Soft-Computing and Human-Centric Approaches for Modelling Complex Manufacturing Systems

ALI MUSTAFA BARAKA BEng, MSc(Eng)

The University Of Sheffield.

A thesis submitted to the University of Sheffield for the degree of

Doctor of Philosophy

Department of Automatic Control and Systems Engineering

January 2017

This thesis is dedicated to my mother, my wife, my brothers and my sisters.

To the memory of

my father MUSTAFA BARAKA

(1947 - 2007)

my brother ABDOU
(1973 - 2005)

ABSTRACT

In systems engineering and especially in manufacturing systems, first-principle models have been widely used for systems modelling. However, advanced manufacturing systems are often complex and information intensive rendering conventional modelling approaches via first-principle models inconvenient for use due to their high computation cost and on some instances limited accuracy. The main objective of this thesis is to develop parsimonious transparent, interpretable and computationally efficient soft-computing techniques and human-centric systems to address challenges associated with modelling complex manufacturing systems such as high-nonlinearity, measurement imprecision and sparsity as well as low process repeatability.

A new data-driven modelling framework based on granular computing (GrC), radial basis function neural fuzzy (RBF-NF) system and conflict measure is proposed in order to allow for the quantification of the uncertainty present during the initial structure identification of a RBF neural fuzzy system. Such framework can be easily translated into human language via simple linguistic rules in order to describe the underlying dynamics behaviour of complex industrial processes with good generalisation capability, tolerance to input imprecision and low computational cost.

A new perpetual learning approach for neural-fuzzy systems is proposed. The proposed perpetual learning framework combines more advanced system's features such as the ability to continuously learn from batch data and periodically update its structure to accommodate new data/information without significantly disturbing the previously gained knowledge and, therefore, allowing for the ability to have an open structure while taking into consideration the trade-off between interpretability and accuracy.

To confirm the effectiveness of each of proposed frameworks in this thesis, a rigorous set of simulation results is presented on well-known benchmark functions as well as real industrial case studies. The perpetual learning concept has great potential for successful implementation in systems where lifelong learning is required.

ACKNOWLEDGEMENTS

All praise and thanks are due to the Almighty Allah, the sole Lord of the universe, for giving me health, strength and grace to go step head to this point in my life.

"O My Lord, increase me in knowledge", Quran 20:114 – Surat Taha

It is impossible to acknowledge adequately everyone who has lent a hand and made contributions, great or small, to this thesis. So, I will attempt to give everyone his/her due here, and I sincerely apologise for any omissions.

First and foremost I would like to express my sincere appreciation and immense gratitude to my supervisor, Dr. George Panoutsos for giving me the opportunity to study for a PhD degree under his supervision. I do appreciate his patience, invaluable guidance, constructive advices, work attitude, and unflinching encouragement during the course of my PhD studies. I am also indebted to the Department of Automatic Control and Systems Engineering, The University of Sheffield and The Welding Institute (TWI Ltd.), TWI Technology Centre (Yorkshire), United Kingdom for their sponsorship and aid through the ACSE departmental waiver and TWI Studentship Award to pursue my degree.

I would like to thank Stephen Cater of the TWI Technology Centre (Yorkshire), United Kingdom for many discussions that helped me to understand many areas related to friction stir welding process and invited me over to visit the company, attend many useful courses and participate in many events. I appreciate and really value all his feedback and guidance.

I sincerely appreciate the support and encouragement received from my second supervisor Professor Mahdi Mahfouf during my PhD studies. I would like to express my thanks to my thesis examiners: Dr Maysam F. Abbod and Dr Hua-Liang Wei for their constructive comments and corrections in my thesis.

I would also like to thank all my colleagues at our intelligent systems and human-centric systems research groups for their friendship, particularly, Raymond Muscat, and Drs. Musa Abdulkareem, Olusayo Obajemu, and Adrian Rubio Solis. My thanks are also extended to the staff members of the Department of Automatic Control & Systems Engineering at The University of Sheffield for their unlimited help and guidance throughout the course of my studies.

Special thanks also go to my friends for their whole-hearted prayers and motivation during my PhD period. I am so lucky to have you in my life.

A big thank you to all my faithful friends who are indeed my second family in the UK.

I wish to thank my beloved wife, Egbal for her patience, support, understanding and encouragement during the period of my studies. She motivated me and gave me extra energy to finish this thesis as soon as possible.

Finally, I am deeply indebted to my mother, brothers and sisters who have been a great source of inspiration and cooperation throughout my life. I am thankful to their unwavering and unequivocal support, understanding, and prayers. Thank you all for made it possible for me to pay full attention on my studies and I hope that this achievement is well-worth to their sacrifices during the period that I have been away from home.

Thank you all very much!

ALI MUSTAFA BARAKA

Sheffield, January 2017

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

General Abbreviations

ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
ART1	Adaptive Learning Theory
BEP	Back Error Propagation
CA	Centre Average
CFD	Computational Fluid Dynamics
CM	Conflict Measure
COA	Centre of Area
COG	Centre of Gravity
COS	Centre of Set
DDCI	Data-Driven Computational Intelligence
DFT	Discrete Fourier Transform
FEA	Finite Element Analysis
FFT	Fast Fourier Transform
FIS	Fuzzy Inference System
FOU	Footprint of Uncertainty
FRBS	Fuzzy Rule-Based System
FSSW	Friction Stir Spot Welding
FSW	Friction Stir Welding
FT	Fourier Transform
GA	Genetic Algorithm
GrC	Granular Computing
HCCI	Human-Centric Computational Intelligence
IG	Information Granulation
IT2-FLS	Interval Type-2 Fuzzy Logic System
IT2-RBF-NFS	Interval Type-2 Radial Basis Function Neural Fuzzy System
KDD	Knowledge Discovery in Databases
LMF	Lower Membership Function

LMS	Least Means Square
LVQ	Learning Vector Quantisation
MAE	Mean Absolute Error
MCP	McCulloch-Pitts
MF	Membership Function
MIMO	Multi-Input Multi Output
MISO	Multi-Input Single Output
MLFNN	Multi-Layer Feed Forward Neural Network
MLP	Multi-Layer Perceptron
MM	Maximum Membership
MOM	Middle of Maxima
MLR	Multiple Linear Regression
NFS	Neural Fuzzy System
PI	Performance Index
RBF-NN	Radial Basis Function Neural Network
RCC	Recurrent Cascade Correlation
RMSE	Root Mean Square Error
SD	Standard Deviation
SOM	Self-Organising Map
T1-FS	Type-1 Fuzzy Set
T1-RBF-NFS	Type-1 Radial Basis Function Neural Fuzzy System
T2-FS	Type-2 Fuzzy Set
TnFS	Type-n Fuzzy Set
TR	Type-Reducer
TU	Total Uncertainty
TWI	The Welding Institute
UMF	Upper Membership Function

Important Symbols

β	Learning rate
γ	Momentum factor
δ	Threshold for the rate of the relative performance index
μ_A	Membership function of type-1 of set A
$\mu_{\tilde{A}}(x,u)$	Secondary membership of $\tilde{A}$
A_i^+, A_i^-	The upper and lower boundaries for granule A respectively
$\tilde{A}$	Embedded Type-2 Fuzzy Set
$Cardinality_{(A,B)}$	The number of granule of the resulting granule
$Cardinality_{MAX}$	The total number of granules in the raw input data
$compat_{(A,B)}$	The compatibility between A and B
$Con(A,B)$	Conflict function between the granules of interest A and B
$d_{(A,B)}$	Distance between granule A and granule B
e_H	Shannon entropy
e_k	Training error of the kth data point
F_x	Traverse Force
F_z	Axial (Downward) Force
h_d	Decreasing factor
h_i	Increasing factor
$h(.)$	Activation Function
$iter$	Iteration number index
$I_{n(A,B)}$	Inclusion of granule A into granule B
J_x	Primary membership
$Length_{(A,B)}$	The length of the resulting information granule in multi-dimension
$Length_{MAX}$	The maximum length of an information granule in the raw input data
U	Universe of discourse
m	Basic probability assignment
$m(A)$	Degree of belief of any element that belongs to the set A
$\|\|.\|\|$	Euclidean norm

CHAPTER 1 - INTRODUCTION

1.1. INTRODUCTION

Across a wide variety of fields and disciplines such as in social systems, political systems, medical systems, advanced manufacturing systems, etc., large amounts of data is being routinely generated, collected and stored in large databases [1, 2]. These databases are a possible 'treasure trove' of useful and valuable source of hidden information. However, in real situations when dealing with information, uncertainty often emerges in form of deficiency in information. It appears in almost every system's measurement as a result of incompleteness, impression, fragmentary, vagueness, and contradiction associated with measurements [3]. Due to the aforementioned reasons, information obtained from a system is often not fully reliable [4].

This leads to a need for computational theories and tools to assist humans in quantifying information and then extracting useful knowledge out of uncertain events. From a historical point of view, the study of uncertainty has not been adopted within the scientific community until the late nineteenth century when physicists noticed that Newtonian mechanics did not address the issue of uncertainty [5]. Subsequently, they went on and developed new methods known as statistical mechanics that could capture a form of uncertainty, which could be described as a probability theory or generally referred to as a random uncertainty [5]. The development of the statistical mechanics triggered the need for the developments of methodologies that consider the influence of uncertainty on real-world problems [6]. Such methodologies aimed to improve the robustness of the models. In this sense, a credible solution to the problem is achieved and at the same time the amount of uncertainty is quantified.

1.1.1. TECHNIQUES FOR HANDLING UNCERTAINTY

For more than two centuries, the issue of uncertainty has been a vital research area in order to aid making decisions and develop models that mimic the human cognition when dealing with complex systems. Until the late twentieth century, the probability theory had been the leading theory for processing and quantification of information uncertainty. However, as a consequence of the gradual evolution of the correct processing and quantification of uncertainty, the probability theory was challenged by many authors. First, it was challenged by Max Black in 1937 [7] when he studied a type of uncertainty associated with linguistic information called vagueness. Then Lotfi Zedeh in 1965 [8] challenged not only the probability theory but also the theory (i.e., classical binary logic) upon which probability theory was based. Moreover, Zadeh went on and introduced fuzzy sets theory in his pioneering paper [8]. On one hand, Zadeh extended the notion that instead of the classical set {0, 1}, the mapping of the membership function is defined in the real unit interval [0, 1]. To put it in another way, the notion of fuzzy set denoted by a group of elements where each element in a universal set is characterised by a membership function that takes values in the interval [0, 1]. Fuzzy set theory was introduced to deal with a certain types of uncertainty associated with intuitive information or linguistic information. On the other hand, in 1976 Shafer extended the theory of evidence introduced by Dempster 1967 to produce a complete mathematical theory of evidence which allows to make decisions based on information from more than one source [9]. The concept of fuzzy set theory was extended by Negoita et al. [10] to the possibility theory which is dedicated to deal with incomplete information [11]. In 1982, Pawlak introduced another mathematical theory known as rough sets theory to deal with uncertainty associated with linguistic information [12]. A paper was published by Klir in 1993 in which the author reviewed the advances and shortcomings in the development of theories for measuring of uncertainty [13].

Despite a number of theoretical frameworks have developed to deal with uncertainty [8, 9, 12, 14-16], there has been a debate about which is the best theoretical framework that is adequate to correctly capture, process and then effectively describe uncertain situations. It is highly difficult to answer which is the best theoretical framework for the quantification of uncertainty. Since the main sources of uncertainties can be broadly categorised into: a) random event; b) experimental error; c) uncertainty in judgement; d) lack of evidence and e) lack of certainty in evidence [4], it is obvious that several types of uncertainties exist. Therefore, the selection of which theory can be used to quantify uncertainty is problem-dependent [3].

1.1.2. KNOWLEDGE BASED-SYSTEMS IN SYSTEMS ENGINEERING

Specific to systems engineering particularly in the fields of data analysis, systems modelling, reasoning under uncertain situations, and decision-making, in the last three decades a lot of attention has been paid from some researchers and practitioners in order to capture, process and understand the nature of uncertainty associated with complex systems. Uncertainty and imprecision usually arise as a result of incomplete information and lack of knowledge reflected in system's structure, inputs, and parameters. In systems engineering, uncertainty can be broadly categorised into two types, namely, aleatory uncertainty and epistemic uncertainty [17, 18]. On one hand, the aleatory uncertainty refers to the inherent randomness in nature as a consequence of system natural variability. This type of uncertainty cannot be reduced or eliminated by collecting more information or knowledge. It is sometimes also called irreducible uncertainty, random uncertainty, natural variability, or real-world uncertainty. Probability theory is the most widely used theoretical framework in dealing with this type of uncertainty [17, 18]. On the other hand, the epistemic uncertainty refers to uncertainty that emerges as a lack knowledge (information) of the physical world as well as a lack of ability to measure and model the physical world. Unlike the former type of uncertainty, the epistemic uncertainty can be reduced or eliminated by collecting more knowledge about the problem and appropriate methods. It is sometimes also referred to as knowledge uncertainty, information incompleteness, or subjective uncertainty [17, 18]. Fuzzy sets theory has high potential to handle human ambiguity by modelling epistemic uncertainty through fuzzy sets and their corresponding membership functions.

In this context, the essence of uncertainty-based information relies on the theoretical framework within which uncertainty relating to solve various real-world problems is formalised. In systems engineering, there has been an increasing interest on system features with a special focus on transparency and interpretability. Such features play an important role for a better understanding of complex and poorly understood systems. To put it in another way, the more interpretable the information of a system under investigation, the better its comprehension. Therefore, extracting meaningful information, processing, and then translating it to ‘easy to interpret’ information is a critical step towards developing computational models, especially in case when dealing with complex and highly nonlinear systems. However, in some of advanced and more complex systems, obtaining knowledge about systems from experts is difficult. This is mainly due to complex nature and lack of understanding of the systems themselves. To overcome the aforementioned limitation, during the last two decades data mining has been used to derive knowledge from process data. Data mining also termed knowledge discovery in databases is one of the most pursuits that humans perform almost on a daily basis. It is a power technique for data processing, analysing, and summarising these data in order to extract meaningful knowledge. Part of the knowledge may be known to experts, but the other part is completely new to non-experts. The extracted knowledge aids people understand and recognise some of the intricacies associated with complex systems.

Owing to the complexity and uncertainty associated with many real-world systems, conventional approaches to systems modelling that usually based on first-principle mathematical models or differential equations often offer poor modelling performance. In this sense, fuzzy systems including fuzzy sets theory and fuzzy logic appear to be an effective and a suitable tool for accurate modelling/representing of complex systems and at the same time has the ability to provide system transparency, which is primarily due to their ability to utilise interpretable linguistic rules that are extracted from process data/information. These linguistic rules can be used to facilitate the understanding and analysis of the system under investigation in a qualitative or semi-qualitative manner near to human reasoning [6]. For this reason, in the literature efforts have been devoted for developing fuzzy systems with a good balance between interpretability and complexity. According to [19], in fuzzy systems interpretability and accuracy are two conflicting requirements. That means, a fuzzy system with a low degree of interpretability often more accurate or vice versa.

In a deeper context, several studies have primarily concentrated on the establishment of functional equivalence between a class of fuzzy logic systems and a type of neural networks called radial basis function (RBF) [20, 21]. RBF-NN is a prominent non-linear input/output mapping for complex systems [22]. It has shown a great success in performing many tasks such as exact functions approximation, regularisation, noisy interpolation, and pattern recognition [23]. Despite the RBF-NN is regarded as a black-box model, it is a powerful modelling tool when it is combined with fuzzy logic (FL) by taking advantages of both neural networks (learning capability) and fuzzy logic (transparency and interpretability) [24]. That means, the initial structure of the RBF-NN identification can be achieved similarly to that employed in FL systems [25, 26]. In other words, the RBF-NN parameters which represent the consequent and premise parameters in fuzzy systems are estimated systematically from observational data via a clustering approach and then the parameters are adjusted more precisely via a gradient-based approach to complete the modelling process [26]. In this context, neural-fuzzy systems appear as a modelling paradigm that combines the advantages of FL systems regarding transparency with the advantages of NNs in terms of learning capabilities.

1.1.3. APPLICATIONS OF RULE-BASED SYSTEMS IN MANUFACTURING SYSTEMS

A large and growing body of literature has applied the neural fuzzy systems to modelling of complex manufacturing processes such as in metal processing [26-29]. Such manufacturing processes are known in industry for their high complexity and nonlinearity. Therefore, the interpretability of neural fuzzy systems (NFSs) is a very important property, since it allows the transformation of process data into human knowledge. The extracted knowledge can also be combined with expert knowledge to aid understanding the dynamic behaviour of the system as well as to confirm the system's validity [27]. For this reason, the interest of the researchers and practitioners in interpretability of NFSs has grown, which has led to the appearance of a great quantity of research work with the intention of developing more interpretable rule-based models without losing the overall accuracy of the models.

Studies relating to interpretability improvement in NFSs have been focusing on developing data-driven computational intelligence modelling frameworks that usually include the initial structure identification, partition validation, input selection, rule-base simplification, and constrained parametric optimisation. However, studies on improving system's feature including simplicity, interpretability and transparency as well as on strong link with human reasoning (human-centricity) are quite few [27, 30]. For instance, in [27] Panoutsos and Mahfouf developed a data-driven computational framework based on granular computing theory. The purpose of this work was to provide good modelling performance as well as to maintain the overall transparency during the modelling process. More recently, Solis and Panoutsos presented a systematic NFS based on granular computing, and neutrosophic logic [29, 31]. These studies aimed at reducing uncertainly during the information granulation process in order to enhance the system transparency and interpretability

1.1.4. TYPE-2 FUZZY LOGIC SYSTEMS

Quite recently, researches on type-2 fuzzy logic systems (T2-FLSs) have attracted much attention [32-34]. This is due to their ability to better handle the measurement noise and linguistic uncertainties. Therefore, particular efforts have been devoted to develop neural fuzzy systems by taking advantages of type-2 fuzzy sets with the view of handling the linguistic uncertainty that cannot be handled via the fuzzy sets of type-1. For instance, in [35] the authors published a paper in which they established the functional equivalence between the IT2-FLS and interval type-2 radial basis function neural network.

1.1.5. MODELLING OF FRICTION STIR WELDING

As discussed above, due to the high degree of complexity and nonlinearity associated with many manufacturing systems, the resulting process data are often vague, imprecise, and uncertain. It is often difficult to extract meaningful knowledge out of such data and use this knowledge to create computationally efficient and interpretable models. Among complex manufacturing systems is Friction Stir Welding (FSW), which was invented in 1991 by Wayne Thomas at TWI, United Kingdom [36] as a practical and non-conventional solid-state welding technique. The success of the process and the excellent weld quality produced by the process is evident by the number of applications [37]. Due to thermo-mechanical coupling of the process, the relationship between the process conditions (inputs), internal process variables and the final quality in the friction welding process is nonlinear and complex. For this reason, several attempts have been made by researchers and metallurgists to describe these phenomena. For instance, researchers and metallurgists often conduct a number of experimental trials to study/analyse the influences of process parameters on weld quality and then design the standard/optimal parameters according to the obtained laws. However, this technique is costly and requires a lot of efforts, experience, and time which make it very difficult to establish a precise first-principle mathematical model. In addition to the lack of physical knowledge, complexity and uncertainties in the process conditions, the actual effects of the process input parameters and internal process variables on the final weld quality are hard to quantify. Therefore, this case demonstrates the need for better techniques for modelling the FSW process. On one hand, process models, both mathematical and physical, are efficient tool to analyse and predict the performance of the process. And on the other hand, establishing the relationship between the process conditions (inputs), and internal process variables and then relate them to the final post-weld prosperities is of paramount importance for such process. Therefore, studying these correlations could be beneficial to design a practical, safe, and optimal FSW process as well as process monitoring and control.

Numerous studies have attempted to develop models via analytical and numerical modelling approaches including finite elements analysis (FEA) models and computational fluid dynamics (CFD) in order to identify the link between process input parameters, internal process variables and post-weld properties. A comprehensive review of the latest developments in the field of analytical and numerical modelling of FSW process, microstructures, and mechanical properties can be found in [38]. Despite the intensive researches on FSW modelling via analytical and numerical modelling approaches, the main drawbacks and limitations lie in their high computation cost, and limited accuracy [38]. For these reasons, analytical and numerical modelling approaches are not feasible for real-time use.

In the last decade, a fewer community of researches and metallurgists have embraced the use of data-driven computational intelligence (CI) models via developing soft-computing techniques for modelling of FSW. Mathematical models via data-driven computational models were developed to remedy the aforementioned drawbacks and limitations of analytical and numerical models. These data-driven CI methods include artificial neural networks (ANNs), fuzzy logic systems (FLSs), genetic algorithms (GAs), etc. Such data-driven CI models are widely used in the area of engineering and material science. For instance, in [39] an ANN was used to model and study the effects of the FSW process parameters on the weld mechanical properties for aluminium sheets. In more recent studies [40] [41, 42], systematic data-driven modelling frameworks to model the FSW process for aluminium alloys were developed. The aim of these studies was to develop interpretable, accurate, and robust neural fuzzy systems for modelling the correlations between the input parameters, the internal process variables (namely bending forces), and the final weld quality.

Reviewing the past and on-going researches that focus on the area of data-driven modelling, monitoring, and control of FSW, a very limited number of researches has been reported but all of them are focused of FSW for aluminium, however in the field of steel friction stir welding, no previous research has been conducted on this field.

1.2. PROBLEM STATEMENT

In the field of soft-computing, several theoretical frameworks have been proposed in order to capture, process and then describe various types of uncertain and incomplete or imprecise information. Among them is fuzzy sets theory, which is considered as the main theoretical framework for dealing with uncertainty in a very intuitive and natural manner. A considerable amount of literature has been published on the development of fuzzy systems with a good balance between interpretability and accuracy. These studies include data clustering methodologies to extract fuzzy rules from data while maintaining the interpretability of the FLS during learning via optimisation algorithms and after learning via rule-base simplification approaches [19, 26, 43-45].

More studies can be found in literature with a particular focus on achieving a trade-off between interpretability and accuracy. For instance, in [46] Zhou and Gan presented a taxonomy of interpretability according to the different components of the FLS and they divided the interpretability into two categories namely: low-level interpretability and high-level interpretability. The former can be obtained with regard to semantic criteria of fuzzy sets by optimising the membership functions (MFs). While the later can be obtained when dealing with the consistency, completeness, and coverage on fuzzy rules after the parametric optimisation with regard to the criteria on fuzzy rules.

It appears from the aforementioned literature that numerous studies have been conducted for improving interpretability of FLSs. However, few of them focused on the field of neural fuzzy system, especially in RBF-NN modeling. Therefore, more research is needed for further investigation of the interpretability of neural fuzzy systems. Moreover, in most of the advanced manufacturing systems, human operators are often an integral part of the manufacturing process chain. On the one hand, human operators routinely require performing cognitive tasks such as reasoning and visual information processing. On the other hand, soft-computing techniques and data-mining methods are very efficient at information processing and knowledge extraction. This rapid advances in manufacturing systems and soft-computing techniques culminated in the need for human-centric computational frameworks. Such computational frameworks aid the efficient integration of human knowledge and skills into the manufacturing process chain. This may involve human-machine interaction, human-like information capture, as well as human-centric platforms designed to support 'collaboration' with humans.

The salient motivation behind the use of neural fuzzy system is that it has the ability to deal with incomplete or imprecise and uncertain information in away akin to the human reasoning. It also has the ability to granulate information and the granulated information can be used to focus and facilitate the analysis and interpretability of complex systems on aspects of interest to the user. These intriguing traits are preferable when the system is too complex to be analysed via first-principle models and can be merely interpreted qualitatively. However, traditional approaches in designing the neural fuzzy systems are unfortunately over-dependent on expert-knowledge and usually entail tedious manual interventions. Furthermore, repeating the whole modelling process each time a new batch of data becomes available is often a laborious and non-automated process as well as time-consuming. Consequently, there is no guarantee that the new process model will keep a good performance comparable to the original model. Research to date has tended to focus on adaptive learning methods for type-1 FLSs and type-2 FLSs with application to time-varying data [47-49] [50-53]. To our knowledge, no research has been carried out on batch incremental learning in the field of type-2 FLSs. There is a need to include more advanced system's feature. Such features include ability to learn from an initial batch of data (with the help of an appropriate training algorithm) and periodically adapt to new process data when these are available. An additional need is to include the capability to interact with a changing environment in a perpetual fashion and also to have an open structure; this entails to dynamically expand the system's structure to fit in/accommodate new data/information – without significantly disturbing the initial model structure.

The rest of the thesis work will be focused on the use of various concepts developed in soft-computing and human-centric intelligent systems including information granulation, fuzzy sets theory, fuzzy logic systems, artificial neural networks and incremental learning for modelling purposes.

From process perspective, the focus of this thesis work will be on developing parsimonious, interpretable and computationally efficient real-time process models that have the capability of taking advantage of expert-knowledge to handle imprecision, inconsistency, incompleteness, sparsity, quantity and complexity of the associated process data. On the one hand, process models that mimic the ability of humans in using simple linguistic interpretable rules extracted from raw data in order to describe complex systems. On the other hand, process models that can be used for real-time prediction and on-line monitoring - through process modelling and optimisation - to aid the process operator in making effective decisions as well as to guide the process into optimum operation conditions by take into account a set of desirable objectives.

Such efficient computational frameworks are tested against well-known benchmark data sets as well as a real industrial case study of steel friction stir welding for the first time using real manufacturing data.

1.3. RESEARCH AIM AND OBJECTIVES

The central aim of this research work is to develop parsimonious, transparent, interpretable and computationally efficient soft-computing techniques and human-centric systems for complex manufacturing processes with particular application in modelling of steel friction stir welding.

In order to achieve its aim, this research work accomplishes the following objectives:

- The goal of the first study is to quantify the uncertainty during the initial structure identification of the RBF neural fuzzy system. According to [46], the interpretability in data-driven fuzzy systems can be categorised into two main components namely: a) low-level interpretability and b) high-level interpretability. These two components can be used to facilitate the analysis and gain a deep insight into the factors involved into developing data-driven fuzzy systems. This study takes benefit of the mathematical equivalence between the RBF-NN and a class of fuzzy inference systems called type-1 FLS under certain conditions to investigate the relationship between uncertainty during the initial structure identification phase and the interpretability of the RBF-NF system as well as the overall system accuracy.
- The second study aims to systematically develop an interval type-2 RBF neural fuzzy (IT2-RBF-NF) system. The IT2-RBF-NF system is mathematically equivalent to an interval type-2 fuzzy logic system (IT2-FLS) under certain conditions and optimised via an adaptive back-error propagation (adaptive-BEP) approach. On the one hand, the advantages of principles of iterative human-like information capture in granular computing (GrC) and the additional degree of freedom from the footprint of uncertainty (FOU) in type-2 fuzzy sets (T2-FSs) are used to extract meaningful information and handle the linguistic uncertainties associated with meaning of words contained in the rule-base of the IT2-RBF-NF system. And on the other hand, the proposed system is used to develop a new generalised model-based real-time process monitoring framework in steel Friction Stir Welding. The proposed real-time process monitoring framework relies on frequency-based information is capable of discriminating the quality of welds produced by friction stir welding

process and providing real-time feedback to the process operator(s) in linguistic format (natural language – rule-based system) on the performance of the process. The intention of this study is also to execute a number of simulation examples in order to confirm the appropriateness and efficiency of the proposed IT2-RBF-NF system against different data-driven models including multiple linear regression (MLR) and multilayer perceptron neural network (MLP-NN) models as well as type-1 radial basis function neural fuzzy (T1-RBF-NF) system.

- Finally, a new perpetual (incremental) learning framework is developed. The objective of this study is to include the ability of the IT2-RBF-NF system to continuously learn from batch data and periodically update its structure to accommodate new data/information without significantly disturbing the previously gained knowledge. Therefore, the ability of the system to have an open structure with a particular focus on achieving the trade-off between interpretability and accuracy to make it feasible for lifelong learning use. The intention of this study is also to execute a number of simulation examples in order to evaluate the performance of the proposed perpetual learning framework on well-characterised benchmark functions as well as a real industrial case study.

1.4. ACHIEVEMENTS AND CONTRIBUTIONS

1.4.1. ACHIEVEMENTS

The main achievements of this thesis can be listed under the following headings:

- In Chapter 3, a data-driven computational intelligence modelling framework based on the well-known modelling framework of adaptive neuro-fuzzy inference systems (ANFIS) and subtractive clustering optimised via a gradient decent approach is provided as a benchmark modelling paradigm for preliminary modelling analysis. The proposed modelling framework is applied to a real industrial data related to the prediction of spindle peak torque in steel friction stir welding. The data sets were obtained from The Welding Institute (TWI Ltd), Technology Centre (Yorkshire), United Kingdom.

- Chapter 4 presents a new conflict measure during the iterative human-like information capture in granular computing (GrC) that was initially proposed in [27] in order to estimate/evaluate the uncertainty arises as a result of conflict during the iterative data granulation process. The proposed conflict measure is calculated via Shannon entropy theory to extract information related to the data uncertainty while carrying out the granulation process and guide the iterative granulation process into merging information granules with low uncertainty. Thus improving the interpretability of the information granules at the low-level component of interpretability (in form of distinguishability between fuzzy sets). The resulting information granules are used to construct the initial parameters of a Radial Basis Function (RBF) based neural-fuzzy model optimised via the adaptive back-error propagation (BEP) algorithm. The effectiveness of the proposed GrC based RBF-NF system is verified by using a well-known benchmark data set and applied to modelling of steel friction stir welding. Finally, the predictive performance and interpretability of the proposed modelling framework with and without conflict measure is compared with the preliminary results presented in Chapter 3.
- In Chapter 5, a twofold contribution is presented; firstly, it is introduced an interval type-2 RBF neural fuzzy (IT2-RBF-NF) system that is mathematically equivalent to an interval type-2 fuzzy logic system (IT2-FLS) under certain conditions and optimised via an adaptive back-error propagation (adaptive-BEP) approach. On the one hand, the advantages of principles of iterative human-like information capture in granular computing (GrC) and the additional degree of freedom from the footprint of uncertainty (FOU) in type-2 fuzzy sets (T2-FSs) are used to extract meaningful information and handle the linguistic uncertainties associated with meaning of words and linguistic propositions contained in the rule-base. Secondly, a new generalised model-based real-time process monitoring framework in steel Friction Stir Welding is developed. The

proposed real-time process monitoring framework relies on frequency-based information is capable of discriminating the quality of welds produced by friction stir welding process and providing real-time feedback to the process operator (s) in linguistic format (natural language – rule-based system) on the performance of the process. Finally, a number of simulation examples are carried out in order to confirm the appropriateness and efficiency of the proposed IT2-RBF-NF system against multiple linear regression (MLR) and multilayer perceptron neural network (MLP-NN) models as well as a type-1 radial basis function neural fuzzy (T1-RBF-NF) system, which is similar to the one developed in Chapter 4.

- In Chapter 6, a new perpetual learning framework that is based on the IT2-RBF-NF system is proposed. A number of simulation examples are carried out in order to evaluate the performance of the proposed perpetual learning framework on well-known benchmark functions as well as a real industrial case study.

1.4.2. PUBLICATIONS

Conferences and Journals

- **Ali Baraka**, George Panoutsos, and Stephen Cater, 'Perpetual Learning Framework based on Type-2 Fuzzy Logic System for a Complex Manufacturing Process', 17th IFAC Symposium on Control, Optimization and Automation in Mining, Mineral and Metal Processing, August 31-September 2, 2016, Vienna, Austria.
- **Ali Baraka**, George Panoutsos, and Stephen Cater, 'Real-time quality monitoring for friction stir welding AA2219-T845 aluminium aerospace alloy via model-based spectral analysis', 11th International Symposium on Friction Stir Welding, 17-19 May 2016, Cambridge, UK.
- **Ali Baraka**, George Panoutsos, and Stephen Cater, 'A real-time quality monitoring framework for steel friction stir welding using computational intelligence', Journal of Manufacturing Procesess, 20, pp.137-148, 2015.

- **Ali Baraka**, George Panoutsos, Mahdi Mahfouf and Stephen Cater, 'A Shannon Entropy-Based Conflict Measure For Enhancing Granular Computing-Based Information Processing', the 10th IEEE International Conference on Granular Computing (GrC), 22nd -24th October 2014, Noboribetsu, Hokkaido, Japan.

- **Ali Baraka**, Adriana A. Gonzalez-Rodriguez, George Panoutsos, Kathryn Beamish and Stephen Cater, 'Manufacturing Informatics and Human-in-the-loop: A case of study on Friction Stir Welding', the 3rd EPSRC Manufacturing the future conference, 23rd-24th September 2014, Glasgow, UK.

Seminars:

- **A. Baraka** and G. Panoutsos M. Mahfouf, and Stephen Cater, 'Steel FSW: Data mining, modelling, and support decision systems', TWI Colloquium, 05 June 2014, Rotherham, UK.
- **A.M. Baraka**, A. Rubio Solis, A.A. Gonzalez-Rodriguez, J.C. De Alejandro and G. Panoutsos, 'Human-Centric Approaches for Modelling Complex Processes', University of Sheffield Engineering Symposium (USES), 20 May 2013, Sheffield, UK.

Journal Papers in Preparation:

- **Ali Baraka** and George Panoutsos, 'Perpetual Learning for Type-2 Neural-Fuzzy Systems', to be submitted to a soft-computing journal.
- Autonomous Rule-based Systems and their applications to advanced manufacturing systems, to be submitted to soft-computing and materials science journals respectively based on the material from Chapter 6.

1.5. THESIS OUTLINE

The thesis is structured into 7 chapters and one appendix. The next paragraphs will describe a chapter-by-chapter overview of the contents of this thesis.

Chapter 1, titled ‘Introduction’, provides background information and the basic notions necessary to understand this research work and examines the important contemporary challenges of advanced manufacturing systems. The concepts of soft-computing and human-centric intelligent systems are briefly introduced. The emphasis in this chapter has been on drawing the attention of the reader to the most important trends and to highlight the modelling of complex manufacturing systems via soft-computing and human-centric approaches.

Chapter 2, titled ‘Modelling of Complex Systems via Soft-Computing and Human-Centric Approaches’, provides a review about the existing techniques found in soft-computing and human-centric computational intelligence systems a particular focus will be put on fuzzy sets, fuzzy logic systems, fuzzy and neural fuzzy modelling, artificial neural networks and information granulation. As far as fuzzy logic systems are concerned in the development of this research, a general survey on different types of information uncertainty is provided. This is mostly due to the type of topics related to the modelling of uncertainty considered in this research work.

Chapter 3, titled ‘Friction Stir Welding and Process Modelling’, provides a literature review about the current developments in the field of Friction Stir Welding process. Fundamental knowledge and basic understanding of the process are presented. These include the principal of operation, development and reasons for using this welding technique with relevance to the advantage of this process for manufacturing of metal joining. This chapter also covers a comprehensive literature review on the main areas related to the analytical and numerical modelling approaches, and data-driven modelling approaches as well as monitoring and control of the FSW process. The chapter concludes by demonstrating the possibility of using data-driven modelling approaches for modelling of steel FSW for real-time applications. Furthermore, preliminary modelling results for the Friction Stir Welding (FSW) prediction of internal process variables, namely the spindle peak torque by using the well-known modelling framework of adaptive neuro-fuzzy interference systems (ANFIS) and subtractive clustering are provided.

Chapter 4, titled 'Interpretability Measures in RBF-NF Systems using Iterative Granular Computing', presents a new conflict measure during the iterative human-like information capture in granular computing (GrC) in order to estimate/evaluate the uncertainty emerges as a result of conflict during the iterative data granulation process. On one hand, the proposed conflict measure is calculated via Shannon entropy theory to extract information related to the data uncertainty while carrying out the granulation process. Such information is used to guide the iterative information granulation process into condensing (merging) the granules (data) with low conflict, and therefore producing better quality information granules. On the other hand, the resulting information granules are employed to construct the initial parameters of a RBF-NF system optimised via the adaptive back-error propagation (BEP) algorithm and applied to modelling of steel friction stir welding. Finally, a comparative study is carried out to compare the performance of the proposed granular computing based RBF-NF modelling framework with and without conflict measure against the preliminary results presented in Chapter 3.

Chapter 5, titled 'An Interval Type-2 Neural Fuzzy System: IT2-RBF-NFS', concentrates on the development of an interval type-2 radial basis function neural fuzzy system (IT2-RBF-NFS) that is mathematically equivalent to interval type-2 fuzzy logic systems (IT2-FLSs). The main focus of this chapter is twofold, on the one hand, a new modelling framework is presented by taking advantages of principles of iterative human-like information capture in granular computing (GrC) and the extra degree of freedom from the footprint of uncertainty (FOU) in type-2 fuzzy sets to take into account for the linguistic uncertainties associated with meaning of words and linguistic propositions contained in the rule-base. An adaptive back-error propagation (adaptive-BEP) approach is used to optimise the initial structure of the proposed modelling framework. And on the other hand, a new generalised systematic human-centric model-based real-time process monitoring framework in steel Friction Stir Welding is developed based on the presented IT2-RBF-NFS. The proposed real-time process monitoring framework relies on discrete frequency-based analysis of key internal process variables (namely axial (Fz) and traverse (Fx) forces) that is capable of providing real-time feedback to the process operator (s) in linguistic format (natural language – rule-based system) on the performance of the process. The proposed model-based monitoring framework is also used to forecast in real-time (during welding) quantitative markers of weld quality extracted from the welding tool feedback forces. The chapter concludes by comparing the performance of the proposed model-based approach against a baseline (multiple linear regression), multilayer perceptron (MLP) as well as RBF-NF models.

Chapter 6, titled 'A New Perpetual Learning Framework for IT2-RBF-NFS', concentrates on the development of a new perpetual learning framework based on the iterative human-like information capture in Granular Computing (GrC) and IT2-RBF-NF system presented in Chapter 5. The proposed framework has the ability to evolve through incremental and structural parametric learning. Such framework relies on the creation of new rules, which are added to the original model to update its structure. The updated model is then optimised during the incremental process. An iterative rule pruning strategy is also included in the structure in order to remove any inconsequential rules as a result of the incremental update routine. The strength of such as a framework is that this framework uses rule growing/pruning strategy, which makes the proposed framework feasible for lifelong learning mode. The chapter concludes by evaluating the performance of the proposed perpetual framework on complex benchmarking functions in system identification as well as a real-industrial case of steel FSW for the modelling of spindle peak torque.

Chapter 7, titled 'Conclusions and Future Work', includes the main findings of this research work. It also presents recommendations for future research.

CHAPTER 2 - MODELLING OF COMPLEX SYSTEMS VIA SOFT-COMPUTING AND HUMAN-CENTRIC APPROACHES

2.1. INTRODUCTION

This chapter includes a literature review on the existing techniques found in soft-computing and human-centric computational intelligence systems. A particular focus will be put on the fields of fuzzy logic systems, fuzzy and neural fuzzy modelling, artificial neural networks and information granulation.

2.2. SOFT-COMPUTING AND HUMAN-CENTRIC COMPUTATIONAL INTELLIGENCE SYSTEMS

The term ‘soft-computing’, sometimes referred to as computational intelligence (CI) was coined by Lotfi Zadeh, the inventor of fuzzy sets theory [54]. Soft-computing is the integration of soft-computing techniques and tools such as fuzzy logic (FL) [55], neural networks (NNs) [56], evolutionary genetic algorithms (GAs) [57], and probabilistic reasoning that are designed to model or deal with and enable solutions to real-world problems which are not modelled or too difficult to model mathematically via conventional (hard) computing techniques such as classical sets theory. Moreover, this fusion of different methodologies aims to exploit the human tolerance for imprecision, approximate reasoning, partial truth and uncertainty in order to achieve tractable, robust and low-cost solutions [54]. Each constituents of soft-computing has its inherited advantages and disadvantages, for example, FL is based on knowledge-driven reasoning and mainly concerned with approximate reasoning and imprecision; and NNs are data-driven learning and curve-fitting approaches [58]. In this regard, the integration of different techniques allows for the combination of domain knowledge and empirical data in order to develop a flexible computing paradigm and solve imprecisely and precisely formulated complex problems.

The rapid advances in computer science and information communication technologies culminated in the development of truly widespread human-centric computing frameworks. Such computing frameworks would require platforms that have the ability to support different operators or users working in varying environments. These computing platforms would be required to engage and play more active roles in performing a wide range of tasks such as intelligent data analysis, data management and information sharing, human/machine interaction, decision-making, etc. It would be also required to develop seamless and transparent computational systems based on the concept of human-centredness that are able to adjust by humans by being moral natural, intuitive to use and consistently integrated within environment [59]. A human-centric system approach involves performing research on understanding the interaction, collaboration, and coordination between users and machines, and machines and machines. Some of the recent existing trends in the development of human-centric systems include, but not limited to systems modelling, intelligent data summarisation and analysis, ubiquitous computing, intelligent interface, etc. [60].

In systems modelling, data-driven computational intelligence (DDCI) models such as fuzzy rule-based systems (FRBSs) [61], neural-fuzzy systems (NFS) [26], artificial neural networks (ANNs) [56] as well as evolutionary genetic algorithms (GAs) [57] have shown great successes in solving various real-world problems such as those associated with industrial, academic, and medical applications. However, recently, there has been a growing demand on system features with a special focus on simplicity, interpretability and transparency as well as on strong link with human reasoning (human-centricity) [27, 30]. Fuzzy systems are powerful data-driven modelling techniques when combined with optimisation techniques such as back-error-propagation and multi-objective optimisation [41]. Fuzzy rule-based systems (FRBS) use natural descriptive language to describe complex processes in a very transparent way and the modelling outcome is more interpretable (characteristic of fuzzy logic). While ANNs are regarded as black-box modelling paradigms because they are geared towards processing of raw numeric data. As a consequence, ANNs are very difficult to be analysed, interpreted, and understood.

Over the past few decades, fuzzy systems have received significant interest from various fields and many studies have clearly demonstrated that fuzzy systems are considered to be a very popular modelling technique in soft-computing [60]. This is due to their use of concept of human reasoning approaches to mimic human tolerance for incompleteness, uncertainty, imprecision and fuzziness in the processes of making effective decision. The concept of human reasoning in fuzzy systems adds the strength of being close to human knowledge and interpretation [62, 63]. One of the advantages of fuzzy systems theory is to be used as universal approximates to approximate mathematical functions or to approximate real systems where analytic functions or numerical relations are not available to govern the relation between the input(s) and output(s). Thus, fuzzy systems have high potential to model and comprehend complex systems such as social systems, political systems, medical systems, complex manufacturing processes, etc. Such systems are often characterised by information incompleteness, imprecision, vagueness or unreliability. Owning to the high degree of vagueness and imprecision in many real-world problems, it is often difficult to model/represent them via accurate first-principle mathematical models. Fuzzy systems including fuzzy sets theory and fuzzy logic provide an effective method to represent uncertainties and deal with conditions which are inherently ill-defined and imprecise [6].

Since the concept of granular of computing (GrC) was first introduced by Zadeh in [64, 65] and Lin in [66] as a conceptual computing paradigm that concerns with the processing of complex information entities – in form of information granules. In essence, information granules arise in the process of data abstraction and knowledge extraction from information [30]. In the past two decades, several mathematical frameworks and tools geared for processing complex information granules have been proposed in conjunction with other methodologies such as interval analysis, rough sets and fuzzy sets [30, 67]. Subsequently, the role of the human-like information processing of granular computing has become profoundly visible in the development of transparent human-centric systems [60]. The integration of the human-like information processing paradigm of GrC (human-like feature), and fuzzy logic (interpretability feature) incorporated also with neural networks (learning feature) [27] leads to improved system transparency and accuracy. Consequently, the general data-driven computational intelligence (DDCI) paradigm provides interpretability, accuracy, and user-friendliness as well as facilitates the activity and visible role of the system's developer (transparency) [68]. The very essence of developing human-centric systems focuses on how to develop systems that support interaction, interpretation/understanding of numeric raw data/ information with descriptive, linguistic and qualitative input arriving from the user(s). In addition, the user-based information processing ability has to be taken into consideration when developing human-centric computational intelligence (HCCI) models.

2.3. FUZZY LOGIC AND FUZZY SYSTEMS

As pointed out in the previous section, due to high degree of vagueness and complexity associated with real-world problems, it becomes more difficult to build precise mathematical models to model/represent the underlying physical process of a system [6]. Fuzzy logic systems have high potential to model/represent complex systems and provide system transparency, which is primarily due to their ability to utilise simple linguistic interpretable rules in the form of IF-THEN statements. These linguistic rules are extracted from process data and can be employed to facilitate understanding and analysis of the system under investigation in a qualitative or semi-qualitative manner near to human reasoning [6]. It has been proved that any fuzzy system can be considered as a nonlinear universal approximator [69, 70].

2.3.1. FUZZY SETS

Fuzzy sets are types of sets (classes) that allow their elements to have different degrees of membership. In classical bivalent set theory, any element either belongs to that set (class) or does not belong to that set (class). For instance, a given set A, the membership function has a value of $\mu_A(x)$ to each $x \in U$ i.e.

$$\mu_A(x) = \begin{cases} 1 & if \quad x \in U \\ 0 & if \quad x \notin U \end{cases} \tag{2-1}$$

or

$\mu_A: U \rightarrow \{0, 1\}$

where U is the universe of discourse

In [8], Zadeh extended the notion that instead of the classical set {0,1}, the mapping of the membership function is defined in the real unit interval [0, 1]. A fuzzy set is defined as “A fuzzy set in the universe of discourse U is characterised by a membership function $\mu_A(x)$ that takes values in the interval [0, 1]"[71] given by the relationship:

$$A = \left\{ \frac{x, \mu_A(x)}{x} \in U \right\} \qquad 2\text{-}2$$

$0 \leq \mu_A(x) \leq 1$

where $\mu_A(x)$ is the membership function which maps each element x of the universal set U to a membership degree between 0 and 1 [5].

Generally, different shapes/types of membership functions (MF) can be employed to map a particular fuzzy logic system, the selection of which is an application dependant. The most common membership function shapes are: Gaussian functions, trapezoidal-shape functions, bell-shaped functions, and S-shaped function as shown in Fig. 2.1. The mathematical expressions for the membership functions can be expressed as [6, 71]:

- **Gaussian MF**

$$f(x; c, \sigma) = e^{-\left[\frac{x-c}{\sigma}\right]^2} \qquad 2\text{-}3$$

The graphical representation of the Gaussian membership function is shown in Fig. 2.1 (a).

- **Triangular MF**

$$f(x; a, b, c) = \begin{cases} 0, & x \leq a \\ \dfrac{x-a}{b-a}, & a \leq x \leq b \\ \dfrac{c-x}{c-b}, & b \leq x \leq c \\ 0, & x \geq c \end{cases} \qquad 2\text{-}4$$

The graphical representation of the triangular membership function is shown in Fig. 2.1 (b).

- **Trapezoidal MF**

$$f(x;a,b,c,d) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \dfrac{d-x}{d-c}, & c \le x \le d \\ 0, & x \ge d \end{cases} \qquad 2\text{-}5$$

The graphical representation of the trapezoidal membership function is shown in Fig. 2.1 (c).

- **Generalised Bell MF**

$$f(x;\mathrm{a},\mathrm{b},c) = \frac{1}{1+\left|\dfrac{x-c}{a}\right|^{2b}} \qquad 2\text{-}6$$

The graphical representation of the generalised bell membership function is shown in Fig. 2.1 (d).

- **Sigmoid (S-shaped) MF**

$$f(x;a,b) = \begin{cases} 0, & x \le a \\ 2\left(\dfrac{x-a}{b-a}\right)^2, & a \le x \le \dfrac{a+b}{2} \\ 1-2\left(\dfrac{x-b}{b-a}\right)^2, & \dfrac{a+b}{2} \le x \le b \\ 1, & x \ge b \end{cases} \qquad 2\text{-}7$$

The graphical representation of the sigmoid (S-shaped) membership function is shown in Fig. 2.1 (e).

- **Z-shaped MF**

$$f(x;a,b) = \begin{cases} 1, & x \le a \\ 1-2\left(\dfrac{x-b}{b-a}\right)^2, & a \le x \le \dfrac{a+b}{2} \\ 2\left(\dfrac{x-a}{b-a}\right)^2, & \dfrac{a+b}{2} \le x \le b \\ 0, & x \ge b \end{cases} \qquad 2\text{-}8$$

The graphical representation of the Z-shaped membership function is shown in Fig. 2.1 (f).

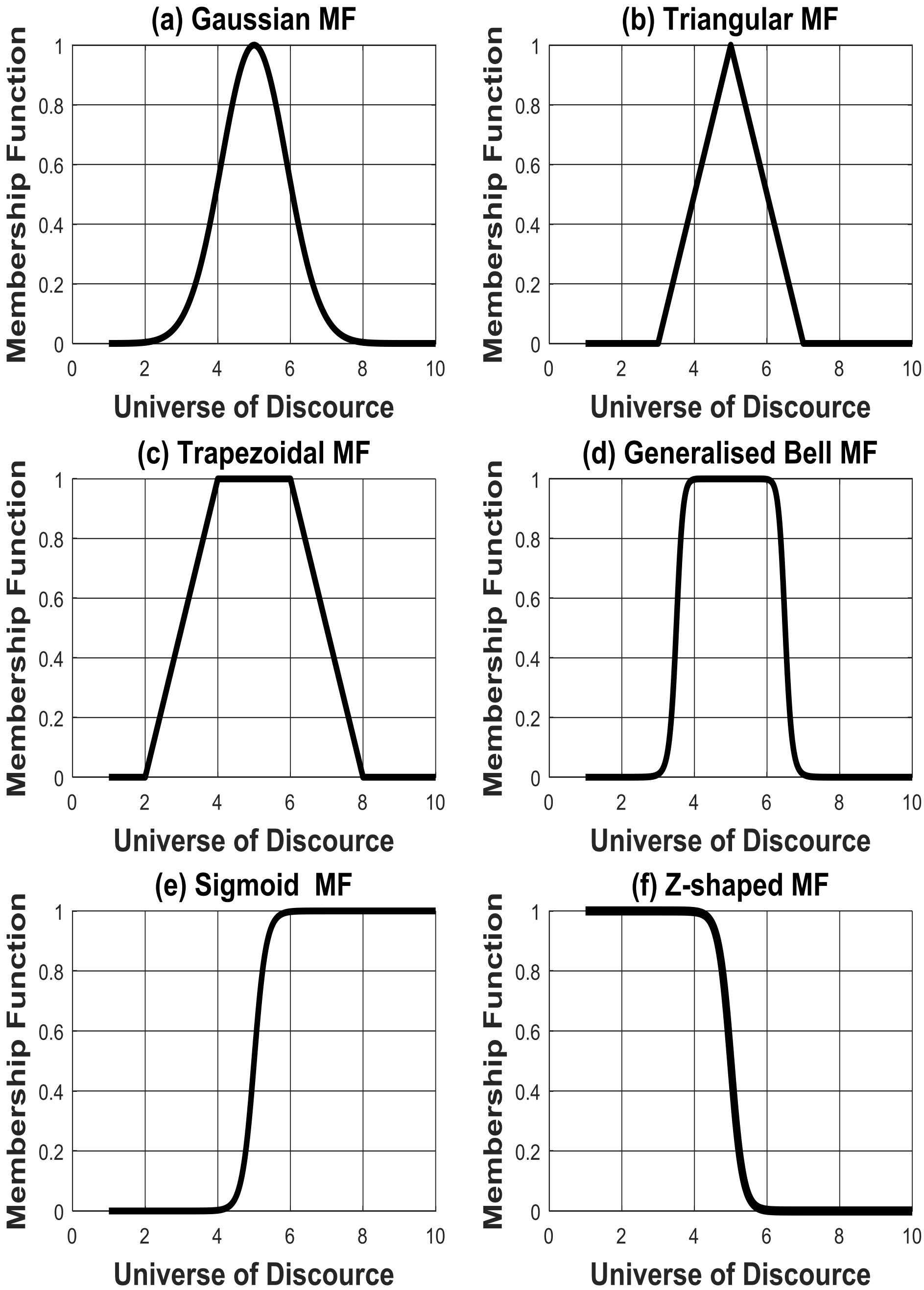


Figure 2.1. Shapes of membership functions.

2.3.2. FUZZY SYSTEMS

A fuzzy system as shown in Fig. 2.2 basically composes of four components: fuzzification, fuzzy inference engine, fuzzy rule base and defuzzification [71].

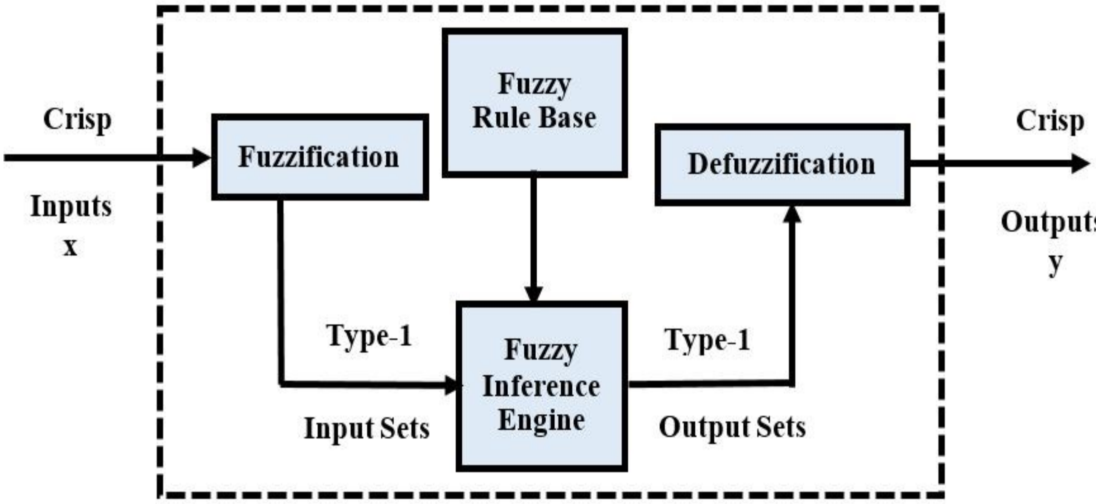


Figure 2.2. T1-FLS block diagram.

The fuzzification is a component that transforms the real value of the input variable x to a fuzzy input. It simply calculates a set of membership functions for the crisp value x. There are three types of fuzzification methods are frequently used, which are Gaussian, singleton, and triangular fuzzification method [71]. The defuzzification, in contrast to fuzziffication, is a way of mapping a fuzzy quantity from the fuzzy inference engine into a quantifiable value to be used in the real-world. There are numerous defuzzification methods, but the most prevalent used are centre of gravity (COG), centre of area (COA), maximum membership (MM), middle of maxima (MOM), and centre average (CA) defuzzifiers [6, 71].

The fuzzy rule-base (knowledge base) is the heart of a fuzzy system, which consists of a set of fuzzy rules in form of IF-THEN rules. The most popular knowledge bases are: Mamdani-type [72] and Sugeno-type (TSK) [73]. A Mamdani-type fuzzy logic system rule has the following form:

$$Rule_r: \boldsymbol{IF}\ x_1 \boldsymbol{is}\ A_1^r \boldsymbol{and} \dots \dots \boldsymbol{and}\ x_d \boldsymbol{is}\ A_n^r\ \boldsymbol{THEN}\ y^r\ is\ B^r \qquad 2\text{-}9$$

where $r = 1, 2, \dots, M$ and M is the number of fuzzy rules in the rule-base; A_n^r and B^r are fuzzy sets in the input space $U_d \subset R$ and $V \subset R$ respectively and $x_1, \dots, x_d \in U_d$ and $y^r \in V$ are the fuzzy system's input and output respectively. A Sugeno fuzzy model is different from Mamdani-type in the way that the consequent parts are deterministic. A Sugeno-type can be expressed as a set of linguistic IF…THEN rules as follows:

$$Rule_r\text{: } \boldsymbol{IF}\ x_1 \boldsymbol{is}\ A_1^r \boldsymbol{an} \dots \boldsymbol{and}\ x_d \boldsymbol{is}\ A_n^r\ \boldsymbol{THEN}\ y^r = f_r(x) = \sum_{r=1}^{n} a_r x_r + a_{r0} \qquad \text{2-10}$$

where the premise part is the same as Mamdani-type, however, the consequent part is deterministic. $f_r(\mathrm{x})$ can be a linear or quadratic function and a_r and a_{r0} are the linear parameters of the consequent part of the Sugeno model. The defuzzifier is not used in this case.

The fuzzy inference engine is the process in which the fuzzy logic operations are used to combine the fuzzy input sets from the input space X with the IF-THEN rules from the fuzzy rule-base to form the fuzzy output sets in Y. The fuzzy logic operations include intersection, union, complement, containment, and Cartesian product. More details on fuzzy sets operations can be found elsewhere [71].

2.3.3. NEURO-FUZZY SYSTEMS

The construction of classical fuzzy system (mostly Mamdani-type) involves the process of finding appropriate MFs and fuzzy rules. This process is based on expert-knowledge and experience and it is not always easy and often a tiring process of trial and error, which will restrict the applications of the classical fuzzy systems. Particularly when the expert knowledge is lacking or not available, thus traditional knowledge-based fuzzy model cannot be generated. Therefore, the concept of applying learning mechanisms to classical fuzzy systems was introduced in the early 1990s. The neuro-fuzzy system was first proposed by Jang [74], where an adaptive-neuro-fuzzy inference system (ANFIS) framework was introduced. ANFIS is a hybrid combination of ANNs and fuzzy inference systems (FIS) [75]. Since it integrates both artificial neural networks and fuzzy logic systems, it has potential to capture the learning capability of the neural networks and transparency and interpretability of fuzzy logic system. The structure of ANFIS is flexible, easy to modify, and learn its parameters adaptively. The ANFIS can construct a mapping between input-output based on both expert knowledge and numerical data. Currently, there are several neuro-fuzzy systems exist in the literature. Most notable are fuzzy adaptive learning control proposed by Lin and Lee [76], NEuro-Fuzzy CONtrol (NEFCON) developed by Nauck et al. [77] and other variants from these development [78-80].

Classical fuzzy systems can also be combined with genetic algorithms (GAs) to enhance the learning ability of the fuzzy systems, the scope of fuzzy systems applications can be expanded. The learning process in genetic fuzzy systems is regarded as an optimisation process of the rule-base of fuzzy systems. There are three genetic algorithms available for optimising the rule-base: the Michigan approach [81], the Pittsburgh approach [82], and the iterative rule learning approach [83].

As the field of neuro-fuzzy systems matures and grows in visibility, there is an increasing concern about the development of more sophisticated neuro-fuzzy systems to solve different kinds of problems in various applications [84-87]. Some of the more recent existing trends in the development of neuro-fuzzy systems can be found in [88-92].

2.4. TYPE-2 FUZZY SETS AND SYSTEMS

As just discussed, the concept of fuzzy (type-1) logic was introduced by Lotfi Zadeh in 1965 [8] as an extension of classical sets to mimic human actions in its use of approximate reasoning in the process of decision making. The goal is to represent uncertainty and vagueness mathematically in order to provide a formalised tool for dealing with information incompleteness, imprecision, and inconsistency of many complex real-world problems. Since knowledge can be expressed in form of linguistic rules (natural language) by using fuzzy sets, many real-world complex problems can be greatly simplified.

Type-1 fuzzy logic systems (T1-FLSs) have been extensively applied in many real world applications [93-95]. However, many real-world applications exhibit measurements noise and modelling uncertainties. The effects of all types of uncertainties cannot be minimised and modelled by using type-1 fuzzy logic systems. Thus, the concept of general type-2 fuzzy sets (T2-FSs) is also introduced by Lotfi Zadeh in 1975 [96] as an extension of the ordinary type-1 fuzzy sets. Since the degree of membership functions in T2-FLSs are themselves fuzzy, while the degree of membership functions in T1-FLSs are crisp. This property provides extra degree of freedom and flexibility in modelling the uncertainties frequently encountered in real-world modelling tasks. Thus, type- 2 fuzzy sets can better model uncertainties and minimize their effects. In [33] Karnik et al. have developed the theory of type-2 fuzzy sets. More details on the theoretical background and design principles of type-2 fuzzy system (T2-FLS) are described in [34]. T2-FLSs measure the entire systems uncertainty and thus they appear to be a more promising method for handling uncertainties (e.g. in a noisy changing environments) than their type-1 counterparts. The studies in [97-99] confirmed the effectiveness of the T2-FLSs in better handing the measurement noise and modelling uncertainties.

When a system is characterised by a large amount of uncertainties, a desired level of accuracy may not be achieved via the use of T1-FLSs. In such cases, the use of T2-FLSs is a preferable methodology as appeared in the literature in many applications where uncertainties occur, such as time series forecasting [100], decision making [101, 102], solving fuzzy relation equations [103], data pre-processing [104], survey processing [99], fuzzy logic control of mobile robot [98], and many more [105, 106].

2.4.1. GENERAL TYPE-2 FUZZY SETS

Type-2 fuzzy set (T2-FS) as introduced by Lotfi. A. Zadeh [96] has a membership function (MF) that is itself a fuzzy set in [0, 1] [96, 107], unlike a normal fuzzy set (T1-FS) where the membership function has a crisp number in [0, 1]. In [108], Mendel defines the definitions of T2-FSs as:

A type-2 fuzzy set (T2-FS) in a universe of discourse U is denoted as $\widetilde{A}$ which is characterised by its membership function (MF) $\mu_{\tilde{A}}(x,u)$, i.e,

$$\tilde{A} = \left\{\left((x,u), \mu_{\tilde{A}}(x,u)\right)\middle| \forall x \in U, \forall x \in J_x \subseteq [0,1]\right\} \qquad 2\text{-}11$$

in which $0 \leq \mu_{\tilde{A}}(x,u) \leq 1$.

Eq. 2-11 can also be expressed as

$$\tilde{A} = \int_{x \in U} \int_{u \in J_x} \mu_{\tilde{A}}(x, u)/(x, u), \quad J_x \subseteq [0, 1] \qquad 2\text{-}12$$

where x is the primary variable of $\tilde{A}$ and $x \in U$. u is the secondary variable of $\tilde{A}$ and $u \in [0, 1]$. J_x is called the primary MF and the amplitude of $\mu_{\tilde{A}}(x, u)$, called a secondary grade of $\widetilde{A}$.

Shown in Fig. 2.3 is a general type-2 Gaussian membership function built from synthetic random data in three-dimensional representation. As can be seen from Fig. 2.3, the membership value or membership grade for each element of the type-2 fuzzy set is a fuzzy set in [0, 1], whereas the membership grade of type-1 fuzzy set (normal fuzzy set) is crisp value in [0, 1].

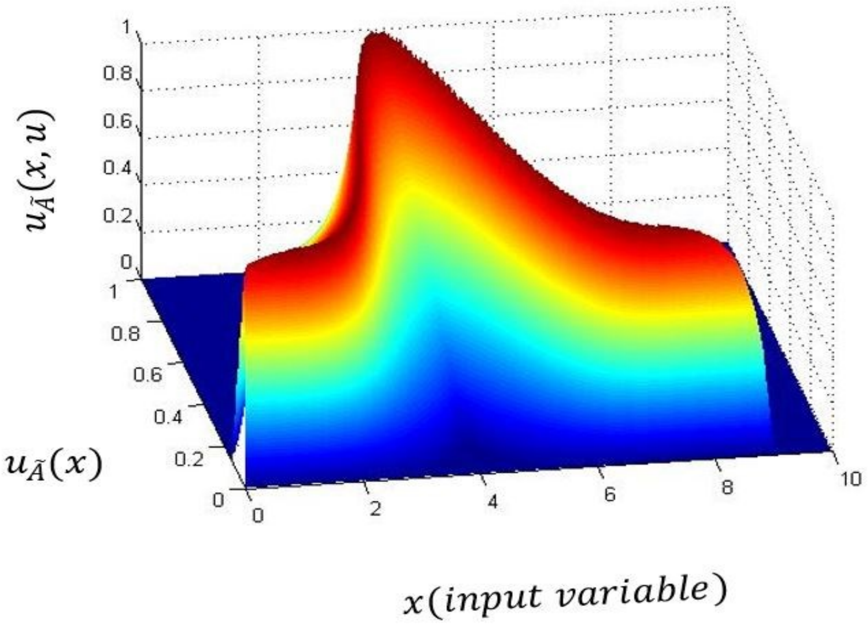


Figure 2.3. Three-dimensional representation of general type-2 membership function.

2.4.2. GENERAL TYPE-2 FUZZY LOGIC SYSTEMS

General type-2 fuzzy logic system (T2-FLS) has similar structure to the type-1 fuzzy logic system (T1-FLS), which is still based on expert knowledge [108]. In T1-

FLS, the expert knowledge is represented by IF-THEN linguistic rules and implied uncertainty, and thus a T2-FLS is also characterised by linguistic terms having uncertain premise part and/or consequent part which are then translated into T2-FSs. The T2-FLS block diagram shown in Fig. 2.4 consists of 5 stages [108]:

1. ***Fuzzification:*** is the process of mapping crisp inputs into T2-FLS to activate the inference engine.
2. ***Rule base:*** the rules in a T2-FLS represent fuzzy relation between the input space X and the output space Y. The rules have similar multiple-antecedent multiple-consequent IF-THEN form as in T1-FLS, however, the antecedents and consequents are type-2 fuzzy sets.
3. ***The inference engine:*** uses type-2 fuzzy set-theoretic operations to combine fuzzy rules from the rule base into a mapping from input type-2 fuzzy sets to output type-2 fuzzy sets. In T2-FSs theory [108], new set-theoretic operations of *join* ($\sqcup$) and *meet* operators ($\sqcap$) are introduced and used instead of *union* ($\cup$) and *intersection* ($\cap$) set-theoretic operators in type-1 fuzzy logic theory. The inference engine can be divided into three further stages: antecedent calculation, implication, and aggregation or combination. The output of the inference engine is type-2 fuzzy sets (known as combined or aggregated sets).
4. ***Type-reduction:*** is the process of transforming the output T2-FSs from the inference engine into T1-FSs known as *the type-reduced sets*.
5. ***Defuzzification:*** is the process of transforming the output of the type reduction process into a defuzzified crisp number.

General T2-FLSs are computationally intensive as compared to their type-1 FLSs counterparts due to the added type-reduction (TR) stage which relies on computing the centroids of a large number of T1-FSs (i.e. embedded sets) into which the T2-FS is decomposed.

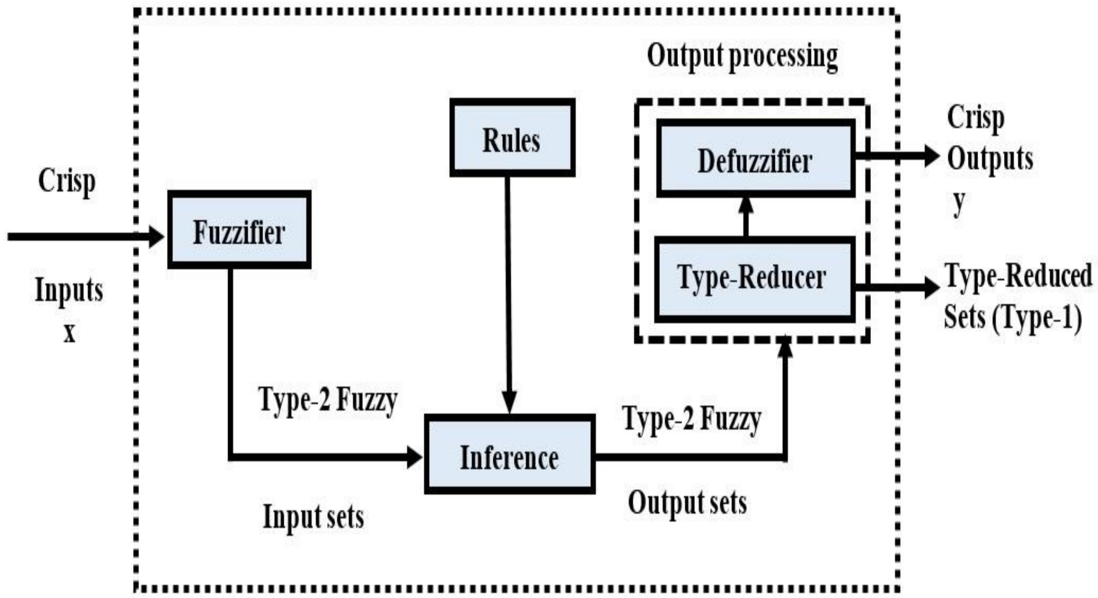


Figure 2.4. General type-2 fuzzy logic system block diagram [108].

Although the T2-FLS structure in Fig. 2.4 brings some advantages while dealing with uncertainties, it also increases the computational cost.

2.4.3. INTERVAL TYPE-2 FUZZY SETS

When $\mu_{\tilde{A}}(x,u) = 1, \forall J_x \subseteq$ [0, 1] in Eq. 2-13, then the secondary membership functions are interval sets such that $u_{\tilde{A}}(x,u)$ can be called an interval type-2 membership function (IT2-MF) [108]. Therefore, the T2-FS $\tilde{A}$ can be shown as:

$$\tilde{A} = \int_{x\in U} u_{\tilde{A}}(x)/x = \int_{x\in U} \left[\int_{u\in J_x} 1/u\right]/x \,, \forall J_x \subseteq \; 0,1 \qquad \text{2-13}$$

A Gaussian primary membership function having uncertain mean (centre) in m, $m \in [m_1, m_2]$ and fixed standard deviation (spread) σ having an interval type-2 secondary membership function can be called an interval type-2 Gaussian membership function as in Eq. 2-14. The interval type-2 Gaussian membership function can be expressed as [109]:

$$u_{\tilde{A}}(x) = exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right], \quad m \in [m_1, m_2] \qquad \text{2-14}$$

where m is the uncertain mean for the Gaussian function in the interval $[m_1, m_2]$, m_1 and m_2 are the means (centres) for the lower and upper membership functions respectively, and σ is the standard deviation.

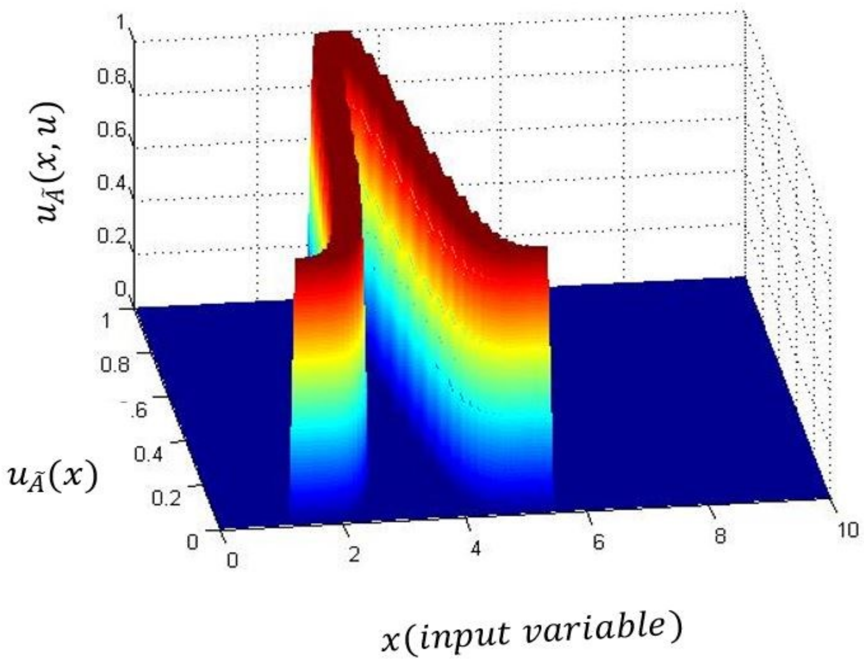


Figure 2.5. Three-dimensional representation of interval type-2 membership function.

From the three-dimensional representation of the IT2-MF in Fig. 2.5, it is obvious that the T2-FS is bounded by lower and upper MFs, which are denoted by $\underline{\mu}_{\tilde{A}}(x)$ and $\bar{\mu}_{\tilde{A}}(x)$ respectively. The region between was called by John and Mendel the footprint of uncertainty (FOU) [33]. Both of the lower and upper membership functions are type-1 membership function (T1-MF). Hence, the Eq. 2-15 can be re-expressed as:

$$\tilde{A} = \int_{x\in U} \left[\int_{u\in\left[\underline{\mu}_{\tilde{A}}(x),\ \bar{\mu}_{\tilde{A}}(x)\right]} 1/u \right] \Big/ x \qquad 2\text{-}15$$

2.4.4. INTERVAL TYPE-2 FUZZY LOGIC SYSTEMS

Due to the great computational complexity involved in processing T2-FLSs especially during the output processing unit (type-reduction and defuzzification stage), the interval type-2 fuzzy logic system (IT2-FLS) is the most widely used type of T2-

FLS [33, 107, 108]. Here the secondary membership function is an interval i.e. the secondary grades are all equal to unity. Using this type of fuzzy set makes the computation of meet and joint operations relatively easy and hence a considerable reduction in the type reduction stage [110]. It also distributes the uncertainty equally among all acceptable primary memberships.

The type-reduction process is an extension of the various forms T1 sets defuzzification process that we have. This is possible because of the Zadeh's Extension Principle [111]. There are many T2-FS TR methods [110], the most widely used are:

1. Centroid Type Reduction.
2. Centre of Sets Type Reduction.
3. Height Type Reduction.

The great computational resource required is clear from the centroid type reduction algorithm (exhaustive method) [110]. However if this was an interval type-2 sets (IT2-FS), the computational resource required is greatly reduced as shown by Karnik and Mendel in [111]. Variations of aforementioned methods exist and are now widely researched upon in order bring about a more efficient and computationally feasible type-reduction and defuzzification process. Some of the methods/representations developed in simplifying the output processing stage include, but not limited to, Wu-Mendel approximation [112], N-T algorithm [113], the Collapsing defuzzification method [114], and the sampling method [115].

In this thesis, the research activities are focused on the IT2-FLSs.

2.5. ARTIFICIAL NEURAL NETWORKS

An ANN, also known as a connectionist model is computational model that is inspired by structure and functionalities of neurobiological nervous systems such as the brain and process information [116]. It consists of a number of independent, simple processors – *the neurons*. These neurons are connected with each other via weighted connections – *the synaptic weights*. Thus, an ANN is actually parallel distributed processing system. Each neuron is a processing unit in an ANN. It is a model that performs the basic mathematical operation of an ANN and produces an output according to an activation function. The activation function expresses a nonlinear or linear function that maps the input into output and is denoted by $h(.)$. Fig. 2.6 illustrates the simple neuron model introduced by McCulloch and Pitts [117]. The McCulloch-Pitts (MCP) model is used in most of ANN models. The output of the neuron can be expressed by the following two equations [117]:

$$v_j = \sum_{i=1}^{M} w_{ji} x_i + b_j \qquad j = 1, \dots, K \tag{2-16}$$

The variables v_j are then transformed by the activation function $h(.)$ to give output values y_j

$$y_j = h(v_j) \tag{2-17}$$

where $x_1, \dots, x_M$ are the ANN inputs, $w_{j1}, w_{j2}, \dots, w_{jM}$ are the synaptic weights, b_j are the bias parameters, K is the number of neurons, M is the number of inputs, and $h(.)$ is the activation function.

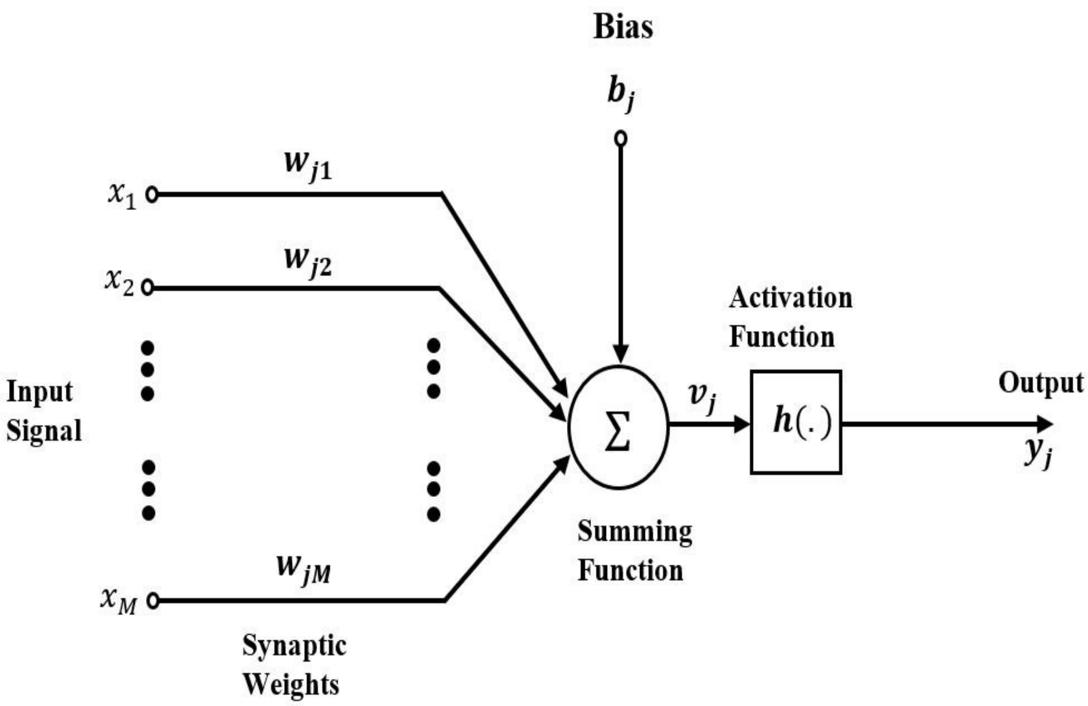


Figure 2.6. Artificial neural network general structure.

The choice of activation function is determined by the nature of the data and usually a continuous or discrete function that maps real numbers into the interval [-1, 1] or [0, 1]. The most popular activation functions used in ANNs are illustrated in Fig. 2.7.

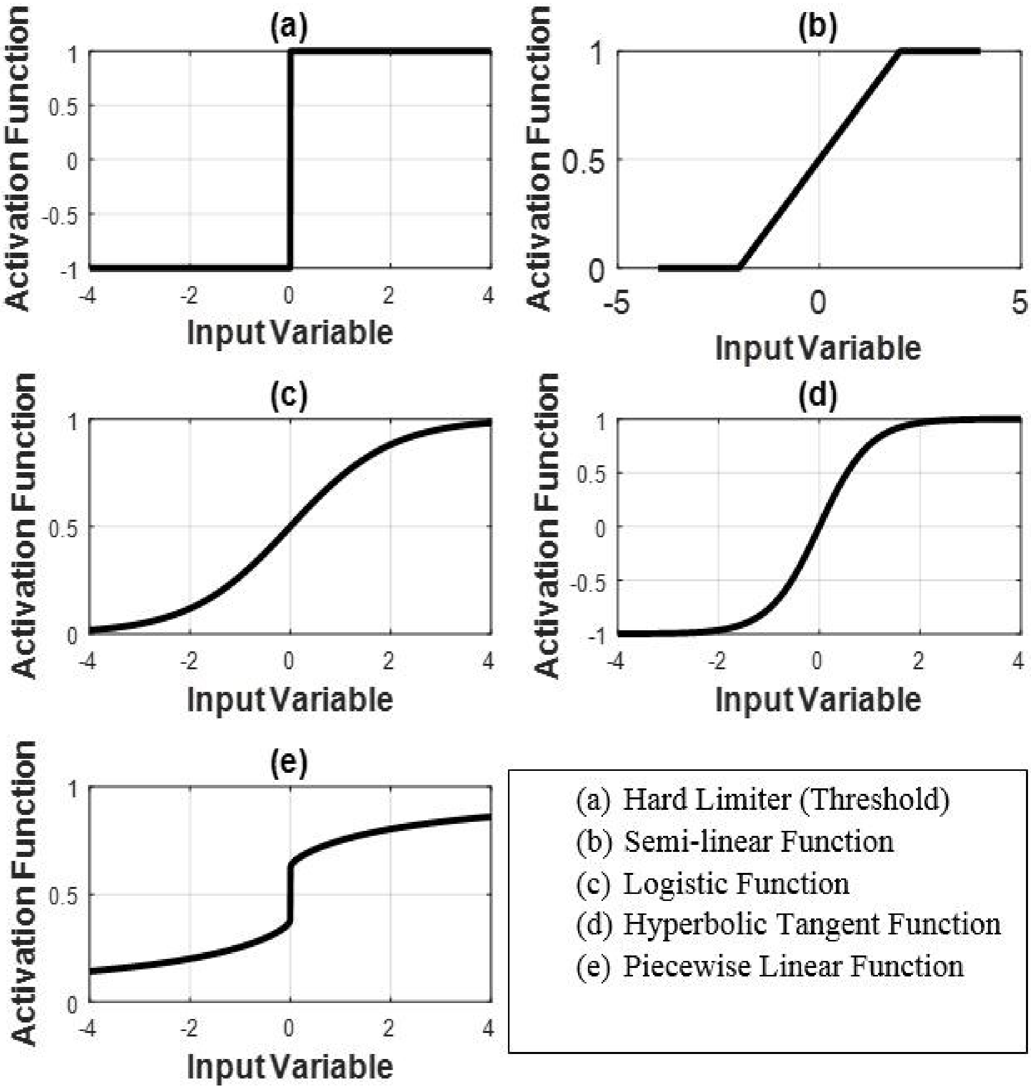


Figure 2.7. Activation functions.

The mathematical expressions for the activation functions can be expressed as [117]:

- **Hard limiter (threshold)**

$$h(x) = \begin{cases} +1, & x \geq 0 \\ -1, & x < 0 \end{cases} \qquad \text{2-18}$$

- **Semi-linear function**

$$h(x) = \begin{cases} 1, & x > a \\ \frac{x}{2a} + \frac{1}{2}, & -a \leq x \leq a \\ 0, & x < -a \end{cases} \qquad 2\text{-}19$$

- **Logistic function**

$$h(x) = \frac{1}{1 + e^{\beta x}} \qquad 2\text{-}20$$

- **Hyperbolic tangent function**

$$h(x) = tanh(\beta x) \qquad 2\text{-}21$$

- **Piecewise linear function**

$$h(x) = \begin{cases} \frac{1}{b - x + 0.5x^2}, & x < 0 \\ 1 - \frac{1}{b + x + 0.5x^2}, & x \geq 0 \end{cases} \qquad 2\text{-}22$$

The ANN architecture shown in Fig 2.7 is the simplest type of feed-forward neural networks, and it is known as a single hidden layer network (perceptron). The number of layers of adaptive weights is important to determine the network properties. A neural network with multiple layer of artificial neurons (computation nodes), where only one forward weighted connections from the input terminals towards the network's output are used, is known as multilayer perceptron (MLP) or multilayer feed-forward neural networks (MLF-NNs) [118]. Each MLP consists of input terminals, a number of hidden neurons connected between the input terminals (input layer of source nodes) and output layer. As mentioned in [119], multiple layers of artificial neurons with nonlinear activation functions allow the neural network (NN) to map nonlinear and linear relationships between input and output (i.e., universal function approximators).

ANNs have a powerful learning ability, that is, they can be trained to map any input-output behaviour. The weighted connections are adjusted to capture/encode the problem information from the raw data. A variety of parametric identification methodologies (learning techniques) can be employed to tune the weight of each connection in order to reduce the error between the actual output and calculated output form the network. This can be achieved via minimising a predefined cost function. In general, there are three types of learning strategies [120]: supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning is the process of providing the NN with a series of input data patterns and comparing the network's output with the correct "target" output. Some of the supervised learning algorithms include learning vector quantisation (LVQ), recurrent cascade correlation (RCC), back-error propagation (BEP), etc. [118, 120].

In ***unsupervised learning*** strategy, the target output is not known for the training input vectors. The NN modifies its weighted connections so that the most similar input vector is assigned to the same output unit. The learning process extracts the statistical properties of the training patterns and group similar vectors into clusters (classes). Common unsupervised learning algorithms include Kohonen self-organising map (SOM), Binary adaptive learning theory (ART1), etc. [118, 120].

Reinforcement learning, in this methodology the target output "teacher" is assumed to be presented but the correct answer is not presented to the NN. The NN is only presented with an indication of whether the output answer caused an error or not. The NN uses this information to improve its performance. The reinforcement learning strategy can be used in case where the knowledge required to use supervised learning is not available. If sufficient information is available, the reinforcement learning can easily handle a specific problem. However, it is usually better to employ the other two strategies because they are direct and their underlying mathematical formulation is usually well understood. This learning strategy is common in robotics [120].

The back-error-propagation (BEP) learning algorithm is the most computationally straightforward learning algorithm for training the feed-forward neural network (FNN) topologies such MLP [121]. The BEP algorithm is a generalisation of the delta rule, known as the least means square (LMS) algorithm. Hence, it is also called the generalised delta rule. This kind of learning methodology uses a gradient optimisation technique to minimise a cost function (objective function) normally equivalent to the mean square error (MSE) between the correct and predicted outputs. The BEP indicates backward propagation of the error signal between the correct output and the network's output. Once an input pattern is provided to the network, the output of the network is calculated and compared with a particular output [118, 120]. The error of each output unit is also calculated. The error signal is then back-propagated, and a closed-loop control system is therefore established. The weight parameters are adjusted such that the error signal decreases after a number of iterations. The gradient of the objective function can be used as a termination criterion so that the algorithm can be terminated when the gradient or the change in the objective function is sufficiently small per each iteration [116].

2.6. RADIAL BASIS FUNCTION NEURAL NETWORKS

The Radial Basis Function neural network (RBF-NN) is a non-linear feed-forward neural network that has been used as an alternative approach to the MPL, since it has a simpler structure and a much faster learning process. The idea of RBF-NN derives from the theory of function approximation [22]. Some remarkable differences between the RBF-NN and MLP can be listed [116, 122]:

- The architecture of RBF-NN consists of only a single hidden layer, while the MLP-NN consists of more than one hidden layer.
- The activation function of each hidden unit (neuron) in an MLP-NN process the inner product of the input vector and weighted connections "weight parameters" vector of that hidden unit. Conversely, the activation function of each hidden unit in a RBF-NN processes the Euclidean norm distance between the input vector and the prototype "centre" of that hidden unit.
- The RBF-NN uses exponentially decaying localised nonlinear activation function to construct local approximations to nonlinear input-output mapping. Consequently, the learning is fast. On the other hand, the MLP-NN uses global activation function to construct global approximations to nonlinear input-output mapping.
- Due to the localised activation function in the RBF-NN and its influence to its neighbourhood is determined by its variance, the localised feature prevents the RBF-NN from extrapolation beyond the learning patterns. On the contrary, the MLP-NN has greater generalisation capability in the regions of the input space

of the training patterns where little or no training data is available, and thus the MLP-NN is a good candidate for extrapolation.

- In the localised RBF-NN, only a few receptive-field units in the hidden layer have significant activations for a given input pattern, meaning that the network adjusts the weight parameters only in the neighbourhood of same point and maintains constant weights in the other regions, thus less learning time for convergence of error surface. However, the MLP-NN based on the gradient-search techniques requires large number of iterations, each iteration involves a large amount of computation to converge and also often gets stuck into a local minima [123].
- The RBF-NN typically requires much more training patterns and more number of receptive-field units in the hidden layer to achieve a desired level of accuracy similar to that of the MLP-NN, which results in the problem of curse of dimensionality.

The general architecture of the RBF-NN is composed of one hidden layer of special receptive-field units, which pre-processes the input pattern and feed a single layer perceptron (see Fig. 2.8). Each receptive-field unit k in the hidden layer consists of the centre x^k of a given cluster (partition) k of the input space. The corresponding nonlinear activation function $\varphi_k(.)$ represents the similarity between any input pattern x and the centre x^k by means of a distance measure.

$$\varphi_k(x) = \varphi_k(\|x - x^k\|), \qquad k = 1, \dots, M \tag{2-23}$$

where M is the number of nonlinear functions which are known as radial basis functions, and $\|.\|$ denotes a *norm* that is usually Euclidean distance.

For a given set of N data points in a d-dimensional space, the input $x^k \in \mathcal{R}^d$ is mapped from the input space into the i-th dimension of the output space t^d according to the following mapping function:

$$f(x) = \sum_{k=1}^{M} \omega_k \varphi_k(\|x - x^k\|) \tag{2-24}$$

where ω_k represents the weight connections between the hidden units k and the output unit i.

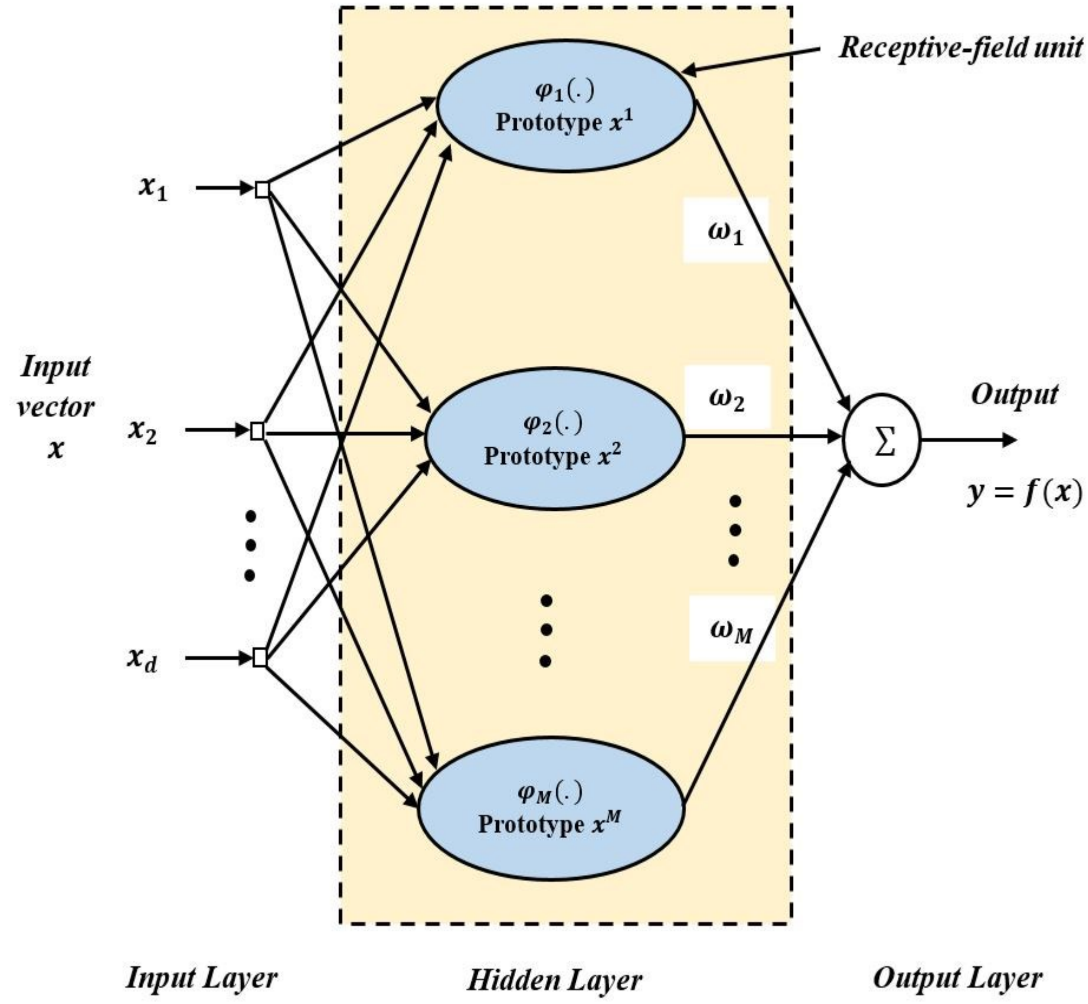


Figure 2.8. Structure of a RBF-NN based on interpolation theory.

According to the theory of multivariable interpolation in high-dimensional space, the interpolation problem can be expresses as follows [124]:

$$f(x^d) = t^d \tag{2-25}$$

where $x^d$is d-dimensional space for the input data x and t^d is d-dimensional space for the output data (target) t.

The interpolating surface is constrained to pass through all the training data points. When the interpolation conditions from Eq. 2-26 is inserted into Eq. 2-25, a set of simultaneous linear equations for the weight parameters ω_k can be obtained:

$$\begin{bmatrix} \varphi_{11} & \varphi_{12} & . & . & . & \varphi_{1M} \\ \varphi_{21} & \varphi_{22} & . & . & . & \varphi_{2M} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \varphi_{M1} & \varphi_{M2} & . & . & . & \varphi_{MM} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ . \\ . \\ . \\ \omega_M \end{bmatrix} = \begin{bmatrix} t_1 \\ t_2 \\ . \\ . \\ . \\ t_M \end{bmatrix} \qquad 2\text{-}26$$

where

$t = [t_1, t_2, \dots, t_M]^T$,

$\omega = [\omega_1, \omega_2, \dots, \omega_M]^T$ and

$$\Phi = \begin{bmatrix} \varphi_{11} & \varphi_{12} & . & . & . & \varphi_{1M} \\ \varphi_{21} & \varphi_{22} & . & . & . & \varphi_{2M} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \varphi_{M1} & \varphi_{M2} & . & . & . & \varphi_{MM} \end{bmatrix}$$

Equation 2-25 can be written:

$$\Phi\omega = t \qquad 2\text{-}27$$

where Φ is the interpolation matrix, ω is weights vector, and t is the target output.

According to Micchelli's theorem [125], for a set of N data points, the $N - by - N$ interpolation matrix Φ, whose its $ikth$ element $\varphi_{ik} = \varphi(\|x_i - x^k\|)$ is non-singular. Provided the inverse matrix Φ^{-1} of the interpolation matrix Φ. The weights vector ω can be represented as

$$\omega = \Phi^{-1}t \qquad 2\text{-}28$$

A number of nonlinear functions can be used as a radial basis function, the following functions are the most popular (see Fig. 2.9) [116]:

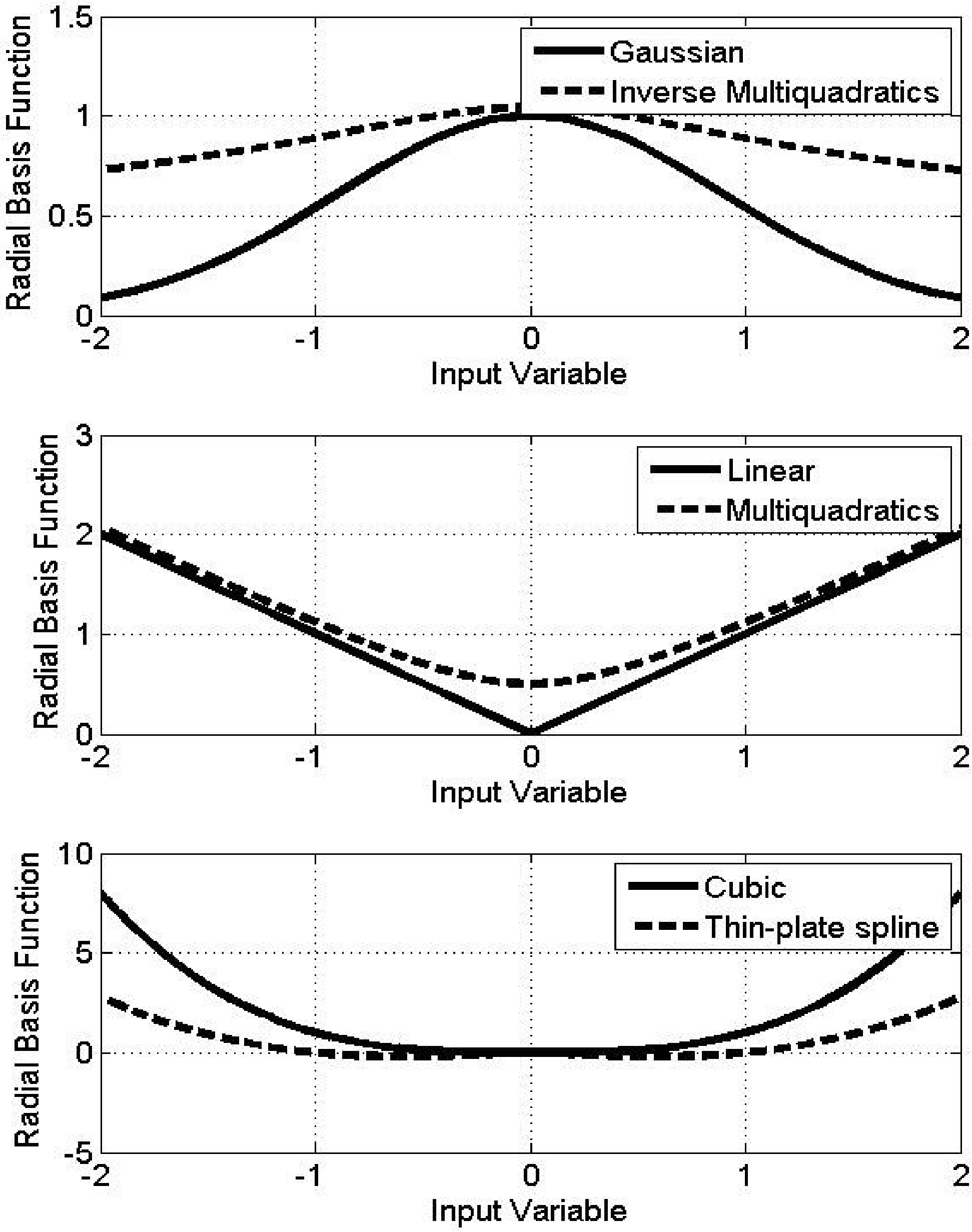


Figure 2.9. Plots of radial basis functions.

The mathematical expressions for the radial basis functions can be expressed as [116]:

- **Gaussian**

$$\varphi(r) = e^{-\left[\frac{r^2}{2\sigma^2}\right]} \qquad 2\text{-}29$$

- **Inverse Multi-quadratics**

$$\varphi(r) = \frac{1}{[\sigma^2 + r^2]^\alpha}, \quad \alpha > 0 \qquad 2\text{-}30$$

- **Multi-quadratics**

$$\varphi(r) = [\sigma^2 + r^2]^\alpha, \qquad 0 < \alpha < 1 \qquad 2\text{-}31$$

- **Linear**

$$\varphi(r) = r \qquad 2\text{-}32$$

- **Cubic**

$$\varphi(r) = r^3 \qquad 2\text{-}33$$

- **Thin-plate spline**

$$\varphi(r) = r^2\, ln(r) \qquad 2\text{-}34$$

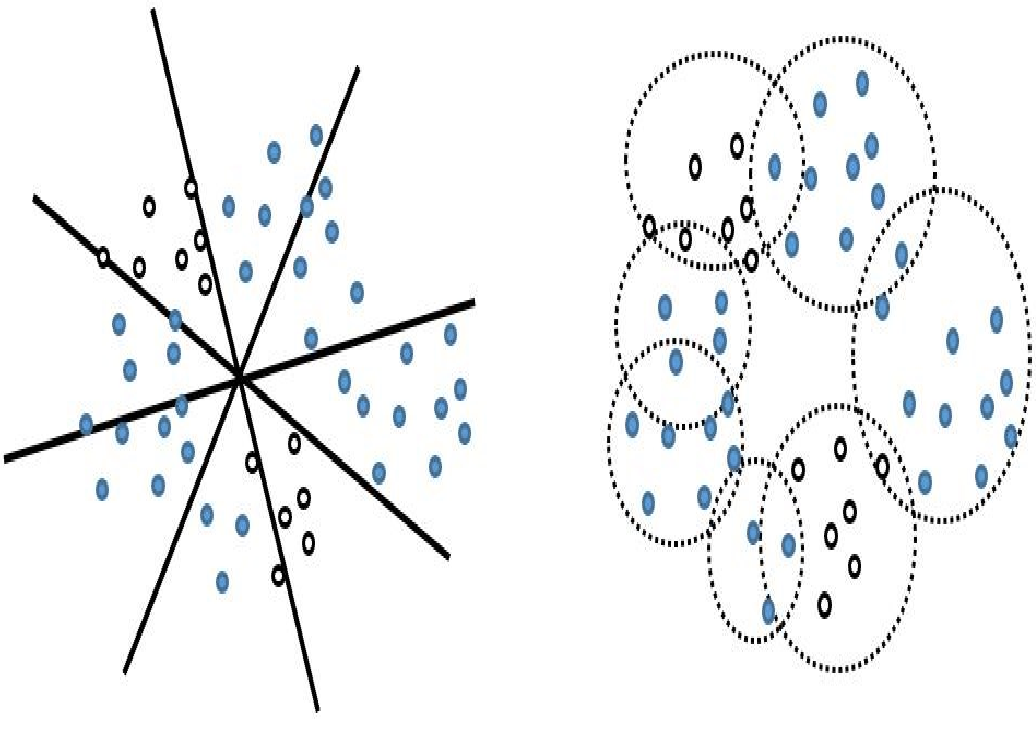
where $r > 0$ represents the distance from any input pattern x to a centre x^k for the same input variable, the parameter σ is used to constraint the smoothness of the interpolating radial basis function.

The most commonly used RBF in modelling and function approximation is usually the unnormalised multivariate Gaussian function. The Gaussian function can be expressed as follows:

$$\varphi(x, x^k) = exp\left(-\frac{\|x-x^k\|^2}{2{\sigma_k}^2}\right) \qquad 2\text{-}35$$

where x represents any input pattern, x^k and σ_k are the centre and spread for the same input variable.

The use of Gaussian function is motivated from the point of view of kernel regression and kernel density estimation. The Gaussian function possesses the property of locality that is $\varphi(x, x^k) \longrightarrow 0$ as $\|x\| \longrightarrow \infty$, which makes it highly desirable [118]. Consequently, the final results of RBF-NN can be easily interpreted, in contrast to the distributed nature of knowledge representation in MLP-NN, which makes the final results difficult to interpret. In addition, the Gaussian function is the only RBF that can be factorised, which makes it desirable for hardware implementation of the RBF-NN [118]. The RBF-NN with a localised Gaussian function is a receptive-field network, which provides the most accurate output when the input x is close the centre x^k of a receptive-field unit (neuron). For a suitably optimised/trained RBF-NN structure, similar input vectors (i.e., similar features), namely, input vectors that are near to each other, always produce similar outputs, while distant input vectors generate nearly independent outputs. Meaning that the RBF-NN partition the input data space into a number of clusters and each cluster is represented by only a few receptive-field units. The input layer activates the receptive-field units representing the input cluster, to which the input data pattern x belongs. The output layer linearly combines the outputs of the receptive-field units to generate the appropriate output vector. This constructs separation surfaces among classes of data in the input space (see Fig. 2.10).



(a) (b)

Figure 2.10. An example of separated surfaces generated by a) a MLP-NN b) a RBF-NN [118].

Since the output of the RBF-NN is a linear combination of Gaussian probabilities, the interpretation of the RBF-NN becomes visible from a statistical point of view. The Gaussian basis function as the posterior probability $p(cluster\ k|x)$ of input vector x belong to cluster k. In a similar way, the output weights $\omega_k$can be interpreted as the posterior probability $p(x^k|cluster\ k)$ of class x^k given the presence of cluster k [126, 127]. The learning of the RBF-NN requires the determination of the Gaussian radial basis function (x^k and σ^k) and the output weights parameters by minimising a suitable objective function. Several learning algorithms can be employed to optimise the RBF-NN parameters, one possible learning strategy could be the BP algorithm in which the partial derivatives of the cost function with respect to the RBF-NN's parameters have to be evaluated via the gradient descent procedure. The BP algorithm is a supervised learning strategy that yields high accuracy performance but with some disadvantage: a) it is a nonlinear optimisation problem which in general requires more computational effort and can trap into a local minima of the cost function, b) if the width parameters σ^k are not constrained to be small, the highly desirable feature of the basis functions can be lost, with a high drawback for both interpretation and computational speed. More efficient hybrid learning strategies have been developed using a two-stage strategy. The first stage specifies and fixes suitable centres for the Gaussian RBF and their respective widths, and then the second stage adjusts the output weights. These learning strategies have proven to be much fast than the BEP algorithm [128].

2.7. GRANULAR COMPUTING FOR HUMAN-CENTRIC INFORMATION PROCESSING

The concept of granular of computing (GrC) was firstly proposed by Zadeh in [64, 65] and Lin in [66] as a unified and coherent computational paradigm based on the human cognition of constructing, describing, and processing information granules. In a broad sense, GrC covers any methodologies, techniques, theories, and tools that make the use of information granule to solve complex problems [129]. Although the term GrC is relatively new, the first appearance of the concept was introduced by Zadeh in 1979 under the term information granulation (IG) in his pioneering paper [130]. There are many reasons for the study of information granulation. From a philosophical and theoretical perspective, many researchers argued that information granulation is very essential to human thinking and problem solving, and hence has a very significant effect on the design and implementation of human-centric intelligent systems. Since the GrC concerns human cognition (thinking) and problem solving, the fundamental three concepts underline this ability are: a) granulation, b) organisation, and c) causation. Granulation involves the process of dividing a large granule into smaller parts; organisation refers to the ability of forming small parts into a large granule; and causation refers to the association of causes with effects [64]. From a more practical perspective, the significance of information granulation and simplicity derived from information granulation in problem solving are perhaps some of the main reasons. In many situations, when a problem associates with incompleteness, uncertainty, or vagueness information, it may be difficult to differentiate distinct boundaries between information granules. For example, the granules of an image of any landscape consists of continents, countries, and oceans. If more detailed processing is required with a high level of abstraction, smaller information granules are sought with a certain criterion, and then these information granules are rearranged to constitute another finer information granules such as regions, provinces, states, seas, etc [30].

The concept of information granulation is inspired by the ways in which humans granulate and manipulate information [64]. The point departure in information granulation is the concept of a generalised constraint in which the granules can be a) crisp (c-granular) or b) fuzzy (f-granular). The latter represents perceptions in the sense that: a) the boundaries of the perceived values of variables are not sharply defined and b) the perceived values of variables are partitioned into small parts (information granules), with a granule being a bunch of points/objects drawn together by the criteria of similarity, indistinguishability, coherency, proximity or functionality [131]. For instance, Figs. 2.11 and 2.12 show crisp and fuzzy granules of the variable temperature (a perception of temperature) respectively and both of the granules can be described as very hot, hot, cold, etc.

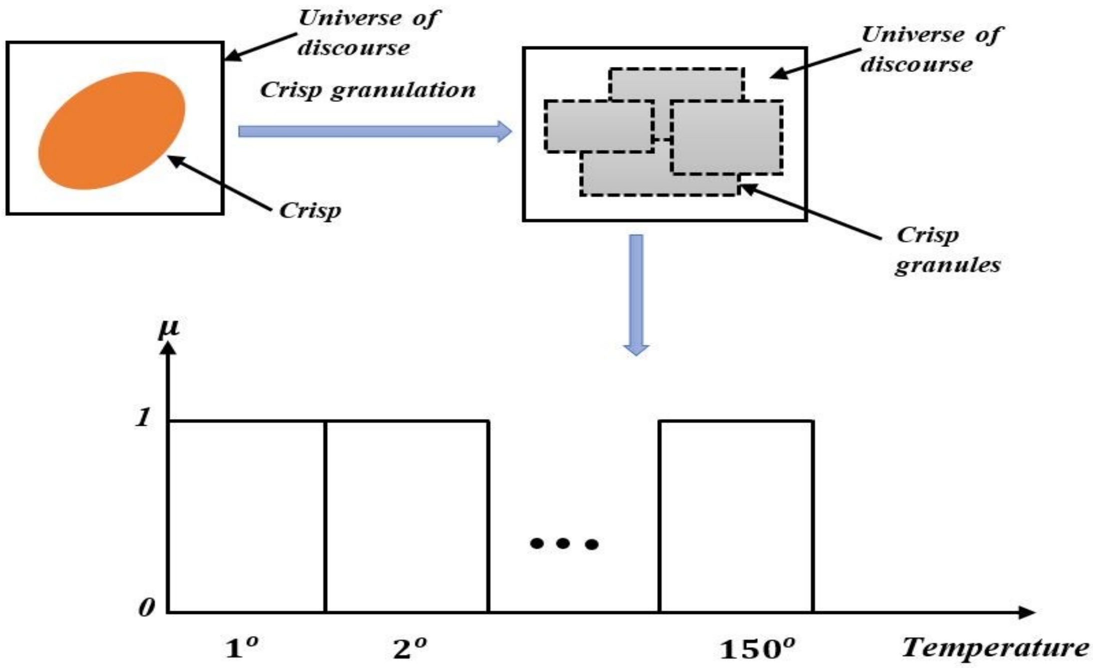


Figure 2.11. Crisp granulation of temperature.

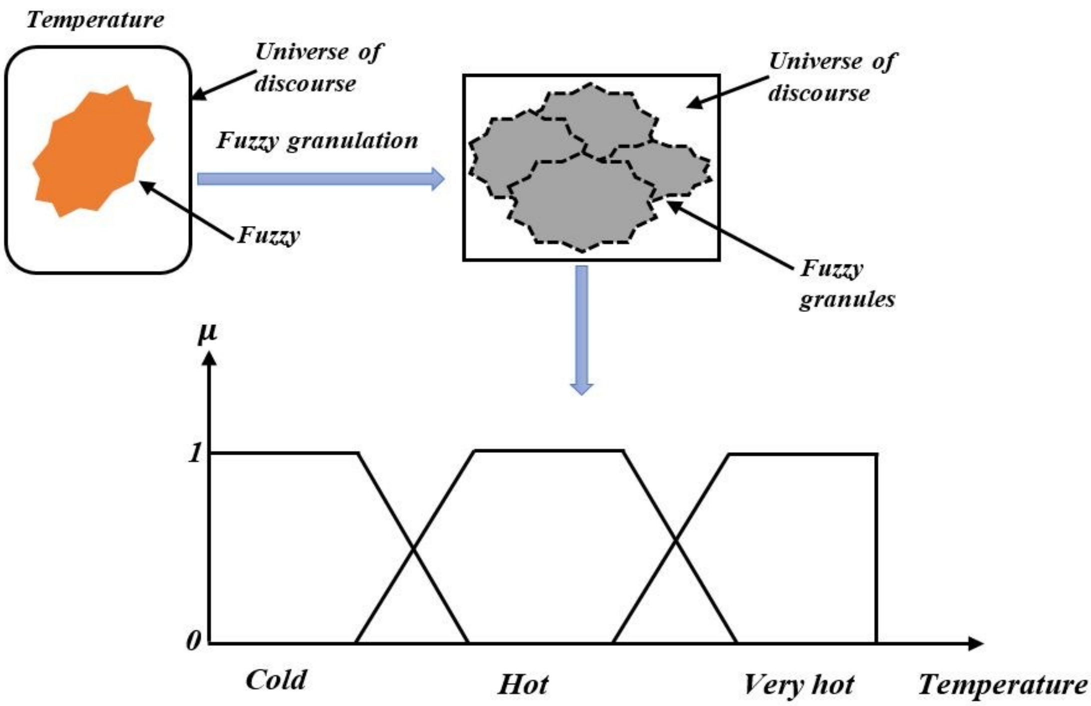


Figure 2.12. Fuzzy granulation of temperature.

The main modes of generalisation in IG are fuzzification (f-generalisation); granulation (g-generalisation); and fuzzy granulation (f.g-generalisation), which refers to combination of f-generalisation and g-generalisation. Figs. 2.13, 2.14, and 2.15 illustrate the concepts of fuzzification, granulation, and fuzzy granulation respectively. F.g-generalisation underlines the fundamental concept of linguistic variables, IF-THEN rules, and fuzzy graphs in fuzzy logic systems (FLSs). Crips granules (c-granular) have been successfully applied in conjunction with other methodologies such as set theory and interval analysis [132], probabilistic reasoning [131], decision trees [133], Dempster-Shafer theory [134], etc. However, they suffer from the ability to deal with entities/objects. Thus, the f-granular can be used to reason with entities/objects because it is inspired by the remarkable human ability to operate on and reason with perception-based information.

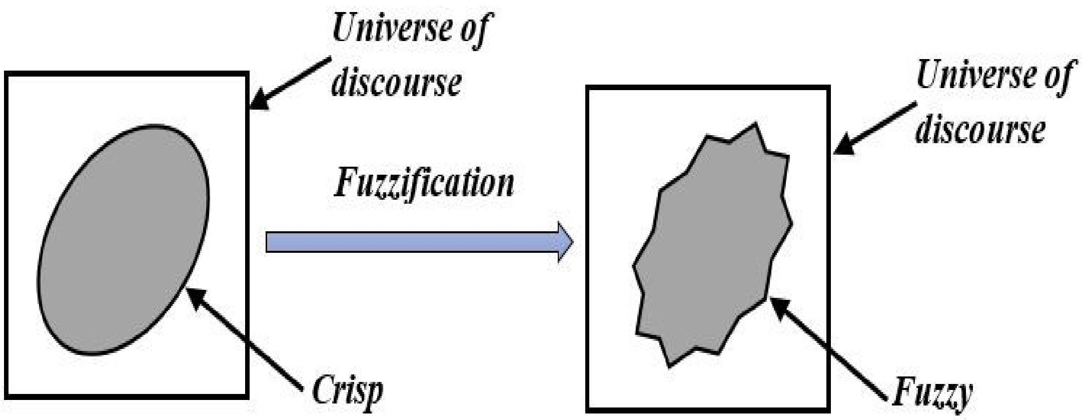


Figure 2.13. Fuzzification: crisp set → fuzzy set

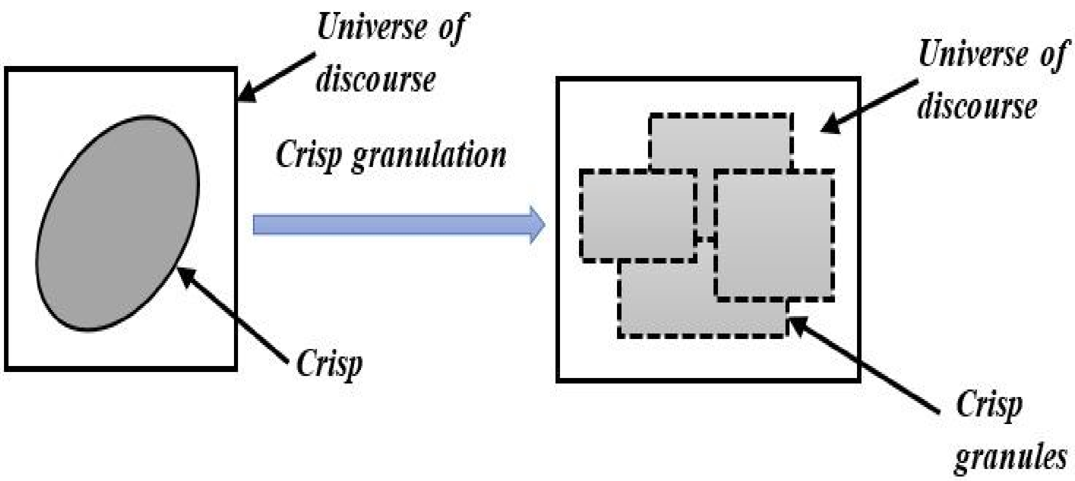


Figure 2.14. Crisp granulation: crisp set → crisp granules.

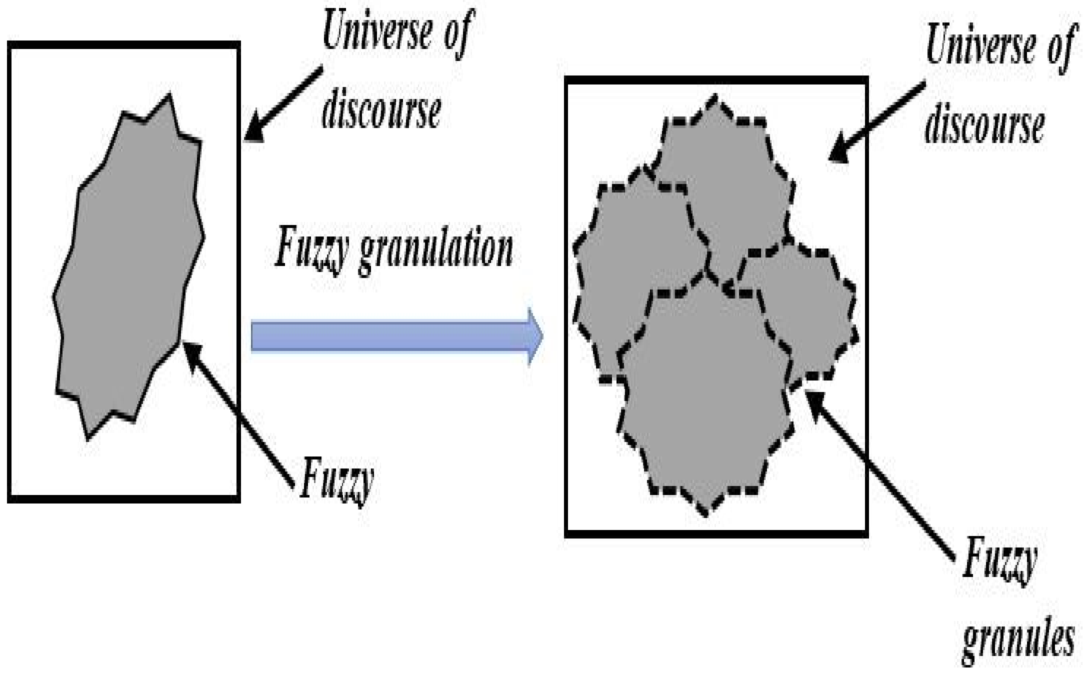


Figure 2.15. Fuzzy granulation: crisp or fuzzy set→ fuzzy granules.

In the context of information granulation, emerging human-like frameworks such as granular computing (GrC) are proposed as computational mechanisms that process complex information entities [135] [30]. In other words, GrC processes and extracts information out of numerical data to mimic the ability of the human beings to build information granules, manipulate them and communicate the final results to external environment. It may easily establish some connections between the concept of GrC and those of concept formation, knowledge discovery in databases (KDD), and data mining. The concept of formation can be viewed as the representation, characterisation, description, and interpretation of information granules representing certain concepts. While the concepts of knowledge discovery in databases and data mining refer to establishment of relationships between information granules such as association and causality.

2.8. CHALLENGES AND RESEARCH DIRECTIONS

In many real-world problems, it is often difficult and very challenging to derive reliable, and accurate first-principle model to represent the dynamics behaviours of complex systems. This is due to the high non-linearity and complexity associated with the systems. It is even more challenging to formulate real-time efficient models to approximate the dynamic behaviours of the systems and at the same time to provide a comprehensive understanding of influence of the process parameters on the system's response. In such a situation, data-driven computational intelligence (CI) modelling strategies can be exploited to build not only adequate and accurate process models, also transparent and interpretable process models based on experimental data. However, according to the theory of data modelling and knowledge discovery, quality of the data has a major impact on the performance of the model as the databases are rich source of hidden information. Knowledge derivation and information extraction out of the databases is regarded as the most critical step towards developing of human-oriented efficient (i.e., mimic the human cognition) CI models, since it aids to extract meaningful information and provide knowledge that can be easily understood by human operators as well as non-experts. The interaction between the derived knowledge and a personnel user can be performed via the use of simple linguistic interpretable rules (rule-based systems). To construct a computationally efficient and interpretable process model, a good quaintly data is required to extract knowledge from and use it in the modelling process. However, in reality most of the processes produce a small quantity of data that are often characterised by the presence of imprecision, inconsistency, incompleteness and sparsity. For example, a company manufactures aerospace products, FSW at its early stages, and some special bladder cancer with a few number medical records.

The main focus of this research work will be on the development of systems that can augment the cognitive capabilities of humans i.e. human-centricity or human-centric intelligent systems (HCISs). HCISs are different from traditional systems in many aspects in the way that they aid humans in performing cognitive tasks.

2.9. SUMMARY

This chapter has reviewed various key theoretical concepts of fuzzy sets such as different shapes of membership functions, fuzzy sets operations, fuzzy logic systems, rule-base, fuzzy inference engine, fuzzification, defuzzification, and comparison between Mamdani type fuzzy model and Sugeno type fuzzy model.

Since Zadeh first introduced the notion of a fuzzy set in 1965 [8] and afterwards went on to extend the concept via the idea of linguistic terms and introduced type-n fuzzy sets (TnFSs), which includes T2-FSs [136], the popularity and use of FLSs has been extraordinary. The concept of T2-FSs and T2-FLS, both general and interval are described. The footprint of uncertainty (FOU) was introduced in [33] to provide a very convenient verbal description for the entire domain of support of all the secondary grades of a T2-MF to better handle the uncertainties associated with the meaning of words. The general components of T2-FLSs including, the fuzzification, rule-base, inference engine processes, type reduction and defuzzification, and operators are also presented. A theoretical background related to ANNs with a particular focus on Radial Basis Functions Neural networks is also included.

In addition, the basics of granular computing for human-centred of information processing and different types of uncertainty measures related to uncertain based-information are presented since it is of great importance for the work of this theses.

In the next chapter, on the one hand an overview on Friction Stir Welding including the process description that is helpful for understanding the process and identifying some of the challenges will be provided. A comprehensive review of the process modelling, monitoring and control will also be presented. And, on the other hand a neural fuzzy modelling framework based on the Adaptive Neuro-Fuzzy Inference (ANFIS) and subtractive clustering is applied to FSW data, which were collected from TWI Ltd., Technology Centre (Yorkshire), United Kingdom.

CHAPTER 3 - FRICTION STIR WELDING AND PROCESS MODELLING

3.1. INTRODUCTION

The main objective of this chapter is to present Friction Stir Welding as a manufacturing process and review recent developments. In this chapter fundamental knowledge will be presented and in so doing will provide insight to some basic operational issues, which relates to the welding process. The principal of operation, development and reasons for using this welding technique with relevance to the advantage of this process for manufacturing of metal joining, will be discussed. In addition, a comprehensive literature review on the main areas related to the analytical and numerical modelling approaches, data-driven modelling approaches, and monitoring and control of the FSW process will be covered. Some preliminary results for the FSW prediction of internal process variable namely spindle peak torque, by using the well-known modelling framework of adaptive neuro fuzzy interference systems (ANFIS) and subtractive clustering are provided.

3.2. PROCESS DESCRIPTION

3.2.1. PRINCIPLE OF OPERATION

Friction Stir Welding (FSW) is a practical and non-conventional solid-state welding technique which was invented in 1991 by Wayne Thomas at TWI, United Kingdom [36]. In FSW, a constantly rotating, cylindrical shouldered tool with a profiled pin rotates and traverses at a constant rate along a pre-specified joint line, the tool’s pin is slowly plunged into butting edges of two firmly clamped sheets of the parent material. Material in the joint line is plasticised by frictional heat generated between the tool’s pin and the work-piece. The tool is moved along the weld joint when the material has been sufficiently softened. The plasticised material is transferred from the advancing side of the tool’s pin to the retreating side and vice versa forming a solid joint on cooling [137]. In order to provide a stable welding process, the use of a backing plate and applied side clamping forces are important. Fig. 3.1 shows the schematic drawing of FSW process.

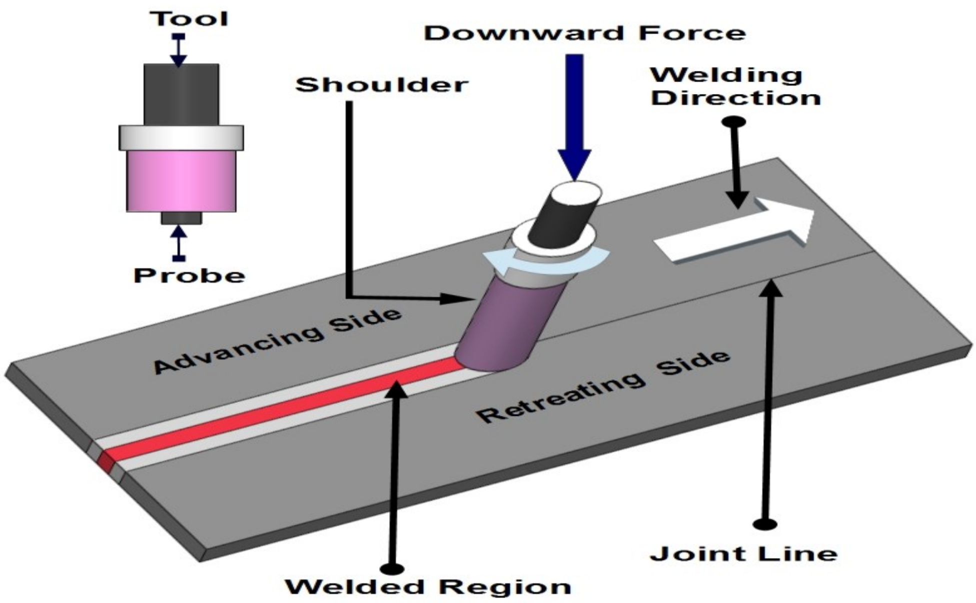


Figure 3.1. Schematic diagram of FSW process.

The FSW process involves four stages [138];

1. ***Pin's initial contact phase***

In the first stage, the non-consumable rotating tool's pin makes initial contact and starts penetrating into the work pieces. This contact generates sufficient frictional heat which results in making the surroundings around the tool's pin become viscous. The main controllable process parameters of this stage are the penetration angle, penetration depth, penetration rate, and tool rotational speed.

2. ***Dwelling phase***

The second stage begins once the shoulder makes contact with the surface of the materials being welded, the tool dwells inside the material for some time. This time known as the dwell time and it depends on the process operator.

3. ***Welding phase***

The third stage is the steady state stage or welding stage in which the tool advances along a predetermined welding path. As the tool travels along the welding line, viscous material from the advancing side of the pin is transferred to the retreating side and vice versa.

4. ***Pulling out phase***

Finally, once the welding stage completes, the tool is pulled out and this is known as the pulling out stage.

3.2.2. TOOL DESIGN

Tool design is the most important aspect of FSW process development [139, 140]. The tool geometry has a critical impact on the frictional heat generation, plastic material flow and the stirring action in the joint line and consequently controls the traverse rate at which FSW can be conducted. The FSW tool has dual functions:

- At the initial stage of the tool penetration, the heating mainly occurs as a result of the friction between the tool's pin and surface of the parent material to plasticise the work piece material. Additional heat is generated by plastic deformation of the material [137].
- The second function of the tool is to 'stir' and 'move or transfer' the material from the advancing side to retreating side and vice versa. The microstructural features and process loads are controlled by the tool design [137].

The design and choice of the material from which is manufactured must take into account the following features [140]:

- ability to reduce welding forces
- ability to flow plasticised material easily
- ability to facilitate the axial auguring effect
- ability to increase the interface between the tool's pin and the plasticised material, thereby more heat will be generated.

For FSW of steels, the pcBN and hybrid pcBN/WRe tool families are currently among the best performing tools [137].

3.2.3. ADVANTAGES OF FRICTION STIR WELDING

The main advantages of FSW process result from the fact that materials joining occurs in the solid phase without liquefying the parent material avoiding possible metallurgical complications such as porosity, cracking or detrimental metallurgical changes [141]. When compared to other conventional welding techniques, FSW is well-recognised in industry for its being [141, 142]

- environmentally friendly
- versatile
- lower energy consumption
- time efficiency
- the excellent post weld mechanical properties.

Friction stir welds have low distortions and shrinkage, no filler wire required, no gas shielding required for aluminium alloys, no arc or toxic fumes, no porosity, and no spatter. Additionally, there is no requirement for welder certifications, no post weld grinding, brushing or pricking required, and the appropriateness for automation and adaptable robot use to produce complex curvature welds. Friction Stir Welding has some limitations over other joining methods such as the moderately slow feed rate (welding speed), rigid side clamping forces device and backing plate, and exist hole at the end of the welds [140, 143].

3.2.4. APPLICATIONS OF FRICTION STIR WELDING

FSW has been successful in welding different grades and thicknesses of aluminium alloys and its extension to weld a variety of materials including plastics [144], metal matrix composites, titanium [145], magnesium [146], and copper [147, 148]. Since its patenting, FSW is considered to be one of the most important developments in materials welding and its applications are widespread in many industrial sectors such as in aerospace components, railway industry, construction industry, land transportation, automotive, spacecraft, marine, etc. [149]. Whilst most of the efforts to date have been concentrated on FSW of aluminium alloys and the process is well established, there is a considerable interests on it for different grades of steels. Due to the lower heat inputs associated with the FSW process (as compared to conventional welding techniques), it is expected to minimise the distortion and residual stress which is extremely important in welding of thicker materials, such as in shipbuilding and heavy manufacturing industries [37, 150].

Due to the high cost and relatively poor performance of the tool required, the transfer of the FSW process into the steel sector is less advanced. The initial technology barrier used to be the high cost and the relatively poor performance of the tool required, however this case is now improving due to the development of tools that are capable of producing industrially useful lengths of welds in steel. FSW of steel is performed at temperatures of up to 1,100˚C (measured at the tool), hence the tool must retain its strength at these high temperatures while being subjected to complex bending, rotational and fatigue loads. An additional complexity is that the FSW of steel is characterised by the presence of phase transformations which deem the process optimisation even more challenging [151].

3.2.5. TYPES OF POTENTIAL FSW FLAWS AND DEFECTS

Although the friction stir welding process usually produces fewer and generally less serious flaws and defects than conventional welding, it can still produce imperfect welds. Therefore, it is necessary to distinguish between defects and flaws [151, 152]. A defect is regarded as an imperfection in a welded part whose appearance is deemed harmful to the integrity of the weld and it must be removed. The common imperfections in friction stir welds are joint line remnants, lack of penetration, poor quality surface and inadequate plastic flow [153, 154]. While a flaw is considered as an unpremeditated imperfection in a welded part whose appearance may or may not affect the integrity of the welded part. After a careful inspection, it could be considered as a tolerable flaw and can be accepted or as a defect and must be removed. Some of common defects and flaws are:

A. Joint line remnant

The joint line remnant, sometimes called oxide entrapment or lazy S is a curve line that sometimes can be observed in all or part of the weld when inspected in cross-section [151]. It results from the oxide interface between the material surfaces during the formation of the friction stir weld is inadequately disrupted in order to form a joint line. Thus, it appears to consist of oxide particles from the work-pieces surfaces, drawn into and distributed throughout the weld during the material flow around the tool. If the oxide particles are small and widely distributed the oxide entrapment has little or no impact on weld strength. A continuous entrapment through the weld is however unacceptable [143].

B. Lack of penetration

If the tool pin is too short or the tool plunge depth is incorrectly set, or there is poor alignment of the tool relative to the weld line, there will be insufficient plasticisation and a lack of material flow at the bottom of the weld resulting in partial welding. In some cases, this will be an indication for a poor weld, often referred to as a kissing bond, where the parent materials are in close contact but not completely welded. The kissing bond defect is hardy to be detected by both visual inspection and by non-destructive testing techniques such as radiography. In more serious cases, the discontinuity between the parent materials that were welded is clearly visible. Since the lack of penetration can be a potential source of many factors that affect the joint line life such as fatigue cracks and corrosion. It is considered to be one of the most serious types of defect in friction stir welding [155, 156].

C. Poor quality surface (galling, excess flash, under-cutting)

A surface galling is type of flaws that occurs while welding some grade of alloys, especially at high tool rotational speeds. This is due to the high shear rates experienced at high tool rotational speeds. Also, the collection and redistribution of very soft material generated by the tool. Irregular surface that results can be an initiation source for cracks, particularly where cyclical loading is applied [157].

The flash is a type of flaw that occurs due to the softening of material by the excessive tool/shoulder frictional heat during the process. Surface flash is quite usual at the start of the weld since the tool must displace its own volume of material upon penetration to the material. This can be reduced by using a pilot hole at the weld start [153].

Undercutting occurs if the downward force applied to the weld is too high, especially under a combination of low welding speed and high tool rotation speed which results in very high plasticised material in the weld region, the weld surface will be depressed below the surrounding parent material. The undercutting can be a potential site for the initiation of cracking [151].

D. ***Inadequate plastic flow (voids and wormholes)***

An adequate level of plastic flow is required to make a friction stir weld. Inadequate or abnormal material forging (mixing or stirring) results in voids [143]. For example, with the tool rotating and travelling, the plasticised material around the tool's pin transfers from the advancing side to the retreating side, hence the tool will leave a cavity behind the advancing side. If the plasticised material flowing back from the retreating side is not sufficient to fill the vacated region completely and instantly before it gets cool, such volumetric defects will occur. Voids can be located at any part of the weld and they can be reduced or eliminated by a) increasing the axial force; b) increasing the tool rotational speed (or temperature) and c) reducing the shear strain rate on the welded material [151].

3.2.6. WELD QUALITY ASSESSMENT

Surface quality evaluation in a friction stir weld is a relatively simple and fast process, replying on a visual inspection [151]. In general, it provides a reasonably accurate indication to the quality of weld made. For instance, from an expert knowledge, when the surface quality of the weld is good and there no internal defects visible at the tool exit hole, it is a valid indication that the friction stir welding process is performing well. When the process starts to deteriorate, the surface quality changes perceptibly, thus, the weld surface quality is a primary factor used to assess the final quality of the welds made. The weld quality assessment is based on a visual inspection of the weld by taking into account the following factors [151]:

- Weld surface smoothness;
- consistency of the characteristic ripple markings formed by the FSW process;
- the appearance or nonappearance of undercutting at the welded region edges;

- the appearance or nonappearance of flash;
- and the appearance or nonappearance of detectable defects such as surface breaking voids;
- evidence of heating of the steel adjacent to the welded region.

Fig. 3.2 shows a weld made in 6 mm thick DH36 steel grade at a moderately low welding speed of 156 mm/minute. It can be observed that the weld surface is quite smooth with the rippled surface effect characteristic of conventional friction stir welding demonstrating an even spacing as the tool progresses from image right to left. The surface shows no signs of hot tears or 'stickiness' associated with the steel at the weld surface being too hot and over plasticised, nor does it show evidence of the cold tearing and machining characteristic of a friction stir weld being too cold. The weld surface is not undercut, i.e, depressed below the surface of the parent material. This particular weld would be assessed as having good to excellent surface quality (i.e. absence of the fine fill form flash would rate it as excellent).

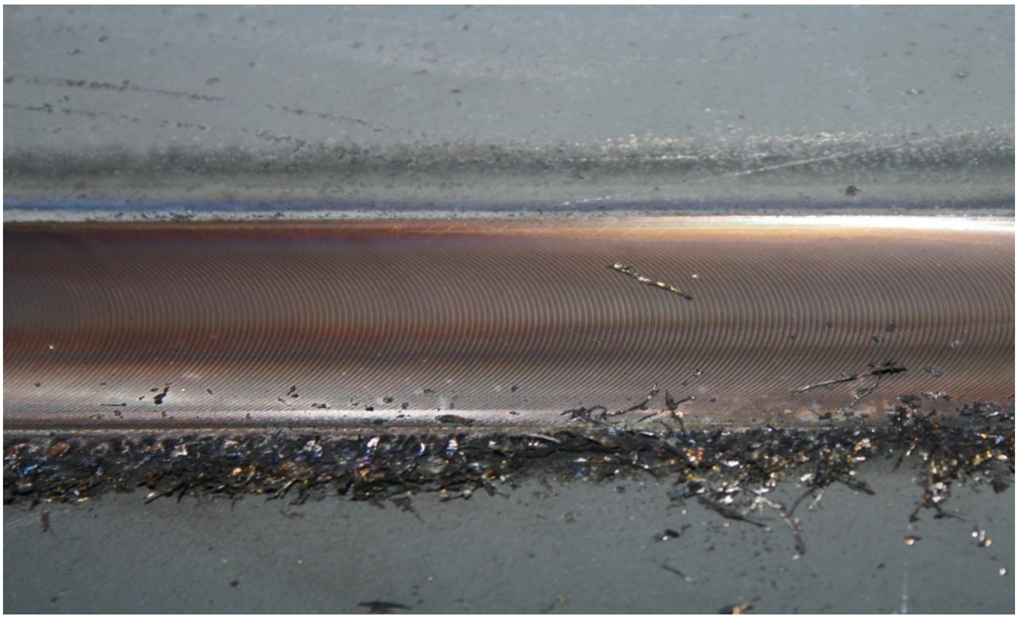


Figure 3.2. A weld made in 6 mm thick steel at a traverse speed of 156 mm/min and a tool rotation rate of 200 rpm.

By way of comparison, Fig. 3.3. shows a weld with poor surface quality. It can be seen that there is a considerable degree of thick flash appear on the retreating side of the weld, and the weld surface has a cyclical variation superimposed over the usual smooth ripple features. The large quantity of flash is characteristic of the weld being hot and the tool unable to contain the plasticised steel effectively. However, there is no visible undercutting, nor signs of surface voids and / or hot tearing. The primer alongside the weld also remains relatively unaffected by the heat of the welding process, i.e displays no burning or blistering. This would be considered a poor friction stir weld.

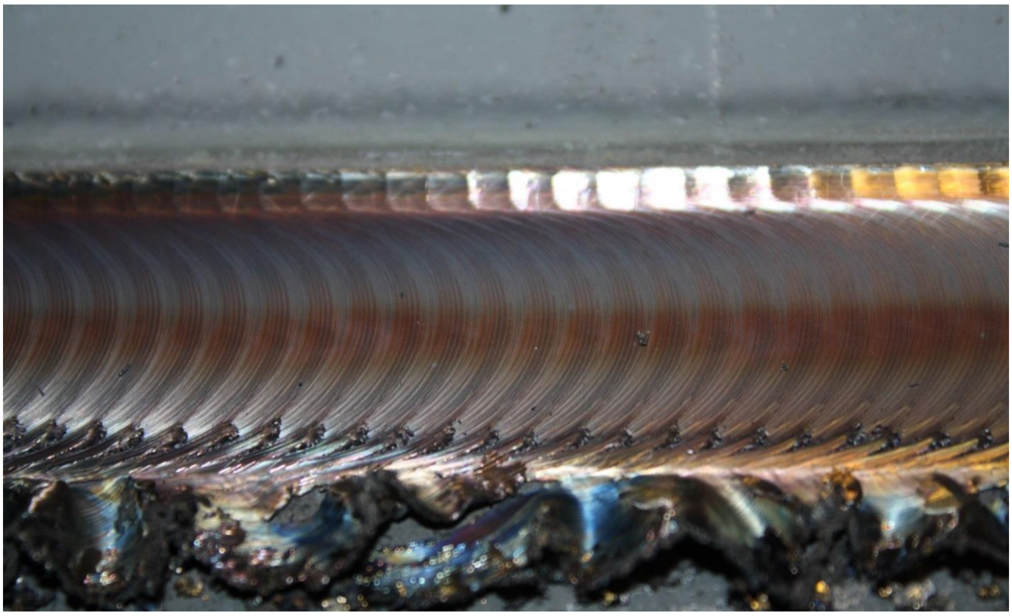


Figure 3.3. A weld made in 6 mm thick at a traverse speed of 400 mm/min and a tool rotation rate of 550 rpm.

Fig. 3.4 shows two samples of the welds produced using the same process conditions of traverse speed at 400 mm/minute and tool rotational speed of 450 rpm. In Weld WD 119 sample, the surface is relatively smooth and welded region is of even width along its length and there is no enormous flash produced and the weld surface is free from macro defects such as surface breaking voids and ‘stick and slip’ scuffing. Therefore, this weld would be considered a high quality friction stir weld. In contrast, the weld in WD 122 specimen has a wider welded region at the start of the weld and over-plasticised with considerable quantities of flash departing from under the tool shoulder. The friction stir welding process also exhibits hot tearing which results in surface breaking voids as the weld transitions from the plunge stage to the welding stage. Although the same process conditions were used to make the weld, the friction stir welding process produced a poor quality weld.

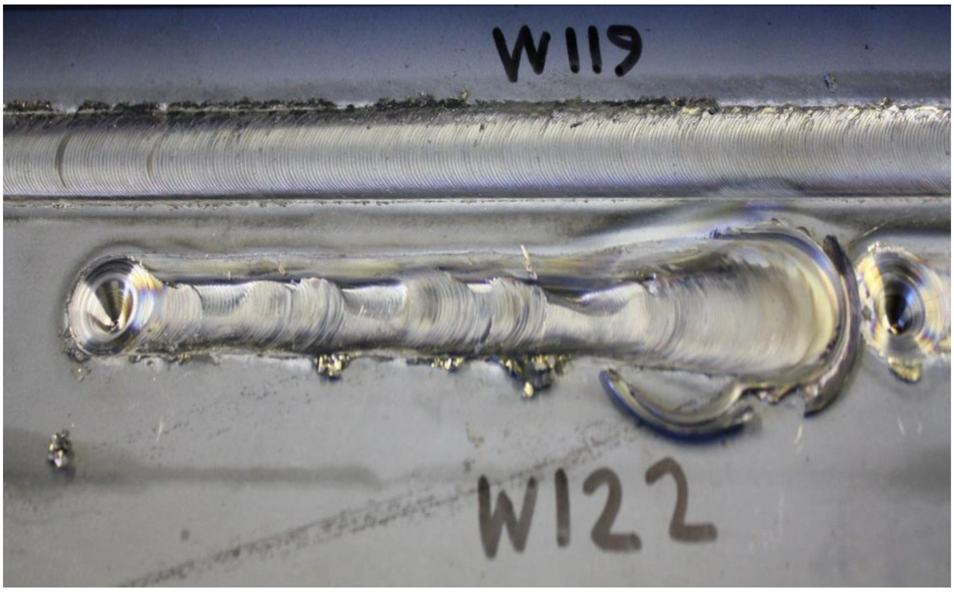


Figure 3.4. Butt weld of 6 mm thick DH36 steel grade using pcBN tool with tool rotational speed 400 rpm and welding speed of 325 mm/min: A WD119 sample with good quality (upper) and WD 122 sample with poor weld quality (lower) (Figure taken from [158]).

3.3. MODELLING, MONITORING AND CONTROL OF FRICTION STIR WELDING

3.3.1. OVERVIEW

Modelling is the process of constructing a simplified mathematical model from a complex physical system. In this process, numerical methods or models are used to describe a complex physical system [159, 160]. Modelling the FSW process is a very complicated task as the behaviour of the process includes physical couplings, very large plastic deformations and strain rates, material flow, mechanical stirring, surface interaction between the tool and the parent material, and high temperature within and around the stirred zone [161]. It also produces complex dynamic microstructural evolution, high shear forces in the plastically deformed material, and high temperature around the tool to nearly the temperature of the material melting point. The thermo-mechanical mechanisms occurring during the process can scarcely be fully quantified, in particular those concerning the contact conditions between the tool and parent material. In addition, there is no consensus on the type of law or equation (first principle model) to be used for the constitutive behaviour of the material flow at high temperatures and under high strain rates [137]. An understanding of the temperature history is essential, as it affects the microstructure and mechanical properties of the weld. Therefore, FSW is highly thermo-mechanically coupled process which makes it very challenging for researchers attempting to describe these phenomena using three types of modelling approaches (analytical, numerical and data-driven methods) [38, 160, 162].

3.3.2. CURRENT RESEARCH

3.3.2.1. ANALYTICAL AND NUMERICAL MODELLING OF FRICTION STIR WELDING

An analytical modelling approach is a mathematical model that uses exact theorems to present a closed form solutions of the governing differential equations which describes the process. The solutions to the equations or formulas that are used to describe changes in a system can be expressed as a mathematical analytical function [162]. A numerical modelling approach is a mathematical model that uses some sort of incremental time-stepping procedure to determine the state of a model and obtain its behaviour over time. The results obtained from such model is purely numerical, so that if someone wants to investigate/study the model's behaviour at different initial conditions, different initial values have to be fed in at the start of the modelling process and run the same incremental time-stepping procedure all over again which is time-consuming and computationally expensive [163].

Over the past few years analytical and numerical modelling methods have been intensively used for elucidating various aspects of the complex thermo-mechanical coupling phenomena associated with FSW. These computational models could also be helpful to better understand and visualise the influence of process input parameters on FSW process. The first numerical models appeared in the literature [164-166] to investigate the temperature history and they are based on Rosenthal's equation to describe the quasi-steady temperature field assuming a constant uniform pressure between the tool and the surface of parent material. The studies reported in [167] and [168] developed a three dimensional thermo-mechanical model based on FEA approaches. This model was used to investigate/study the thermal history and thermo-mechanical behaviour in the butt-joining of aluminium alloy 6061-T6. Their developed model can be extended to optimise the FSW process parameters to minimise the residual stress of the weld. In [169], Chen and Kovacevic continued their previous study and successfully studied the thermal history, stress and compute the mechanical forces in the lateral, vertical, longitudinal and directions. Other researchers have used CFD approaches to study the material flow and spatial velocity filed around the rotating tool's pin [170, 171]. There are other several analytical methods have been developed to calculate the heat transfer rate and materials flow but they are computational expensive [172-179]. A comprehensive review of the latest developments in the field of analytical and numerical modelling of FSW process, microstructures, and mechanical properties can be found in [38]. For instance, in the study reported in [169], the temperature was calculated based on Fourlier's equation:

$$\rho c \frac{\mathrm{d}T}{\mathrm{d}t} = \mathrm{div}(\kappa.\mathrm{grad}T) + q \text{ in } \Omega \tag{3-1}$$

where ρ is the material density, c is the heat capacity, T is the temperature, κ is the conductivity, and q is the power generated by friction between the tool and the top of the workpiece and by the plastic deformation work of the central weld zone.

The main source of heat in FSW is generally considered to be the friction between the rotating tool and the welded materials, and the "cold work" in the plastic deformation of material in the vicinity of the tool. The heat generation rate at the contact surface between the tool shoulder and the top surface of workpiec can be derived from the friction in the element at radius r is:

$$\mathrm{d}\dot{q} = 2\pi\omega.r^2\mu(T)p(T)\mathrm{d}r \tag{3-2}$$

The heat generation rate over the entire interface of the contact is calculated by:

$$\dot{q} = \int_{r_0}^{R_0} 2\pi\omega.r^2\mu(T)p(T)\mathrm{d}r = \frac{2}{3}\pi\omega\mu(T)p(T)(R_0^3 - r_0^3) \tag{3-3}$$

The heat generation rate at the interface between the shoulder and the top of the workpiece surface is a function of the coefficient of friction μ, angular velocity ω, and radius r. As the $\mu(T)$ and $p(T)$ are dependent on the local temperature and the radius r, Eq. 3-3 is difficult to evaluate. As the temperature increases, the friction coefficient is expected to decrease, and the plastic formation of the weld region increases. In this model, the p is an experimental value and the friction coefficient was kept constant in order to approximate the study effect of both factors of thermal and plastic effects during FSW, and the performance of model was verified against measured temperature history.

In spite of the intensive researches on FSW modelling by using approaches based on analytical models such as FEA and CFD models, however, all of these models have their own drawbacks and limitations, also the high computation cost, and limited accuracy which make it difficult for real-time use. Over the last few years, another solution has been suggested for modelling and optimisation of FSW, which is the use of data generated from the process. Consequently, data-driven modelling methods such as fuzzy logic systems, neuro-fuzzy systems, neural networks, and genetic algorithms are expected to provide a better solution for the modelling of FSW. In addition, low computation cost associated with the data-driven modelling approaches (as compared to the analytical and numerical modelling approaches) makes them feasible for real-time use. In this thesis, the research work is focused on the use of data-driven computational intelligence models for the modelling of friction stir welding. Therefore, a comprehensive review on data-driven modelling of FSW is provided in the next section.

3.3.2.2. DATA-DRIVEN MODELLING FOR FRICTION STIR WELDING

Based on the above motivations for the use of data-driven modelling to model the FSW, a comprehensive literature review related to this area is presented in this section. In [39], Okuyucu et al. used an ANN to model and study the effects of the FSW process parameters (traversing speed and rotation speed) on the weld mechanical properties namely: tensile strength, elongation, yield strength, and hardness for aluminium sheets. Moreover, in [178] Lakshminarayanan and Balasubramanian developed a three layer ANN to predict tensile strength of AA7039 aluminium welds, the developed model is capable of predicting the tensile strength within the range that it has been trained, but the generalisation ability was not good. The same model was used for the same modelling purpose by Jayaraman et al. [180]. The studies in [181] and [182] introduced a genetically optimised neural network model for the selection of optimum process conditions for the Friction Stir Spot Welding (FSSW), in these studies the process conditions are utilised as inputs to the model and the outputs are selected as the weld's penetration load and tensile force. Additional experiments were used to verify the obtained optimum process parameters.

The majority of the previous studies [183-187] based on neural networks modelling approaches focused on FSW of aluminium alloys, but the effects of the FSW process parameters on the microstructure, mechanical properties, and final weld quality are still needed to be further studied and interpreted. Moreover, there is a demand for reliable and accurate FSW models. In [40] Zhang et al. introduced a systematic data-driven modelling framework to model the FSW process for AA 5083 aluminium alloy, the proposed modelling framework confirmed to be interpretable, accurate, and robust and thus it can be further utilised to aid the optimal design of FSW process. Also, in [188] Zhang et al. studied the correlations between the input parameters, the internal process variables (namely bending forces), and the final weld quality. They have successfully developed a quality indicator to monitor the process in an on-line manner. In [41] Zhang et al. extended their previous work and successfully developed a multi-objective modelling approach to model the intricate FSW behaviours. The proposed modelling approach is used to forecast internal process variables, grain size as well as mechanical properties.

3.3.2.3. MONITORING AND CONTROL OF FRICTION STIR WELDING

Weld quality of a friction stir weld is determined by the combination of process parameters and internal process variables [151]. Therefore, to produce a steady FSW, process parameters (e.g. feed rate, spindle speed, tilt angle, plunge depth, etc.) and internal process variables (e.g. temperature, torque, forces, etc.) have to be monitored and controlled during the process. Measuring techniques of tool torque and feedback forces (Fx, Fy, and Fz) have to be implemented. A comprehensive investigation was made to establish the best practical methods for monitoring the crucial parameters during the FSW process and then selecting the best and most economical method. For all the various feedback forces, and tool torque measurements will be conducted by a method of strain gauging in one form or another. On the other hand, temperature history of the tool could be measured using a simply embedded thermocouple in the rotating tool or infrared thermal imaging camera, embedded sensors in the parent material. All these options also have their limitations, mainly being cost.

In the field of FSW monitoring and control, classical closed and open loop force and position control mechanism have been introduced in some researches [189, 190]. In the study reported in [191] a multi-input/multi-output (MIMO) neuro-fuzzy controller for nonlinear and complex curvature surface was proposed. The proposed monitoring/ control scheme consists of integrated sensor monitoring, fuzzy logic controller trained with back error propagation algorithm to generate on-line fuzzy rules. Despite the simulation results showed that the process conditions (temperature and torque in this study) were well maintained within limited ranges from their reference values, the proposed control scheme needs to be investigated in wide range of input parameters and process conditions and process parameters in order to prove its feasibility.

A recent study in [192], has investigated the effects of the number of tool rotations on the quality of friction stir spot weld of an aluminium alloy. The authors concluded that there is a linear relationship between the number of tool rotations during the spot weld of an aluminium alloy and the resulting tensile shear strength. In addition, a modified open-loop position control was proposed to monitor and limit the energy generated during the welding by regulating the dwell time. In [193], Su et al. proposed a methodology for measuring some of the process internal variables (namely the traverse force (X-axis), axial force (Z-axis) and the tool torque) simultaneously under different welding conditions for the FSW of AA2024-T4 aluminium alloys. The studies reported in [194] [42] investigated the frequency spectra of the tool feedback forces in X, Y, and Z axes and it was concluded that the frequency spectra of the feedback forces is more likely to contain useful information about the weld quality. Therefore, it would useful to use the information from frequency domain to build monitoring tools. In [194], the authors proposed a model-based classification algorithm that takes advantages of frequency domain information to build a weld quality marker for aluminium alloy. However, this model-based approach was proved to be not feasible as the changing process parameters also change the behaviour of the frequency domain information. Hence, the modelling performance cannot be generalised for different process conditions, as the model needs to be retrained every time the process condition is changed which would limit the usability of the model in real-time. Most of the research on the area of FSW monitoring is focused on tool design, parent materials, basic/fundamental understanding of the process, and post-weld properties [183-187].

3.3.3. CHALLENGES AND RESEARCH DIRECTIONS

Reviewing the past and on-going researches that focus on the area of data-driven modelling, monitoring, and control of FSW, a very limited number of researches has been reported but all of them are focused of FSW for aluminium, however in the field of friction stir welding of steel, no previous research has been conducted for steel friction stir welding. The focus of this thesis work will be in the field of data-driven modelling, process monitoring, and control for steel friction stir welding.

The previous studies reported and covered in the previous sections revealed a number of challenges, which makes the modelling process of FSW more complicated. This is mainly due to the following reasons:

- The complex thermo-mechanical behaviour makes the process highly non-linear.
- Measurements uncertainty of the real industrial data (noise, operator errors, etc.).
- Low process repeatability due to constraints on the quantity and quality of the real industrial data that results in similar statistical properties.
- High interaction between the multiple input parameters.
- Sparse and complex data space.
- The quality of the weld produced by FSW is influenced by a number of different factors in combination. These factors cannot be gauged due to some practical reasons, hence further complicating and creating more challenging process modelling conditions (for instance persistent tool tip temperature).

The focus of this thesis work will be on developing simple and computationally efficient real-time process models that have the capability of taking advantage of expert-knowledge to handle imprecision, inconsistency, incompleteness, sparsity, quantity and complexity of the associated process data. On the one hand, CI models that mimic the ability of human beings in using simple linguistic interpretable rules extracted from raw data in order to describe complex systems. On the other hand, CI models that can be used for on-line monitoring and real-time prediction (i.e. through process modelling and optimisation) to aid the process operator in making effective decisions as well as to guide the process operator to prevent overheating issues and problems related to tool wear leading to poor performance and hence poor weld quality.

The rest of the thesis work will be focused on the use of various concepts developed in human-centric intelligence systems including fuzzy sets theory, fuzzy logic systems and ANNs for modelling purposes.

3.4. SOFT-COMPUTING AND HUMAN-CENTRIC SYSTEMS FOR MODELLING OF FRICTION STIR WELDING

Friction stir welding is widely used in the industry for joining of metals and the success of the process is evident by the number of applications [37]. To guarantee the desired weld quality, it is crucial to have a complete control over the relevant process parameters on which the quality of a friction stir weld is based. Therefore, it is very important to select, design, control and optimise the FSW parameters for obtaining the desired weld quality. However, due to some uncertainty and thermo-mechanical phenomena (such as very large plastic deformations and strain rates, material flow, mechanical stirring, etc.), the relationship between the process parameters (inputs), internal process variables and the final quality in the friction welding process is nonlinear and complex, and it is very difficult to establish a precise first-principle mathematical model. Generally, researchers and metallurgists have to make a lot of welding experiments to study/analyse the influences of process parameters on weld quality and then design the standard/optimal parameters according to the obtained laws. Since the invention of the process, intensive researches on FSW modelling via analytical and numerical modelling approaches such as FEA and CFD models have been conducted, however, all of these models have their own drawbacks and limitations, also the high computation cost, and limited accuracy which make it difficult for real-time use [38]. The last decade, a fewer community of researches metallurgists have embraced the use of data-driven computational intelligence (CI) models via developing soft-computing techniques for modelling of FSW.

Whilst a research of the literature revealed few studies, which applied data-driven computational intelligence models in FSW in general, no single study exists reported for the data-driven modelling of FSW for steel. The main purpose of such CI models is to construct from numeric process data a relationship between the process input parameters, internal process variables, and final weld quality. In contrast to other data-driven CI models appeared in the literature, neural-fuzzy model approaches offer good level of accuracy (precision) and transparency (interpretability) due to their ability in combining the learning capability of neural network and capability of fuzzy systems in utilising simple linguistic interpretable rules extracted from complex high dimensional raw data.

The real industrial data used in this research work is a collection of a set of experimental trials conducted at TWI Ltd., Technology Centre, Yorkshire, United Kingdom. In order to be familiar with the process under investigation and the associated process data, several meetings were held at TWI Ltd with an expert to discuss the project's background and provide an insight of the data acquisition. The data set consists of 191 measurements on welding of two similar DH36 steel plates, which is used for shipbuilding applications [195]. The raw data set is provided in the Appendix. The experiments were conducted by using different levels of process parameters (rotation speed and welding speed) in order to obtain the best process parameters. During each experiment the data evolution for the internal process variables (namely, tool torque, axial force and traverse force) was recorded. Full details on data pre-processing and preliminary analysis is provided in Chapter 5.

In FSW, two input parameters are used to control the process: the rotation speed of the tool in the direction of clockwise or counter-clockwise and welding speed along the weld line. For the design of stable, safe, and practical steel FSW, it is very important to study relationships between the process conditions (inputs) and internal process variables. Establishing the correlations is helpful in avoiding the overheating and tool wear problems during the process. Due to the reasons mentioned in Section 3.4, and the complexity and high non-linearity of the data space and its sparsity; there are areas of low data density and areas of high data density (i.e. process operation envelope). By way of illustration, Fig. 3.5 shows the histogram of tool rotation speed and welding speed.

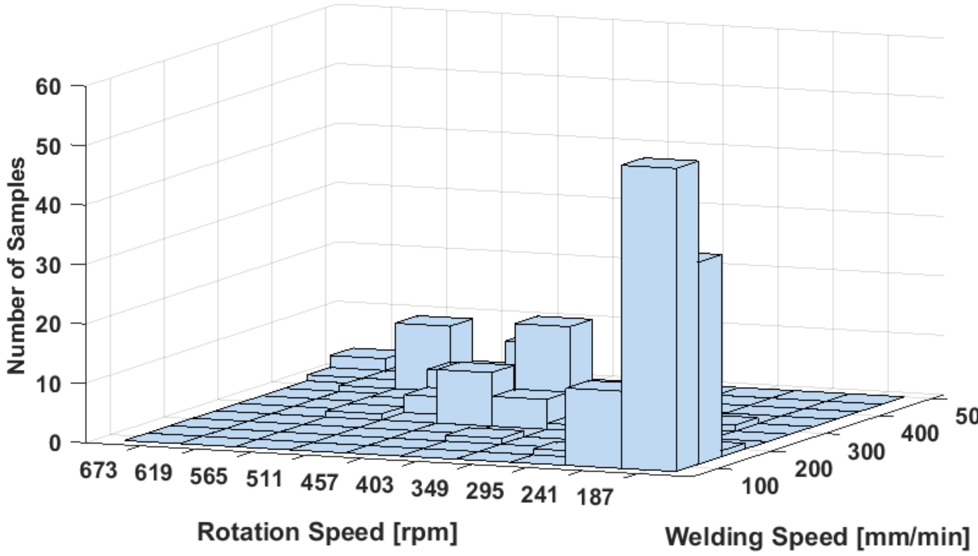


Figure 3.5. Data density example.

3.4.1. MODELLING OF FSW USING A NEURAL-FUZZY APPROACH

In a narrow sense, in the field of soft-computing a neural-fuzzy approach refers to the use of various technologies such as fuzzy systems, neural networks (NNs), and evolutionary computation to constitute some form of hybrid architecture. NFS is a computing paradigm that utilises on the one hand the capabilities of FLSs such as FSs, linguistic fuzzy rules, fuzzification, FIS, and defuzzification in order to develop interpretable yet transparent efficient process models. On the other hand, a NFS maintains the learning capabilities as well as generalisation properties of NNs to perform functional approximation for complex and highly non-linear systems. Moreover, a NFS has the ability to describe real system by using a set of linguistic rules and quantifying the uncertainty in a simple way, which can be easily translated into human language (rule-based systems).

As one of the earliest architectures for neural-fuzzy modelling approach, the ANFIS model [74] primarily represents a Sugeno-type fuzzy system in a special fivelayered feed-forward neural network [77]. ANFIS implements linguistic IF-THEN rules of the form

$$R_r: \boldsymbol{IF}\ x_1\ \boldsymbol{is}\ A_{j1} \boldsymbol{AND} \dots \boldsymbol{AND}\ x_n\ \boldsymbol{is}\ A_{jn} \boldsymbol{THEN}\ y = p_0{}^{(r)} + p_1{}^{(r)} x_1 + \cdots + p_n{}^{(r)} x_n \qquad 3\text{-}4$$

where $A_{j1}, \dots, A_{jn}$ denotes the membership functions for the antecedent part, $x = (x_1, \dots, x_n)^T$ is the input vector and $p_0, \dots p_n$ are the linear parameters of the consequent part of the Sugeno model and r denotes the rth rule. Fig. 3.6 illustrates the ANFIS architecture:

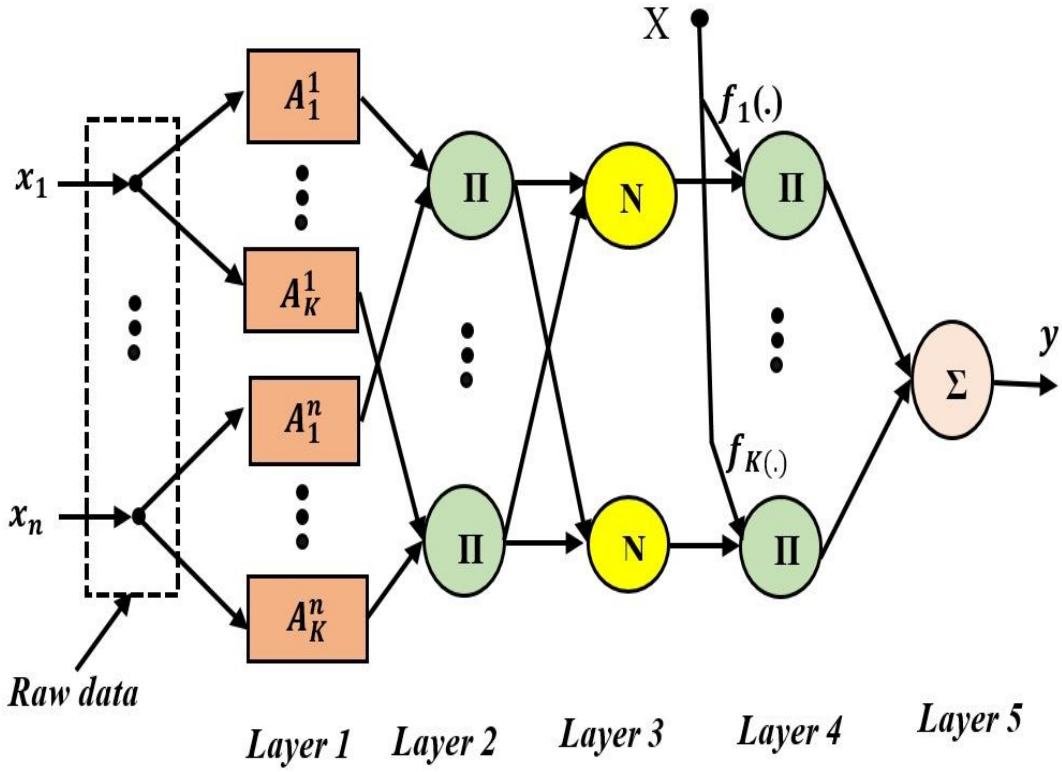


Figure 3.6. ANFIS architecture [74].

The functions of the various ANFIS layers are given below:

Layer 1 computes the membership degree of each input value.

Layer 2 uses the set-theoretic operator of t-norm "AND" to perform the aggregation and to determine the degree of fulfilment of each rule.

Layer 3 normalises each fuzzy rule and the outputs of this layer are called normalised firing strengths.

Layer 4 performs the consequent part "THEN".

Layer 5 computes the final output y of the ANFIS architecture.

More details on adaptive neuro-fuzzy systems can be found in [75, 116].

In ANFIS, the number of MFs for each input variable has to be decided in advance. However, the selection of an appropriate number of MFs for each input variable is based on a priori knowledge. Thus, it would be required to use an offline systematic selection approach to determine the appropriate number of membership functions for each input variable and therefore select the best structure of the fuzzy model in terms of accuracy and generalisation. The process of designing an ANFIS model usually consists of two steps, namely rule generation or initial model creation and system optimisation. Rule generation refers to the partitioning of the input space and identifying the corresponding set of rules, while system optimisation can be the optimisation/tuning the membership parameters and rule base. One of the methods that may be used to automatic generation of initial structure of an ANFIS model is clustering. Clustering is concerned with the partitioning/grouping of a numerical data set into several groups based on the similarity within a cluster/group to produce a concise representation of a system's behaviour. K-means clustering [196], Fuzzy C-means [197] are among other clustering methods that have been used to aid the design of neural fuzzy systems. However, in these methods the number of clusters must be determined in advance regardless of whether the obtained clusters are meaningful or not. Correct specification of the number of cluster is very important, however, because a large number of clusters results in an unnecessarily complicated rule-base, while a small number of clusters leads a poor model. Consequently, subtractive clustering algorithm [198] has been proposed for determining the "optimal" number of rules in a rule-base. Subtractive clustering is a fast and one-pass clustering algorithm for estimating/selecting the appropriate number of clusters and the cluster centres in a data set. More details on subtractive can be found elsewhere in [198].

Once the initial structure of the fuzzy model is specified via the subtractive clustering algorithm, the membership functions of the antecedents and consequent parameters are adjusted. The BEP learning algorithm can be utilised to tune both the antecedents and consequent parameters [199]. Fig. 3.7 illustrates the overall flow diagram of the data-driven model based on ANFIS model and subtractive clustering.

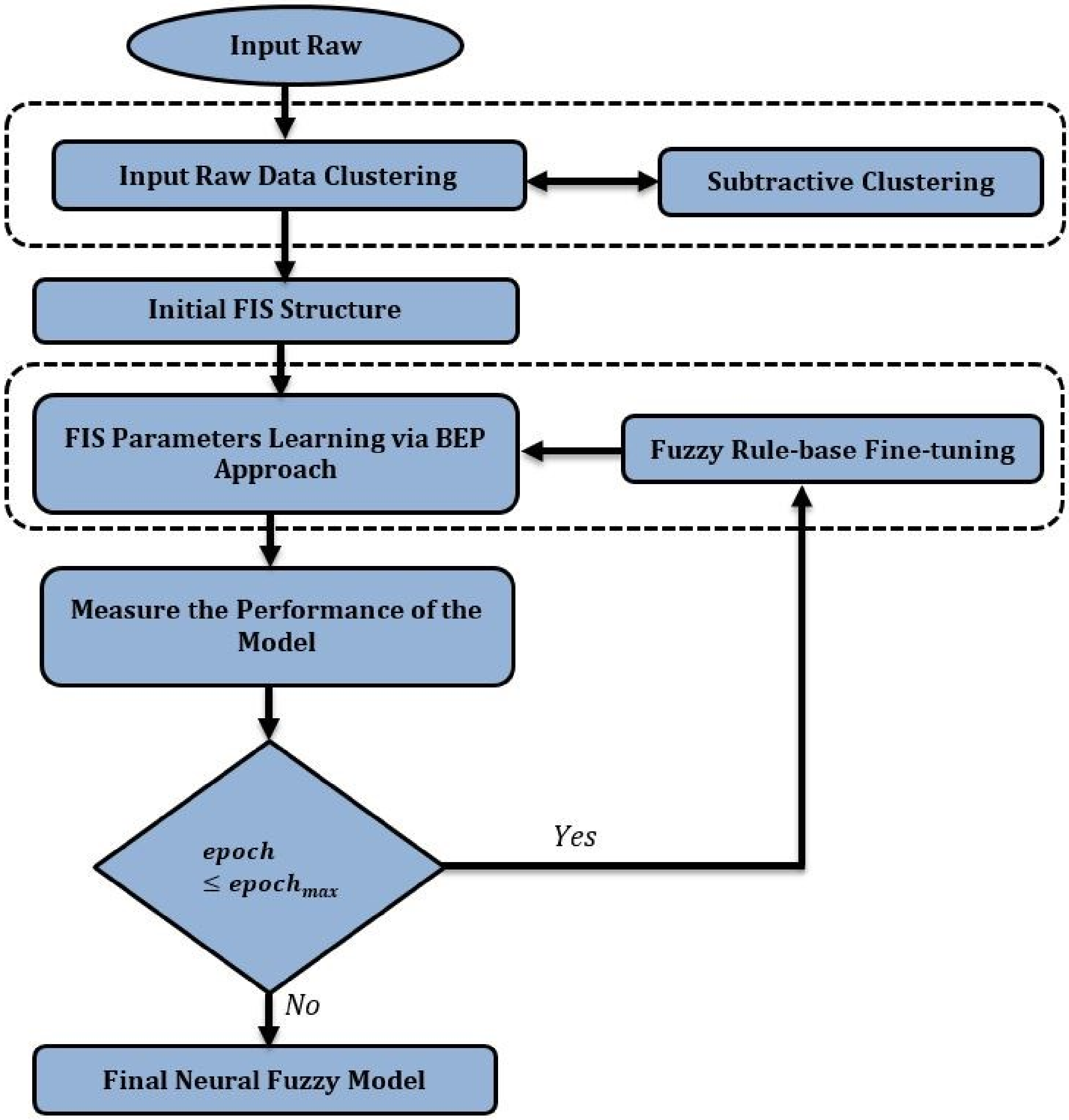


Figure 3.7. Data-driven model based on ANFIS and subtractive clustering.

3.4.2. PRELIMINARY MODELLING RESULTS

The proposed data-driven modelling framework is used as a benchmark model to model and predict the internal process variables namely spindle peak torque in this research work as shown in Fig. 3.8.

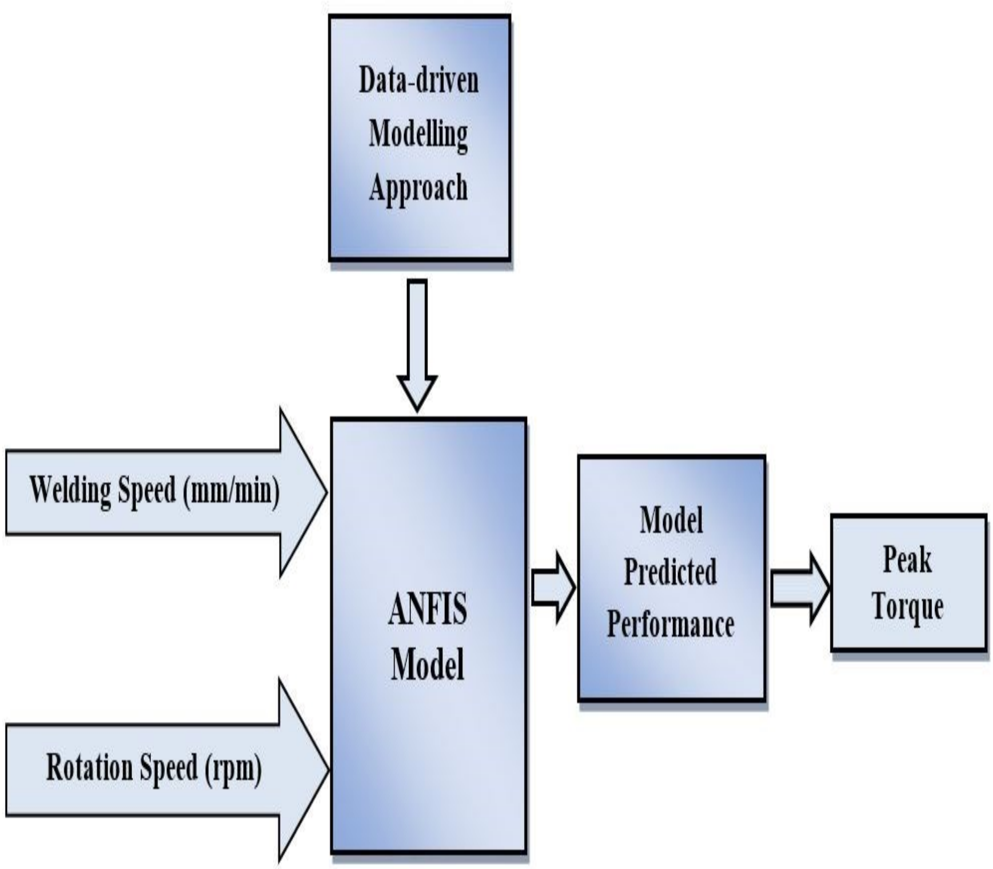


Figure 3.8. FSW modelling via ANFIS model.

The proposed data-driven model is used to derive the correlation between the FSW parameters of the DH36 steel plates and internal process variables. In the following modelling case relating to modelling of the spindle peak torque, the spindle torque is the amount of torque required by the shoulder to maintain the tool plunge depth into the joint line of the work-pieces, rotation rate, and applied force on the tool in the traverse direction. For the purpose of comparison, a Multiple Regression Linear (MRL) [126] model was utilised as a baseline and the results are provided in the Appendix. From the analysis of the results obtained it can be confirmed that the linear model provides only a basic level of prediction performance.

A more advanced nonlinear data-driven CI model of ANFIS technique was then used, for cross validation purposes, the data set was split into two sets, 133 (70%) data points to train the ANFIS model, and 58 (30%) data points to test the generalisation capability of the final model. By using this cross validation method, the interpolation capability of the model is expected to be improved with the risk of over-fitting. To avoid the over-fitting problem and improve the generalisation capability of the model, k-fold cross validation strategy is used in which the training set is split into k (i.e. 10 folds in this study) equally sized subsets and each time, one of the k subsets is used as the testing set and the other subsets are combined together to form the training set. Then the average error across all k runs is calculated. While the k-fold method systematically results in relatively low variance as the algorithm needs to run k times, thus becoming computationally expensive. Following the flow diagram in Fig. 3.7, the clustering algorithmic procedure employed for the initial structure identification of the ANFIS model is the subtractive clustering, which groups similar data points based on a density measure.

Since there is no prior knowledge about the number of clusters, with a number of systematic simulations (increased/reduced the radius cluster), it was established that the appropriate cluster radius specification was set to 0.40 which produced five clusters (five rules in the rule-base). The ANFIS model having less fuzzy rules (large cluster radius) achieved less accuracy performance but the rube-base is simpler in structure and more interpretable. However, the ANFIS model having more fuzzy rules (small cluster radius) captured more information about the dynamics of the process being modelled and achieved better accuracy performance with lack of interpretability and simplicity.

To discuss the results obtained from the Subtractive clustering and hence in relation to the initial structure of fuzzy rule-base, it would be worth to provide an illustrative example of the final shape of the membership functions (MFs) after subtractive clustering. Therefore, in Fig. 3.9 the initial universe of discourse after the application of Subtractive clustering for the dimensions that linguistically describe the welding speed and rotation speed is presented. The initial rule-base can be employed to describe the complex and non-linear behaviour between welding speed, rotation speed and the predicted spindle peak torque, which can be achieved by taking advantage of fuzzy logic systems. The corresponding fuzzy rules in linguistic format is as follows:

Rule 1: IF** Welding Speed **is** very small **AND** Rotation Speed **is** very small, **THEN** Peak Torque **is $y = p_0^{(1)} + p_1^{(1)}x_1 + \cdots + p_n^{(1)}x_n$

Rule 2: IF** Welding Speed **is** high **AND** Rotation Speed **is** high, **THEN** Peak Torque **is $y = p_0^{(2)} + p_1^{(2)}x_1 + \cdots + p_n^{(2)}x_n$

Rule 3: IF** Welding Speed **is** small **AND** Rotation Speed **is** Small, **THEN** Peak Torque **is $y = p_0^{(3)} + p_1^{(3)}x_1 + \cdots + p_n^{(3)}x_n$

Rule 4: IF** Welding Speed **is** Very High AND Rotation Speed **is** Very High, **THEN** Peak Torque **is $y = p_0^{(4)} + p_1^{(4)}x_1 + \cdots + p_n^{(4)}x_n$

Rule 5: IF** Welding Speed **is** Medium **AND** Rotation Speed **is** Medium, **THEN** Peak Torque **is $y = p_0^{(5)} + p_1^{(5)}x_1 + \cdots + p_n^{(5)}x_n$

where $x = (x_1, \ldots, x_n)^T$ is the input vector and $p_0, \ldots p_n$ are the linear parameters of the consequent part of the Sugeno model.

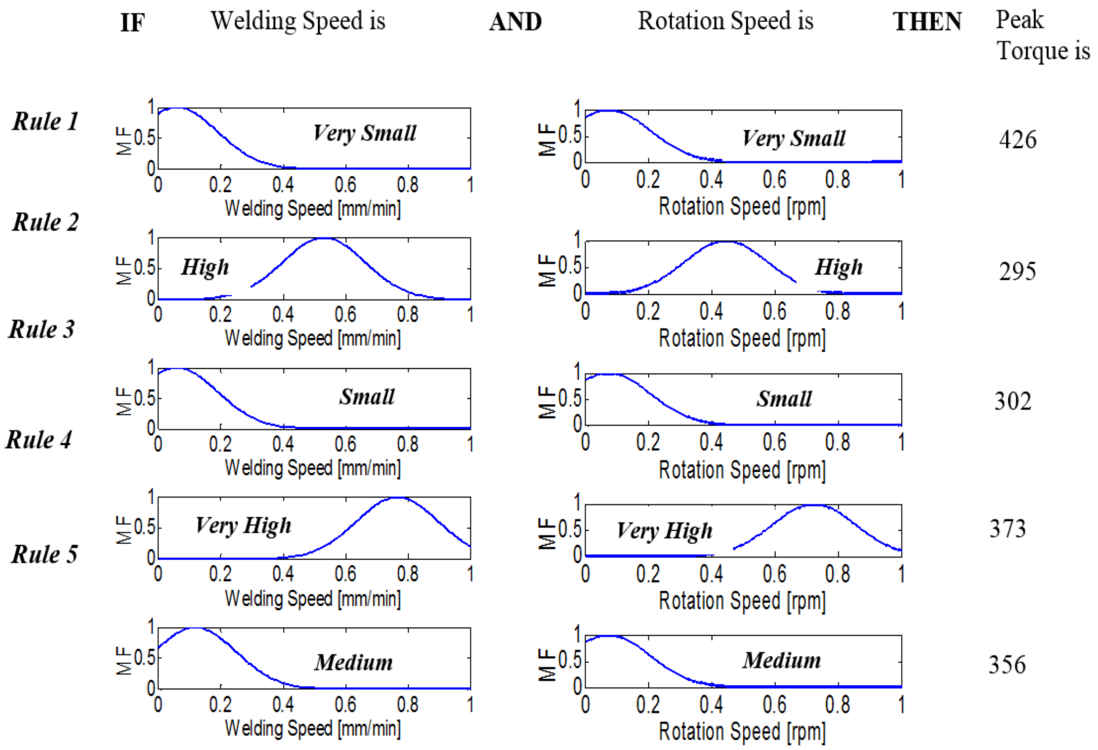


Figure 3.9. The initial rule-base of ANFIS model constructed by using subtractive clustering.

Due to the variability produced by the subtractive clustering, a systematic number of simulations were conducted to select the appropriate number of clusters (fuzzy rules) it was established that the optimum number of fuzzy rules is 5 for the prediction of spindle peak torque, hence through 10-fold cross-validation experimentation for training the initial rule-base structure of the model. The initial rule-base was optimised via back error propagation approach based on gradient descent procedure due to its efficiency in optimising the proposed type of neural fuzzy systems [199]. Table 3-1 and Fig. 3.10 show the effects of increasing the number of clusters (fuzzy rules) on the mean squared prediction accuracy for peak torque prediction. Meaning that the accuracy performance of the ANFIS model was compared by evaluating the root-mean-square error (RMSE) with its standard deviation (SD) and mean-absolute error (MAE%) between the current model output and desired output. The results obtained after a number of systematic simulations in the range between 3-17 clusters are summarised in Table 3-1. This table contains the information required to make a judgement on which model could have a good balance between prediction accuracy and interpretability of the rule-base. Fig. 3.11 illustrates the measured versus predicted initial model performance for spindle peak torque for the training and testing respectively. Visible in the simulation plots also 2 × standard deviation bands.

Table 3.1. RMSE and MAE% for the ANFIS modelling framework.

	RMSE±SD	**RMSE±SD**	**MAE%±SD**	**MAE% ±SD**
Number of Clusters	Training	Testing	Training	Testing
3	46.18±7.48	48.23±9.11	11.48±1.56	13.92±3.75
5	45.96±5.46	47.13±6.01	11.42±1.32	13.15±3.04
6	42.60±3.14	51.34±7.71	10.86±1.24	12.98±2.87
8	42.11±2.06	51.17±8.34	9.97±1.18	12.89±2.65
11	41.34±1.15	63.55±8.87	9.62±1.03	14.51±6.90
17	35.65±0.96	239.55±15.10	8.09±0.96	27.63±17.73

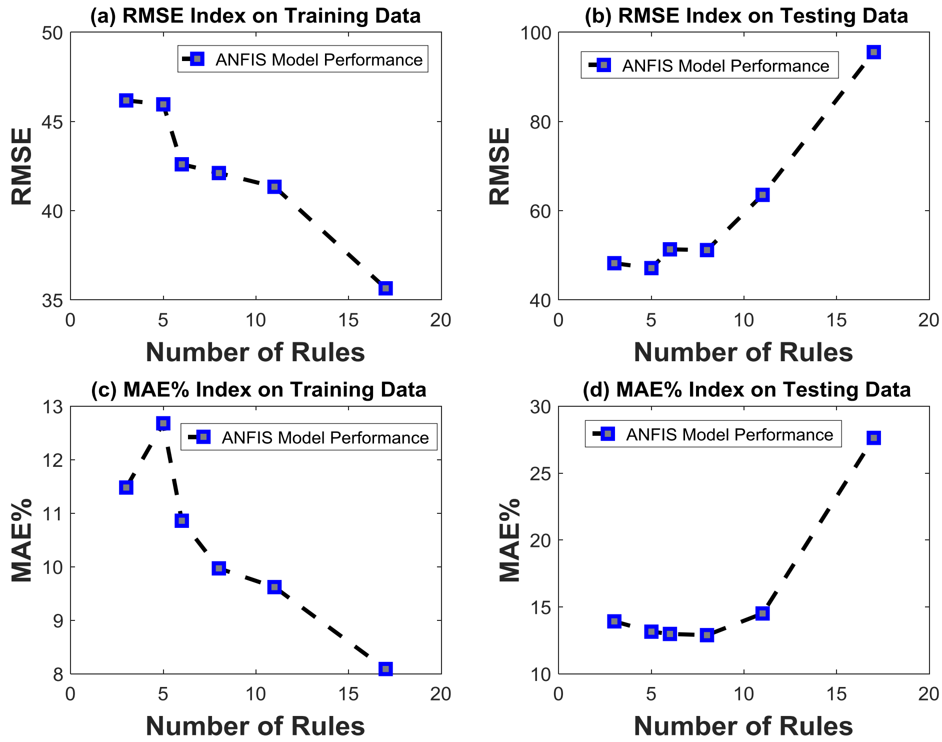


Figure 3.10. Comparison of performance indices based on the RMSE and MAE% for ANFIS modelling framework with different number of rules.

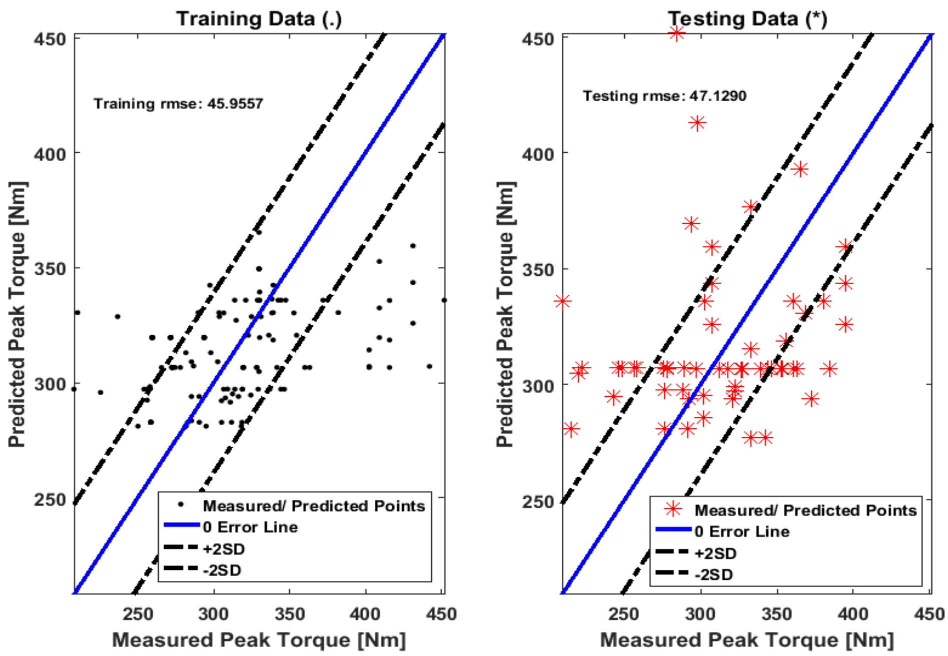


Figure 3.11. Data fit, peak torque prediction by using subtractive clustering to construct the initial fuzzy rule-base.

From the preliminary modelling results obtained from the ANFIS modelling framework, it is evident that the model was not able to provide good prediction accuracy and some data points were not correctly predicted. Despite k-fold cross-validation strategy was used to construct the final model, the model still suffers from over-fitting problem and there is a non-linear relationship between the process parameters and spindle peak torque. This may due the learning algorithm or complexity, sparsity and inconsistency associated with the process data.

3.5. SUMMARY

Despite considerable interests in the Friction Stir Welding (FSW) technology in past decades, the basic physical understanding of the process is still insufficient. Clearly, the complete understanding of the material flow around the rotating tool is crucial to optimise parameters of the FSW (including tool rotational speed, welding speed, spindle tilt angle, and penetration depth) and design of tool geometry. In this chapter, various topics related to the FSW process were presented including principle of operation, tool design, advantages and applications of the process, and weld quality assessment. A comprehensive literature review on FSW modelling (analytical, numerical and data-driven computational intelligence approaches) was also covered and the recent developments in the field of process monitoring and control were presented.

Preliminary modelling results based on the well-known ANFIS modelling framework were also described, including the subtractive clustering approach and application of back error propagation approach based on gradient descent procedure. Finally, some modelling results for the prediction of peak torque variable in the complex manufacturing process of FSW were discussed.

In the next chapter, a data-driven modelling framework based radial basis function neural networks (RBF-NNs), human-like information capture of granular computing (GrC) with an application of a conflict measure for the evaluation and analysis of uncertainty during the information granulation (IG) will be introduced.

CHAPTER 4 - INTERPRETABILITY MEASURES IN RBF-NF SYSTEMS USING ITERATIVE GRANULAR COMPUTING

This chapter presents a new conflict measure during iterative information capture in granular computing (GrC), in order to estimate/evaluate the uncertainty present during the iterative data granulation process. On one hand, the conflict measure is calculated via Shannon entropy theory to extract information related to the data uncertainty while carrying out the granulation process. Subsequently, this information is used to guide the iterative granulation process into condensing (merging) the granules (data) with low conflict, therefore producing better quality information granules. On the other hand, the resulting information granules are utilised to construct the initial parameters of a radial basis function (RBF) neural-fuzzy structure to model the steel friction stir welding process.

The predictive performance and interpretability of the proposed modelling framework is compared to the preliminary results obtained in Chapter 3 and results in recent literature.

4.1. INTRODUCTION

In systems engineering, one of the main objectives of fuzzy logic modelling is to develop interpretable and computationally efficient models. These may describe real systems or natural phenomena with nonlinear behaviour through the construction of a linguistic rule-base. The rule-base description of a given system can be achieved by establishing relationships between the relevant system inputs and outputs in the form of IF-THEN linguistic rules. Each fuzzy rule maps a fuzzy region from the antecedent space to another fuzzy region in the consequent space. The linguistic interpretability of fuzzy logic models can hardly be maintained when using ANNs [26] such as MLPs [24]. In fact, fuzzy models combine some distinctive properties that make them particularly interesting such as the facility to formulate the system knowledge based on transparent and interpretable linguistic rules, the capacity of integrating linguistic information from human experts with observational data, and the ability of performing universal function approximation for complex and nonlinear functions with simpler computational models [44].

In spite of the rapid development of hybrid modelling approaches based on FLSs [26], ANNs [56] and GAs [57] and their successful application to a variety of real-world problems, most existing fuzzy modelling approaches focus on model accuracy, rather than interpretability and simplicity of the obtained models, which are considered a primary advantage of fuzzy logic rule-based systems (FLRBSs). In many situations, users require the model not only to perform good global predictions but also to provide meaningful linguistic descriptions of the system behaviour. Such descriptions can be extracted and possibly combined with expert knowledge, not only to aid understanding the system, but also to validate the model obtained from process data via a parsimonious but understandable rule-base.

RBF-NN is one of the data-driven CI modelling paradigms that is often used for modelling of nonlinear and complex systems [22]. RBF-NN is a non-linear input/output mapping which is used for performing many tasks such as exact functions approximation, regularisation, noisy interpolation and pattern recognition of process data in a multi-dimensional space due to its ability to learn from data and improve its performance by adapting to the changes in the environment [23]. Although, the RBF-NN is a black-box model, it is a powerful modelling tool when it is combined with fuzzy logic (FL) by taking advantage of both Neural Networks (learning capability) and fuzzy logic (transparency and interpretability) [24].

Several studies have established the mathematical equivalence between the RBF-NN called T1-RBF-NF system and a class of FISs [20, 21] and thus the RBF-NN can be considered as a fuzzy inference model of type-1 under certain conditions. That means, the initial structure of the RBF-NN identification can be achieved similarly to that employed in FLSs [25, 26]. In other words, the RBF-NN parameters which represent the consequent and premise parameters in fuzzy systems are estimated systematically from observational data by using a clustering procedure and then the parameters are adjusted more precisely to complete the modelling process. In FLSs theory [200, 201], parsimony is a very important feature as it is closely related to the interpretability of the fuzzy model as a result of a good distinguishable fuzzy rule-base that determines the degree of transparency in the FIS. In order to construct parsimonious FLSs, it is required to generate distinguishable fuzzy sets and at the same time preserve global model accuracy. The distinguishability of fuzzy sets is considered as one of the most important aspects for the interpretability of FL systems. By the way of comparison between RBF-NN and FL systems, the former often suffers from loss of interpretability due to the parametric optimisation process, which is usually performed via the use of a gradient-based procedure [26].

To produce distinguishable fuzzy sets, in [200, 201] the authors suggested a procedure to design interpretable fuzzy systems based on a collection of interpretability constraints, i.e., a group of formal properties that are imposed on the components of fuzzy models (fuzzy sets, rules, etc.) to prevent unintelligible configurations. For example, a metric similarity measure to quantify distinguishability as a result of the degree of overlapping between two or more FSs and this measure is used to evaluate how much is the interpretability of a fuzzy system influenced by the degree of overlap between fuzzy sets. Meaning that, high degree of overlap between the fuzzy sets results in less distinguishable fuzzy sets and therefore less interpretable rule-base. In [46] Zhou and Gan divided the interpretability into two categories based on the two learning phases (i.e. initial structure identification phase and parametric optimisation phase) that involves in the design of FLSs. In a deeper context, the authors described the interpretability during the initial structure identification phase and it was dubbed low-level interpretability; while the interpretability during the parametric optimisation phase was denoted as high-level interpretability.

The low-level of interpretability of fuzzy models is the interpretability that can be obtained with regard to semantic criteria on fuzzy sets by optimising the membership functions (MFs). Whilst the high-level of interpretability is the fuzzy model interpretability that associated with the parametric optimisation phase and it can be obtained when dealing with the consistency, completeness, and coverage of the fuzzy rules with reference to the criteria on fuzzy rules. The transparency, parsimony, and readability as well as simplicity of the fuzzy rules are among the criteria that can also be taken into consideration at the high-level interpretability.

A prevalent approach to design the RBF-NN is to first select the number and locations of centres (prototypes) in the hidden layer via an unsupervised learning procedure to discover a structure in the input training data [202]. In particular, popular clustering algorithms to partition (cluster) the input space can be used – such as the hard C-means clustering or k-means clustering algorithm [203, 204], the Fuzzy C-means (FCM) clustering [197], the subtractive clustering [198] and recently the iterative data granulation algorithm of granular computing (GrC) [27]. Almost all the existing clustering algorithms that appeared in the literature operate on numeric data and produce prototypes (i.e., data representatives) that are again completely numeric [30]. In this sense, the form of these prototypes does not reflect the number of data points they represent and how the initial data distribution looks like [30].

Particularly the emerging computing paradigms of granular computing (GrC) have been used for processing of information in forms of 'information granules' in a transparent and interpretable way [135]. The information granules are used to estimate the initial parameters of the RBF-NN at the low level interpretability. In the field of GrC, granulation is a computational paradigm that is different from the aforementioned data clustering algorithms in its ability to imitate the human ability in terms of grouping/arranging similar objects together based on predefined compatibility measures. In the field of human-centric intelligent systems (HCISs), GrC is an emerging computing framework of human-like information processing that deals with processing and representing information in form of information granules. The process of forming these information granules is called iterative information granulation [205]. Each information granule consists of complex information entities that are grouped/arranged together during the information granulation due to their similarity (shape, orientation), indistinguishability, functional adjacency (distance, size, volume), coherency, cardinality, and density.

In data-driven computational intelligence models, the process of abstraction of raw data, knowledge acquisition, identifying valid, and potentially meaningful information in numeric data is a nontrivial step towards the development of efficient HCIS since it provides to CI models with the ability of acquiring knowledge that can be understood/communicated to users. The acquired knowledge is also used to estimate the initial parameters (structure) of the RBF-NN model. However, most of the real-world data that are available often associated with uncertainty because of measurement imprecision, inaccuracy, scatter, scarcity, high non-linearity, limited data size, low process repeatability, or other errors. The information explosion/extraction in such data has raised an urgent demand for the development of knowledge discovery approaches that have the ability to handle such challenges.

A number of statistical and machine learning approaches (i.e., CI approaches) that depend on data pre-processing techniques aim at representing uncertainties such as those involve noise reduction and outliers detection as well as errors elimination [206, 207]. In some cases, outliers and noise are produced as a consequence of the process itself and they contain useful information that represent part of the process dynamics. These information can be employed in the modelling process. In addition, noise reduction or outliers detection (removing some data points) is not always feasible/desirable in practice [27].

The iterative human-like computational framework of GrC in the form proposed in [27] represents an explanatory and effective data analysis tool and a useful data clustering technique. It has also demonstrated its efficiency as a tool for constructing the initial structure of the RBF-NF system. Even through, it has proven its efficiency, during the iterative granulation adding/inserting a new information granule to the input space in particular in the case where two or more information possess a similar compatibility measure results in a degree of uncertainty about which information granules to be merged. This phenomenon produces a grade of conflict uncertainty among the new information granules and accordingly the final information granules do not represent the accurate distribution of the input space of the process under study (i.e., producing low quality information granules).

Since the acquired knowledge (in the form of granulated data) from the iterative information granulation can be utilised to estimate the initial parameters of the RBF-NF system (fuzzy logic rule-base), however generating information granules with high level of uncertainty will result in FL rule-base with high degree of overlap between FSs and less distinguishable FSs (i.e., low quality FL rule-base). Less distinguishability leads to loss of transparency and then the interpretability of rule-base might be lost.

According to [29] [31, 208-210], a number of information uncertainties may present during the iterative granulation process, for instance uncertainty caused due to lack of boundaries between the information granules (fuzziness), conflict between granules, incompleteness , etc. In this sense, uncertainty can broadly be categorised into three basic types: a) fuzziness (shapelessness or vagueness), b) non-specificity, and c) discord (dissonance or conflict). Researchers from different fields have proposed mathematical frameworks to model/deal with uncertain data, for instance fuzzy sets theory [8], possibility theory [14, 15], evidence theory [9], the theory of fuzzy measures [16] and rough sets theory [12].

This chapter addresses the issue of uncertainty present during the iterative IG process where a new measure that is able to capture the uncertainty caused by conflict is proposed. In this context, a Shannon-based uncertainty measure is introduced to measure the conflict during the iterative information granulation and use it as a guide to enhance the IG process. This measure provides an extra degree of freedom, which makes the compatibility creation able to drive the compatibility search into merging the information granules with low conflict. Thus, better quality information granules, more transparent, interpretable and consistent FL rule-base is obtained.

The main contributions of this chapter are twofold, on the one hand, the information granulation process is utilised to estimate the initial location of the centres (prototype) in the field-receptive units in the hidden layer of the RBF-NF model. On the other hand, a new modelling framework based on the granulation process presented in [27] and the Shannon entropy-based conflict measure [209, 211] is proposed in order to estimate the uncertainty/conflict between the granules during the granulation process to improve the quality of the information granules at low-level interpretability (i.e., initial structure identification of a RBF-NN). Therefore, the main objectives of this chapter can best be treated under two headings:

- The two levels of interpretability (low-level and high-level of interpretability) during the development of RBF-NF system are described in full detail. A particular focus will be put on the low-level interpretability process of information granulation for estimating the initial RBF-NF system parameters (initial FL rule-base).

- A systematic data-driven neural-fuzzy (NF) modelling framework based on the iterative human-like information capture of granular computing (GrC) and RBF-NF system is proposed. A new uncertainty measure that takes into account the uncertainty present during the iterative IG process is introduced. This uncertainty measure is used to evaluate/quantify the degree of conflict among information granules during the iterative IG process. The initial structure of the RBF-NF system is optimised via the use of the adaptive back-error propagation (BEP) to improve its performance.

4.2. INTERPRETABILITY MEASURES IN THE RBF-NF SYSTEM STRUCTURE

According to [20], T1-FLSs and RBF-NNs are functionally equivalent under some moderate conditions. Thereby, interpretability property from fuzzy logic systems and learning property of ANNs can be exploited and explored from a unified modelling framework (known as neural fuzzy systems). That implies the RBF-NN may be easily translated into human language (rule-based systems) in the fuzzy logic system and vice versa. Nevertheless, this interpretability is often ignored and most of the researches have been devoted to approximation capabilities of neural-fuzzy systems.

When neural fuzzy systems are developed from process data to approximate nonlinear functions, learning methods are used to optimise the FLSs and improve their approximation accuracy. However, most learning methods are set-up for accuracy, which often causes a lack of interpretability in the constructed rule-base. Over the past years, several methods have been proposed to improve the interpretability of FL systems [212-215]. In [216], the authors discuss the difference between the RBF-NN and interpretable T1-FLSs. The authors also proposed some conditions that should be fulfilled for rule-based systems to be considered interpretable:

- *Completeness and distinguishability of rule-based systems:* Fuzzy rules that are generated based on clustering of all the input variables in the rule-base should be complete and well distinguishable. Fuzzy partitioning corresponds to initial structure identification for rule-base. Each fuzzy partition represents a linguistic variable in the rule-base to facilitate the easy interpretation of complex systems. The completeness of rule-based systems means that for each input variable, at least one fuzzy partition (fuzzy set) is fired.

- *Consistency of the rule-based systems:* Rule-base consistency means that there is no any contradictory rules in the rule-based systems in the sense that rules with similar premise/antecedent parts should have similar conclusion/consequent parts. For instance, if any two antecedents in a rule-base are similar but produce a completely different consequent, as a result, there is an inconsistency in the rule-based system.
- *The number of features or variables in the antecedent parts and the number of fuzzy rules in the rule-based systems:* The number of variables in the antecedent parts and the number of fuzzy rules in the rule-base both should be as small as possible.

For instance in [46], Zhou and Gan proposed a taxonomy of interpretability for fuzzy linguistic rule-based systems in terms of low-level interpretability and high-level interpretability. On the one hand, the authors in [46] proposed several semantic criteria (semantics-based interpretability) that can be applied on the fuzzy set level by optimising membership functions (MFs) to achieve low-level interpretability. The low-level interpretability improvement relies upon the modification of the MFs by imposing some constraints towards better distinguishability, a moderate number of MFs, better coverage or completeness of the fuzzy partitions in the input space, as well as complimentary and normalisation. On the other hand, high-level interpretability of the linguistic fuzzy rule-based systems is achieved on the fuzzy rule level by carrying out overall complexity reduction in terms of complexity-based interpretability criteria such as a moderate number fuzzy rules and features (variables). In addition, high-level interpretability can be achieved by imposing semantic-based interpretability constraints on the fuzzy rules such as consistency of the rule-base, parsimony and its transparency of rule-base structure, fuzzy rules fired at the same time, and cointension. However, the taxonomy presented in [46] is only used to analyse the interpretability of linguistic rule-based systems.

According to [19], when dealing with the trade-off of interpretability-accuracy of the obtained model, two main requirements may be considered:

1. *Linguistic fuzzy modelling (LFM):* This field of study is mainly devoted to build fuzzy models with good interpretability through the use of linguistic fuzzy rule-based systems. Such fuzzy systems are heavily based on simple linguistic interpretable rules (or Mamdani) extracted from raw data/information in order to describe complex systems. The antecedent and consequent parts of the Mamdani-type fuzzy logic systems make the use of linguistic variables with their linguistic terms (meanings) and the corresponding MFs.
2. *Precise fuzzy modelling (PFM):* This field of study is mainly focused on the construction of fuzzy models with good accuracy through the use of Takagi-Sugeno FRBSs. A Takagi-Sugeno FRBS is different from Mamdani-type in the way that the consequent parts are deterministic without an associated meaning.

Since the T1-FLSs and RBF-NNs are mathematically equivalent under some moderate conditions, a RBF-NN can be regarded as a FLS of type-1. The development of a type-1 FLS generally involves two learning stages, the structure identification stage (initial rule-base) stage and the parametric optimisation stage. These two stages are often carried out sequentially; the structure identification stage is used to build the initial structure of rule-base and then the parametric optimisation stage is used to tune the parameters of each linguistic fuzzy rule. Thus, the taxonomy of interpretability in the RBF-NN can also be categorised into two different levels namely a) low-level interpretability during the structure identification of the RBF-NN via the use of clustering algorithm, and b) high-level interpretability during the parametric optimisation stage of the initial rule-base (i.e., initial parameters of the RBF-NN) via the application of a gradient descent approach. Fig. 4.1 illustrates the two levels of interpretability at the RBF-NN modelling framework.

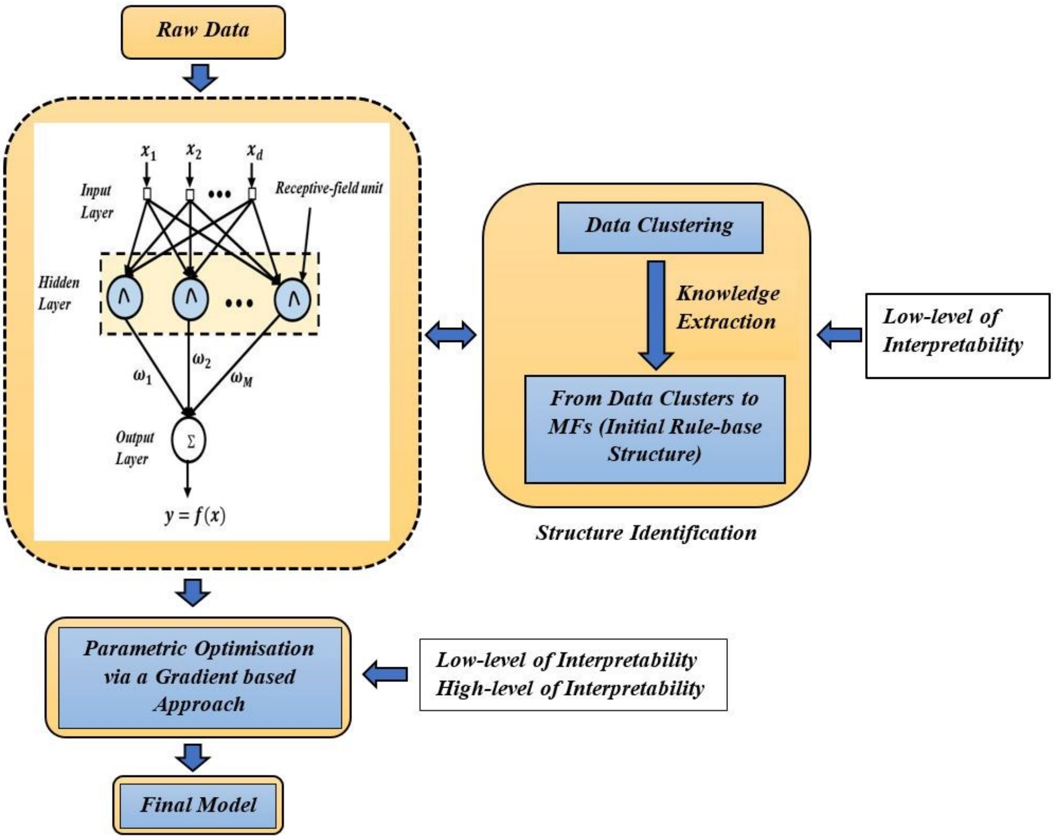


Figure 4.1.Interpretability levels for the RBF-NF system.

As pointed out in [46], the semantic criteria that may be considered at the low-level interpretability of the RBF-NN include:

1. *Distinguishability of the MFs:* Distinguishability is an essential and basic criterion for obtaining interpretable rule-base, because, in each input space partitioning distinctive boundaries between the fuzzy sets should be clearly defined in the universe of discourse of a variable. Each MF corresponds to a linguistic variable with a clear semantic meaning therefore it becomes hard to assign distinct linguistic variables (labels) to indistinguishable fuzzy sets. According to [217], a clustering algorithm is required to estimate parameters of the radial basis function (RBF) (locations of the centres). The initial parameters of the RBF are then used to construct the initial rule-base. The initial rule-base is optimised via the use of a gradient-based approach to construct the final rule-base system. For this reason, the

initial locations of the RBF centres and their associated distinguishability play an important role in the construction of the final rule-based systems.

2. *Normalisation of the MFs:* In the universe of discourse at least one data point should have a MF equal to one and all the MFs in the universe of discourse of a variable should be normal.
3. *Moderate number of MFs:* The number of MFs (linguistic labels) for each input variable in the rule-base should be a small as possible while preserving a good level of accuracy. A smaller number of fuzzy rules (number of RBFs in the hidden unit of the RBF-NN) allows us to better understand the semantic meaning of each MF.
4. *Completeness or coverage of fuzzy partitioning:* The entire universe of discourse of a variable should be partitioned in a way that all the MFs generated should cover the entire universe of discourse and every data point in universe of discourse should belong to at least one of the fuzzy partitions (fuzzy sets) and should be linguistically represented by a fuzzy set.
5. *Complementary:* For each element in the universe of discourse, the sum of all its corresponding MFs should be close or equal to unity. This guarantees the meanings among the all elements are uniformly distributed.

To achieve interpretable MFs (low-level interpretability) [46] in the linguistic fuzzy rule-based systems in terms of semantic criteria on input space fuzzy partitioning, several constructive techniques have been applied, such constructive techniques include: (1) applying regularisation techniques in the objective function for estimating parameters of the MFs; (2) multi-objective optimisation techniques for estimating the antecedent parameters of the linguistic FRBSs by taking into consideration the good trade-off between low-level interpretability and global model performance; (3) Fuzzy set merging approaches and (4) user-oriented interactive approach.

In a similar fashion, high-level interpretability of the RBF-NN is achieved on the optimised RBF-NN structure (i.e., fuzzy rule level) by carrying out overall complexity reduction and taking into consideration the following factors:

a) *Consistency of the rule-base:* The degree of consistency in the RBF-NN is measured by the absence of contradictory rules in the rule-base, i.e., fuzzy rules with similar antecedent parts should have similar consequent parts.
b) *Parsimony and simplicity of rule-base structure:* According to [46], the best model is the simplest in its structure and the one that fits the system behaviours well – meaning that the number of fuzzy rules in the rule-base must be as small as possible while preserving a satisfactory level of model performance. A larger number of fuzzy rules would result in a lack of global interpretability of the system. For high dimensional systems, a parsimonious linguistic FRBS is very desirable.
c) *Transparency of rule-base structure:* According to [43], transparency is a measure of the ability of extracted fuzzy rules from training the RBF-NN to characterise human knowledge in an interpretable way (i.e., transformation of process data (information) into (human) knowledge) in order to permit a deeper understanding of the system under study.
d) *Completeness:* For any input vector to the RBF-NN, at least one fuzzy rule should be fired in order to prevent the FLS from breaking the inference. Due to the nature of the RBF-NN, an input space fuzzy partition with very low activation may deteriorate the overall interpretability of the model. Therefore, a tolerance threshold is required to be taken into consideration for rule-activation during FSs and rule generation to avoid activating the rule-base at very low level.

A detailed overview of the existing interpretability measures and methods for achieving more interpretable FLS is presented in [19]. Therefore, this study aims to provide a computational framework in order to deal with two contradictory requirements in the development of RBF-NN (interpretability and accuracy).

4.3. ITERATIVE DATA GRANULATION

In the field of knowledge discovery in databases and data mining, data clustering can be broadly split into two main categories namely; a) partitioning clustering [218] and b) hierarchical clustering [219]. The former groups data based on optimising a certain criterion function and the number of clusters has to be determined beforehand.

One of the disadvantages of partitioning clustering is that it is sensitive to the initial locations of the clusters centres and sensitive to noise and outliers [116]. The latter groups similar objects in hierarchical structure based on some measures of similarity or distance. In addition, the number of clusters does not need to be specified and outliers can be easily identified. Moreover, it is static and not sensitive to the initial locations of the centres [220]. In this context, the IG process aims at grouping data points with similar features and characteristics. To achieve the iterative IG, a compatibility measure based on the granular similarity that calculates 'compatibility index' is usually employed.

Even through, the term 'granule' was initially introduced by Zadeh in [130] into the field of fuzzy logic theory under the term information granulation, the term 'Granular of Computing' (GrC) was first proposed by Zadeh in [64, 65] and Lin in [66]. GrC is a unified and coherent computational paradigm that imitates human cognition in terms of constructing, describing and grouping similar objects together according to their features and characteristics [30, 221]. GrC aims at processing of complex information entities – information granules – that are formed by condensing data at numeric level and then extracting meaningful knowledge. Information granulation plays a significant role in extracting previously unknown, implicit and meaningful knowledge out of data [222, 223]. The use of information granulation in this research work is motivated from the ability to represent the information granules in form of hyper-boxes located in a high dimensional data space [67, 129].

Several methodologies, techniques, theories, and tools geared towards the processing of information granules have been proposed to solve complex problems [129]. Set theory and interval analysis, fuzzy sets, rough sets, shadowed sets are some of the existing formal frameworks in which information entities (granules) are constructed [30]. In [67], Pedrycz proposed a mathematical formalism based on interval analysis for the analysis of information density of the granular structures. These granular structures arise as a result of the granulation process. The proposed framework represents a clustering algorithm that operates on a numeric raw data and granulates through it. The proposed methodology aims at capturing the information through the process of organising and abstracting data based on their similarities and characteristics in the form of granules. According to Pedrycz [67], a high level of data abstraction can be achieved via the use of a clustering approach based on granulation through a process of condensing the original data into granules. The granular clustering approach as described in [30, 67] is a dual-stage iterative algorithm that can be carried out as follows:

- Find the most two closet (compatible) information granules according to a predefined compatibility measure and merge them together to build a new information granule. The new information granule contains both original information granules. Thus, the size of the data set is reduced and the same time the clustering algorithm condenses data.
- Repeat the first stage of finding the most two closest information granules and building a new information granule until a satisfactory data condensation is achieved or a predefined criterion is met.

The compatibility criterion between any two information granules can be calculated according to their similarity, functionality, coherency, compactness, or indistinguishability. To illustrate the concept of compatibility measure, consider two information granules (hyper-boxes) A and B as depicted in Fig. 4.2.

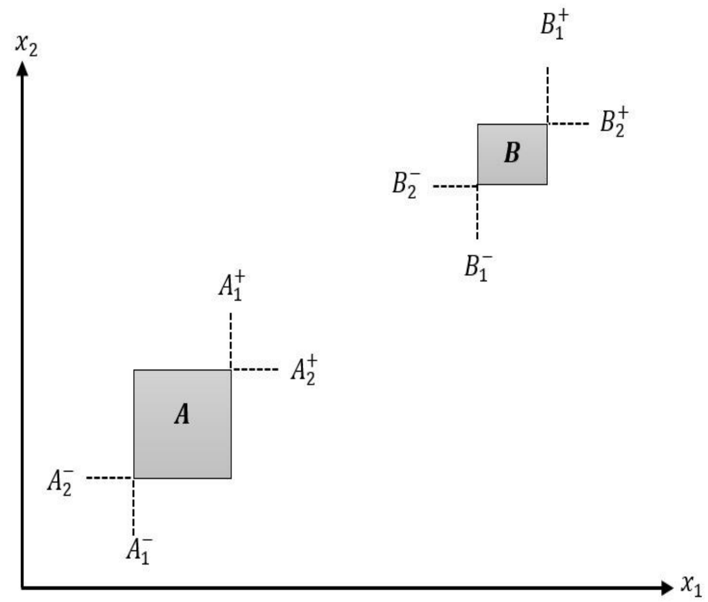


Figure 4.2. Two information granules A and B in a 2-dimensional space.

To point at the locations of the information granules in the space, in this studyA^-, A^+denote to the lower and upper bound of granule A respectively while B^-, B^+denote to the lower and upper bound of granule B respectively. The compatibility between A and B, $compat_{(A,B)}$ can be expressed in two components: a) the distance between granule A and granule B, $d_{(A,B)}$; b) the size of the newly formed information granules that resulted as a consequence of merging A and B. Fig. 4.3 illustrates some geometric properties of a resulting information granules C formed as a result of merging two compatible information granules A and B.

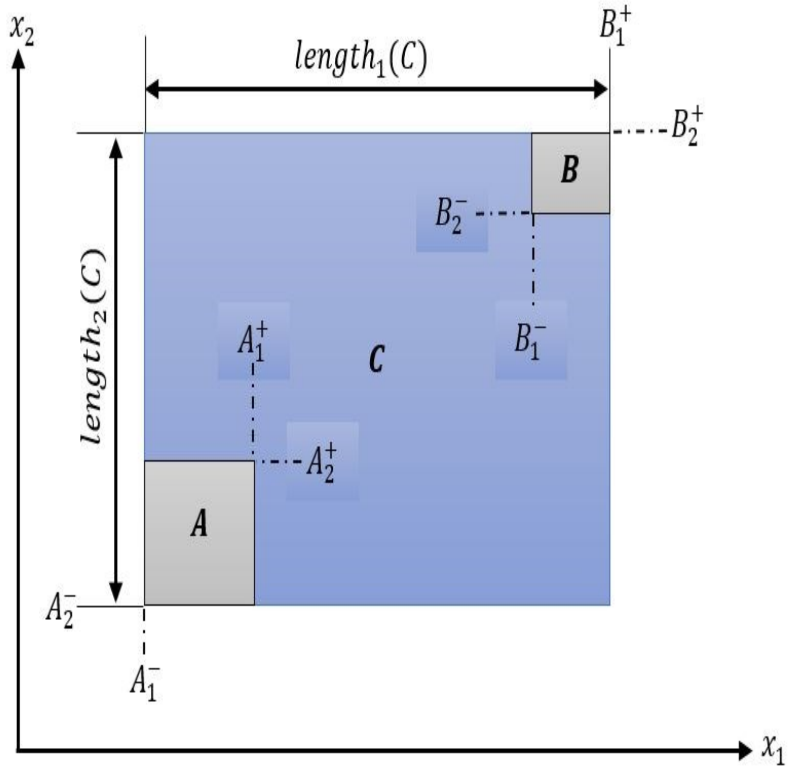


Figure 4.3. Information granule C formed as result of merging A and B.

Therefore, the compatibility criterion can be calculated in the form

$$compat_{(A,B)} = 1 - d_{(A,B)} e^{-\alpha V(C)} \quad 4\text{-}1$$

where

$$d_{(A,B)} = \frac{\|A^- - B^-\| + \|A^+ - B^+\|}{L_p} \quad 4\text{-}2$$

with ‖.‖ is an L_p distance between two objects (i.e., distance between two numeric vectors in this case), $p > 1$. The value of p can be varied through the spectrum of well-known distances (Hamming distance ($p = 1$), Euclidean distance ($p = 2$), etc).

$$V(C) = volume(C) = \prod_{i=1}^{n} length_i(C) \tag{4-3}$$

$$length_i(C) = max(A_i^+, B_i^+) - min(A_i^-, B_i^-) \tag{4-4}$$

Note that the compatibility measure in Eq. 4-1 is not only based on merging the closest information granules but also the resulting information granule should be compact. Compactness means the size of the information granule in every dimension is nearly equal. The maximum value of the computability measure ($compat_{(A,B)} = 1$) can be attained when the volume of the resulting granule $V(C)$ is reduced to zero. Therefore, the compactness factor $e^{-\alpha V(C)}$ guarantees that only dense and compact information granules are being formed. However, to calculate the compatibility measure in Eq. 4-1, the original data set has to be normalised in the interval [0, 1].

The definition of compatibility criterion can vary between authors. It can be purely based on geometry of granules (size, volume, or distance), similarity between granules (shape, orientation), or density (ratio between cardinality and volume). In [27], Panoutsos and Mahfouf extended the compatibility measure that was introduced by Pedyrcz [67] to a compatibility measure that instead of using the volume of the resulting granule it is defined as a function of distance between two information granules and the density of the resulting granule formed from merging of two information granules. The density of the resulting granule can be expressed as a ratio between cardinality of each granule and the length of the resulting information granule in multi-dimension. Therefore, in this chapter the compatibility measure that is used during the granulation process is defined in the following set of equations:

$$compat_{\,(A,B)} = d_{MAX} - d_{(A,B)}.compactness\ factor \tag{4-5}$$

where

d_{MAX} is the maximum distance in the input data set; and $d_{(A,B)}$ is weighted the average distance between two information granules A and B calculated in multi-dimension; such as:

$$compactness\ factor = e^{-\alpha R} \qquad 4\text{-}6$$

$$R = \left(\frac{Cardinality_{(A,B)}/Cardinality_{MAX}}{Length_{(A,B)}/Length_{MAX}}\right) \qquad 4\text{-}7$$

and

$$d_{(A,B)} = \frac{\sum_{i=1}^{d} \omega_i (max(A_i^+ - B_i^+) - min(A_i^- - B_i^-))}{d} \qquad 4\text{-}8$$

With A_i^-, A_i^+ are the lower and upper boundaries for information granule A in dimension i; and ω_i is the dimensional weighting importance factor for dimension i and d is the total number of dimensions; α the weight in the interval [0, 1] that is used in order to balance the requirements between distance and compactness; $Cardinality_{MAX}$ represents the total number of granules in the input data set; the term $Cardinality_{(A,B)}$ is the number of granule of the resulting information granule; $Length_{MAX}$ is the maximum length of an information granule in the input data set; $Length_{(A,B)}$ is the length of the resulting information granule in multi-dimension; such that:

$$Length_{(A,B)} = \sum_{i=1}^{d} (max(A_i^+ - B_i^+) - min(A_i^- - B_i^-)) \qquad 4\text{-}9$$

The extended version of the compatibility measure proposed by Panoutsos in [27] geared towards merging of compact granules with a high cardinality. In addition, to calculate the compatibility measure in Eq. 4-5, there is no need to normalise the original data set, since the reference is the maximum distance in the raw input data d_{MAX}.

To illustrate the concept of the compatibility measure during data granulation process, an artificial 2-dimensional raw data is considered containing of 200 data set [195], as shown in Fig. 4.4. As the iterative IG process advances coarser information granules are created as illustrated in Fig. 4.4. The top snapshot of the figure shows the original pre-granulated raw data consisting of 200 data points. As the information granulation evolves 50 information granules are created as shown in the middle snapshot and finally 5 information granules are created.

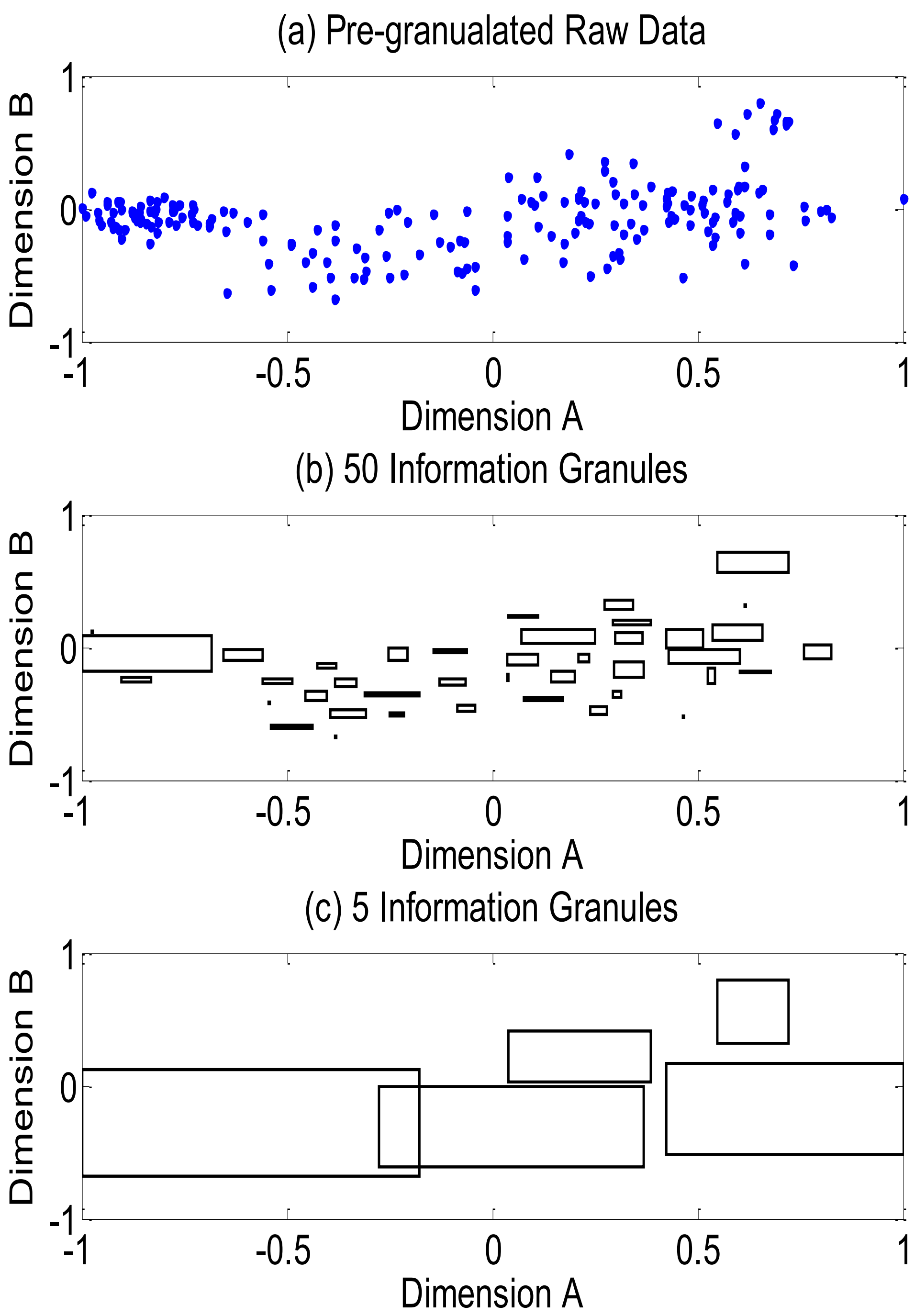


Figure 4.4 Iterative data granulation process – two dimensional data example.

The evolution of compatibility measure during the data granulation cycle is shown in Fig. 4.5, where the weighting importance factor $\alpha = 0.5$. It can be noted that from the evolution curve at the beginning of the granulation algorithm, the gradient of the compatibility criterion is relatively small which implies that highly compatible information granules are being merged. By way of contrast, at some point in the data granulation cycle (i.e., at the later stage), the gradient of the compatibility criterion becomes larger which indicates that highly incompatible granules are being merged. When the compatibility measure falls under a certain threshold, it can be concluded that the information granules are no longer more compatible. This effect can be used as a criterion to terminate the granulation process and to establish the minimum or optimal number of granules required for capturing the dynamics of a particular process. This can be used as a termination criterion for the granulation process.

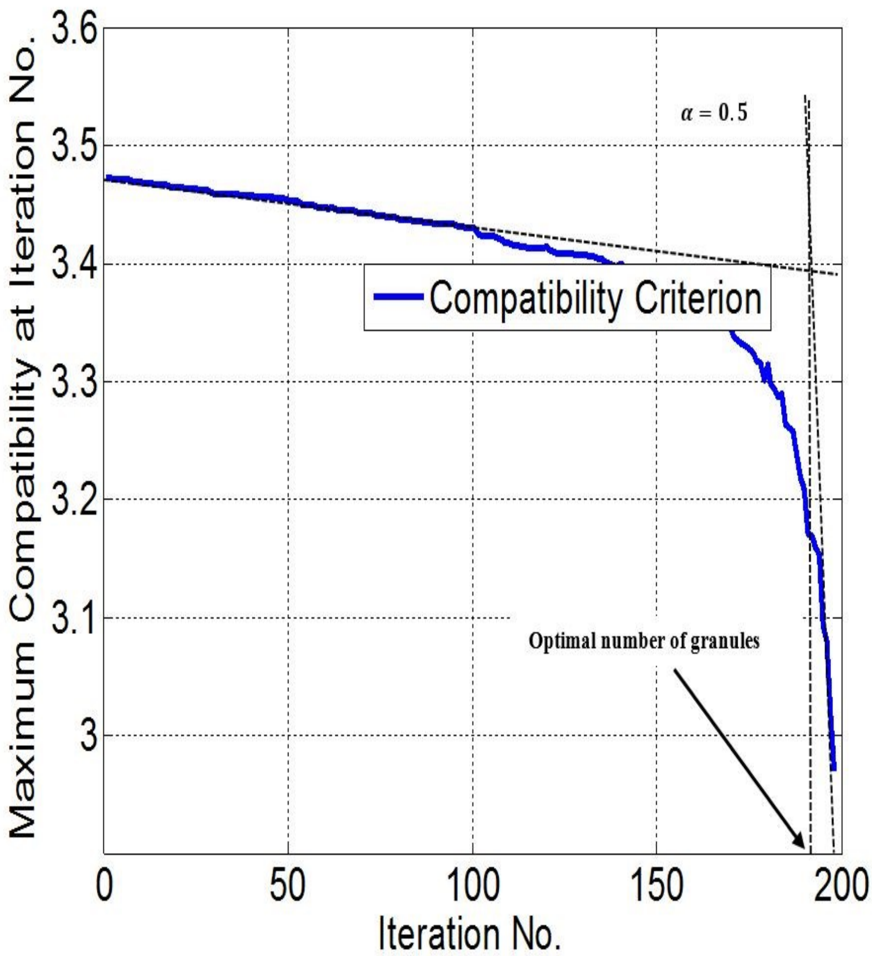


Figure 4.5 Compatibility measure example.

As it is pointed out in [67], the optimal number of information granules that represent the dynamics of a particular process can be obtained from the intersection of two gradient lines as shown in Fig. 4.5.

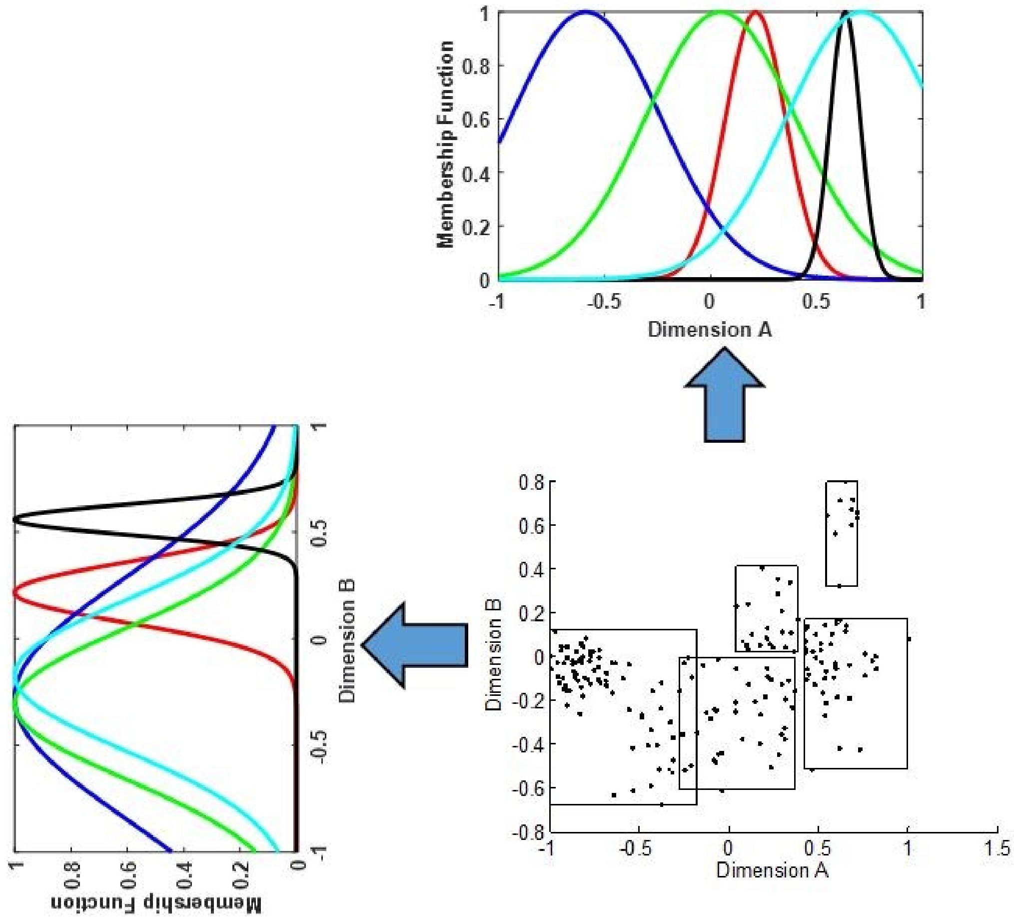


Figure 4.6. Initial fuzzy model creation.

According to Fig. 4.6 such final information granules from Fig. 4.4 are employed to elicit the initial multidimensional fuzzy rule base. The geometrical lower and upper boundaries of the final information granules are employed to estimate the initial parameters of the RBF-NF system.

The average hyper-box limits of each information granule can be used to calculate the initial centres C_i^j as follows:

$$C_i^j = [C_1^1, \dots, C_i^j, C_K^M] \qquad 4\text{-}10$$

where is the ith input and jth hidden unit. K is the total number of input data points.

with

$$C_i^j = \frac{1}{2}(max(A_i^+ - B_i^+) - min(A_i^- - B_i^-)) \qquad 4\text{-}11$$

$$C_i^j = \frac{1}{2}(max_{Xi} - min_{Xi}) \qquad 4\text{-}12$$

where min_{Xi} and max_{Xi} are the minimum and maximum multidimensional length of the resulting granule at dimension i.

The width parameter of the Gaussian membership function in the RBF-NF system is be calculated via the following equation [116]:

$$\sigma_i^j = \frac{d_{max}}{\sqrt{2M}} \qquad 4\text{-}13$$

where M is the number of final information granules (centres) and d_{max} is the maximum distance between the two centres of information granules of interest.

4.4. A HUMAN-CENTRIC APPROACH FOR MODELLING OF COMPLEX MANUFACTURING SYSTEMS

This section presents the results obtained by using the human-like iterative granulation process described above for the granulation of input raw data in order to estimate the initial structure of the fuzzy rule base, which is then tuned/optimised via the use of an adaptive back-error propagation (BEP) algorithm that is fully described in Section 4.4.3.

4.4.1. NEURAL-FUZZY GENERAL ARCHITECTURE

The raw input data are compressed (clustered) across each dimension to capture a particular process dynamics. The granulated data (final condensed information granules) are employed to construct the initial structure of fuzzy rule-base. As it is suggested in Chapter 2, the neural-fuzzy model that is considered in this research work is a 3-Layer RBF-NN and has the structure shown in Fig. 2.8 (in Chapter 2) [26].

Fig. 2.8 represents a multi-input and single output FLS described by Eq.4-14. The FLS with a single consequent part of a Mamdani-type fuzzy logic system and Gaussian membership functions (MFs) in the antecedent and consequent parts. The product operation is used in the inference engine and output is mapped via centre of gravity defuzzification method [26]. The input/output relationship of the FLS can be expressed in the following form:

$$y(x) = \sum_{j=1}^{M} \omega_j (\prod_{i=1}^{n} \mu_{A_i^j}(x_i)) \Big/ \sum_{j=1}^{M} (\prod_{i=1}^{n} \mu_{A_i^j}(x_i)) \qquad 4\text{-}14$$

with $\mu_{A_i^j}(x_i)$ is a membership function of a Gaussian type of x_i that belongs to the jth linguistic rule.

$$\mu_{A_i^j}(x_i) = e^{-\left[\frac{x_i - C_i^j}{\sigma_i^j}\right]^2} \qquad 4\text{-}15$$

$i = 1, \ldots, n$ and $j = 1, \ldots, M$.

where σ_i^j and C_i^j are the width and centre parameters of each fuzzy MF respectively, M is number of fuzzy rules, and n is the number of inputs.

4.4.2. INITIAL RULE-BASE CREATION BASED ON ITERATIVE DATA GRANULATION

As is described above, the IG is an iterative process that calculates the compatibility measure between data points (granules) and finds the two data points (granules) with the highest compatibility measure at each iteration, and then merge them together geometrically to form a new information granule that contains both the original granules until a predefined number of information granules is achieved or a specified termination criterion is met.

The relationship between an information granule in multi-dimension and a linguistic fuzzy rule is one-to-one relationship as shown in Fig. 4.6. The initial structure of the FIS can be expressed as a collection of linguistic fuzzy rules (granules) in the following IF-THEN rules:

$$\boldsymbol{Rule_i}: \boldsymbol{IF}\ x_1\ is\ A_1^i\ and, \dots, and\ x_d\ is\ A_d^i, \boldsymbol{THEN}\ y\ is\ B^i \tag{4-16}$$

where $x_1, \dots, x_d$, are the input vectors, $A_1^i, \dots,\ A_d^i$ are the fuzzy sets, $i = 1, \dots, M$, M is the number of rules, i is the index of the rules.

4.4.3. PARAMETRIC STRUCTURE OPTIMISATION

After the initial structure of the model is constructed from the iterative data granulation process and its initial parameters are estimated, a number of training algorithms can be utilised to parametrically tune the model and optimise its structure. Among them is the adaptive back-error propagation (adaptive-BEP) learning algorithm which is proven in the past to be very efficient in optimising similar granular based models [26]. The conventional BEP learning algorithm requires large number of iterations, each iteration involves a large amount of computation to converge and also often gets stuck into a local minima [123]. To circumvent this issue, the adaptive version of BEP is utilised, where the learning rate and momentum factor of the learning algorithm are adjusted during the optimisation process in order to enhance its convergence properties. Hence, a performance index PI can be defined as follows:

$$PI = \frac{1}{K}\sum_{k=1}^{K} e_k^2 \tag{4-17}$$

$$e_k = \left(y_k - y_k^d\right) \tag{4-18}$$

The update rule for the centre estimation:

$$\Delta C_i^j(iter+1) = \gamma \Delta C_i^j(iter) - \beta e_k g_i\left(y_k - y_k^d\right)\left(\left(x_i - C_i^j(iter)\right)/\left(\sigma_i^j\right)^2\right) \tag{4-19}$$

The update rule for the width parameter estimation:

$$\begin{aligned}\Delta \sigma_i^j(iter+1) &= \gamma \Delta \sigma_i^j(iter) \\ &\quad - \beta e_k g_i\left(y_k - y_k^d\right)\left(\sum_{i=1}^{2}\left(x_i - C_i^j(iter)\right)^2 \Big/ \left(\sigma_i^j\right)^3\right)\end{aligned} \tag{4-20}$$

The update rule for the output weight estimation:

$$\Delta w_i(iter+1) = \gamma \Delta w_i(iter) - \beta e_k g_i \qquad 4\text{-}21$$

with

$$g_i = \mu_i(x) \Big/ \sum_{i=1}^{K} \mu_i(x) \qquad 4\text{-}22$$

where K is the total number of training data points, β learning rate, γ momentum factor, e_k is the training error of the kth data point, $iter$ is the iteration number index, y_k^d is the kth true data point, and y_k is the kth model's output.

The performance index PI is used for the continuous adaptation of the algorithm as follows:

- If $PI(iter+1) \geq PI(iter)$ then $\beta(iter) = h_d \beta(iter), \gamma(iter+1)$
- If $PI(iter+1) < PI(iter)$ and $\left|\frac{\Delta PI}{PI(iter)}\right| < \delta$ then $\beta(iter+1) = h_i \beta(iter), \gamma(iter+1) = \gamma_0$

If $PI(iter+1) < PI(iter)$ and $\left|\frac{\Delta PI}{PI(iter)}\right| \geq \delta$ then $\beta(iter+1) = \beta(iter), \gamma(iter+1) = \gamma(iter)$ 4-23

where δ is the threshold for the rate of the relative performance index, and h_i and h_d are the increasing and decreasing factors respectively. The performance index PI follows the behaviour of a RMSE energy function where the following constraints are imposed:

$$0 < h_d < 1$$

$$h_i > 1 \qquad 4\text{-}24$$

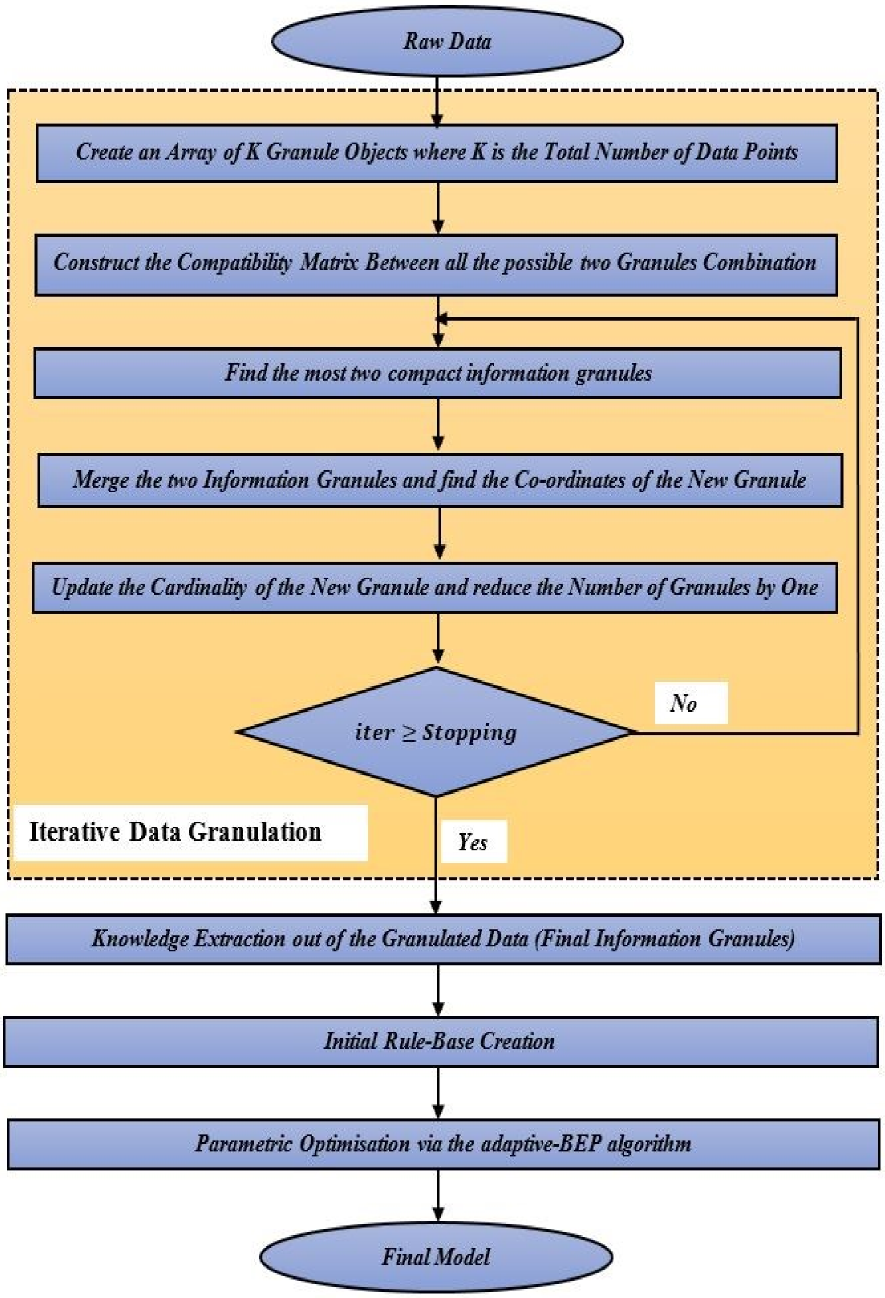


Figure 4.7. Granular computing based RBF-NN modelling framework

4.5. A NEW CONFLICT MEASURE FOR REDUCING UNCERTAINTY AND IMPROVING THE INTERPRETABILITY DURING THE ITERATIVE DATA GRANULATION

The iterative data granulation described above aims at extracting of implicit information, identifying valid, previously unknown patterns and then deriving meaningful knowledge from data. The iterative data granulation is a critical step towards the development of efficient data-driven CI models. This is because the extracted knowledge from the granulated data (final granules) is utilised to create the initial structure (initial rule-base) of the model. During the granulation process, at each iteration the two information granules with the highest compatibility criterion are merged to create a new information granule. However, during the iterative information process uncertainty could present in different forms. For example, uncertainty could occur as a result of: a) when more than two granules have similar values of the compatibility criterion, b) when one of the information granules of interest is an outlier but not necessarily an erroneous information granule, or c) the distance between the two most compact granules is large which leads to information incompleteness while merging. As such, the iterative IG process could ultimately lead to the creation of granules that do not represent the correct distribution of the input space of the process under study (i.e., sparse distribution within the resulting information granule).

Since the acquired knowledge (in the form of information granules) from the iterative data granulation process is utilised to estimate the initial parameters of the RBF-NN (fuzzy logic rule-base), but creating highly uncertain information granules (low quality information granules) will lead to the creation of highly overlapped or less distinguishable FSs in the rule-base (i.e., low quality fuzzy logic rule-base). Less distinguishability leads to loss of transparency and then the overall interpretability of rule-base might be lost (low-level of interpretability and high-level of interpretability). The uncertainty during the iterative granulation process is not taken into consideration in the compatibility measure proposed by Pedrycz in [67] and the extended compatibility measure introduced by Panoutsos in [27].

In this research work, a computational paradigm in which uncertainty present as a consequence of conflict during the iterative IG process is quantified and used to direct the IG process into merging the granules with low-conflict. More specifically, the conflict caused during the iterative IG process can represent a loss of distinguishability and transparency in the rule-base structure (linguistic fuzzy rules) and their characterisation. For instance, Fig. 4.8 shows a case where uncertainty caused by conflict occurred and its effect on the construction of the final linguistic rules. As a consequence of the conflict during the iterative IG process, the final information granules can be misinterpreted and hence lack of overall model parsimony. The proposed uncertainty measure uses the Shannon entropy theory to capture and process conflict uncertainty present during the iterative merging operation of the granules. The conflict type of uncertainty is very common in some manufacturing processes such as friction stir welding, for instance information conflict may present as a consequence of low process repeatability where two similar process parameters in the input space produce slightly different results in the output space.

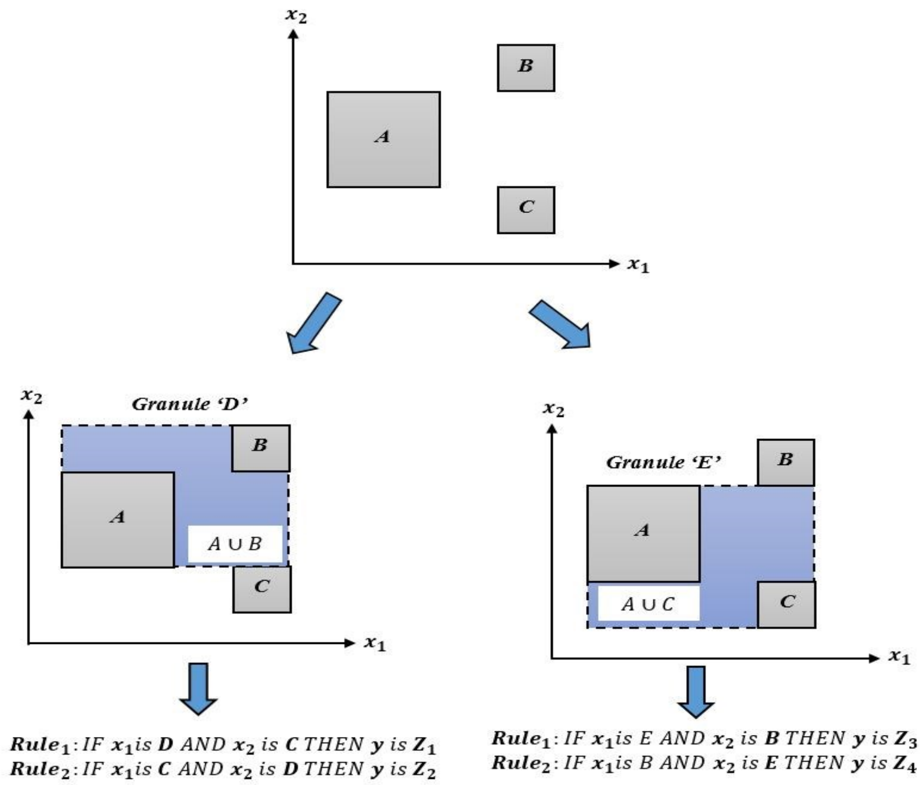


Figure 4.8. Uncertainty caused by conflict during granulation and its effect to the rule-based structure.

4.5.1. IMPROVING LOW-LEVEL INTERPRETABILITY IN THE RBF-NF SYSTEM VIA MEASURING CONFLICT DURING THE ITERATIVE GRANULATION PROCESS

In information theory, uncertainty is a type of deficiency that usually emerges when dealing with information. The sources and kinds of uncertainties may be classified as: a) random event; b) experimental error; c) uncertainty in judgement; d) lack of evidence and e) lack of certainty in evidence [4]. The information obtained from a system is often not fully reliable due to incompleteness, impression, fragmentary, vagueness, and contradiction associated with measurements [3]. In general, a number of theories have been proposed to model and deal with uncertainty. They include fuzzy sets theory [8], possibility theory [14, 15], evidence theory [9], the theory of fuzzy measures [16] and rough sets theory [12]. The essence of information uncertainty relies on the mathematical theory within which uncertainty relating to solve various real-world problems is formalised [3].

In information theory, three basic types of uncertainty have been identified by Klir and Yuan [5]: a) fuzziness, b) non-specificity, and c) discord. Fuzziness refers to lack of definite or sharp boundaries (i.e. imprecise boundaries). Discord or strife expresses the disagreement in choosing among several alternatives and non-specificity or imprecision, which occurs when two, or more alternatives are left unspecified. The latter one has been studied within context of plausibility measures and belief measures [9], while the former two have been studied under fuzzy sets theory [224]. It is noted that in [5], uncertainty can be widely partitions into two facets: fuzziness and non-fuzziness (ambiguity). The general two facets of uncertainty can be viewed in Fig. 4.9. Fuzziness occurs when the boundary of a set is not sharply defined. Many studies have been devoted on measures of fuzziness [211, 225, 226]. On the other hand, non-fuzziness occurs due to randomness and/or non-specificity associated with a system. In this case the output of a system is unambiguously (crisply) defined. The total non-fuzziness comprises at least two components: uncertainty due to randomness and to non-specificity. Several researchers have proposed different uncertainty measures for these two facets of non-fuzziness. In [227] Hohle proposed a measure to estimate the level of conflict present in a body of evidence, whereas Yager [228] introduced a measure called dissonance or conflict. However, in [229] authors highlighted some limitations associated with the measure of conflict that was proposed by Hohle and they proposed a new conflict measure. In [230], authors introduced an uncertainty measure to quantify the non-specificity in a possibility distribution and Dubios and Prade went on and extended the measure to any body of evidence [231].

A list of different non-fuzzy existing measures is summarised in [232, 233]. In [234] Lamata and Moral highlighted that the uncertainty of a system is a composite of two different type of uncertainties, called global uncertainty measures. Another type of composite measures called total uncertainty (TU) that also defined as the sum of non-specificity and a new conflict measure they called discord was introduced by Klir and Ramer [229].

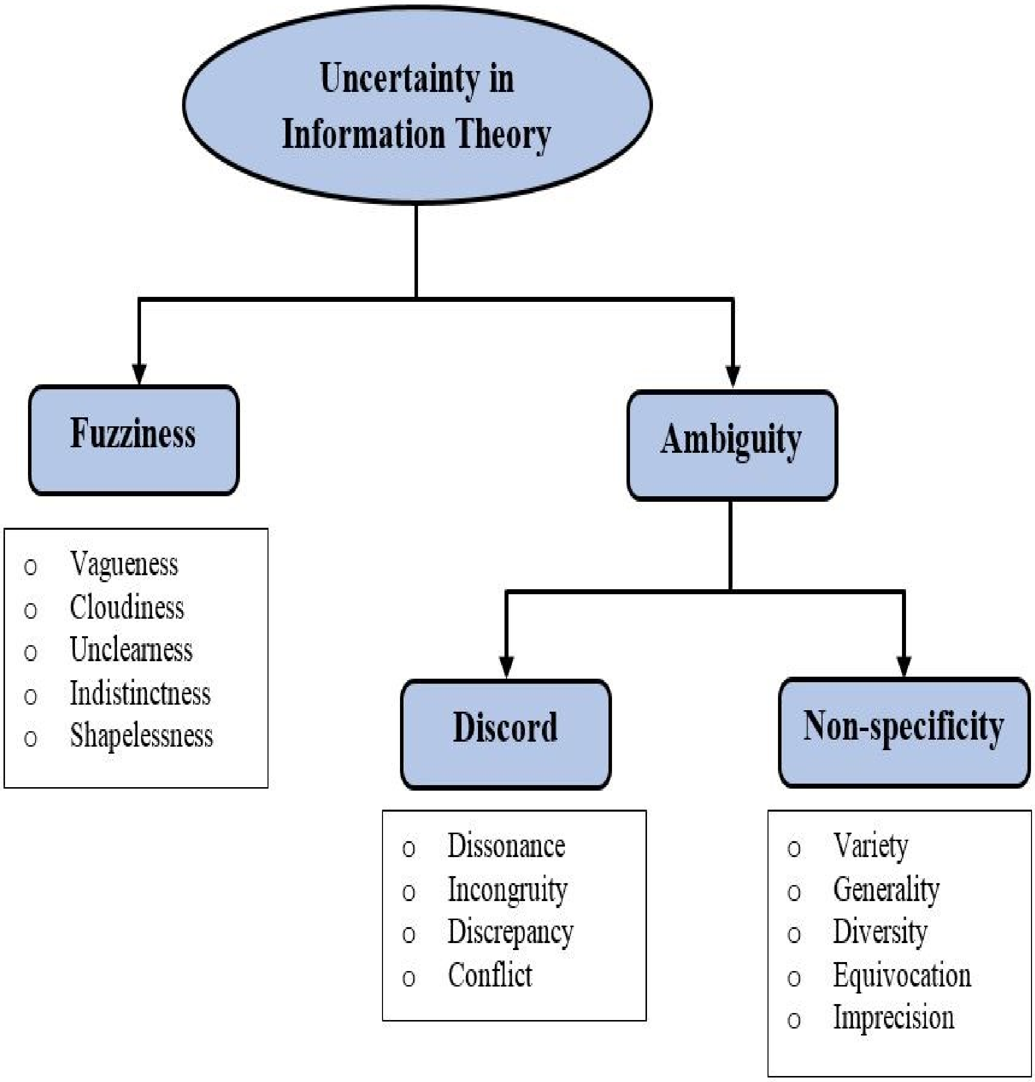


Figure 4.9.Three basic types of uncertainty measures [5].

In order to quantify the degree the conflict during the granulation process and then attenuate its effect. The conflict measure (CM) is defined within the theory of evidence framework [9]. The evidence theory also known as Dempster-Shafer theory is one of the most prevalent mathematical frameworks for dealing with uncertainty [235].

Various definition of uncertainties have been introduced [224, 232, 233], the uncertainty measure considered in this study is defined as [209]:

$$Con\big(m_1(A), m_2(B)\big) = -log_2(1-k) \qquad 4\text{-}25$$

with X is the universe of discourse, m is basic probability assignment, $m(A)$ is the degree of belief or evidence that the element of interest belongs to the set A, $m_1(A), m_2(B)$ are basic assignments representing two bodies of evidence. For each set $A \in X$, the value of $m(A)$ expresses the degree of evidence supporting the evidential claim that a specific element of the universe of discourse X belongs to the set A. The conflict function takes the values from 0 to ∞ and it is a monotonic function increasing with the value of k, and

$$k = \sum_{\substack{i,j \\ A_i \cap B_j = \phi}} m_1(A_i).m_2(B_j) \qquad 4\text{-}26$$

$Con\big(m_1(A), m_2(B)\big) = 0$ if only if $m_1(A)$ and $m_2(B)$ do not conflict at all (i.e., $k = 0$), and $Con\big(m_1(A), m_2(B)\big) = \infty$ only if they conflict totally (i.e., $k = 1$).

In this research work, the basic assignment function is defined as the degree of inclusion between the granules of interest [195]. The degree of inclusion $I_{n(A,B)}$ between two information granules can be expressed as the ratio of two volumes as follows [67]:

$$Inclusion(A, B) = I_{n(A,B)} = \frac{Volume(A \cap B)}{Volume(A)} = \frac{V(A \cap B)}{V(A)} \qquad 4\text{-}27$$

where $I_{n(A,\mathrm{B})}$ is the inclusion of granule A into granule B, and A ∩ B is the degree of overlapping of granule A upon granule B.

$$V(A \cap B) = \prod_{i=1}^{n} length_i(A \cap B) \qquad 4\text{-}28$$

$$length_i(A \cap B) = max(A \cap B) - min(A \cap B) \qquad 4\text{-}29$$

$$V(A) = volume(C) = \prod_{i=1}^{n} length_i(A) \tag{4-30}$$

$$length_i(A) = max(A_i^-, A_i^+) - min(A_i^-, A_i^+) \tag{4-31}$$

The inclusion measure in Eq. 4-27 $I_{n(A,\mathrm{B})}$ is a monotonic function that satisfies the following conditions: $I_{n(A,I)} = 1$ and $I_{n(A,\emptyset)} = 0$ where I and $\emptyset$ are the unit hyper-box and the empty set in the multidimensional space respectively [30]. Fig. 4.11 shows how to calculate the inclusion between two information granules A and B and all the cases for one-dimensional information granules with the corresponding values of the inclusion measure.

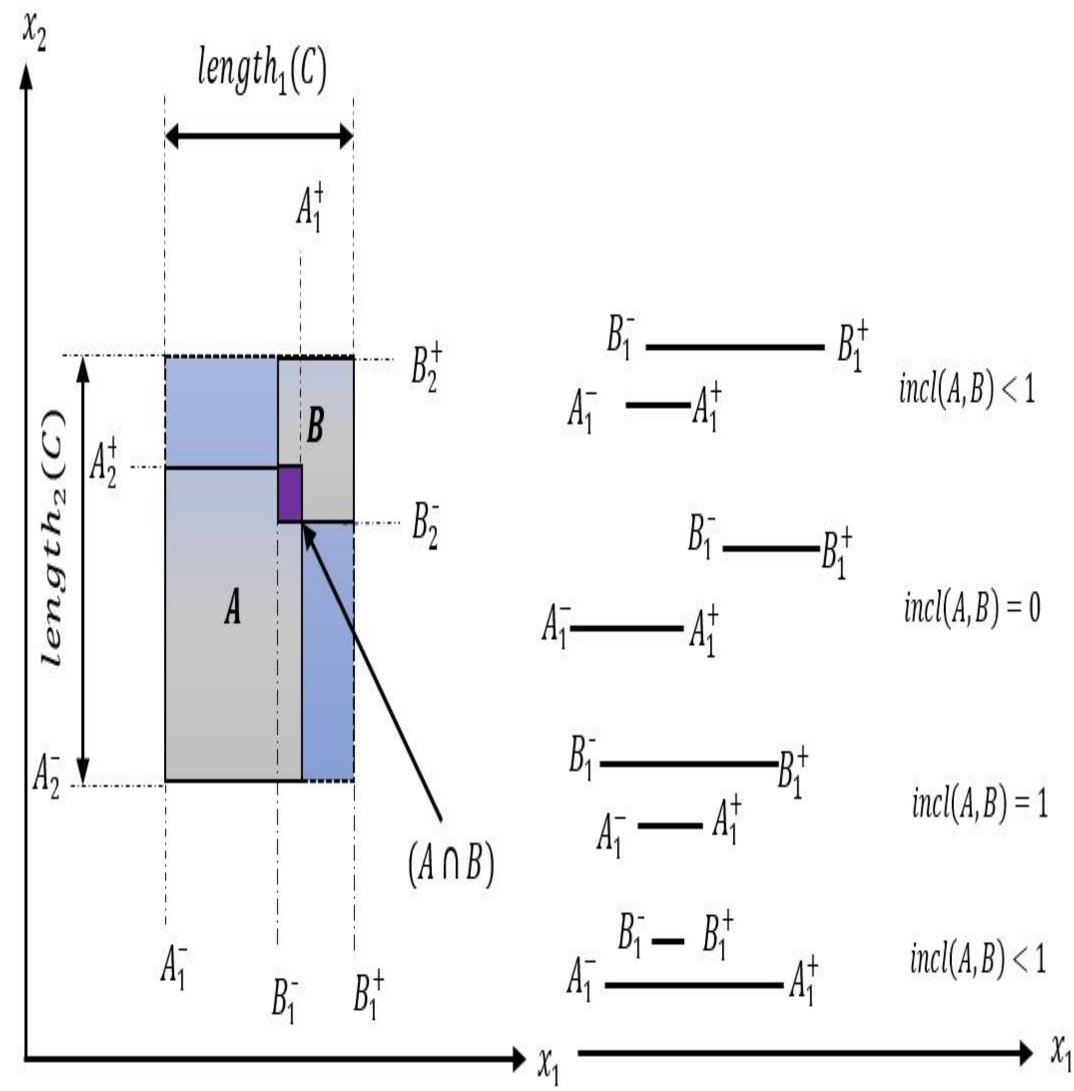


Figure 4.10. Computing the inclusion using the intervals of the granules A and B.

From Eq. 4-25, the conflict function between the granules of interest A and B can be written as follows:

$$Con(A, B) = -log_2(1 - (k)) \quad \text{4-32}$$

$$k = \sum_{\substack{i,j \\ A_i \cap B_j = \phi}} I_{n(A,B)} . I_{n(B,A)} \quad \text{4-33}$$

To measure the uncertainty associated with the conflict function, the conflict measure (CM) can be expressed via the Shannon entropy (e_H) [236], in Eq. 4-34 shown below.

$$e_H = -\sum_{i=1}^{n} (p_i) \log(p_i) \quad \text{4-34}$$

The term p_i represents the probability of occurrence of an event i, where $0 \leq p_i \leq 1$ and

$$\sum_{i=1}^{n} p_i = 1 \quad \text{4-35}$$

Here p_i is computed as the conflict function that indicates the degree of inclusion between two information granules. Therefore, the conflict measure can be stated as:

$$CM = -\sum_{i=1}^{n} (Con(A, B)) log_2(Con(A, B)) \quad \text{4-36}$$

The extended compatibility criterion proposed in [27] is used to calculate the compatibility between information granules to allow merging the most compact information granules while at the same time the conflict uncertainty present during the granulation is taken into consideration to direct the iterative IG process into merging the information granules with low degree of conflict. Fig. 4.11 illustrates the proposed GrC-CM based RBF-NF system modelling framework.

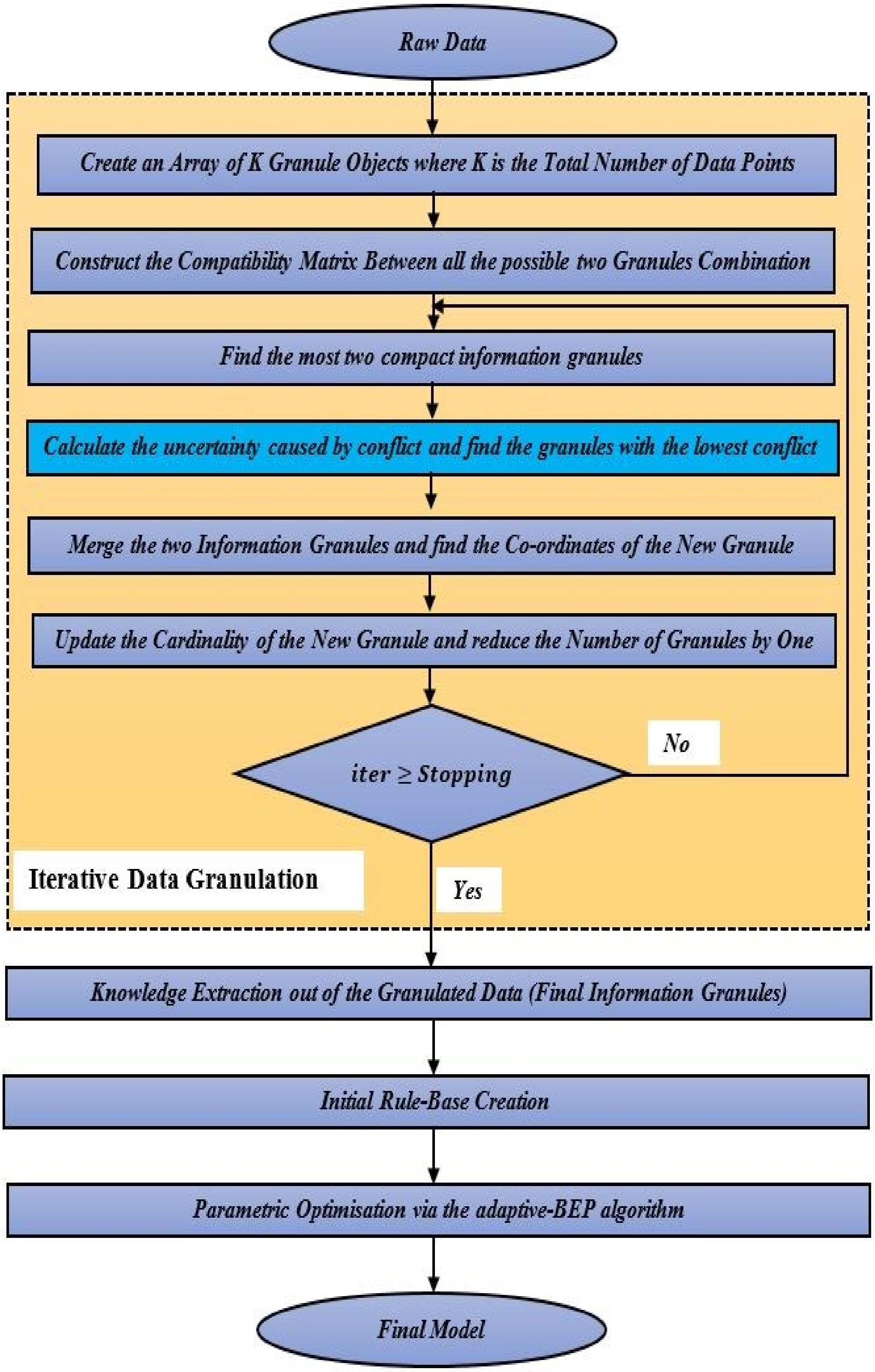


Figure 4.11. Granular computing based RBF-NN modelling framework and conflict measure.

4.5.2. HIGH-LEVEL OF INTERPRETABILITY IN THE RBF-NF SYSTEM

As is discussed earlier, there is still no standard interpretability measure to assess how good overall model interpretability is and it depends on many factors such as consistency, parsimony, simplicity, and completeness of the rule-base structure. Consistency of the rule-base has received much attention by researchers and practitioners this is because seriously inconsistent fuzzy rules will deteriorate the performance and make the fuzzy system inexplicable [19, 24, 237, 238]. The degree of consistency in the rule-base is measured by the absence of contradictory rules in the rule-base, i.e., fuzzy rules with similar antecedent parts should have similar consequent parts. The consistency index (Con) between two linguistic fuzzy rules $Rule_i$ and $Rule_k$ is calculated as follows [239, 240]:

$$\boldsymbol{Rule_i}: \boldsymbol{IF}\ x_1\ is\ A_1^i\ and, \ldots, and\ x_d\ is\ A_d^i, \boldsymbol{THEN}\ y\ is\ B^i \qquad 4\text{-}37$$

$$\boldsymbol{Rule_k}: \boldsymbol{IF}\ x_1\ is\ A_1^k\ and, \ldots, and\ x_d\ is\ A_d^k, \boldsymbol{THEN}\ y\ is\ B^k \qquad 4\text{-}38$$

$$Con(Rule_i, Rule_k) = exp\left[-\frac{\left(\frac{SRP(Rule_i, Rule_k)}{SRC(Rule_i, Rule_k)} - 1.0\right)^2}{\left(\frac{1}{SRP(Rule_i, Rule_k)}\right)^2}\right] \qquad 4\text{-}39$$

where $A_1^i, A_j^i, \ldots, A_d^i$ and B^i are the antecedent and consequent parts of the ith fuzzy rule respectively. $A_1^k, A_j^k, \ldots, A_d^k$ and B^k are the antecedent and consequent parts of the kth fuzzy rule respectively. SRP and SRC are the similarity of consequent premise (antecedent) and the similarity of the rule consequent respectively. They are defined as follow:

$$SRP(Rule_i, Rule_k) = min_{r=1}^{M} S\left(A_j^i, A_j^k\right) \qquad 4\text{-}40$$

$$SRC(Rule_i, Rule_k) = S\left(B^i, B^k\right) \qquad 4\text{-}41$$

where M is the number of rules, r is the index of the rules, and $S(.,.)$ is the similarity measure between two fuzzy sets A and B and is defined as follows [241]:

$$S(A, B) = \frac{|A \cap B|}{|A \cup B|} \qquad 4\text{-}42$$

where $\cap$ and $\cup$ represent the intersection and union operators respectively. $|.|$ denotes the cardinality of the resulting fuzzy set.

From Eq. 4-39, the degree of consistency between two fuzzy rules tends to be high when the values of SRP and SRC are in proportion, provided that the value of SRP is high. If the fuzzy rules possess similar premise and similar consequent, the degree of consistency reaches its highest value of one. If the fuzzy rules possess similar premise but different consequent, then the degree of consistency takes the value in the range $[0,1]$. Additionally, if the SRP of two fuzzy rules is very low the degree of consistency is always high regardless of how the relation between SRP and SRC changes [19].

4.6. SIMULATION RESULTS

To demonstrate the benefits of the proposed modelling framework, in this section a comparative study between the proposed GrC-RBF-NF and GrC-RBF-NF with conflict measure for two different simulation examples is provided. The first simulation example uses the IRIS plant database [242], which is probably one of the most popular and widely used best-known benchmarking databases in pattern recognition. In the second simulation example, the proposed framework is applied to a real-industrial problem. The prediction of spindle peak torque of steel Friction Stir Welding is investigated. Steel FSW is a complex thermo-mechanical process that exhibits highly non-linear and complex as well as sparse databases as it was discussed in Chapter 3.

4.6.1. EXAMPLE 1: IRIS PLANT DATASET PATTERN CLASSIFICATION

4.6.1.1. MODELLING RESULTS BY USING DATA GRANULATION

In this example, the proposed granular computing RBF-NF modelling framework is employed for classification of the IRIS database which was obtained from UCI Machine Learning Repository and generated by R.A. Fisher [242]. The IRIS data set consists of three classes, viz; 1) IRIS Setosa, 2) IRIS Versicolour and 3) IRIS Virginica of 50 objects each, where each class of objects refers to a type of an IRIS plant. Four numeric attributes namely 1) sepal length, 2) sepal width, 3) petal length and 4) petal width are used to predict the belongingness to the class of IRIS plant. The first class is linearly separable/distinguishable from the remaining two classes and the latter two classes are not linearly separable from each other. For the purpose of modelling, the initial dataset (150 instances) was split randomly into two sub-sets: 105 (70%) data points to train the model and 45 (30%) data points to test the generalisation capability of the final model.

Two simulation studies were carried out to compare the performance of the proposed modelling framework with a different number of fuzzy rules (three and five in this research work). The same simulation conditions were also considered in order to fairly evaluate the performance of the models. The initial structure of RBF-NF system is identified via the iterative granulation process described in Section 4.3. The initial structure of the rule-base is optimised via the adaptive-BEP described in Section 4.4.3. The performance indices based on the RMSE and MAE% of the final models are shown in Table 4-1. Fig. 4.12 depicts the data fit of the 5-rule RBF-NF model for the training and testing respectively. It is obvious from results obtained that model with more rules produced better prediction classification performance with more than 98% accuracy. The results obtained from the proposed modelling framework are better than those reported in [35] for the same classification problem with the same number of linguistic rules.

The fuzzy rule-base of the obtained model can be expressed as a collection of linguistic rules in the following form:

$$\boldsymbol{Rule_i}\text{: } \boldsymbol{IF}\ Sepal\ Length\ \boldsymbol{is}\ A_1^i\ \boldsymbol{AND} Sepal\ Width\ \boldsymbol{is}\ A_2^i\ \boldsymbol{AND}\ Petal\ Length\ \boldsymbol{is}\ A_3^i$$
$$\boldsymbol{AND}\ Petal\ Width\ \boldsymbol{is}\ A_4^i, \boldsymbol{THEN}\ the\ IRIS\ Plant\ \boldsymbol{is}\ B^i \qquad \text{4-43}$$

$A_1^i, \dots,\ A_4^i$ are the FSs, $i = 1, \dots, M$, M is the number of rules, i is the ith fuzzy rule.

Table 4.1 Performance of the GrC RBF-NF model classification in Example 1.

Number of rules	**RMSE±SD**		**MAE%±SD**	
	Training	Testing	Training	Testing
3	0.10±0.01	0.16±0.01	1.01±0.02	1.99±0.02
5	0.07±0.00	0.10±0.00	0.66± 0.00	1.17±0.01

The consistency matrix representation for the final rule-base is shown in Table 4-2. The consistency index is used to analyse the degree of consistency among the fuzzy rules. The higher the value the more consistent the rules [19].

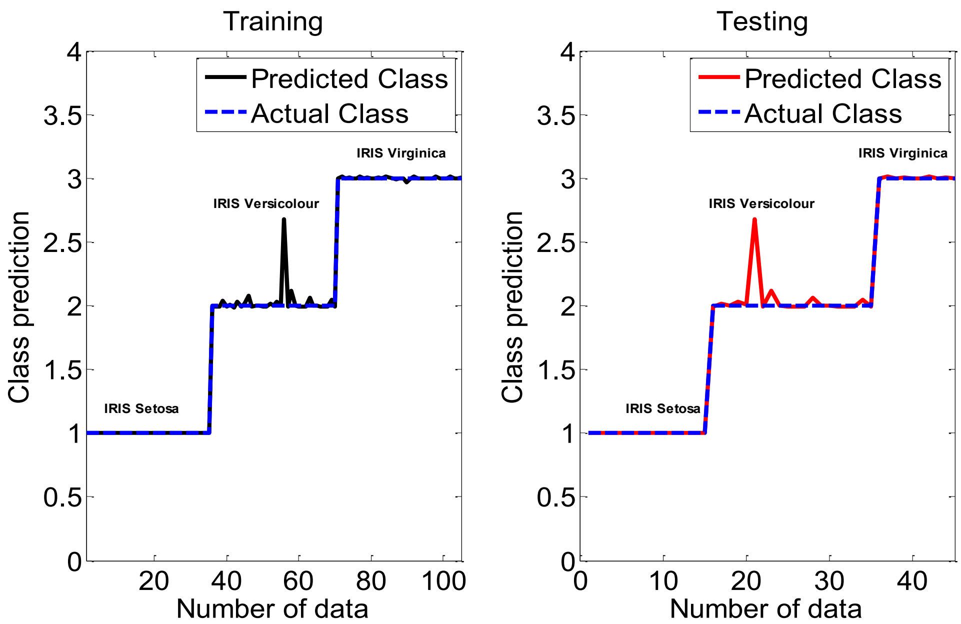


Figure 4.12. Data fit of the 5-rule RBF-NF model for the classification prediction of IRIS dataset after iterative data granulation.

Table 4.2 Consistency matrix for the 5-rule RBF-NF model for the classification prediction of IRIS dataset after iterative data granulation.

Rule no.	1	2	3	4	5
1	1.0000	0.7320	0.8237	0.5934	0.9249
2	0.7320	1.0000	0.8014	0.7821	1.0000
3	0.8237	0.8014	1.0000	0.9123	0.7675
4	0.5934	0.7821	0.9123	1.0000	0.8231
5	0.9249	1.0000	0.7675	0.8231	1.0000

Even though, the proposed granular based modelling framework produced a good level of generalisation capability, the corresponding final rule-base suffers from the lack of distinguishability between MFs in the input space (particularly in petal length and petal width dimensions) and output space as shown in Fig. 4.13. This is due to the iterative granulation process described in Section 4.3 doses not take in account the uncertainty presents as a result of the conflict during the granulation process. Lack of distinguishable in the FSs of each rule in the rule-base degrades the overall interpretability of the rule-based system. Therefore, in the following Section the results obtained in the case where the iterative granulation process is modified to include a conflict measure (CM) will be presented.

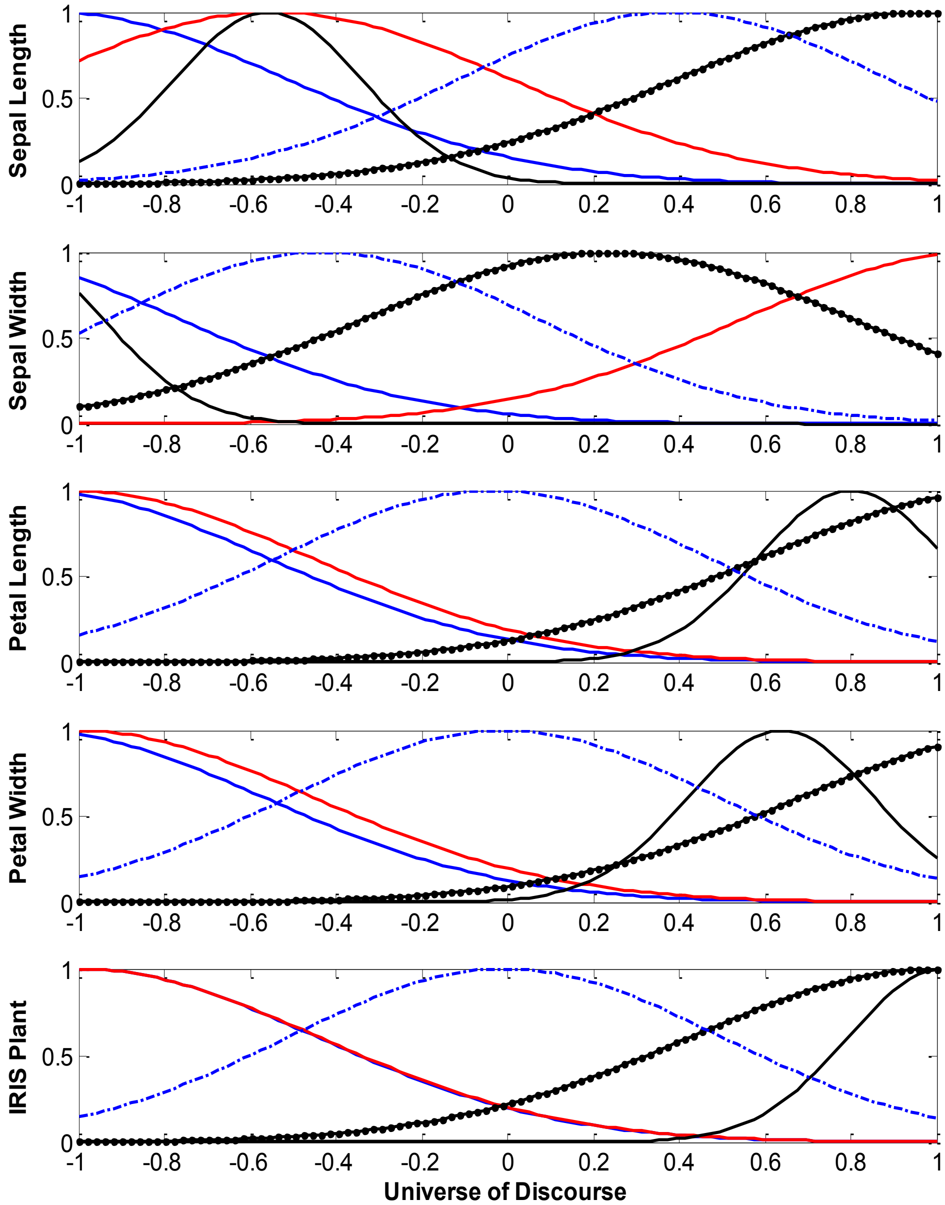


Figure 4.13. The final rule-base of RBF-NF model constructed by using granulation and adaptive-BEP.

4.6.1.2. MODELLING RESULTS BY USING IMPROVED INTERPRETABILTY GRANULATED DATA

In like manner, the same modelling procedures described in Section 4.4 were carried out. The 150 IRIS plant data points for training and testing purposes and the initial structure identification for the RBF-NF were identical to those used in the previous Section. The performance indices based on the RMSE and MAE% of the granular based conflict measure (GrC-CM) modelling framework are summarised in Table 4-3. The simulation results obtained by using the iterative data granulation process based on conflict measure and the adaptive-BEP for the 5-rules RBF-NF model is depicted in Fig. 4.14.

Compared with the previously obtained results from the previously developed framework without including the conflict measure, it is clear that the training performance and the generalisation capability of the proposed modelling framework based GrC-CM is better than the GrC-RBF-NF in terms of accuracy. Furthermore, from the final rule-based system constructed by using the proposed modelling framework based GrC-CM and adaptive-BEP shown in Fig. 4.15, it is obvious that the MFs in the input space (particularly in petal length and petal width dimensions) and output space are more distinguishable than of those in Fig. 4.13. Moreover, the consistency indices among the fuzzy rules in Table 4-4 are higher than those reported in Table 4-2. It can be concluded that in general the GrC-RBF-NF-(CM) outperforms the GrC-RBF-NF, while the proposed iterative granulation based on conflict measure better quality information granules/rules that are more distinguishable to be optimised in order to construct more interpretable rule-based system as compared to the iterative data granulation for setting the initial structure of the rule-base.

Table 4.3 Performance of the GrC RBF-NF model based conflict measure classification in Example 1.

Number of rules	**RMSE±SD**		**MAE%±SD**	
	Training	Testing	Training	Testing
3	0.10±0.01	0.16±0.01	1.01±0.01	1.99±0.02
5	0.06±0.00	0.09±0.00	0.59±0.00	1.07±0.01

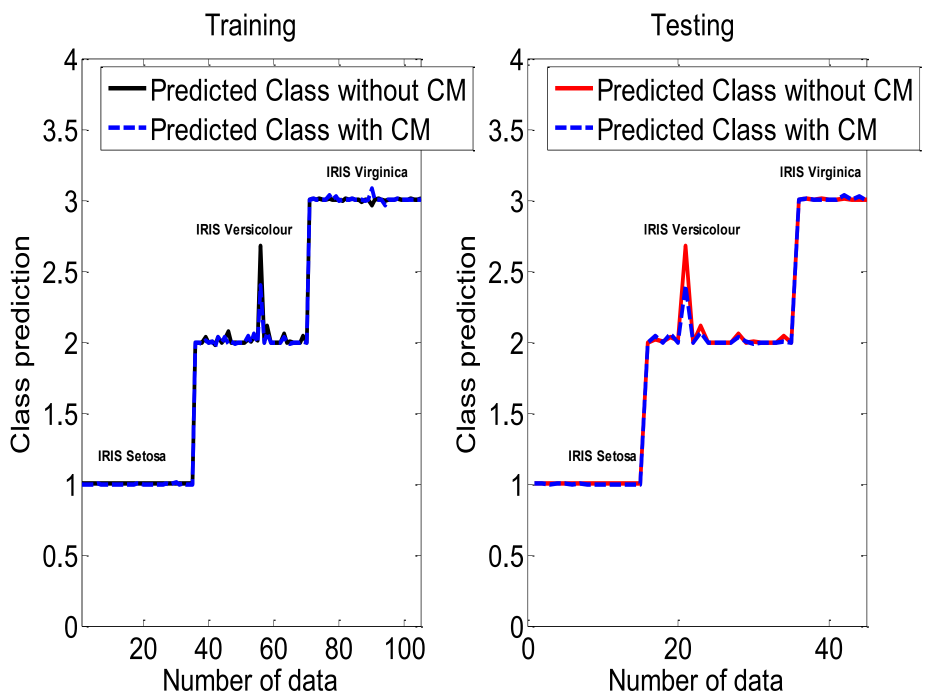


Figure 4.14. Data fit of the 5-rule RBF-NF model for the classification prediction of IRIS dataset with and without conflict measure.

Table 4-4 shows the consistency matrix representation for the final 5-rule model for the classification prediction of IRIS dataset after iterative data granulation with conflict measure.

Table 4.4 Consistency matrix for the 5-rule RBF-NF model for the classification prediction of IRIS dataset after iterative data granulation with conflict measure.

Rule no.	**1**	**2**	**3**	**4**	**5**
1	1.0000	0.7320	0.8137	0.8034	0.9239
2	0.7320	1.0000	0.8014	0.7821	1.0000
3	0.8137	0.8014	1.0000	0.9123	0.7675
4	0.8034	0.7821	0.9123	1.0000	0.8231
5	0.9239	1.0000	0.7675	0.8231	1.0000

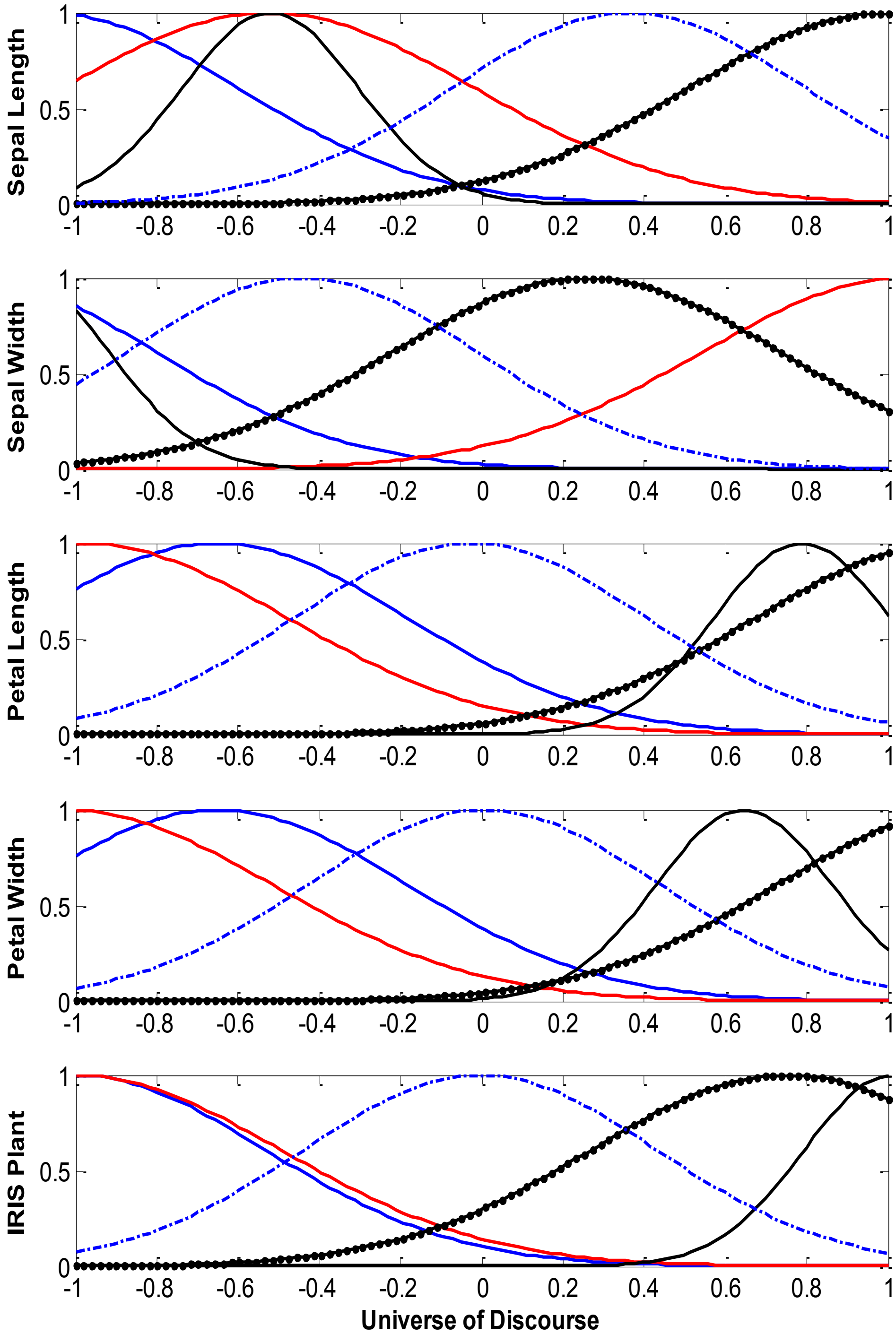


Figure 4.15. The final rule-base of RBF-NF model constructed by using granulation based on conflict measure and adaptive-BEP.

4.6.2. EXAMPLE 2: STEEL FRICTION STIR WELDING DATA

4.6.2.1. MODELLING RESULTS BY USING DATA GRANULATION

For cross validation purposes and also for comparison reasons, the data set used in this Chapter was split identically into two sets similar to that presented in Chapter 3, 133 (70%) data points to train the RBF-NF model, and 58 (30%) data points to test the prediction performance of the final model. By using this cross validation method, the interpolation capability of the model is expected to be improved with the risk of over-fitting. To circumvent the risk of over-fitting and to improve the generalisation capability of the model, 10-folds cross validation approach described in Chapter 3 is used.

Since there is no prior knowledge about the number of clusters (granules), a systematic of number simulations (increased/reduced the number of granules) was carried out similar to those results obtained from the subtractive clustering approach in Chapter 3. It was found that the appropriate number of information granules is between 3 and 17 granules. However, the simulation results with the model having less fuzzy rules (small number of granules) achieved less accuracy performance but the rube-base is simpler in structure and more interpret/able. In contrast to the model with more fuzzy rules (large number of granules) captures more information about the dynamics of the process being modelled and achieved better accuracy performance with lack of interpretability and simplicity.

The geometrical boundaries of the final information granules are unitised to estimate the initial parameters of the RBF-NF model and then to construct the initial multidimensional rule-base as described above. A number of simulation runs were carried out in order to determine the value of the weighting importance factor α at which compact and distinguishable information granules can be obtained. The final compatibility criterion during the iterative granulation cycle is presented in Fig. 4.16.

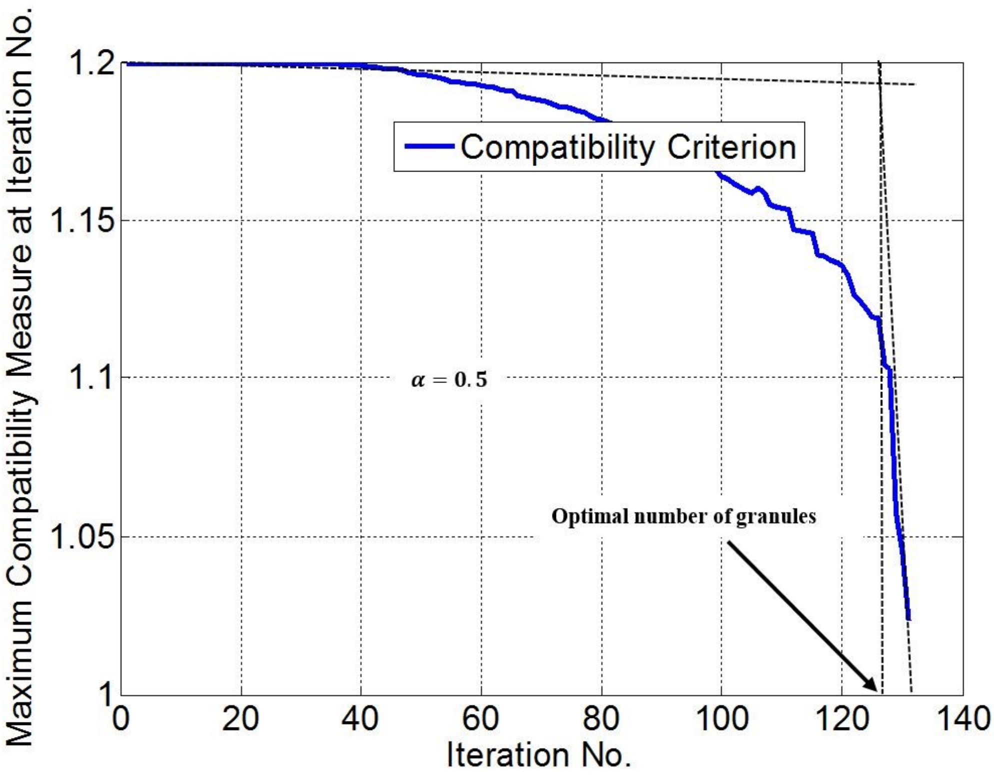


Figure 4.16. Compatibility criterion evolution.

A comparison of the performance index for the RBF-NF using the human-like information capture of granular computing with a different number of fuzzy rules (centres/granules) varies in the range between 3 and 17 is shown in Table 4-5. The performance index based on the RMSE is used to measure the training and testing performance. The performance index based on the MAE is used to measure the overall performance of the model for training and testing in percentage (MAE%). The number of fuzzy rules (centres/granules) is suggested to 5 granules. The evolution of the RMSE, learning rate, and momentum rate are depicted in Fig. 4.17. In Fig. 4.18 plots of simulation results obtained by using the iterative data granulation process and the adaptive-BEP are illustrated. It is evident from the simulation results that the model was not able to provide good predictions accuracy and some data points were not correctly predicted. This is mainly due to the peak torque of the rotational tool is a highly non-linear characteristic in relation to the FSW process parameters, and then the peak torque dataset produced from the process is sparse and difficult to be mapped (modelled) as a consequence of the conflicting data points that exist.

Table 4.5. RMSE and MAE% for the GrC-RBF-NF modelling framework.

Number of Clusters	RMSE±SD		MAE%±SD	
	Training	Testing	Training	Testing
3	46.16±7.00	48.22±7.78	12.68±2.73	11.47±1.59
5	42.83±3.14	49.27±4.04	10.70±1.30	11.08±1.42
6	43.66±3.20	43.46±3.17	11.07±1.45	11.16±1.63
8	43.05±2.98	43.01±3.01	10.81±1.38	10.66±1.49
11	41.27±0.63	44.26±0.71	10.08±1.19	11.29±1.78
17	42.63±1.43	44.81±2.57	10.56±1.23	11.54±1.83

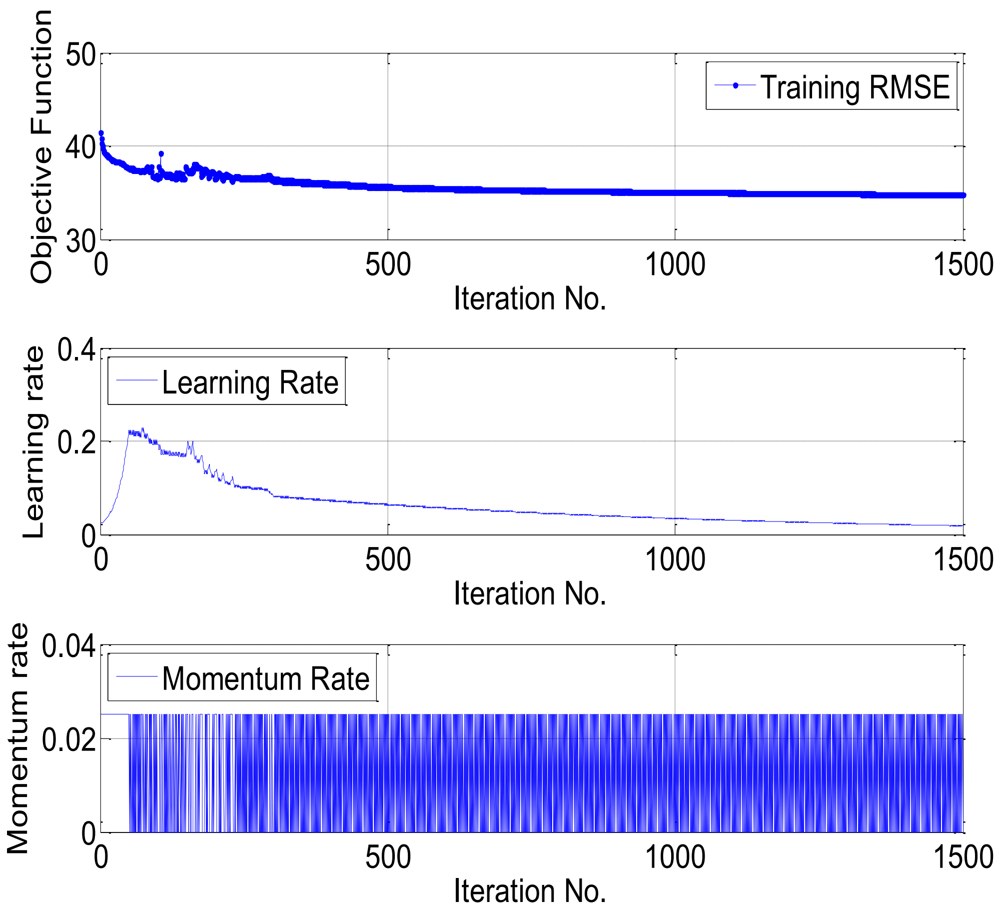


Figure 4.17. Root mean square error, learning rate and momentum rate during the adaptive BEP algorithm.

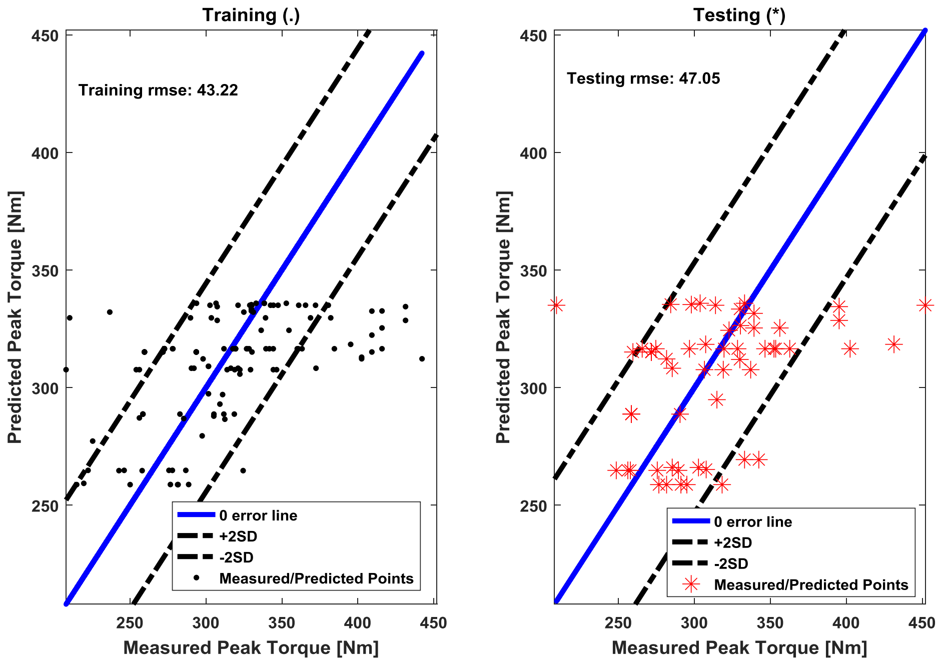


Figure 4.18. Data fit, peak torque prediction by using data granulation to construct the initial fuzzy rule-base.

The final rule-base can be used to describe the complex and non-linear behaviour between welding speed, rotation speed and the predicted spindle peak torque, which can be achieved by taking advantage of fuzzy logic systems. The corresponding fuzzy rules in linguistic format is as follows:

Rule 1: IF *Welding Speed* ***is*** *medium* ***AND*** *Rotation Speed* ***is*** *medium,* ***THEN*** *Peak Torque* ***is*** *very small*

Rule 2: IF *Welding Speed* ***is*** *high* ***AND*** *Rotation Speed* ***is*** *high,* ***THEN*** *Peak Torque* ***is*** *small*

Rule 3: IF *Welding Speed* ***is*** *very small* ***AND*** *Rotation Speed* ***is*** *very small,* ***THEN*** *Peak Torque* ***is*** *medium*

Rule 4: IF *Welding Speed* ***is*** *high* ***AND*** *Rotation Speed* ***is*** *high,* ***THEN*** *Peak Torque* ***is*** *high*

Rule 5: IF *Welding Speed* ***is*** *small* ***AND*** *Rotation Speed* ***is*** *small,* ***THEN*** *Peak Torque* ***is*** *very high*

From a statistical point of view, the neural-fuzzy model based on the human-like iterative information of GrC to elicit the initial rule-base structure outperformed the neural-fuzzy model (ANFIS) based on the subtractive clustering described in Chapter 3. Furthermore, the iterative data granulation offers more transparency, simplicity and efficiency in grouping similar data points together based on their features. Unlike the subtractive clustering, the final information granules constructed by the iterative granulation process does not depend on the number and initial location of centres. Thus, the granulation process involves transparency and distinguishability at the low level of interpretability. However, it is obvious from the fuzzy rules of the GrC based neural-fuzzy model shown above, *Rule 2* and *Rule 4* are two conflicting rules. These fuzzy rules have same antecedents but different consequents. As a result of these conflicting rules, the overall interpretability of the rule-base is lost. Therefore a conflict measure is introduced in order to improve the quality of the final rules. In the next section, the simulation results obtained by using the iterative data granulation process based on conflict measure will be presented.

4.6.2.2. MODELLING RESULTS BY USING IMPROVED INTERPRETABILTY GRANULATED DATA

For comparison reasons, the same modelling procedures described in Section 4.4 were carried out. The data points for training and testing purposes and the initial parameters for the RBF-NF were identical to those used in Section 4.6.2.2. In Table 4-6 and Fig. 19, the performance indices based on the RMSE and MAE% of the previously developed models obtained via Subtractive based ANFIS, GrC-RBF-NF and GrC-RBF-NF-(CM) are illustrated. The training performance and the generalisation capability of the granular based conflict measure (GrC-CM) modelling framework is better than the previously obtained results from the subtractive based ANFIS and GrC models. In Fig. 4.20 the simulation results obtained by using the iterative data granulation process based on conflict measure and the adaptive-BEP for 5-rules model are shown.

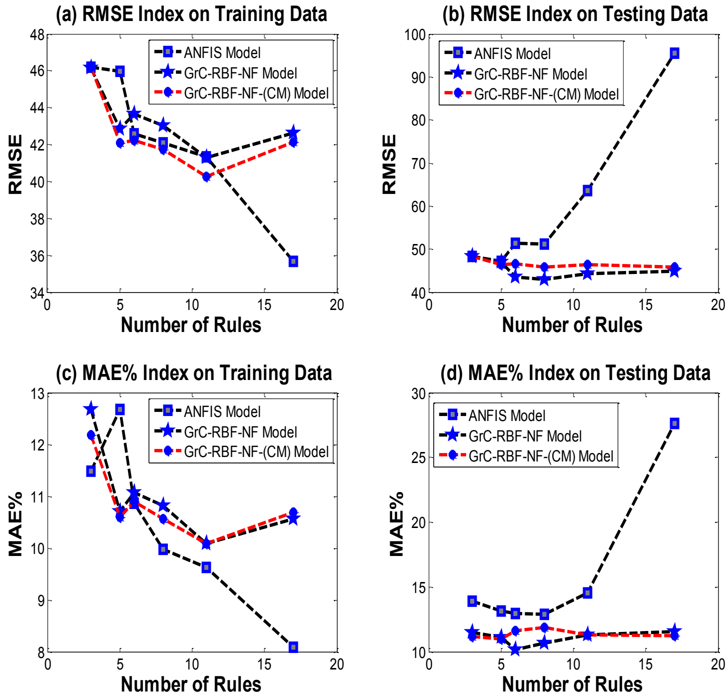


Figure 4.19. Comparison of performance indices based on the RMSE and MAE% for ANFIS, GrC-RBF-NF and GrC-RBF-NF-(CM) modelling frameworks with different number of rules.

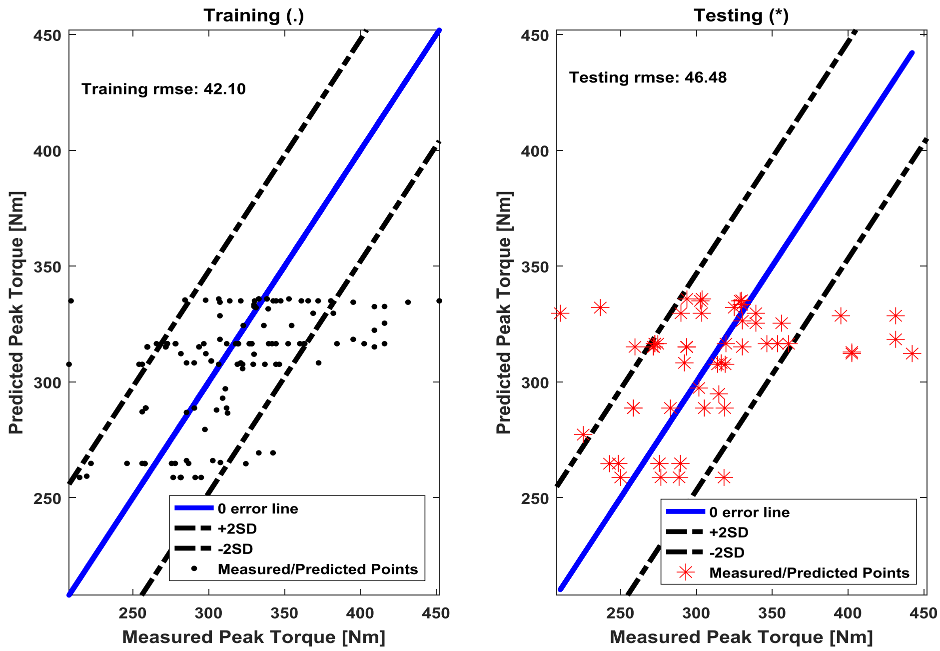


Figure 4.20. Data fit, peak torque prediction by using data granulation based on conflict measure to construct the initial fuzzy rule-base.

Table 4.6. RMSE and MAE% for the ANFIS, GrC-RBF-NF and GrC-RBF-NF-(CM) modelling frameworks.

Number of Clusters	**Data-driven model type**	**RMSE±SD**		**MAE%±SD**	
		Training	Testing	Training	Testing
3	ANFIS	46.18±7.48	48.23±9.11	11.48±1.59	13.92±3.75
	GrC	46.16±7.00	48.22±7.78	12.68±2.73	11.47±1.59
	GrC-CM	46.19±7.69	48.22±8.10	12.19±2.50	11.20±1.14
5	ANFIS	45.96±5.46	47.13±6.01	11.42±1.32	13.15±3.04
	GrC	42.83±3.14	49.27±4.04	10.70±1.30	11.08±1.42
	GrC-CM	42.10±2.02	46.48±2.52	10.60±1.28	11.00±1.39
6	ANFIS	42.60±3.14	51.34±7.71	10.86±1.24	12.98±2.87
	GrC	43.66±3.20	43.46±3.17	11.07±1.45	10.16±1.63
	GrC-CM	42.24±1.89	46.56±2.09	10.89±1.22	11.59±1.55
8	ANFIS	42.11±2.06	51.17±8.34	9.97±1.18	12.89±2.65
	GrC	43.05±2.98	43.01±3.01	10.81±1.38	10.66±1.49
	GrC-CM	41.75±0.98	45.86±1.02	10.56±0.87	11.86±1.14
11	ANFIS	41.34±1.15	63.55±8.87	9.62±1.03	14.51±6.90
	GrC	41.27±0.63	44.26±0.71	10.08±0.75	11.29±1.02
	GrC-CM	40.25±0.41	46.36±0.78	10.08±1.19	11.30±1.78
17	ANFIS	35.65±0.96	239.55±15.10	8.09±0.96	27.63±17.73
	GrC	42.63±1.43	44.81±2.57	10.56±1.23	11.54±1.83
	GrC-CM	42.15±1.43	45.84±3.09	10.69±1.10	11.26±1.18

In general, the proposed conflict measure scheme proved to be more efficient and robust in terms of granulating most compact and at the same time low conflict information granules. The proposed CM framework offers more distinguishable information granules, this is due to it is ability to direct the merging operation into the two information granules which have the highest compatibility criterion and lowest conflict measure at each iteration of the granulation process. Since the final information granules can be regarded as a fuzzy model representation due to the ability of extracting simple linguistic interpretable rules from the final granules in order to describe complex systems. However, the linguistic rule-based system obtained by using the iterative granulation process and shown in Fig. 4.21 suffers from the lack of distinguishability between the MFs in the input and output dimensions.

The iterative granulation process never takes into account the uncertainty caused by conflict while merging, hence producing information granules that are less transparent and less distinguishable. Lack of transparency and distinguishability affects negatively the overall interpretability of the linguistic rule-based system. Fig. 4.22 shows the final rule-base of RBF-NF model constructed by using granulation based on Conflict Measure and adaptive-BEP. It is evident that from the consistency indices and Fig. 4.22, the MFs in the final rule-base of the GrC-RBF-NF-(CM) model are more distinguishable of those in the rule-base of the GrC-RBF-NF. This is due to the ability of the proposed CM in producing better quality information granules. The corresponding fuzzy rules in linguistic format is as follows:

Rule 1: IF *Welding Speed* ***is*** *medium* ***AND*** *Rotation Speed* ***is*** *medium,* ***THEN*** *Peak Torque* ***is*** *very small*

Rule 2: IF *Welding Speed* ***is*** *very high* ***AND*** *Rotation Speed* ***is*** *very high,* ***THEN*** *Peak Torque* ***is*** *small*

Rule 3: IF *Welding Speed* ***is*** *very small* ***AND*** *Rotation Speed* ***is*** *very small,* ***THEN*** *Peak Torque* ***is*** *medium*

Rule 4: IF *Welding Speed* ***is*** *high* ***AND*** *Rotation Speed* ***is*** *high,* ***THEN*** *Peak Torque* ***is*** *high*

Rule 5: IF *Welding Speed* ***is*** *small* ***AND*** *Rotation Speed* ***is*** *small,* ***THEN*** *Peak Torque* ***is*** *very high*

The consistency matrices representation for the 5-rule RBF-NN model with and without conflict measure are shown in Tables 4-7 and 4-8 respectively.

Table 4.7 Consistency matrix for the 5-rule RBF-NF model constructed by using granulation and adaptive-BEP.

Rule no.	1	2	3	4	5
1	1.0000	0.7614	0.6517	0.8514	0.8071
2	0.7614	1.0000	0.7105	0.7921	0.6798
3	0.6517	0.7105	1.0000	1.0000	0.9643
4	0.8514	0.7921	1.0000	1.0000	0.5127
5	0.8071	0.6798	0.9643	0.5127	1.0000

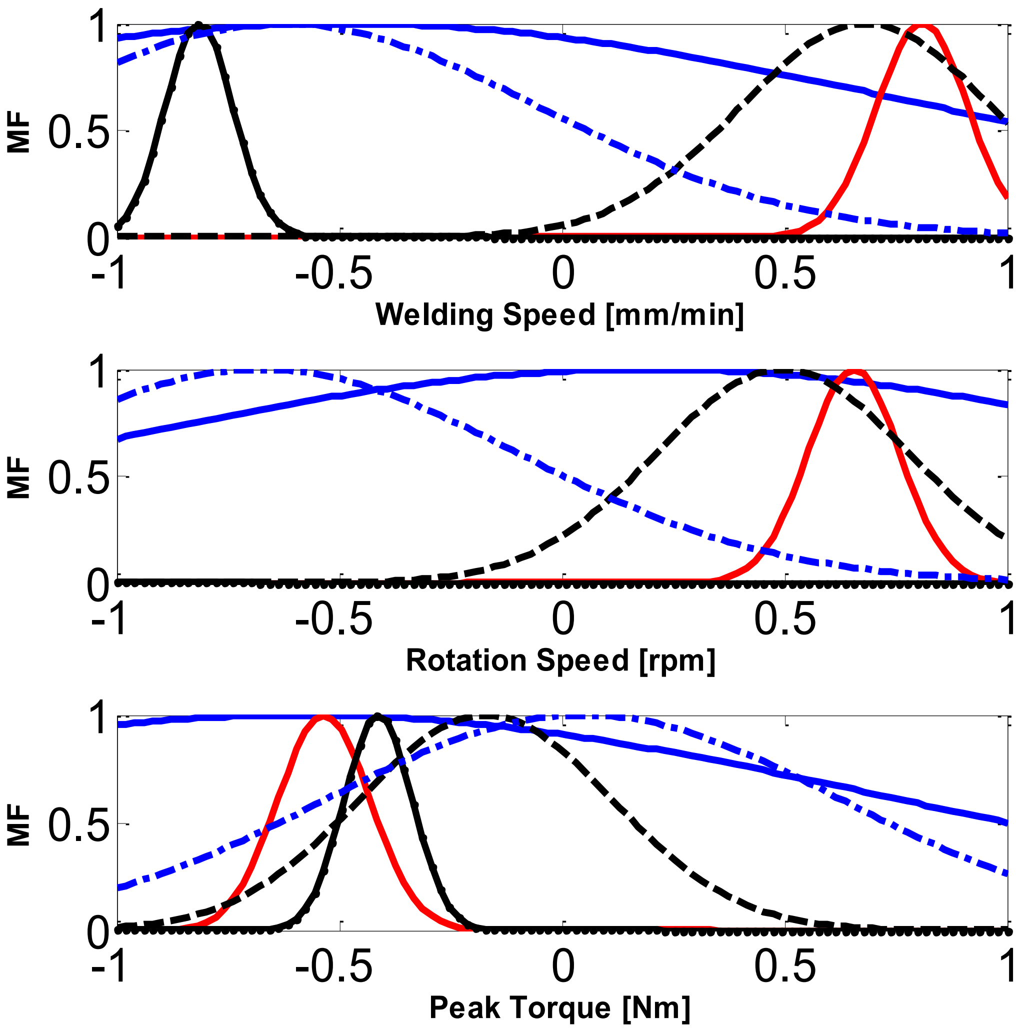


Figure 4.21. The final rule-base of RBF-NN model constructed by using granulation and adaptive-BEP.

Table 4.8 Consistency matrix for the 5-rule RBF-NN model constructed by using granulation based on conflict measure and adaptive-BEP.

Rule no.	1	2	3	4	5
1	1.0000	0.9815	0.6517	0.8514	0.8071
2	0.9815	1.0000	0.7105	0.7921	0.6798
3	0.6517	0.7105	1.0000	1.0000	1.0000
4	0.8514	0.7921	1.0000	1.0000	0.9643
5	0.8071	0.6798	0.9643	0.8721	1.0000

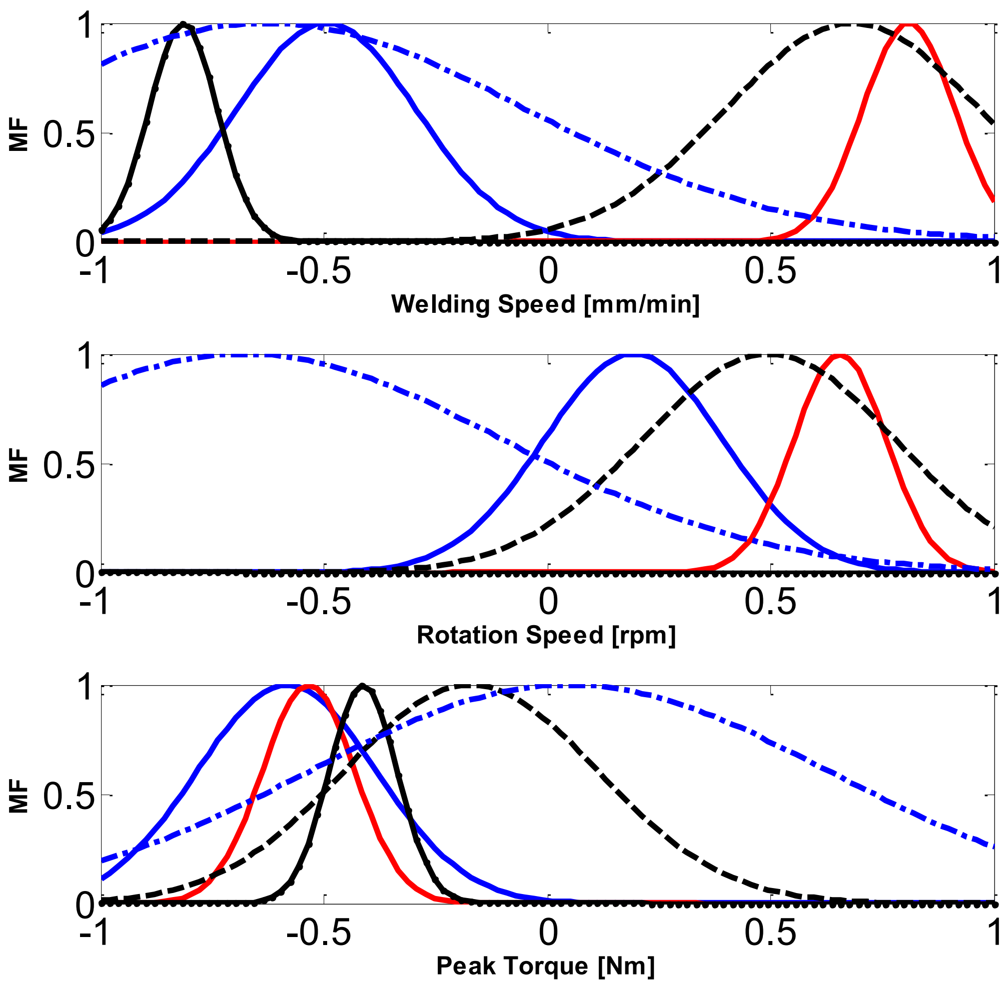


Figure 4.22. The final rule-base of RBF-NF model constructed by using granulation based on conflict measure and adaptive-BEP.

In summary, the iterative granulation process based on a predefined compatibility criterion is modified to include a conflict measure. The conflict measure is utilised to capture the uncertainty during the granulation in order to enhance the overall quality of the granulated data (representation of underlying process data) and interpretability of the linguistic fuzzy rules. The proposed conflict measure does not affect the ability of the iterative human-like information capture of granular computing in grouping similar data according to their features. It is only used as an index to guide the information merging process towards granules with less conflict.

4.7. SUMMARY

In this Chapter, a new conflict measure during the iterative human-like information capture in granular computing (GrC) in order to estimate/evaluate the conflict uncertainty present during the iterative data granulation process is introduced. The uncertainty index is calculated via Shannon entropy theory in order to capture the uncertainty and to serve as a guide into solutions to low-conflict information granules. The conflict measure is employed to direct the iterative IG process into merging (condensing) the granules (data) with low conflict, and therefore producing better quality information granules. The resulting information granules are used to construct the initial parameters of a radial basis function (RBF) based neural-fuzzy model optimised via the adaptive back-error propagation (adaptive-BEP) algorithm.

The main trait of the proposed methodology is used to estimate/quantify the uncertainty as a result of conflict occurred during the iterative data granulation. This results in an improved granulated information set (better quality information granules) in terms of transparency and distinguishability as well as interpretability. The proposed framework was successfully applied to modelling of real-industrial data of steel Friction Stir Welding. This manufacturing process is known of its complexity, highly non-linear behaviour and the production of sparse and limited measurements.

The simulation results show an enhanced granulated information capture (model interpretability) and an achieved generalisation performance compared with similar data-driven modelling approaches. The proposed methodology achieved prediction accuracy with more than 90% in forecasting the spindle peak torque based on the process input parameters of welding speed (mm/min) and tool rotation speed (rpm). The improved information granules from the raw input data are used to extract simple linguistic interpretable rules that can be exploited by the process operator to describe the dynamic behaviour of the process. In addition, to better understand the effects of the process input parameters on the internal process variables as well as aid the process operator (s) in optimising the process.

The results obtained by the proposed methodology led to the writing of articles that were presented at the *3^rd EPSRC Manufacturing The Future Conference in Glasgow, United Kingdom* and the *10^th IEEE International Conference on Granular Computing in Noboribetsu, Hokkaido, Japan* respectively. The results were also presented at the *2013 University of Sheffield Engineering Symposium in Sheffield, United Kingdom* and the *2014 TWI Colloquium in Rotherham, United Kingdom* respectively.

In the next chapter, a new model-based approach relies on an interval type-2 radial basis function neural fuzzy (IT2-RBF-NF) system is proposed in order to take into consideration for uncertainties associated with the linguistic meaning and input noises. The proposed model-based IT2-RBF-NF system will be used to develop a generalised model-based real-time quality monitoring for steel friction stir welding.

CHAPTER 5 - AN INTERVAL TYPE-2 NEURAL FUZZY SYSTEM: IT2-RBF-NFS

This chapter introduces an interval type-2 radial basis function neural fuzzy (IT2-RBF-NF) system that is mathematically equivalent to interval type-2 fuzzy logic systems (IT2-FLSs).

The main contribution of the work presented in this chapter is twofold, on the one hand a new modelling framework is presented by taking advantages of principles of iterative human-like information capture in granular computing (GrC) and the extra degree of freedom from the footprint of uncertainty (FOU) in type-2 fuzzy sets to take into account for the linguistic uncertainties associated with meaning of words. The proposed modelling framework is optimised via an adaptive back-error propagation (adaptive-BEP) approach.

And on the other hand, the proposed modelling framework is used to develop a new generalised systematic human-centric model-based real-time process monitoring framework in steel Friction Stir Welding is presented. The proposed real-time process monitoring framework relies on discrete frequency-based analysis of key internal process variables (namely axial (Fz) and traverse (Fx) forces) that is capable of providing real-time feedback to the process operator (s) in linguistic format (natural language – rule-based system) on the performance of the process. The proposed model-based monitoring framework is also used to forecast in real-time (during welding) quantitative markers of weld quality extracted from the welding tool feedback forces.

5.1. INTRODUCTION

As it was pointed out in [24, 69, 243], fuzzy logic systems are powerful modelling techniques when they are combined with neural networks due to their ability to handle numerical data and linguistic information. However, dealing with linguistic information is problematic especially the determination of the MFs in the antecedent and consequent parts of the rule-base [69]. This is due to controversy between two or more experts in meaning of words [69]. As stated by Mendel in [110] *'Words can mean different things to different people'*. For instance, an expert might intend that 50 miles per hour is high speed with membership grade of 0.70 another might say 50 miles per hour is not fast enough and the membership grade is 0.84. This non-uniqueness/ difference in selecting size and shape of the membership function arises the need for a better modeller (in terms of an interval) to capture the opinion of different experts. In other words, a modeller that takes into account the linguistic uncertainties/non-uniqueness in the rule-base. Another sources of uncertainty can occur in the FLSs are the measurements and process noise [33] that are used to tune the antecedent and consequent parameters in a FLS. Other sources of uncertainty that make fuzzy sets of higher order pertinent are elucidated in [110].

Since knowledge can be expressed in form of linguistic rules (natural language) by using fuzzy sets [8], many real-world complex problems can be greatly simplified. However, many real world applications exhibit measurements noise and modelling uncertainties [93, 94, 244-247]. The effects of all types of uncertainties cannot be minimised and modelled by using type-1 fuzzy logic systems [93, 94, 244-247]. Thus, the concept of general type-2 fuzzy sets (T2-FSs) was introduced by Lotfi Zadeh in 1975 [96] as an extension of the ordinary type-1 fuzzy sets. Since the degree of membership functions in T2-FLSs are themselves fuzzy, while the degree of membership functions in T1-FLSs are crisp [32, 107, 110]. This property provides extra degree of freedom from the type-2 footprint of uncertainty (FOU) and flexibility in modelling the uncertainties frequently encountered in real-world modelling tasks [248]. Thus, T2-FSs can better model uncertainties and minimise their effects [97]. In addition, T2-FLSs use large number of T1-FSs which are embedded within the FOUs of the T2-FSs [97]. The use of such a large number of embedded T1-Fs to describe the relationship between the input and output variables will result in much smaller number of fuzzy rules with greater accuracy, hence, T2-FLSs lead to overall better accuracy, simplicity, information interpretability, and then to permit a deeper understanding of the system under investigation [34, 249]. In [33] Karnik et al. have developed the theory of T2-FSs. More details on the theoretical background and design principles of type-2 fuzzy system (T2-FLS) are described in [34]. T2-FLSs measure the entire systems uncertainty and thus they appear to be a more promising method for handling uncertainties (e.g. in a noisy changing environments) than their type-1 counterparts

[108]. The studies in [97-99, 248] confirmed the effectiveness of the T2-FLSs in better handing the measurement noise and modelling uncertainties.

In practice, T2-FSs theory and T2-FLSs is more computationally expensive compared to their T1-FSs and T1-FLSs counterparts [108]. This is mainly due to the complexity and computational load results from the large number of computations that are needed to obtain type-2 MFs for each variable in the input space. In addition to the number of iterations required to perform a type-reduction from T2-FSs to reduced T1-FSs [250, 251]. In this sense, IT2-FSs become more prevalent and the most widely used type of T2-FSs. The use of IT2-FSs is motivated from the fact that the secondary grades in the secondary MF is an interval and all equal to unity which considerably reduces the computational burn in the type-reduction stage to some extent [109]. Moreover, the concept of interval T2-FSs offers a mathematical tool to handle the linguistic uncertainty in the rule-base and then better understanding of real complex systems from a linguistic perspective. In this research work, an interval type-2 fuzzy logic system (IT2-FLS) modelling framework is used to simplify and minimise the computational cost in order to produce a real-time capable system. The proposed IT2-FLS is of type interval type-2 radial basis function neural fuzzy (IT2-RBF-NF) model as they are mathematically equivalent in their design [35, 108].

This chapter is devoted to the development of a human-centric rule-based modelling framework based on an interval type-2 radial basis function neural fuzzy (IT2-RBF-NF) system optimised via an adaptive-BEP approach. The proposed framework takes advantage of the mathematical equivalence between the T1-FLS and RBF-NN as described in Chapter 4 to develop a new IT2-RBF-NF system that is mathematically equivalent to an interval type-2 fuzzy logic system (IT2-FLS). The initial rule-base structure and FOU of the IT2-RBF-NF system are estimated directly from the iterative data granulation approach used in Chapter 4. The proposed IT2-RBF-NF system incorporates IT2-FSs within the RBF (fuzzification) layer of the NN in order to take into account the linguistic uncertainty associated with the system's variables. The antecedent and consequent parts of each linguistic fuzzy rule in the IT2-RBF-NF system is an IT2-FS and the consequent part is of the Mamdani type fuzzy model, and the output of the network is calculated via average of the interval type-reduced set obtained from the Karnik-Mendel iterative type-reduction approach [108, 109].

Specific to the field of FSW quality monitoring via data-driven computational intelligence (CI) models, a very limited number of studies have been reported [40] [188] [41] but none of the CI models based on type-2 fuzzy sets theory. In [192], Cox et al. have recently investigated the effects of the number of tool rotations on the quality of friction stir spot weld of an aluminium alloy. The authors concluded that there is a linear relationship between the number of tool rotations during the spot weld of an aluminium alloy and the resulting tensile shear strength. In another study [193], Su et al. proposed a measuring approach to measure some of the internal process variables (namely the traverse force (X-axis), axial force (Z-axis) and the tool torque) simultaneously under different welding conditions for the FSW of AA2024-T4 aluminium alloys.

The studies reported in [194] [42] investigated the frequency spectra of the tool feedback forces in X, Y, and Z axes and it was concluded that the frequency spectra of the feedback forces is more likely to contain useful information about the weld quality. Therefore, it would useful to use the information from frequency domain to build monitoring tools. In [194], the authors proposed a model-based classification algorithm based on an ANN structure that takes advantages of frequency domain information to build a weld quality marker for aluminium alloy. However, this model-based approach was proved to be not feasible as the changing process parameters also change the behaviour of the frequency domain information. Due to different process conditions (including input parameters, materials, tool, etc.) different internal process variables (different dynamic behaviour) will be generated and therefore different welding performance (post-weld properties). As a consequence, the modelling performance cannot be generalised for different process conditions, as the model needs to be re-trained or re-calibrated every time the process condition is changed which would limit the usability of the model in real-time.

Building on from the development of the IT2-RBF-NF system, a new generalised systematic human-centric model-based real-time process monitoring framework in steel Friction Stir Welding is presented. The proposed real-time process monitoring framework relies on discrete frequency-based analysis of key internal process variables (namely axial (Fz) and traverse (Fx) forces) that is capable of providing real-time feedback to the process operator (s) in linguistic format (natural language – rule-based system) on the performance of the process. The proposed approach relies on a dynamic IT2-RBF-NF model, that instead of forecasting directly the weld quality, it forecasts a 'moving threshold' that can be used by the operator as an indicator of weld quality. Thus, there is no need to re-tune or re-calibrate the model for every time the welding conditions change.

The main contributions of this chapter can best be categorised as follows:

- A new model-based approach that relies on an IT2-RBF model that is mathematically equivalent to an IT2-FLS in its design, is developed. The IT2-RBF model can be described via simple linguistic interpretable rules extracted from raw data in order to describe dynamic behaviour of the process. The initial parameters of the IT2-RBF-NF (initial rule-base structure) and the FOU are estimated directly via the iterative data granulation approach used in Chapter 4. The iterative information-capture of granular computing has demonstrated its ability in capturing meaningful information from raw process data where only non-linear, complex, and scarce measurements are available and exploit such information to build a linguistic rule-base. The initial parameters of an IT2-RBF based interval type-2 neural-fuzzy model are then optimised via the adaptive back-error propagation (BEP) algorithm and applied to modelling of steel friction stir welding.
- The proposed model-based IT2-RBF-NF system is used to forecast in real-time (during welding) quantitative markers of weld quality. Part quality thresholds are extracted from the frequency spectra of the feedback forces (namely axial (Fz) and traverse (Fx) forces). As already mentioned in Chapter 3, due to the dynamic behaviour of the process, a static model predicting the weld quality threshold would need updating (tuning) for every time the welding conditions change [194]. Therefore, the proposed approach relies on a dynamic model, that instead of predicting directly the weld quality, it predicts a 'moving threshold' that can be used by the operator as an indicator of weld quality. Thus, there is no need to re-tune or re-calibrate the model. The proposed monitoring

framework also takes advantage of the proposed model-based approach to provide continuous linguistic-based feedback to the operator(s) – rule-based human-centric system – on the performance of the process.

- The effectiveness of the proposed model-based approach to better handle uncertainties and produce reasonable predictive performance is tested against multiple linear regression (MLR) and multilayer perceptron neural network (MLP-NN) models as well as type-1 radial basis function neural fuzzy (T1-RBF-NF) model.

5.2. IT2-RBF-NF RULE-BASED MODELLING STRUCTURE

In the literature, there are several studies established the mathematical equivalence between the RBF-NN denoted as T1-RBF-NN and a type of FIS denoted as T1-FLS under some certain conditions [20, 21, 252], and accordingly this mathematical equivalence demonstrates that the RBF-NN can be considered as a T1-FLS sharing characteristics such as universal function approximation, fuzzy rule-base, low and high level of interpretability, etc. Therefore, advances in FSs theory and FLSs may be applied on RBF-NNs under some certain conditions. Additionally, the same conditions may be extended to type-2 fuzzy set theory and an interval type-2 radial basis function neural network (IT2-RBF-NN) [35] can also be considered as mathematically equivalent to an IT2-FLS [107] when the following conditions are met [21]:

1) The number of the receptive-field unites in the hidden layer of the RBF is equal to the number of linguistic IF-THEN rules in the FLS.
2) The MFs within each fuzzy rule are selected as Gaussian MFs.
3) The T-norm operator used to calculate each fuzzy rule's firing strength is multiplication.
4) The outputs of both the RBF-NN and the FIS of FLS are computed using the same defuzzification method (i.e., either COG or weighted sum).

Another line of thought on the mathematical equivalence between the IT2-RBF-NN and IT2-FLS demonstrates that the IT2-RBF-NN initial parameters which represents the antecedent and consequent parameters in the IT2-FLS can be estimated systematically via a data clustering approach and then the parameters are optimised to complete the modelling process. To further understand the concept, this section introduces an interval type-2 neural fuzzy system (IT2-NFS) by incorporating the iterative human-like information granulation in granular computing (GrC) used in Chapter 4, a six-layered interval type-2 radial basis function neural fuzzy (IT2-RBF-NF) model and parametric optimisation via an adaptive back-error propagation (adaptive-BEP) approach.

5.2.1. INTERVAL TYPE-2 RADIAL BASIS FUNCTION NEURAL-FUZZY SYSTEM

5.2.2. IT2-RBF-NF SYSTEM ARCHITECTURE

Since the proposed IT2-RBF-NF system represents a type of extension of IT2-FLSs and inherits some properties from NNs such as universal function approximation, regularisation, interpolation, adaptation, generalisation and learning capabilities [253]. The RBF-NN uses exponentially decaying localised non-linear activation functions (basis functions or T1-MFs) to construct local approximations to nonlinear input/output mapping [22]. Likewise, the IT2-RBF-NN may be seen as a non-linear input/output mapping that uses interval type-2 non-linear activation functions (IT2-MFs). Therefore, the components of the proposed IT2-RBF-NF can be considered as a six-layered multi-input/single output (MISO) model. The configuration of the six-layered IT2-RBF-NF system is illustrated in Fig. 5.1. The proposed structure can be seen as an IT2-FLS with a singleton fuzzifier whose T-norm is the product operator, the antecedent parts use the IT2-FSs (Gaussian IT2-MFs) having fixed means (centres) and uncertain standard deviations (spreads), the consequent part of each linguistic fuzzy rule is of the Mamdani type, and the KM type-reducer which is proposed by Karnik and Mendel [108, 109]. Each linguistic fuzzy rule has the following IF-THEN form:

$$\boldsymbol{Rule_i}\text{: } \boldsymbol{IF}\ x_1\ is\ \tilde{A}_1^i\ \boldsymbol{AND}, \dots, \boldsymbol{AND}\ x_d\ is\ \tilde{A}_d^i, \boldsymbol{THEN}\ y\ is\ \tilde{B}^i \tag{5-1}$$

where $x_1, \dots, x_d$, are the input vectors, $\tilde{A}_1^i, \dots, \ \tilde{A}_d^i$ are the interval type-2 fuzzy sets, $i = 1, \dots, M$, M is the number of rules, i is the index of the rules. The detailed mathematical functions of each layer of the six-layered IT2-RBF-NF are described below.

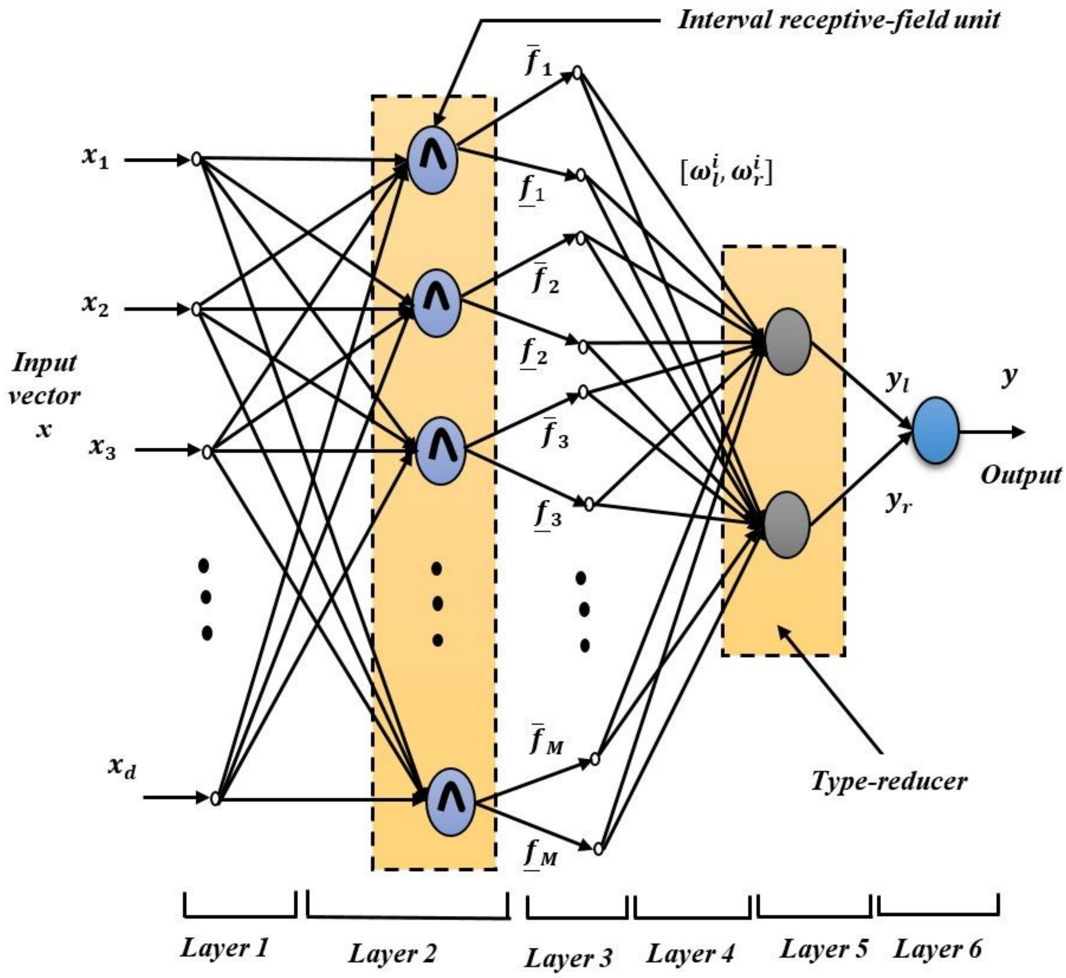


Figure 5.1 IT2-RBF-NF system general structure.

1) Layer 1 (Input Layer):

This layer only transmits the current input values to the next layer directly without performing any computation. Each node in this layer represents one crisp variable from the multidimensional input data $\vec{x} = [x_1, \dots, x_d] \in R^d$, where d is the number of input variable.

2) Layer 2 (Fuzzification Layer):

Each node in this layer uses an interval type-2 MF to perform the fuzzification process in order to produce the upper and lower intervals $[\underline{\mu}_{\tilde{A}_j^i}\ ,\ \bar{\mu}_{\tilde{A}_j^i}]$. With the choice of a Gaussian primary MF having fixed mean m_j^i and uncertain standard deviation that the value in the interval $\sigma_j^i \in [\sigma_{j1}^i, \sigma_{j2}^i]$ can be stated as (see Fig. 5.2):

$$\tilde{A}_j^i(x_j) = exp\left[-\frac{1}{2}\left(\frac{x_j - m_j^i}{\sigma_j^i}\right)^2\right] \equiv N(m_j^i, \sigma_j^i; x_j), \quad \sigma_j^i \in [\sigma_{j1}^i, \sigma_{j2}^i] \qquad 5\text{-}2$$

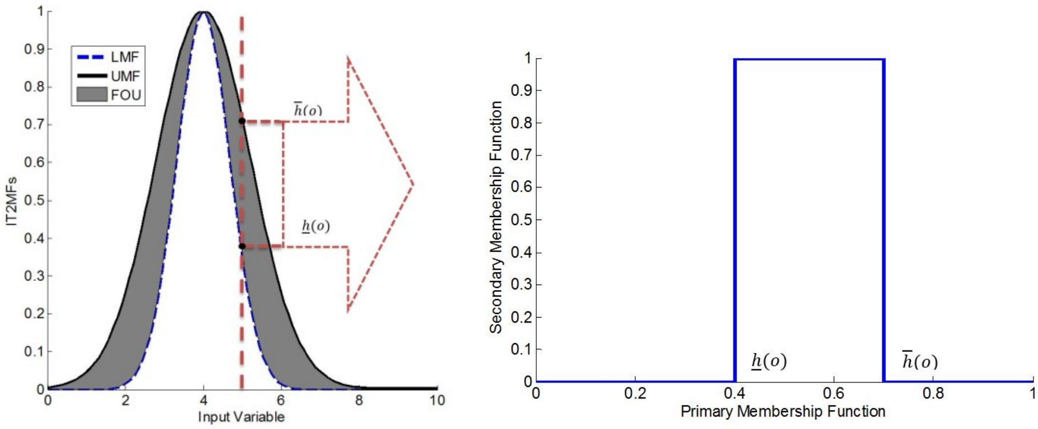
where $\tilde{A}_j^i(x_j)$ is the ith fuzzy set in input variable x_j. It is clear that the IT2-FS is bounded by the upper MF $\overline{\mu}_{\tilde{A}_j^i}$ and lower MF $\underline{\mu}_{\tilde{A}_j^i}$, and the area in between is called the footprint of uncertainty (FOU). The upper membership function (UMF) is

$$\overline{\mu}_{\tilde{A}_j^i} = N(m_j^i, \sigma_{j1}^i; x_j) \qquad 5\text{-}3$$

and the lower membership function (LMF) is

$$\underline{\mu}_{\tilde{A}_j^i} = N(m_j^i, \sigma_{j2}^i; x_j) \qquad 5\text{-}4$$

Each node in this layer corresponds to a linguistic variable (e.g. fast, very fast, etc.) and the output of each node can be represented as an interval (FOU) $[\underline{\mu}_{\tilde{A}_j^i} \, , \, \overline{\mu}_{\tilde{A}_j^i}]$.



(a) (b)

Figure 5.2. The interval type-2 Gaussian MF with uncertain standard deviation in (a), which has lower (thick dashed line) and upper (thick solid line) boundaries. The primary MF at input variable of 5 is an interval $[\underline{h}(o), \overline{h}(o)]$, with uniform interval secondary MF in (b). The shaded region in (a) is the FOU.

3) __Layer 3 (Rules Firing Layer):__

This layer performs the join ⊔ and meet ⊓ operations, which are new concepts introduced in type-2 fuzzy logic theory that are used instead of intersection and union operators in type-1 fuzzy logic theory [112]. The output of this layer is an interval type-1 fuzzy set (i.e. rule node firing strength). The node rule firing strength F^i is calculated by using an algebraic product operation as follows:

$$F^i = \left[\underline{f}^i(x_j), \overline{f}^i(x_j)\right] \qquad 5\text{-}5$$

Where $\underline{f}^i(x_j)$ and $\overline{f}\,(x_j)$ can be written, where * denotes the meet operation under product t-norm:

$$\overline{f}^i = \overline{\mu}_{\tilde{A}_1^i}(x_1) * \dots * \overline{\mu}_{\tilde{A}_p^i}(x_p) = \prod_{j=1}^{n} \overline{\mu}_{\tilde{A}_j^i} \qquad 5\text{-}6$$

$$\underline{f}^i = \underline{\mu}_{\tilde{A}_1^i}(x_1) * \dots * \underline{\mu}_{\tilde{A}_p^i}(x_p) = \prod_{j=1}^{n} \underline{\mu}_{\tilde{A}_j^i} \qquad 5\text{-}7$$

4) ***Layer 4 (Compensatory Firing Layer):***

Each node in this layer has its corresponding firing strength, which is calculated from the previous layer 3. The layer defines the consequences of the rule nodes and the links between this layer and the next layer consist of interval weighing factors $[\omega_l^i, \omega_r^i]$ which will decide outputs of this network.

5) ***Layer 5 (Type-reducer Layer):***

This layer generates a T1-FS output, which is then converted to a numeric output through the defuziffication layer. This T1-FS is also an interval set $[\omega_l^i, \omega_r^i]$, which is determined by its two end points (i.e, left end-point l, and right end-point r). The centroid of the type-2 interval consequent set $\tilde{B}$ in Eq. 5-5, for the case of centre-of-sets COS type-reduction method [108],

$$Centriod_{\tilde{B}^i} = \int_{\theta_1 \in J_{y_1}} \dots \int_{\theta_N \in J_{y_N}} \frac{1}{\sum_{i=1}^{N} y_i\theta_i / \sum_{i=1}^{N} \theta_i} = \left[\omega_l^i, \omega_r^i\right] \qquad 5\text{-}8$$

The extended output is computed as follows:

$$Y_{COS} = \left[y_l^i, y_r^i\right] = \int_{\omega^1 \in [\omega_l^1, \omega_r^1]} \dots \int_{\omega^M \in [\omega_l^M, \omega_r^M]} \int_{f^1 \in \left[\underline{f}^1, \overline{f}^1\right]} \dots$$

$$\int_{f^M \in \left[\underline{f}^M, \overline{f}^M\right]} \frac{1}{\sum_{i=1}^{M} f^i \omega^i / \sum_{i=1}^{M} f^i} \qquad 5\text{-}9$$

Each node in this layer calculates this interval output. The outputs y_l and y_r can be computed as follows:

$$y_l = \frac{\sum_{i=1}^{M} f_l^i \omega_l^i}{\sum_{i=1}^{M} f_l^i} \qquad \text{5-10}$$

$$y_r = \frac{\sum_{i=1}^{M} f_r^i \omega_r^i}{\sum_{i=1}^{M} f_r^i} \qquad \text{5-11}$$

According to [108, 109], the interval type-reduced sets in Eq. 5-10 and Eq. 5-11 can be expressed as:

$$y_l = \frac{\sum_{i=1}^{L} \overline{f}^i \omega_l^i + \sum_{i=L+1}^{M} \underline{f}^i \omega_l^i}{\sum_{i=1}^{L} \overline{f}^i + \sum_{i=L+1}^{M} \underline{f}^i} \qquad \text{5-12}$$

$$y_r = \frac{\sum_{i=1}^{R} \underline{f}^i \omega_r^i + \sum_{i=L+1}^{M} \overline{f}^i \omega_r^i}{\sum_{i=1}^{R} \underline{f}^i + \sum_{i=R+1}^{M} \overline{f}^i} \qquad \text{5-13}$$

where M is the number of rules in the rule base of the IT2-RBF, i is the index of the rules, and $[\omega_l^i, \omega_r^i]$ represents the centroid interval of the consequent type-2 FS of the ith rule. ω_l^i, ω_r^i are also called the weighting factors of the consequent part of the IT2-RBF [254]. The values of L and R can be obtained from the iterative Karnik-Mendel type-reduction method [108, 109].

6) Layer 6 (Defuzzification Layer):

Once y_l and y_r are obtained by using the KarnikMendel iterative type-reduction approach, the type-reduced set can be defuzzified to compute the output values of the system (crisp). For an interval type-reduced set, the defuzzified output y. In this layer, the defuzzified output is then computed by the average of y_l and y_r.

$$y = \frac{1}{2}(y_l + y_r) \qquad \text{5-14}$$

5.2.3. INITIAL STRUCTURE IDENTIFICATION OF THE IT2-RBF-NF SYSTEM

The iterative human-like information granulation of granular computing (GrC) described in Chapter 4 is used to group similar entities in the input space according to a predefined similarity measure. The iterative GrC is used to process numeric data in order to obtain the information relating to clusters (granules). This algorithm has shown its efficiency and simplicity in extracting information out of raw data, inspired by the human cognition of grouping similar objects together better than other well-known data clustering methods such as the mountain [255] and fuzzy C-means (FCM) algorithms [256]. The information that these clusters (granules) provide is then used to estimate the initial structure of the IT2-RBF-NF model.

The granulated input space determines the number of linguistic rules extracted from the input raw data as well as the number of FSs on the universe of discourse of each input variable. In this approach, the relationship between an information granule in multi-dimension and a fuzzy rule is one-to-one relationship. Geometrically, one information granule corresponds directly to one fuzzy linguistic rule; the centres of the MFs m_j^i are defined by calculating the average hyper-box limits of each information granule. Other parameters relating to the MFs (spread parameters $\sigma_j^i \in [\sigma_{j1}^i, \sigma_{j2}^i]$) to automatically generate IT2-FMFs from training data can be defined using three common methods. Common methods are based on histograms, heuristics, and interval type-2 fuzzy C-means (IT2-FCM) [257]. The histogram method uses a suitable parameterised function chosen to model the smoothed histograms of sample data. The heuristic method simply generates the IT2-FMF using heuristics type-1 fuzzy membership function (T1-FMF) and a scaling factor. The IT2-FCM method is the derived formulas of the IT2-FMFs similar to the well-known fuzzy C-means clustering (FCM) algorithm. A detailed description of each method is discussed in [257]. In this research work, the heuristic method based on Gaussian membership function is used to represent the distribution of the training data due to its simplicity.

The spread parameters for type-1 Gaussian membership functions σ_j^i is estimated by the method proposed in [258] and hence its calculation is done via the following equation:

$$\sigma_j^i = \frac{1}{r}\left(\sum_{k=1}^{r} \left\| m_k^i - m_j^i \right\|^2\right)^{\frac{1}{2}} \qquad 5\text{-}15$$

where m_k^i are the r-nearest neighbours of centroid m_j^i in the dimension i. A suggested value for r is 2.

Once a T1-FMF is selected to represent a given data set, the IT2-FMFs are obtained by scaling the T1-FMFs by a factor k_v between 0 and 1. The factor k_v controls the interval between the LMF and UMF of the FOU in which the width parameters vary as follows:

For LMF

$$\sigma_{j1}^i = k_v \, \sigma_j^i \quad \text{5-16}$$

And for UMF

$$\sigma_{j2}^i = \frac{1}{k_v} \, \sigma_j^i \quad \text{5-17}$$

The scaling factor is constrained [259] in the range $k_v \in [0.3, 1]$. The factor also controls the area of the FOU. The smaller the k_v, the larger the FOU, which insinuates the greater uncertainty in the IT2-FMFs. The FOU adds an extra degree of freedom in the T2-FLSs to account for uncertainty.

5.2.4. PARAMETRIC OPTIMISATION

Once the initial structure of the IT2-RBF-NF model is constructed and its initial parameters m_j^i, $\sigma_j^i \in [\sigma_{j1}^i, \sigma_{j2}^i]$, and $[\omega_l^i, \omega_r^i]$ are estimated, a number of training algorithms can be utilised to parametrically tune the model and optimise its structure. Among them is the adaptive-BEP learning algorithm which is proven in the past to be very efficient in optimising similar granular based models [35]. In the BEP learning algorithm [260], the updating of the IT2-RBF-NF parameters depends on the status of x_k as compared $\sigma_{j1}^i, \sigma_{j2}^i$ and m_j^i as well as the status of rule i as compared to L and R calculated during the type-reduction stage. The conventional BEP learning algorithm requires large number of iterations, each iteration involves a large amount of computation to converge and also often gets stuck into a local minima [254, 260]. To circumvent this issue, the adaptive version of BEP is utilised in previous study [35], where the learning rate and momentum factor of the learning algorithm are adjusted during the optimisation process in order to enhance its convergence properties. By using the BEP, for K input-output training data points $(\vec{x}_k: y_k^d), k = 1, \dots, K$. Hence, a performance index PI or cost function (error function) can be defined as follows:

$$PI = \frac{1}{K}\sum_{k=1}^{K} e_k^2 \tag{5-18}$$

$$e_k = (y_k - y_k^d) \tag{5-19}$$

The rule update for the IT2-RBF-NFS's parameter θ^i can be written as:

$$\theta^i(iter + 1) = -\beta\frac{\partial e_k}{\partial \theta^i} + \gamma\theta^i(iter) \tag{5-20}$$

where

$$e_k = \frac{1}{2}\left(y_k - y_k^d\right)^2 \tag{5-21}$$

The error function derivatives with respect to parameter θ^i can be calculated by using the chain rule as follows:

$$\frac{\partial e_k}{\partial \theta^i} = \frac{\partial e_k}{\partial y_k}\frac{\partial y_k}{\partial y_l}\frac{\partial y_l}{\partial \theta^i} + \frac{\partial e_k}{\partial y_k}\frac{\partial y_k}{\partial y_r}\frac{\partial y_r}{\partial \theta^i} \tag{5-22}$$

where

$$\frac{\partial e_k}{\partial y_k} = (y_k - y_k^d) \tag{5-23}$$

$$\frac{\partial y_k}{\partial y_l} = \frac{\partial y_k}{\partial y_r} = \frac{1}{2} \tag{5-24}$$

$$\frac{\partial e_k}{\partial \theta^i} = \frac{1}{2}\left(y_k - y_k^d\right)\left[\frac{\partial y_l}{\partial \theta^i} + \frac{\partial y_r}{\partial \theta^i}\right] \tag{5-25}$$

$$\frac{\partial y_l}{\partial \theta^i} = \frac{\partial y_l}{\partial \overline{f}^i}\frac{\partial \overline{f}^i}{\partial \theta^i} + \frac{\partial y_l}{\partial \underline{f}^i}\frac{\partial \underline{f}^i}{\partial \theta^i} \tag{5-26}$$

$$\frac{\partial y_r}{\partial \theta^i} = \frac{\partial y_r}{\partial \overline{f}^i}\frac{\partial \overline{f}^i}{\partial \theta^i} + \frac{\partial y_r}{\partial \underline{f}^i}\frac{\partial \underline{f}^i}{\partial \theta^i} \tag{5-27}$$

$$\frac{\partial \overline{f}^i}{\partial \theta^i} = \left[\prod_{j=1}^{n} \overline{\mu}_{A_j^{\sim i}}\right] * \frac{\partial \overline{\mu}_{A_j^{\sim i}}}{\partial \theta^i} \qquad 5\text{-}28$$

$$\frac{\partial \underline{f}^i}{\partial \theta^i} = \left[\prod_{j=1}^{n} \underline{\mu}_{A_j^{\sim i}}\right] * \frac{\partial \underline{\mu}_{A_j^{\sim i}}}{\partial \theta^i} \qquad 5\text{-}29$$

According to [35, 260], when computing the derivatives of y_l and $y_l$with respect to the parameters of the antecedent and consequent parts it is necessary to know their locations. Therefore, the derivatives can be computed as follows:

$$\frac{\partial y_l}{\partial \overline{f}^i} = \begin{cases} \dfrac{(\omega_l^i - y_l)}{(\sum_{i=1}^{L} \overline{f}^i + \sum_{i=L+1}^{M} \underline{f}^i)}, & i \le L \\ 0, & i > L \end{cases} \qquad 5\text{-}30$$

$$\frac{\partial y_l}{\partial \underline{f}^i} = \begin{cases} \dfrac{(\omega_l^i - y_l)}{(\sum_{i=1}^{L} \overline{f}^i + \sum_{i=L+1}^{M} \underline{f}^i)}, & i > L \\ 0, & i \le L \end{cases} \qquad 5\text{-}31$$

$$\frac{\partial y_r}{\partial \overline{f}^i} = \begin{cases} \dfrac{(\omega_r^i - y_r)}{(\sum_{i=R+1}^{M} \overline{f}^i + \sum_{i=1}^{R} \underline{f}^i)}, & i > R \\ 0, & i \le R \end{cases} \qquad 5\text{-}32$$

$$\frac{\partial y_r}{\partial \underline{f}^i} = \begin{cases} \dfrac{(\omega_r^i - y_r)}{(\sum_{i=R+1}^{M} \overline{f}^i + \sum_{i=1}^{R} \underline{f}^i)}, & i \le R \\ 0, & i > R \end{cases} \qquad 5\text{-}33$$

Let θ^i to be m_j^i to calculate the partial derivatives with respect to centre estimation: From Eqs. 5-2, 5-28 and 5-29

$$\frac{\partial \underline{\mu}_{A_j^{\sim i}}}{\partial m_j^i} = \frac{(x_j - m_j^i)N\left(m_j^i, \sigma_{j1}^i; x_j\right)}{(\sigma_{j1}^i)^2} \qquad 5\text{-}34$$

$$\frac{\partial \overline{\mu}_{A_j^{\sim i}}}{\partial m_j^i} = \frac{\left(x_j - m_j^i\right)N\left(m_j^i, \sigma_{j2}^i; x_j\right)}{(\sigma_{j1}^i)^2} \qquad 5\text{-}35$$

The update rule for the centre estimation can be obtained by substituting Eqs. 5-26, 5-27, 5-28, 5-29, 5-30, 5-31, 5-32, 5-33, 5-34, and 5-35 into 5-25 and then into Eq. 5-37.

$$\Delta m_j^i(iter+1) = -\beta \frac{\partial e_k}{\partial m_j^i} + \gamma \Delta m_j^i(iter) \quad 5\text{-}36$$

For the update rule for the width parameter estimation:

For σ_{j1}^i, from Eqs. 5-2, 5-28 and 5-29

$$\frac{\partial \underline{\mu}_{\tilde{A}_j^i}}{\sigma_{j1}^i} = (x_j - m_j^i) N(m_j^i, \sigma_{j1}^i; x_j), \quad \frac{\partial \overline{\mu}_{\tilde{A}_j^i}}{\sigma_{j1}^i} = \quad 0 \quad 5\text{-}37$$

Substituting Eqs. 5-26, 5-27, 5-28, 5-29, 5-30, 5-31, 5-32, 5-33 and 5-34 into 5-25 and then into Eq. 5-37.

$$\Delta \sigma_{j1}^i(iter+1) = -\beta \frac{\partial e_k}{\partial \sigma_{j1}^i} + \gamma \Delta \sigma_{j1}^i(iter) \quad 5\text{-}38$$

For σ_{j2}^i, from Eqs. 5-2, 5-28 and 5-29

$$\frac{\partial \overline{\mu}_{\tilde{A}_j^i}}{\sigma_{j2}^i} = \quad (x_j - m_j^i) N(m_j^i, \sigma_{j2}^i; x_j), \quad \frac{\partial \underline{\mu}_{\tilde{A}_j^i}}{\sigma_{j2}^i} = 0 \quad 5\text{-}39$$

Substituting Eqs. 5-26, 5-27, 5-28, 5-29, 5-30, 5-31, 5-32, 5-33 and 5-39 into 5-25 and then into Eq. 5-38.

$$\Delta \sigma_{j2}^i(iter+1) = -\beta \frac{\partial e_k}{\partial \sigma_{j2}^i} + \gamma \Delta \sigma_{j2}^i(iter) \quad 5\text{-}40$$

Similarly the corresponding rule update form $[\omega_l^i, \omega_r^i]$ can be obtained from the following equations:

The update rule for the output weight estimation:

$$\Delta \omega_l^i(iter+1) = -\beta \frac{\partial e_k}{\partial \omega_l^i} + \gamma \Delta \omega_l^i(iter) \quad 5\text{-}41$$

$$\Delta \omega_r^i(iter+1) = -\beta \frac{\partial e_k}{\partial \omega_r^i} + \gamma \Delta \omega_r^i(iter) \quad 5\text{-}42$$

where K is the total number of training data points, β learning rate, γ momentum factor, e_k is the training error of the kth data point, $iter$ is the iteration number index, y_k^d is the kth true data point, and y_k is the kth model's output.

The performance index PI is used for the continuous adaptation of the algorithm as follows:

- If $PI(iter+1) \geq PI(iter)$ then $\beta(iter) = h_d\beta(iter), \gamma(iter+1)$
- If $PI(iter+1) < PI(iter)$ and $\left|\frac{\Delta PI}{PI(iter)}\right| < \delta$ then $\beta(iter+1) = h_i\beta(iter), \gamma(iter+1) = \gamma_0$

If $PI(iter+1) < PI(iter)$ and $\left|\frac{\Delta PI}{PI(iter)}\right| \geq \delta$ then $\beta(iter+1) = \beta(iter), \gamma(iter+1) = \gamma(iter)$ 5-43

where δ is the threshold for the rate of the relative performance index, and h_i and h_d are the increasing and decreasing factors respectively. The performance index PI follows the behaviour of a RMSE energy function where the following constraints are imposed:

$$0 < h_d < 1$$

$$h_i > 1 \qquad 5\text{-}44$$

5.3. FRICTION STIR WELDING PRELIMINARY DATA PRE-PROCESSING AND ANALYSIS

As already stated in Chapter 3 and due to the complexity and non-linearity nature of Friction Stir Welding process, establishing relationship between the process conditions (inputs) and internal process variables and then relate them to the final post-weld properties is of paramount importance for the manufacturing industry. Therefore, studying these correlations could be beneficial to design a practical, safe, and optimal FSW process. In addition, real-time online monitoring of the internal FSW variables could provide a source of information on the underlying process conditions and expose possible causes of defect formation.

Some important recent findings [194] [42] indicate that the frequency spectra of the tool feedback forces in X, Y, and Z axes and it was concluded that the frequency spectra of the feedback forces is more likely to contain useful information about the weld quality. Therefore, this section is devoted to cover the steel friction stir welding experimental trials including, the process parameters, the parent material, the welding tool, the interval process variables, the experimental settings, the data acquisition, frequency analysis of the internal process variables, statistical correlation analysis and the post-weld tests for quantifying the weld quality.

5.3.1. PROCESS PARAMETERS, MATERIAL, TOOL, EXPERIMENTAL TRIALS, AND DATA ACQUISITION

Friction Stir Welding is a multi-input and multi-output (MIMO) system which involves process parameters (tool rotational speed, welding speed, plunge depth, dwell time and tilt angle), and process conditions (tool geometry, machine stiffness, parent materials properties) [191]. These process parameters determine the internal process variables (axial (downward) force F_z, forward (traverse) force Fx, side force Fy, spindle torque, temperature, tool power, heat generation, etc.). All these parameters and internal process variables influence the thermo-mechanical state in the matrix, e.g. velocities, strains, temperatures and stresses. Furthermore, they influence the sliding/sticking or partial sliding/sticking condition at the tool/matrix interface [160] [38, 162, 261].

5.3.1.1. PROCESS PARAMETERS

The tool rotational speed (rpm), the welding speed (mm/min), axial pressure, the tool design, and the tilt angle of the tool are the main independent variables usually employed to control the FSW process [139, 140]. The tool rotational speed results in stirring of material around the tool's pin while the translation of the tool moves the stirred material from the advancing side to the retreating side of the tool's pin. The applied axial pressure on the tool also has influence on the quality of the weld. Meaning that a very high pressure results in overheating and thinning of the joint, while a very low pressure results in insufficient heating and voids [140, 143]. The tool's tilt angle is also an important process parameter, especially to aid producing welds with "smooth" tool shoulders. It is measured with respect to the parent material surface [137].

Fig. 5.3 presents a systematic example of FSW parameters, it is clear that at the beginning of the process the tool rotational speed in clockwise or counter-clockwise direction and welding speed (feed rate) has to be small to generate sufficient heat to plasticise the material in order to avoid tool fracture and void formation. As the thickness of the parent material increases, the tool rotational speed has to be increased to match this increase in the heat conduction away from the weld line until a stabilised temperature is obtained [262]. This can be clearly observed in the temperature history of the parent material in Fig. 5.4. The tool rotational speed is regarded as one of the most momentous process parameters.

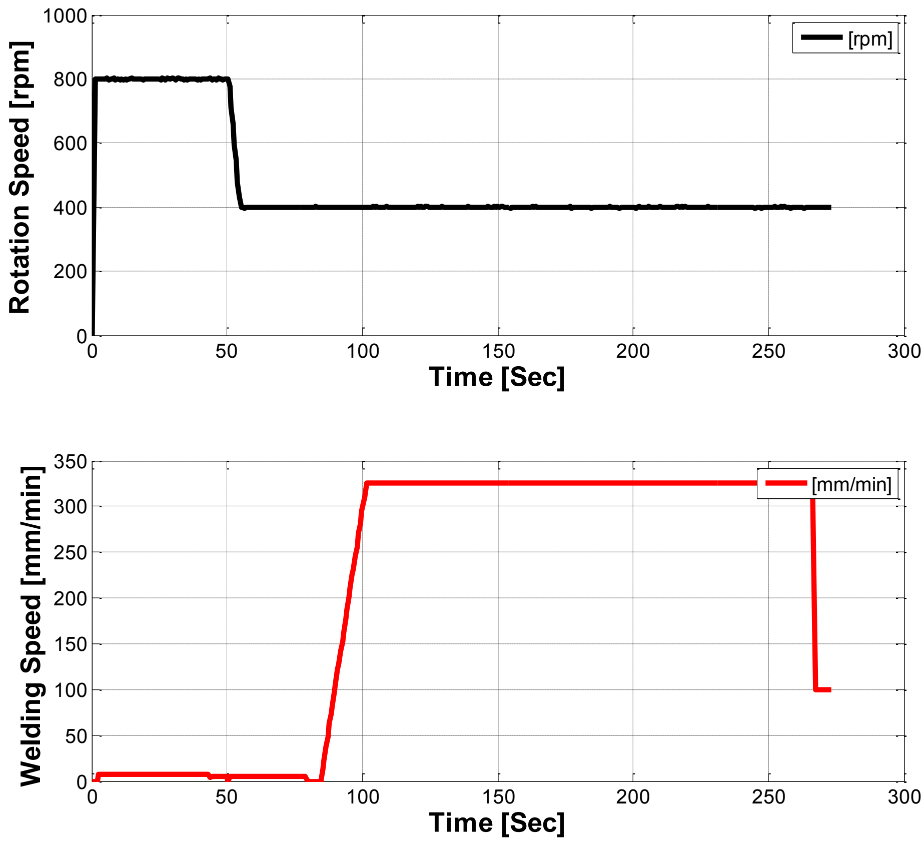


Figure 5.3. Example of FSW input process parameters for DH36 butt weld, 6 mm thick welded via pcBN tool with tool rotational speed of 400 rpm and welding speed of 325 mm/min.

5.3.1.2. INTERNAL PROCESS VARIABLES

The main independent variables in the FSW determine the peak temperature, axial force (F_z), traverse force (F_x), the torque required by the spindle to maintain the rotation rate, and the power features of the process. As the tool rotational speed and axial pressure increase, the temperature increases significantly. While the temperature decreases slightly with increasing the welding speed. The tool rotational speed is used to stir and mix the material around the rotating toot's pin, and then the translation of tool shifts the stirred material around the tool's pin from the advancing side to the retreating side of the tool's pin, and completes the welding process. Higher temperatures can be generated with higher tool rotational speed, this is due to higher friction heating. Thus, more intense stirring and mixing of parent material. The mechanism of heat generated in FSW is based on generating friction between the tool and parent material. The heat generated is conducted to both the parent material as well as the tool. The amount of heat conducted into the plate dictates the temperature distribution of the parent material. This in turn affects the material flow, microstructure evolution, and mechanical properties of the weld region [38, 162, 261].

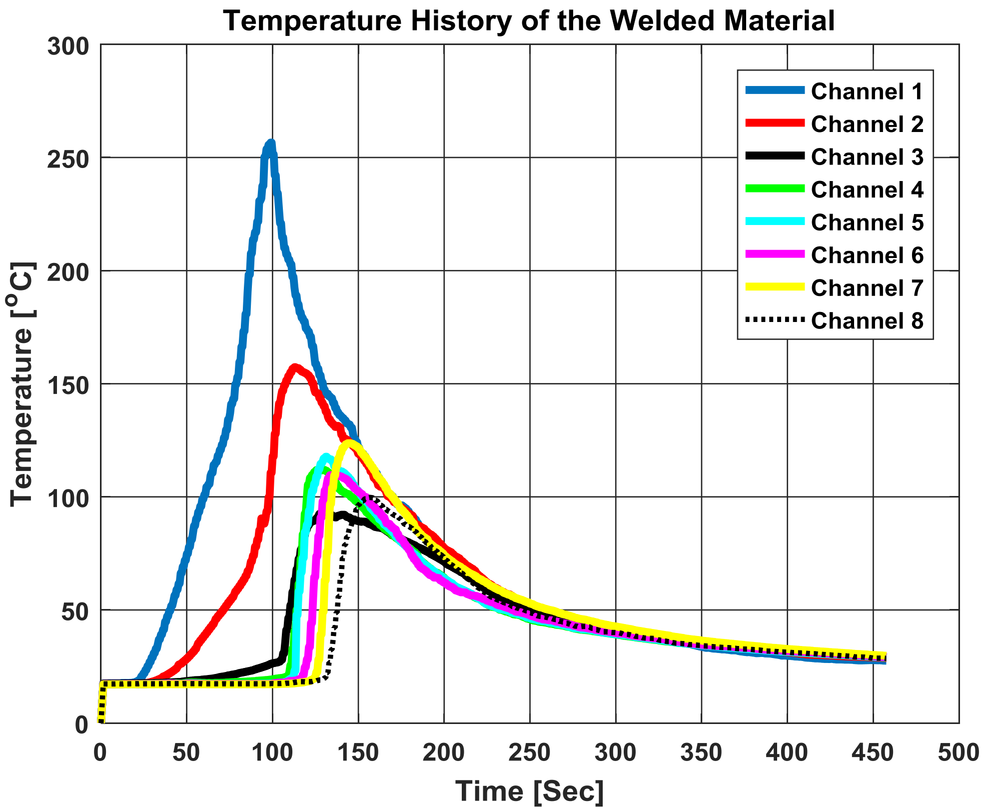


Figure 5.4. Example of FSW temperature history recorded from 8-therocouples: DH36 butt weld, 6 mm thick welded via pcBN tool with tool rotational speed of 400 rpm and welding speed of 325 mm/min.

The axial (downward) force F_z significantly depends on the tool plunge depth. Insufficient axial force results in insufficient heat at the tool/material interface, whereas excessive axial force produces excessive shoulder penetration into the material, which causes excessive flash at the edges of the shoulder [263]. It was also reported that [264], the axial force influences the surface texture and fatigue performance of FSW. The side force Fy is the force from the clamping system that holds and prevents parent materials from being pushed away by the force generated by the rotating tool (Fz). The traverse force Fx increases with the thickness of the parent material because the dimensions of the tool's pin increases at the same time. Insufficient traverse force causes void formation below the surface on the advancing (front) side of the weld [265].

Different materials result in different welding performance due to their difference in chemical composition and mechanical properties. Fig. 5.5 shows an example data record for the spindle torque and feedback forces (F_x and F_z) of butt weld of 6 mm DH36 steel grade using pcBN tool with tool rotational speed of 400 rpm and welding speed of 325 mm/min. The four different welding phases can be identified (plunging, dwelling, welding, and pulling out). Once the tool is plunged into the parent material, the feedback forces (axial and traverse forces) and spindle torque rise steady during the first part of the plunge phase. The initial smooth rise in the feedback forces and spindle torque is caused by the tool's pin permeating the matrix. In the dwelling phase, the tool can no longer maintain in a vertical displacement and it starts to rotate until the surrounding material reaches a stabilised temperature (below the melting point of the parent material) in order to prevent the tool wear. As the tool travels along the joint line, the material cools and solidifies resulting in a solid state joining of materials. After the initial response period, the feedback forces and spindle torque remain relatively stable, thus indicating the process has reached its stability (welding phase). It is within this phase where the process is stable and useful welds are made.

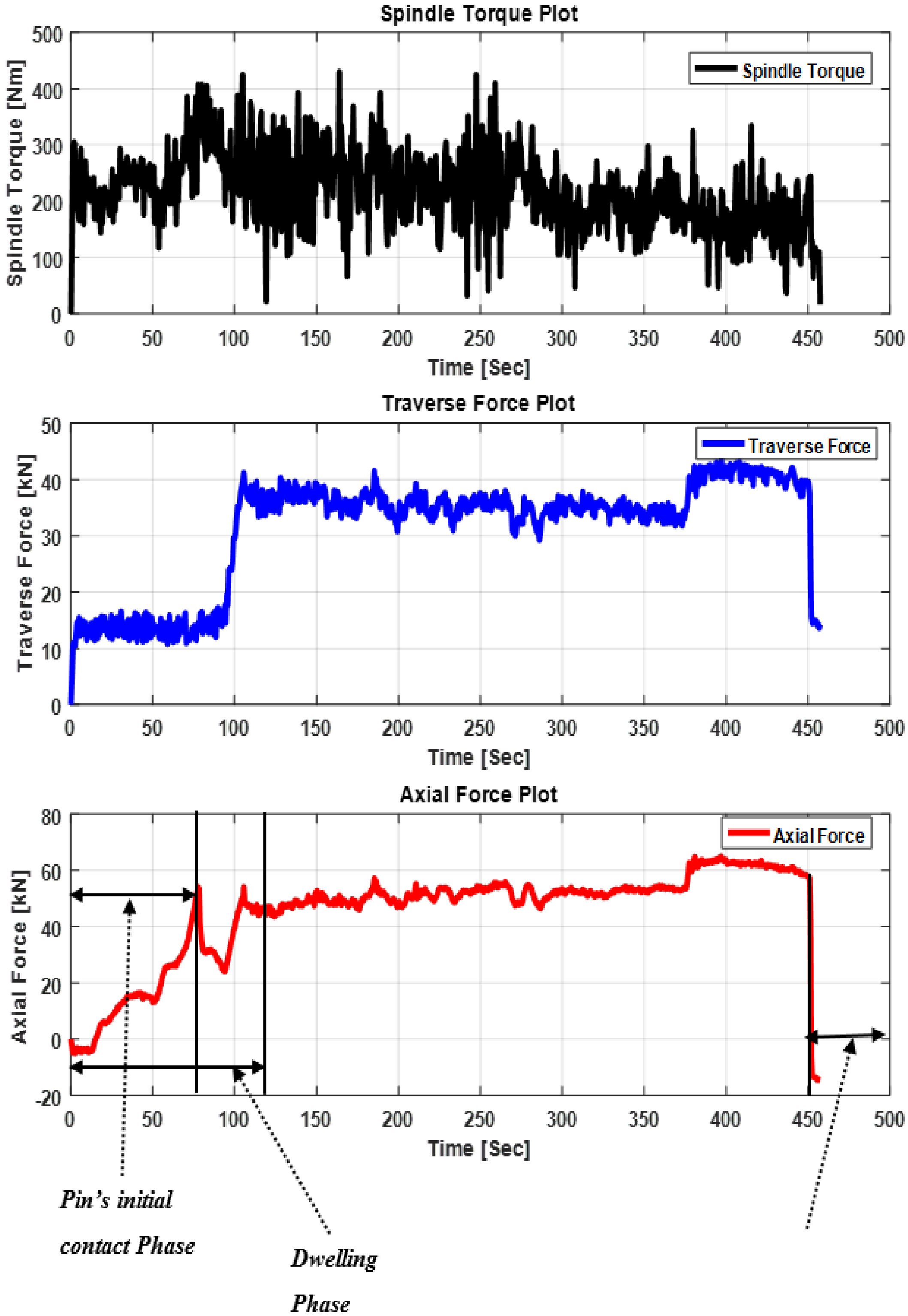


Figure 5.5. Example of FSW internal process variables: DH36 butt weld, 6 mm thick welded via pcBN tool with tool rotational speed of 400 rpm and welding speed of 325 mm/min.

5.3.1.3. DATA ACQUISITION

Since the internal process variables are those features that can be monitored during the FSW and can provide a rich source of information about the undergoing process and the weld. A collection of data (25 trials) on the axial force (Fz) and traverse force (Fx) under different process conditions have been provided by TWI Ltd., Technology Centre (Yorkshire), United Kingdom, for the welding of two similar DH36 steel plates 6 mm thick. This material is used for shipbuilding applications. All the welds were made on the same material with a pcBN weld tool. The welding trials were made using several different levels of tool rotational speed, and welding speed in order to obtain experimentally suitable welding conditions (see Fig. 5.6).

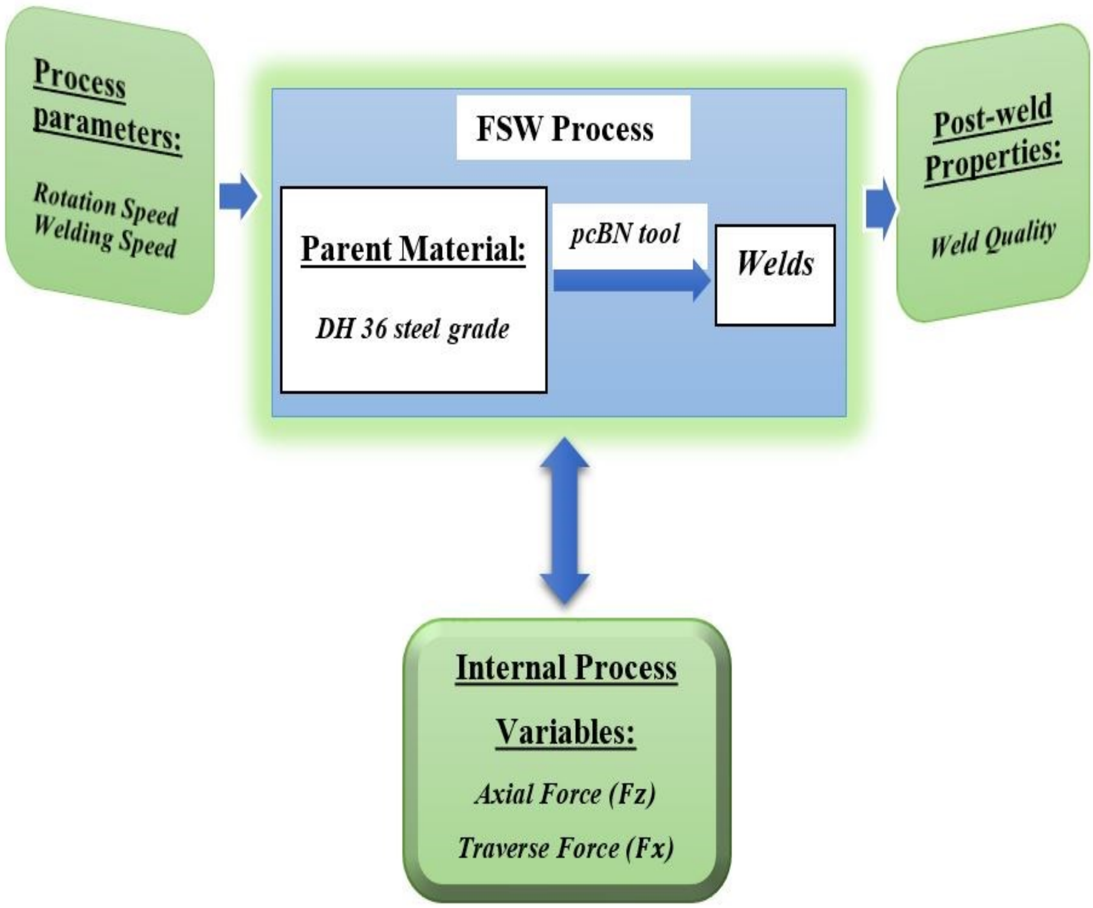


Figure 5.6. Essential parameters, internal process variables and post-weld properties involved in FSW.

The data acquisition rate (sampling rate) for feedback force signals was 1 Hz. The studies in [42, 188, 194] reported that the amplitude of the low frequency components of feedback forces (Fz and Fx) has some direct correlation with the weld quality in the same process, whilst higher acquisition rate offers no significant benefits. In addition, one of the objectives in this research work is to build a data-driven computational framework that can be employed in real-time, inline to the process, with minimum computational cost; using low sampling rate also aids with this objective. Due to the specific requirements of obtaining experimentally suitable welding conditions, the overall data space for the FSW process is very complex and sparse. There are areas of low density and areas of high density (i.e. process operation envelope) (see Fig. 5.7). A good model of the process should be able to handle the complexity and sparsity associated with the data space and provide reliable predications in all density areas [158].

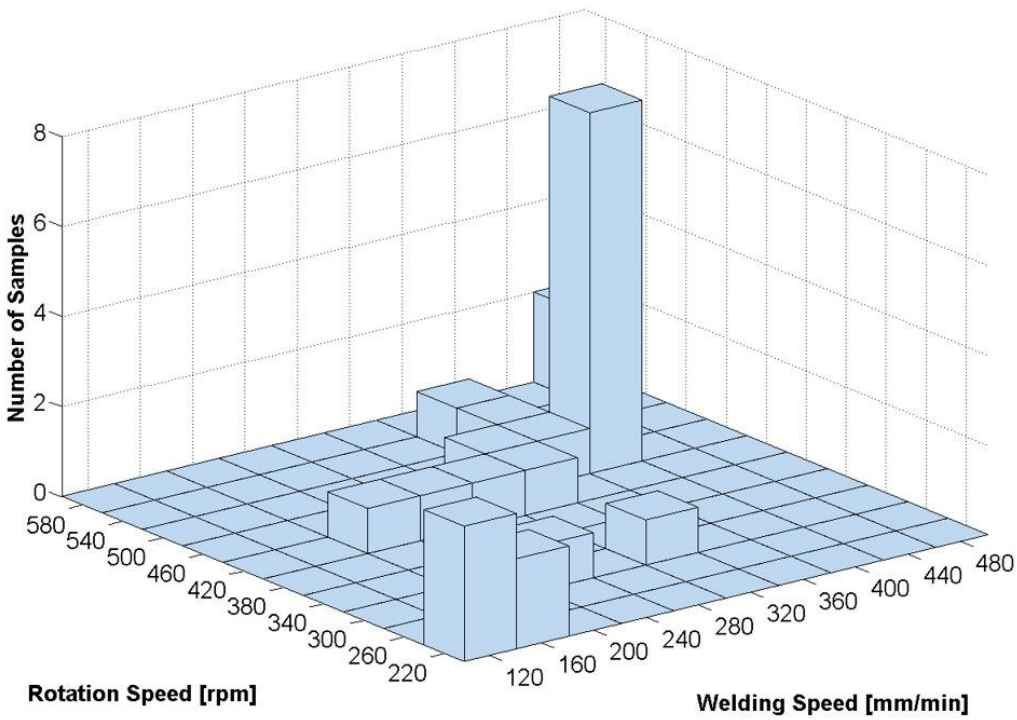


Figure 5.7. FSW data density of experimental trials.

5.3.2. FREQUENCY ANALYSIS

Frequency analysis is a common and powerful tool for analysing unwanted variations (energy) in a time series data. By decomposing a time series signal into frequency domain, it would be easier to define both the predominant modes of variability and how those modes vary in time. The process of decomposing a time series into simpler Fourier series and analysing the series is called Fourier analysis. In signal processing, the Fourier transform (FT) often decomposes a time series into simple sinusoidal components of different frequencies known as frequency spectrum. Therefore, the Fourier analysis is also considered as frequency analysis. The FT's advantage lies in its ability to analyse a signal in the time domain for its frequency component [266]. The Fourier transform works by first converting a function in the time domain into a function in the frequency domain. The time domain signal can then be analysed for its frequency component because the Fourier coefficients of the converted function represents the contribution of each sine and cosine function at each frequency component. To estimate the FT of a function from a finite length of its sampled points, the discrete Fourier transform (DFT) is often employed [267]. The sampled points are intended to be typical of what the original signal looks like at all other times. A variety of techniques can be used to calculate the discrete Fourier transform. One of them is the fast Fourier transform (FFT) algorithm [268]. FFT is a non-parametric technique which is used to quickly evaluate the power spectral density (PSD) of a signal from the signal itself.

For FSW, the research studies in [269-271] investigated the tool forces in X, Y, Z planes (Fx, Fy and Fz) and it was revealed that useful information about the weld quality relevant to 'wormhole' defects is more likely to be obtained from the frequency spectra of the feedback forces. A wormhole defect is considered to be one of the most common, and difficult to detect, FSW defects and it refers to a continuous void underneath the weld surface. The study in [272] also found that the feedback forces oscillations can be useful for detecting the gaps at the weld line. The related study in [194] also highlights the importance of the frequency spectra of the force X and Y planes in detecting a wormhole defect. Many FSW related studies [273-276] showed that the plasticised material flow is highly periodic in its behaviour to the tool movement. For that reason, the stability of the spindle frequency (tool rotational speed) oscillation in the traverse force (Fx) and lateral (side) force (Fy) is a valid indication of a good weld without wormhole defect. Therefore, the frequency spectra of the feedback forces could provide a rich source of information for analysis. In this work, the Fast Fourier Transform (FFT) algorithm was used to determine the DFT for the axial (Fz) and Traverse (Fx) force signals. The DFT can either be computed for the entire signal in time domain or else the signal can be windowed first. Windowing the signal is the process of selecting segments from the time domain signal in order to get a perspective of what the frequency spectrum is at a particular moment in time. In this research work, all the signals in time domain are windowed for 4 seconds during the welding phase. The sampling rate determines the bandwidth of the FFT, so with a higher sampling rate, higher frequencies can analysed. But this offers no significant benefits as stated earlier.

To evaluate the quality of welds, a visual inspection on the surface of the welds was conducted by an expert. For every single weld, an overall quality index is designed to represent the general situation for weld quality. The value of this index could be 0 or 1, where 0 indicates defect free (good quality) while 1 indicates poor quality (defective weld). More details on the weld quality assessment and the factors that needed to take into consideration while assessing the weld quality are fully presented in Chapter 3.

As an illustration, Figs. 5.8 and 5.9 show the axial force signals and traverse force signals for 8 specimens at tool rotational speed of 450 rpm and welding speed of 400 mm/min respectively. All the force signals are windowed for 4 seconds during the steady state welding phase. Then the DFT of the windowed signals was calculated by using the FFT algorithm. Figs. 5.10 and 5.11 show an example of the DFT of the axial and traverse forces for 8 different specimens respectively. In these figures, it is clear that the defect-free specimens produce smaller frequency-components (amplitudes in the frequency spectra) at the lower frequency spectrum including the DC component (i.e. at zero frequency), whilst the specimens with some defects have larger amplitudes at the lower frequency components. Consequently, the amplitudes of the low frequency components in the frequency spectra may have a correlation with weld quality.

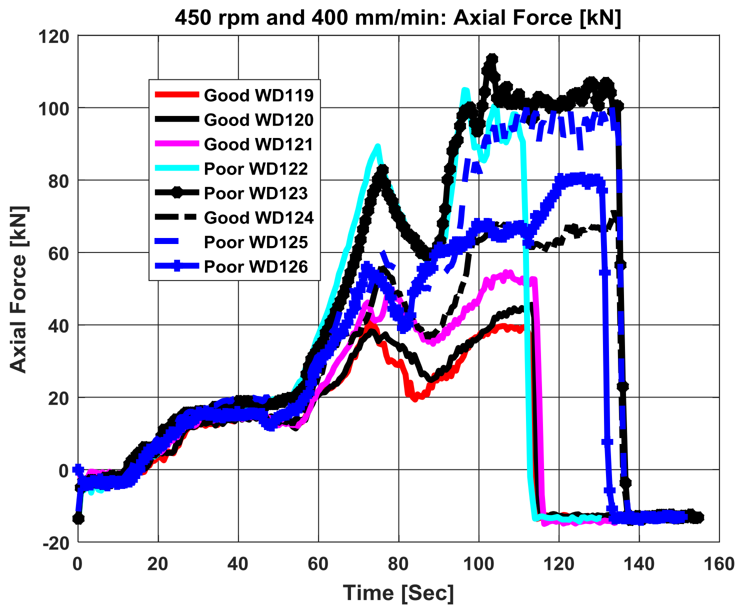


Figure 5.8. Examples of axial force signals recorded for Butt weld of 6 mm thick DH36 steel grade welded via pcBN tool with tool rotational speed 450 rpm and welding speed of 400 mm/min, WD119, WD120, WD121 and WD124 were reported as good welds and WD122, WD123, WD125 and WD126 were reported as poor welds.

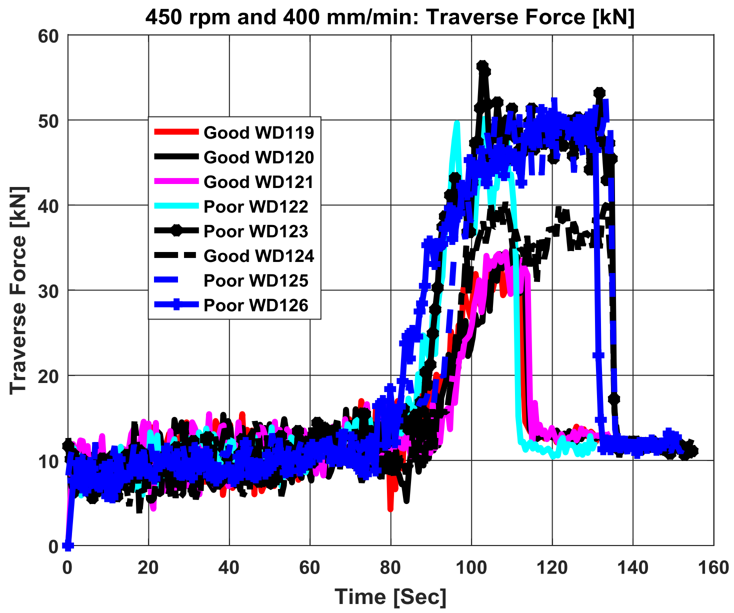


Figure 5.9. Examples of traverse force signals recorded for Butt weld of 6 mm thick DH36 steel grade welded via pcBN tool with tool rotational speed 450 rpm and welding speed of 400 mm/min, WD119, WD120, WD121 and WD124 were reported as good welds and WD122, WD123, WD125 and WD126 were reported as poor welds.

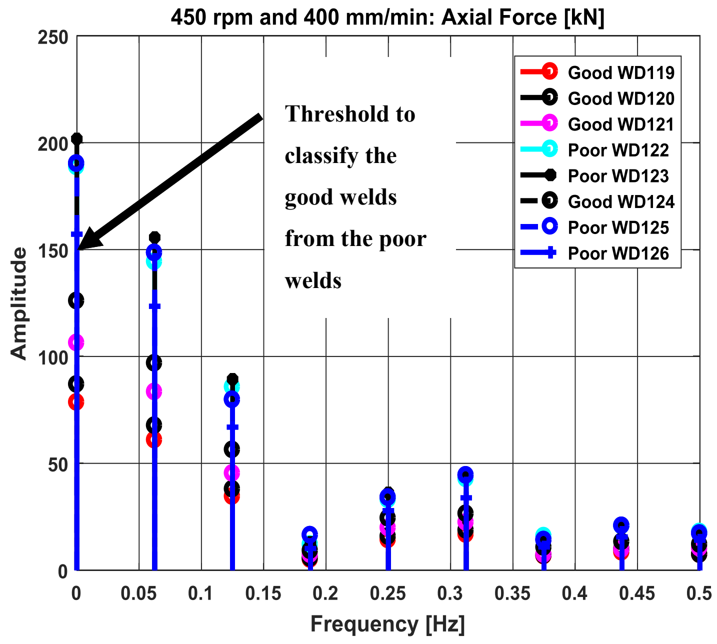


Figure 5.10. Discrete Fourier Transform of axial force signals for Butt weld of 6 mm thick DH36 steel grade welded via pcBN tool with tool rotational speed 450 rpm and welding speed of 400 mm/min: WD119, WD120, WD121 and WD124 were reported as good welds and WD122, WD123, WD125, and WD126 were reported as poor welds.

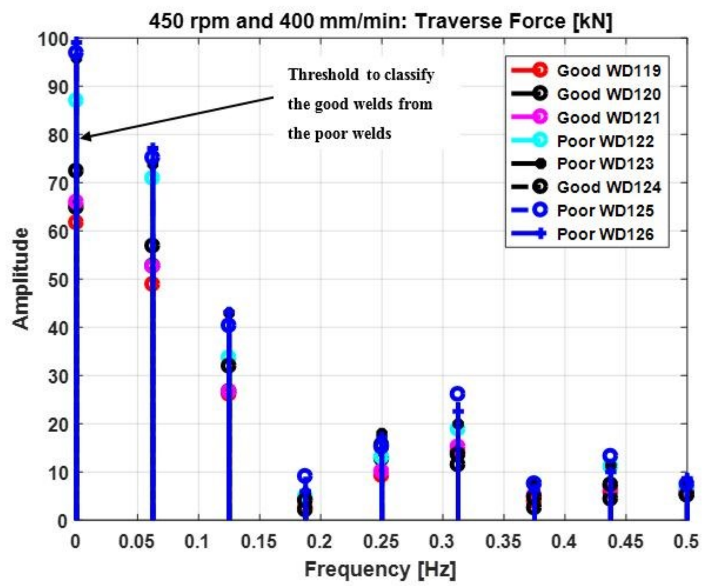


Figure 5.11. Discrete Fourier Transform of traverse force signals for Butt weld of 6 mm thick DH36 steel grade welded via pcBN tool with tool rotational speed 450 rpm and welding speed of 400 mm/min, WD119, WD120, WD121 and WD124 were reported as good welds and WD122, WD123, WD125, and WD126 were reported as poor welds.

5.3.3. CORRELATION AND REGRESSION ANALYSIS

To further understand/investigate the relationship between the quality of the welds produced by FSW and the moving threshold extracted from the frequency spectra of the tool feedback forces, this section presents correlation and regression analysis. Correlation and regression refer to the relationship between two causally correlated variables, the independent variable is the one that is capable of affecting the other, and the dependent variable is the one that is capable of being affected by the other [277]. For instance, in FSW the tool rotational speed will tend to increase the temperature, whereas a change in temperature will not affect the tool rotational speed. In this relationship between tool rotational speed and temperature, the tool rotational speed is the independent variable and the temperature is the dependent variable. The welding parameters are independent variables and both the internal process variables and weld quality are dependent variables. In statistics, the primary measure of linear dependence (correlation) between two variables is the Pearson's product–moment correlation coefficient r [277]. Its range $[-1.0, +1.0]$, where $r = -1.0$ for a perfect negative correlation and $r = +1.0$ for a perfect positive correlation. The midpoint of its range, $r = 0.0$ corresponds to no correlation. Values falling in the range $[-1.0, 0.0]$ represent varying degrees of correlation in the negative direction, while those falling in the range $[0.0, +1.0]$ represent varying degrees of correlation in the positive direction. For purposes of interpretation, the correlation coefficient can be translated into terms of percentages ($i.e., correlation\ in\ percentage = r \times 100$).

In this research work, the correlation coefficient r was employed to study the dependency between the feedback forces (Fz and Fx) and the weld quality. The Pearson's product–moment correlation coefficient r is calculated between the amplitudes of the DC frequency components in the frequency spectra of the feedback forces and the weld quality. As shown in Figs. 4.12 and 4.13 show the scatter plot and regression line of 8 specimens: overall weld quality index (0 indicates defect-free or good quality, whereas 1 indicates poor weld quality) versus the weld quality threshold from the frequency spectra of the axial and traverse force signals respectively. The correlation coefficients r for the frequency spectra of the axial and traverse force signals are 0.93 (93%) and 0.96 (96%) respectively. The high value of r indicates a high linear correlation between the amplitudes in the frequency spectra and the overall weld quality.

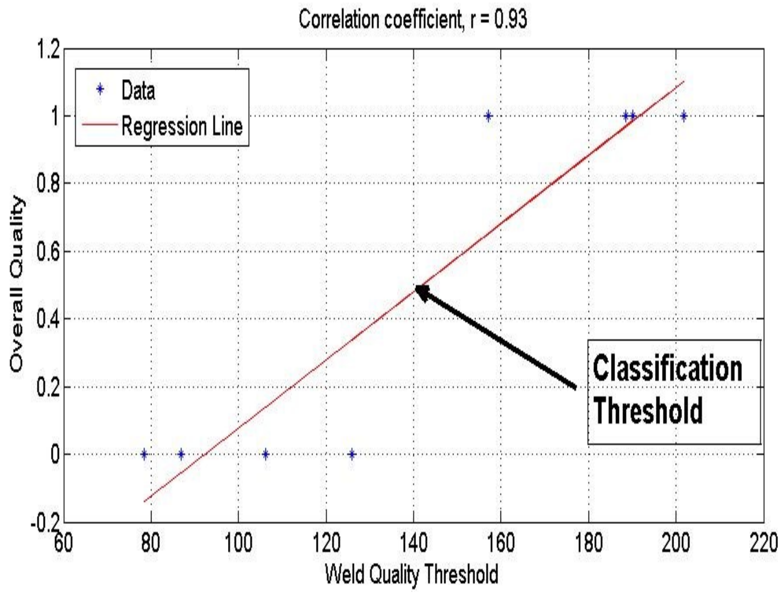


Figure 5.12. Scatter plot and regression line of 8 FSW specimens: overall weld quality index (0 means defect free or good quality while 1 indicates poor weld quality) versus extracted thresholds from frequency spectra of the axial force signals.

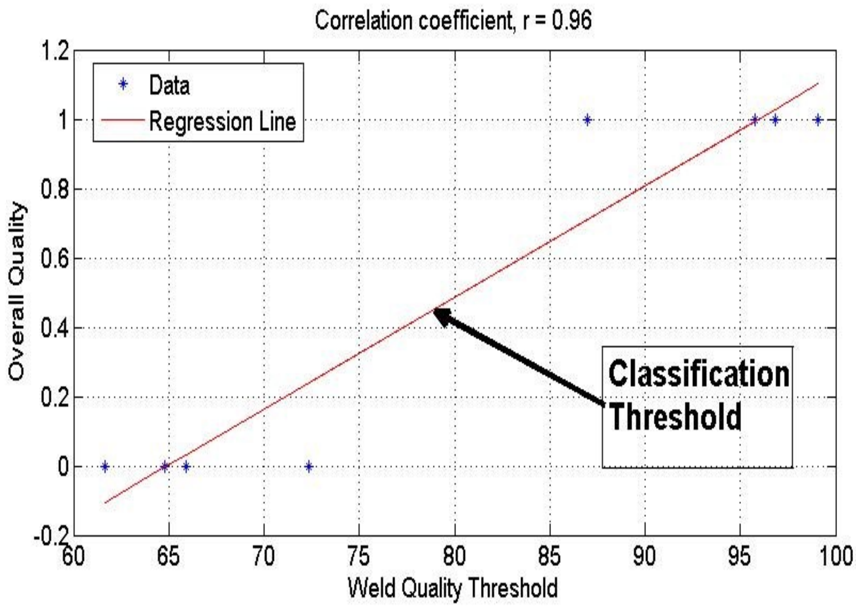


Figure 5.13. Scatter plot and regression line of 8 FSW specimens: overall weld quality index (0 means defect free or good quality while 1 indicates poor weld quality) versus extracted thresholds from frequency spectra of the traverse force signals.

Based on the frequency analysis and statistical correlation analysis, useful information can be extracted from the frequency spectra of the feedback force signals and create a threshold (marker) that can be used to categorise (classify) the good welds from the poor welds. However, in reality this threshold cannot be generalised for different process parameters because different process parameters result in different feedback force signals' behaviour, and therefore different thresholds. The relationship between the change in process parameters and thresholds is non-linear. As is discussed earlier in this chapter, the model-based approach that is used for the monitoring framework relies on the prediction of the non-linear behaviour of the change of this threshold. Successful forecasting of this threshold enables the process operator(s) to classify (discriminate) weld quality.

5.4. A NEW MODEL-BASED REAL-TIME MONITORING OF FRICTION STIR WELDING VIA SPECTRAL ANALYSIS

This section presents a new generalised and systematic model-based real-time monitoring framework relies on the previously IT2-RBF-NF model. The proposed monitoring framework takes the advantage of principle of human-like information granulation in granular computing (GrC) used in Chapter 4 to extract meaningful knowledge out of raw data. The extracted knowledge is utilised to elicit the initial structure of the IT2-RBF-NF model, and the initial structure is then optimised via the adaptive-BEP to improve its performance in predicting in real-time (during welding) quantitative markers of weld quality. Part quality thresholds are extracted from the frequency spectra of the feedback forces (axial (Fz) and traverse (Fx) forces) as shown in Section 5.2. The proposed model relies on a dynamic model that instead of predicting directly the weld quality, it predicts a moving threshold that can be used by the operator(s) as an indicator to discriminate the good welds from the bad welds. Thus, there is no need to re-tune the model when the process condition is changed. The proposed real-time model-based monitoring framework (see Fig. 5.14) is validated in different material, tool, and weld geometry combinations. The first case study is related to shipbuilding grade steel friction stir welding and the second case study is related to the FSW-based joining of an aluminium alloy. The results are provided in the following sections.

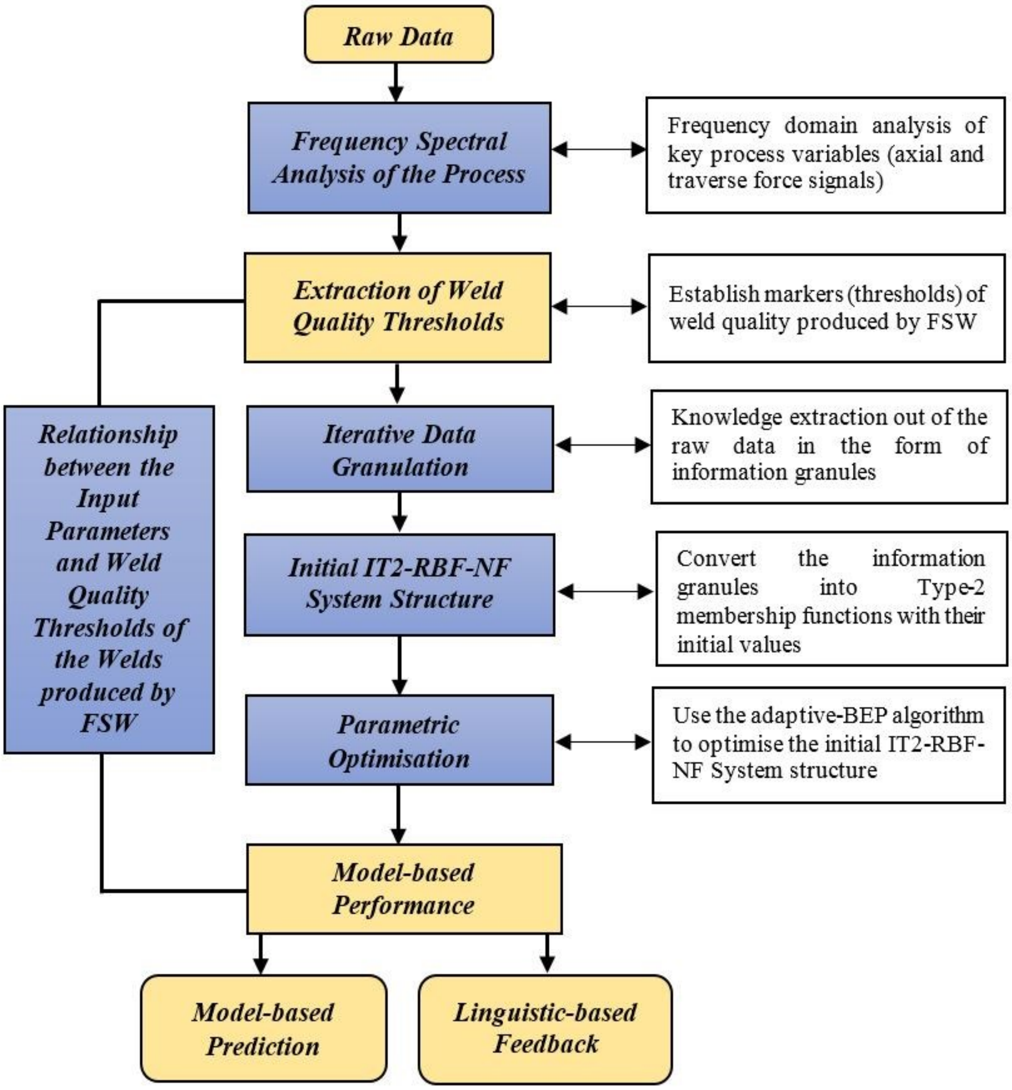


Figure 5.14. Flow chart of the GrC-based IT2-RBF-NF proposed monitoring framework.

5.4.1. APPLICATION TO STEEL FRICTION STIR WELDING

The FSW data set is used consisting of 25 measurements (See Table 5-1). The measurements include the process parameters (welding speed and tool rotation speed) and weld quality thresholds extracted from the frequency spectra of feedback force signals (axial and traverse forces) as described in Section 5.3. During the first stage of the modelling process the data set has been split into two sets, 25 data points to train the model and 5 data points for validation purposes. The training raw data is then compressed into information granules as described Chapter 4 in via the iterative data granulation approach that are true representatives of the process dynamics. The compressed information granules are used to construct the initial structure of the model (number of rules and initial parameters $m_j^i, \sigma_j^i \in [\sigma_{j1}^i, \sigma_{j2}^i]$, and $[\omega_l^i, \omega_r^i]$). The initial GrC-IT2-RBF-NF model is trained via the adaptive-BEP algorithm as shown in the previous Section. After a number of systematic simulations (increased/reduced the number of granules), it was established that the predictive performance of the model was acceptable in the range between 3 to 7 linguistic rules.

The predictive performance of the model having 3 rules is less compared to the model having 5 rules. In contrast to the model having 7 linguistic rules, the predictive performance is better but not a substantial predictive performance improvement compared with the IT2-RBF-NF model with 5 linguistic rules. Hence, the 5-rule IT2-RBF-NF model was selected as the best model (optimal number of rules is five) in terms of balancing simplicity (in structure) and good predictive performance. In general, the rule-based human-centric models with more linguistic rules (hence more antecedent and consequent parameters to be optimised) often lead to better predictive performance due to their ability to capture more information about the dynamics of the process (with the danger of over-fitting during the parametric learning phase). However, this results in a lack of simplicity and interpretability, while models with few number of linguistic rules and parameters result in simpler models that facilitate the easier interpretation, albeit with lower predictive performance. In addition, one of the main objectives of this research work was to maintain the overall system structure as simple as possible in order to create a low cost computational paradigm that can be feasible for real-time use.

To avoid over-fitting during the parametric optimisation phase and to establish the best model training regime, the k-fold cross-validation approach ($k = 10$ in this research work) was utilised where the modelling process was repeated ten times. Performance indexes based on RMSE and MAE between the true threshold and model predicted threshold were used to evaluate the performance of the IT2-RBF-NF models developed. For the purpose of comparison and to confirm the appropriateness of the proposed model, a performance analysis was performed against a baseline (multiple linear regression (MLR) Model) and well-known multilayer perceptron neural network (MLP-NN). The performance of the proposed IT2-RBF-NF model is also tested against its type-1 radial basis function neural network (T1-RBF-NN) model counterpart that is described in details in Chapter 4 using the same number of linguistic rules. The proposed monitoring framework is applied to two cases related to axial and traverse forces used as the main internal process variables for the process real-time monitoring. The two cases are detailed below:

Table 5.1 Process paramerers of 25 experimental trials used for modelling (including training and testing data). If the value of the indicator is below the threshold, the weld can be classified as a good weld ortherwise the weld is of poor quality [158].

Weld Sample	Welding Speed [mm/min]	Rotation Speed [rpm]	FFT Threshold Values for Axial Force Signals	FFT Threshold Values for Traverse Force Signals
1	100	200	138.41	53.55
2	125	200	95.35	57.33
3	110	200	71.60	26.01
4	143	200	71.77	24.35
5	200	400	108.36	40.16
6	350	450	115.13	60.37
7	250	300	112.54	69.83
8	500	600	107.10	76.74
9	500	575	175.58	99.12
10	325	400	160.58	83.09
11	275	400	103.03	48.64
12	250	400	79.35	30.47
13	325	450	115.40	64.18
14	400	450	141.52	79.66
15	400	450	141.52	79.66
16	400	450	141.52	79.66
17	400	450	141.52	79.66
18	400	450	141.52	79.66
19	400	450	141.52	79.66
20	400	450	141.52	79.66
21	400	450	141.52	79.66
22	156	200	137.89	51.13
23	350	500	119.03	52.79
24	350	550	119.88	66.79
25	300	300	91.46	42.90

5.4.1.1. TOOL AXIAL FORCE CASE

The following case relating to the axial force case, as it is illustrated in Fig. 5.15 the prediction performance of the 5-rule IT2-RBF-NF model for the training and testing performance respectively. Fig. 5.15 shows the experimentally measured and the corresponding predicted values from the model. The simulation results show a relatively good performance in the prediction of weld quality threshold with more than 80% of the predicted values lie within the 80% confidence limit.

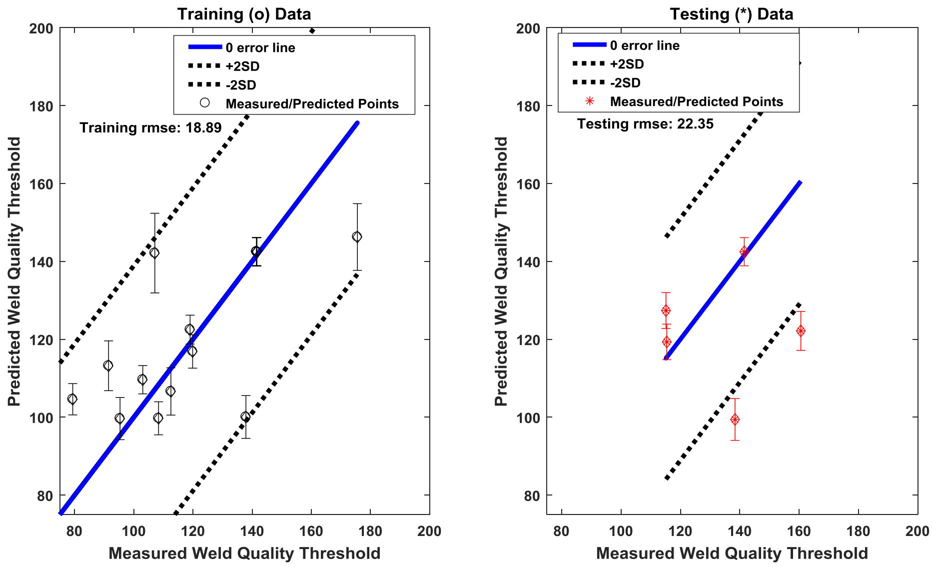


Figure 5.15. The 5-rule IT2-RBF-NF weld quality threshold model for the axial force.

The system's simulated behaviour between the process parameters and the predicted weld quality threshold; the mean IT2-RBF-NF model along with its standard deviation (SD) is depicted in Fig. 5.16. From the three-dimensional surface, it can be seen that the simulated behaviour of the threshold for establishing weld quality produced by FSW is in a non-linear relationship with respect to the process parameters.

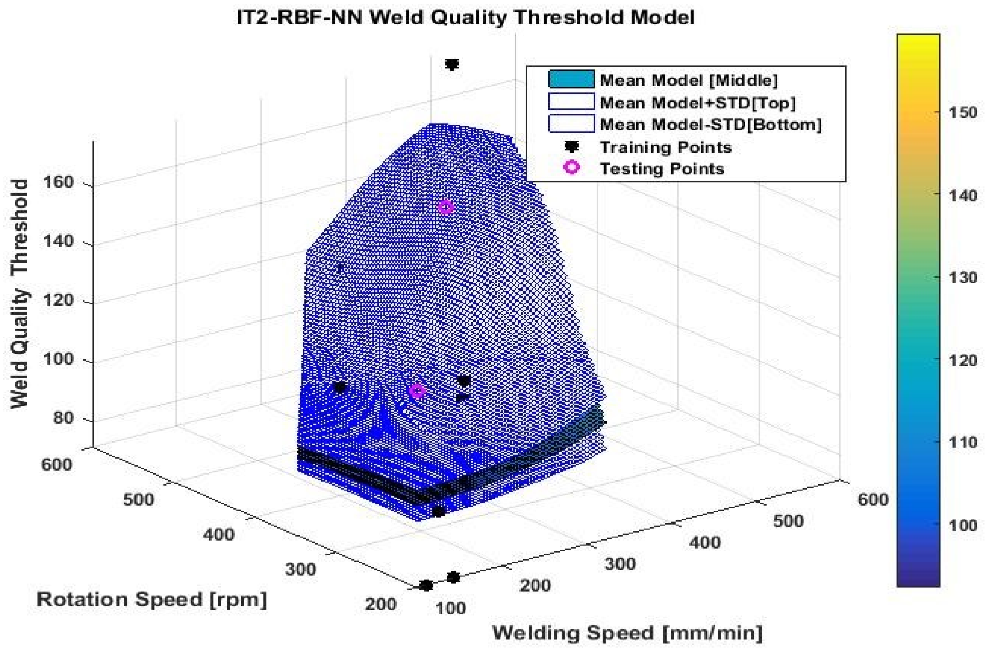


Figure 5.16. The response surface of the 5-rule IT2-RBF-NF weld quality threshold model for the axial force (mean +/- SD).

The improved interpretability of the IT2-RBF-NF model can be demonstrated in the linguistic human-centric rule-based structure generated to forecast the weld quality threshold. Fig. 5.17 shows the fuzzy rule-base of the 5-rule IT2-RBF-NF model (each row represents one rule).

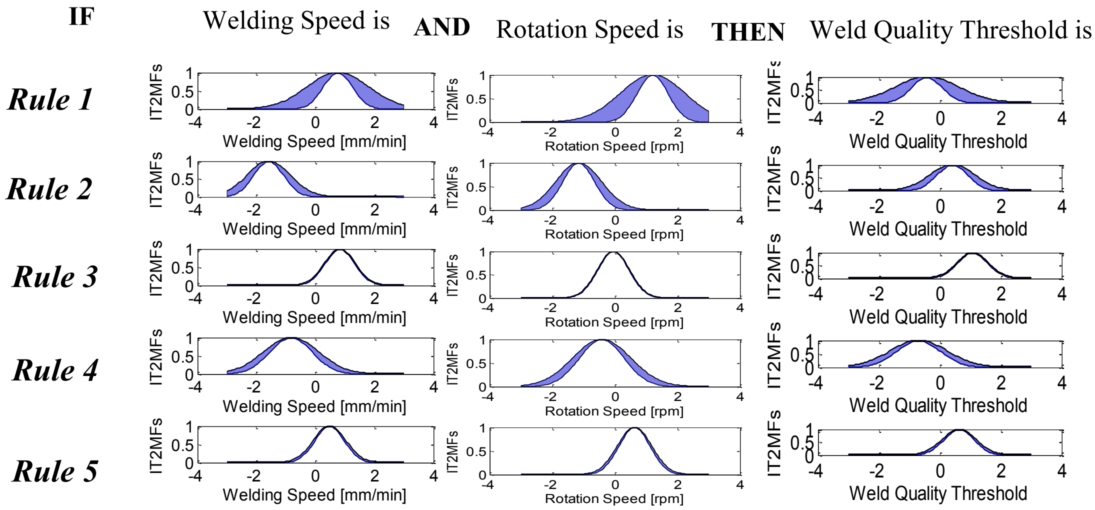


Figure 5.17. The rule-base of the 5-rule IT2-RBF-NFweld quality threshold model for the axial force.

The corresponding linguistic rule-base in Fig. 5.17 can be employed to describe the non-linear relationships between the welding speed, rotation speed and the predicted weld quality threshold produced by FSW, which can be achieved via the linguistic hedges approach from the interval type-2 Fuzzy Logic structure of the model. Such a rule-base structure from the axial force IT2-RBF-NF model can be interpreted in a linguistic format as follows:

Rule 1: IF *Welding Speed* ***is*** *high* ***AND*** *Rotation Speed* ***is*** *very high,* ***THEN*** *Weld Quality Threshold* ***is*** *low*

Rule 2: IF *Welding Speed* ***is*** *very low* ***AND*** *Rotation Speed* ***is*** *very low,* ***THEN*** *Weld Quality Threshold* ***is*** *medium*

Rule 3: IF *Welding Speed* ***is*** *very high* ***AND*** *Rotation Speed* ***is*** *medium,* ***THEN*** *Weld Quality Threshold* ***is*** *Medium*

Rule 4: IF *Welding Speed* ***is*** *low* ***AND*** *Rotation Speed* ***is*** *low,* ***THEN*** *Weld Quality Threshold* ***is*** *very small*

Rule 5: IF *Welding Speed* ***is*** *medium* ***AND*** *Rotation Speed* ***is*** *high,* ***THEN*** *Weld Quality Threshold* ***is*** *high*

A multiple linear regression (MLR) model described in more details in [126] was employed as a baseline for comparison purposes. From the performance analysis of the modelling results obtained, it can be confirmed that the MLR model provides only a basic level of modelling performance (see Tables 5-2 and 5-3). Two more advanced model-based non-linear modelling methods were then used, a MLP-NN model which has been proven to provide good classification accuracy in the same process in the past [194], as well as a type-1 NF system [42], which also has been applied in the past with good prediction performance. In addition, it provides a useful comparison between the proposed type-2 FLS and its type-1 FLS counterpart. The MLP-NN used for comparison purposes consists of two hidden layers, each hidden layer composed of 100 hidden units having the hyperbolic tan transfer function trained via the adaptive-BEP approach. A detailed description about the MLP-NN modelling approach is given elsewhere [278].

Similarly to the previously presented modelling results, the regression line between the experimentally measured and the corresponding predicted weld quality threshold values for the developed MLP (Fig. 5.18) and T1-RBF-NN (Fig. 5.19) models for the axial force, along with summative comparison tables (see Table 5-2 and 5-3) including the RMSE and MAE% for training and testing performance of all the developed models.

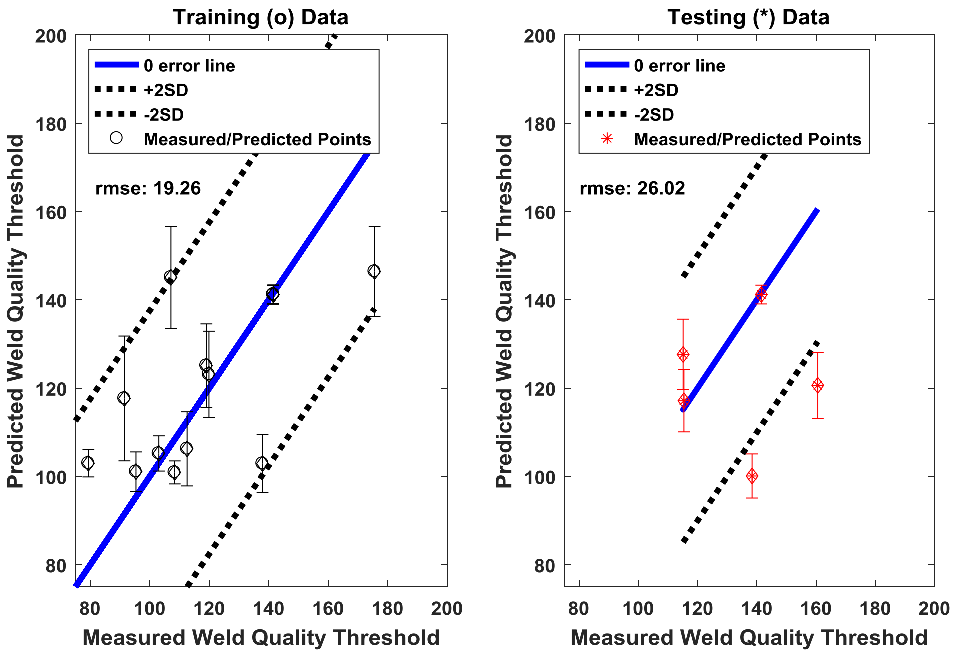


Figure 5.18. The 2-hidden layers, 100 hidden units weld quality threshold model for the axial force.

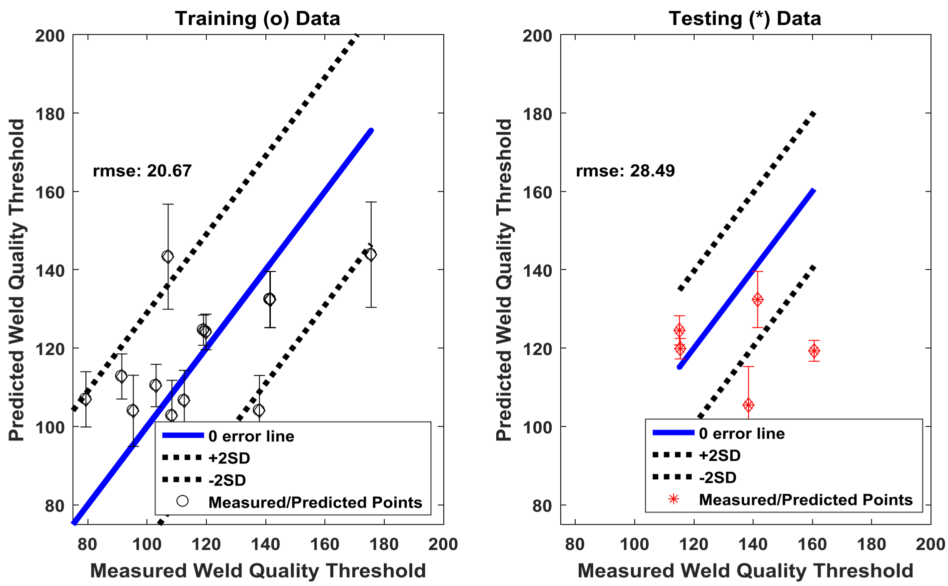


Figure 5.19. The 5-rule T1-RBF-NN weld quality threshold model for the axial force.

Table 5.2 Performance of the multiple linear regression and multilayer perceptron weld quality threshold models for the axial force.

Performance Index	MRL Model		MLP-NN Model	
	Training	Testing	Training	Testing
RMSE±SD	26.20±5.43	31.03±10.93	19.26 ±2.40	26.02 ±3.80
MAE %±SD	20.15±6.12	28.27±9.39	13.16±2.18	13.73±2.31

Table 5.3 Performance of the T1-RBF-NN and IT2-RBF-NF weld quality threshold models for the axial force.

Model Type	**T1-RBF- NN**				**IT2-RBF-NF**			
Number of Rules	**RMSE±SD**		**MAE%±SD**		**RMSE±SD**		**MAE%±SD**	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
3	25.13 ±5.10	32.59 ±7.64	20.66 ±2.96	25.04± 3.73	21.09 ±2.17	24.30 ±3.07	15.43± 1.03	17.11± 1.06
5	20.67 ±2.18	28.49 ±3.18	15.63 ±2.14	20.34± 2.83	18.89 ±0.95	22.35 ±1.02	12.02± 0.83	12.72± 0.91
7	19.23 ±2.00	26.14 ±2.98	14.77 ±2.00	18.79± 2.75	17.45 ±0.90	21.19 ±0.94	11.98± 0.64	12.06± 0.67

From the performance analysis and regression plots of the MLP-NN, T1-RBF-NN, and IT2-RBF-NF models, it is clear that all the non-linear model-based methods have achieved relatively good predictions accuracy (RMSE, MAE %) within the process envelope (operating window). However, the IT2-RBF-NFS has a smaller RMSE and MAE% as well as a much smaller standard deviation (SD) intervals compared to the other models, indicating that the IT2-RBF-NF model demonstrates its ability to handle effectively the uncertainties associated with the input data. The extra degree of the freedom from the IT2-MF FOU of the IT2-RBF-NF model has high tolerance to the input noise and thus its ability to take into account the uncertainties associated with the meaning of words. The additional flexible parameters provided by the IT2-MF FOU in the IT2-RBF-NFS is used to better handle the numerical uncertainty associated with system inputs and outputs. Therefore, the IT2-RBF-NF model has the potential to achieve better predictive performance (generalisation/ testing performance) and robust (low SD) response.

5.4.1.2. TOOL TRAVERSE FORCE CASE

In a like manner to the axial force, the modelling results presented in this section include the measured/predicted performance for training and testing (Fig. 5.20), as well as the three-dimensional simulated model behaviour (Fig. 5.21) and its corresponding linguistic rule-base (Fig. 5.22).

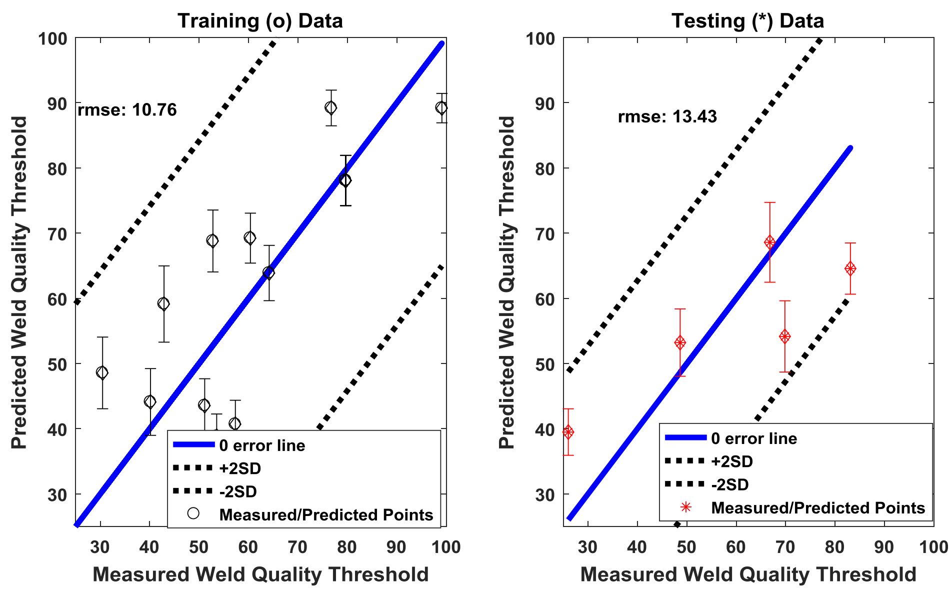


Figure 5.20. The 5-rule IT2-RBF-NF weld quality threshold model for the traverse force.

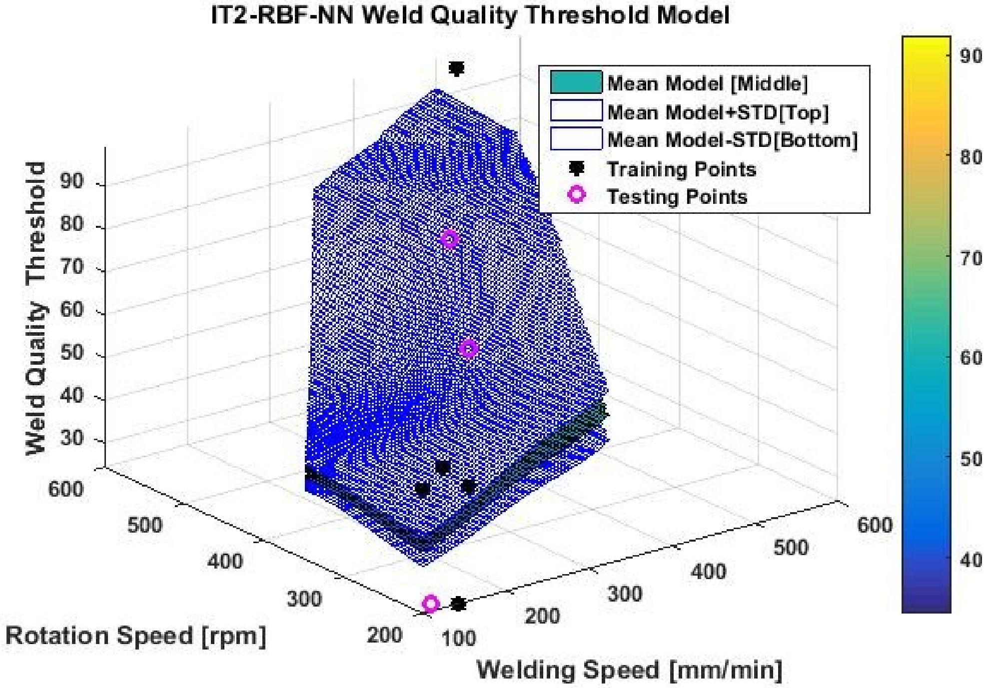


Figure 5.21. The response surface of the 5-rule IT2-RBF-NF weld quality threshold model for the traverse force (mean +/- SD).

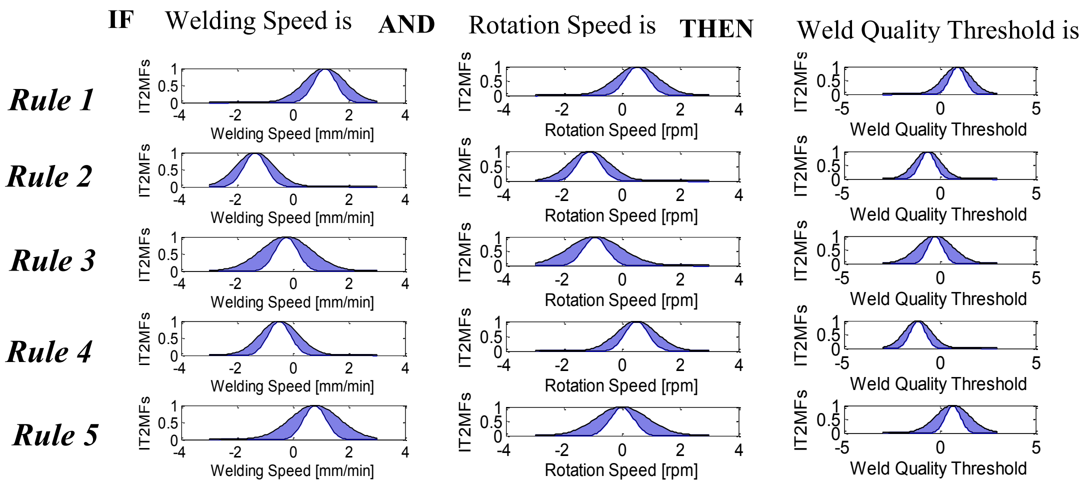


Figure 5.22. The rule-base of the 5-rule IT2-RBF-NF weld quality threshold model for the traverse force.

The linguistic rule-base structure from the traverse force IT2-RBF-NF model presented in Fig. 5.22 can be interpreted in a linguistic form as follows:

Rule 1: IF *Welding Speed* ***is*** *very high* ***AND*** *Rotation Speed* ***is*** *very high,* ***THEN*** *Weld Quality Threshold* ***is*** *very high*

Rule 2: IF *Welding Speed* ***is*** *very low* ***AND*** *Rotation Speed* ***is*** *very low,* ***THEN*** *Weld Quality Threshold* ***is*** *low*

Rule 3: IF *Welding Speed* ***is*** *medium* ***AND*** *Rotation Speed* ***is*** *low,* ***THEN*** *Weld Quality Threshold* ***is*** *medium*

Rule 4: IF *Welding Speed* ***is*** *low* ***AND*** *Rotation Speed* ***is*** *high,* ***THEN*** *Weld Quality Threshold* ***is*** *very low*

Rule 5: IF *Welding Speed* ***is*** *high* ***AND*** *Rotation Speed* ***is*** *medium,* ***THEN*** *Weld Quality Threshold* ***is*** *high*

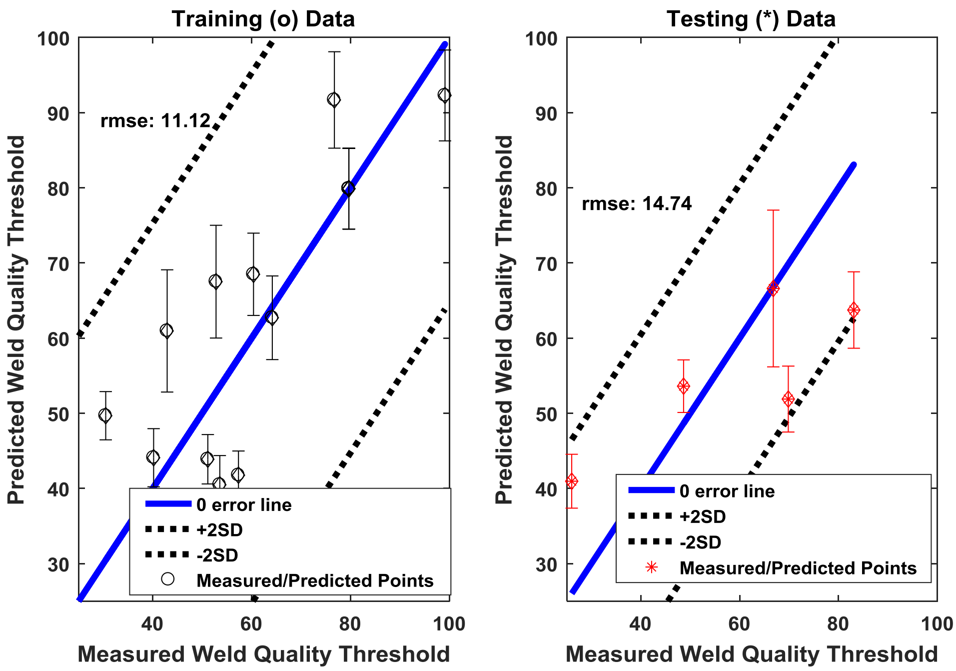


Figure 5.23. The 2-hidden layers, 100 hidden units weld quality threshold model for the traverse force.

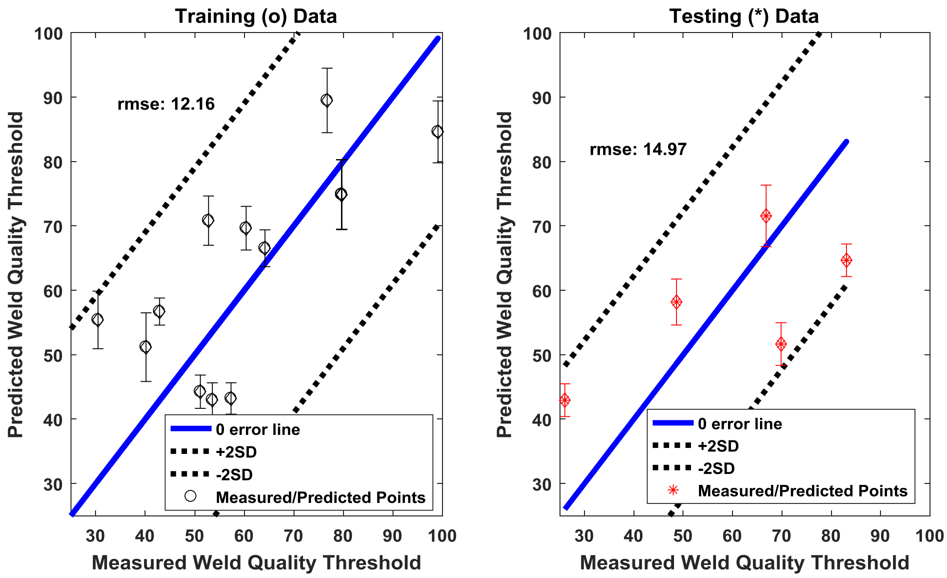


Figure 5.24. The 5-rule T1-RBF-NN weld quality threshold model for the traverse force.

The same comparative modelling procedures were utilised for the traverse force case. Tables 5-4 and 5-5 show the overall summative performance analysis of the developed models. As in the axial force case, it is evident that the prediction performance on the traverse force models is very similar in terms of their overall model accuracy, i.e. the MRL model (linear method) fails to capture the process dynamics, but other non-linear methods (MLP-NN, T1-RBF-NN, and IT2-RBF-NF models) are almost equally accurate. The main difference again, is the robustness of the proposed IT2-RBF-NF model, which exhibits the lowest SD intervals of prediction.

Table 5.4 Performance of the multiple linear regression and multilayer perceptron weld quality thresholds models for the traverse force.

Performance Index	**MLR Model**		**MLP-NN Model**	
	Training	Testing	Training	Testing
RMSE±SD	20.12±8.23	20.12±8.29	11.12 ±1.37	14.74 ±2.88
MAE%±SD	27.08±7.16	27.08±7.20	18.26 ±14.83	25.67±5.16

Table 5.5 Performance of the T1-RBF-NN and IT2-RBF-NF weld quality threshold models for the traverse force.

Model Type	**T1-RBF- NN**				**IT2-RBF-NF**			
Number of Rules	**RMSE±SD**		**MAE%±SD**		**RMSE±SD**		**MAE%±SD**	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
3	17.35 ±6.99	22.00 ±5.88	26.86 ±3.21	33.14± 10.71	14.36 ±2.97	17.29 ±3.09	21.60 ±2.52	24.48 ± 2.89
5	12.16 ±2.41	14.97 ±3.18	21.31 ±1.93	28.29± 2.13	10.76 ±0.72	13.43 ±0.83	18.13 ±0.54	20.15 ±0.61
7	11.76 ±2.19	13.79 ±2.78	20.99 ±1.84	25.69± 1.91	9.86 ±0.64	12.23± 0.78	16.95 ±0.23	19.79 ±0.28

Overall, from the frequency spectral analysis, the statistical correlation analysis and the modelling results it is confirmed that indeed there is a non-linear relationship between the weld quality threshold in the frequency domain and the input process parameters. This non-linear behaviour is modelled successfully via the IT2-RBF-NF model, while having the ability to also provide continuous linguistic feedback to the operator(s) about the performance of the process. The successful modelling of this moving threshold enables the creation of a process monitoring system that would be capable of classifying in real-time the quality of produced welds in steel FSW based on axial and traverse force signals. To put it more simply, from the linguistic rule-base of the final IT2-RBF-NF model, continuous feedback in a linguistic format (rule-based human-centric system) can be provided to the process operator(s). The computational demands and efficiency of the IT2-RBF-NF model are only modest (almost instantaneous single weld quality threshold prediction on a standard PC), therefore the real-time application (while welding) of the proposed reliable monitoring system is also feasible.

5.4.2. APPLICATION TO ALUMINIUM FRICTION STIR WELDING

In this section the generalised model-based real-time monitoring framework that relies on the IT2-RBF-NFS is validated in different material and weld tool combinations. The dataset in this research work has been obtained with the help of TWI Ltd, South Yorkshire, UK, where 31 experimental trials were conducted for the welding of an aerospace-grade aluminium plates in 76 mm thickness (AA2219-T845) welded via a 38 mm TriFlat weld tool [279]. All the welds were made with a 38 mm TriFlat weld tool. All welds were made with different levels of tool rotational speed and welding speed in order to determine the process envelope for welding AA2219-T845 aluminium alloy. The data acquisition rate for all relevant signals was set to 1 Hz. Figs. 5.25 and 5.26 show data records for the 31 experimental trials for the traverse and axial force signals respectively. The histogram of tool rotational speed and welding speed for the overall data space for the FSW process is plotted in Fig. 5.27. It can be shown that the data space is complex and sparse; there are areas of low data density and areas of high data density (i.e. process operation envelope).

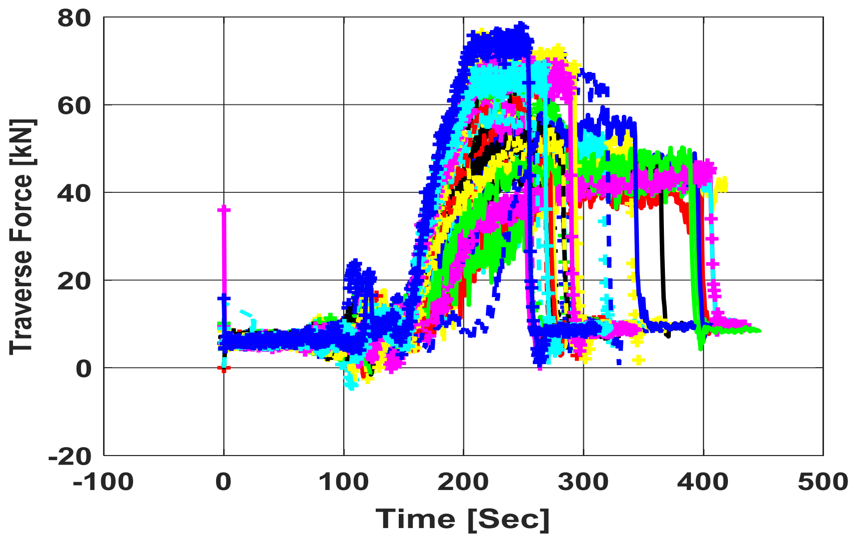


Figure 5.25. Traverse force signals recorded during 31 experimental trials using a 38 mm TriFlat weld tool with different levels of tool rotational speed and welding speed.

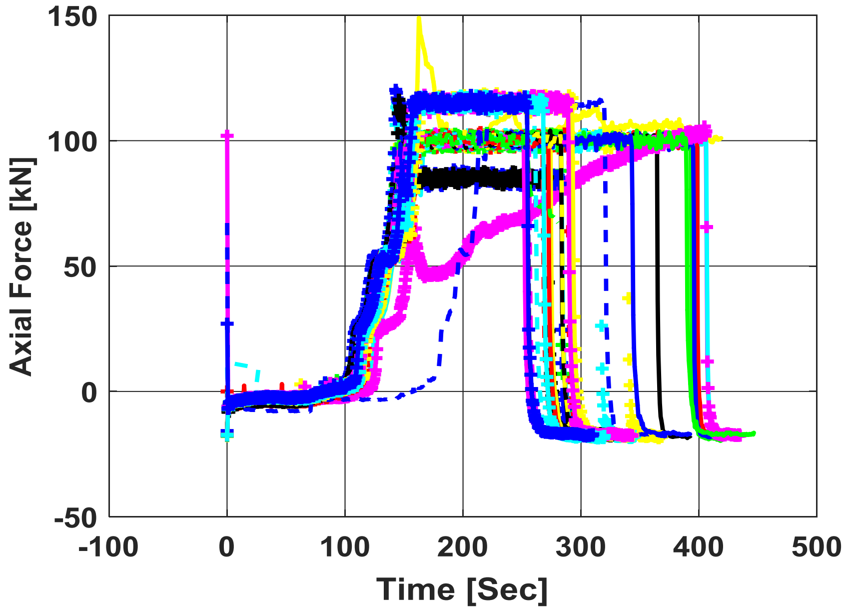


Figure 5.26. Axial force signals recorded during 31 experimental trials using a 38 mm TriFlat weld tool with different levels of tool rotational speed and welding speed.

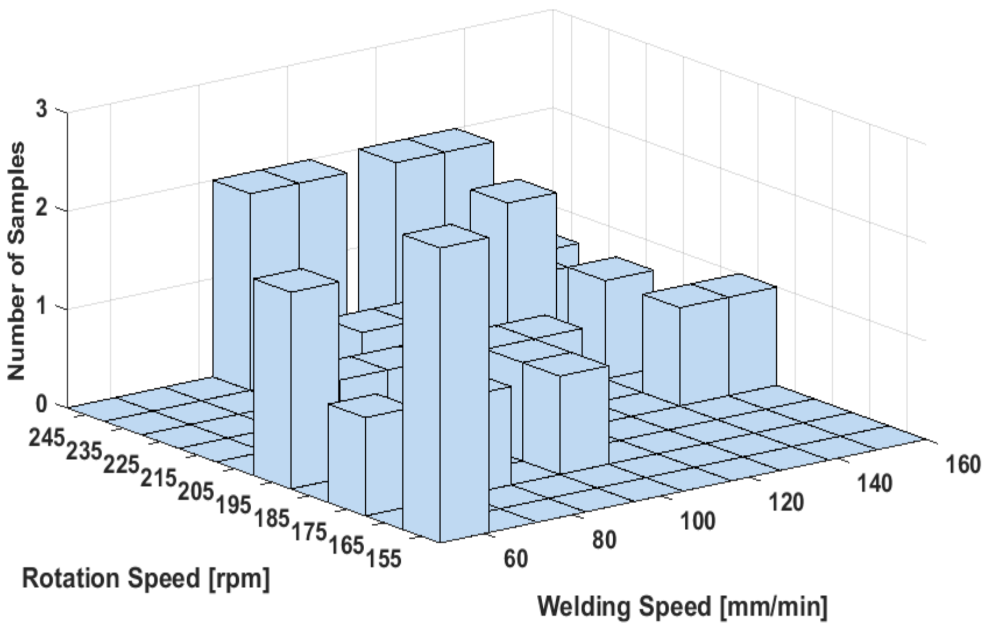


Figure 5.27. FSW data density of 31 experimental trials.

The quality of the welded samples was quantified by process experts via visual assessment of the surface area. Based on surface assessment, an overall quality index is obtained to represent the general situation for weld quality, whose value could be 0 indicates excellent weld quality or 1 indicates poor weld quality.

In the same way to the work conducted in the previous section, the DFT for the feedback force signals is calculated. By way of illustration, Figs. 5.28 and 5.29 show the DFT of the traverse and axial force signals for 31 welds at different levels of tool rotational speed and welding speed respectively. All the force signals are windowed for the whole steady state welding stage. In these figures, it is obvious that the defect-free welds produce smaller frequency components at the lower frequency spectrum including the DC component, whilst the defective welds produces larger amplitudes at the lower frequency components. Consequently, the amplitudes of the low frequency components may have a correlation with weld quality.

Based on the frequency and correlation analysis, useful information can be extracted from the frequency spectra of the feedback force signals. The frequency spectra information can be used to establish a threshold in order to separate (classify) the good welds from the poor welds. However, in practice different process parameters result in different thresholds (i.e. different force signals 'behaviour') and therefore this threshold cannot be generalised for different process parameters. The relationship between the process conditions and the resulting weld quality threshold is nonlinear. As discussed earlier, the model-based approach that is used for the monitoring framework depends on the prediction of the nonlinear behaviour of the threshold for different process conditions. Successful prediction of this marker enables the process user to classify weld quality.

Table 5-6 presents the FSW data set used consisting of 31 measurements. The measurements include the process parameters (welding speed and tool rotation speed) and weld quality thresholds extracted from the feedback force signals (traverse and axial forces) as described earlier. During the first stage of the modelling process the data set has been split into two sets, 25 data points to train the model and 6 data points for validation purposes. After a number of systematic simulations (increased/reduced the number of granules), it was established that the optimal number of granules is five. These five granules are used to construct the initial structure of the model (number of rules and initial parameters). The initial structure of the IT2-RBF-NF model is trained via the adaptive back-propagation (BEP) algorithm. Performance indexes based on RMSE and MAE% between the true threshold and model predicted threshold was used to evaluate the performance of the neural fuzzy models developed.

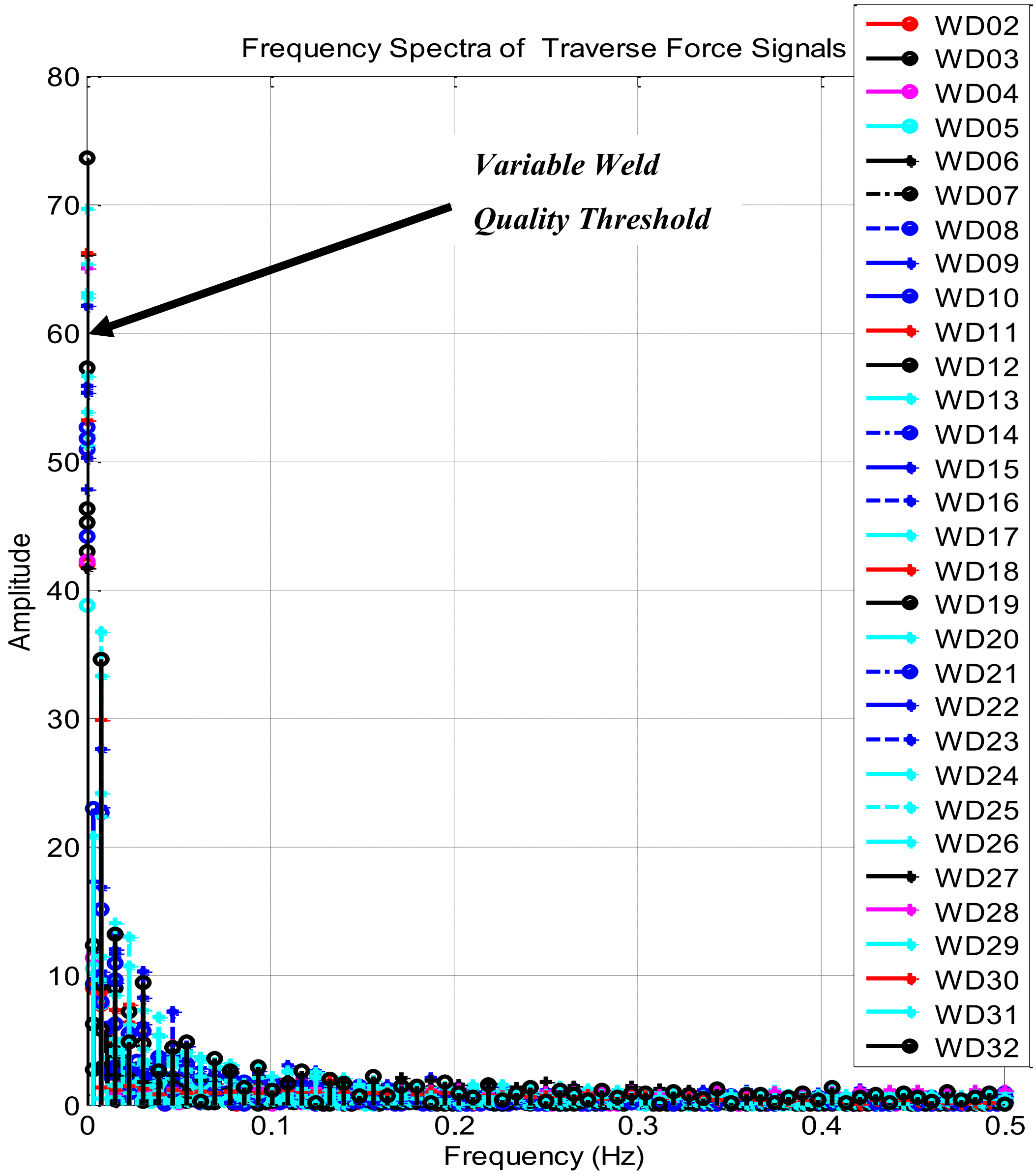


Figure 5.28. Discrete Fourier Transform of the traverse force signals.

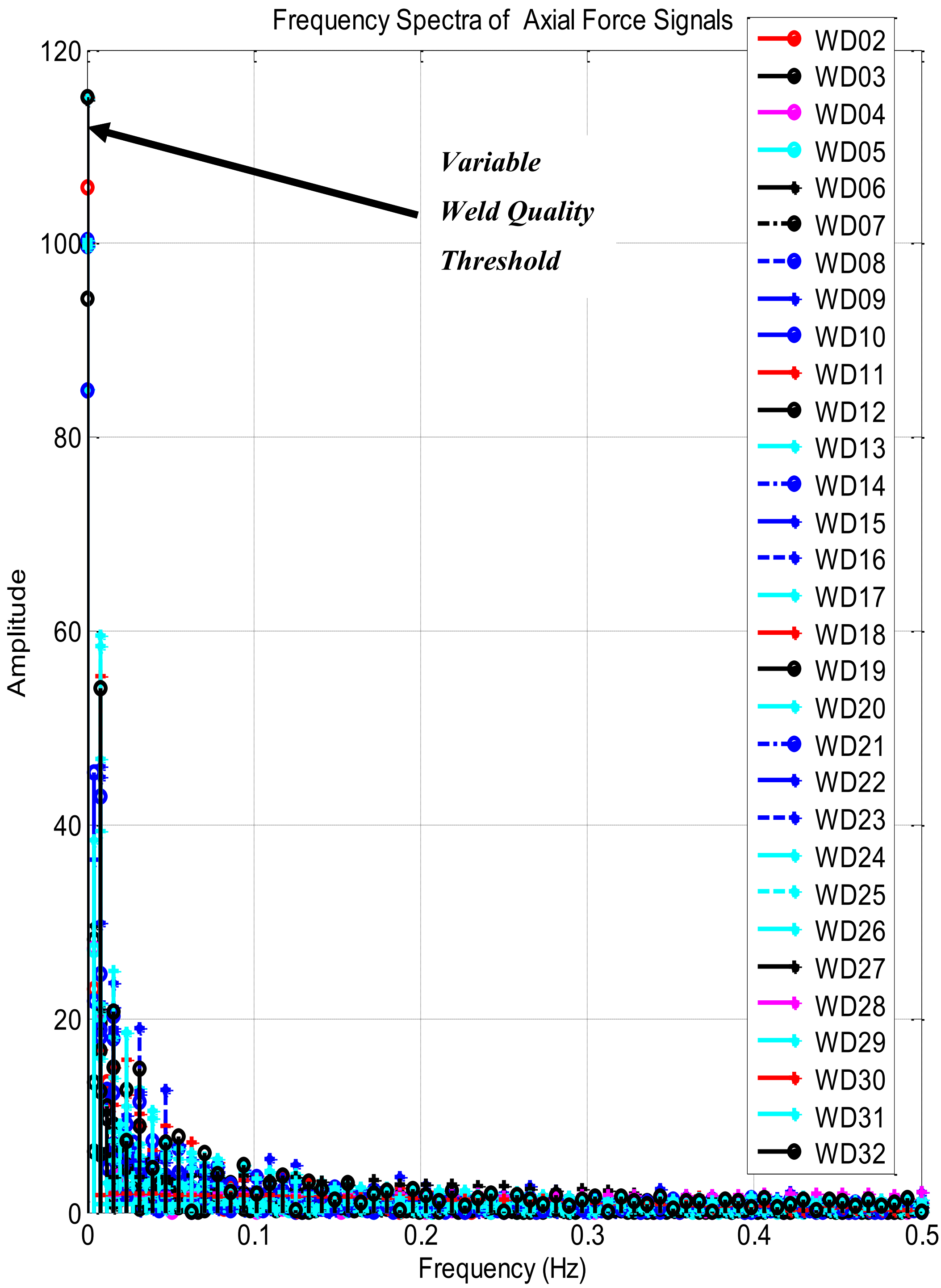


Figure 5.29. Discrete Fourier Transform of the axial force signals.

Table 5.6 Process parameters of 31 experimental trials used for modelling (including training and testing data). If the value of the indicator is below the threshold, the weld can be classified as a good weld otherwise the weld is of poor quality.

Weld Sample	Welding Speed [mm/min]	Rotation Speed [rpm]	Threshold Values for Traverse Force	Threshold Values for Axial Force
WDNo.02	50	150	41.9892	105.8636
WDNo.03	50	150	43.0203	94.3488
WDNo.04	50	150	42.2466	100.0358
WDNo.05	50	170	38.7740	100.1951
WDNo.06	50	190	41.7373	99.9398
WDNo.07	60	190	45.3140	99.8043
WDNo.08	70	190	44.2371	100.2622
WDNo.09	80	190	47.8541	100.0881
WDNo.10	90	190	50.9214	99.8174
WDNo.11	90	210	50.3309	100.0018
WDNo.12	90	250	57.2531	100.1225
WDNo.13	90	250	51.2287	84.9336
WDNo.14	100	250	52.6585	84.8498
WDNo.15	100	250	55.8986	99.7639
WDNo.16	120	250	62.0890	99.7583
WDNo.17	100	190	52.1112	99.7003
WDNo.18	110	190	53.2582	99.9864
WDNo.29	80	170	46.3304	100.0211
WDNo.20	100	170	53.8700	99.9506
WDNo.21	100	180	51.8012	100.3040
WDNo.22	100	210	50.3391	100.1607
WDNo.23	120	210	55.3761	99.7559
WDNo.24	120	210	56.6503	99.9345
WDNo.25	120	250	62.7591	99.5400
WDNo.26	130	240	63.0390	100.2984
WDNo.27	130	240	66.1035	114.8705
WDNo.28	130	220	64.9960	115.0037
WDNo.29	140	230	69.6559	115.2312
WDNo.30	140	210	66.1871	115.0800
WDNo.31	140	190	65.3498	114.9401
WDNo.32	160	190	73.5770	115.0809

The proposed framework was applied to two cases related to traverse and axial forces used as the main signals for the process monitoring; the two cases are detailed below:

5.4.2.1. TOOL TRAVERSE FORCE CASE

In the following case relating to the traverse force signals for the process monitoring. Using the proposed model-based framework, a 5-rule fuzzy model was developed after structure identification and parametric optimisation. The rule-based fuzzy model is presented in Fig. 5.30.

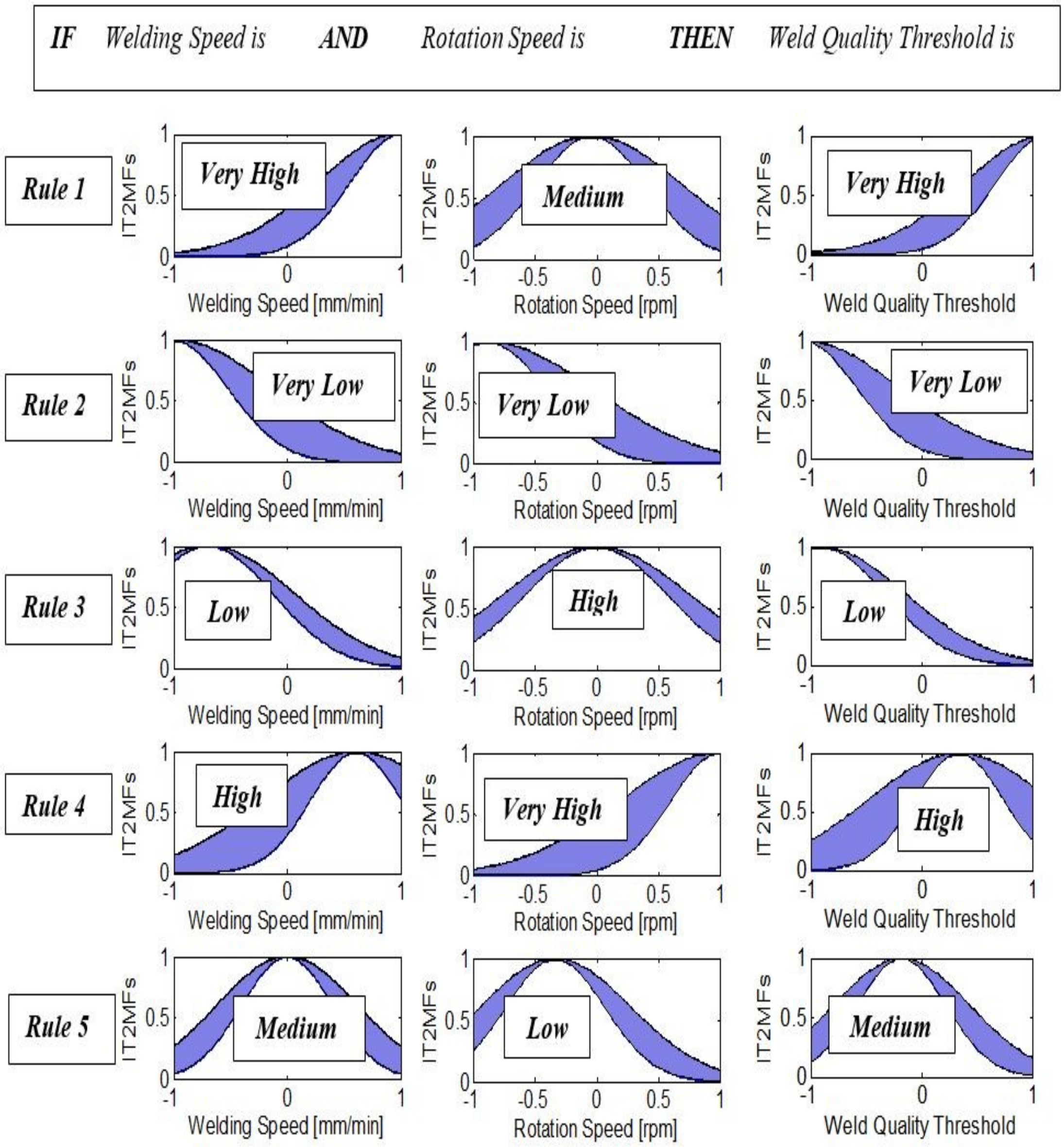


Figure 5.30. The rule-base of the 5-rule IT2-RBF-NF weld quality threshold model for the traverse force.

From the rule-based fuzzy model, the corresponding linguistic rules can be written as follows:

Rule 1: IF *Welding Speed* ***is*** *very high* ***AND*** *Rotation Speed* ***is*** *medium* ***THEN*** *Weld Quality Threshold* ***is*** *very high*

Rule 2: IF *Welding Speed* ***is*** *very low* ***AND*** *Rotation Speed* ***is*** *very low* ***THEN*** *weld quality threshold* ***is*** *very low*

Rule 3: IF *Welding Speed* ***is*** *low* ***AND*** *Rotation Speed* ***is*** *high* ***THEN*** *weld quality threshold* ***is*** *low*

Rule 4: IF *Welding Speed* ***is*** *high* ***AND*** *Rotation Speed* ***is*** *very high* ***THEN*** *weld quality threshold* ***is*** *high*

Rule 5: IF *Welding Speed* ***is*** *medium* ***AND*** *Rotation Speed* ***is*** *low* ***THEN*** *weld quality threshold* ***is*** *medium*

The prediction performance of the obtained model: RMSE = 1.79, 1.78 and MAE % = 2.82, 2.23 for model training and testing respectively. Simulation results of the model for training and testing are shown in Fig 5.31. According to the simulation results, the model gives good prediction and generalisation with 97.78 % accuracy.

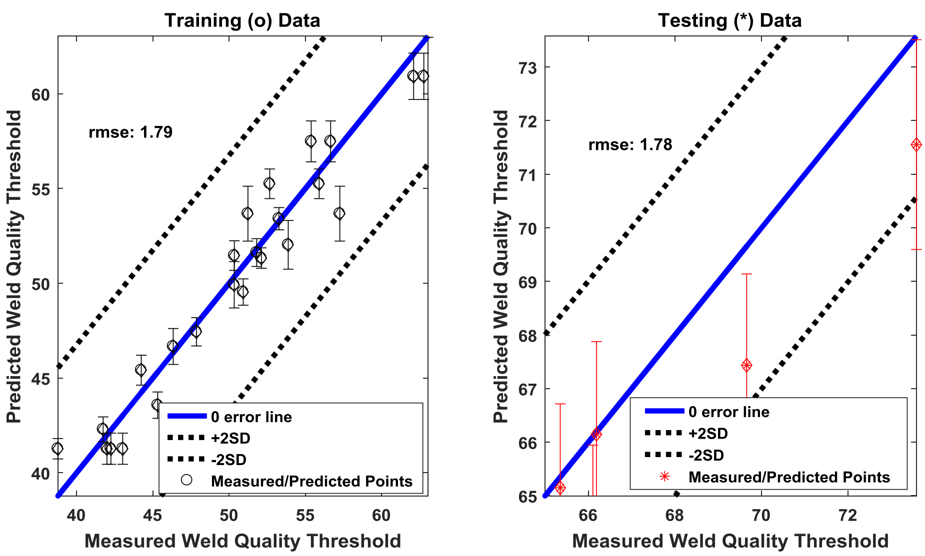


Figure 5.31. The 5-rule IT2-RBF-NF weld quality threshold model for the traverse force.

To provide more details and also to verify the physical interpretation of the obtained model, Fig. 5.32 illustrates the system's simulated behaviour between the welding speed, rotation speed and the predicted weld quality threshold. From the surface plot, it can be clearly seen that there is a nonlinear relationship between how the threshold changes with respect to the process conditions. It can also be observed that, with increasing rotation speed, the weld quality threshold tends to increase. This trend follows the expected behaviour from the knowledge experts.

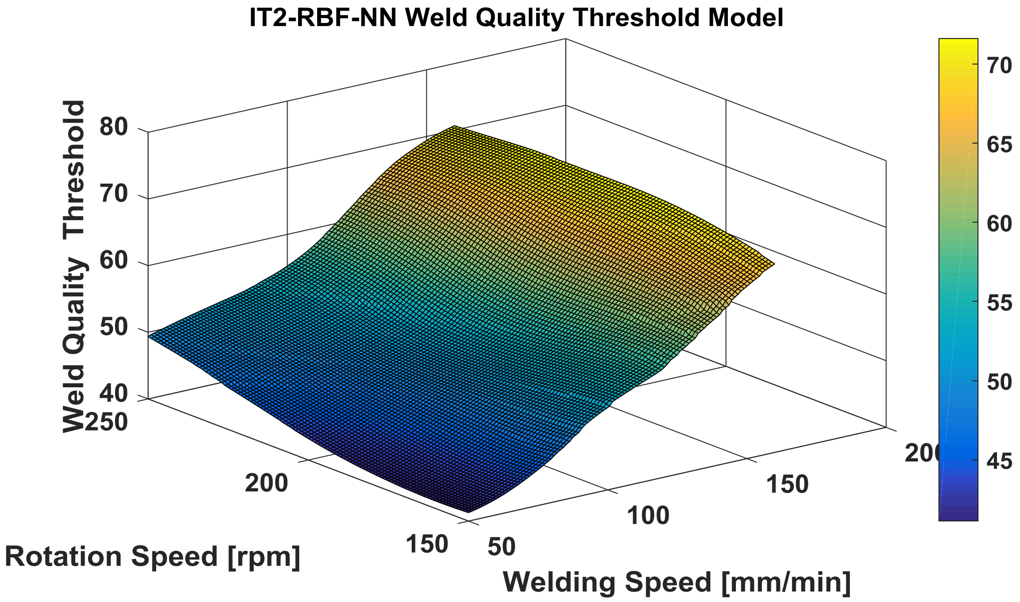


Figure 5.32. The response surface of the 5-rule IT2-RBF-NF weld quality threshold model for the traverse force: If the value of the indicator is below the threshold, the weld can classified as a good weld, otherwise the weld is of poor quality.

5.4.2.2. TOOL AXIAL FORCE CASE

Similarly, a 5-rule fuzzy model was developed for the axial force case using the proposed framework. After structure identification and model optimisation, the rule-based fuzzy model is presented in Fig. 5.33.

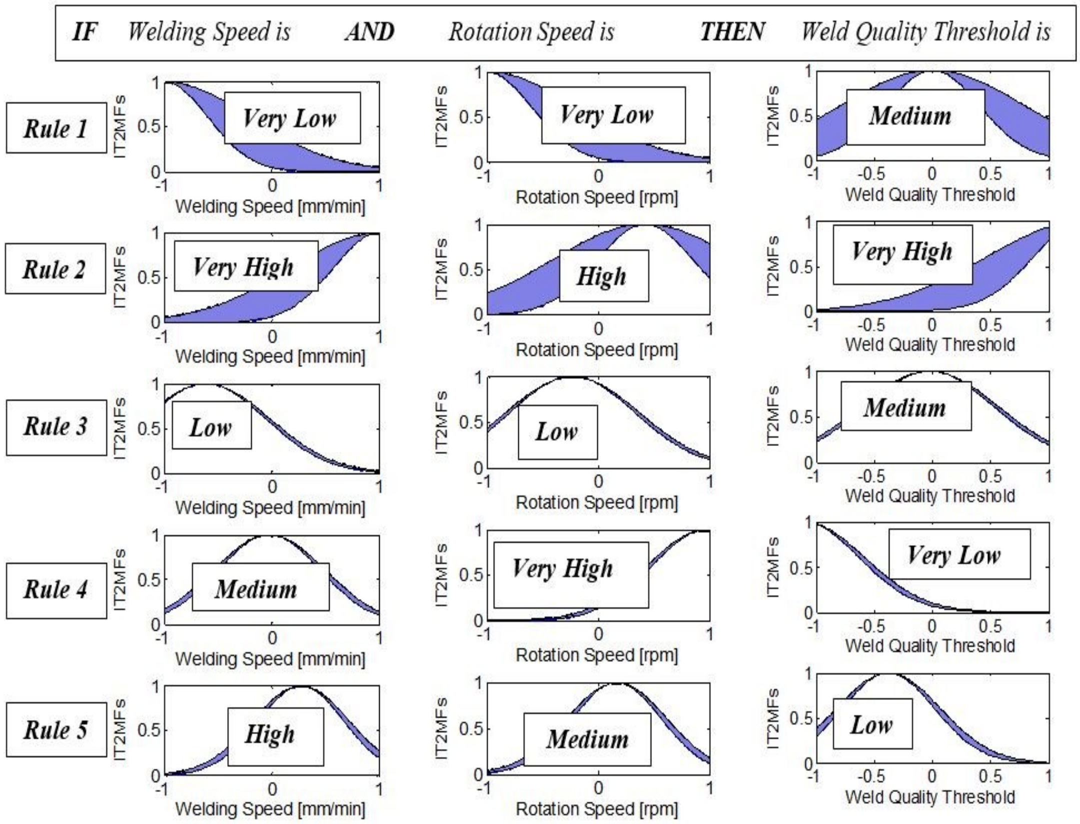


Figure 5.33. The rule-base of the 5-rule IT2-RBF-NF weld quality threshold model for the axial force.

The corresponding linguistic format is as follows:

Rule 1: IF *welding speed is very low* ***AND*** *rotation speed is very low* ***THEN*** *weld quality threshold is medium*

Rule 2: IF *welding speed is very high* ***AND*** *rotation speed is high* ***THEN*** *weld quality threshold is very high*

Rule 3: IF *welding speed is low* ***AND*** *rotation speed is low* ***THEN*** *weld quality threshold is medium*

Rule 4: IF *welding speed is medium* ***AND*** *rotation speed is very high* ***THEN*** *weld quality threshold is very low*

Rule 5: IF *welding speed is high* ***AND*** *rotation speed is medium* ***THEN*** *weld quality threshold is low*

The prediction performance of the obtained model: RMSE = 1.79, 1.78 and MAE % = 3.32, 3.23 for model training and testing respectively. Simulation results of the model for training and testing are shown in Fig. 5.34. According to the simulation results, the model gives good prediction and generalisation with 96.77% accuracy. The simulated model behaviour is shown in Fig. 5.35.

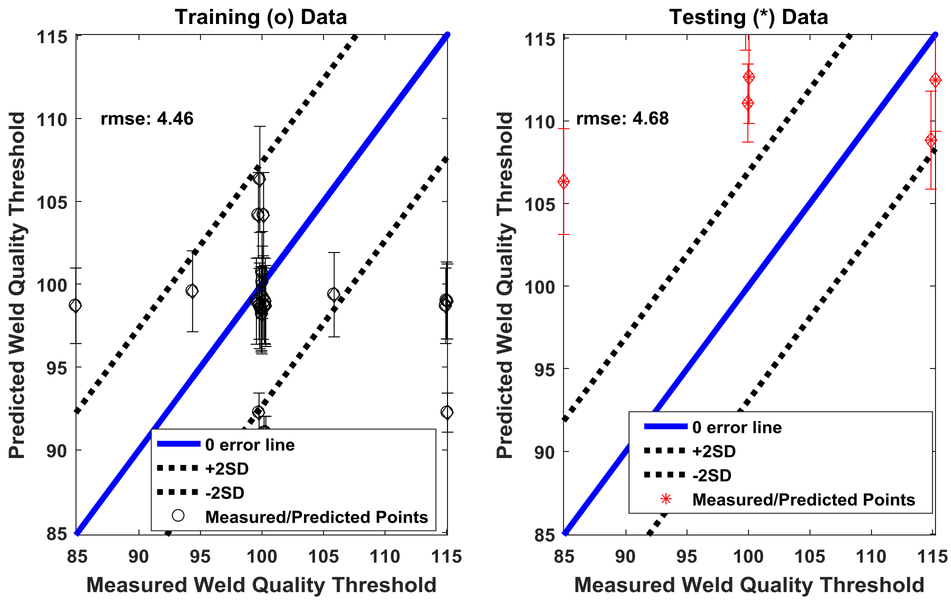


Figure 5.34. The 5-rule IT2-RBF-NF weld quality threshold model for the axial force.

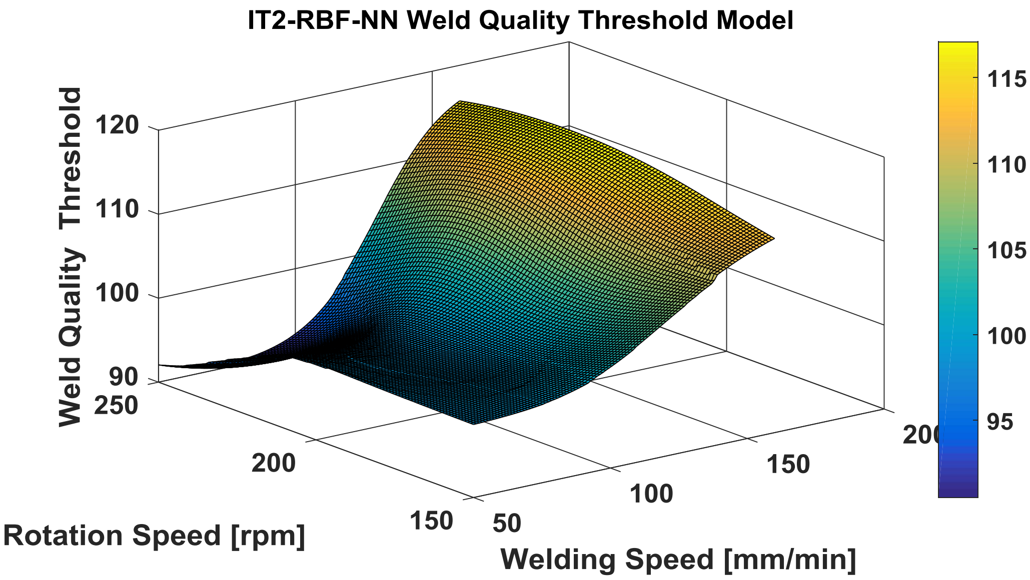


Figure 5.35. The response surface of the 5-rule IT2-RBF-NF weld quality threshold model for the axial force: If the value of the indicator is below the threshold, the weld can classified as a good weld, otherwise the weld is of poor quality.

The simulation results show the effectiveness of the IT2-RBF-NFS approach to handle linguistic uncertainties and achieve reasonable process performance markers (~98% accuracy in testing data). The successful application of the proposed model-based framework enables the creation of a process monitoring system that would be capable of classifying in real-time the quality of produced welds in aluminium FSW based on axial and traverse force signals.

From the simulations results presented in the previous sections, it is clear that the proposed IT2-RBF-NFS outperformed the MRL and MLP-NN models as well as its type-1 radial basis function neural network counterpart mainly in generalisation terms. It also proved its efficiency and high prediction accuracy for fitting data in the presence of high input uncertainty.

In practical terms, it is clear that the data space for steel FSW is more complex and highly non-linear than the data space for aluminium FSW. This is mainly due to steel FSW is performed at high temperatures of up to 1,100˚C (measured at the tool), hence the tool must retain its strength at these high temperatures while being subjected to complex bending, rotational and fatigue loads. An additional complexity is that the FSW of steel is characterised by the presence of phase transformations which deem the process optimisation even more challenging [151]. Therefore, the proposed model-based monitoring framework has great potential for successful implementation in different materials, weld tools, and tool geometry combinations.

5.5. SUMMARY

In this chapter, a new systematic IT2-RBF-NF modelling framework is presented. The aim was to develop a parsimonious, robust and computationally efficient human-centric rule-based neural fuzzy modelling framework that can be easily translated into human language via simple linguistic rules in order to describe the underlying dynamics behaviour of complex industrial processes with good generalisation capability, tolerance to input imprecision and low computational cost.

The proposed modelling framework was used to develop a new generalised model-based real-time process-monitoring framework to analyse the performance and behaviour of the steel FSW process and further validated in aluminium FSW process. On one hand, the proposed monitoring framework relied on frequency domain analysis of key process variables to establish markers (thresholds) of weld quality. On the other hand, such markers (thresholds) are subject to change as the weld input parameters also change. It was also demonstrated that how the feedback forces (axial (Fz) and traverse (Fx) forces) on tool during welding can be used to forecast/estimate the resulting weld quality threshold on different levels of process conditions. The proposed monitoring framework has proven its computational efficiency and robustness. It can also be used in line to the process, in real-time, to non-destructively (indirectly) evaluate the performance of the FSW process and the quality of the welds produced.

The modelling results showed an achieved prediction accuracy of more than 80% and 98% in the prediction of the weld quality thresholds on different levels of process conditions for welding of 6 mm thick DH36 shipbuilding-steel grade, welded via a pcBN weld tool and 76 mm thickness (AA2219-T845) aerospace aluminium grade welded via a 38 mm TriFlat weld tool respectively. In addition, the interpretability attribute of the developed human-centric rule-based neural fuzzy model can be used to provide continuous linguistic feedback on the performance of the process to the process operator(s). The use of linguistic rules can also be beneficial to the process operator(s) to gain an insight into the effects of the process conditions on the final weld quality and to monitor and prevent performance deterioration and overheating problems related to significant tool wear conditions.

This chapter is concluded by comparing the performance of the presented model-based approach against a linear model – as a baseline – as well as two common non-linear model-based methods, which exhibited good predictive performance in the past in complex manufacturing processes. The modelling results showed that while the linear model poorly models the relationship between the input parameters and weld quality threshold, as expected, the non-linear model-based methods provide similar levels of good predictions accuracy. It is noticeable that the proposed model-based approach outperforms all others in terms of the robustness of the predictions accuracy, as the standard deviation obtained in the predictions is the lowest among all of the other models. This confirms the ability of the IT2-RBF-NF model to perform better in the presence of high input uncertainty due to the extra degree of the freedom from the IT2-MF FOU of the IT2-RBF-NF has high tolerance to the input noise and thus its ability to take into account the linguistic uncertainty in the fuzzy rules. Additionally, the extra flexible parameters provided by the IT2-MF FOU in the IT2-RBF-NF system helps handle the numerical uncertainty associated with system inputs and outputs.

The results obtained from this chapter led to the publication of an article in the peer reviewed journal '*Journal of Manufacturing Processes*' with the title: *A real-time quality monitoring framework for steel friction stir welding using computational intelligence*, and an article that was presented at the *11th International Symposium on Friction Stir Welding in Cambridge, United Kingdom* with the title: *Real-time quality monitoring for friction stir welding AA2219-T845 aluminium aerospace alloy via model-based spectral analysis.*

In the next chapter, a new perpetual learning framework based on the IT2-RBF-NF system that has been developed in this chapter is proposed in order to accommodate new input-output mappings and new classes of data as well as make the system/model feasible for lifelong leaning mode.

CHAPTER 6 - A NEW PERPETUAL LEARNING FRAMEWORK FOR IT2-RBF-NFS

This chapter presents a new perpetual learning framework based on the iterative human-like information capture in granular computing (GrC) and IT2-RBF-NF model as described in Chapter 5.

On one hand, the proposed framework evolves through incremental and structural parametric learning. Such framework relies on the creation of new rules, which are added to the original model to update its structure. The updated model is then optimised during the incremental process.

On the other hand, an iterative rule pruning strategy is used to remove any inconsequential rules as a result of the incremental update routine. The strength of such perpetual learning framework is that this framework uses rule growing/pruning strategy, which makes the proposed framework feasible for lifelong learning mode.

6.1. INTRODUCTION

In the field of data-driven computational intelligence modelling, the learning methodologies can be broadly divided into two main categories: *offline (batch) learning* and *online learning* [280-282]. The former method assumes that all the data points are available to the model before the learning process starts and can be accessed repeatedly, while the latter assumes one data point arrives at a time from a possibly infinite stream [283, 284]. In a broad sense, on one hand, the offline leaning methods are concerned with real-world problems when a set of data is obtained and used to train/learn an approximating function before the function is used in the application [285]. On the other hand, online learning methods are applied to complex and continuously changing characteristics real-world problems, such as chaotic time series prediction [286], dynamic auditory-visual signal analysis [287], and ambient intelligent environments [288]. Such dynamic in nature case studies require sophisticated computational frameworks that are able to adapt incrementally in an online manner [283], learn and generalise from data automatically [289], and dynamically change their structure and create new rules [290].

In the literature, some of the frameworks that have been introduced to deal with such problems for instance adaptive decision-making modelling and control systems [291], multi-modal information processing systems [287] intelligent agent-based systems [292, 293]. A number of adaptive learning approaches exist in Computational Intelligence (CI), namely online learning [47-49], incremental learning [294-296], lifelong learning [48], and knowledge-based learning neural networks [297-299]. Recent research on online learning concentrates on adaptive learning approaches to follow time-varying distributions [283, 299, 300]. Incremental learning is the process of repeatedly training a model with new data without completely disturbing the old model [301]. Lifelong learning also termed perpetual "continuous learning" addresses learning through the whole lifespan of a system [302]. In [47] a dynamic evolving neural fuzzy system for the prediction of time-series data was proposed; this approach requires to normalise all data prior to training, which infers that all the data points must be present prior to any training.

Other dynamic neuro-fuzzy system approaches include a self-organising fuzzy neural network [303], a self-constructing neural fuzzy inference network [304], a dynamic fuzzy neural network [305-307] and a dynamic parsimonious fuzzy neural network [290]. However, the aforementioned methods never prune the rules once generated, regardless of their relevance. Therefore, a large number of inconsequential rules may be generated each time new data sets (information) are available. Other alternative approaches for online learning which adapt the feature of rule creation and pruning mechanisms exist, such as the sequential adaptive fuzzy inference system [281], and the most recent sequential probabilistic learning for adaptive fuzzy inference system [283]. In on-line learning, only one data pattern is provided at a time and then discarded after the learning process has been completed. The online methodology is not very demanding on computing resources and at the same time it fits well with dynamically changing environments.

To this point the focus has been on the adaptive learning methods, particularly in addressing the recent advances in the field of type-1 fuzzy logic system. However, as already discussed in Chapter 5 recent research on type-2 fuzzy logic systems (T2-FLSs) have attracted much attention [32-34]. This is due to their ability to better handle the measurement noise and modelling uncertainties. Therefore, particular efforts on online learning methodology have also been recently focused on type-2 fuzzy logic systems. In [308], a self-evolving interval type-2 fuzzy neural network is proposed, in which the structure and parameters learning are carried out in an online manner. The parameters of antecedent and consequent parts are optimised via the gradient descent algorithm and rule-ordered Kalman filter algorithm respectively; its performance was validated for time-varying systems. In [309], the authors proposed a mutually recurrent interval type-2 neural fuzzy system for the identification of non-linear and time-varying systems. The proposed structure also used the gradient descent and rule-ordered Kalman filter algorithm for parameters tuning.

In more recent studies [50-53] online leaning methods have been proposed. In [50], a TSK-Type-based self-evolving type-2 fuzzy neural network is proposed to improve the system's robustness in noisy environments [50], and a Mamdani-type IT2-NFS with on-chip incremental learning ability is introduced [51]. This system [51] utilises a simplified type-reduction operation into an interval Type-2 NFSs to reduce the computational cost. In [52], Lin et al. proposed an interval type-2 NFS for online system identification and feature elimination. The proposed structure possesses a self-organising property that can automatically generate fuzzy rule and optimised via a gradient descent based approach.

The aforementioned adaptive learning methodologies that have been proposed so far incrementally evolve their structure and optimise the parameters. The online adaptive learning ability of such methodologies makes it suitable for learning data streams that are generated from non-stationary environments, for example in processes where time-series data are generated (i.e., evolving data) [47]. However, in some applications, data streams are generated periodically from non-varying environments (i.e., static, non-dynamic relationship-based data) [27].

The development of a human-centric rule-based system from static data is a dual-stage process, which includes the structure learning stage and the parameter learning phase. These two stages are often performed sequentially; the first stage is employed to construct the initial structure of rule-base. This stage can be carried out via the iterative human-like information granulation of GrC as described in Chapter 4. Then in the second stages, a parametric optimisation approach is used to optimise/tune the parameters of each rule in the rule-base as well as the inference mechanism. However, developing efficient data-driven computational models require significant effort and the training process is over-dependant on expert knowledge. Repeating the whole modelling process is often a laborious and non-automated process as well as time-consuming. Consequently there is no guarantee that the new model will maintain a good performance comparable to the original model. The batch learning approach is often assumed, where the model structure uses all training data simultaneously and allowed to employ them as often as required.

In some industrial/manufacturing applications, obtaining batch data is a slow and expensive process and the data set can only be obtained periodically. Therefore, there is a need to for the system to have the ability to learn from an initial batch of data (with the help of an appropriate training algorithm) and periodically adapt to new data sets when these are available. An additional need is to include the capability to interact with a changing environment in a 'perpetual' continuous fashion and also to have an open structure; this entails to dynamically expand the system's structure to accommodate new data/information – without significantly disturbing the initial model structure. A rule pruning mechanism would also be required in order to remove/prune insignificant and/or redundant rules that have limited contribution to the system's performance.

In this chapter, a new perpetual learning framework is proposed, where the initial model can incrementally update its structure to accommodate new data sets in type-2 fuzzy logic systems. The proposed perpetual learning framework is able to detect effectively any change in the data distribution via a novelty detection process and expands its structure to cover the new input data space without significant loss on the overall performance of the model (including original and new data combined). The proposed system is realised based on an Interval Type-2 Radial Basis Function Neural Fuzzy (IT2-RBF-NF) system that has been developed in Chapter 5. The proposed perpetual learning system has also the ability to improve its structure periodically by removing inconsequential rules.

The motivation for the creation of this mathematical framework that is in classical approaches to data-driven CI modelling, the model training process needs to be completed initially by the user. If new process data become available, the model training exercise needs to be carried out again, thus creating a new model from scratch. This is often a laborious and non-automated process, which needs significant expert knowledge. The proposed perpetual learning framework has the ability to acclimate/adapt to new information in an additive and lifelong learning fashion in order to accommodate new input-output mappings and new data clusters.

Therefore, the main contributions of this chapter can best be treated under the following headings:

a) Develop a new perpetual (incremental) learning framework that is based on granular computing IT2-RBF-NF model optimised via the adaptive back-error propagation (BEP) algorithm. This modelling framework has been developed and fully described in Chapter 5. By using such a human-centric modelling framework, it is possible to capture meaningful information out of non-linear, complex, and scarce process data to build a rule-based system with good modelling performance and at the same time good overall system transparency (interpretability).
b) The proposed structure has the ability to continuously learn from new process data – in an incremental learning fashion by modifying (adapting) the existing model. This is usually achieved via expanding the structure of the existing model by creating new rules to accommodate the new data without significantly disturbing the existing model. An iterative rule pruning strategy is used as the main feature that prunes/removes the inconsequential/redundant fuzzy rules after each incremental step, which allows the model to be used in a lifelong learning mode.
c) The performance of the proposed structure is demonstrated using a number of simple as well as complex non-linear benchmark functions. Simulation results show that the performance of the original model structure is maintained and it is comparable to the overall model performance after the final incremental model updating routine. The efficiency and effectiveness of the proposed

structure is also evaluated in case where more frequent/periodic model updates are required with good prediction performance. Finally, this chapter is concluded by applying the proposed methodology to a real-industrial problem. The prediction of spindle peak torque of steel Friction Stir Welding is investigated. As was mentioned in Chapter 3, FSW is a complex thermo-mechanical process that involves highly non-linear and complex as well as sparse databases.

6.2. A NEW PERPETUAL LEARNING FRAMEWORK WITH RULE GROWING AND PRUNING

This section explores theoretical approaches to the proposed perpetual learning framework. As was explained in Chapter 5, the development of an IT2-RBF-NF system generally involves two learning stages, the structure identification stage and the parametric optimisation stage. One characteristic of this model-based approach is that it is suitable when sufficient amount of process data is collected and used to construct the rule-based system. Once the model is constructed, it has a fixed structure (fixed rule-base) with the desired level of accuracy. However, the model having a fixed structure cannot always extrapolate well each time new process data become available. This is due to different characteristics/dynamics of a complex system under different input conditions. To improve the extrapolation capability of a model when new process data are available, in general there are two strategies to achieve this. The first strategy is to develop an entirely new model taking into consideration the specific features of the new data, which are not covered by the 'old' model. In this case, it is required to use the two learning procedures in order to develop a new model that covers the new input data patterns. However, in practice, developing an entirely new model often requires a lot of efforts and it is over-dependent on expert knowledge. The second strategy is concerned with modifying (adapting) the existing model. This is usually achieved via a dual-step procedure: in the first step, the structure of the existing model is expanded by creating new fuzzy rules to accommodate the new data without significantly disturbing the original model and in the second step the updated model is further fine-tune to improve its generalisation capability.

The work presented in this chapter is focused on the idea of offline perpetual (incremental) learning in which additional knowledge is added to the original model based on the second strategy. In the proposed framework, the modelling scheme is designed to learn from an initial dataset (via the two learning stages) but at the same time incrementally updates its structure to adapt to new process data when these are available without deteriorating the performance of the original (core) model and significantly disturbing the knowledge acquired from the initial database. Additional system's characteristics include the ability of the system to interact with the external environment in a perpetual mode (i.e. life-long learning mode) and having a dynamically expandable structure (i.e. an open structure) in which the system has the ability to add/create new rules and remove/prune redundant rules (knowledge maintenance).

In this research work, a number of key characteristics of incremental (perpetual) learning are adopted as follows [294, 301, 310]:

a) An initial data set is trained to construct the initial/original model
b) Each time, a new data set is sequentially – batch by batch – made available to the system
c) The system dynamically expands its structure to accommodate the uncovered data by the original model without significantly disturbing the original structure
d) There is the ability to 'memorise' the knowledge acquired from the original model
e) There is the ability to improve the system's structure over time by pruning rules that evolved to be inconsequential and/or redundant, which allows the model to be used in a perpetual learning mode.

In order to explain and understand the proposed perpetual learning framework, it would be useful to show its overall structure. Fig. 6.1 shows the overall structure of the proposed perpetual learning framework.

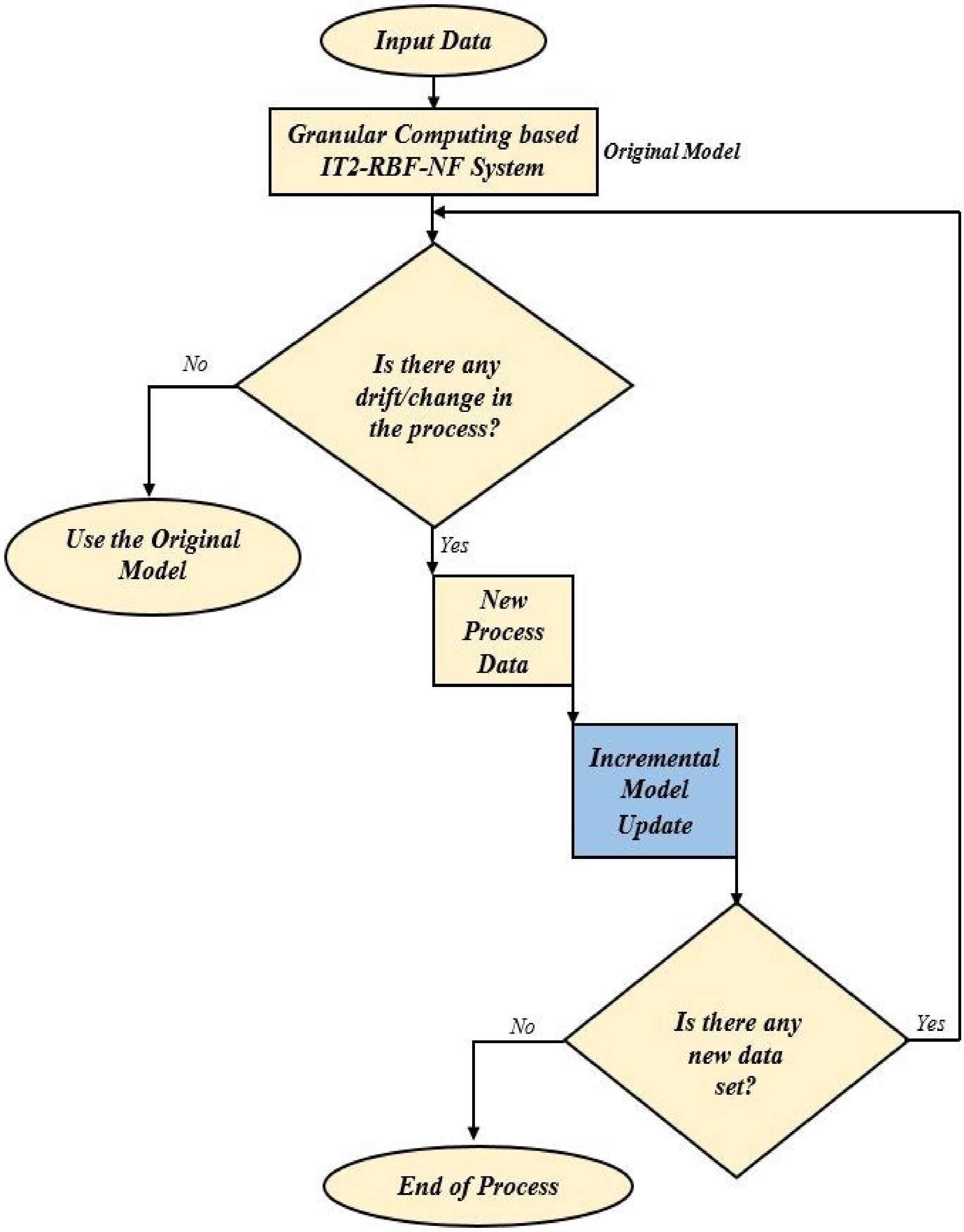


Figure 6.1. The structure of the perpetual learning framework.

The overall perpetual learning structure can be divided into components, which are described as follows:

- *Granular computing based IT2-RBF-NF System:* The iterative information-capture of granular computing is employed to elicit the initial structure of the IT2-RBF-NF model as it was described in Chapter 5. In this approach, one information granule corresponds directly to a fuzzy rule. The parameters of each rule are defined and then optimised via the adaptive-BEP algorithm.

- *New process data:* This includes the data that are belong to a totally or partially new input space – different data distribution – as compared to those used to construct the original model.
- *Incremental model update:* When the new process data are available to the model. The model can dynamically expand (modify) its structure to cover these data without significantly disturbing the original model.

To achieve the designed perpetual (incremental) learning framework, the following steps are introduced:

- *Novelty detection and the creation of new rules:* This step is used to filter and classify the new input data based on a pre-defined threshold into novel and partially novel data. The novel data are used to create new rules to cover the input space of these data. The new rules are optimised and then added to/merged with the original model rules to update the structure of the original model (old and new rules combined).
- *Interpretability improvement via iterative rule-pruning mechanism:* This aims at removing the inconsequential/redundant rules as a result of the incremental update (rule growing) process in order to improve overall the interpretability and to prevent the massive rules growth.
- *Accuracy improvement via constrained optimisation:* The pruned incrementally updated model is fine-tuned via a constrained parametric optimisation algorithm to improve its accuracy.

Fig. 6.2. depicts the incremental model update and a detailed description about the incremental model update process is provided in the following sections.

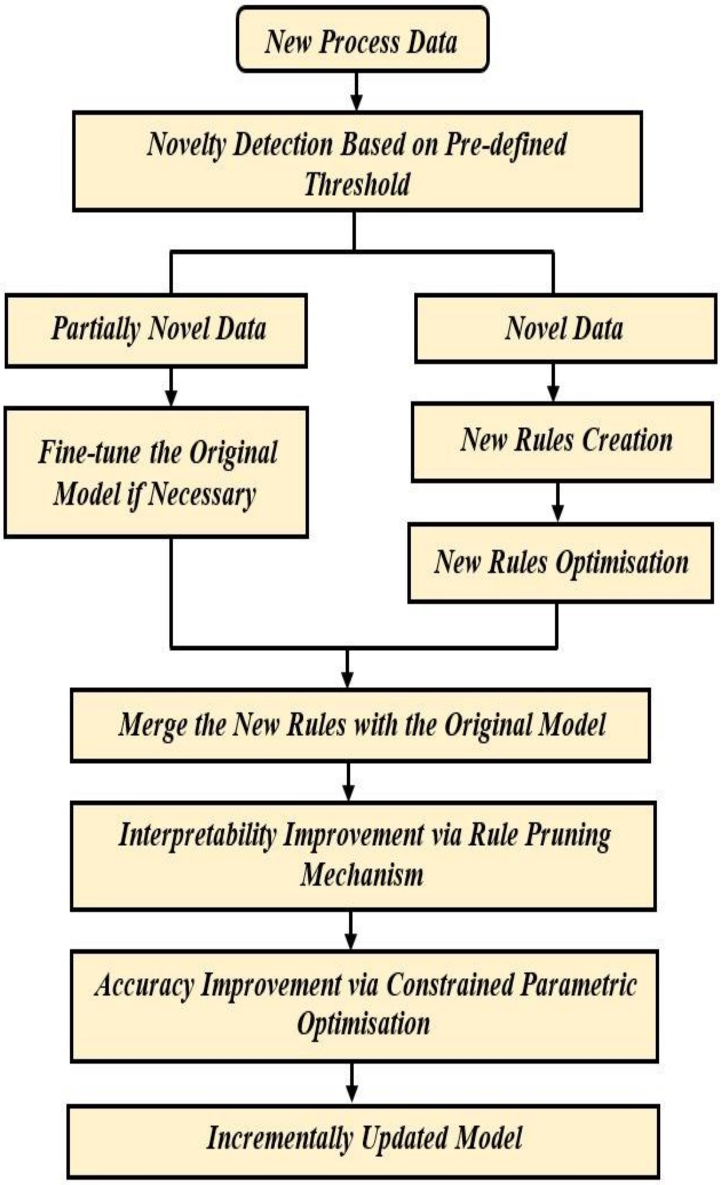


Figure 6.2. The structure of the incremental model update framework.

6.2.1. NOVELTY DETECTION AND NEW RULES CREATION

To start the perpetual learning process, when new data are available to the original model, they pass through a novelty detection unit before they are fed to the incremental learning process. Novelty detection is the process of identifying the new or partially new data that the original model is not aware of during the training process. Several statistical and neural networks based approaches can be used to estimate whether a test dataset belongs to the same distribution or not [311, 312]. In this research work, a simple novelty detection approach based on multidimensional Euclidean distance is used where the novelty is assessed by calculating the distance of each data sample from the cluster centres of previously elicited model (original model) [313]. The new data are then split into two data sets (namely novel and partially/non-novel data) based a predefined threshold. The novel data consist of the data that are totally belong to a different data distribution – new input space – as compared to the old/original process data, while the partially/non-novel data consist of the data that close or belong to the data distribution of the original data (i.e. mostly covered by the input space of the original data).

Each of the two data sets is treated differently by the incremental learning process. The partially/ non-novel data are fed to the existing model and if the model performance on these data is acceptable, then nothing to be modified to the existing model. Otherwise, the existing is fine-tuned without disturbing the existing structure (constrained tuning) to improve the performance of the original model on the partially/non-novel data. Since the input space of the non-novel/partially data is mainly covered by the original model (by one or more rules), there is no need to generate a new rule but only fine-tune the previously developed (existing) model. The novel data are utilised to generate new rules to cover the input space of the novel data, using the same GrC-IT2-RBF modelling approach. The new rules are optimised and then merged with the rest of the IT2-RBF-NF system rules to form a new IT2-RBF-NF model (see Fig. 6.2).

After the initial structure of the model is expanded to accommodate both the old data and new data, now the incremental learning structure contains all the knowledge required by the system. However, each time the model receives new data it generates new fuzzy rules and incrementally updates its structure. The proposed perpetual learning framework never removes the rules once created regardless of their importance. As a consequence of the incremental model updating process, a number of inconsequential/redundant rules will be generated. To circumvent the massive rule growth, an iterative rule pruning mechanism is introduced in order to dispose of the insignificant rules and improve the incrementally updated structure (i.e., interpretability improvement) as well as make the proposed framework feasible for lifelong learning.

6.2.2. INTERPRETABILITY IMPROVEMENT VIA RULE PRUNING MECHANISM

After the rule-base structure of the IT2-RBF-NF model is updated by introducing new rules to accommodate the new data, thus creating a model with more rules (updated model). As a result of the incremental updating process, the updated model often contains redundant information in terms of fuzzy rules. In addition, the continuous rule growth after each model-updating rule contradicts with the main the requirements to allow the incremental updating routine to be used in a perpetual learning mode. In this light, an iterative rule-based pruning approach is used to minimise/reduce the number of fuzzy sets in the universe of discourse of each input variable and eliminate possible inconsequential/redundant rules after each incremental updating routine. The proposed iterative rule-pruning mechanism can be achieved via a four-step procedure, which includes

1) removing redundant fuzzy sets,
2) merging similar fuzzy sets,
3) removing redundant fuzzy rules, and
4) merging similar fuzzy rules.

These four operations are controlled by thresholds $Th_1 - Th_4$. Fig. 6.1 shows a flowchart of the iterative rules pruning algorithm. A detailed description about the iterative rule-pruning algorithm is provided in the following sections.

The obtained rule-based system is improved in its structure, including the variation of the fuzzy sets and fuzzy rules, considering the simplicity and interpretability issues. Pruning the inconsequential rules results in distinguishable fuzzy sets and thus simplified fuzzy rules. It also controls the growth in the number of rules over time.

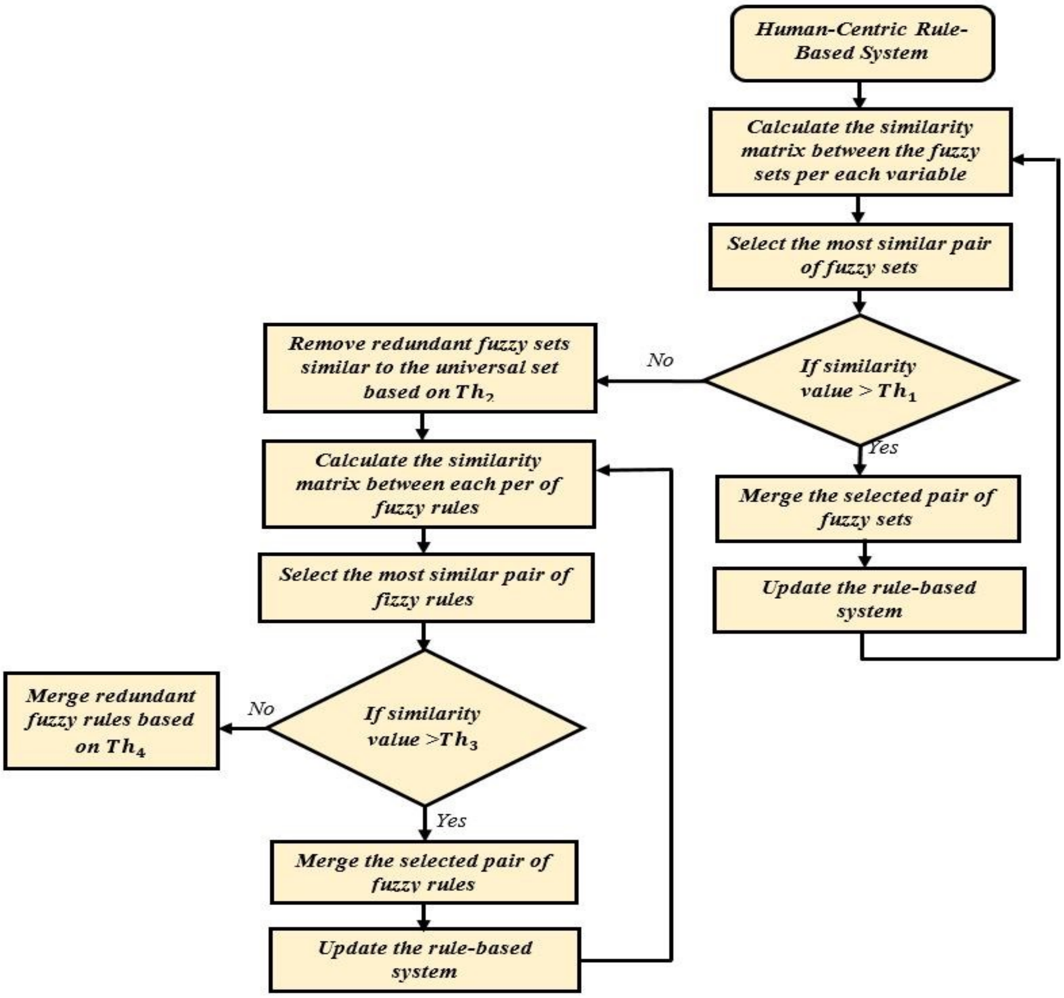
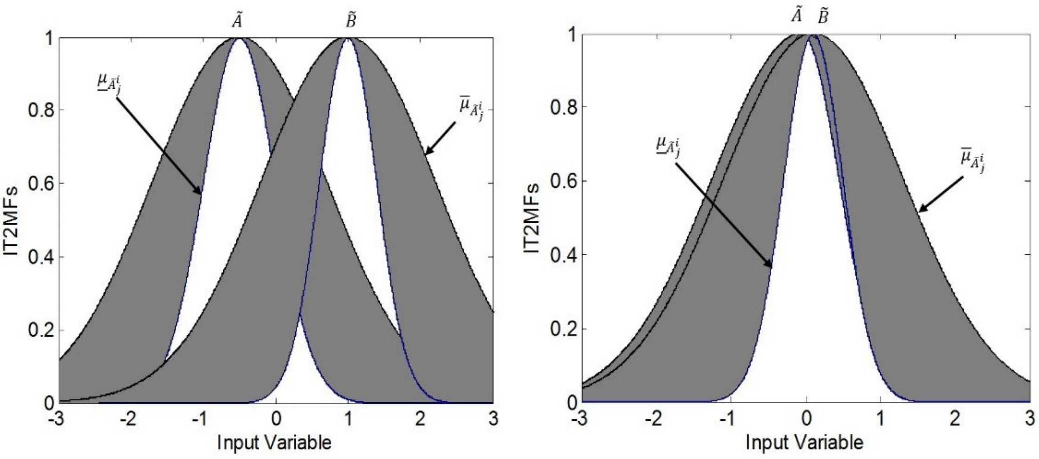


Figure 6.3. Iterative rule pruning algorithm.

6.2.2.1. MERGING SIMILAR FUZZY SETS

According to [241], when the rule-base is acquired from process data, it may consist of redundant/superfluous information in the form of similarity between fuzzy sets. A rule-based system with many similar fuzzy sets becomes superfluous, unnecessarily complex and computational expensive [44]. Since the linguistic interpretability of such a model lies in the idea of assigning qualitatively meaningful variables to fuzzy sets. However, it is difficult to assign qualitatively meaningful linguistic variables to highly similar fuzzy sets. Similarity between FSs can be defined as the degree to which the FSs are equal. For instance, Figs. 6.4(a) and 6.4(b) depicts two indistinguishable and distinguishable interval type-2 fuzzy sets, respectively. In Fig. 6.4(a), IT2-FSs are distinguishable with less degree of overlap, while in Fig. 6.4(b) IT2-FSs are seriously overlapped with high degree of similarity.



(a) (b)

Figure 6.4. Example of IT2 fuzzy sets: (a) distinguishable fuzzy sets. (b) Indistinguishable (highly overlapped) fuzzy sets.

To measure the degree of overlap between two fuzzy sets, a similarity measure is generally used. Although a quite extensive research has been carried out in the area of type-1 fuzzy sets similarity measures [237, 241, 314], only a few number of similarity measures for T2-FSs have appeared to date [315-317]. For two IT2-FSs $\widetilde{A}$ and $\tilde{B}$, calculation of their similarity degree $S(\widetilde{A}, \tilde{B})$ is much more complex than of their type-1 fuzzy sets counterparts, particularly for those with primary Gaussian MFs. In this research work, the Jaccard's similarity measure [316] is used to measure the similarity between two type-2 FSs. The Jaccard's similarity measure between two IT2FSs $\widetilde{A}$ and $\tilde{B}$ is defined as

$$S_J(\widetilde{A}, \tilde{B}) = \frac{f(\widetilde{A} \cap \tilde{B})}{f(\widetilde{A} \cup \tilde{B})} \quad 6\text{-}1$$

where f is a function satisfying $f(\widetilde{A} \cup \tilde{B}) = f(\widetilde{A}) + f(\tilde{B})$ for disjoint $\widetilde{A}$ and $\tilde{B}$.

For simplicity, the function f is chosen as the cardinality (Card). Then the Jaccard's similarity measure can be written as

$$S_J(\widetilde{A}, \tilde{B}) = \frac{Card(\widetilde{A} \cap \tilde{B})}{Card(\widetilde{A} \cup \tilde{B})} \quad 6\text{-}2$$

$$S_J(\tilde{A},\tilde{B}) = \frac{\int_X \min\left(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}}(x)\right) dx + \int_X \min\left(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}}(x)\right) dx}{\int_X \max(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}}(x)) dx + \int_X \max(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}}(x)) dx} \quad 6\text{-}3$$

The similarity measure can be used to quantify/estimate the degree of similarity between IT2-FSs in the rule base. If the similarity value $S_J(\tilde{A},\tilde{B})$ is larger than a predefined threshold $S_{J_{Th}}$ then $\tilde{A}$ and $\tilde{B}$ are considered as being highly overlapped fuzzy sets. Therefore, $\tilde{A}$ and $\tilde{B}$ can be merged to form a new FS that is representative of the merged FSs. The choice of a suitable threshold $S_{J_{Th}}$ is an application dependent. The lower the value of the threshold, the more fuzzy sets are merged. Generally, there are three possible methods for merging highly overlapped fuzzy sets [241]:

1) replace $\tilde{A}$ by $\tilde{B}$;
2) replace $\tilde{B}$ by $\tilde{A}$;
3) replace both $\tilde{A}$ and $\tilde{B}$ by a new fuzzy set $\tilde{C}$.

In this research work, the third method is used, where the newly merged IT2-FS $(\tilde{C})$ is formed by the combination of the FOUs of both $\tilde{A}$ and $\tilde{B}$ in order not to lose any information from the merging process. By substituting this new fuzzy set representative of the merged sets for each input variable in the rule-base, the number of FSs to constitute the rule-based system is reduced as shown in Fig. 6.5.

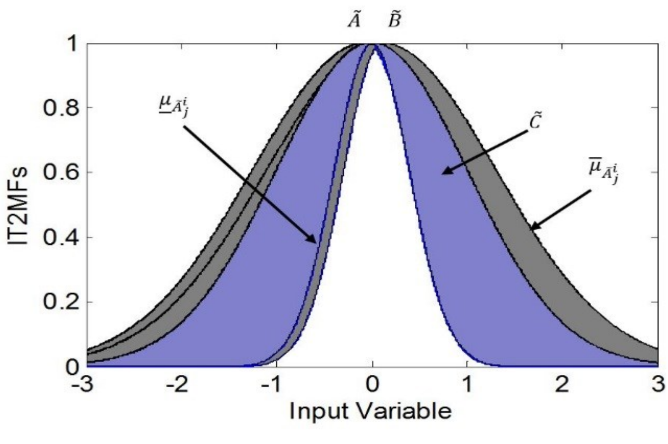
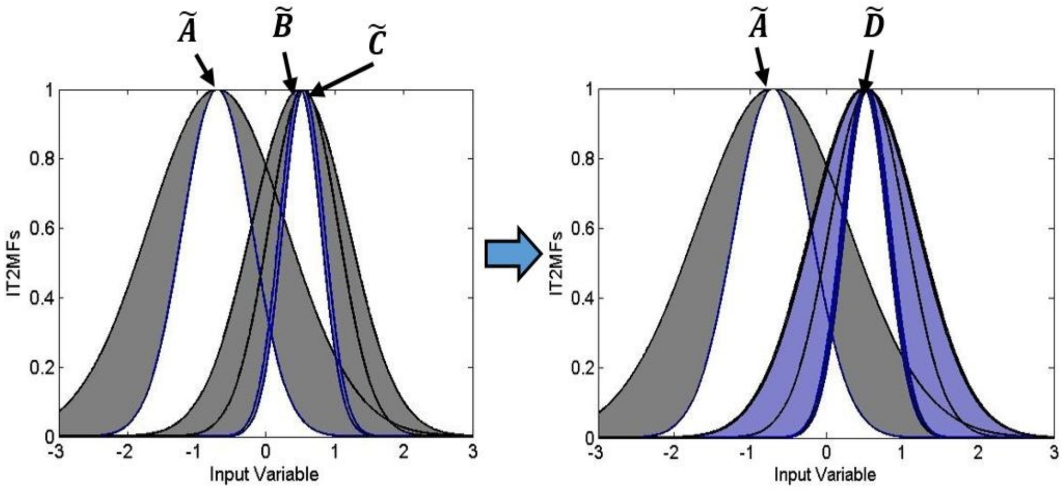


Figure 6.5. Merging similar fuzzy sets.

To illustrate the concept, Table 6-1 shows the similarity matrix representation for the IT2-FSs in Fig. 6.6(a). A threshold Th_1 for merging similar FSs is then defined, where $Th_1 \in [0,1]$. If $S_J(\tilde{A}, \tilde{B}) > Th_1$, i.e., the FSs $\tilde{A}$ and $\tilde{B}$ are highly overlapped, then these two FSs should be merged into one new FSs. With $Th_1 = 0.75$ and from Table 6-1, $S_J(\tilde{B}, \tilde{C}) > Th_1$, then the FSs $\tilde{B}$ and $\tilde{C}$ are merged to form a new FS $\tilde{D}$ without losing any information from both $\tilde{B}$ and $\tilde{C}$. The resulting FSs in the same universe of discourse after the merging process is shown in Fig. 6.6(b).



(a) (b)

Figure 6.6. Example of three IT2 fuzzy sets: (a) IT2-FSs before the merging operation. (b) Resulting IT2-FSs after the merging operation.

Table 6.1 Similarity matrix for the three IT2-FSs in Fig. 6.4 (a) when the Jaccard similarity matrix is used.

IT2-FS	$\tilde{A}$	$\tilde{B}$	$\tilde{C}$
$\tilde{A}$	1.0000	0.2277	0.1791
$\tilde{B}$	0.2277	1.0000	0.7645
$\tilde{C}$	0.1791	0.7645	1.0000

6.2.2.2. REMOVING REDUNDANT FUZZY SETS

A fuzzy set in the antecedent part of linguistic rule is said to be redundant if it has a MF $\mu_{\tilde{A}}(x) \approx 1, \forall x \in X$, it is similar to the universal set $U(\mu_U(x) = 1)$ and can be removed [241]. The similarity of a FS $\widetilde{A}$ to the universal set U is calculated by $S_J(\tilde{A}, U)$. If the $S_J(\tilde{A}, U)$ is greater than a predefined threshold Th_2, then this FS is considered as a redundant FS. As a result, the corresponding FS should be removed. Fig. 6.7 depicts an illustration of the IT2-MFs relating to a redundant FS.

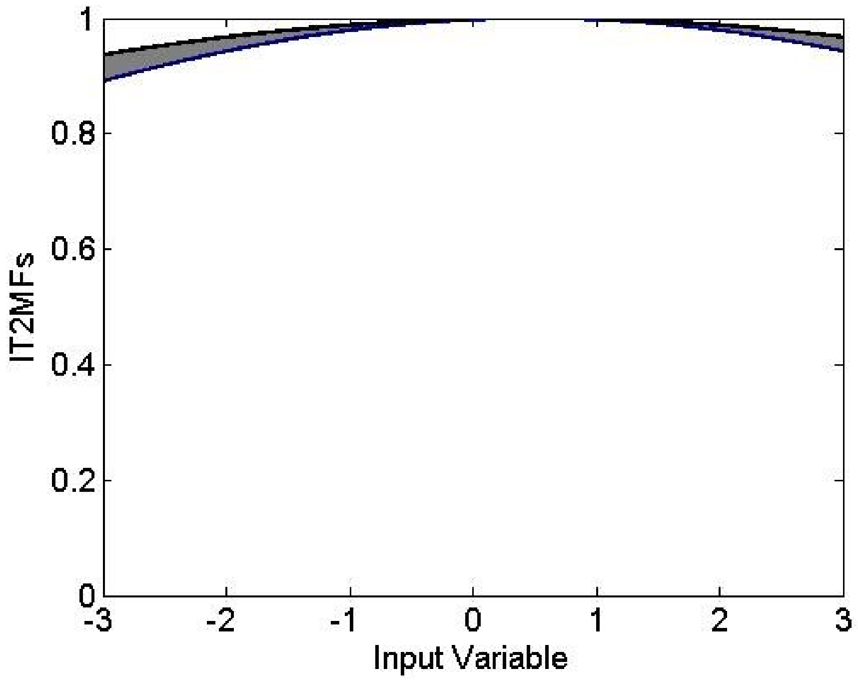


Figure 6.7. An illustration of the IT2-MFs relating to a redundant fuzzy set.

6.2.2.3. MERGING SIMILAR FUZZY RULES

Two fuzzy rules are said to be similar enough for merging if only the antecedent of the rules are equal and the consequents do not [318]. Two fuzzy rules with different consequents but very similar antecedent parts usually indicates conflicting rules [318]. Therefore, conflicting rules are either merged together to form a new rule or one of them is removed. For instance, the following linguistic rules may be regarded as similar:

$$\boldsymbol{Rule_1}: \boldsymbol{IF}\ x_1\ is\ 'low'and\ x_2\ is\ 'high', \boldsymbol{THEN}\ y\ is\ 'high'$$
$$\boldsymbol{Rule_2}: \boldsymbol{IF}\ x_1\ is\ 'low'and\ x_2\ is\ 'high', \boldsymbol{THEN}\ y\ is\ 'medium'$$

The above two rules can be merged into one rule as follows:

$$\boldsymbol{Rule_{new}}: \boldsymbol{IF}\ x_1\ is\ 'low'and\ x_2\ is\ 'high', \boldsymbol{THEN}\ y\ is\ 'slightley\ high'$$

To evaluate the similarity degree between two linguistic fuzzy rules, the similarity measure of every FS pair has to be calculated [318]. For the jth fuzzy rule $Rule_j$, the corresponding IT2-FSs are $\tilde{A}_1^j, \tilde{A}_i^j, \dots, \tilde{A}_d^j$. In similar fashion, the corresponding antecedent parts of the kth fuzzy rule $Rule_k$, the are $\tilde{B}_1^k, \tilde{B}_i^k, \dots, \tilde{B}_d^k$. Therefore, the similarity measure can be expressed as follows:

$$S_J(Rule_j, Rule_k) = \prod_{i=1}^{d} S_J(\tilde{A}_i^j, \tilde{B}_i^k) \tag{6-4}$$

where $S_J(\tilde{A}_i^j, \tilde{B}_i^k)$ is the Jaccard similarity measure of two IT2-FSs $\tilde{A}_i^j$ and $\tilde{B}_i^k$ and it is defined in Section 6.2.3.1. If $S_J(Rule_j, Rule_k)$ is greater than a predefined threshold Th_3, then the FS pairs of these two fuzzy rules are similar. Therefore, these two rules are also considered to be similar and then merged into a new rule $Rule_{new}$. The antecedent and consequent parts of the new rule $Rule_{new}$ are obtained via the merging operation described in Section 6.2.3.1.

6.2.2.4. REMOVING REDUNDANT FUZZY RULES

If the membership value of an IT2-FS is always near zero over its entire universe of discourse, i.e., $\mu_{\tilde{A}}(x) \approx 0, \forall x \in X$, its corresponding rule is considered redundant [241]. Since this redundant rule will almost never be fired, which means its output is always near zero. A threshed Th_4 is also defined to determine whether the rule is redundant or not. If $\mu_{\tilde{A}}(x) < Th_4$, then the corresponding fuzzy rule is deemed redundant. Therefore, the redundant rule should be removed [241]. In general, Th_4 is defined in the range [0, 0.01].

The rule-pruning algorithm depicted in Fig. 6.3 is an iterative algorithmic process where at each iteration, the similarity measure between all pairs of IT2-FSs for each input variable is calculated. The pairs of IT2-FSs having the highest similarity value $S_J(\widetilde{A}, \tilde{B}) > Th_1$ are merged to form a new IT2-FS. Then the rule-base of the IT2-RBF-NF model is updated by substituting this new IT2-FS for the IT2-FS merged to form it. The process of calculating the similarity measure on the updated rule-base structures continues until there are no more IT2-FSs for which $S_J(\widetilde{A}, \tilde{B}) \geq Th_1$. Thereafter, the

IT2-FSs that have similarity $S_J(\widetilde{A}, \widetilde{B}) \geq Th_2$ to the universal set U are removed. Finally, the similarity measure between all pairs of linguistic rules for entire rule-base is computed. The pairs of fuzzy rules having the highest similarity value $S_J(\widetilde{A}, \widetilde{B}) > Th_3$ are merged to form a new IT2-FS. Repeat the process of merging similar fuzzy rules until there are no more rules for which $S_J(\widetilde{A}, \widetilde{B}) \geq Th_3$. Then the redundant rules are removed based a predefined threshold Th_4. The rule-pruning algorithm is summarised as follows:

Iterative rule pruning algorithm:

Given a linguistic rule-base $\boldsymbol{R} = \{\boldsymbol{Rule_j}\}_{\boldsymbol{j=1}}^{\boldsymbol{N}}$, where $\boldsymbol{Rule_j}$ is the $\boldsymbol{jth}$ rule, choose thresholds $\boldsymbol{Th_1 - Th_4}$

$$\boldsymbol{Rule_j: IF\ x_1\ is\ \widetilde{A}_1^j\ and, \dots, and\ x_d\ is\ \widetilde{A}_d^j, THEN\ y\ is\ \widetilde{B}_1^j}$$

Step 1: Calculate the similarity matrix in $\boldsymbol{R}$, $\boldsymbol{S_J(\widetilde{A}_i^j, \widetilde{B}_i^k)}$, $\boldsymbol{i = 1, \dots, d}$, $\boldsymbol{j = 1, \dots, N}$, $\boldsymbol{k = 1, \dots, N}$. Select two most similar IT2-FSs $\boldsymbol{S_{Jmax}(\widetilde{A}_i^j, \widetilde{B}_i^k) = max_{j \neq k}\{S_{Jmax}(\widetilde{A}_i^j, \widetilde{B}_i^k)\}}$.

Step 2: If $\boldsymbol{S_{Jmax}(\widetilde{A}_i^j, \widetilde{B}_i^k) \geq Th_1}$, then merge the two IT2-FS to form a new fuzzy set. Continue until: no more IT2-FSs have similarity measure such that $\boldsymbol{S_J(\widetilde{A}_i^j, \widetilde{B}_i^k) \geq Th_1, j \neq k}$.

Step 3: Calculate similarity measure of a FS $\boldsymbol{\widetilde{A}_i^j}$ to the universal set $\boldsymbol{S_J(\widetilde{A}_i^j, U)}$. If the similarity value $\boldsymbol{S_J(\widetilde{A}_i^j, U) \geq Th_2}$, then the $\boldsymbol{\widetilde{A}_i^j}$ is considered to be a redundant fuzzy set and should be removed from the antecedent of $\boldsymbol{Rule_j}$.

Step 4: Calculate the similarity matrix between the rules in $\boldsymbol{R}$,

$$\boldsymbol{S_J(Rule_j, Rule_k) = \prod_{i=1}^{d} S_J(\widetilde{A}_i^j, \widetilde{B}_i^k)}$$

where $\boldsymbol{S_J(\widetilde{A}_i^j, \widetilde{B}_i^k)}$ is the Jaccard similarity measure of two IT2-FSs $\boldsymbol{\widetilde{A}_i^j}$ and $\boldsymbol{\widetilde{B}_i^k}$. If $\boldsymbol{S_J(Rule_j, Rule_k) \geq Th_3}$, then the IT2-FS pairs of these two fuzzy rules are similar and merged into a new rule $\boldsymbol{Rule_{new}}$.
Continue until: no more rules have similarity measure such that $\boldsymbol{S_J(Rule_j, Rule_k) \geq Th_3, j \neq k}$.

Step 5: Remove the redundant fuzzy rules if $\boldsymbol{\mu_{\widetilde{A}_i^j}(x) \leq Th_4, \forall x \in X}$.

6.2.3. ACCURACY IMPROVEMENT VIA CONSTRAINED OPTIMISATION

After rule-base pruning achieved via rule pruning mechanism shown above, the obtained rule-based system is structurally simpler and interpretably more tractable. However, with less modelling performance compared to the originally created model. In rule-based systems, interpretability and accuracy are two contradictory and conflicting modelling requirements [19], as improving interpretability of rule-based systems – pruning the inconsequential rules – generally degrades the performance of the model and vice versa. To improve the accuracy of the pruned model, a parametric fine-tuning based on the same adaptive-BEP in the parametric fine-tuning process is introduced. A good trade-off between model accuracy and interpretability requires imposing constraints on the parametric optimisation. To better understand this trade-off, Fig. 6.8 depicts a trade-off relation between accuracy and interpretability based on constraints.

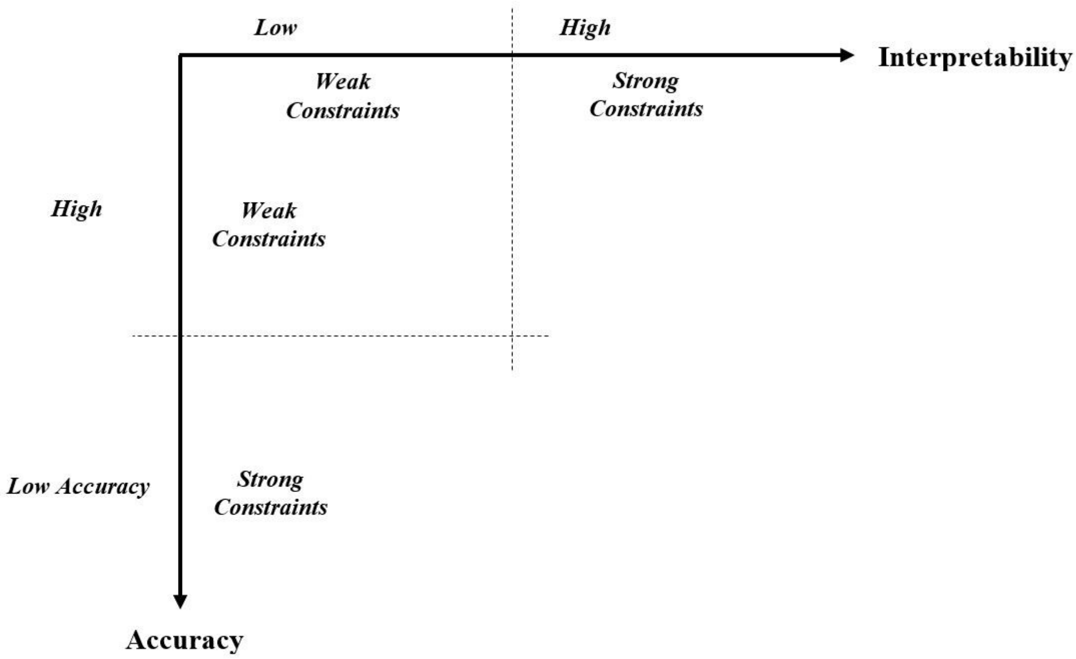


Figure 6.8. A typical trade-off relation between accuracy and interpretability based on imposing constraints.

Clearly from the trade-off relation above, imposing strong constraints on the parametric optimisation leads to models that are very poor in terms of accuracy (i.e., keep some of the information from the pruned model), even though the results obtained for interpretability can be satisfactory, the overall model accuracy is in general very poor. In contrast to imposing relaxed constraints (i.e., keep most of the information from the pruned model), a satisfactory level of model accuracy can be achieved. However, as a consequence of imposing the relaxed constraints, model interpretability may not be good enough.

To further understand the proposed framework, an algorithmic procedure of the proposed perpetual learning framework is illustrated as follows:

Perpetual Learning Algorithm:

Step 1: The first batch of data is available.
Step 2: Generate the initial interval type-2 fuzzy rules:
- ***Step 2-1:*** Determine the antecedent parameters from the iterative data granulation process.
- ***Step 2-2:*** Initialise the values of the consequent parameters.
- ***Step 2-3:*** Use the adaptive back error propagation to optimise the initial IT2-RBF-NF structure.
- ***Step 2-4:*** Measure the performance of the model using RMSE.

Step 3: A new batch of data is available.
Step 4: Pass the new batch of data set into the novelty detection algorithm using a predefined threshold.
- ***Step 4-1:*** Construct new rules to accommodate the novel data.
- ***Step 4-1-1:*** Generate the initial new rules using the iterative data granulation process.
- ***Step 4-1-2:*** Optimise the initial structure of the new rules by using the adaptive back error propagation algorithm.

Step 5: Merge/Integrate the original model with the generated rules to expand the structure of the original model.
Step 6: Apply the iterative pruning mechanism described in Section 6.2.2 to manage the redundant rules and improve the structure of the incrementally updated model.
Step 7: Optimise the pruned structure with constraints and measure its performance using RMSE on both old data and the new data.
Step 8: If new incoming data points are available go to Step 4; otherwise stop.

6.3. SIMULATION RESULTS

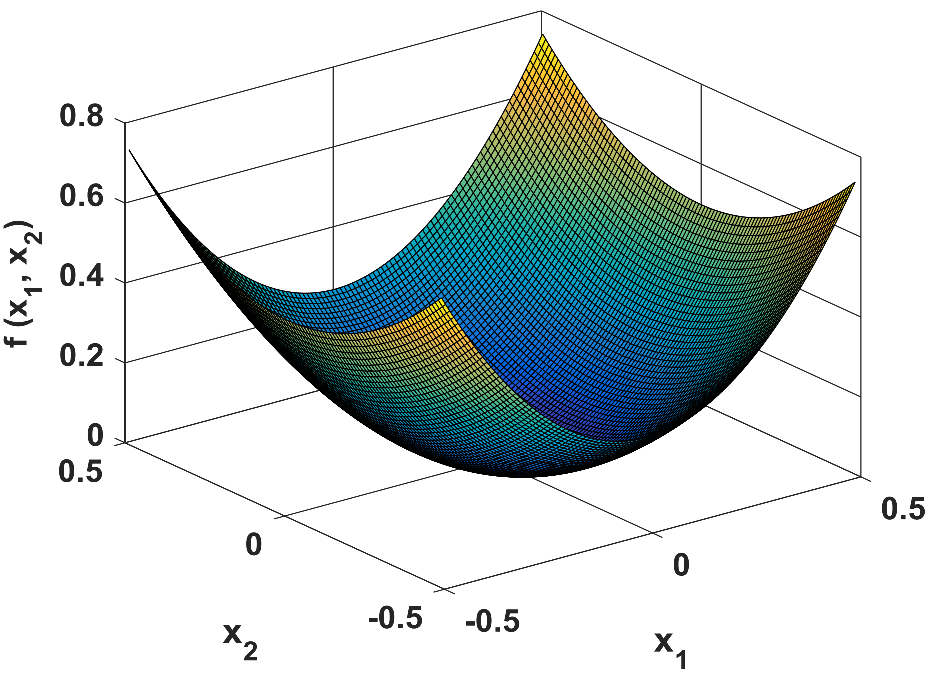
To confirm the efficiency of any modelling framework, it is common to test its performance on well-characterised benchmark functions. Therefore, in the first simulation the performance of the proposed framework is demonstrated using a number of simple uni-modal non-linear benchmark function as well as two complex non-linear benchmark functions. These non-linear functions are probably one of the most popular benchmarking functions in system identification. The second simulation uses a multi-modal non-linear function to confirm the sustainability of the proposed framework in case where more frequent model updates are required with good prediction performance. Finally, the real-industrial case study of steel friction stir welding for the predictive modelling of the spindle peak torque is used. FSW exhibits highly non-linear and complex as well as sparse databases as a consequence of its thermo-mechanical complexity.

6.3.1. EXAMPLE 1: UNI-MODAL FUNCTION IDENTIFICATION

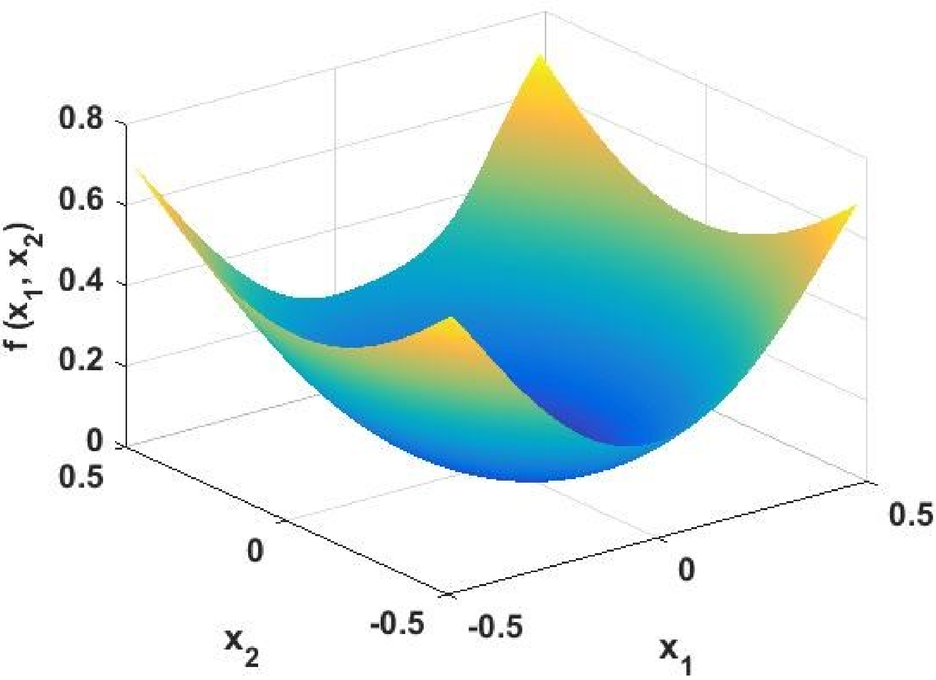
The system to be identified is represented by an equation as follows:

$$f(x_1, x_2) = 2x_1^2 + x_2^2 \qquad \text{6-5}$$

The non-linear static system is taken from [319]. One hundred data points were generated randomly from $-0.5 \leq (x_1, x_2) \leq 0.5$ and the corresponding output data were obtained from Eq. 6-5. The data set has been divided into 75 (75%) data points to train the model and 25 (25%) data points to test the prediction performance of the final model. The training raw data are granulated into 5 information granules (optimal number of information granules in this case) via the iterative data granulation process as shown in Chapter 4. The extracted optimal information granules are then mapped into linguistic type-2 fuzzy logic rules to elicit the initial structure of the IT2-FLS rule-base. Once the initial structure of the IT2-FLS rule-base (5 fuzzy rules) is obtained, the initial IT2-RBF-NF structure (see Chapter 4) is optimised via the adaptive-BEP algorithm. After structure identification and parametric optimisation, a 5-rule-based model was produced. The prediction error of RMSE = 0.0040 and 0.0031 for training and testing respectively. The system approximation results using the produced fuzzy model are shown in Fig. 6.9 and 6.10.



(a) Actual system output.



(b) Output of the IT2-RBF-NF system.

Figure 6.9. Non-linear system approximation.

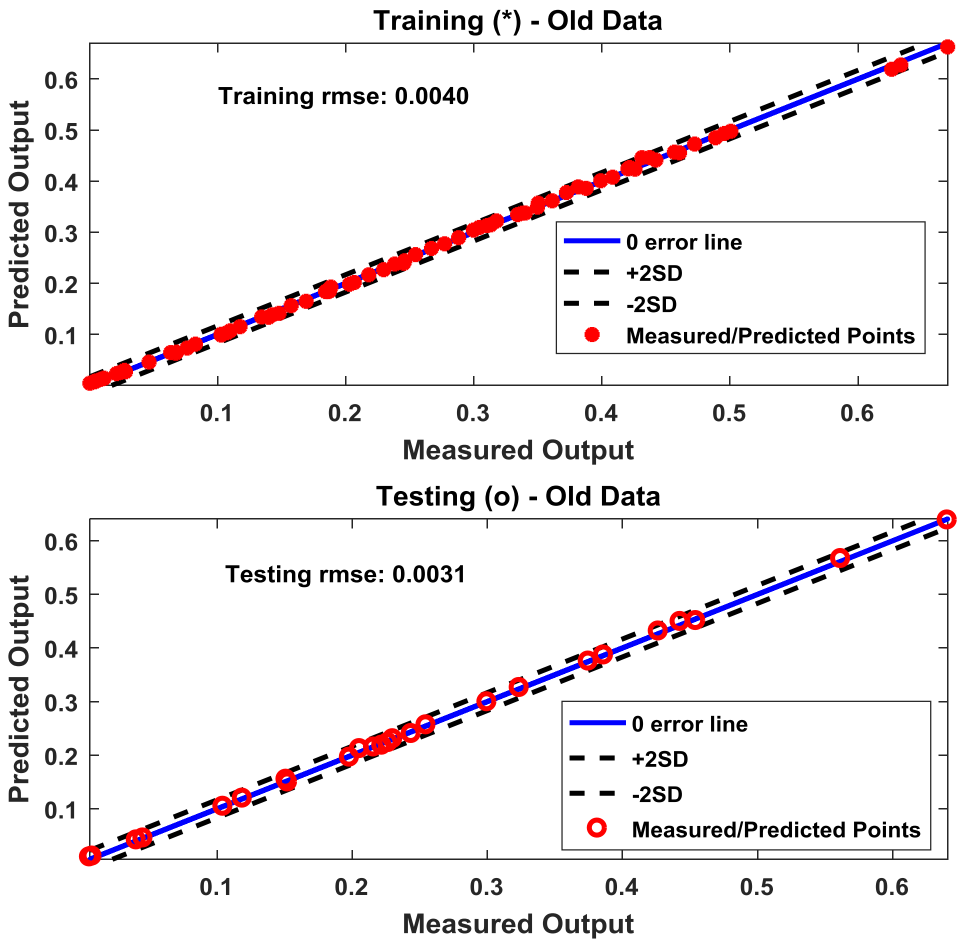


Figure 6.10. Measured versus predicted output of the IT2-RBF-NF model for training and testing.

Perpetual Learning Performance:

In this modelling scenario, a hundred random data points in the range between $-0.65 \leq (x_1, x_2) \leq 0.65$ (i.e. generate new data that are not covered by the original data) was generated. When the new data are available to the developed model, they pass through a novelty detection unit before they are fed to the incremental learning process. The novelty detection unit based on multidimensional Euclidean distance where the novelty is assessed by measuring the distance of each data sample from the cluster centres of the core model. The new data are then split into two data sets (namely novel and partially/non-novel data) based a predefined threshold (set to 0.5 in this case). The novel data contain the data that belong to a completely new data distribution –

totally new input space – as compared to the original data, while the partially/non-novel data consist of the data that close or belong to the data distribution of the original data (i.e., mostly covered by the input space of the original data). The performance of the model is tested on the new data (both partially novel and totally novel data), as it was expected, the performance of the model on the novel data is far worse than its performance on the partially novel data since the input space of the novel data is totally new and never seen by the original model. The performance of the model on the partially novel and totally novel is presented in Table 6-2.

Table 6.2 Performance of the original model on the new data for non-linear function approximation in Example 1.

IT2-RBF-NF system	**Initial Model (Core Model)**
Number of rules	5
Number of parameters	20
RMSE for partially novel data	0.01
RMSE for totally novel data	0.08

Each of the two data sets is treated differently by the incremental learning algorithm. The partially/ non-novel data are fed to the existing model directly, and if the model performance on this data set is acceptable, then nothing to be modified to the existing model. Otherwise, the existing is fine-tuned without significantly disturbing the existing structure (constraint tuning) to improve the performance of the core model on the partially/non-novel data. The novel data set is used to generate new rules to cover the input space of the novel data set, using the same GrC-IT2-RBF modelling approach. The new rules are then added to the rest of the existing IT2-RBF-NF model rules. The new data set produced 4 rules, and they are combined with the existing 5 rules to construct a new IT2-FLS with 9 rules.

Using the proposed iterative rule-pruning mechanism described in Section 6.2.2, two rules were merged based on merging thresholds $Th_1 - Th_3$ of 0.8 and Th_4 of 0.01 in this case. The simplified 8 rules FLS is then fined-tuned to improve its accuracy. The performance of the updated model is tested on the old/original data set as well as the new data set (training and testing). The results are shown in Fig. 6.11.

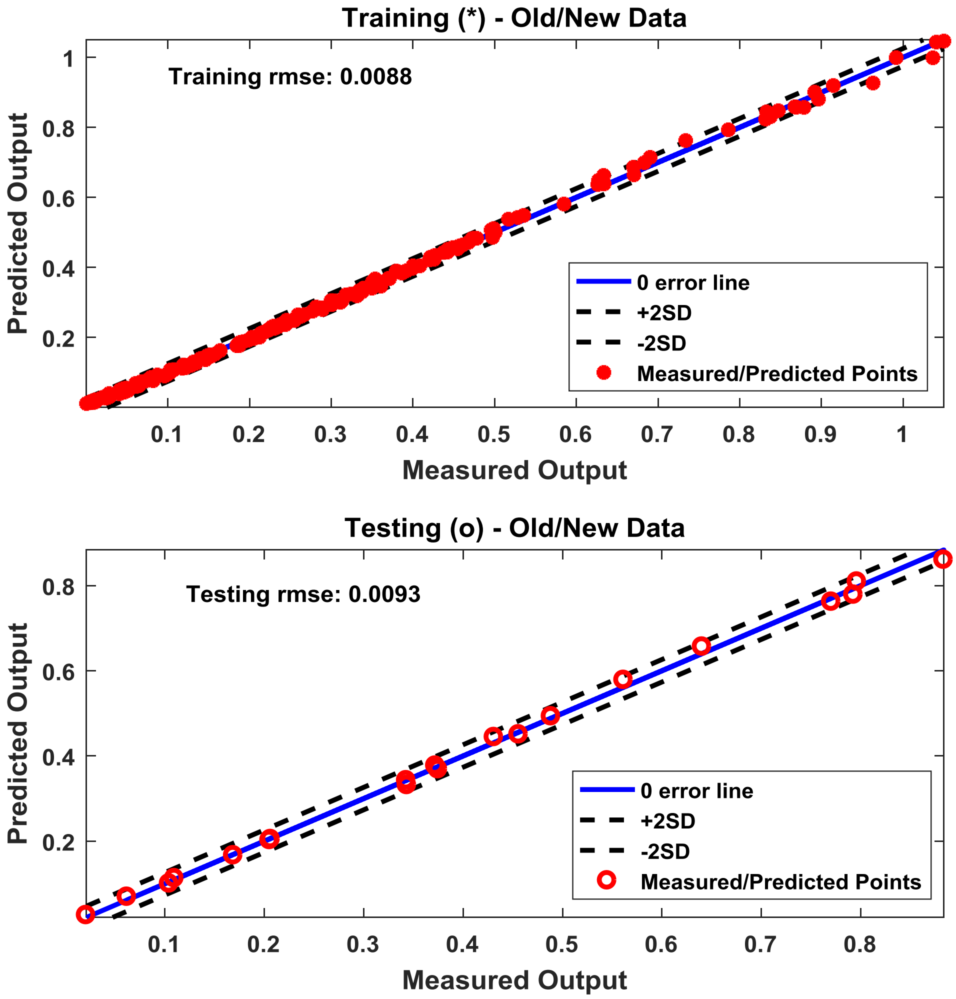


Figure 6.11. Performance of the updated model on the whole data set.

From the simulation results, it is clear that the incremental updating algorithmic process provides a reliable model updating procedure that results in an open structure (i.e., dynamically expandable structure) without neglecting any previously gained knowledge. It is shown from Table 6-3 that the proposed framework is able to model/learn from an initial process data and incrementally updates its structure when needed and at the time improves its structure by removing the inconsequential rules. The performance of the updated model on the original data set is maintained, and its performance on the new data after the incremental update is comparable to the original performance (see Fig. 6.11).

Table 6.3 Performance of the original model and updated model for non-linear function approximation in Example 1.

IT2-RBF-NF system	Initial Model (Core Model)	Incrementally updated model
Number of rules	5	8
Number of parameters	20	26
Training RMSE	0.0040	0.0089
Testing RMSE	0.0031	0.0093

6.3.2. EXAMPLE 2: MULTI-MODAL NON-LINEAR FUNCTION IDENTIFICATION

This example employs the proposed framework to model a well-known complex multi-modal benchmark function, which is taken from [320]. The multi-modal function is generated from the following equation:

$$f(x_1, x_2) = sin\left(2\pi.\sqrt{(x_1^2 + x_2^2)}\right) \qquad 6\text{-}6$$

One thousand data points were generated randomly from $-0.5 \leq (x_1, x_2) \leq 0.5$ and the corresponding output data were obtained from Eq. 6-6. The data set has been divide into 75 (75%) data points to train the model and 25 (25%) data points to test its prediction performance. The training raw data are granulated into 13 information granules (optimal number of information granules) via the iterative IG approach. The granulated data are then mapped into linguistic type-2 fuzzy rules to elicit the initial structure of fuzzy rule-base. After the initial structure of IT2-FLS (13 fuzzy rules) is obtained, the initial IT2-RBF-NF structure is optimised via the adaptive-BEP approach. After structure identification and parametric optimisation, a 13-rule model was produced. The simulation results of the system approximation, with RMSE = 0.0090 and 0.0079 for training and testing respectively are shown in Figs. 6.12 and 6.13.

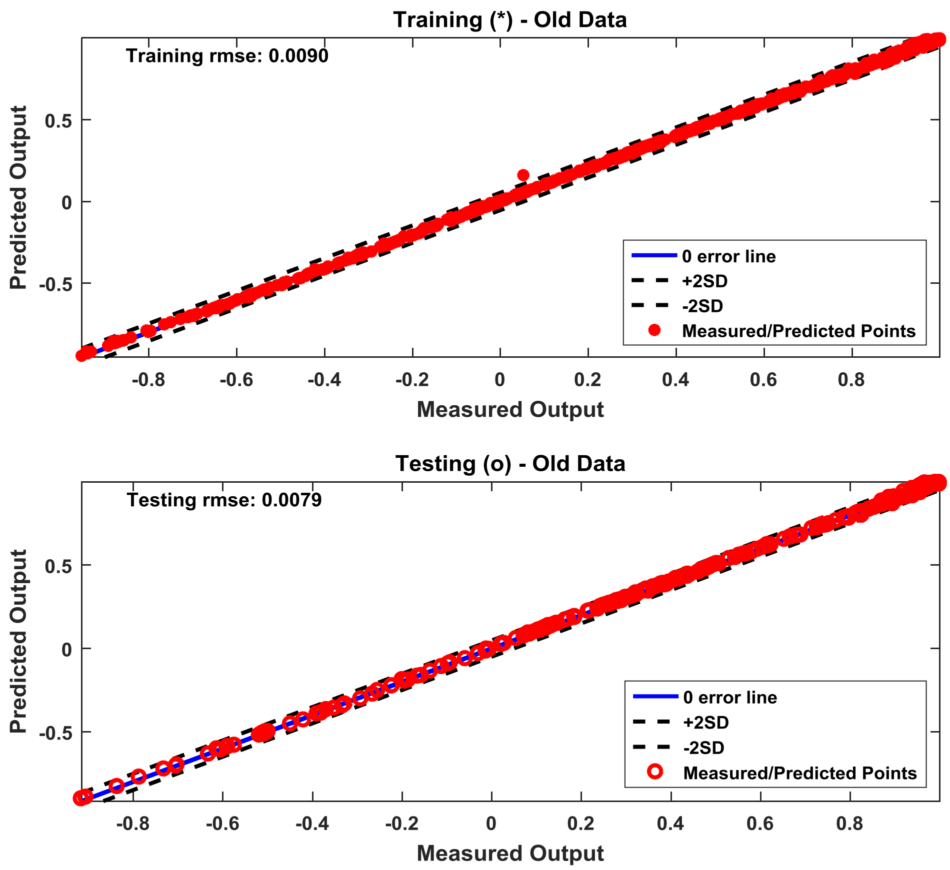
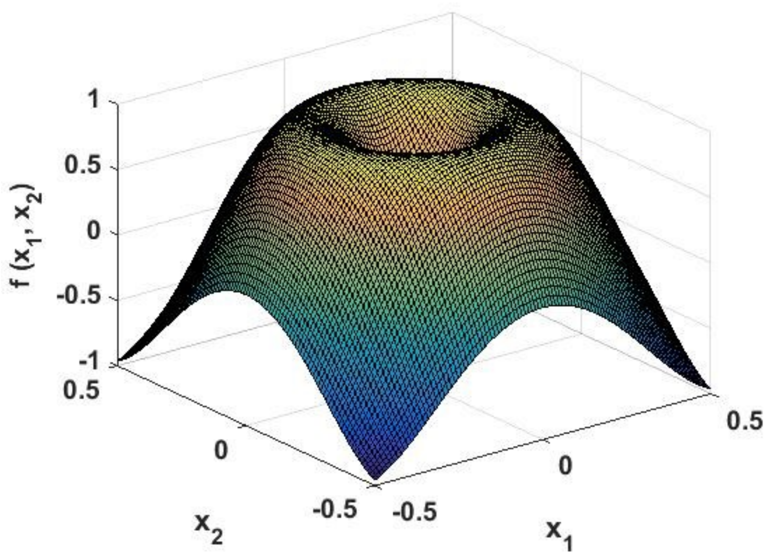
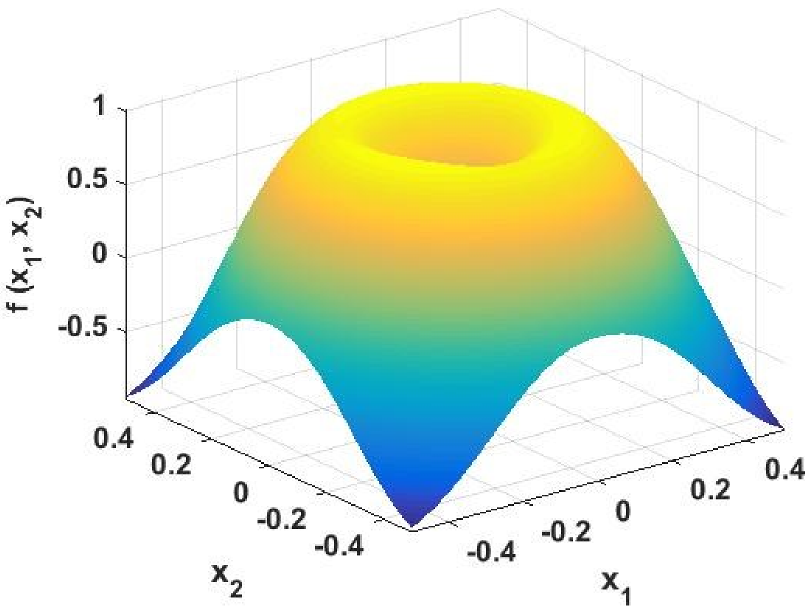


Figure 6.12. Measured versus predicted output of the IT2-RBF-NF model for training and testing.



(a) Actual system output.



(b) Output of the IT2-RBF-NF system.

Figure 6.13. Non-linear system approximation.

Perpetual Learning Performance:

To test the generalisation ability of the proposed incremental learning structure in a more complex system identification problem, a synthetic new data set that contains both novel and partially novel data in the range $-0.65 \leq (x_1, x_2) \leq 0.65$ was generated. The performance of the original model on the new data is presented in Table 6-4.

Table 6.4 Performance of the original model on the new data for non-linear function approximation in Example 2.

IT2-RBF-NF system	Initial Model (Core Model)
Number of rules	13
Number of parameters	52
RMSE for partially novel data	0.0092
RMSE for totally novel data	0.1431

After the new data are passed though the incremental update process, 14 new rules are generated from the iterative data granulation process. The updated model (27 rule-based system) is then pruned and fine-tuned to construct the final updated model (20 rules), when the pruning thresholds were set to $Th_1 - Th_3$ of 0.8 and Th_4 of 0.01. Fig. 6.14 shows the performance of the updated model on both the old and new data. For the purpose of comparison, Table 6-5 also shows the performance of the original model (13 rule-based system) and the updated model (20 rule-based system). As it can be seen, the incremental learning framework has an adaptive behaviour by incrementally updating itself to accommodate the unseen input data set. The incrementally updated model is able to retain a good performance without ignoring any previously gained knowledge.

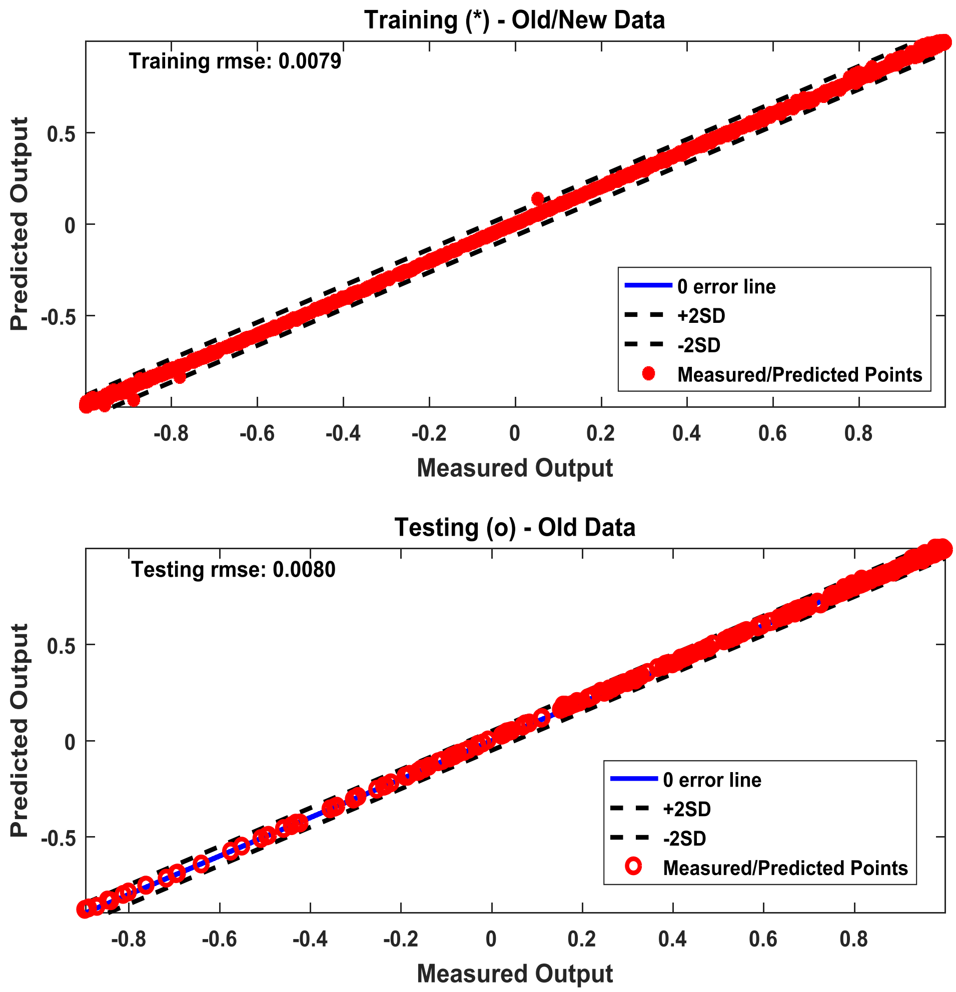


Figure 6.14. Performance of the updated model on the whole data set.

Table 6.5 Performance of the original model and updated model for non-linear function approximation in Example 2.

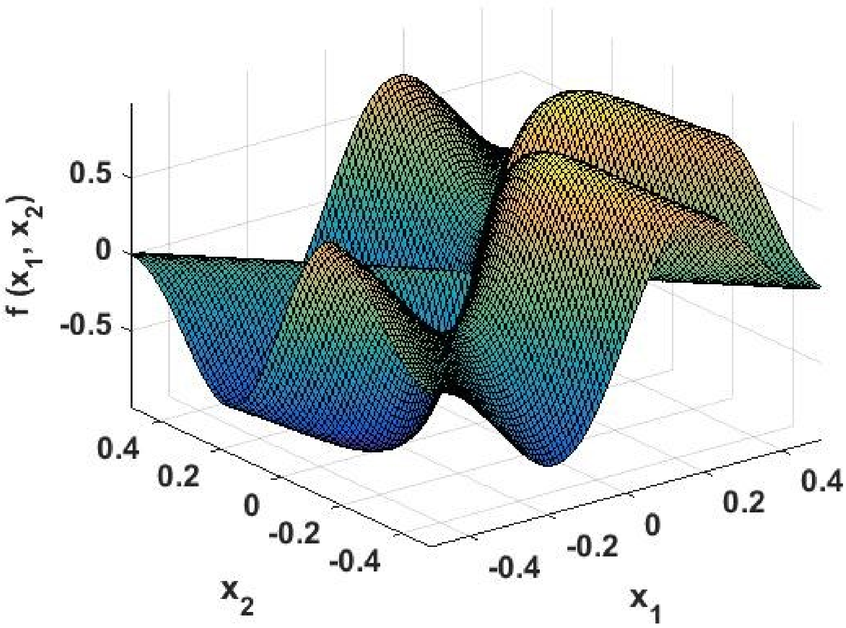
IT2-RBF-NF system	Initial Model (Core Model)	Incrementally updated model
Number of rules	13	20
Number of parameters	52	80
Training RMSE	0.0090	0.0079
Testing RMSE	0.0079	0.0080

6.3.3. EXAMPLE 3: MUTI-MODAL BUTTERFLY FUNCTION IDENTIFICATION

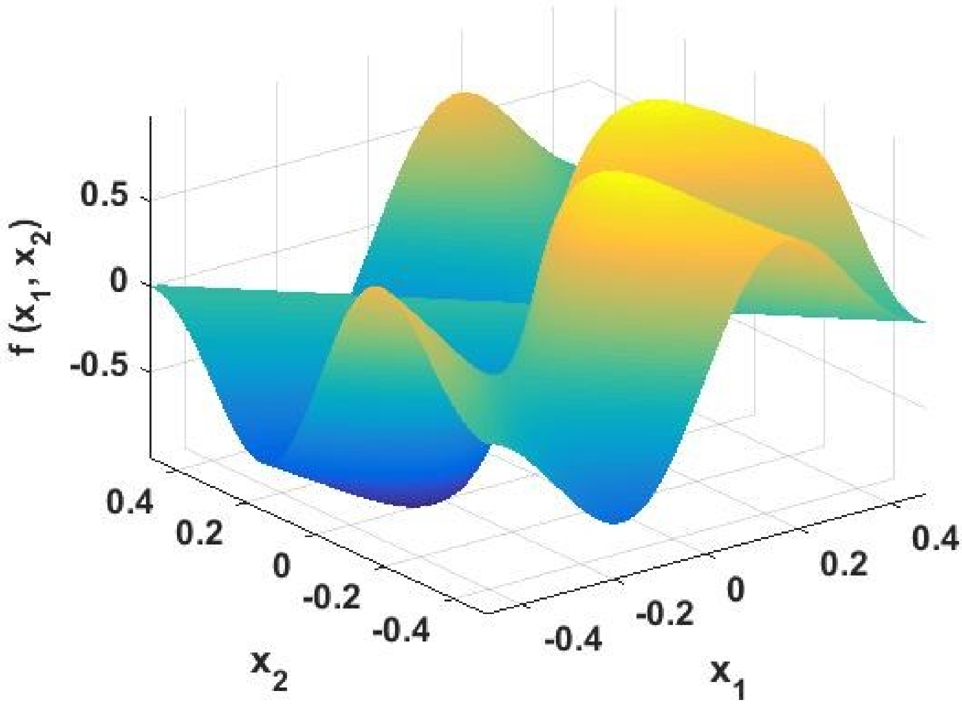
In this example, a double-input and single-output static complex multi-modal butterfly function is chosen to be a target system for the proposed incremental learning strategy. The function is taken from [321] and represented as

$$f(x_1, x_2) = (x_1^2 - x_2^2).\frac{sin(x_1^2 + x_2^2)}{(x_1^2 + x_2^2)}, \qquad -0.5 \leq (x_1, x_2) \leq 0.5 \qquad 6\text{-}7$$

For which 1000 data points are generated. The same modelling procedures were adopted. A 15-rule-based model is developed. Figs. 6.15 and 6.16 show the results of the obtained model for the butterfly function approximation.



(a) Actual system output.



(b) Output of the IT2-RBF-NF system.

Figure 6.15. Multi-modal butterfly system approximation.

Perpetual Learning Performance:

In this simulation, the same incremental learning process was adopted by generating a new data set in the range of $-0.65 \leq (x_1, x_2) \leq 0.65$ to make the function more complex. Tables 6-6 and 6-7 show the performance of the original model (15 rule-based system) on the new data and the updated model (24 rule-based system) respectively.

Table 6.6 Performance of the original model on the new data for non-linear function approximation in Example 3.

IT2-RBF-NF system	**Initial Model (Core Model)**
Number of rules	15
Number of parameters	60
RMSE for partially novel data	0.0528
RMSE for totally novel data	0.3759

Table 6.7 Performance of the original model and updated model for non-linear function approximation in Example 3.

IT2-RBF-NF system	Initial Model (Core Model)	Incrementally updated model
Number of rules	15	24
Number of parameters	60	96
Training RMSE	0.0496	0.0464
Testing RMSE	0.0507	0.0474

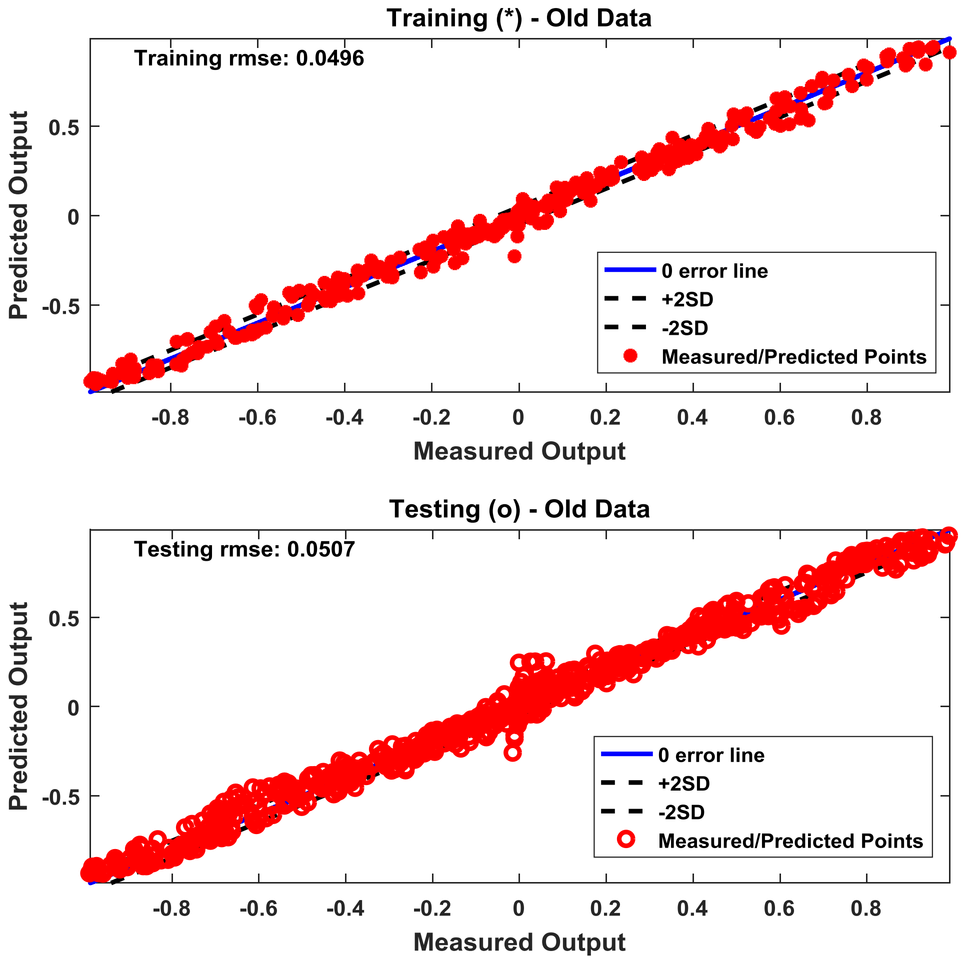


Figure 6.16. Measured versus predicted output of the IT2-RBF-NF model for training and testing.

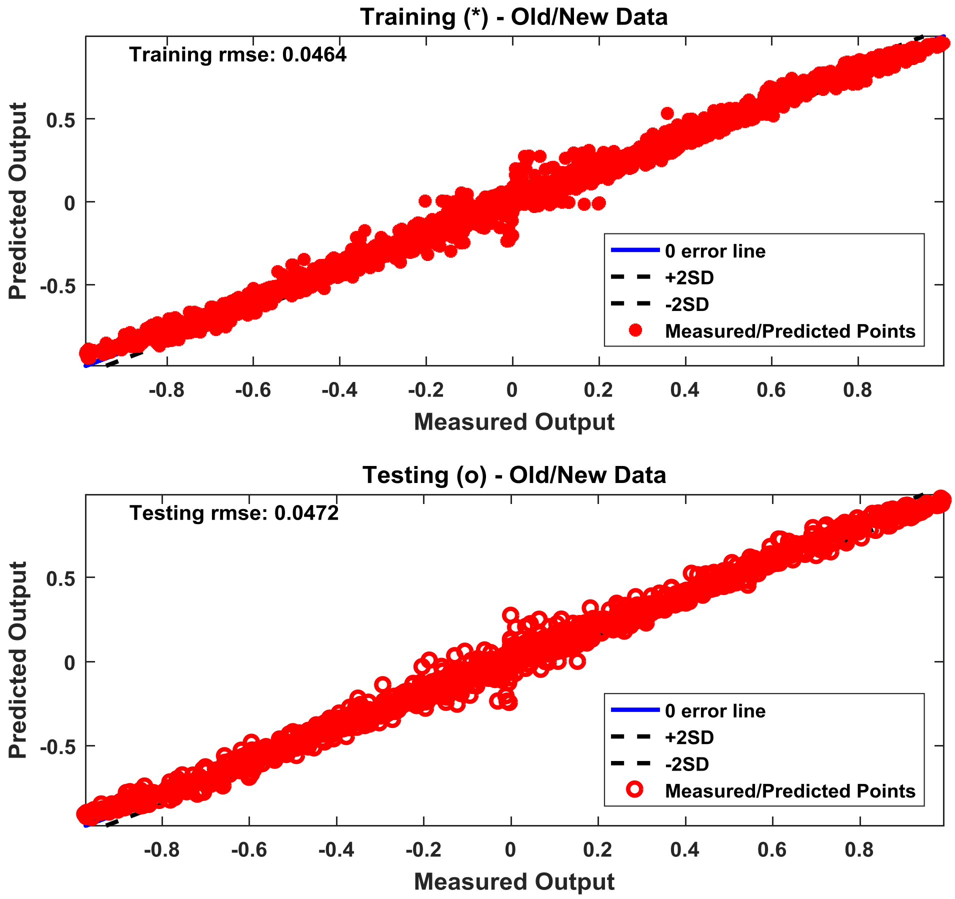


Figure 6.17. Performance of the updated model on the whole data set.

Clearly from the simulation results, the incremental learning structure has a dynamic behaviour by updating itself to accommodate the unseen input data set. The obtained updated model is able to maintain a good performance (as compared to the original model) without ignoring any previously gain knowledge. The updated model has the ability to capture the complex dynamic of the system well as depicted in Fig. 6.17.

In the next simulation example, on one hand, the sustainability of the perpetual framework in case where more frequent/periodic incremental model updates are required will be demonstrated. And on the other hand, two simulation studies are carried out: in the first case the perpetual learning framework is geared towards accuracy by imposing weak/relaxed constraints while in the second case the perpetual learning framework is geared towards interpretability by imposing strong constraints.

6.3.4. EXAMPLE 4: ITERATIVE INCREMENTAL LEARNING FOR MULTI-MODAL NON-LINEAR FUNCTION IDENTIFICATION

In this example, it will be shown that how the proposed perpetual learning framework can produce a sustainable incremental updates - iterative incremental update - that can handle the changing in the input conditions. A hypothetical case of many steps (2% incremental step in this study) for the non-linear complex function in example 2 has been created. The incremental step was kept at 2% in order not to generate a highly novel data that the model will find it difficult to capture the core dynamics of the post-updated system. Bigger incremental step needs more data sets to cover the input space of the system and thus generating more rules which results in complex structure. An incremental update for 10 iterative incremental steps was carried out such that after each step the model expands its structure by adding rules to accommodate the novel data and at the same time improves its structure by pruning the inconsequential rules over time.

In the following case relating to iterative incremental updates where weak/relaxed constraints are imposed on the iterative rule pruning algorithm (0.6 in this case) and constrained parametric optimisation (0.6 in this case). Consequently, a large number of inconsequential rules is generated after the final incremental step. In other words, the perpetual learning framework is geared towards accuracy. The performance of the iterative incremental updates is summarised in Table 6-8. As shown, the number of rules grows but not by a substantial increment due the pruning mechanism applied to the incrementally updated architecture at each step. The incremental learning architecture is able to maintain a good performance without ignoring any previously gained knowledge. From Table 6-8, it is concluded that the proposed incremental framework produces good accuracy after 10 incremental steps. Fig. 6.18 shows the simulation results of the incrementally updated model after 10 steps. Although incrementally updated model after 10 steps with nearly three times rule base size (38 rule-based system) of that of the core model (13 rule-based system), the updated structure is able to learn more accurately over time. In summary, the results show that the incremental learning framework achieves good balance between model accuracy (by inserting new rules) and complexity (by deleting insufficient rules), while yielding sustainable and reliable incremental update architecture that can be adapted incrementally in a lifelong learning mode (continuous rule growing and pruning).

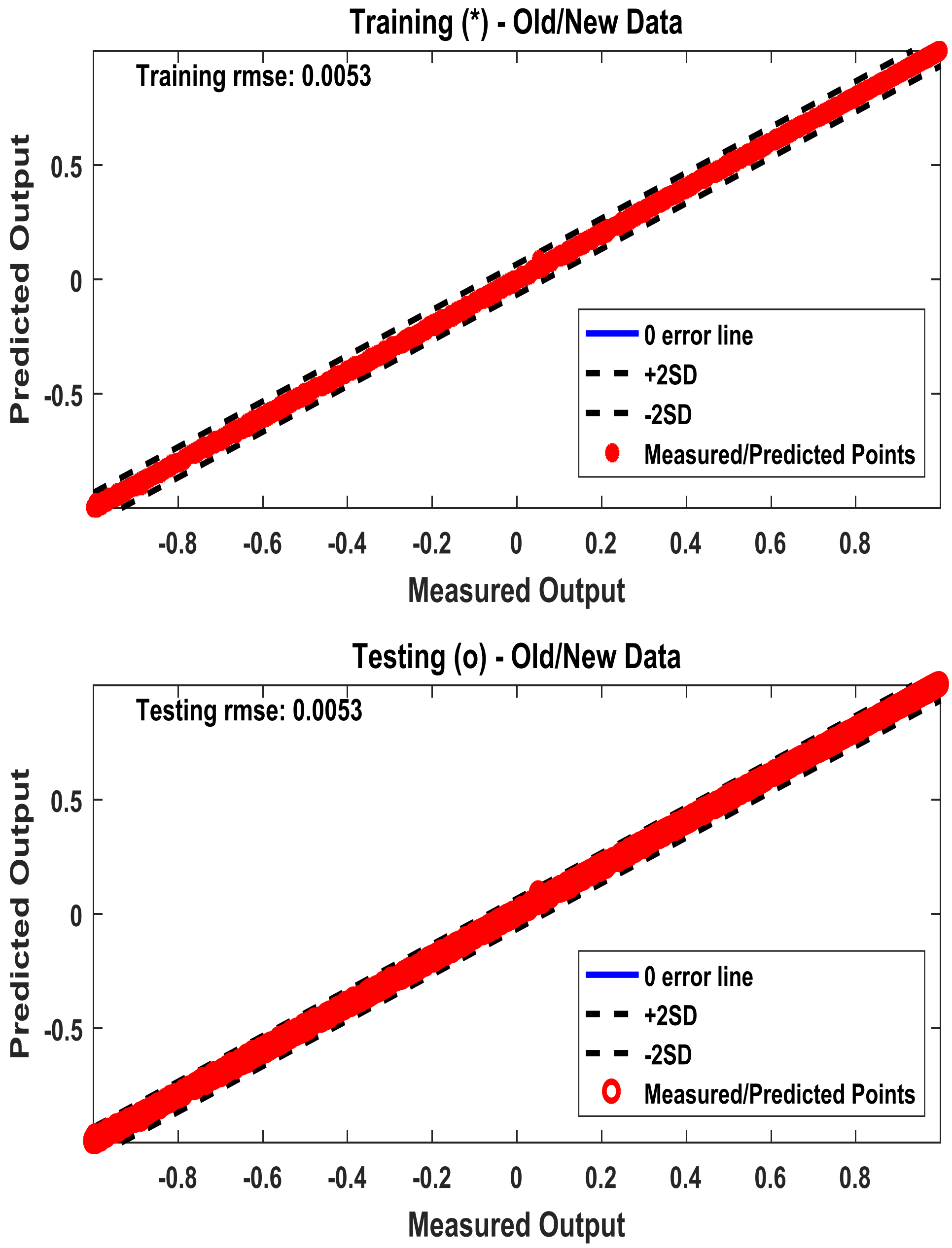


Figure 6.18. Performance of the updated model after 10 incremental steps (with weak/relaxed constraints): Regression line for training and testing.

Table 6.8 Performance of the updated model during 10 iterative incremental updates (with weak/relaxed constraints) for multi-modal function approximation in Example 2.

Step No.	**Number of Rules**		**Training RMSE**		**Testing RMSE**		**No. of added rules**
	Before incremental update	*After incremental update*	*Before incremental update*	*After incremental update*	*Before incremental update*	*After incremental update*	
1	13 ***(Core Model)***	15	0.0090 ***(Core Model)***	0.0099	0.0079 ***(Core Model)***	0.0093	3
2	15	17	0.0099	0.0063	0.0093	0.0062	2
3	17	19	0.0063	0.0058	0.0062	0.0058	2
4	19	21	0.0058	0.0083	0.0058	0.0081	2
5	21	24	0.0083	0.0068	0.0081	0.0061	3
6	24	27	0.0068	0.0108	0.0061	0.0107	3
7	27	30	0.0108	0.0061	0.0107	0.0065	3
8	30	32	0.0061	0.0055	0.0065	0.0054	2
9	32	35	0.0055	0.0072	0.0054	0.0057	3
10	35	38 ***(Final updated Model)***	0.0072	0.0053 ***(Final updated Model)***	0.0057	0.0053 ***(Final updated Model)***	3

In the following case, strong constraints are imposed on the iterative rule-pruning algorithm (0.9 in this case) and constrained parametric optimisation (0.9 in this case). On one hand, imposing strong constraints deteriorates the performance of the perpetual learning over time (i.e., after a number of incremental steps), and on the other hand it preserves a low number of rules over time with a good level of interpretability. Fig. 6.19 shows the simulation results of the incrementally updated model after 10 steps. Table 6-9 summarises the performance index based on RMSE for the 10 incremental steps. It is evident that as a result of the strong constraints, the rule-base size of the updated model (18 rule-based system) is not big as compared to the rule-base of the original model (13 rule-based system), but the final performance of the updated structure is not as good as the final updated model in case where weak/relaxed constraints are imposed.

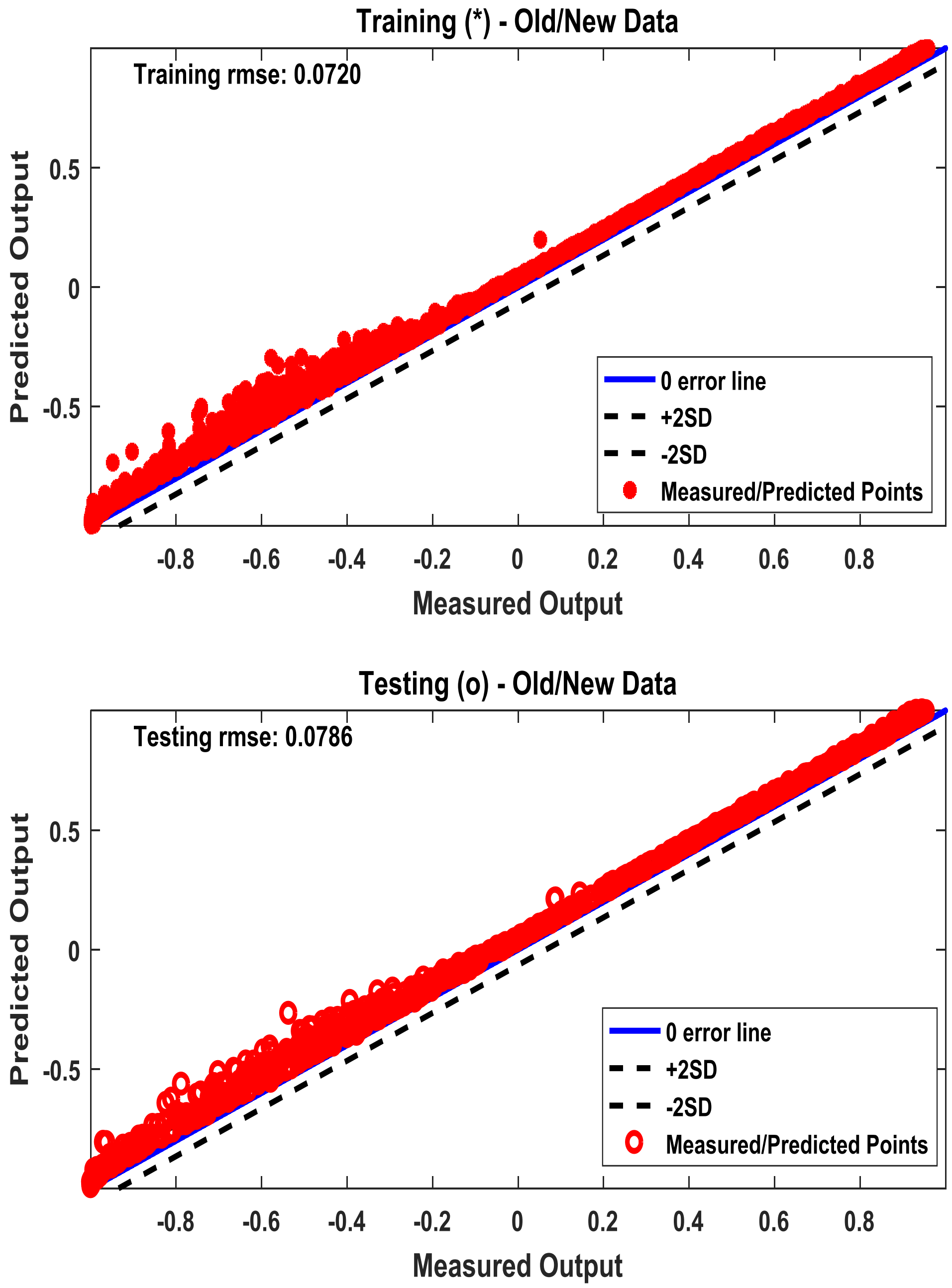


Figure 6.19. Performance of the updated model after 10 incremental steps (with strong constraints): Regression line for training and testing.

Table 6.9 Performance of the updated model during 10 iterative incremental updates (with strong constraints) for multi-modal function approximation in Example 2.

Step No.	**Number of Rules**		**Training RMSE**		**Testing RMSE**		**No. of added rules**
	Before incremental update	*After incremental update*	*Before incremental update*	*After incremental update*	*Before incremental update*	*After incremental update*	
1	13 ***(Core Model)***	14	0.0090 ***(Core Model)***	0.0114	0.0079 ***(Core Model)***	0.0125	1
2	14	14	0.0114	0.0147	0.0125	0.0180	0
3	14	14	0.0147	0.0201	0.0180	0.0234	0
4	14	15	0.0201	0.0289	0.0234	0.0270	1
5	15	16	0.0289	0.0356	0.0270	0.0319	1
6	16	16	0.0356	0.0468	0.0319	0.0491	0
7	16	17	0.0468	0.0542	0.0491	0.0580	1
8	17	18	0.0542	0.0604	0.0530	0.0510	1
9	18	18	0.0604	0.0693	0.0510	0.6810	0
10	18	18 ***(Final updated Model)***	0.06913	0.0720 ***(Final updated Model)***	0.6810	0.0786 ***(Final updated Model)***	3

In general, from the above simulation examples when dealing with rule-based system, overall system accuracy and interpretability are two conflicting requirements, as improving global system accuracy of FL models generally degrades overall system interpretability of FL models, and vice versa. Hence one challenging problem is how to select the thresholds ($Th_1 - Th_4$) in the iterative rule pruning algorithm to remove the inconsequential rules as a result of the incremental update process. A careful selection of these thresholds determines the number of redundant rules to be removed. Thus, improving the overall interpretability of the updated model. Small thresholds result in a large number of redundant rules to be removed, and then more interpretable fuzzy model but less global model accuracy. In addition, the constraints on the parametric optimisation after the incremental updating algorithm play an important role on the overall system accuracy. Imposing strong constraints leads to less accurate modelling performance but more interpretable fuzzy rule. While imposing weak constraints leads to good global system accuracy but this requires abolishing all the knowledge maintained from the original model.

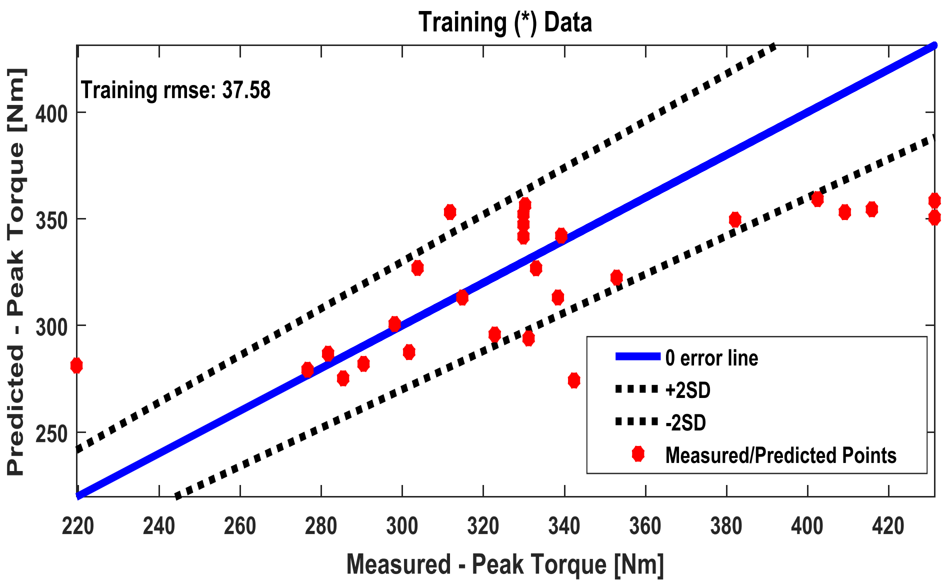
6.3.5. APPLICATION TO STEEL FRICTION STIR WELDING

In this section, the effectiveness of the proposed perpetual learning framework is verified over a real industrial case study. Simulation studies are carried out in this part in order to apply the proposed perpetual framework to the modelling of steel friction stir welding data. A number of experimental trials for welding of a shipbuilding-grade steel (DH36), were carried out at TWI Ltd, Technology Centre (Yorkshire), United Kingdom. The aim was to determine the process operating window by changing the levels of the welding parameters (i.e., tool rotational speed and welding speed). The proposed perpetual learning framework is applied to the prediction of spindle peak torque. The raw data consists of in total 55 data samples, the data set is shown in the Appendix. For the purpose of modelling, a care has taken to split the database so that the new dataset covers mostly an input space that is not covered by the initial 'old' dataset (i.e., new input process parameters that are not covered by the initial 'old' dataset). The process dataset has been split as follows:

a) Initial dataset (36 data points) which has been divided into 27 (75%) data points to train the model and 9 (25%) data points to test the prediction performance of the original model;
b) New dataset (19 data points), which has been used to test the incremental modelling framework.

6.3.5.1. INITIAL MODEL PERFORMANCE

The initial dataset (including training and testing data sets) is used to train and then test the performance of the initial model respectively. After a number of simulations, it was established that the optimal number of granules is five. After performing the iterative data granulation to the training dataset, the initial training dataset is granulated into five information granules. The extracted granulated data are then mapped into linguistic fuzzy rules to construct the initial rule-base structure. Finally, after the parametric optimisation stage, a 5-rule IT2-FLS is created. The prediction performance obtained from the 5-rule IT2-FLS is: MAE = 7.51% and 8.04% for training and testing respectively. Fig. 6.20 illustrates the regression line between measured and predicted peak torque within the 2 × standard deviation band for the modelling performance on training and testing respectively.



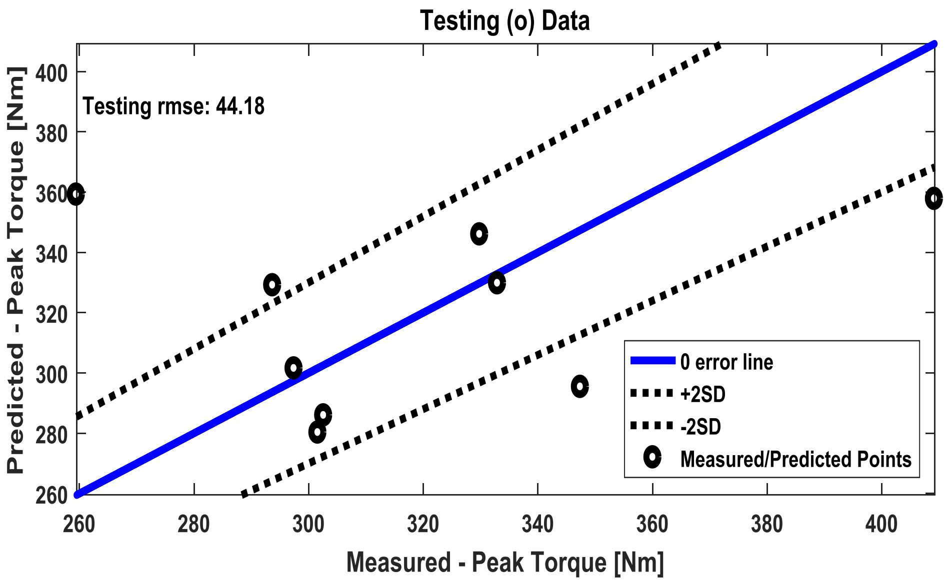


Figure 6.20. Performance of the initial model.

6.3.5.2. INCREMENTAL LEARNING PERFORMANCE

The new dataset (19 data points) is split into two subsets (training and testing). The new training set is then presented to the incremental learning framework. The performance of the original model on the new dataset is shown in Table 6-10. The incremental learning framework classifies the training data into novel and partially novel data set. The latter is used to fine-tune the initial model via the adaptive error-propagation algorithm and the former is used to generate new fuzzy rules that represent the new dataset. 4 new rules are obtained from the granular computing algorithm (i.e. optimal number of rules to cover the new dataset). The new fuzzy rules are trained via the same algorithmic procedure as the initial IT2-RBF-NF model. Subsequently, the newly generated fuzzy rules are combined with the rest of the fuzzy rules in the modelling structure.

Following the rule-pruning process (with $Th_1 - Th_3 = 0.9$ and $Th_4 = 0.01$), two rules are removed based on a pre-defined similarity thresholds; the resulting rule-base is further optimised (fine-tuned) as described in Section 6.2.3.

The resulting model is tested for its performance on the initial dataset as well as new dataset (training and testing). The results are shown in Fig. 6.20. The prediction performance obtained from the 7-rule IT2-FLS is: MAE = 8.61% and 9.84% for training and testing respectively. As illustrated in the model fit plot for the training data set, the incremental learning framework is able to preserve a good prediction performance, in fact it is able to correctly construct input-output mappings similar to the original model. Similar behaviour is observed for the testing data set for old and new datasets. The model is able to predict correctly the new – unseen – input patterns.

Table 6.10 Performance of the original model on the new dataset.

	Number of rules	**Number of parameters**	**RMSE for partially novel data**	**RMSE for totally novel data**	**MAE% for partially novel data**	**MAE% for totally novel data**
IT2-RBF-NF system	5	20	38.13	56.07	6.61	20.31

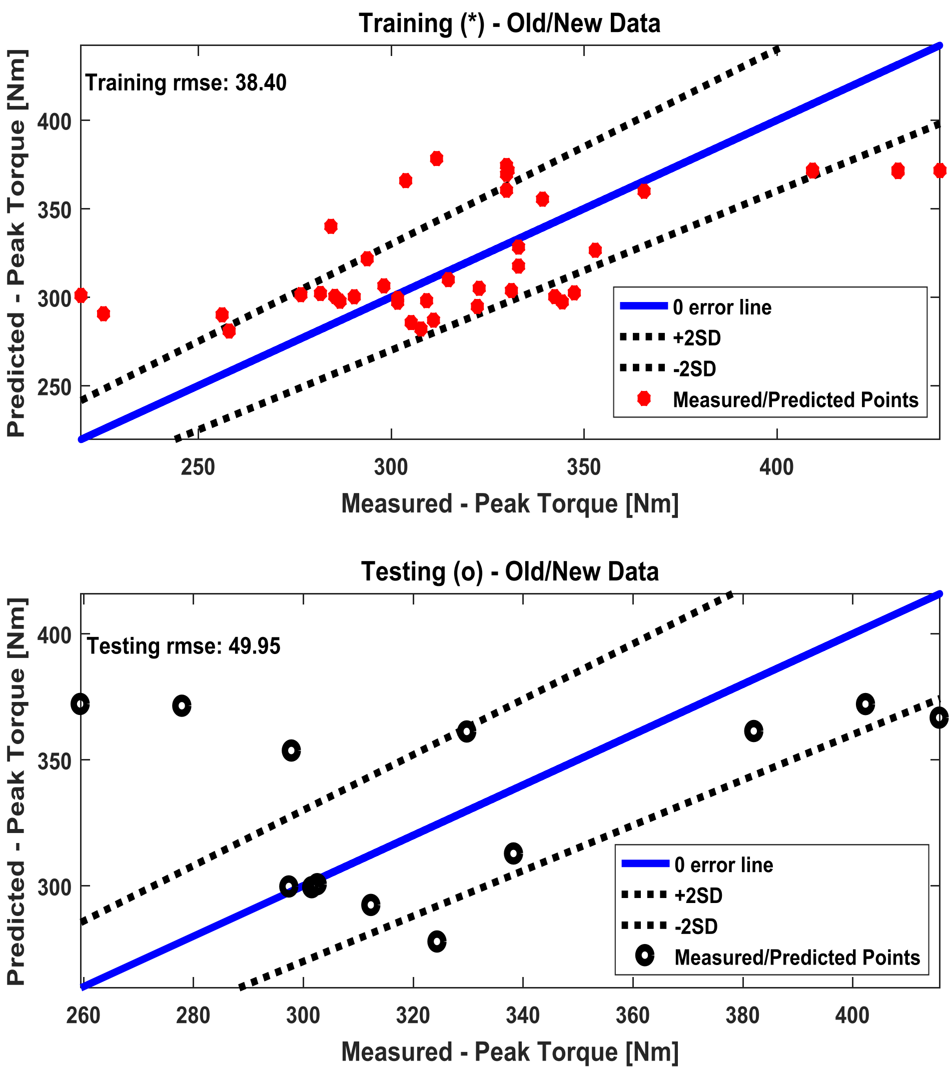


Figure 6.21. Performance of the incrementally updated model on the old/new data.

From the simulation results, the proposed perpetual learning framework has the ability to achieve a good balance between accuracy and interpretability without ignoring any previously gained knowledge from the original model. An iterative rule pruning mechanism is used as the main feature that removes the redundant fuzzy rules after each incremental step, which allows the model to be used in a lifelong learning mode. Moreover, no pervious perpetual learning framework that is based on granular computing interval type-2 neural fuzzy systems for offline (batch) learning have been reported and this was the first attempt to develop such perpetual (incremental) learning structure.

6.4. SUMMARY

The work in this chapter is devoted to the development of a new perpetual learning framework. On the one hand, a perpetual learning framework was presented to provide a reliable incremental model updating routine that resulted in an open - dynamically expandable - structure without ignoring any previously obtained knowledge from the original model. On the other hand, the proposed framework has satisfied the requirements needed for continuous learning as it is able to handle the short and long-term change in the input conditions in a lifelong learning mode by incrementally updating its structure to accommodate the change in the process input data distribution. The incremental updating algorithm also incorporated an iterative rule pruning strategy to remove the inconsequential rules as the result of the incremental update process with compromising the prediction accuracy.

The proposed framework was tested against four case studies, which include three well-known non-linear benchmark function and one real industrial case study. In each case study, the simulation results of the proposed methodology was verified by increasing the complexity of the function. The results showed that the ability of the proposed methodology to update/create new rule in order to accommodate the unseen input data set and maintain good overall prediction performance on the system (initial dataset, and initial/new dataset combined) and at the same time without ignoring any previously gained knowledge from the original model. The sustainability (i.e., iterative model updates) of the proposed incremental methodology was also tested against a multi-modal complex function where more frequent/periodic model updates are required while maintain overall system accuracy. It was demonstrated that the proposed framework was able to periodically/incrementally model the new data when these are available. The overall system prediction performance on the initial dataset is preserved and the performance on the new dataset is comparable to the overall original performance.

The results obtained from the chapter led to the writing of two articles: an article that was presented at the *17th IFAC Symposium on Control, Optimisation and Automation in Mining, Mineral and Metal Processing in Vienna, Austria* and another article that is ready to be submitted to a soft-computing jouranl.

In the next chapter, conclusions of the research work presented in thesis will be drawn and the future work related to this research work will be discussed.

CHAPTER 7 - CONCLUSIONS AND FUTURE WORK

This chapter summarises several major contributions in this thesis and provides recommendation for future works.

7.1. CONCLUSIONS

The main objective of this research work is to develop parsimonious, transparent interpretable and computationally efficient soft-computing techniques and human-centric systems with applicability in various engineering and scientific domains, where in this particular work, the developed techniques are applied to the modelling of steel friction stir welding data obtained from TWI Ltd., Technology Centre (Yorkshire), United Kingdom. The developed computational frameworks aim to deal with various challenges such as uncertainty, imprecision, inconsistency, incompleteness and spasity arise in complex systems, including friction stir welding. Therefore, this thesis makes important contributions towards developing efficient human-centric systems.

Chapter 4 developed a systematic data-driven neural-fuzzy (NF) modelling framework based on the iterative human-like information capture of granular computing (GrC) and RBF-NF system. The iterative human-like computational framework of granular computing (GrC) in the form proposed in [27] represents an explanatory and effective data analysis tool and a useful data clustering technique. It has also demonstrated its efficiency as a method for constructing the initial structure of the RBF-NF system. Even through, it has proven its efficiency, there was no an uncertainty measure to quantify the degree of conflict among information granules in case where two or more information possess a similar compatibility measure. As was pointed out in this chapter, as a result of conflict among information granules during the iterative granulation process, the granulated data do not represent the accurate distribution of the input space of the process under study (i.e., producing low quality information granules). Accordingly, the two levels of interpretability (low-level and high-level of interpretability) during the development of RBF-NF system were fully described. The low-level interpretability can be obtained with regard to semantic criteria on fuzzy sets such as consistency and distinguishability during the formation of fuzzy sets (granulated data). While the high-level interpretability can be achieved on the final fuzzy rules after the parametric optimisation process with regard to criteria on fuzzy rules such as the consistency, completeness, and coverage.

A particular focus were put on the low-level interpretability during the information granulation process for estimating the initial RBF-NF system parameters (initial FL rule-base) and a new uncertainty measure calculated via Shannon entropy theory was proposed in order to take into consideration the uncertainty as a consequence of conflict among information granules during the iterative information granulation process. This uncertainty measure was used to evaluate/quantify the degree of conflict among information granules during the granulation process and to guide the merging operation into forming better quality information granules in terms of their distinguishability. The initial structure of the RBF-NF system was then optimised via the use of an adaptive back-error propagation (BEP) approach to improve its performance. Finally, the effectiveness of the proposed framework was tested against a popular benchmark data set as well as applied for the first time to the prediction of peak torque data in steel friction stir welding. It was concluded that the degree of distinguishability among the information granules that are formed from the iterative information granulation has a significant influence on the interpretability of the initial FL rule-base at low-level and at high-level of interpretability as well as transparency of the final FL rule-base.

It was also concluded that the integration of the human-like information capture in granular computing [67, 135, 322] and fuzzy logic theory [8, 61] in modelling systems engineering will add more transparency to the overall system. Since the iterative human-like information of granular commuting uses the principle of information granulation in order to mimic the human cognition in terms of grouping/arranging similar objects together. This intriguing human-centricity feature can be used to focus and facilitate the analysis and interpretability of complex systems in a transparent way on aspects of interest to the user.

In the first part of Chapter 5, a new data-driven modelling approach relied on a six-layered IT2-RBF structure that is mathematically equivalent to an IT2-FLS in its design was developed. In the same way to the design of an IT2-FLS, the proposed IT2-RBF-NFS can be seen as an IT2-FLS under some certain conditions having a singleton fuzzifier whose T-norm is the product operator, the antecedent parts use the Gaussian IT2-MFs, and the consequent part of each fuzzy rule of the Mamdani type. The IT2-RBF-NFS can be described via simple linguistic interpretable rules extracted from raw data in order to describe dynamic behaviour of the process. The initial parameters of the IT2-RBF-NF (initial rule-base structure) and the FOU were estimated directly via the iterative data granulation approach used in Chapter 4. Despite the important advances have been made and a number of algorithms were developed to simplify the computational cost in the type-reduction stage, in the research work, the Karnik-Mendel type reduction was used as the starting point to reduce the T2-FSs from the inference engine to T1-FSs and then the defuzzified crisp output was calculated via the average of the interval type-reduced sets. The initial parameters of an IT2-RBF based interval type-2 neural-fuzzy model were optimised via the adaptive back-error propagation (BEP) algorithm.

In the second part of Chapter 5, the proposed model-based IT2-RBF-NF system was used to develop a new generalised monitoring system for the first time in FSW in order to forecast in real-time (during welding) quantitative markers of weld quality. On the one hand, part quality thresholds were extracted from the frequency spectra of the feedback forces (namely axial (Fz) and traverse (Fx) forces). On the other hand, a dynamic model, that instead of predicting directly the weld quality, it predicts a 'moving threshold' that can be used by the operator as an indicator of weld quality was developed. By using such model there is no need to re-tune or re-calibrate the model every time the welding conditions change. This intriguing generalisation feature remedied the drawbacks and limitations of the static models and also made the monitoring systems feasible for real-time use. The proposed monitoring framework also takes advantage of the proposed model-based approach to provide continuous linguistic-based feedback to the operator(s) – rule-based human-centric system – on the performance of the process. Finally, the effectiveness of the proposed model-based approach to better handle uncertainties and produce reasonable predictive performance was tested against multiple linear regression (MLR) and multilayer perceptron neural network (MLP-NN) models as well as its type-1 radial basis function neural fuzzy (T1-

RBF-NF) model counterpart. Some important conclusions drawn from this research work include:

- The ability of the proposed IT2-RBF-NFS to handle effectively the linguistic uncertainty in the rule-base.
- The proposed IT2-RBF-NFS has high tolerance to the input noise and the ability to produce an accurate and robust performance.
- A good prediction performance compared to MRL and MLP models as well as its type-1 RBF-NFS counterpart.
- The ability to provide a generalised reliable and online monitoring system with modest computational cost and real-time feasibility.
- The proposed data-driven framework could be extended to various industrial applications for modelling, process monitoring and control purposes.

Lastly, in Chapter 6, a new perpetual (incremental) learning framework that is based on granular computing presented in Chapter 4 and the IT2-RBF-NF system presented in Chapter 5 was developed. An adaptive back-error propagation (BEP) algorithm was used to optimise the initial IT2-RBF-NFS. The main motivations for developing such mathematical framework was to include advanced system's features such as the ability to accommodate new data when these made available to the system without significantly disturbing the existing knowledge from the initial model. Therefore, a perpetual learning framework was developed to include the ability of the IT2-RBF-NF system to continuously learn from new process data – in an incremental learning fashion – by creating new IT2 fuzzy rules and modifying (adapting) the existing model. An iterative rule pruning strategy was also used as the main feature to prune/remove the inconsequential/redundant fuzzy rules as a result of the incremental model update process, which allows the model to be used in a lifelong learning mode.

The performance of the proposed structure was demonstrated using a number of simple and complex non-linear benchmark functions as well as real industrial case. Simulation results showed that the performance of the original model structure is maintained and it is comparable to the overall model performance after the final incremental model updating routine. The efficiency and effectiveness of the proposed structure was also evaluated in case where more frequent/periodic model updates are required with good balance between interpretability and accuracy.

In conclusion, this thesis has achieved its aims and objectives.

7.2. SPECIFIC FUTURE WORK FOR THIS THESIS

Despite the proposed soft-computing techniques and human-centric systems prove to be promising, more research is required in some areas for further improvements. Therefore, this section presents several suggestions for future works from both modelling systems engineering perspective and process perspective:

- Further investigation and experimentation into the two levels of interpretability in neural fuzzy systems (low-level of interpretability and high-level of interpretability) is strongly recommended. A number of possible future studies are needed to assess and include different types of uncertainty that may present during the iterative human-like information granulation for instance, as a result of discord, conflict, ambiguity, and congruency. More research work should be carried out to develop the theoretical foundations of these types of uncertainty. These theoretical aspects are important, which could help us to further improve the quality of the granulated data. Thus improving the interpretability on both levels (on the fuzzy sets and the final fuzzy rules). So far, however, too little attention has been paid in this area; for instance, in [29, 31] Solis and Panoustos studied two types of uncertainty (namely fuzziness and ambiguity) during the information granulation using neutrosophic sets theory.
- The interpretability studies in the development of RBF-NF system recommended that studying the effects of uncertainty during the information granulation could be extended to IT2-RBF-NFS. This includes investigating the interpretability at low-level and high level of interpretability in the IT2-RF-NFS. From real-time applicability point of view, the computational overhead associated with the computation of type-reduced sets using the iterative KM algorithm can be further reduced by using approximate type-reduction

methods; for instance the Wu-Mendel uncertainty bounds approximation method [112] is one of a wide range of type-reducers appeared in literature.

- The proposed IT2-RBF-NFS takes advantages of principles of human-like information granulation in granular computing to extract meaningful knowledge out of uncertain data and the additional degree of freedom from the T2-FS FOU to handle the linguistic uncertainties associated with meaning of words and linguistic propositions contained in the rule- base. Such computational framework offered an efficient data-driven modelling approach to deal with epistemic uncertainty (i.e. knowledge uncertainty). Consequently, further research could be to combine the IT2-RBF-NFS with the existing theoretical frameworks such as Mote Carlo method and Bayesian method that have been developed to deal with aleatory uncertainty (i.e. random uncertainty) in order to take into account both epistemic and aleatory uncertainties in the process of making decisions in complex systems.
- The development of the perpetual learning framework may open up a new field of research to develop fully autonomous or semi-autonomous human-centric systems for model updating routine without any involvement from the system's designer. It appears from the research work conducted in this thesis, no attempt was made to develop model-based autonomous or semi-autonomous human-centric systems.
- The proposed perpetual learning framework assumes that the new input data that are presented to the model are valid to be used in the incremental learning process. The perpetual learning framework also does not take into account a case where the new input data that are available to the incremental learning framework include 'noise' data points. Therefore, further research could be conducted on checking the validity of new input data before they are fed to the incremental process. Further work should be focused on the performance of the framework in the presence of 'noise' data points.
- Specific to the friction stir welding process, future research aspects could include: a) collect and analyse high resolution data for various internal process variables e.g., vibration of the machine and temperature of the tool and then relate them to the post-weld properties (including mechanical properties,

microstructure features and weld quality); b) develop real-time model-based autonomous or semi-autonomous systems for quality control to – non-destructively – evaluate the resulting weld quality; and c) develop real-time decision support systems (DSSs) for tool design.

7.3. FUTURE RESEARCH DIRECTIONS

Advanced manufacturing systems are awash with a mammoth amount of data that are being generated from daily process routine and stored in databases. Some of the data are carrying meaningful information and others may simply be noises. The data are often characterised with the presence of several issues such as high dimensionality, nonlinearity, collinearity, uncertainty, missing measurements, outlier, etc. Hence, these issues should be taking into consideration while carrying out the information extraction process. The manufacturing industry is looking for techniques to deliver innovations that address these issues. Since the extracted information could be employed in the production lines and/or to develop data analytics for prediction, process monitoring, soft-sensing, fault diagnosis, process optimisation, and control.

Advanced manufacturing systems are information intensive and process operators are often play an important role in making decisions during the manufacturing process chain. On the one hand, human operators are required to perform tasks in dangerous environments and/or to perform cognitive tasks that are difficult to perform computationally. On the other hand, ICT services, offline/real-time data usage, data generation, data architecture, data-assist systems, and data-based optimisation are very efficient at capturing, processing and then quantifying information from process data. The manufacturing industry roadmap should focus in the areas such as human-machine interaction, linguistics-based techniques to data mining and information extraction, human-in-the-loop robotics, and real-time decision support systems designed to collaborate/interact with humans and to aid making decisions at multiple levels in the manufacturing process chain.

The stricter demands for manufacturing process certification and quality assurance in many applications such as automotive, aerospace, rail industry, food production, etc and the worldwide overcapacities raised the need for efficient process monitoring tools in line to the process. Additionally, in order to ensure a constant and reproducible product quality, the development of autonomous or semi-autonomous systems for process monitoring and quality control will be indispensable for maintaining the complex manufacturing process chain with high productivity. Therefore, data-driven computational intelligence concepts and human-centric systems within the automation systems will be the basis for a reliable, exceptional and reproducible manufacturing systems.

REFERENCES

[1] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth *et al.*, *Advances in knowledge discovery and data mining*: AAAI press Menlo Park, vol.21, 1996.

[2] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI magazine,* vol. 17, no. 3, pp. 37, 1996.

[3] G. J. Klir, and M. J. Wierman, *Uncertainty-based information: elements of generalized information theory*: Springer Science & Business Media, 1999.

[4] I. Graham, and P. L. Jones, *Expert systems: knowledge, uncertainty, and decision*: Chapman & Hall, Ltd., 1988.

[5] G. Klir, and B. Yuan, *Fuzzy sets and fuzzy logic*: Prentice Hall New Jersey, 1995.

[6] T. J. Ross, *Fuzzy logic with engineering applications*: John Wiley & Sons, 2009.

[7] M. Black, "Vagueness. An exercise in logical analysis," *Philosophy of science,* vol. 4, no. 4, pp. 427-455, 1937.

[8] L. A. Zadeh, "Fuzzy sets," *Information and control,* vol. 8, no. 3, pp. 338-353, 1965.

[9] G. Shafer, *A mathematical theory of evidence*: Princeton university press Princeton, 1976.

[10] C. Negoita, L. Zadeh, and H. Zimmermann, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy sets and systems,* vol. 1, no. 3-28, pp. 61-72, 1978.

[11] D. Dubois, and H. Prade, "Possibility theory: qualitative and quantitative aspects," *Quantified representation of uncertainty and imprecision*, pp. 169-226: Springer, 1998.

[12] Z. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences,* vol. 11, no. 5, pp. 341-356, 1982.

[13] G. J. Klir, "Developments in uncertainty-based information," *Advances in computers,* vol. 36, pp. 255-332, 1993.

[14] D. Dubois, "Possibility theory and statistical reasoning," *Computational statistics & data analysis,* vol. 51, no. 1, pp. 47-69, 2006.

[15] L. A. Zadeh, "Key roles of information granulation and fuzzy logic in human reasoning, Concept formulation and computing with words." *Proceedings of the Fifth IEEE Intenational Conference on Fuzzy Systems*, vol. 1, pp. 1-1, 1996.

[16] K. Ishii, and M. Sugeno, “A model of human evaluation process using fuzzy measure,” *International Journal of Man-Machine Studies,* vol. 22, no. 1, pp. 19-38, 1985.

[17] Y. Li, J. Chen, and L. Feng, “Dealing with uncertainty: a survey of theories and practices,” *IEEE Transactions on Knowledge and Data Engineering,* vol. 25, no. 11, pp. 2463-2482, 2013.

[18] A. Der Kiureghian, and O. Ditlevsen, “Aleatory or epistemic? Does it matter?,” *Structural Safety,* vol. 31, no. 2, pp. 105-112, 2009.

[19] M. J. Gacto, R. Alcalá, and F. Herrera, “Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures,” *Information Sciences,* vol. 181, no. 20, pp. 4340-4360, 2011.

[20] J.-S. R. Jang, and C.-T. Sun, “Functional equivalence between radial basis function networks and fuzzy inference systems,” *Neural Networks, IEEE Transactions on,* vol. 4, no. 1, pp. 156-159, 1993.

[21] H. Anderson, A. Lotfi, L. Westphal *et al.*, “Comments on" Functional equivalence between radial basis function networks and fuzzy inference systems"[with reply],” *Neural Networks, IEEE Transactions on,* vol. 9, no. 6, pp. 1529-1532, 1998.

[22] T. M. Cover, “Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition,” *Electronic Computers, IEEE Transactions on*, no. 3, pp. 326-334, 1965.

[23] P. Liu, and H. Li, “Efficient learning algorithms for three-layer regular feedforward fuzzy neural networks,” *IEEE Transactions on Neural Networks,* vol. 15, no. 3, pp. 545-558, 2004.

[24] A. Riid, and E. Rüstern, “Identification of transparent, compact, accurate and reliable linguistic fuzzy models,” *Information Sciences,* vol. 181, no. 20, pp. 4378-4393, 2011.

[25] M.-Y. Chen, and D. A. Linkens, "Rule-base self-generation and simplification for data-driven fuzzy models.", *The 10th IEEE International Conference on Fuzzy Systems*, vol. 1, pp. 424-427, 2001.

[26] M.-Y. Chen, and D. A. Linkens, “A systematic neuro-fuzzy modeling framework with application to material property prediction,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 31, no. 5, pp. 781-790, 2001.

[27] G. Panoutsos, and M. Mahfouf, “A neural-fuzzy modelling framework based on granular computing: Concepts and applications,” *Fuzzy Sets and Systems,* vol. 161, no. 21, pp. 2808-2830, 2010.

[28] W. Yu, M. Li, J. Luo *et al.*, "Prediction of the mechanical properties of the post-forged Ti–6Al–4V alloy using fuzzy neural network," *Materials & Design,* vol. 31, no. 7, pp. 3282-3288, 2010.

[29] A. R. Solis, and G. Panoutsos, "Granular computing neural-fuzzy modelling: A neutrosophic approach," *Applied Soft Computing,* vol. 13, no. 9, pp. 4010-4021, 2013.

[30] A. Bargiela, and W. Pedrycz, *Granular computing: an introduction*: Springer, 2003.

[31] A. Rubio-Solis, and G. Panoutsos, "Fuzzy uncertainty assessment in RBF Neural Networks using neutrosophic sets for multiclass classification.", *2014 IEEE International Conference on Fuzzy Systems*, pp. 1591-1598, 2014.

[32] S. Greenfield, and R. John, "Stratification in the type-reduced set and the generalised Karnik-Mendel iterative procedure.", *Proceedings of IPMU*, vol. 8, pp. 1282-1289, 2008.

[33] N. N. Karnik, J. M. Mendel, and Q. Liang, "Type-2 fuzzy logic systems," *IEEE Transactions on Fuzzy Systems,* vol. 7, no. 6, pp. 643-658, 1999.

[34] J. M. Mendel, and R. I. B. John, "Type-2 fuzzy sets made simple," *Fuzzy Systems, IEEE Transactions on,* vol. 10, no. 2, pp. 117-127, 2002.

[35] A. Rubio-Solis, and G. Panoutsos, "Interval Type-2 Radial Basis Function Neural Network: A Modeling Framework," *Fuzzy Systems, IEEE Transactions on,* vol. 23, no. 2, pp. 457-473, 2015.

[36] W. Thomas, E. Nicholas, J. Needham *et al.*, "'Friction stir butt welding', Int. GB Patent Application No. 9125978-8," *US Patent No,* vol. 5460317, 1991.

[37] R. S. Mishra, and Z. Ma, "Friction stir welding and processing," *Materials Science and Engineering: R: Reports,* vol. 50, no. 1, pp. 1-78, 2005.

[38] X. He, F. Gu, and A. Ball, "A review of numerical analysis of friction stir welding," *Progress in Materials Science,* vol. 65, pp. 1-66, 2014.

[39] H. Okuyucu, A. Kurt, and E. Arcaklioglu, "Artificial neural network application to the friction stir welding of aluminum plates," *Materials and Design,* vol. 28, no. 1, pp. 78-84, 2007.

[40] Q. Zhang, M. Mahfouf, G. Panoutsos *et al.*, "Multiple characterisation modelling of friction stir welding using a genetic multi-objective data-driven fuzzy modelling approach.", *2011 IEEE International Conference on Fuzzy Systems*, pp. 2288-2295, 2011.

[41] Q. Zhang, M. Mahfouf, G. Panoutsos *et al.*, "Knowledge discovery for friction stir welding via data driven approaches Part 2 - Multiobjective modelling using

fuzzy rule based systems," *Science and Technology of Welding and Joining,* vol. 17, no. 8, pp. 681-693, 2012.

[42] G. P. Alicia A. Gonzalez-Rodriguez, Kirsten Sinclair, Mahdi Mahfouf, Kathryn Beamish, "Model-based process monitoring in Friction Stir Welding," *in 9th International Symposium on Friction Stir Welding ALABAMA*, USA, 2012.

[43] R. P. Paiva, and A. Dourado, "Interpretability and learning in neuro-fuzzy systems," *Fuzzy sets and systems,* vol. 147, no. 1, pp. 17-38, 2004.

[44] H. Wang, S. Kwong, Y. Jin *et al.*, "Agent-based evolutionary approach for interpretable rule-based knowledge extraction," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* vol. 35, no. 2, pp. 143-155, 2005.

[45] H. Wang, S. Kwong, Y. Jin *et al.*, "Multi-objective hierarchical genetic algorithm for interpretable fuzzy rule-based knowledge extraction," *Fuzzy sets and systems,* vol. 149, no. 1, pp. 149-186, 2005.

[46] S.-M. Zhou, and J. Q. Gan, "Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling," *Fuzzy Sets and Systems,* vol. 159, no. 23, pp. 3091-3131, 2008.

[47] N. K. Kasabov, and Q. Song, "DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction," *Fuzzy Systems, IEEE Transactions on,* vol. 10, no. 2, pp. 144-154, 2002.

[48] D. Saad, *On-line learning in neural networks*: Cambridge University Press, 2009.

[49] L. Liao, and S. Li, "On-line TS Fuzzy model identification with growing and pruning rules," *Advances in Neural Networks–ISNN 2007*, pp. 505-511: Springer, 2007.

[50] Y.-Y. Lin, J.-Y. Chang, and C.-T. Lin, "A TSK-type-based self-evolving compensatory interval type-2 fuzzy neural network (TSCIT2FNN) and its applications," *Industrial Electronics, IEEE Transactions on,* vol. 61, no. 1, pp. 447-459, 2014.

[51] C.-F. Juang, and C.-Y. Chen, "An interval type-2 neural fuzzy chip with on-chip incremental learning ability for time-varying data sequence prediction and system control," *Neural Networks and Learning Systems, IEEE Transactions on,* vol. 25, no. 1, pp. 216-228, 2014.

[52] C.-T. Lin, N. R. Pal, S.-L. Wu *et al.*, "An Interval Type-2 Neural Fuzzy System for Online System Identification and Feature Elimination," *IEEE transactions on neural networks and learning systems,* vol. 26, no. 7, pp. 1442-1455, 2015.

[53] A. K. Das, S. Sundaram, and N. Sundararajan, “A self-regulated interval type-2 neuro-fuzzy inference system for handling non-stationarities in EEG signals for BCI,”, *IEEE Transactions on Fuzzy Systems*, 2016.

[54] L. A. Zadeh, “Fuzzy logic, neural networks, and soft computing,” *Communications of the ACM,* vol. 37, no. 3, pp. 77-85, 1994.

[55] P. P. Bonissone, “Soft computing: the convergence of emerging reasoning technologies,” *Soft computing,* vol. 1, no. 1, pp. 6-18, 1997.

[56] J. Tenner, “Optimisation of the heat treatment of steel using neural networks,” The University of Sheffield 1999.

[57] Y. Yang, D. Linkens, and M. Mahfouf, “Genetic algorithms and hybrid neural network modelling for aluminium stress—strain prediction,” *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering,* vol. 217, no. 1, pp. 7-21, 2003.

[58] R. Seising, "Soft concepts” for Soft Computing in “soft sciences” on 20 years of “Soft Computing.", *2010 IEEE International Conference on Fuzzy Systems*, pp. 1-8, 2010.

[59] W. W. Smari, S.-y. Kim, W. K. McQuay *et al.*, "Toward a distributed collaborative human-centric decision-making system: An information grid approach.", *Computers, Communications, & Signal Processing with Special Track on Biomedical Engineering, 2005. CCSP 2005. 1st International Conference on. IEEE*, pp. 271-277, 2005.

[60] W. Pedrycz, and F. Gomide, *Fuzzy systems engineering: toward human-centric computing*: John Wiley & Sons, 2007.

[61] L. A. Zadeh, *Towards a theory of fuzzy systems*: World Scientific Publishing Co., Inc., 1996.

[62] B. Kosko, “Fuzzy systems as universal approximators,” *Computers, IEEE Transactions on,* vol. 43, no. 11, pp. 1329-1333, 1994.

[63] K. M. Passino, S. Yurkovich, and M. Reinfrank, *Fuzzy control*: Citeseer, 1998.

[64] L. A. Zadeh, “Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic,” *Fuzzy sets and systems,* vol. 90, no. 2, pp. 111-127, 1997.

[65] L. A. Zadeh, “Fuzzy logic= computing with words,” *Fuzzy Systems, IEEE Transactions on,* vol. 4, no. 2, pp. 103-111, 1996.

[66] T. Y. Lin, “Granular computing: Fuzzy logic and rough sets,” *Computing with words in information/intelligent systems,* vol. 1, pp. 183-200, 1999.

[67] W. Pedrycz, and A. Bargiela, “Granular clustering: a granular signature of data,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 32, no. 2, pp. 212-224, 2002.

[68] W. Pedrycz, “Granular Computing as a Framework of System Modeling,” *Journal of Control, Automation and Electrical Systems,* vol. 24, no. 1-2, pp. 81-86, 2013.

[69] J. M. Mendel, “Fuzzy logic systems for engineering: a tutorial,” *Proceedings of the IEEE,* vol. 83, no. 3, pp. 345-377, 1995.

[70] L.-X. Wang, "Fuzzy systems are universal approximators.", *1992 IEEE International Confernce on Fuzzy Systems*, pp. 1163-1170, 1992.

[71] L.-X. Wang, *A course in fuzzy systems*: Prentice-Hall press, USA, 1999.

[72] E. H. Mamdani, "Application of fuzzy algorithms for control of simple dynamic plant.", *Electrical Engineers, Proceedings of the Institution of* 121.12 (1974), pp. 1585-1588, 1974.

[73] M. Sugeno, and T. Yasukawa, “A fuzzy-logic-based approach to qualitative modeling,” *IEEE Transactions on fuzzy systems,* vol. 1, no. 1, pp. 7-31, 1993.

[74] J.-S. R. Jang, “ANFIS: adaptive-network-based fuzzy inference system,” *Systems, Man and Cybernetics, IEEE Transactions on,* vol. 23, no. 3, pp. 665-685, 1993.

[75] J. Jang, C. T. Sun, and E. Mizutani, “Neuro-fuzzy and soft computing. 1997,” *PTR Prentice Hall*, 1997.

[76] C.-T. Lin, and C. G. Lee, “Neural-network-based fuzzy logic control and decision system,” *Computers, IEEE Transactions on,* vol. 40, no. 12, pp. 1320-1336, 1991.

[77] D. Nauck, F. Klawonn, and R. Kruse, *Foundations of neuro-fuzzy systems*: John Wiley & Sons, Inc., 1997.

[78] C.-T. Lin, and C. G. Lee, “Neural fuzzy systems,” *PTR Prentice Hall*, 1996.

[79] J. S. Jang, “R., Sun CT, Mizutani E., Neuro-Fuzzy and Soft Computing,” *America: Prentice-Hall*, 1997.

[80] M.-Y. Chen, and D. A. Linkens, “A fuzzy modelling approach using hierarchical neural networks,” *Neural Computing & Applications,* vol. 9, no. 1, pp. 44-49, 2000.

[81] J. H. Holland, and J. S. Reitman, “Cognitive systems based on adaptive algorithms,” *ACM SIGART Bulletin*, no. 63, pp. 49-49, 1977.

[82] S. F. Smith, “A learning system based on genetic adaptive algorithms,” University of Pittsburgh, 1980.

[83] V. G. Sia, "A supervised inductive algorithm with genetic search for learning attributes based concepts.", *European conference on machine learning*. Springer Berlin Heidelberg, pp. 280-296, 1993.

[84] T. Xie, “Neuro-fuzzy System,” *Intelligent Systems*, 2016.

[85] L. Suganthi, S. Iniyan, and A. A. Samuel, “Applications of fuzzy logic in renewable energy systems–a review,” *Renewable and Sustainable Energy Reviews,* vol. 48, pp. 585-607, 2015.

[86] S. Kar, S. Das, and P. K. Ghosh, “Applications of neuro fuzzy systems: A brief review and future outline,” *Applied Soft Computing,* vol. 15, pp. 243-259, 2014.

[87] M. B. Gorzalczany, *Computational intelligence systems and applications: neuro-fuzzy and fuzzy neural synergisms*: Physica, 2012.

[88] K. Łapa, K. Cpałka, and L. Wang, "New Approach for Interpretability of Neuro-Fuzzy Systems with Parametrized Triangular Norms." *International Conference on Artificial Intelligence and Soft Computing*. Springer International Publishing, pp. 248-265, 2016.

[89] Y. Liu, and D. Yang, “Convergence analysis of the batch gradient-based neuro-fuzzy learning algorithm with smoothing L 1/2 regularization for the first-order Takagi–Sugeno system,” *Fuzzy Sets and Systems*, 2016.

[90] W. Lu, L. Zhang, W. Pedrycz *et al.*, “The granular extension of Sugeno-type fuzzy models based on optimal allocation of information granularity and its application to forecasting of time series,” *Applied Soft Computing,* vol. 42, pp. 38-52, 2016.

[91] W. Pedrycz, A. Jastrzebska, and W. Homenda, “Design of fuzzy cognitive maps for modeling time series,” *IEEE Transactions on Fuzzy Systems,* vol. 24, no. 1, pp. 120-130, 2016.

[92] W. Pedrycz, "Fuzzy Modelling and Fuzzy Collaborative Modelling: A Perspective of Granular Computing," *Seminal Contributions to Modelling and Simulation*, pp. 73-89: Springer, 2016.

[93] C. Coza, and C. Macnab, "A new robust adaptive-fuzzy control method applied to quadrotor helicopter stabilization." *NAFIPS 2006-2006 Annual Meeting of the North American Fuzzy Information Processing Society. IEEE*, pp. 454-458.

[94] G. Feng, “A survey on analysis and design of model-based fuzzy control systems,” *Fuzzy systems, IEEE Transactions on,* vol. 14, no. 5, pp. 676-697, 2006.

[95] B. Kosko, “Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence/book and disk,” *Vol. 1, Prentice hall*, 1992.

[96] L. A. Zadeh, “The concept of a linguistic variable and its application to approximate reasoning—I,” *Information sciences,* vol. 8, no. 3, pp. 199-249, 1975.

[97] F. Doctor, H. Hagras, and V. Callaghan, “A type-2 fuzzy embedded agent to realise ambient intelligence in ubiquitous computing environments,” *Information Sciences,* vol. 171, no. 4, pp. 309-334, 2005.

[98] H. Hagras, and C. Wagner, “Introduction to interval type-2 fuzzy logic controllers—towards better uncertainty handling in real world applications,” *IEEE Systems, Man and Cybernetics eNewsletter,* vol. 27, 2009.

[99] N. N. Karnik, and J. M. Mendel, "Applications of type-2 fuzzy logic systems: handling the uncertainty associated with surveys.", *Fuzzy Systems Conference Proceedings, 1999. FUZZ-IEEE'99. 1999 IEEE International, vol.3,* pp. 1546-1551, 1999.

[100] N. N. Karnik, and J. M. Mendel, “Applications of type-2 fuzzy logic systems to forecasting of time-series,” *Information Sciences,* vol. 120, no. 1, pp. 89-111, 1999.

[101] J. Chaneau, M. Gunaratne, and A. Altschaeffl, “An application of type-2 sets to decision making in engineering,” *Analysis of Fuzzy Information,* vol. 2, pp. 145-151, 1987.

[102] R. R. Yager, “Fuzzy subsets of type II in decisions,” *Cybernetics and System,* vol. 10, no. 1-3, pp. 137-159, 1980.

[103] M. Wagenknecht, and K. Hartmann, “Application of fuzzy sets of type 2 to the solution of fuzzy equations systems,” *Fuzzy Sets and Systems,* vol. 25, no. 2, pp. 183-190, 1988.

[104] R. John, P. Innocent, and M. Barnes, "Type 2 fuzzy sets and neuro-fuzzy clustering of radiographic tibia images.", *Fuzzy Systems Conference Proceedings, 1999. FUZZ-IEEE'99. 1999 IEEE International,* vol.3*,* pp. 1373-1376, 1999.

[105] D. Wu, and W. W. Tan, "A type-2 fuzzy logic controller for the liquid-level process.", *Proceedings of 2004 IEEE International Conference on Fuzzy Systems,* vol.2, pp. 953-958, 2004.

[106] J. Mendel, H. Hagras, W.-W. Tan *et al.*, *Introduction to type-2 fuzzy logic control: theory and applications*: John Wiley & Sons, 2014.

[107] N. N. Karnik, and J. M. Mendel, "Introduction to type-2 fuzzy logic systems.", *Fuzzy Systems Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on,* vol.2, pp. 915-920, 1998.

[108] J. M. Mendel, *Uncertain rule-based fuzzy logic system: introduction and new directions*, Upper Saddle River, NJ: Prentice-Hall, 2001.

[109] Q. Liang, and J. M. Mendel, "Interval type-2 fuzzy logic systems: theory and design," *Fuzzy Systems, IEEE Transactions on,* vol. 8, no. 5, pp. 535-550, 2000.

[110] J. M. Mendel, "Type-2 fuzzy sets and systems: an overview," *Computational Intelligence Magazine, IEEE,* vol. 2, no. 1, pp. 20-29, 2007.

[111] N. N. Karnik, and J. M. Mendel, "Centroid of a type-2 fuzzy set," *Information Sciences,* vol. 132, no. 1, pp. 195-220, 2001.

[112] H. Wu, and J. M. Mendel, "Uncertainty bounds and their use in the design of interval type-2 fuzzy logic systems," *IEEE Transactions on Fuzzy Systems,* vol. 10, no. 5, pp. 622-639, 2002.

[113] M. Nie, and W. W. Tan, "Towards an efficient type-reduction method for interval type-2 fuzzy logic systems.", Fuzzy *Systems, 2008. FUZZ-IEEE 2008.(IEEE World Congress on Computational Intelligence). IEEE International Conference on. IEEE*, pp. 1425-1432, 2008.

[114] S. Greenfield, F. Chiclana, S. Coupland *et al.*, "The collapsing method of defuzzification for discretised interval type-2 fuzzy sets," *Information Sciences,* vol. 179, no. 13, pp. 2055-2069, 2009.

[115] R. I. J. Sarah Greenfield, and Simon Coupland, "A novel sampling method for type-2 defuzzification.", *Proc. UKCI* 6 (2005), pp. 120-127, 2005.

[116] K.-L. Du, and M. N. Swamy, *Neural networks in a softcomputing framework*: Springer, 2006.

[117] W. S. McCulloch, and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics,* vol. 5, no. 4, pp. 115-133, 1943.

[118] M. Berthold, and D. J. Hand, *Intelligent data analysis: an introduction*: Springer Science & Business Media, 2003.

[119] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural networks,* vol. 4, no. 2, pp. 251-257, 1991.

[120] S. Sivanandam, and S. Deepa, *Introduction to neural networks using Matlab 6.0*: Tata McGraw-Hill Education, 2006.

[121] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning internal representations by error propagation*, DTIC Document, 1985.

[122] S. S. Haykin, *Neural networks and learning machines*: Pearson Education Upper Saddle River, 2009.

[123] D. Lowe, “Multi-variable functional interpolation and adaptive networks,” *Complex Systems,* vol. 2, pp. 321-355.

[124] P. J. Davis, *Interpolation and approximation*: Courier Corporation, 1975.

[125] C. A. Micchelli, *Interpolation of scattered data: distance matrices and conditionally positive definite functions*: Springer, 1984.

[126] C. M. Bishop, *Pattern recognition and machine learning*: springer New York, 2006.

[127] C. M. Bishop, *Neural networks for pattern recognition*: Oxford university press, 1995.

[128] J. Moody, and C. J. Darken, “Fast learning in networks of locally-tuned processing units,” *Neural computation,* vol. 1, no. 2, pp. 281-294, 1989.

[129] J. T. Yao, A. V. Vasilakos, and W. Pedrycz, “Granular computing: perspectives and challenges,” *Cybernetics, IEEE Transactions on,* vol. 43, no. 6, pp. 1977-1989, 2013.

[130] L. Zadeh, "Fuzzy sets and information granurity, Advances in Fuzzy Set Theory and Applications", M. Gupta, RK Ragade, RR Yager," North-Holland Publishing Company, pp3-18, 1979.

[131] L. A. Zadeh, “Toward a perception-based theory of probabilistic reasoning with imprecise probabilities,” *Journal of statistical planning and inference,* vol. 105, no. 1, pp. 233-264, 2002.

[132] R. E. Moore, *Interval analysis*: Prentice-Hall Englewood Cliffs, 1966.

[133] W. Pedrycz, and Z. A. Sosnowski, “The design of decision trees in the framework of granular data and their application to software quality models,” *Fuzzy Sets and Systems,* vol. 123, no. 3, pp. 271-290, 2001.

[134] S. Butenkov, "Granular computing in image processing and understanding." *Proc. IASTED Int. Conf. Artificial Intelligence and Applns*. pp. 811-816, 2004.

[135] W. Pedrycz, *Granular computing: an emerging paradigm*: Springer Science & Business Media, 2001.

[136] L. A. Zadeh, “Quantitative fuzzy semantics,” *Information sciences,* vol. 3, no. 2, pp. 159-176, 1971.

[137] R. Nandan, T. DebRoy, and H. Bhadeshia, “Recent advances in friction-stir welding–Process, weldment structure and properties,” *Progress in Materials Science,* vol. 53, no. 6, pp. 980-1023, 2008.

[138] J. M. Rice, S. Mandal, and A. A. Elmustafa, "Microstructural investigation of donor material experiments in friction stir welding," *International Journal of Material Forming*, pp. 1-11, 2012.

[139] D. M. Neto, and P. Neto, "Numerical modeling of friction stir welding process: a literature review," *The International Journal of Advanced Manufacturing Technology,* vol. 65, no. 1-4, pp. 115-126, 2013.

[140] S. Guerdoux, "Simulation numérique du soudage par frottement malaxage," Paris, ENMP, 2007.

[141] S. Packer, T. Nelson, C. Sorensen *et al.*, "Tool and equipment requirements for friction stir welding ferrous and other high melting temperature alloys.", *Proceedings of the Fourth International Symposium on Friction Stir Welding,* 2003.

[142] N. Dialami, M. Chiumenti, M. Cervera *et al.*, "An apropos kinematic framework for the numerical modeling of friction stir welding," *Computers & Structures,* vol. 117, pp. 48-57, 2013.

[143] P. Threadgill, A. Leonard, H. Shercliff *et al.*, "Friction stir welding of aluminium alloys," *International Materials Reviews,* vol. 54, no. 2, pp. 49-93, 2009.

[144] Z. Kiss, and T. Czigány, "Applicability of friction stir welding in polymeric materials," *Periodica Polytechnica. Engineering. Mechanical Engineering,* vol. 51, no. 1, pp. 15, 2007.

[145] A. J. Ramirez, and M. C. Juhas, "Microstructural evolution in Ti-6Al-4V friction stir welds.", *Materials Science Forum*. Vol. 426. Trans Tech Publications pp. 2999-3004, 2003.

[146] N. Afrin, D. Chen, X. Cao *et al.*, "Microstructure and tensile properties of friction stir welded AZ31B magnesium alloy," *Materials Science and Engineering: A,* vol. 472, no. 1, pp. 179-186, 2008.

[147] H. Pashazadeh, A. Masoumi, and J. Teimournezhad, "Numerical modelling for the hardness evaluation of friction stir welded copper metals," *Materials & Design,* vol. 49, pp. 913-921, 2013.

[148] T. Sakthivel, and J. Mukhopadhyay, "Microstructure and mechanical properties of friction stir welded copper," *Journal of materials science,* vol. 42, no. 19, pp. 8126-8129, 2007.

[149] W. Thomas, and E. Nicholas, "Friction stir welding for the transportation industries," *Materials & Design,* vol. 18, no. 4, pp. 269-273, 1997.

[150] T. Lienert, W. Stellwag Jr, B. Grimmett *et al.*, "Friction stir welding studies on mild steel," *Welding Journal-New York,* vol. 82, no. 1, pp. 1-S, 2003.

[151] S. Cater, *Fundamentals of Friction Stir Welding*: TWI Ltd, 2014.

[152] W. R. Longhurst, “Force control of friction stir welding,” Vanderbilt University, 2009.

[153] T. Santos, P. Vilaça, and L. Quintino, “Developments in NDT for detecting imperfections in friction stir welds in aluminium alloys,” *Welding in the World,* vol. 52, no. 9-10, pp. 30-37, 2008.

[154] A. Leonard, and S. Lockyer, "Flaws in friction stir welds."

[155] B. Li, Y. Shen, and W. Hu, “The study on defects in aluminum 2219-T6 thick butt friction stir welds with the application of multiple non-destructive testing methods,” *Materials & Design,* vol. 32, no. 4, pp. 2073-2084, 2011.

[156] S. Lomolino, R. Tovo, and J. Dos Santos, “On the fatigue behaviour and design curves of friction stir butt-welded Al alloys,” *International Journal of Fatigue,* vol. 27, no. 3, pp. 305-316, 2005.

[157] Y. Kim, H. Fujii, T. Tsumura *et al.*, “Three defect types in friction stir welding of aluminum die casting alloy,” *Materials Science and Engineering: A,* vol. 415, no. 1, pp. 250-254, 2006.

[158] A. Baraka, G. Panoutsos, and S. Cater, “A real-time quality monitoring framework for steel friction stir welding using computational intelligence,” *Journal of Manufacturing Processes,* vol. 20, pp. 137-148, 2015.

[159] S. L. Barbour, and J. Krahn, “Numerical modelling–Prediction or process,” *Geotechnical News,* vol. 22, no. 4, pp. 44-52, 2004.

[160] B. Gibson, D. Lammlein, T. Prater *et al.*, “Friction stir welding: process, automation, and control,” *Journal of Manufacturing Processes,* vol. 16, no. 1, pp. 56-73, 2014.

[161] R. S. Mishra, P. S. De, and N. Kumar, *Fundamentals of the Friction Stir Process*: Springer, 2014.

[162] M. J. Russell, and H. Shercliff, “Analytical modelling of friction stir welding,” *Analytical modelling of friction stir welding Russell, MJ; Shercliff, R Inalco,* vol. 98, pp. 197-207, 1999.

[163] T. Cebeci, J. P. Shao, F. Kafyeke *et al.*, *Computational fluid dynamics for engineers: from panel to navier-stokes methods with computer programs*: Springer Berlin Heidelberg New York, 2005.

[164] Z. Feng, J. Gould, and T. Lienert, “A heat flow model for friction stir welding of aluminum alloys,” *Hot Deformation of Aluminum Alloys II*, pp. 149-158, 1998.

[165] J. E. Gould, and Z. Feng, “Heat flow model for friction stir welding of aluminum alloys,” *Journal of Materials Processing and Manufacturing Science*, no. 7, pp. 185-194, 1998.

[166] M. Russell, H. Shercliff, and P. Threadgill, "Development and application of an analytical process model for friction stir welding.", *Automotive Aluminum 2001: Proceedings of a Conference of the TMS Annual Meeting*. TMS: The Minerals, Metals and Materials Society, pp. 225-234, 2001.

[167] C. Chen, and R. Kovacevic, “Finite element modeling of friction stir welding—thermal and thermomechanical analysis,” *International Journal of Machine Tools and Manufacture,* vol. 43, no. 13, pp. 1319-1326, 2003.

[168] H. Zhang, Z. Zhang, and J. Chen, “The finite element simulation of the friction stir welding process,” *Materials Science and Engineering: A,* vol. 403, no. 1, pp. 340-348, 2005.

[169] C. Chen, and R. Kovacevic, “Thermomechanical modelling and force analysis of friction stir welding by the finite element method,” *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science,* vol. 218, no. 5, pp. 509-519, 2004.

[170] X. Deng, and S. Xu, “Two-dimensional finite element simulation of material flow in the friction stir welding process,” *Journal of manufacturing processes,* vol. 6, no. 2, pp. 125-133, 2004.

[171] P. A. Colegrove, and H. R. Shercliff, “3-Dimensional CFD modelling of flow round a threaded friction stir welding tool profile,” *Journal of Materials Processing Technology,* vol. 169, no. 2, pp. 320-327, 2005.

[172] A. Arora, R. Nandan, A. Reynolds *et al.*, “Torque, power requirement and stir zone geometry in friction stir welding through modeling and experiments,” *Scripta Materialia,* vol. 60, no. 1, pp. 13-16, 2009.

[173] H. Atharifar, D. Lin, and R. Kovacevic, “Numerical and experimental investigations on the loads carried by the tool during friction stir welding,” *Journal of Materials Engineering and Performance,* vol. 18, no. 4, pp. 339-350, 2009.

[174] S. Guerdoux, and L. Fourment, “A 3D numerical simulation of different phases of friction stir welding,” *Modelling and simulation in materials science and engineering,* vol. 17, no. 7, pp. 075001, 2009.

[175] Z. Zhang, and H. Zhang, “Numerical studies on the effect of transverse speed in friction stir welding,” *Materials & Design,* vol. 30, no. 3, pp. 900-907, 2009.

[176] G. Buffa, G. Campanile, L. Fratini *et al.*, “Friction stir welding of lap joints: Influence of process parameters on the metallurgical and mechanical

properties," *Materials Science and Engineering: A,* vol. 519, no. 1, pp. 19-26, 2009.

[177] C. Hamilton, A. Sommers, and S. Dymek, "A thermal model of friction stir welding applied to Sc-modified Al–Zn–Mg–Cu alloy extrusions," *International Journal of Machine Tools and Manufacture,* vol. 49, no. 3, pp. 230-238, 2009.

[178] A. K. Lakshminarayanan, and V. Balasubramanian, "Comparison of RSM with ANN in predicting tensile strength of friction stir welded AA7039 aluminium alloy joints," *Transactions of Nonferrous Metals Society of China (English Edition),* vol. 19, no. 1, pp. 9-18, 2009.

[179] D. Jacquin, B. De Meester, A. Simar *et al.*, "A simple Eulerian thermomechanical modeling of friction stir welding," *Journal of Materials Processing Technology,* vol. 211, no. 1, pp. 57-65, 2011.

[180] M. Jayaraman, R. Sivasubramanian, V. Balasubramanian *et al.*, "Application of RSM and ANN to predict the tensile strength of Friction Stir Welded A319 cast aluminium alloy," *International Journal of Manufacturing Research,* vol. 4, no. 3, pp. 306-323, 2009.

[181] H. Atharifar, "Optimum parameters design for friction stir spot welding using a genetically optimized neural network system," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture,* vol. 224, no. 3, pp. 403-418, 2010.

[182] I. N. Tansel, M. Demetgul, H. Okuyucu *et al.*, "Optimizations of friction stir welding of aluminum alloy by using genetically optimized neural network," *International Journal of Advanced Manufacturing Technology,* vol. 48, no. 1-4, pp. 95-101, 2010.

[183] M. H. Shojaeefard, R. A. Behnagh, M. Akbari *et al.*, "Modelling and pareto optimization of mechanical properties of friction stir welded AA7075/AA5083 butt joints using neural network and particle swarm algorithm," *Materials and Design,* vol. 44, pp. 190-198, 2013.

[184] H. Wang, P. A. Colegrove, and J. dos Santos, "Hybrid modelling of 7449-T7 aluminium alloy friction stir welded joints," *Science and Technology of Welding and Joining,* vol. 18, no. 2, pp. 147-153, 2013.

[185] C. Patel, S. Das, and R. G. Narayanan, "CAFE modeling, neural network modeling, and experimental investigation of friction stir welding," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science,* vol. 227, no. 6, pp. 1164-1176, 2013.

[186] H. K. Mohanty, M. M. Mahapatra, P. Kumar *et al.*, "Predicting the effects of tool geometries on friction stirred aluminium welds using artificial neural networks and fuzzy logic techniques," *International Journal of Manufacturing Research,* vol. 8, no. 3, pp. 296-312, 2013.

[187] E. J. Macías, A. S. Roca, H. C. Fals *et al.*, “Neural networks and acoustic emission for modelling and characterization of the friction stir welding process,” *Emisión Acústica y Redes Neuronales para Modelado y Caracterización del Proceso de Soldadura por Fricción Agitación,* vol. 10, no. 4, pp. 434-440, 2013.

[188] Q. Zhang, M. Mahfouf, G. Panoutsos *et al.*, “Knowledge discovery for friction stir welding via data driven approaches part 1 - Correlation analyses of internal process variables and weld quality,” *Science and Technology of Welding and Joining,* vol. 17, no. 8, pp. 672-680, 2012.

[189] M. Soron, and I. Kalaykov, "A robot prototype for friction stir welding." *2006 IEEE Conference on Robotics, Automation and Mechatronics*. IEEE, pp. 1-5, 2006.

[190] X. Zhao, P. Kalya, R. G. Landers *et al.*, "Design and implementation of a nonlinear axial force controller for friction stir welding processes.", MISSOURI UNIV-ROLLA pp. 5553-5558, 2009.

[191] T. I. Van Niekerk, T. Hua, and D. G. Hattingh, "A neuro-fuzzy scheme for process control during complex curvature friction stir welding.", *IFAC Proceedings Volumes,* vol. 39, no.3, pp.113-118, 2006.

[192] C. D. Cox, B. T. Gibson, A. M. Strauss *et al.*, “Energy input during friction stir spot welding,” *Journal of Manufacturing Processes,* vol. 16, no. 4, pp. 479-484, 2014.

[193] H. Su, C. Wu, A. Pittner *et al.*, “Simultaneous measurement of tool torque, traverse force and axial force in friction stir welding,” *Journal of Manufacturing processes,* vol. 15, no. 4, pp. 495-500, 2013.

[194] E. Boldsaikhan, E. M. Corwin, A. M. Logar *et al.*, “The use of neural network and discrete Fourier transform for real-time evaluation of friction stir welding,” *Applied Soft Computing Journal,* vol. 11, no. 8, pp. 4839-4846, 2011.

[195] A. Baraka, G. Panoutsos, M. Mahfouf *et al.*, “A Shannon Entropy-Based Conflict Measure For Enhancing Granular Computing-Based Information Processing,” *in 2014 IEEE International Conference on Granular Computing (GrC)*, Noboribetsu, Japan, 2014, pp. 13 - 18.

[196] J. A. Hartigan, and M. A. Wong, “Algorithm AS 136: A k-means clustering algorithm,” *Journal of the Royal Statistical Society. Series C (Applied Statistics),* vol. 28, no. 1, pp. 100-108, 1979.

[197] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*: Springer Science & Business Media, 2013.

[198] R. R. Yager, and D. P. Filev, “Generation of fuzzy rules by mountain clustering,” *Journal of Intelligent & Fuzzy Systems,* vol. 2, no. 3, pp. 209-219, 1994.

[199] M. Buragohain, and C. Mahanta, “A novel approach for ANFIS modelling based on full factorial design,” *Applied Soft Computing,* vol. 8, no. 1, pp. 609-625, 2008.

[200] C. Mencar, G. Castellano, and A. M. Fanelli, “Distinguishability quantification of fuzzy sets,” *Information Sciences,* vol. 177, no. 1, pp. 130-149, 2007.

[201] H. A. Hefny, “Comments on “Distinguishability quantification of fuzzy sets”,” *Information Sciences,* vol. 177, no. 21, pp. 4832-4839, 2007.

[202] F. Girosi, M. B. Jones, and T. Poggio, “Regularization theory and neural networks architectures,” *Neural computation,* vol. 7, no. 2, pp. 219-269, 1995.

[203] Z. Huang, “Extensions to the k-means algorithm for clustering large data sets with categorical values,” *Data mining and knowledge discovery,* vol. 2, no. 3, pp. 283-304, 1998.

[204] J.-S. R. Jang, C.-T. Sun, and E. Mizutani, “Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence,” Prentice Hall, Upper Saddle River, 1997.

[205] C. Mencar, and A. M. Fanelli, “Interpretability constraints for fuzzy information granulation,” *Information Sciences,* vol. 178, no. 24, pp. 4585-4618, 2008.

[206] Y. Yang, "Noise reduction in a statistical approach to text categorization.", *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, pp. 256-263, 1995.

[207] J. D. Farmer, and J. J. Sidorowich, *Exploiting chaos to predict the future and reduce noise*: World Scientific, Singapore, 1988.

[208] C. Wu, S.-C. Kao, and K. Okuhara, “Examination and comparison of conflicting data in granulated datasets: Equal width interval vs. equal frequency interval,” *Information Sciences,* vol. 239, pp. 154-164, 2013.

[209] Y. Li, and H. Dai, "Reducing uncertainties in data mining.", *Software Engineering Conference, 1997. Asia Pacific... and International Computer Science Conference 1997. APSEC'97 and ICSC'97. Proceedings*. IEEE, pp. 97-105, 1997.

[210] M. Chau, R. Cheng, and B. Kao, "Uncertain data mining: A new research direction.", *Proceedings of the Workshop on the Sciences of the Artificial,* 2005.

[211] N. R. Pal, and S. K. Pal, “Entropy: a new definition and its applications,” *Systems, Man and Cybernetics, IEEE Transactions on,* vol. 21, no. 5, pp. 1260-1270, 1991.

[212] Y. Jin, W. Von Seelen, and B. Sendhoff, "An approach to rule-based knowledge extraction.", Fuzzy *Systems Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Conference on,* vol.2, pp. 1188-1193, 1998.

[213] H. Roubos, and M. Setnes, “Compact and transparent fuzzy models and classifiers through iterative complexity reduction,” *Fuzzy Systems, IEEE Transactions on,* vol. 9, no. 4, pp. 516-524, 2001.

[214] J. Yen, and L. Wang, “Simplifying fuzzy rule-based models using orthogonal transformation methods,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 29, no. 1, pp. 13-24, 1999.

[215] D. Nauck, and R. Kruse, “Obtaining interpretable fuzzy classification rules from medical data,” *Artificial intelligence in medicine,* vol. 16, no. 2, pp. 149-169, 1999.

[216] Y. Jin, and B. Sendhoff, “Extracting interpretable fuzzy rules from RBF networks,” *Neural Processing Letters,* vol. 17, no. 2, pp. 149-164, 2003.

[217] J. Park, and I. W. Sandberg, “Approximation and radial-basis-function networks,” *Neural computation,* vol. 5, no. 2, pp. 305-316, 1993.

[218] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: a review,” *ACM computing surveys (CSUR),* vol. 31, no. 3, pp. 264-323, 1999.

[219] S. C. Johnson, “Hierarchical clustering schemes,” *Psychometrika,* vol. 32, no. 3, pp. 241-254, 1967.

[220] R. Xu, and D. Wunsch, “Survey of clustering algorithms,” *IEEE Transactions on neural networks,* vol. 16, no. 3, pp. 645-678, 2005.

[221] W. Pedrycz, and G. Vukovich, “Granular neural networks,” *Neurocomputing,* vol. 36, no. 1, pp. 205-224, 2001.

[222] A. Bargiela, W. Pedrycz, and K. Hirota, “Granular prototyping in fuzzy clustering,” *IEEE Transactions on Fuzzy Systems,* vol. 12, no. 5, pp. 697-709, 2004.

[223] W. Lu, W. Pedrycz, X. Liu *et al.*, “The modeling of time series based on fuzzy information granules,” *Expert Systems with Applications,* vol. 41, no. 8, pp. 3799-3808, 2014.

[224] G. J. Klir, “Where do we stand on measures of uncertainty, ambiguity, fuzziness, and the like?,” *Fuzzy Sets and Systems,* vol. 24, no. 2, pp. 141-160, 1987.

[225] G. J. Klir, and T. A. Folger, “Fuzzy sets, uncertainty, and information,” Hall,1988.

[226] N. R. Pal, and S. K. Pal, “Higher order fuzzy entropy and hybrid entropy of a set,” *Information Sciences,* vol. 61, no. 3, pp. 211-231, 1992.

[227] U. Höhle, "Fuzzy plausibility measures.", *Proceedings of the 3th International Seminar on Fuzzy Set Theory, Johannes Kepler University, Linz,* pp. 7-30, 1981.

[228] R. R. Yager, “Entropy and specificity in a mathematical theory of evidence,” *International Journal of General System,* vol. 9, no. 4, pp. 249-260, 1983.

[229] G. J. Klir, and A. Ramer, “Uncertainty in the Dempster-Shafer theory: a critical re-examination,” *International Journal of General System,* vol. 18, no. 2, pp. 155-166, 1990.

[230] M. Higashi, and G. J. Klir, “Measures of uncertainty and information based on possibility distributions,” *International Journal of General Systems,* vol. 9, no. 1, pp. 43-58, 1982.

[231] D. Dubois, and H. Prade, “A note on measures of specificity for fuzzy sets,” *International Journal of General System,* vol. 10, no. 4, pp. 279-283, 1985.

[232] N. R. Pal, J. C. Bezdek, and R. Hemasinha, “Uncertainty measures for evidential reasoning I: A review,” *International Journal of Approximate Reasoning,* vol. 7, no. 3, pp. 165-183, 1992.

[233] N. R. Pal, J. C. Bezdek, and R. Hemasinha, “Uncertainty measures for evidential reasoning II: A new measure of total uncertainty,” *International Journal of Approximate Reasoning,* vol. 8, no. 1, pp. 1-16, 1993.

[234] M. T. LAMATA, and S. MORAL, “Measures of entropy in the theory of evidence,” *International Journal Of General System,* vol. 14, no. 4, pp. 297-305, 1988.

[235] A.-L. Jousselme, C. Liu, D. Grenier *et al.*, “Measuring ambiguity in the evidence theory,” *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans,* vol. 36, no. 5, pp. 890-903, 2006.

[236] C. E. Shannon, “A mathematical theory of communication,” *ACM SIGMOBILE Mobile Computing and Communications Review,* vol. 5, no. 1, pp. 3-55, 2001.

[237] D. Dubois, H. Prade, and L. Ughetto, “Checking the coherence and redundancy of fuzzy knowledge bases,” *Fuzzy Systems, IEEE Transactions on,* vol. 5, no. 3, pp. 398-417, 1997.

[238] Y. Jin, “Fuzzy modeling of high-dimensional systems: complexity reduction and interpretability improvement,” *IEEE Transactions on Fuzzy Systems,* vol. 8, no. 2, pp. 212-221, 2000.

[239] Y. Jin, W. Von Seelen, and B. Sendhoff, “On generating FC 3 fuzzy rule systems from data using evolution strategies,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics),* vol. 29, no. 6, pp. 829-845, 1999.

[240] Y. Jin, "Generating distinguishable, complete, consistent and compact fuzzy systems using evolutionary algorithms," *Accuracy improvements in linguistic fuzzy modeling*, pp. 100-118: Springer, 2003.

[241] M. Setnes, R. Babuška, U. Kaymak *et al.*, “Similarity measures in fuzzy rule base simplification,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 28, no. 3, pp. 376-386, 1998.

[242] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” *Annals of eugenics,* vol. 7, no. 2, pp. 179-188, 1936.

[243] J.-S. R. Jang, and C.-T. Sun, *Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence*: Prentice-Hall, Inc., 1996.

[244] E. H. Mamdani, “Application of fuzzy logic to approximate reasoning using linguistic synthesis,” *IEEE transactions on computers,* vol. 100, no. 12, pp. 1182-1191, 1977.

[245] R. Tong, “A control engineering review of fuzzy systems,” *Automatica,* vol. 13, no. 6, pp. 559-569, 1977.

[246] C.-C. Lee, “Fuzzy logic in control systems: fuzzy logic controller. II,” *IEEE Transactions on systems, man, and cybernetics,* vol. 20, no. 2, pp. 419-435, 1990.

[247] B. Kosko, *Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence/book and disk*: Prentice Hall, Upper Saddle River, 1992.

[248] H. A. Hagras, “A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots,” *IEEE Transactions on Fuzzy systems,* vol. 12, no. 4, pp. 524-539, 2004.

[249] D. Wu, and W. W. Tan, "Type-2 FLS modeling capability analysis." pp. 242-247.

[250] N. N. Karnik, and J. M. Mendel, "Type-2 fuzzy logic systems: type-reduction.", *Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on,* vol. 2, pp. 2046-2051, 1998.

[251] D. Wu, and J. M. Mendel, “Enhanced karnik--mendel algorithms,” *Fuzzy Systems, IEEE Transactions on,* vol. 17, no. 4, pp. 923-934, 2009.

[252] K. J. Hunt, R. Haas, and R. Murray-Smith, “Extending the functional equivalence of radial basis functionnetworks and fuzzy inference systems,” *IEEE Transactions on Neural Networks,* vol. 7, no. 3, pp. 776-781, 1996.

[253] B. A. Foss, and T. A. Johansen, "On local and fuzzy modelling.", *Industrial Fuzzy Control and Intelligent Systems, 1993., IFIS'93., Third International Conference on.,* pp. 80-87, 1993.

[254] C.-H. Wang, C.-S. Cheng, and T.-T. Lee, “Dynamical optimal training for interval type-2 fuzzy neural network (T2FNN),” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 34, no. 3, pp. 1462-1477, 2004.

[255] R. R. Yager, and D. P. Filev, “Approximate clustering via the mountain method,” *Systems, Man and Cybernetics, IEEE Transactions on,* vol. 24, no. 8, pp. 1279-1284, 1994.

[256] J. BBezdek, "Pattern Recognition with Fuzzy Objective Algorithms," New York: Plenum Press, 1981.

[257] B.-I. Choi, and F. C.-H. Rhee, “Interval type-2 fuzzy membership function generation methods for pattern recognition,” *Information Sciences,* vol. 179, no. 13, pp. 2102-2122, 2009.

[258] N. Benoudjit, C. Archambeau, A. Lendasse *et al.*, "Width optimization of the Gaussian kernels in Radial Basis Function Networks.", *ESANN,* vol.2, pp. 425-432, 2002.

[259] J. Zeng, L. Xie, and Z.-Q. Liu, “Type-2 fuzzy Gaussian mixture models,” *Pattern Recognition,* vol. 41, no. 12, pp. 3636-3643, 2008.

[260] H. Hagras, “Comments on" dynamical optimal training for interval type-2 fuzzy neural network (T2FNN),” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 36, no. 5, pp. 1206-1209, 2006.

[261] J. Shneider, “Temperature distribution and resulting metal flow,” *Friction stir welding and processing*, pp. 37-49, 2007.

[262] T. North, G. Bendzsak, and C. Smith, "Material properties relevant to 3-D FSW modeling.", *Friction Stir Welding, Proceedings of the 2nd International Conference.*, 2000.

[263] O. Midling, E. Morley, and A. Kluken, "Joining of aluminium constructions by friction stir welding.", *Int. Symp. Advanced Transportation Applications, ISATA27*,1994.

[264] P. Haagensen, O. Midling, and M. Ranes, "Fatigue performance of friction stir butt welds in a 6000 series aluminum alloy," *International Journal of Fatigue,* vol. 18, no. 7, pp. 508-508, 1996.

[265] K. Elangovan, and V. Balasubramanian, "Influences of tool pin profile and welding speed on the formation of friction stir processing zone in AA2219 aluminium alloy," *Journal of materials processing technology,* vol. 200, no. 1, pp. 163-175, 2008.

[266] G. Kaiser, *A friendly guide to wavelets*: Springer Science & Business Media, 2010.

[267] W. T. Cochran, J. W. Cooley, D. L. Favin *et al.*, "What is the fast Fourier transform?," *Proceedings of the IEEE,* vol. 55, no. 10, pp. 1664-1674, 1967.

[268] J. W. Cooley, and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Mathematics of computation,* vol. 19, no. 90, pp. 297-301, 1965.

[269] W. J. Arbegast, "Using process forces as a statistical process control tool for friction stir welds.", *Friction Stir Welding and Processing III as held at the 2005 TMS Annual Meeting*, pp. 193-204, 2005.

[270] W. J. Arbegast, "A flow-partitioned deformation zone model for defect formation during friction stir welding," *Scripta materialia,* vol. 58, no. 5, pp. 372-376, 2008.

[271] A. Logar, E. Corwin, E. Boldsaikhan *et al.*, "Applications of artificial intelligence to friction stir welding.", *National Conference on Recent Advancements in Information Technology*, pp. 9-10,2007.

[272] P. Fleming, D. Lammlein, D. Wilkes *et al.*, "In-process gap detection in friction stir welding," *Sensor Review,* vol. 28, no. 1, pp. 62-67, 2008.

[273] K. Colligan, "Material flow behavior during friction welding of aluminum," *Weld J,* vol. 75, no. 7, pp. 229s-237s, 1999.

[274] A. Reynolds, T. Seidel, and M. Simosen, "Visualisation of Material Flow in an Autogenous Friction Stir Weld in 1st International Symposium on Friction Stir Welding 1999. Thousand Oaks, CA: Abington, Cambridge CB1 6AH, UK," Abington Publishing., 1997

[275] B. London, M. Mahoney, W. Bingel *et al.*, "Material flow in friction stir welding monitored with Al-SiC and Al-W composite markers," *TMS2003, CA, K. Jata et al., eds*, pp. 3, 2003.

[276] H. Schmidt, and J. Hattel, “A local model for the thermomechanical conditions in friction stir welding,” *Modelling and Simulation in Materials Science and Engineering,* vol. 13, no. 1, pp. 77, 2005.

[277] J. Lee Rodgers, and W. A. Nicewander, “Thirteen ways to look at the correlation coefficient,” *The American Statistician,* vol. 42, no. 1, pp. 59-66, 1988.

[278] I. Nabney, *NETLAB: algorithms for pattern recognition*: Springer Science & Business Media, 2002.

[279] A. Baraka, G. Panoutsos, and S. Cater, “Real-time quality monitoring for friction stir welding AA2219-T845 aluminium aerospace alloy via model based spectral analysis,” *in 11th International Symposium on Friction Stir Welding*, Cambridge, UK, 2016.

[280] C.-T. Lin, *Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems*: Prentice hall PTR, 1996.

[281] H.-J. Rong, N. Sundararajan, G.-B. Huang *et al.*, “Sequential adaptive fuzzy inference system (SAFIS) for nonlinear system identification and prediction,” *Fuzzy sets and systems,* vol. 157, no. 9, pp. 1260-1275, 2006.

[282] G.-B. Huang, P. Saratchandran, and N. Sundararajan, “A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation,” *IEEE Transactions on Neural Networks,* vol. 16, no. 1, pp. 57-67, 2005.

[283] R. J. Oentaryo, M. J. Er, S. Linn *et al.*, “Online probabilistic learning for fuzzy inference system,” *Expert Systems with Applications,* vol. 41, no. 11, pp. 5082-5096, 2014.

[284] S. Ben-David, E. Kushilevitz, and Y. Mansour, “Online learning versus offline learning,” *Machine Learning,* vol. 29, no. 1, pp. 45-63, 1997.

[285] R. S. Sutton, and S. D. Whitehead, "Online learning with random representations.", *Proceedings of the Tenth International Conference on Machine Learning*, pp. 314-321, 1993.

[286] G. E. Box, G. M. Jenkins, G. C. Reinsel *et al.*, *Time series analysis: forecasting and control*: John Wiley & Sons, 2015.

[287] N. Kasabov, E. Postma, and J. Van Den Herik, “AVIS: a connectionist-based framework for integrated auditory and visual information processing,” *Information Sciences,* vol. 123, no. 1, pp. 127-148, 2000.

[288] M. Wooldridge, and N. R. Jennings, “Intelligent agents: Theory and practice,” *The knowledge engineering review,* vol. 10, no. 02, pp. 115-152, 1995.

[289] G. G. Yen, and P. Meesad, “An effective neuro-fuzzy paradigm for machinery condition health monitoring,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 31, no. 4, pp. 523-536, 2001.

[290] M. Pratama, M. J. Er, X. Li *et al.*, “Data driven modeling based on dynamic parsimonious fuzzy neural network,” *Neurocomputing,* vol. 110, pp. 18-28, 2013.

[291] J. S. Albus, “A new approach to manipulator control: The cerebellar model articulation controller (CMAC),” *Journal of Dynamic Systems, Measurement, and Control,* vol. 97, no. 3, pp. 220-227, 1975.

[292] H. Hagras, F. Doctor, V. Callaghan *et al.*, “An incremental adaptive life long learning approach for type-2 fuzzy embedded agents in ambient intelligent environments,” *IEEE Transactions on Fuzzy Systems,* vol. 15, no. 1, pp. 41-55, 2007.

[293] W. Shen, Q. Hao, H. J. Yoon *et al.*, “Applications of agent-based systems in intelligent manufacturing: An updated review,” *Advanced engineering INFORMATICS,* vol. 20, no. 4, pp. 415-431, 2006.

[294] G. Panoutsos, and M. Mahfouf, "An incremental learning structure using granular computing and model fusion with application to materials processing," *Intelligent Techniques and Tools for Novel System Architectures*, pp. 139-153: Springer, 2008.

[295] J. Mańdziuk, and L. Shastri, “Incremental class learning approach and its application to handwritten digit recognition,” *Information Sciences,* vol. 141, no. 3, pp. 193-217, 2002.

[296] S. Schaal, and C. G. Atkeson, “Constructive incremental learning from only local information,” *Neural Computation,* vol. 10, no. 8, pp. 2047-2084, 1998.

[297] G. G. Towell, and J. W. Shavlik, “Knowledge-based artificial neural networks,” *Artificial intelligence,* vol. 70, no. 1, pp. 119-165, 1994.

[298] T. Kohonen, “The self-organizing map,” *Proceedings of the IEEE,* vol. 78, no. 9, pp. 1464-1480, 1990.

[299] N. K. Kasabov, “Evolving connectionist systems for adaptive learning and knowledge discovery: Trends and directions,” *Knowledge-Based Systems,* vol. 80, pp. 24-33, 2015.

[300] K. Ganapathy, V. Vaidehi, and J. B. Chandrasekar, “Optimum steepest descent higher level learning radial basis function network,” *Expert Systems with Applications,* vol. 42, no. 21, pp. 8064-8077, 2015.

[301] N. Kasabov, “Evolving Connectionist and Fuzzy Connectionist Systems–Theory and Applications for Adaptive On-line Intelligent Systems,” *Neuro-*

Fuzzy Techniques for Intelligent Information Processing, Kasabov N. and Kozma R.,(Eds.), Physica Verlag, 1999.

[302] F. H. Hamker, "Life-long learning cell structures—continuously learning without catastrophic interference," *Neural Networks,* vol. 14, no. 4, pp. 551-573, 2001.

[303] G. Leng, T. M. McGinnity, and G. Prasad, "An approach for on-line extraction of fuzzy rules using a self-organising fuzzy neural network," *Fuzzy sets and systems,* vol. 150, no. 2, pp. 211-243, 2005.

[304] C.-F. Juang, and C.-T. Lin, "An online self-constructing neural fuzzy inference network and its applications," *Fuzzy Systems, IEEE Transactions on,* vol. 6, no. 1, pp. 12-32, 1998.

[305] S. Wu, and M. J. Er, "Dynamic fuzzy neural networks-a novel approach to function approximation," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 30, no. 2, pp. 358-364, 2000.

[306] S. Wu, M. J. Er, and Y. Gao, "A fast approach for automatic generation of fuzzy rules by generalized dynamic fuzzy neural networks," *Fuzzy Systems, IEEE Transactions on,* vol. 9, no. 4, pp. 578-594, 2001.

[307] N. Wang, M. J. Er, and X. Meng, "A fast and accurate online self-organizing scheme for parsimonious fuzzy neural networks," *Neurocomputing,* vol. 72, no. 16, pp. 3818-3829, 2009.

[308] C.-F. Juang, and Y.-W. Tsao, "A self-evolving interval type-2 fuzzy neural network with online structure and parameter learning," *Fuzzy Systems, IEEE Transactions on,* vol. 16, no. 6, pp. 1411-1424, 2008.

[309] Y.-Y. Lin, J.-Y. Chang, N. R. Pal *et al.*, "A mutually recurrent interval type-2 neural fuzzy system (MRIT2NFS) with self-evolving structure and parameters," *Fuzzy Systems, IEEE Transactions on,* vol. 21, no. 3, pp. 492-509, 2013.

[310] N. K. Kasabov, "On-line learning, reasoning, rule extraction and aggregation in locally optimized evolving fuzzy neural networks," *Neurocomputing,* vol. 41, no. 1, pp. 25-45, 2001.

[311] M. Markou, and S. Singh, "Novelty detection: a review—part 1: statistical approaches," *Signal processing,* vol. 83, no. 12, pp. 2481-2497, 2003.

[312] M. Markou, and S. Singh, "Novelty detection: a review—part 2:: neural network based approaches," *Signal processing,* vol. 83, no. 12, pp. 2499-2521, 2003.

[313] A. R. Webb, *Statistical pattern recognition*: John Wiley & Sons, 2003.

[314] V. V. Cross, and T. A. Sudkamp, *Similarity and compatibility in fuzzy set theory: assessment and applications*: Springer Science & Business Media, 2002.

[315] D. Wu, and J. M. Mendel, “A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets,” *Information Sciences,* vol. 178, no. 2, pp. 381-402, 2008.

[316] D. Wu, and J. M. Mendel, “A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets,” *Information Sciences,* vol. 179, no. 8, pp. 1169-1192, 2009.

[317] Z. Xu, and M. Xia, “Distance and similarity measures for hesitant fuzzy sets,” *Information Sciences,* vol. 181, no. 11, pp. 2128-2138, 2011.

[318] Q. Zhang, “Nature-Inspired Multi-Objective Optimisation and Transparent Knowledge Discovery via Hierarchical FuzzyModelling,” Department of Automatic Control and Systems Engineering, The University of Sheffield, 2008.

[319] M. Jamil, and X.-S. Yang, “A literature survey of benchmark functions for global optimisation problems,” *International Journal of Mathematical Modelling and Numerical Optimisation,* vol. 4, no. 2, pp. 150-194, 2013.

[320] V. Cherkassky, D. Gehring, and F. Mulier, “Comparison of adaptive methods for function estimation from samples,” *Neural Networks, IEEE Transactions on,* vol. 7, no. 4, pp. 969-984, 1996.

[321] X.-S. Yang, Z. Cui, R. Xiao *et al.*, *Swarm intelligence and bio-inspired computation: theory and applications*: Newnes, 2013.

[322] A. Bargiela, and W. Pedrycz, "Recursive information granulation," *Granular Computing*, pp. 161-192: Springer, 2003.

APPENDIX

A.1. STEEL FRICTION STIR WELDING DATA AND PRELIMINARY MODELLING RESULTS

This section presents the process data and preliminary modelling results in Chapter 3.

<u>Steel Friction Stir Welding Data:</u>

Table A.1 presents the data set used in Chapter 3 consisting of 191 measurements, each raw represents a single experiment.

Table A.1 Process conditions of 191 weld samples used for modelling (including training and testing data).

Weld Sample	Welding Speed [mm/min]	Rotation Speed [rpm]	Spindle Peak Torque [Nm]
1	100	240	382.12
2	100	200	278.00
3	100	200	274.56
4	200	200	303.73
5	100	160	311.75
6	100	160	305.21
7	100	160	318.65
8	100	160	290.51
9	100	160	282.85
10	100	160	258.46
11	100	160	258.46
12	100	160	258.46
13	100	160	258.46
14	100	160	258.46
15	100	160	258.46
16	100	200	319.47
17	100	200	319.47
18	100	200	319.47
19	100	200	319.47
20	125	200	259.50
21	125	200	259.50
22	125	200	259.50
23	125	200	259.50
24	125	200	259.50
25	125	200	259.50
26	125	200	259.50
27	125	200	271.66
28	125	200	271.66
29	125	200	271.66

30	125	200	271.66
31	125	200	271.66
32	125	200	271.66
33	125	200	293.61
34	125	200	293.61
35	125	200	293.61
36	125	200	293.61
37	125	200	293.61
38	125	200	293.61
39	100	200	272.61
40	100	200	272.61
41	100	200	272.61
42	100	200	272.61
43	100	200	272.61
44	100	200	272.61
45	75	200	329.87
46	75	200	329.87
47	75	200	329.87
48	75	200	329.87
49	90	240	329.87
50	90	240	329.87
51	100	220	415.87
52	90	240	415.87
53	100	200	415.87
54	100	200	402.46
55	110	200	402.46
56	120	200	402.46
57	130	200	431.36
58	143	200	431.36
59	156	200	431.36
60	200	400	314.78
61	250	400	298.08
62	275	400	322.78
63	350	500	290.40
64	350	500	316.34
65	400	500	309.15
66	325	400	328.84
67	375	400	297.48
68	400	450	286.71
69	350	450	302.65
70	250	300	352.87
71	500	600	305.15
72	500	650	307.70
73	500	575	310.94
74	500	700	324.45
75	325	400	347.44
76	325	400	344.37
77	325	400	342.52
78	325	400	321.11
79	325	400	330.24
80	325	400	372.10
81	275	400	336.33

82	275	400	354.63
83	325	400	313.85
84	110	200	442.22
85	325	400	451.96
86	400	550	344.32
87	400	550	346.49
88	400	550	256.15
89	480	550	256.15
90	400	550	318.96
91	400	550	313.91
92	400	550	329.32
93	400	550	337.00
94	400	550	254.09
95	400	550	318.90
96	450	575	312.41
97	300	400	331.14
98	300	400	324.92
99	100	200	360.07
100	100	200	265.58
101	100	200	363.23
102	100	200	328.36
103	100	200	328.36
104	400	550	306.58
105	100	200	341.72
106	300	400	236.66
107	100	220	339.20
108	100	240	339.20
109	100	260	339.20
110	100	240	289.80
111	100	240	210.26
112	100	240	303.52
113	125	200	330.27
114	140	200	330.27
115	125	200	409.20
116	135	200	409.20
117	150	200	409.20
118	400	550	207.88
119	450	700	257.97
120	450	600	225.40
121	350	450	285.36
122	350	500	285.36
123	315	500	285.36
124	250	300	338.38
125	250	360	338.38
126	250	300	306.61
127	250	450	250.05
128	250	450	294.99
129	250	450	318.20
130	250	450	281.64
131	250	400	281.64
132	250	450	281.64
133	250	450	281.32

134	250	450	291.09
135	250	450	276.57
136	300	450	276.57
137	325	450	219.56
138	300	450	288.58
139	250	450	214.87
140	500	700	242.71
141	100	200	360.78
142	100	200	326.74
143	100	200	384.52
144	100	200	352.34
145	100	200	353.48
146	100	200	346.46
147	100	200	339.48
148	100	200	344.29
149	100	200	362.73
150	100	200	296.74
151	100	200	317.71
152	100	200	311.84
153	100	200	278.23
154	325	400	361.04
155	325	400	302.63
156	325	400	380.52
157	325	400	209.19
158	400	450	289.46
159	400	450	246.13
160	400	450	248.51
161	400	450	275.62
162	400	450	275.51
163	400	450	256.13
164	400	450	257.82
165	400	450	277.87
166	400	450	222.37
167	130	200	395.24
168	143	200	395.24
169	156	200	395.24
170	130	200	307.38
171	143	200	307.38
172	156	200	307.38
173	100	200	327.08
174	100	220	356.18
175	100	240	368.82
176	150	240	365.53
177	150	260	297.94
178	150	300	284.35
179	300	500	342.36
180	350	500	292.09
181	350	500	372.41
182	350	500	321.11
183	400	550	322.31
184	400	600	322.31
185	350	550	301.64

186	315	550	301.64
187	245	550	301.64
188	300	500	332.98
189	300	250	332.98
190	300	300	332.98
191	300	280	293.75

Multiple Regression Linear Model results:

The performance indices based on the RMSE and MAE% of the MRL model are shown in Table A.2. The simulation results obtained by using the MRL model are depicted in Fig. A.1. It is clear that the MRL model provides only a basic level of predication performance and therefore more advanced nonlinear data-driven CI models are needed to model the nonlinear relationship between the process parameters and the internal process variables.

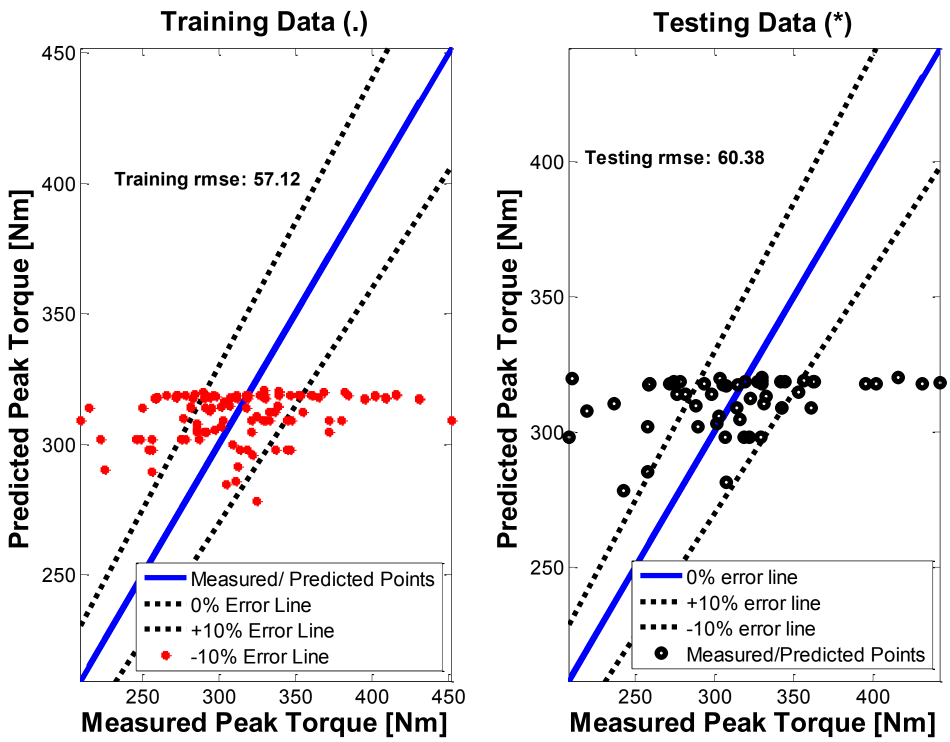


Figure A.1. Data fit, peak torque prediction by using multiple linear regression model.

Table A.2 Performance of the multiple linear regression model for spindle peak torque.

Performance Index	**MLR Model**	
	Training	Testing
RMSE±SD	57.12±23.01	60.38±30.19
MAE %±SD	30.94±21.67	34.85±29.92

A.2. STEEL FRICTION STIR WELDING DATA FOR PERPETUAL LEARNING FRAMEWORK

The initial data set (old data set) used in Chapter 6 to construct the initial model and test its performance is shown in Table A.1, each raw represents a single experiment.

Table A.3 Process conditions of 36 weld samples used for constructing the initial model (including training and testing data).

Weld Sample	**Welding Speed [mm/min]**	**Rotation Speed [rpm]**	**Spindle Peak Torque [Nm]**
1	325	450	219.56
2	140	200	330.27
3	130	200	431.36
4	250	450	281.64
5	300	400	331.14
6	350	550	301.64
7	120	200	402.46
8	200	400	314.78
9	300	300	332.98
10	200	200	303.73
11	350	500	290.40
12	250	300	352.87
13	90	180	329.87
14	100	160	311.75
15	150	200	409.20
16	315	500	285.36
17	156	200	431.36
18	100	260	339.20
19	100	240	382.12
20	275	400	322.78
21	100	220	415.87
22	75	240	329.87
23	90	240	329.87
24	300	500	342.36
25	300	450	276.57
26	250	400	298.08

27	250	360	338.38
28	75	200	329.87
29	125	200	259.50
30	315	550	301.64
31	375	400	297.48
32	300	250	332.98
33	135	200	409.20
34	325	400	347.44
35	350	450	302.65
36	300	280	293.75

Table A.1 presents the new data set used in Chapter 6 to test the prediction performance of the incremental learning framework.

Table A.4 Process conditions of 19 weld samples used for perpetual learning.

Weld Sample	Welding Speed [mm/min]	Rotation Speed [rpm]	Spindle Peak Torque [Nm]
1	**400**	**500**	**309.15**
2	400	450	286.71
3	500	600	305.15
4	500	650	307.70
5	500	575	310.94
6	500	700	324.45
7	400	550	344.32
8	480	550	256.15
9	450	575	312.41
10	450	700	257.97
11	450	600	225.40
12	400	600	322.31
13	110	200	442.22
14	143	200	431.36
15	245	550	301.64
16	150	300	284.35
17	150	240	365.53
18	100	200	278.00
19	150	260	297.94