neuron for Industrial Time Series Forecasting," *IEEE Transactions on industrial informatics.* **Under review.**

D. Zurita-Millán, M. Delgado, J. A. Carino, J. A. Ortega, "Multimodal Forecasting Methodology applied to Industrial Process Monitoring," *IEEE Transactions on industrial informatics.* **Under review.**

D. Zurita-Millán, M. Delgado, J. A. Carino, J. A. Ortega, H. Razik, G. Clerc, "Industrial Process Condition Forecasting by Neo-Fuzzy Neuron and Self-Organizing Maps," *IEEE Access.* **Under review.**

Conferences

D. Zurita-Millán, J. A. Carino, M. Delgado, and J. A. Ortega, "Distributed neuro-fuzzy feature forecasting approach for condition monitoring," *in Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, pp. 1–8, 2014.

D. Zurita-Millán, J. A. Carino, E. Sala, M. Delgado-Prieto, and J. A. Ortega, "Time series forecasting by means of SOM aided Fuzzy Inference Systems," *in 2015 IEEE International Conference on Industrial Technology (ICIT)*, pp. 1772–1778, 2015.

D. Zurita-Millán, J.A. Carino; E. Sala, M. Delgado, M.; Ortega, J.A. Enhanced Time Series Forecasting by means of Dynamics Boosting applied to Industrial Process Monitoring, in *2015 IEEE 10th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)*, pp. 212-218, 2015.

D. Zurita-Millán, J.A. Carino, A. Picot, M. Delgado, J.A. Ortega. "Diagnosis Method based on Topology Codification and Neural Network applied to an Industrial Camshaft". *2015 IEEE 10th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)*, pp. 124-130, 2015.

D. Zurita-Millán, E. Sala, E.; J.A. Carino, M. Delgado, J.A. Ortega. " Industrial Process Monitoring by means of Recurrent Neural Networks and Self Organizing Maps" *Emerging Technology and Factory Automation (ETFA), 2016 IEEE,* pp.1-8, 2016.

E. Sala, **D. Zurita-Millán**, K. Kampouropoulos, M. Delgado-Prieto, and L. Romeral, "Enhanced load forecasting methodology by means of probabilistic prediction intervals estimation," *in 2015 IEEE International Conference on Industrial Technology (ICIT)*, pp. 1299–1304, 2015.

J. A. Carino, **D. Zurita-Millán**, M. Delgado, J. A. Ortega, and R. J. Romero-Troncoso, "Remaining useful life estimation of ball bearings by means of monotonic score calibration," *in 2015 IEEE International Conference on Industrial Technology (ICIT)*, pp. 1752–1758, 2015.

6.2 Publications: Collaborations and other works

Dissemination related to collaborations in the topic

Journals

M. Delgado Prieto, **D. Zurita-Millán**, W. Wang, A. Machado Ortiz, J. A. Ortega Redondo and L. Romeral Martinez, "Self-Powered Wireless Sensor Applied to Gear Diagnosis Based on Acoustic Emission," *in IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 1, pp. 15-24, Jan. 2016.

J. A. Cariño-Corrales, J.J. Saucedo-Dorantes, **D. Zurita-Millán**, M. Delgado-Prieto, R. A. Osornio-Rios, J. A. Ortega-Redondo, R.J. Romero-Troncoso. "Vibration-based adaptive novelty detection method for monitoring faults in a kinematic chain"". *Shock Vib.,* vol. 2016, pp. 1–13, 2016.

J. A. Carino, M. Delgado-Prieto, **D. Zurita-Millán**, M. Millan, J. A. Ortega Redondo and R. Romero-Troncoso, "Enhanced Industrial Machinery Condition Monitoring Methodology Based on Novelty Detection and Multi-Modal Analysis," *in IEEE Access*, vol. 4, pp. 7594-7604, 2016.

M. Delgado, **D. Zurita-Millán**, "Chromatic Monitoring of Gear Mechanical Degradation based on Acoustic Emission," *IEEE Transactions on Industrial Electronics*, vol. PP, no.99, p-1-1, 2017.

Conferences

D. Zurita-Millán, M. Delgado, J. A. Ortega, and L. Romeral, "Intelligent sensor based on acoustic emission analysis applied to gear fault diagnosis," in 2013 9th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives (SDEMPED), pp. 169–176, 2013.

J. A. Carino, **D. Zurita-Millán**, M. Delgado, J. A. Ortega, and R. J. Romero-Troncoso, "Hierarchical classification scheme based on identification, isolation and analysis of conflictive regions," in Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA), pp. 1–8, 2014.

A. Picot, **D. Zurita-Millán**, J. Cariño, E. Fournier, J. Régnier and J. A. Ortega, "Industrial machinery diagnosis by means of normalized time-frequency maps," 2015 IEEE 10th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED), Guarda, pp. 158-164, 2015.

J.A. Carino, **D. Zurita-Millán**, A. Picot, M. Delgado, J. A. Ortega, R.J. Romero-Troncoso. "Novelty Detection Methodology based on Multi-Modal One-class Support Vector Machine", *2015 IEEE 10th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED), Guarda, pp. 184-190, 2015.*

E. Sala, **D. Zurita-Millán**, K. Kampouropoulos, M. Delgado and L. Romeral, "Occupancy forecasting for the reduction of HVAC energy consumption in smart buildings," *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, Florence, 2016, pp. 4002-4007.

Book Chapters

M. Delgado Prieto, J. A. Cariño Corrales, **D. Zurita-Millán**, M. Millán Gonzalvez, J. A. Ortega Redondo and R. J. Romero Troncoso. Evaluation of Novelty Detection Methods for Condition Monitoring applied to an Electromechanical System.

References

References

[1] S. Yin, X. Li, H. Gao, and O. Kaynak, “Data-Based Techniques Focused on Modern Industry: An Overview,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 657–667, Jan. 2015.

[2] S. Wang, J. Wan, D. Zhang, D. Li, and C. Zhang, “Towards smart factory for Industry 4.0: A self-organized multi-agent system with big data based feedback and coordination,” *Computer Networks*, vol. 101, pp. 158–168, 2015.

[3] K. Zhou, “Industry 4.0: Towards future industrial opportunities and challenges,” in *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2015, pp. 2147–2152.

[4] M. Holloway and C. Nwaoha, “Dictionary of Industrial Terms,” in *Dictionary of Industrial Terms*, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2013, pp. 1–165.

[5] K.-S. Wang, “Towards zero-defect manufacturing (ZDM)—a data mining approach,” *Adv. Manuf.*, vol. 1, no. 1, pp. 62–74, 2013.

[6] J. F. Halpin, *Zero defects: a new dimension in quality assurance*. McGraw-Hill, 1966.

[7] E. Westkämper and H.-J. Warnecke, “Zero-Defect Manufacturing by Means of a Learning Supervision of Process Chains,” *CIRP Ann. - Manuf. Technol.*, vol. 43, no. 1, pp. 405–408, 1994.

[8] O. Myklebust, “Zero Defect Manufacturing: A Product and Plant Oriented Lifecycle Approach,” *Procedia CIRP*, vol. 12, pp. 246–251, 2013.

[9] G. Medina-Oliva, P. Weber, and B. Iung, “Industrial system knowledge formalization to aid decision making in maintenance strategies assessment,” *Eng. Appl. Artif. Intell.*, vol. 37, pp. 343–360, Jan. 2015.

[10] R. Zhang, R. Lu, A. Xue, and F. Gao, “New Minmax Linear Quadratic Fault-Tolerant Tracking Control for Batch Processes,” *IEEE Trans. Automat. Contr.*, vol. 61, no. 10, pp. 3045–3051, Oct. 2016.

[11] Aurubis, “Continuous Casting of Copper Rod.” [Online]. Available: https://www.aurubis.com/en/en/corp/products/rod--specialty-wire/rod. [Accessed: 07-May-2017].

[12] W. Y. Wang K, “Data mining for zero–defect manufacturing,” *Tpir Acad. Press London*, 2012.

[13] J. Lee, E. Lapira, B. Bagheri, and H. Kao, “Recent advances and trends in predictive manufacturing systems in big data environment,” 2013.

[14] R. Harrison, D. Vera, and B. Ahmad, “Engineering Methods and Tools for Cyber–Physical Automation Systems,” *Proc. IEEE*, vol. 104, no. 5, pp. 973–985, May 2016.

[15] J. Lee, B. Bagheri, and H.-A. Kao, “A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems,” 2015.

[16] P. J. Mosterman and J. Zander, “Industry 4.0 as a Cyber-Physical System study,” *Softw. Syst. Model.*, vol. 15, no. 1, pp. 17–29, Feb. 2016.

[17] K. Kampouropoulos, F. Andrade, A. Garcia, and L. Romeral, “A Combined Methodology of Adaptive Neuro-Fuzzy Inference System and Genetic Algorithm for Short-term Energy Forecasting,” *Adv. Electr. Comput. Eng.*, vol. 14, no. 1, pp. 9–14, 2014.

[18] M. S. Kan, A. C. C. Tan, and J. Mathew, “A review on prognostic techniques for non-stationary and non-linear rotating systems,” *Mech. Syst. Signal Process.*, vol. 62–63, pp. 1–20, Oct. 2015.

[19] A. K. Mahamad, S. Saon, and T. Hiyama, “Predicting remaining useful life of rotating machinery based artificial neural network,” *Comput. Math. with Appl.*, vol. 60, no. 4, pp. 1078–1087, 2010.

[20] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, “Fault prognostic of bearings by using support

vector data description," in *Prognostics and Health Management (PHM), 2012 IEEE Conference on*, 2012, pp. 1–7.

[21] C. Chen *et al.*, "Machine condition prediction based on adaptive neuro–fuzzy and high-order particle filtering," *IEEE Trans. Ind. Electron.*, vol. 58, no. 9, pp. 4353–4364, 2011.

[22] A. Soualhi, H. Razik, G. Clerc, and D. D. Doan, "Prognosis of Bearing Failures Using Hidden Markov Models and the Adaptive Neuro-Fuzzy Inference System," *IEEE Trans. Ind. Electron.*, vol. 61, no. 6, pp. 2864–2874, Jun. 2014.

[23] A. Widodo and B.-S. Yang, "Machine health prognostics using survival probability and support vector machine," *Expert Syst. Appl.*, vol. 38, no. 7, pp. 8430–8437, 2011.

[24] S. Wooldridge, *Data processing made simple*. 2013.

[25] N. Dokuchaev, "On exact and optimal recovering of missing values for sequences," *Signal Processing*, vol. 135, pp. 81–86, 2017.

[26] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.

[27] M. Lebold and M. Thurston, "Open standards for Condition-Based Maintenance and Prognostic Systems," in *5th Annual Maintenance and Reliability Conference*, 2001.

[28] A. Hess, "Prognostics, from the need to reality-from the fleet users and PHM system designer/developers perspectives," in *Aerospace Conference Proceedings, 2002. IEEE*, 2002, vol. 6, pp. 6-2791-6–2797 vol.6.

[29] ISO13381-1, "Condition Monitoring and Diagnostics of Machines – Prognostics – Part 1: General Guidelines," *Int. Stand. Organ.*, 2004.

[30] G. Vachtsevanos, F. L. Lewis, M. Roemer, A. Hess, and B. Wu, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. Wiley, 2006.

[31] A. Heng, S. Zhang, A. C. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mech. Syst. Signal Process.*, vol. 23, no. 3, pp. 724–739, 2009.

[32] O. Elena Dragomir, R. Gouriveau, F. Dragomir, E. Minca, and N. Zerhouni, "Review of prognostic problem in condition-based maintenance," in *European Control Conference, ECC'09.*, 2009, vol. sur CD ROM, pp. 1585–1592.

[33] J. Z. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mech. Syst. Signal Process.*, vol. 25, no. 5, pp. 1803–1836, 2011.

[34] H. Chestnut, *Systems engineering tools*. John Wiley & Sons, New York, Library of Congress Catalog Card, 1965.

[35] S. Zhang, L. Ma, Y. Sun, and J. Mathew, "Asset health reliability estimation based on condition data," 2007.

[36] B. Saha, K. Goebel, S. Poll, and J. Christophersen, "Prognostics Methods for Battery Health Monitoring Using a Bayesian Framework," *Instrum. Meas. IEEE Trans.*, vol. 58, no. 2, pp. 291–296, 2009.

[37] G. Niu and B.-S. Yang, "Dempster–Shafer regression for multi-step-ahead time-series prediction towards data-driven machinery prognosis," *Mech. Syst. Signal Process.*, vol. 23, no. 3, pp. 740–751, 2009.

[38] W. He, N. Williard, M. Osterman, and M. Pecht, “Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method,” *J. Power Sources*, vol. 196, no. 23, pp. 10314–10321, 2011.

[39] A. Schömig and O. Rose, “On the suitability of the Weibull distribution for the approximation of machine failures,” in *Proceedings of the 2003 Industrial Engineering Research Conference*, 2003.

[40] Q. Zhang, C. Hua, and G. Xu, “A mixture Weibull proportional hazard model for mechanical system failure prediction utilising lifetime and monitoring data,” *Mech. Syst. Signal Process.*, vol. 43, no. 1–2, pp. 103–112, 2014.

[41] L. Jianhui, K. R. Pattipati, Q. Liu, and S. Chigusa, “Model-Based Prognostic Techniques Applied to a Suspension System,” *Syst. Man Cybern. Part A Syst. Humans, IEEE Trans.*, vol. 38, no. 5, pp. 1156–1168, 2008.

[42] C. H. Oppenheimer and K. A. Loparo, “Physically based diagnosis and prognosis of cracked rotor shafts,” 2002, vol. 4733, pp. 122–132.

[43] H. Park, N. Kim, and J. Lee, “Parametric models and non-parametric machine learning models for predicting option prices: Empirical comparison study over KOSPI 200 Index options,” *Expert Syst. Appl.*, vol. 41, no. 11, pp. 5227–5237, 2014.

[44] B. Reusch, Ed., *Computational Intelligence, Theory and Applications*, vol. 33. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[45] A. Abdoos, M. Hemmati, and A. A. Abdoos, “Short term load forecasting using a hybrid intelligent method,” *Knowledge-Based Syst.*, vol. 76, pp. 139–147, Mar. 2015.

[46] S. Barak and S. S. Sadegh, “Forecasting energy consumption using ensemble ARIMA–ANFIS hybrid algorithm,” *Int. J. Electr. Power Energy Syst.*, vol. 82, pp. 92–104, 2016.

[47] A. A. El Desouky and M. M. Elkateb, “Hybrid adaptive techniques for electric-load forecast using ANN and ARIMA,” *Gener. Transm. Distrib. IEE Proceedings-*, vol. 147, no. 4, pp. 213–217, 2000.

[48] M. Orchard, B. Wu, and G. Vachtsevanos, “A Particle Filtering Framework for Failure Prognosis,” in *3rd, World tribology congress*, 2005, p. 2; 883-884.

[49] T. Khan, L. Udpa, and S. Udpa, “Particle filter based prognosis study for predicting remaining useful life of steam generator tubing,” in *Prognostics and Health Management (PHM), 2011 IEEE Conference on*, 2011, pp. 1–6.

[50] W. Jinjiang and R. X. Gao, “Multiple model particle filtering for bearing life prognosis,” in *Prognostics and Health Management (PHM), 2013 IEEE Conference on*, 2013, pp. 1–6.

[51] B. E. Olivares, M. A. Cerda Munoz, M. E. Orchard, and J. F. Silva, “Particle-Filtering-Based Prognosis Framework for Energy Storage Devices With a Statistical Characterization of State-of-Health Regeneration Phenomena,” *Instrum. Meas. IEEE Trans.*, vol. 62, no. 2, pp. 364–376, 2013.

[52] S. Yin, S. X. Ding, X. Xie, and H. Luo, “A Review on Basic Data-Driven Approaches for Industrial Process Monitoring,” *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6418–6428, Nov. 2014.

[53] P. J. Werbos, “Generalization of backpropagation with application to a recurrent gas market model,” *Neural Networks*, vol. 1, no. 4, pp. 339–356, 1988.

[54] F. S. Wong, “Time series forecasting using backpropagation neural networks,” *Neurocomputing*, vol. 2, no. 4, pp. 147–159, 1991.

[55] T. Khawaja, G. Vachtsevanos, and B. Wu, “Reasoning about uncertainty in prognosis: a confidence prediction neural network approach,” in *Fuzzy Information Processing Society, 2005. NAFIPS 2005. Annual Meeting of the North American*, 2005, pp. 7–12.

[56] J. Abdi, B. Moshiri, and B. Abdulhai, “Emotional temporal difference Q-learning signals in multi-agent system cooperation: real case studies,” *Intell. Transp. Syst. IET*, vol. 7, no. 3, pp. 315–326, 2013.

[57] Z. Xiaodong, R. Xu, K. Chiman, S. Y. Liang, X. Qiulin, and L. Haynes, “An integrated approach to bearing fault diagnostics and prognostics,” in *American Control Conference, 2005. Proceedings of the 2005*, 2005, pp. 2750–2755 vol. 4.

[58] M. Dong and D. He, “A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology,” *Mech. Syst. Signal Process.*, vol. 21, no. 5, pp. 2248–2266, 2007.

[59] J. Lee, E. Lapira, B. Bagheri, and H. Kao, “Recent advances and trends in predictive manufacturing systems in big data environment,” *Manuf. Lett.*, vol. 1, no. 1, pp. 38–41, 2013.

[60] B. Wilamowski, C. Cecati, J. Kolbusz, P. Rozycki, and P. Siano, “A Novel RBF Training Algorithm for Short-term Electric Load Forecasting and Comparative Studies,” *IEEE Trans. Ind. Electron.*, vol. PP, no. 99, pp. 1–1, 2015.

[61] F. Guo, L. Lin, and C. Wang, “Novel continuous function prediction model using an improved Takagi–Sugeno fuzzy rule and its application based on chaotic time series,” *Eng. Appl. Artif. Intell.*, vol. 55, pp. 155–164, 2016.

[62] W. Wang, W. Pedrycz, and X. Liu, “Time series long-term forecasting model based on information granules and fuzzy clustering,” *Eng. Appl. Artif. Intell.*, vol. 41, pp. 17–24, May 2015.

[63] X. Deng, X. Zeng, P. Vroman, and L. Koehl, “Selection of relevant variables for industrial process modeling by combining experimental data sensitivity and human knowledge,” *Eng. Appl. Artif. Intell.*, vol. 23, no. 8, pp. 1368–1379, 2010.

[64] M. Khashei and M. Bijari, “Hybridization of the probabilistic neural networks with feed-forward neural networks for forecasting,” *Eng. Appl. Artif. Intell.*, vol. 25, no. 6, pp. 1277–1288, 2012.

[65] Z. Su, J. Wang, H. Lu, and G. Zhao, “A new hybrid model optimized by an intelligent optimization algorithm for wind speed forecasting,” *Energy Convers. Manag.*, vol. 85, pp. 443–452, 2014.

[66] A. Kusiak and Z. Zhang, “Short-Horizon Prediction of Wind Power: A Data-Driven Approach,” *IEEE Trans. Energy Convers.*, vol. 25, no. 4, pp. 1112–1122, Dec. 2010.

[67] I. Koprinska, M. Rana, and V. G. Agelidis, “Correlation and instance based feature selection for electricity load forecasting,” *Knowledge-Based Syst.*, vol. 82, pp. 29–40, 2015.

[68] E. Ceperic, V. Ceperic, and A. Baric, “A Strategy for Short-Term Load Forecasting by Support Vector Regression Machines,” *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4356–4364, Nov. 2013.

[69] K. Kampouropoulos, J. J. Cardenas, F. Giacometto, and L. Romeral, “An energy prediction method using Adaptive Neuro-Fuzzy Inference System and Genetic Algorithms,” in *Industrial Electronics (ISIE), 2013 IEEE International Symposium on*, 2013, pp. 1–6.

[70] A. Soualhi, H. Razik, G. Clerc, and D. D. Doan, “Prognosis of Bearing Failures Using Hidden Markov Models and the Adaptive Neuro-Fuzzy Inference System,” *Ind. Electron. IEEE Trans.*, vol. 61, no. 6, pp. 2864–2874, 2014.

[71] H. Taheri Shahraiyni, S. Sodoudi, A. Kerschbaumer, and U. Cubasch, “A new structure identification scheme for ANFIS and its application for the simulation of virtual air pollution monitoring stations in urban areas,” *Eng. Appl. Artif. Intell.*, vol. 41, pp. 175–182, 2015.

[72] C. Li and J.-W. Hu, “A new ARIMA-based neuro-fuzzy approach and swarm intelligence for time series forecasting,” *Eng. Appl. Artif. Intell.*, vol. 25, no. 2, pp. 295–308, Mar. 2012.

[73] J. S. R. Jang, “ANFIS: adaptive-network-based fuzzy inference system,” *Syst. Man Cybern. IEEE Trans.*, vol. 23, no. 3, pp. 665–685, 1993.

[74] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” *Syst. Man Cybern. IEEE Trans.*, vol. SMC-15, no. 1, pp. 116–132, 1985.

[75] W. Q. Wang, M. F. Golnaraghi, and F. Ismail, “Prognosis of machine health condition using neuro-fuzzy systems,” *Mech. Syst. Signal Process.*, vol. 18, no. 4, pp. 813–831, 2004.

[76] A. Davydenko and R. Fildes, “Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts,” *Int. J. Forecast.*, vol. 29, no. 3, pp. 510–522, Jul. 2013.

[77] C.-H. Su and C.-H. Cheng, “A hybrid fuzzy time series model based on ANFIS and integrated nonlinear feature selection method for forecasting stock,” *Neurocomputing*, vol. 205, pp. 264–273, 2016.

[78] J. H. (John H. Holland, *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*. MIT Press, 1992.

[79] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Elsevier Science, 2008.

[80] M. Delgado Prieto, D. Zurita Millan, W. Wang, A. Machado Ortiz, J. A. Ortega Redondo, and L. Romeral Martinez, “Self-Powered Wireless Sensor Applied to Gear Diagnosis Based on Acoustic Emission,” *IEEE Trans. Instrum. Meas.*, vol. 65, no. 1, pp. 15–24, Jan. 2016.

[81] Y. Ren, P. N. Suganthan, and N. Srikanth, “A Comparative Study of Empirical Mode Decomposition-Based Short-Term Wind Speed Forecasting Methods,” *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 236–244, Jan. 2015.

[82] L.-Y. Wei, “A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting,” *Appl. Soft Comput.*, vol. 42, pp. 368–376, 2016.

[83] N. E. Huang *et al.*, *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*, vol. 454, no. 1971. 1998.

[84] Z. Guo, W. Zhao, H. Lu, and J. Wang, “Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model,” *Renew. Energy*, vol. 37, no. 1, pp. 241–249, Jan. 2012.

[85] S. Dong and T. Luo, “Bearing degradation process prediction based on the PCA and optimized LS-SVM model,” *Measurement*, vol. 46, no. 9, pp. 3143–3152, 2013.

[86] R. Shao, W. Hu, Y. Wang, and X. Qi, “The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform,” *Measurement*, vol. 54, pp. 118–132, Aug. 2014.

[87] T. Kohonen, “The self-organizing map,” *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[88] C. W. Frey, “Monitoring of complex industrial processes based on self-organizing maps and watershed transformations,” in *Industrial Technology (ICIT), 2012 IEEE International Conference on*, 2012, pp. 1041–1046.

[89] L. Weihua, Z. Shaohui, H. Guolin, W. Li, S. Zhang, and G. He, “Semisupervised Distance-Preserving

Self-Organizing Map for Machine-Defect Detection and Classification," *Instrum. Meas. IEEE Trans.*, vol. 62, no. 5, pp. 869–879, 2013.

[90] C. M. Bishop, *Neural Networks for Pattern Recognition*. Clarendon Press, 1995.

[91] B. Yusob, S. M. H. Shamsuddin, and H. N. A. Hamed, "Spiking Self-organizing Maps for Classification Problem," *Procedia Technol.*, vol. 11, pp. 57–64, 2013.

[92] L. Baldacci, M. Golfarelli, D. Lombardi, and F. Sami, "Natural gas consumption forecasting for anomaly detection," *Expert Syst. Appl.*, vol. 62, pp. 190–201, 2016.

[93] D. Zurita-Millan *et al.*, "Vibration signal forecasting on rotating machinery by means of signal decomposition and neuro-fuzzy modeling," *Shock Vib.*, 2016.

[94] Z. Hu, Y. Bao, T. Xiong, and R. Chiong, "Hybrid filter–wrapper feature selection for short-term load forecasting," *Eng. Appl. Artif. Intell.*, vol. 40, pp. 17–27, Apr. 2015.

[95] F. Keynia, "A new feature selection algorithm and composite neural network for electricity price forecasting," *Eng. Appl. Artif. Intell.*, vol. 25, no. 8, pp. 1687–1697, 2012.

[96] A. D. Back and T. P. Trappenberg, "Selecting inputs for modeling using normalized higher order statistics and independent component analysis," *IEEE Trans. Neural Networks*, vol. 12, no. 3, pp. 612–617, May 2001.

[97] F. M. Bianchi, E. De Santis, A. Rizzi, and A. Sadeghian, "Short-Term Electric Load Forecasting Using Echo State Networks and PCA Decomposition," *IEEE Access*, vol. 3, pp. 1931–1943, 2015.

[98] C. Brighenti and M. Á. Sanz-Bobi, "Auto-regressive processes explained by self-organized maps. Application to the detection of abnormal behavior in industrial processes.," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2078–90, Dec. 2011.

[99] M. Domínguez, J. J. Fuertes, I. Díaz, M. A. Prada, S. Alonso, and A. Morán, "Monitoring industrial processes with SOM-based dissimilarity maps," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7110–7120, 2012.

[100] Q. Zhang, F. Liu, X. Wan, and G. Xu, "An Adaptive Support Vector Regression Machine for the State Prognosis of Mechanical Systems," *Shock Vib.*, 2015.

[101] D. Wang, C. Li, A. Widodo, P. K. Kankar, and W. Caesarendra, "Fault Diagnosis and Prognosis of Critical Components," *Shock Vib.*, vol. 2016, 2015.

[102] V. S. Kodogiannis, T. Pachidis, and E. Kontogianni, "An intelligent based decision support system for the detection of meat spoilage," *Eng. Appl. Artif. Intell.*, vol. 34, pp. 23–36, Sep. 2014.

[103] H. A. Zamani, S. Rafiee-Taghanaki, M. Karimi, M. Arabloo, and A. Dadashi, "Implementing ANFIS for prediction of reservoir oil solution gas-oil ratio," *J. Nat. Gas Sci. Eng.*, vol. 25, pp. 325–334, 2015.

[104] Y. Ren, P. N. Suganthan, and N. Srikanth, "A Comparative Study of Empirical Mode Decomposition-Based Short-Term Wind Speed Forecasting Methods," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 236–244, Jan. 2015.

[105] R.-A. Hooshmand, H. Amooshahi, and M. Parastegari, "A hybrid intelligent algorithm based short-term load forecasting approach," *Int. J. Electr. Power Energy Syst.*, vol. 45, no. 1, pp. 313–324, 2013.

[106] L. Cao, A. Mees, and K. Judd, "Dynamics from multivariate time series," *Phys. D Nonlinear Phenom.*, vol. 121, no. 1–2, pp. 75–88, Oct. 1998.

[107] M. Yu, D. Wang, and M. Luo, "An Integrated Approach to Prognosis of Hybrid Systems With Unknown Mode Changes," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 503–515, Jan. 2015.

[108] T. Xiong, Y. Bao, and Z. Hu, “Interval forecasting of electricity demand: A novel bivariate EMD-based support vector regression modeling framework,” *Int. J. Electr. Power Energy Syst.*, vol. 63, pp. 353–362, Dec. 2014.

[109] H. Liu, H. Tian, and Y. Li, “Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms,” *Energy Convers. Manag.*, vol. 100, pp. 16–22, Aug. 2015.

[110] L. Yu, Z. Wang, and L. Tang, “A decomposition–ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting,” *Appl. Energy*, vol. 156, pp. 251–267, Oct. 2015.

[111] J. J. Wang, W. Zhang, Y. Li, J. J. Wang, and Z. Dang, “Forecasting wind speed using empirical mode decomposition and Elman neural network,” *Appl. Soft Comput.*, vol. 23, pp. 452–459, Oct. 2014.

[112] H. Hanli Wei, Y. Yongmao Xu, and R. Re Zhang, “Neural networks based model predictive control of an industrial polypropylene process,” in *Proceedings of the International Conference on Control Applications*, 2002, vol. 1, pp. 397–402.

[113] M. V. V. N. Sriram, N. K. Singh, and G. Rajaraman, “Neuro fuzzy modelling of Basic Oxygen Furnace and its comparison with Neural Network and GRNN models,” in *2010 IEEE International Conference on Computational Intelligence and Computing Research*, 2010, pp. 1–8.

[114] V. Kolodyazhniy, Y. Bodyanskiy, and P. Otto, “Universal Approximator Employing Neo-Fuzzy Neurons,” in *Computational Intelligence, Theory and Applications*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 631–640.

[115] Y. Bodyanskiy, Y. Bodyanskiy, I. Kokshenev, V. Kolodyazhniy, and V. K. Ye. Bodyanskiy, I. Kokshenev, *An Adaptive Learning Algorithm for a Neo Fuzzy Neuron*. 2003, pp. 375–379.

[116] Y. V. Bodyanskiy, O. K. Tyshchenko, and D. S. Kopaliani, “A Multidimensional Cascade Neuro-Fuzzy System with Neuron Pool Optimization in Each Cascade,” *Int. J. Inf. Technol. Comput. Sci.*, vol. 6, no. 8, p. 11, 2014.

[117] Y. Bodyanskiy, I. Pliss, and O. Vynokurova, “Flexible Neo-fuzzy Neuron and Neuro-fuzzy Network for Monitoring Time Series Properties,” *Inf. Technol. Manag. Sci.*, vol. 16, no. 1, pp. 47–52, Jan. 2013.

[118] K. T. Chaturvedi, M. Pandit, and L. Srivastava, “Modified neo-fuzzy neuron-based approach for economic and environmental optimal power dispatch,” *Appl. Soft Comput.*, vol. 8, no. 4, pp. 1428–1438, 2008.

[119] A. Soualhi, G. Clerc, H. Razik, and F. Rivas, “Long-term prediction of bearing condition by the neo-fuzzy neuron,” in *2013 9th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives (SDEMPED)*, 2013, pp. 586–591.

[120] T. Yamakawa, E. Uchino, T. Miki, and H. Kusanagi, “A neo fuzzy neuron and its applications to system identification and prediction of the system behavior,” in *Proc. 2nd Int. Conf. on Fuzzy Logic and Neural Networks*, 1992, pp. 477–483.

[121] E. Uchino and T. Yamakawa, “Soft Computing Based Signal Prediction, Restoration, and Filtering,” in *Intelligent Hybrid Systems*, Boston, MA: Springer US, 1997, pp. 331–351.

[122] T. Miki and T. Yamakawa, “Analog Implementation of Neo-Fuzzy Neuron and Its On-board Learning.”

[123] E. Uchino and T. Yamakawa, “System Modeling by a Neo-Fuzzy-Neuron with Applications to Acoustic and Chaotic Systems,” *Int. J. Artif. Intell. Tools*, vol. 4, no. 01n02, pp. 73–91, Jun. 1995.

[124] A. M. Silva, W. Caminhas, A. Lemos, and F. Gomide, “A fast learning algorithm for evolving neo-fuzzy neuron,” *Appl. Soft Comput.*, vol. 14, pp. 194–209, 2014.

[125] Y. V. Bodyanskiy, O. K. Tyshchenko, and D. S. Kopaliani, “An Extended Neo-Fuzzy Neuron and its Adaptive Learning Algorithm,” *Int. J. Intell. Syst. Appl.*, vol. 7, no. 2, p. 21, 2015.

[126] Z. Wu and N. E. Huang, “Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method,” *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.

[127] D. Camarena-Martinez, J. P. Amezquita-Sanchez, M. Valtierra-Rodriguez, R. J. Romero-Troncoso, R. A. Osornio-Rios, and A. Garcia-Perez, “EEMD-MUSIC-based analysis for natural frequencies identification of structures using artificial and natural excitations.,” *ScientificWorldJournal.*, vol. 2014, p. 587671, 2014.

[128] J. J. Downs and E. F. Vogel, “A plant-wide industrial process control problem,” *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, 1993.

[129] A. Bathelt, N. L. Ricker, and M. Jelali, “Revision of the Tennessee Eastman Process Model,” *IFAC-PapersOnLine*, vol. 48, no. 8, pp. 309–314, 2015.

[130] C. Sankavaram, A. Kodali, K. R. Pattipati, and S. Singh, “Incremental Classifiers for Data-Driven Fault Diagnosis Applied to Automotive Systems,” *IEEE Access*, vol. 3, pp. 407–419, 2015.

[131] J. Huang, G. Chen, L. Shu, S. Wang, and Y. Zhang, “An Experimental Study of Clogging Fault Diagnosis in Heat Exchangers Based on Vibration Signals,” *IEEE Access*, vol. 4, pp. 1800–1809, 2016.

[132] Z. Ge and J. Chen, “Plant-Wide Industrial Process Monitoring: A Distributed Modeling Framework,” *IEEE Trans. Ind. Informatics*, vol. 12, no. 1, pp. 310–321, 2016.

Annexes

The annexes of this thesis are related with the definition of the experimental test benches used for the development of the proposed thesis.

CONTENTS:

A.I Copper rod manufacturing process

In regard with the industrial process considered in this thesis, the copper rod manufacturing process, manufacturing process implies the transformation of the melted input copper in a solid copper rod. Such transformation is based on a controlled solidification and copper roughing process. However, the solidification of the copper is a critical aspect within the manufacturing process, in which the excessive heat should be correctly extracted from the copper bar. In this regard, during the proposed thesis, the following set of signals has been considered in regard with the two different analysed parts of the process.

- **Signal regarding the tundish process**

The tundish process summarizes the information of the melted copper from the shaft furnace to its last step before entering to the casting wheel. The process signals available in regard with the tundish process are listed in **Table AI. 1.**

Table AI. 1. Main signals related with the tundish process.

Nom.	***Target/Auxiliary***	***Description***
$T_{tu}(t)$	*Target signal*	Tundish temperature, input measurement of the temperature of the melted copper before starting the solidification procedure
$W_{tu}(t)$	*Auxiliary signal*	Copper weight in tundish, corresponds to a measure of the total copper weight available in the tundish [Kg].
$O_{tu}(t)$	*Auxiliary signal*	Copper oxygen in copper, it is a measure in regard que the combustion quality of the melted copper inside the tundish [ppm].
$R_{tu}(t)$	*Auxiliary signal*	Ratio of gas/air in the main burner of the tundish. It is used to know the set point of the tundish temperature control.

- **Signals regarding the heat extraction process**

The heath extraction process is located after the entering of the melted copper by the tundish' valve. It considers the casting wheel, in which a water-based refrigeration procedure is in charge of extracting the heath of the melted copper to conform the final copper bar. The main signals available in the heat extraction procedure are listed in **Table AI. 2.**

Table AI. 2. Main signals related with the heat extraction process.

Nom.	*Target/Auxiliary*	*Description*
I_{cw} *(t)*	Target Signal	Heat extraction index of the casting wheel. It is measured as the difference between the input water temperature and the resulting water after the refrigeration of the wheel [ºC].
T_{ba} *(t)*	Target Signal	Temperature of the copper rod bar before entering to the roughing process. It represents the output temperature of the solidification process [ºC].
F_{cw} *(t)*	Target Signal	Total water flow introduced to the solidification process. It shows deviations from the nominal configuration if the heat extraction is not properly done [l/min].
W_{sP1} **(t, 1)**	*Auxiliary signal*	Average temperature of the casting wheel during the solidification process [ºC].
W_{sP1} **(t, 2)**	*Auxiliary signal*	Water flow in the interior part of the wheel [l/min].
W_{sP1} **(t, 3)**	*Auxiliary signal*	Water pressure in the interior of the wheel [kPa]
W_{sP1} **(t, 4)**	*Auxiliary signal*	Water flow in the exterior of the wheel [l/min].
W_{sP1} **(t, 5)**	*Auxiliary signal*	Water pressure in the exterior of the wheel [kPa].
W_{sP1} **(t, 6)**	*Auxiliary signal*	Water flow in the lateral of the wheel [l/min].
W_{sP1} **(t, 7)**	*Auxiliary signal*	Water pressure in the lateral of the wheel [kPa].
W_{sP1} **(t, 8)**	*Auxiliary signal*	Water flow in the center of the wheel [l/min].
W_{sP1} **(t, 9)**	*Auxiliary signal*	Water pressure in the center of the wheel [kPa].
A_{sP2} **(t, 1)**	*Auxiliary signal*	Temperature of the steel band enclosure [ºC]
A_{sP2} **(t, 2)**	*Auxiliary signal*	Acetylene flow in the exterior of the wheel [l/min]
A_{sP2} **(t, 3)**	*Auxiliary signal*	Acetylene pressure in the exterior [kPa].
A_{sP2} **(t, 4)**	*Auxiliary signal*	Acetylene flow in the interior [l/min]
A_{sP2} **(t, 5)**	*Auxiliary signal*	Acetylene pressure in the interior [kPa].
A_{sP2} **(t, 6)**	*Auxiliary signal*	Acetylene flow in the lateral of the wheel [l/min]
A_{sP2} **(t, 7)**	*Auxiliary signal*	Acetylene pressure in the lateral of the wheel [kPa].
A_{sP2} **(t, 8)**	*Auxiliary signal*	Acetylene flow in the center of the wheel [l/min]
A_{sP2} **(t, 9)**	*Auxiliary signal*	Acetylene pressure in the center of the wheel [kPa].

Furthermore, some images of the copper rod manufacturing process are given to illustrate critical parts of the process. In this regard, **Fig. AI.1 (a)** shows a detail of the tundish where the melted copper is flowing from the holding furnace, and **Fig. AI. 1. (b)** illustrates the output copper bar from the casting wheel.

a)

b)

Fig. AI.1. Detail of both tundish process (a) and the resulting semi-solidified copper bar from the casting wheel.

A2. Electro-Mechanical Test Bench

In order to test the partial contributions of this thesis, an electromechanical actuator based test bench has been used experimental vibrations data is shown in **Fig. AII. 1**. This test bench consists on a kinematic chain composed by a three phase 1492 *W* induction motor, WEG 00236ET3E145T-W22, which speed is controlled by a variable frequency drive-VFD, WEG CFW08, the operating speed is fixed to 60 Hz for all experiments. A 4:1 ratio gearbox, BALDOR GCF4X01AA, is used to couple the drive motor to a DC generator, BALDOR CDP3604. The DC motor is used as a non-controlled mechanical load that comprises around 20% of the nominal torque of the driving motor. The DAS is a proprietary low-cost design based on field programmable gate array technology. The output rotational speed is obtained by using a digital encoder; the motor start-up is controlled by a relay in order to automatize the test run. A 12-bit 4-channel serial-output sampling analog-to-digital converter, ADS7841, is used in the on-board data acquisition system (DAS).

Vibration signal from the perpendicular plane of the motor axis is acquired using a tri-axial accelerometer, LIS3L02AS4, mounted on a board with the signal conditioning and anti-aliasing filtering. Sampling frequency is set to 3 kHz for vibration acquisition. The data retrieved by the DAS is stored in a regular computer (PC).

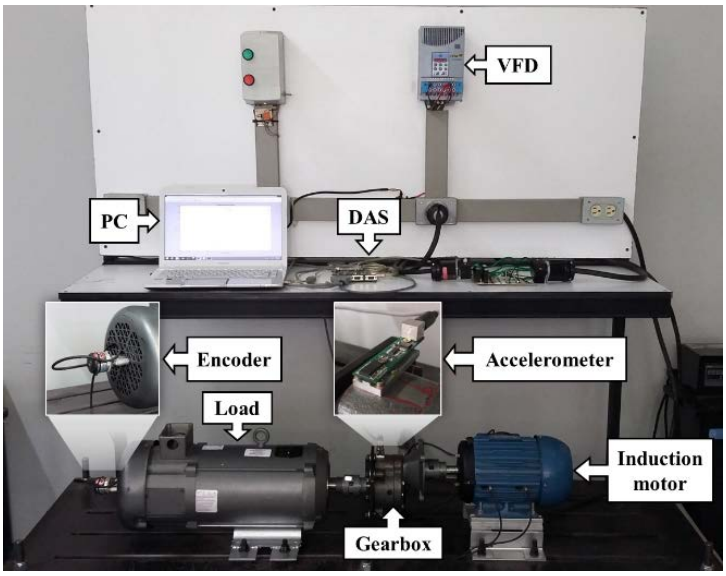

Fig. AII.1. Electromechanical test bench used for experimental validation of the methodologies.

Four scenarios have been considered, that is, the healthy condition, *HC*, a bearing defect condition, *BF*, a half broken rotor bar condition, *HBRB*, and full broken rotor bar, *FBRB*. The detail of the failures is shown in **Fig. AII. 2**. The failure in a 6205-2ZNR bearing has been induced by means of a hole of 1.191 mm ∅ in the outer race; the hole has been produced by a tungsten drill bit. HBRB failure is artificially produced by drilling a 6 mm ∅ hole with a depth of 3 mm that corresponds mostly to the 22% of the section of the rotor bar. Finally, FBRB is produced by a through-hole with a diameter of 6 mm ∅ and a depth of 14 mm, which corresponds to the complete section of the rotor bar.

Two different datasets with the same length have been acquired; the first one is used to train the proposed method, and the second dataset is used to validate the performance of the method. In this regard, periodical acquisitions of 30 seconds containing 90 ksamples are obtained from the accelerometer; the acquisitions are

temporally spaced 30 seconds. The average duration of the experiments for all operating conditions is configured to be 3600 seconds.

A) B) C)

Fig. AII. 2. Detail of the failures produced in the test bench. A) Corresponds to the bearing failure, BF. B) Corresponds to the ½ broken rotor bar, HBRB, and C) to 1 broken rotor bar, FBRB.

The selected duration allows obtaining the complete response of the vibrations in regard with the thermal stability of the electromechanical system. This is done in order to consider the thermal evolution of the vibrations as a part of the system and facilitate the representation of the failure. The proposed approach brings the experimental validation closer to the real behaviour of complex industrial machinery, in which the behaviour of the vibration presents a non-constant dynamic from the starting point since its steady state. As a result, the modelling of the vibration considering the thermal evolution represents an additional challenge to the validation of the proposed method. Furthermore, **Fig. AII. 3.** shows the behaviour of the vibration in healthy condition, HC, in regard with the temperature of the motor from the starting point till the end of the experiment; note that the thermal steady state is reached.

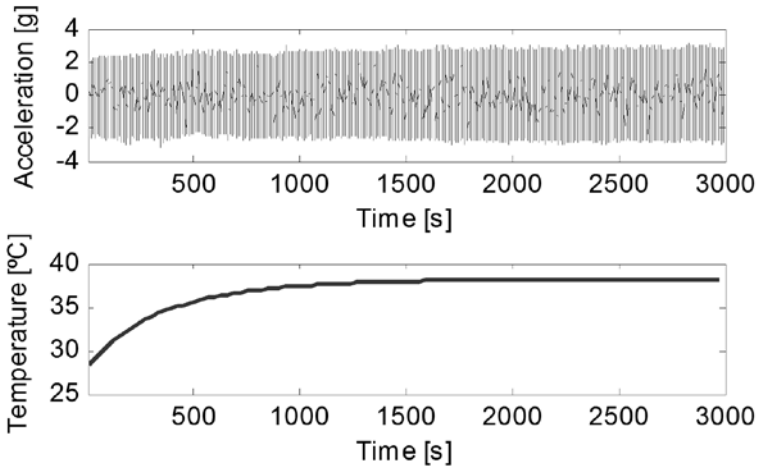

Fig. AII. 3. Evolution of the acceleration signal versus the motor temperature

As can be appreciated in **Fig. AII. 3**, the raw signal of the acceleration gives no important information regarding the condition of the system since it is difficult to detect underlying patterns such as the affectation of the thermal stabilization, since it is modulated in the oscillatory waveform of the vibration. For this reason, the

necessity of calculating a statistical feature representative of the condition of the system is justified after the pre-processing stage.

The vibration signal filtering and the corresponding windowed RMS calculation is carried out by means of FIR filters and considering the rotating speed, the sampling frequency and, also, the theoretical background in regard with mechanical failures. As a result, the cut-off frequencies of the filters in order to isolate the three primarily frequency bands, Z_1, Z_2 and Z_3, are shown in **Table AII. 1**. The attenuation is fixed to get -60 dB in the cutting frequency. It should be noticed that, as it was expected, the order of the filter decreases with the bands since the bandwidth of each filter increases.

Table AII. 1. Design characteristics of the three digital filters used to decompose the vibration signal. Calculations made for a sampling frequency of 3 kHz.

Filter	Cut-off frequency 1	Cut-off frequency 2	Order
$\mathbf{F_{Z1}}$	0 Hz	120 Hz	248
$\mathbf{F_{Z2}}$	120 Hz	350 Hz	136
$\mathbf{F_{Z3}}$	350 Hz	1000 Hz	120

Finally, the RMS feature is calculated for each filter output; the results are 3 RMS signals that summarize the information regarding the vibration of the electromechanical actuator under a concrete failure condition, and are the target signals to be modelled. For this experimental application, the temporal windows is configured to be w_h = 3000 samples, that corresponds to a window of 1 second of actuator operation, and an overlapping factor of ov_{RMS} =25% is selected to smooth the filter response. The resulting signals to be predicted are shown in **Fig. AII. 4** for all considered failures. It can be appreciated how the faulty signal corresponding to a bearing fault is mostly located as it was expected in the second filter, Z_2, from 100 to 500 Hz.

The figure justifies the decomposition of the signal since different spectral behaviour can be observed in each filter output with the presence of failure. This means that each specific failure focus its affectation to a certain frequency band, conclusion that matches with the theoretical considerations exposed previously. The next step in the method deals with the design and the generation of the optimal dedicated ANFIS forecasting model for each filter output.

Finally, to understand the fundamentals of the method, **Fig. AII. 5** shows the characteristic RMS signature of each faulty condition. It demonstrates how the calculated RMS provides the necessary information to distinguish the different failures presented in the system during the thermal stabilization of the actuator. It also justifies the suitability of modelling the RMS of the vibration signal for forecasting and diagnosis purposes. It should be remarked that these RMS are the target signals to be obtained by the models among the defined forecasting horizon p, the final result of the proposed method.

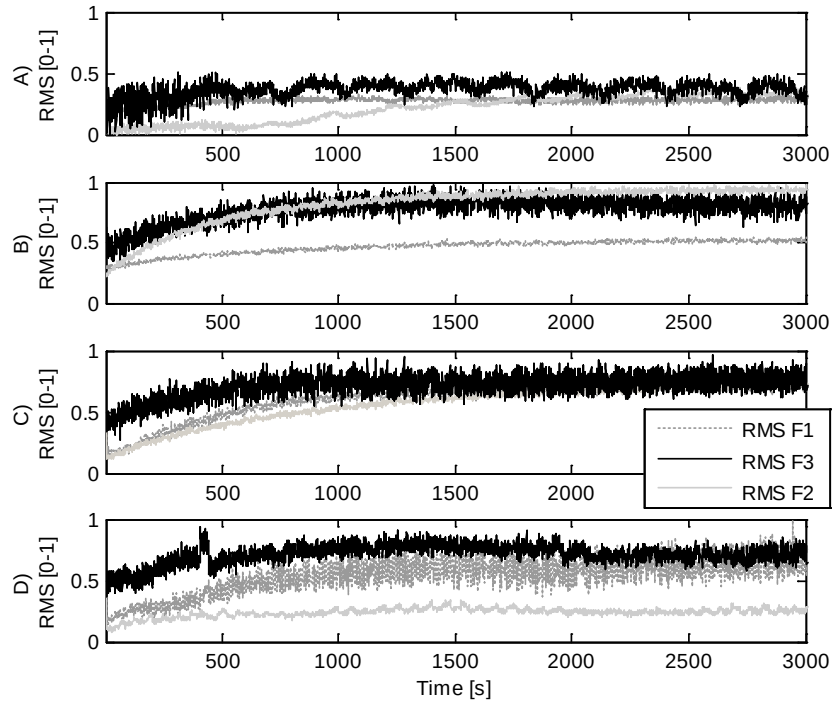

Fig. AII. 4. RMS of each filter output for all operating conditions: A) HS, B) BF, C) FBRB, D) HBRB.

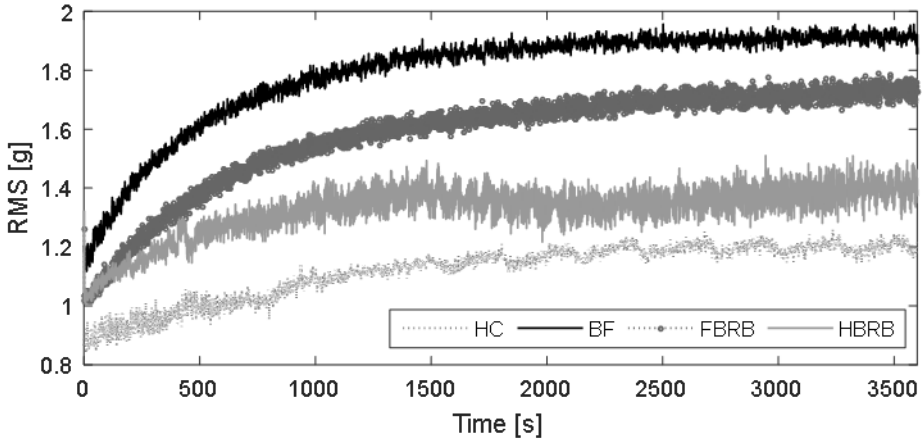

Fig. AII. 5. RMS of the vibration signal during the thermal stabilization period. Note that the different failures can be identified during this period, and as it was expected, the BF case shows the highest vibration.