

**A HYBRID NEURO-WAVELET APPROACH TO
ELECTRIC ARC FURNACE MODELING**

by

Shadan Ghaffaripour

Bachelor of Science in Computer Engineering, Azad Tehran University, Iran, 2010

A thesis

presented to Ryerson University

In partial fulfillment of the
requirements for the degree of

Master of Science

In the program of

Computer Science

Toronto, Ontario, Canada, 2014

© Shadan Ghaffaripour 2014

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

A HYBRID NEURO-WAVELET APPROACH TO ELECTRIC ARC FURNACE MODELING

Shadan Ghaffaripour

M. Sc. in Computer Science, 2014

Ryerson University, Toronto, Canada

Abstract

This thesis proposes a hybrid neuro-wavelet based approach for modeling the dynamic voltage-current characteristics in electrical arc furnaces. This method uses the data obtained from an operational electrical arc furnace exclusively to describe the underlying process, and unlike conventional mathematical techniques it does not rely on presumed model structures or simplified assumptions. A comparison between the results that proceeded from the proposed method and the actual measurements has been made. The proposed method is demonstrated to be capable of modeling the EAF's dynamic voltage-current behaviour accurately.

Acknowledgements

First and foremost, I would like to express my special appreciation to my supervisor, Professor Dr. Alireza Sadeghian, who has supported me generously throughout my graduate studies with his patience, understanding and extensive knowledge. I especially want to thank him for broadening my understanding of research work and for the pleasant learning environment he provided. One simply could not wish for a more supportive supervisor or a better research and learning experience.

I would also like to thank my thesis committee members, Prof. Issac Woungang, Prof. Eric Harley and Prof. Cherie Ding, who have generously given their time and knowledge to improve my thesis document.

Lastly, I would also like to sincerely thank my family, my parents and my sister Taban, for all their unconditional love and support, and all my friends, in particular my best friend Sina, for their tremendous support, encouragement and understanding during the hard times.

Dedication

To my lovely parents, Reza and Tahereh for their endless love and support ...

Table of Contents

List of Tables	ix
List of Figures.....	x
List of Abbreviations	xiii
1 Introduction	1
1.1 Electric Arc Furnaces.....	1
1.2 Motivation.....	3
1.3 Objective	3
1.4 The Modeling Challenge.....	4
1.5 Existing Approaches	4
1.6 Wavelet Transform	8
1.7 The Proposed Approach.....	9
1.8 Thesis Outline	11
2 Wavelet Transform.....	13
2.1 Wavelet Transform vs. Fourier and Short time Fourier Transforms	13
2.2 Wavelet	14
2.3 Wavelet expansion	15
2.4 Continuous Wavelet Transform.....	17
2.5 Multi-Resolution Analysis	19
2.5.1 Multi-Resolution Formulation	19
A. The Scaling Function	20
B. The Wavelet Function	22
2.6 The Discrete Wavelet Transform.....	26
2.6.1 DWT Calculation, from the Signal Processing View	27
A. Analysis - From High Resolution to Low Resolution.....	28
B. Synthesis – From Low Resolution to High Resolution.....	31
3 Electric Arc Furnace Modeling.....	32
3.1 Mathematical methods for electrical arc furnace modeling.....	32
3.1.1 White-Box Models.....	33
3.1.2 Black-Box Models	36

A. Single Scale Black-Box Methods	36
B. Multi-Scale Black-Box Methods.....	40
4 Methodology	46
4.1 Data Acquisition	48
4.2 Preprocessing Methods	48
4.2.1 Wavelet Transform	48
4.2.2 The Non-Decimated Discrete Wavelet Decomposition.....	53
A. The MODWT Wavelet Transform.....	55
4.2.3 Time series Alignment.....	61
4.3 Feature Extraction Method	62
4.4 Hybrid Neuro-Wavelet Modeling Method.....	65
4.4.1 Information Criteria	70
5 Experimental Results	72
5.1 Parameter selection	75
5.2 Results and Discussion	76
5.2.1 The First Experiment	77
5.2.2 The Second Experiment.....	89
5.2.3 The Third Experiment.....	93
5.2.4 The Fourth Experiment	95
5.3 Comparisons to previous work	101
6 Conclusions and Contributions	103
6.1 Summary	103
6.2 Conclusions.....	103
6.3 Contributions.....	105
6.4 Future Work Directions	105

List of Tables

Table 5.1. Properties of the Haar wavelet.....	77
Table 5.2. The results of Neuro-Wavelet models at different decomposition levels.....	85
Table 5.3. The results of neuro-wavelet models under the first and second schemes	93
Table 5.4. The effect of the current decomposition levels on performance	94
Table 5.5. Properties of the Coiflet wavelets	96
Table 5.6. Properties of the Symlet wavelets.....	96
Table 5.7. Number of features extracted from each resolution level, using three different wavelets.....	97
Table 5.8. Comparison of performance of the models created by different wavelets	97
Table 5.9. The results of neuro-wavelet models at different decomposition levels, based on the Coiflet wavelet	97
Table 5.10. The results of neuro-wavelet models at different decomposition levels, based on the Symlet wavelet.....	98

List of Figures

Fig. 1.1: Electric Arc Furnace.....	2
Fig. 2.1: Scaling Function and Wavelet Space	23
Fig. 2.2. Two-Band Analysis Bank.....	30
Fig. 2.3. Three stage two-band analysis tree.....	30
Fig. 2.4. Two Band Synthesis Bank.....	32
Fig. 4.1. The proposed modeling algorithm.....	47
Fig. 4.2. The arc voltage signal.....	51
Fig. 4.3: The arc current signal	51
Fig. 4.4: A close-up view of the voltage signal	52
Fig. 4.5: A close-up view of the current signal.....	52
Fig. 4.6. The Least Asymmetric wavelet and scaling function filters	58
Fig. 4.7. The Best-Localized wavelet and scaling function filters.....	59
Fig. 4.8. The Coiflet wavelet and scaling function filters.....	60
Fig. 4.9. Features used in modeling v-i relationship in the first scheme	64
Fig. 4.10. Features used in modeling v-i relationship in the second scheme.....	65
Fig. 4.11. The proposed neural network structure for modeling v-i relationship in the first scheme	67
Fig. 4.12. The proposed neural network structure for modeling v-i relationship in the second scheme.....	67
Fig. 5.1. The first feature set	73
Fig. 5.2. The second feature set	74
Fig. 5.3. The Haar wavelet.....	77

Fig. 5.4. The decomposed voltage signal into five levels of resolution.....	78
Fig. 5.5. The decomposed voltage signal into four levels of resolution	79
Fig. 5.6. The decomposed voltage signal into three levels of resolution.....	79
Fig. 5.7. The decomposed voltage signal into two levels of resolution.....	80
Fig. 5.8. The decomposed voltage signal into 1 level of resolution	80
Fig. 5.9. The BIC value of the models created with different orders	82
Fig. 5.10. Comparison of train and test errors having different number of neurons	84
Fig. 5.11. Over-fitting occurring during the training process	84
Fig. 5.12. Comparison of Neuro-Wavelet models at different decomposition levels.....	86
Fig. 5.13. The actual measured voltage vs. the estimated voltage.....	86
Fig. 5.14. The neuro-wavelet model response to the test data-case1:decomposition level=1	87
Fig. 5.15. The actual (top) and estimated (bottom) EAF ($v-i$) characteristic.....	88
Fig. 5.16. The neuro-wavelet model response to the test data-case 2: decomposition level=2	88
Fig. 5.17. The neuro-wavelet model response to the test data-case 3: decomposition level=5	89
Fig. 5.18. The decomposed current signal into five levels of resolution	90
Fig. 5.19. The decomposed current signal into four levels of resolution.....	90
Fig. 5.20. The decomposed current signal into three levels of resolution	91
Fig. 5.21. Decomposed current signal into two levels of resolution.....	91
Fig. 5.22. Decomposed current signal into 1 level of resolution	92
Fig. 5.23. Comparison of neuro-wavelet models under the first and second schemes	93
Fig. 5.24. The effect of the current decomposition level on performance.....	94
Fig. 5.25. The first member of the Coiflet family.....	95
Fig. 5.26. The first member of the Symlet family	96

Fig. 5.27. Comparison of performance of the models created by different wavelets	99
Fig. 5.28. Comparison of size of the models created by different wavelets	100

List of Abbreviations

AI	Artificial Intelligence
AIC	Akaike Information Criterion
AICc	Akaike Information Criterion Corrected
AFLS	Adaptive Fuzzy Logic Systems
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
AR	Auto Regression
ARfIMA	Auto Regressive <i>fractionally</i> Integrated Moving Average
BIC	Bayesian Information Criterion
CSI	Common Signal Index
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
EAF	Electric Arc Furnace
FFT	Fast Fourier Transform
FLS	Fuzzy Logic System
LS-SVM	Least Square-Support Vector Machine
MLP	Multi-Layer Perceptron
MODWT	Maximal Overlap Discrete Wavelet Transform
MAR	Multi-Resolution Analysis
NDIE	Non-Dimensional Index Error
PACF	Partial Auto Correlation Function
RBF	Radial Basis Function
SVM	Support Vector Machine
WNN	Wavelet Neural Network

1 Introduction

1.1 Electric Arc Furnaces

The importance of the steel industry to the Canadian economy and its contribution to society are indisputable. According to a research study [1] conducted by the University of Toronto, the steel industry employs directly and indirectly 130,000 men and women across Canada. It produces multi-purpose material that is essential to many key industries, including transportation and physical infrastructure. In addition, it provides \$7 billion per year in exports and plays a major role in the energy and environmental future of Canada. As stated in [1], “if we didn't have the steel industry we have, we wouldn't have the industry or society that we have today.”

Electrical Arc Furnaces (EAF) take an essential role in the Canadian steel industry and account for about half of North American steel production [1]. Broadly, an arc furnace transfers electrical energy to thermal energy to melt the metal scrap by means of a high power electric arc, which is set up between electrodes and the molten bath containing the raw material [2] (see Fig. 1.1). The meltdown process is then followed by a refinement stage, which is the main purpose of EAF operation.

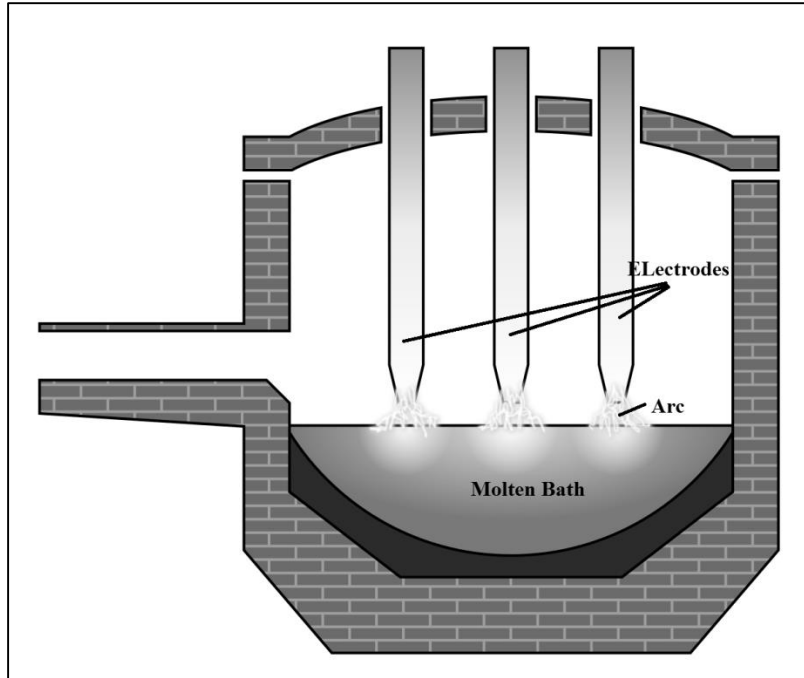


Fig. 1.1: Electric Arc Furnace

During the meltdown process, the vertical movement of the high power electrodes is regulated in order to keep a stable arc in spite of random movement of the melting material. This makes the arc furnace draw both high and varying power from the supply system. As a result, the arc voltage and current waveforms deviate significantly from a perfect sinusoidal shape and no two cycles of them are identical. This type of load, which is highly nonlinear and dynamic in nature, has many adverse impacts on the power quality of the joint network. With a power with such quality, other electrical loads connected to the network may malfunction, fail prematurely or not operate at all. All of that causes the whole electrical system to not function in its intended manner [3]. In particular, voltage flicker and harmonic injection are the two major power quality problems raised in such electrical systems as a result of EAF operation [4]. Flicker phenomena refer to rapid visible changes of light level in lighting equipment caused by voltage fluctuations. Harmonics, on the other hand, refer to variations in the wave shape of the voltage and current signals, all of which are undesirable. The

injected harmonics may propagate into the network and affect other electrical loads adversely. For example, they may result in extra heating in the equipment and conductors, misbehaviour in variable frequency drives, and added torque pulsations in motors [5].

1.2 Motivation

As noted earlier, the operation of EAFs reduces the power quality of the system to which it is connected. However, as the operation of EAF is crucially necessary, remedial solutions should be adopted in order to maintain the quality of power in the interconnected network. In other words, the adverse effects of EAF operation should be compensated appropriately. The most common approach is the installation of power compensators that can dynamically compensate the power consumption of EAFs [6]. However, for this to become feasible, the behaviour of EAFs in their electrical network has to be studied and then modeled as accurately as possible. For the purpose of modeling this behaviour, the EAF voltage-current relationships, referred to as $(v-i)$ characteristics hereafter, are usually determined.

This model can then be used for power quality penetration studies and mitigation designs before any EAF installations or upgrades in the system. As a result, unexpected situations that can impose intolerable financial burden are avoided.

1.3 Objective

The purpose of this study is to investigate new intelligent data-driven modeling techniques, based on time-frequency domain analysis, and to build accurate and reliable model for EAF operation based on such techniques. This model would be capable of being used in power quality penetration studies such as flicker compensation or harmonic injection cancellation. The primary goal of this

study is to justify the application of wavelet transform in the proposed framework and, therefore, explicit attention has been paid to this theory.

1.4 The Modeling Challenge

It is important to realize that if we had an ideal electric arc furnace with linear ($v-i$) characteristics, modeling would be a less challenging task. However, in reality, because EAF load is time variant and highly nonlinear with complex dynamics, the ($v-i$) curve is anything but linear. Hence, producing accurate models of EAFs, based on conventional mathematical techniques and explicit equations that are simplified for such a problem, may not be feasible.

An alternative solution to this problem is the use of black-box and non-parametric techniques that have been recently widely adopted. These techniques do not presume a structure for the system and instead, concentrate on models that describe the behaviour of the system in detail, entirely based on input and output data [4]. In this study, we are employing a variant of black-box methods, in which artificial intelligent and multi-scale time-frequency techniques are combined to model the behaviour of an electric arc furnace accurately.

1.5 Existing Approaches

Early EAF models were predominantly based on explicit mathematical equations, simplified or questionable assumptions and experimental findings, which could not accurately capture the dynamics of such a complicated system [4]. As an example, the conventional system identification techniques have been adopted in many studies such as [7] and [8]. However, they could not achieve accurate results as they presumed the model structure was known, and all that needed to be identified were the corresponding parameters based on some knowledge of the system. In a similar

manner, many time series analysis methods have used differential equations to describe the arc furnace systems. The accuracy of these methods is, however, questioned [9]. Time domain equivalent circuit methods that design equivalent circuits consisting of the voltage source and resistors have also simplified the voltage-current characteristics that affect the accuracy significantly [10]. The frequency domain analysis techniques suffer from quite similar defects. For example in [11], the harmonic voltage source method, which applied the Fourier transform to the arc voltage signal, determined the parameters of the equivalent circuit, using the parameters of the supply system. This is however done by assuming that the power transfer has been made at its maximum capacity. This assumption cannot reflect the arc furnace operating condition and, therefore, constituted the main source of error. Another method is “domain solution of nonlinear differential equation” presented in [12]. This model was developed from energy balance equations, which are nonlinear differential equations of arc radius and arc current. These equations use some parameters, obtained experimentally, to describe the operation of arc furnace.

These arguments suggest that conventional techniques are in essence, not capable of addressing issues such as modeling complex systems. In this respect, black-box and data-driven methods, that fall under the broad area of artificial intelligence (AI), can act as substitutes for such techniques. Over the past few decades, the widespread interest has been in artificial intelligence among researchers, and a mass of literature is devoted to the practice and application of such techniques in non-linear modeling problems. These techniques provide powerful tools for finding solutions to a variety of practical problems that other more traditional methods often fail to solve. Artificial neural networks, fuzzy systems, adaptive fuzzy neural networks and genetic algorithm inspired algorithms are considered to be the most common of such techniques. The rest of this section is

designated to a brief review of major research work regarding EAF modeling in each of these subareas.

Many studies such as [4] and [13] applied the notion of Artificial Neural Networks (ANN) to model the dynamic and highly nonlinear EAF systems, and the computational results obtained from these works were strictly comparable to the existing measurements. In fact, there is extensive evidence for justifying the artificial neural networks' value for the general case of nonlinear modeling. The most important one is their universal approximation capability that allows for estimation of any real continuous function to any specified level of accuracy [4]. This proven capability is indeed a result of their adaptive learning capabilities. In addition, the superiority of neural networks for modeling such complex systems is their use of available data exclusively [4], for describing the system under study. Consequently, they do not need prior knowledge of the underlying process. However, since ANNs are only designed to work with numerical data, it is impossible to use qualitative and linguistic information from the experts. In addition, the input and output mapping of a trained neural network cannot be interpreted into meaningful predictions rules. In other words, neural networks lack the capability of knowledge representation. In fact, this is the underlying reason behind naming these methods as "black-box."

Fuzzy logic, on the other hand, has the capability to deal with linguistic knowledge, as well as numerical data [14]. The fuzzy logic strength especially lies in linguistic knowledge handling, which is burdensome to quantify, according to what traditional mathematics permits [15]. Therefore, they provide alternatives for a system that is too complicated to be described with equations, and hence, the mathematical model is either ill-defined or does not exist at all [15]; this is the case with EAF operation. The distinguishing characteristics of Fuzzy Logic Systems (FLS), namely the universal approximation ability and functional equivalence with Radial Basis

Functions (RFB) [15], give additional support to the effectiveness of such systems to highly nonlinear problems such as EAF modeling. As a final point, this rigorous mathematical discipline is capable of expressing nonlinear relationship between input and output, by a collection of qualitative if-then rules [14], which is a major advantage over ANNs. However, FLSs lack an effective learning capability and furthermore, the rule definition depends on intuitive experience of experts. Another disadvantage is that the *a-priori* or heuristic knowledge is needed in order to trim the vast number of rules in the rule base. In addition, the distribution of data in input-output space and the natural grouping of the data cannot be considered [15]. To compensate for these shortcomings, the fuzzy logic systems can be augmented with neural networks. This arrangement, which has been presented in [15-20], combines the precision and adaptability of neural networks along with the fast training and generalization ability of fuzzy logic systems in one single package [20].

Genetic Algorithms (GA) are another area in AI, which have been used to enhance models in many different disciplines. Although still not applied directly to the arc furnace modeling problem, diverse areas in power systems such as load modeling have profited from these methods [21, 22]. The popularity of GAs is due to their practical and robust optimization and search capability. As a result of their robustness in complex parameter search spaces, they are regularly used together with other artificial intelligence techniques such as neural networks and fuzzy logic to optimize their parameters. These methods have a high probability of locating the global solution optimally in the search space with several local minima [23]. However, they are computationally intensive, a property that is not desirable.

1.6 Wavelet Transform

The wavelet transform is an analysis method, which differs from the well-known Fourier transform by employing short waves instead of long sine waves as the analysis function or basis [24]. Wavelet analysis can shed light to attractive aspects of data that is ignored by other analysis techniques [24]. Broadly, a wavelet is a small wave that oscillates in a limited time and satisfies certain mathematical criteria (see section 2.1). The transform that uses such wavelets, transforms data from the original time domain to a joint time-frequency domain, by an expansion in orthonormal bases, generated by dilation (or contraction) and translation of wavelets [24].

This transformation is a powerful mathematical tool that decomposes data into various frequency bands and then examines each band, with a resolution appropriate for its scale [11]. Furthermore, it preserves both time and frequency information in data. However, the Heisenberg uncertainty imposes restrictions on the simultaneous resolution of this time and frequency information. The bottom line is that the wavelet transform can potentially provide a useful and highly informative mathematical representation of any existing phenomena in the world, including EAF operation indeed.

Wavelets have many attractive properties, by which many data analysis tasks including regression and modeling are solved effectively. The most important properties include: compact support [25], which guarantees a localized transform and is especially useful for signals with local irregularities in either domain; vanishing moments, which results in a transform that can distinguish between essential and nonessential information [25], leading to dimension reduction; hierarchical and multi-resolution structure; linear time and space complexity of transformation [26]; smoothness [25]; ability to act as a generator of orthonormal bases of function space L^2 [25], and possession of various basis functions [27].

In addition, wavelets transform leads to de-correlation in coefficients, that is the wavelet coefficients are less correlated than original data points [25, 28]. This implies that this transform has the potential of reducing the complexity of a problem in the time domain and producing a simpler problem in the wavelet domain [25].

Furthermore, wavelet transform facilitates performing operations at different resolutions and also localizes the operation in both time and frequency domains. According to [25], manipulating the wavelet transformed coefficients and then constructing the results is more efficient than working on the original domain data points in the first place. This is only possible because of another important property of wavelets, namely reversibility. Being reversible in this context means that the representation in the original domain, usually time domain, is exactly equivalent to the representation in the wavelet domain (under certain circumstances), and it is possible to go back and forth between the two domains, without any loss of information.

The above mentioned properties make wavelet transform a good candidate for studying and analysing a complex system such as EAF. As we proceed to the next chapters, the usefulness of each property will become clearer for the particular case of EAF data.

1.7 The Proposed Approach

This research builds on existing knowledge in the fields of artificial intelligence, soft computing and digital signal processing and in particular, proposes a method to model the complex dynamic behaviour of electric arc furnaces by a thoughtful integration of artificial neural networks and wavelet transformation.

Briefly, wavelet transform is firstly applied to the electric arc voltage and current signals for the extraction of essential features. These features are then fed into a neural network with the intention of learning of the underlying process. Integrating the rigorous theory of wavelets, with the adaptability of neural networks in learning, has many advantages, among the most important, are higher chances of generalization and faster convergence [25].

In a more detailed view, the proposed methodology for modeling the (v - i) characteristics in electric arc furnaces is built on the approach taken in [4] and supplemented with the suggestions in [29]. In [4] a functional mapping is found between the electric arc voltage and current by means of artificial neural networks. The rate of change of arc current in time (derivative of current with respect to time) is also considered in the inputs along with previous values of the electric arc voltage. The work in [29] is related to a multi-scale linear approach to time series prediction, based on some coefficients of the wavelet transform of the previous values. The transform results in a decomposition of the time series into frequency bands of different scales. The prediction is then made using few coefficients on each of these scales.

In our work, the same functional mapping as in [4] has been found. However, with the incorporation of the feature extraction method explained in [29], the need for including the arc current derivative is eliminated. In [29], a fixed number of coefficients are selected at each resolution level (or scale). However, in our approach, the number of coefficients to be selected at each resolution level is determined based on statistical methods. These methods include “Akaike Information Criterion” (AIC), “Akaike Information Criterion Corrected” (AICc) and “Bayesian Information Criterion” (BIC), which are suggested by the authors of [29] and also, “Partial Auto Correlation Function” (PACF). Besides, we used a variant of wavelet transform, namely Maximal

Overlap Discrete Wavelet Transform (MODWT). This algorithm has been explained in [28] in great detail.

A two-layer feed-forward neural network is then used to model the electrical arc voltage in terms of electric arc current, based on the extracted features. The weight and biases of this network are determined by a variant of the recognized back-propagation learning algorithm. This neural network is capable of generalizing through the use of wavelet transformed coefficients, instead of the original data.

To the best of our knowledge, this compound method has never been applied to model the behavior of electric arc furnaces, in despite the substantial benefits that wavelet transform can bring. The only work that employs wavelet transform in this area is [6]. However, wavelet transform is only used for distinguishing EAF operation stages and not for modeling purpose.

1.8 Thesis Outline

The remainder of this thesis is organized as follows: Chapter 2 is devoted to the establishment of wavelet theory and multi-resolution. It also explains the notion of wavelets and their main properties. Moreover, the most common wavelet transforms, namely continuous and decimated discrete wavelet transforms are formulated. Chapter 3 provides background information on previous research work, conducted in modeling EAFs. These works fall under two broad categories of black-box and white-box models. The incompetence and inadequacy of white-box models for the application of the EAF modeling is justified, and the rationale behind focusing on black-box models, especially those that are also multi-scaled is given. Chapter 4 elaborates on the proposed solution for modeling the time-varying ($v-i$) characteristics of EAFs. In Chapter 5, the summaries

of simulation results are presented and discussed. The conclusions and recommendations for future work are then given in Chapter 6.

2 Wavelet Transform

2.1 Wavelet Transform vs. Fourier and Short time Fourier Transforms

Most signals have characteristics that change in both time and frequency domains [30]. However, the well-known Fourier transform can only represent the frequency content of the signal globally [24], over the signal's existence time, and it does not provide any information about the signal's spectral changes during this period. In other words, the time information is lost and, therefore, it is not a satisfactory analysis tool for non-stationary signals. To avoid this restriction on locality in time, another transformation, namely Short Time Fourier Transform (STFT) is introduced. This transform is acquired by simply applying the Fourier transform to consecutive parts of the signal, by employing a sliding window of limited size [30].

STFT suffers from a major drawback, that is, once a particular window is selected, the time and frequency resolutions become invariable during the whole analysis procedure [30]. In other words, STFT analyzes components of a signal, which are a mixture of time and frequency information with a predefined frequency and time resolution that is not changeable as a result of the fixed window length. This implies that the analysis of low-frequency and high-frequency components of signals are both performed using the same windows. In view of the fact that in practice, high-resolution low-frequency analysis is made using long windows in the time domain and short windows in the frequency domain (and vice-versa for high-frequency analysis) [30], this transform is by no means efficient. The wavelet transform, on the other hand, uses variable size analysis windows. In this setting, long time domain and short frequency domain windows is used for the analysis of low-frequency components of the signal and vice-versa for the analysis of high-frequency components of that signal [30]. Therefore, what wavelet analysis provides, is a better

trade-off between time resolution and frequency resolution [30] that is imposed by Heisenberg uncertainty principle. According to this principle, it is not possible to examine a part of a signal with unlimited frequency and time resolution simultaneously.

2.2 Wavelet

As suggested by the name, a wavelet is a little wave. It oscillates up and down just like any other wave. However, it has its energy concentrated in time [27]. In other words, it rises and falls only within a short period, hence the name. Percival and Walden in [28] quantified the notion of a wavelet as a real-valued function, $\psi(\cdot)$, that is defined over the real axis, and satisfies two essential properties of (2.1) and (2.2):

The integral of $\psi(\cdot)$ being zero:

$$\int_{-\infty}^{\infty} \psi(u) du = 0 \quad (2.1)$$

and the square of $\psi(\cdot)$ integrating to unity:

$$\int_{-\infty}^{\infty} \psi^2(u) du = 1 \quad (2.2)$$

Should (2.2) hold, then for any ε , which is between zero and one ($0 < \varepsilon < 1$), an interval $[-T, T]$ of limited size exists [28] so that:

$$\int_{-T}^T \psi^2(u) du > 1 - \varepsilon \quad (2.3)$$

This equation suggests that $\psi(\cdot)$ can only deviate slightly from zero outside of $[-T, T]$, on condition that ε is very close to zero. Considering that the length of this interval is vanishingly small in comparison with the unlimited size of the whole real axis, the non-zero activity of $\psi(\cdot)$ should be

very limited and small [28]. Therefore, while (2.2) implies that $\psi(\cdot)$ has to rise above or fall below zero in some parts, (2.1) enforces any part that is above zero to be cancelled out by those below zero. Consequently, $\psi(\cdot)$ must look like a little wave [28].

In summary, a wavelet can be regarded as any function that integrates to zero and is square integrable, by a very broad definition [28]. However, it is necessary to impose conditions beyond (2.1) and (2.2) for practical use of wavelets. An extensive mathematical research has been done in order to determine what conditions are required to yield particular type of analysis with wavelets [28].

2.3 Wavelet expansion

By expressing a function $f(t)$, as a linear decomposition (see (2.4)), more comprehensive analysis, description and processing of the signal is possible [27]:

$$f(t) = \sum_l a_l \psi_l(t) \quad (2.4)$$

Where a_l is the expansion coefficient and $\psi_l(t)$ are a collection of real-valued functions named as expansion set. For a wavelet expansion, an arrangement with two parameters is made such that (2.4) changes to

$$f(t) = \sum_k \sum_j a_{j,k} \psi_{j,k}(t) \quad (2.5)$$

where the two parameters of j and k are integer and $\psi_{j,k}(t)$ are the wavelet expansion sets that are usually orthogonal functions. This two-dimensional representation leads to localization of the signal in both time and frequency [27].

The wavelet expansion set is by no means unique, and there are numerous wavelets that can be used effectively. But, they all share some general characteristics, according to [27]. To begin with, all wavelet bases are collection of building blocks to form or present a signal or function. More importantly, wavelet expansions localize the signal in both time and frequency. This suggests that the majority of the signal energy can be represented appropriately, by only a few coefficients in the expansion. Another significant trait of wavelet analysis is that it is especially appropriate for transient signals. This is because the localizing property of wavelets enables modeling of temporal event with only a few coefficients.

Wavelet bases are all generated from a single wavelet, referred to as the mother wavelet, by simple scaling (contraction or dilation) and translation (or shifting) operations [27]. The two-dimensional parameterization is obtained from the function $\psi(t)$ (usually called the mother wavelet), by (2.6).

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k) \quad j, k \in \mathbb{Z} \quad (2.6)$$

The change of parameter k , moves the wavelet along the time axis and enables localization of the events in time. In contrast, the change of parameter j , changes the shape of the wavelet by either contraction or dilation to modify its scale and allows representation of a particular resolution [27].

For the multi-resolution formulation, a fundamental concept that will be explained later, two basic functions are required, each of which is closely related to the other. That is the wavelet functions are supplemented with another set of basic functions named scaling functions, $\varphi(t)$. Using a combination of both functions, a wide variety of signals can be expressed as (2.7) [27].

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \varphi(t - k) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_{j,k} \psi(2^j t - k) \quad (2.7)$$

Wavelet expansions are undoubtedly very efficient and effective in analyzing a variety of signals and phenomena in the world. The properties that give this effectiveness are as follows according to [27]: Firstly, the wavelet expansion coefficients drop quickly from large to negligible amounts with j and k , for most of the signals. As a result of this, wavelets have successful applications in signal and image compression, de-noising and detection. Secondly, with the wavelet expansion, signal characteristics can be described locally with high precision. They can also be separated accurately. What's more, a wavelet expansion coefficient is local in essence, and it is easier to be interpreted. As a result, the wavelet expansion is capable of separating components of a signal in both time and frequency that overlap with one another. Lastly, wavelets are adaptable and adjust themselves to conform to the signal. Therefore, they can be designed in a particular way to meet individual applications' needs. It is worthwhile mentioning that some of these properties are, in fact, same things that are viewed from different perspectives.

2.4 Continuous Wavelet Transform

The Continuous Wavelet Transform (CWT) is particularly useful for transforming a function or signal $x(\cdot)$ that is defined over continuous time [31]. The calculation of the continuous wavelet transform is a very straightforward process. Intuitively, we calculate how closely correlated the scaled and translated wavelets are with different parts of the signal; the more similarity, the higher the value of the coefficient. In this transform, both parameters λ (scaling parameter) and t (translation parameter) change continuously. Note that λ and t are the same as j and k in the context of discrete wavelet transform that had been briefly explained in the previous section.

Mathematically, CWT is the integral over infinite time of the signal multiplied by scaled, shifted (or translated) mother wavelet. CWT generates wavelet coefficients as a function of scale and time [31]:

$$C_{\psi} = \langle x, \psi_{\lambda,t} \rangle = \int_{-\infty}^{\infty} \psi_{\lambda,t}(u) x(u) du \quad (2.8)$$

where $\psi_{\lambda,t}(u)$ is a modification to the mother wavelet $\psi(u)$, and is formulated as:

$$\psi_{\lambda,t}(u) = \frac{1}{\sqrt{\lambda}} \psi\left(\frac{u-t}{\lambda}\right) \quad (2.9)$$

In this equation, $\lambda \in \mathbb{R}^+$ and $t \in \mathbb{R}$ are the scaling and translation parameters respectively. By varying λ , a picture of how the wavelet function fits the signal from one scale to another, can be built up. By varying t , how the nature of the signal changes over time can be seen [31]. The collection of coefficients $\{\langle x, \psi_{\lambda,t} \rangle | \lambda > 0, -\infty < t < \infty\}$ is called the CWT of $x(\cdot)$.

Interestingly, the CWT preserves all the information in the original signal, provided that the wavelet function $\psi(\cdot)$ satisfies the admissibility condition [28], which is fulfilled by the following constraint:

$$\Psi(0) = \int_{-\infty}^{+\infty} \psi(t) dt = 0 \quad (2.10)$$

Where $\Psi(\cdot)$ denotes the Fourier transform of $\psi(\cdot)$. According to this condition, the wavelet does not have zero frequency components.

If the signal $x(\cdot)$ also satisfies:

$$\int_{-\infty}^{\infty} x^2(t) dt < \infty \quad (2.11)$$

Then, $x(\cdot)$ can be perfectly reconstructed from its CWT, by the following inverse transform [28]:

$$x(t) = \frac{1}{C_\psi} \int_0^\infty \left[\int_{-\infty}^\infty \langle x, \psi_{\lambda,u} \rangle \psi_{\lambda,u} du \right] \frac{d\lambda}{\lambda^2} \quad (2.12)$$

As a matter of fact, the signal $x(\cdot)$ and its CWT coefficients are different ways of representing the same mathematical entity. However, the CWT presents the signal in a new way, which enables gaining additional insight into the signal [28].

2.5 Multi-Resolution Analysis

Multi-Resolution Analysis (MRA) is an imperative concept in wavelet systems. By using this concept, it is possible to analyze various frequency components of a signal with different resolutions [32]. The overall idea in MRA is that by having a sequence of embedded subspaces, capable of approximating $L^2(R)$, one can focus on particular subspaces for a specific application.

For employing multi-resolution concept, two sets of functions, namely scaling function and wavelet function need to be defined.

2.5.1 Multi-Resolution Formulation

A multi-resolution analysis is a sequence of subspaces belonging to $L^2(R)$, and satisfying three conditions [33]. The first condition is containment, which requires a nesting of the spanned spaces as:

$$0 \subset \cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \subset L^2 \quad (2.13)$$

or:

$$V_j \subset V_{j+1} \quad \text{for all } j \in \mathbb{Z} \quad (2.14)$$

which indicates that every subspace ν_j is contained in the subsequent wider subspace ν_{j+1} . This implies that the space that contains high-resolution signals accommodates lower resolution signals as well [27]. The next condition imposes the following to be true:

$$\nu_{-\infty} = \{0\}, \quad \nu_{\infty} = L^2 \quad (2.15)$$

which indicates that at zero resolutions, where $j = -\infty$, only a zero finite energy signal exists. Whereas, at infinite resolution where $j = +\infty$, a perfect reproduction of the signal can be obtained [27]. Another condition to be satisfied is scaling, which indicates that any subspace has to meet the reasonable scaling condition of:

$$f(t) \in \nu_j \Leftrightarrow f(2t) \in \nu_{j+1} \quad (2.16)$$

This ensures any element in a space, is a scaled versions of that element in the next space [27]. Altogether, these aforementioned conditions define a multi-resolution scheme.

A. The Scaling Function

Theoretically, obtaining a function $\varphi(t) \in \nu_0$, known as scaling function, whose integer translates is an orthonormal basis of ν_0 is feasible. The collection of these scaling functions, which are defined in terms of integer translates of $\phi(t)$, known as father wavelet, are formulated in [27] as:

$$\varphi_k(t) = \varphi(t - k) \quad k \in \mathbb{Z} \quad \varphi \in L^2 \quad (2.17)$$

The subspace belonging to $L^2(\mathbb{R})$ that is spanned by the generated functions in (2.17) is defined as (2.18), for all possible integer values of k , from minus infinity to infinity.

$$\nu_0 = \overline{\text{Span}_k \{\phi_k(t)\}} \quad (2.18)$$

Equation (2.18) can be interpreted as:

$$f(t) = \sum_k a_k \varphi_k(t) \quad \text{for any} \quad f(t) \in \nu_0 \quad (2.19)$$

The subspace spanned can be expanded by manipulating the scale of the primary scaling function. A two-dimensional class of basis functions is produced from the father wavelet or $\phi(t)$, using scaling and translation operations [27]:

$$\varphi_{j,k}(t) = 2^{j/2} \varphi(2^j t - k) \quad (2.20)$$

This set is an orthonormal basis for the subspace ν_j that can span:

$$\nu_j = \overline{\text{Span}_k \{ \varphi_k(2^j t) \}} = \overline{\text{Span}_k \{ \varphi_{j,k}(t) \}}$$

for all integers $k \in \mathbb{Z}$. In other words, if $f(t) \in \nu_j$, it can be decomposed as follows:

$$f(t) = \sum_k a_k \varphi(2^j t + k) \quad (2.21)$$

The set of nested subspaces requires that on condition that $\varphi(t)$ is in ν_0 , it is also in ν_1 , the subspace spanned by $\varphi(2t)$. This suggests that it is possible to formulate $\varphi(t)$ in terms of a weighted sum of translated $\varphi(2t)$ [27] as:

$$\varphi(t) = \sum_n h(n) \sqrt{2} \varphi(2t - n), \quad n \in \mathbb{Z} \quad (2.22)$$

In this equation, $h(n)$ is a series of real-valued numbers known as the scaling coefficients or low-pass filter coefficients. All the scaling functions and their corresponding scaling coefficients satisfy this equation. The design process of any wavelet system is in fact, selection of appropriate coefficients $h(n)$ [27].

B. The Wavelet Function

Wavelet functions, $\psi_{j,k}(t)$, are capable of identifying and describing essential characteristics of any signal more accurately than scaling functions. These functions span disjoint difference of the subspaces spanned by the various scales of the scaling function [27].

The wavelet spanned subspace w_0 , is defined such that

$$v_1 = v_0 \oplus w_0 \quad (2.23)$$

which can simply extend to

$$v_2 = v_0 \oplus w_0 \oplus w_1 \quad (2.24)$$

and generalize to the following:

$$L^2 = v_0 \oplus w_0 \oplus w_1 \oplus \dots \quad (2.25)$$

where v_0 is the subspace spanned by the integer translates of the primary scaling function $\varphi(t-k)$

. Fig. 2.1 illustrates the nested spaces, related to the scaling function of various scales j ; and also, those related to the wavelets that are the disjoint differences [27].

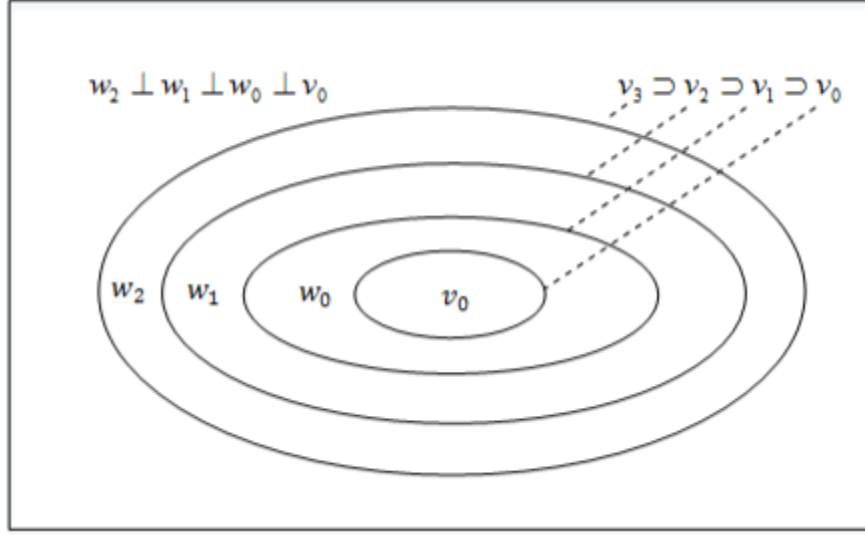


Fig. 2.1: Scaling Function and Wavelet Space

The relation of the different subspaces can be seen from the expression (2.13). From (2.13), it is possible to start with any subspace [27], for example at $j=0$, so that

$$v_0 \subset v_1 \subset v_2 \subset \dots \subset L^2 \quad (2.26)$$

to have

$$L^2 = v_0 \oplus w_0 \oplus w_1 \oplus \dots \quad (2.27)$$

The scale of initial space can be selected at finer scale of, for example $j=10$ to have

$$L^2 = v_{10} \oplus w_{10} \oplus w_{11} \oplus \dots \quad (2.28)$$

or at a coarser scale such as $j=-5$ to have

$$L^2 = v_{-5} \oplus w_{-5} \oplus w_{-4} \oplus \dots \quad (2.29)$$

Or even at zero resolutions, $j = -\infty$, where (2.29) become:

$$L^2 = \cdots \oplus w_{-2} \oplus w_{-1} \oplus w_0 \oplus w_1 \oplus w_2 \oplus \cdots \quad (2.30)$$

Therefore, another way of expressing the relation between v_0 and wavelet subspaces is writing [27]

$$w_{-\infty} \oplus \cdots \oplus w_{-1} = v_0 \quad (2.31)$$

In general, according to Fig. 2.1, the additional information contained in the finer approximation space, v_{j+1} , is included in another subspace of v_{j+1} , such that

$$v_{j+1} = v_j \oplus w_j \quad (2.32)$$

The subspace, known as complementary or detail space of v_{j+1} , is orthogonal to the subspace v_j . Furthermore, it is orthogonal to all the other detail subspaces at different resolutions. However, it is only orthogonal to the approximation subspaces of lower resolution [27]. This can be confirmed by looking at Fig. 2.1

Theoretically, it is always feasible to find a function, wavelet function $\psi(t)$, such that its scaled (at any resolution j) and integer translates form an orthonormal basis capable of spanning the w_j subspace.

Considering that the wavelets exist in the subspace spanned by the next finer-scale scaling function, meaning that $w_0 \subset v_1$, it is possible to represent them by a weighted sum of translated scaling function $\varphi(2t)$ [27]:

$$\psi(t) = \sum_n h_1(n) \sqrt{2} \varphi(2t - n), \quad n \in Z \quad (2.33)$$

Where $h_1(n)$ denotes wavelet filter coefficients.

The function constructed by (2.33) provides the prototype, or mother wavelet for the following basis functions [27]:

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k) \quad (2.34)$$

The whole set $\{\psi_{j,k}(t)\}$ is an orthonormal basis for L^2 (see (2.30)). A set of functions $\varphi_k(t)$ and $\psi_{j,k}(t)$ can also span all of $L^2(R)$. According to (2.27), any function $g(t) \in L^2(R)$ can be expressed as a series expansion in terms of translated scaling functions at a particular scale and various scaled and translated wavelets [27]:

$$g(t) = \sum_{k=-\infty}^{\infty} c(k) \varphi_k(t) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} d(j,k) \psi_{j,k}(t) \quad (2.35)$$

In (2.35), the first summation provides a low-resolution picture or a coarse approximation of $g(t)$. In the second summation, for each increase in parameter j , a finer detail is added to $g(t)$ [27].

A more general formulation for the expansion (2.35) can be given by:

$$g(t) = \sum_k c_{j_0}(k) 2^{j_0/2} \varphi(2^{j_0} t - k) + \sum_k \sum_{j=j_0}^{\infty} d_j(k) 2^{j/2} \psi(2^j t - k) \quad (2.36)$$

or by:

$$g(t) = \sum_k c_{j_0}(k) \varphi_{j_0,k}(t) + \sum_k \sum_{j=j_0}^{\infty} d_j(k) \psi_{j,k}(t) \quad (2.37)$$

The value of j_0 determines the lowest resolution whose subspace is spanned by $\varphi_{j_0,k}(t)$. The remainder of $L^2(R)$ is spanned by the wavelets that provide the higher resolution information of the signal. The expansion coefficients in (2.37) are referred to as the Discrete Wavelet Transform (DWT) of the signal $g(t)$. These coefficients can be calculated by the following equation [27]:

$$c_j(k) = \langle g(t), \varphi_{j,k}(t) \rangle = \int g(t) \varphi_{j,k}(t) dt \quad (2.38)$$

and

$$d_j(k) = \langle g(t), \psi_{j,k}(t) \rangle = \int g(t) \psi_{j,k}(t) dt \quad (2.39)$$

2.6 The Discrete Wavelet Transform

The continuous wavelet analysis of signals provides an immense abundance of data. This statement is justified by considering the two-dimensional nature of the outcome of the continuous wavelet analysis, and the fact that this result proceeds from only a one-dimensional signal. Therefore, it is self-evident that an enormous excess of data is in the CWT. As a result, in this case, only subsamples of CWT that retain certain key features, could be considered, instead [28].

Mallat's Filter banks algorithm, which will be explained in the next section, can be used to perform the Discrete Wavelet Transform (DWT). In this way, DWT is formulated entirely in its own right without explicitly connecting it to CWT. However, DWT can also be regarded as a solution for retaining the essential features of the CWT and discarding the rest, in an efficient manner. Taking this view, the DWT is conceived, as a thoughtful subsampling of the continuous wavelet analysis coefficients, which exclusively examines the dyadic scales. (i.e. λ is selected to take the form of $\lambda = 2^{j-1}$, $j = 1, 2, 3, \dots$ and then, within a particular dyadic scale 2^{j-1} , times t that are located at multiples of 2^j are selected [28].)

Interestingly, the DWT of a data consisting of N values, also consists of N values called DWT coefficients. These coefficients can be organized into $\log_2^N + 1$ series. \log_2^N number of these series are called wavelet coefficients and are associated with scale one up to scale $2^{\log_2^N - 1}$. For scale 2^{j-1} , there are $N_j \equiv \frac{N}{2^j}$ wavelet coefficients and their matching points in time are of the form

$(2n+1)2^{j-1} - \frac{1}{2}$, $n = 0, 1, \dots, N_j - 1$. With appropriate normalization [28], the n^{th} wavelet coefficient at scale 2^{j-1} can be regarded as an approximation to the CWT coefficient at scale 2^{j-1} and time $(2n+1)2^{j-1} - \frac{1}{2}$. The remainder of coefficients are known as scaling coefficients. The scaling coefficients equal (if proportionally adjusted) averages of the original data over a scale of $2^{\log_2 N}$, whereas, the wavelet coefficients equal (if proportionally adjusted) differences of averages over that scale. Hence, the scaling coefficients exhibit a similar trend to that of the data.

Similar to CWT, the perfect recovery of the data from its DWT coefficients is feasible. Sub-sampling the CWT coefficients at just the dyadic scales might imply that a large portion of the information has been discarded. However, surprisingly a signal and its corresponding DWT coefficients are, in fact, two different expressions for the same mathematical entity [28]. Therefore, no information is lost by dropping down from CWT to DWT. In fact, the set of dyadic scales is often sufficient to concisely characterize physical processes, particularly in scientific areas [28].

2.6.1 DWT Calculation, from the Signal Processing View

There is a link between wavelets and digital signal processing, which is defined by Mallat and Daubechies. A signal processing view makes the implementation of wavelet transform much simpler than what the mathematical theory suggests. In this section, we describe how wavelet-transform of a discrete signal is calculated in Mallat's filter bank algorithm, using a set of high-pass and low-pass filters in combination with down-samplers. This algorithm can calculate a signal's wavelet transformed coefficients at any specific finite resolution. The corresponding algorithm for reconstruction will also be described in the continuation.

A. Analysis - From High Resolution to Low Resolution

The approximation of a function at the resolution level $j+1$, corresponding to its projection on the subspace v_{j+1} , can alternatively be expressed as the summation of its projections on the subspaces v_j and w_j [30]. Additionally, since $\{\varphi_{j,k}\}$ and $\{\psi_{j,k}\}$ are orthonormal basis of subspaces v_j and w_j respectively, it is possible to express the projection of a function $x(t)$ in v_j and w_j in the following respective forms: $\sum_k c_j(k)\varphi_{j,k}$ and $\sum_k d_j(k)\psi_{j,k}$. In a similar manner, as clearly articulated in [30]:

“...a given approximation of the signal $x(t)$ on the subspace v_j can be decomposed into a coarser approximation, which is its projection on the subsequent subspace v_{j-1} and details given by its projection on the complementary subspace w_{j-1} .”

Considering these facts, it is apparent that the coefficients at a lower resolution can be calculated from the coefficients at a higher resolution. This is done by an algorithm, whose structure resembles a tree, and it is called filter bank. This algorithm not only calculates the expansion coefficients efficiently but also relates wavelet transform to digital signal processing techniques. The relationship between the scaling function coefficients at a lower scale level in terms of those at a higher scale can be formulated as follows [27]:

$$c_j(k) = \sum_m h(m-2k)c_{j+1}(m) \quad (2.40)$$

The equivalent relationship for the wavelet coefficients is given by

$$d_j(k) = \sum_m h_1(m-2k)c_{j+1}(m) \quad (2.41)$$

For better understanding of (2.40) and (2.41), it is worthwhile to make two points: Firstly, in the discipline of digital signal processing, the filtering of a sequence of numbers is achieved by convolving the sequence with another set of numbers, called the filter coefficients or impulse response. The filtered output sequence $y(n)$ is calculated by (2.42), given an input sequence $x(n)$ and filter coefficients $h(n)$.

$$y(n) = \sum_{k=0}^{N-1} h(k)x(n-k) \quad (2.42)$$

In (2.42), the total number of filter coefficients is denoted by N . The design problem is always the choice of the $h(n)$, to obtain some desired effects [27].

Secondly, two basic operations in multi-rate filters are the down-sampler and the up-sampler. The down-sampler, known as decimator takes an input $x(n)$ and produces an output of the form $y(n) = x(2n)$. The down-sampling operation, obviously poses a risk of losing some valuable information, now that a large portion of the data is discarded, this procedure can affect the information in the frequency domain adversely. This phenomenon is called aliasing that is a mixing up of frequency components.

Digital filtering and down-sampling are precisely what (2.40) and (2.41) do [27]. In other words, two digital filters that are characterized by coefficients $h(-n)$ and $h_1(-n)$, filter the coefficients at level j . Following this, the down-sampling operation then results in the scaling function and wavelet coefficients at the next lower resolution. Mallat and Daubechies are the mathematicians who showed the relation of wavelet coefficient calculation and filter banks. The implementations of (2.40) and (2.41) are illustrated in Fig. 2.2, where the down pointing arrow denotes decimation by two and the other boxes denote filtering or a convolution by $h(-n)$ or $h_1(-n)$.

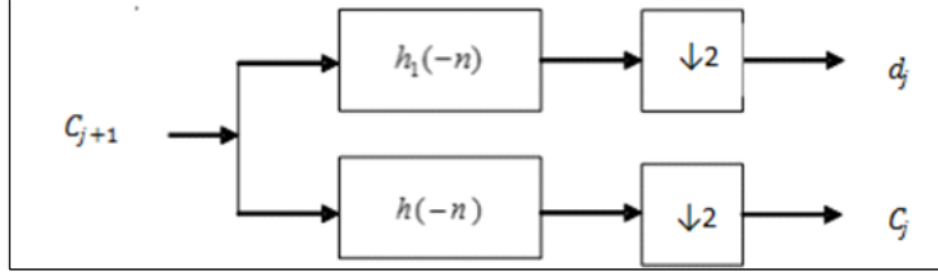


Fig. 2.2. Two-Band Analysis Bank

The filter implemented by $h(-n)$ is a low-pass filter, and the one implemented by $h_1(-n)$ is a high-pass filter. Note that total number of data points out of this system is equal to the number of data points into the system. This means chances are that no information is lost and, therefore, perfect reconstruction of the original signal is possible [27], and this is indeed the case under certain circumstances.

Repetition of the three tasks of splitting, filtering and decimation is possible on the scaling function coefficients to iterate the filter bank [27]. For example, the three scale structure can be seen in Fig. 2.3.

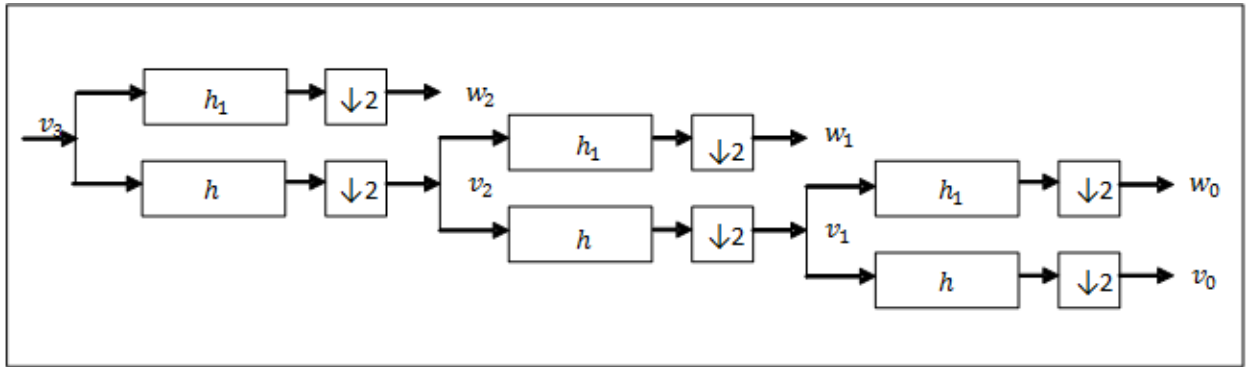


Fig. 2.3. Three stage two-band analysis tree

The first stage in Fig. 2.3, filters $c_{j+1}(k)$ with a high-pass and a low-pass filter. In other words it decomposes $c_{j+1}(k)$ into a low-pass and high-pass band, which yields the scaling coefficients and

wavelet coefficients at a coarser scale: $c_j(k)$ and $d_j(k)$. Afterwards, the second stage decomposes the low-pass band from the previous stage, into another low-pass and high-pass band, which essentially results in a lower low-pass band and a band-pass band [27]. The first stage divides the whole frequency range into two equal parts, while the second stage divides the lower half into two more equal halves and so on.

B. Synthesis – From Low Resolution to High Resolution

The original fine scale (high resolution) coefficients of the signal are constructed by merging the scaling function and wavelet coefficients at a lower resolution, according to the following formula [27]:

$$c_{j+1}(k) = \sum_m c_j(m)h(k-2m) + \sum_m d_j(m)h_1(k-2m) \quad (2.43)$$

For the synthesis in the filter bank, a series of first up-sampling or stretching and then filtering is needed. These operations are precisely what the synthesis equation in (2.43) does. This equation is evaluated by up-sampling the coefficient sequence at level j , $c_j(k)$, then convolving it with the scaling coefficients, $h(n)$. The same process takes place for the wavelet coefficient sequence at level j . Finally, the summation, results in the scaling function coefficients at the next finer scale or at level $j+1$. This structure is illustrated in Fig. 2.4, where $g(n) = h(n)$ and $g_1(n) = h_1(n)$. This merging process can be continued to any level by combining the appropriate wavelet coefficients. The two-scale tree is shown in Fig. 2.4.

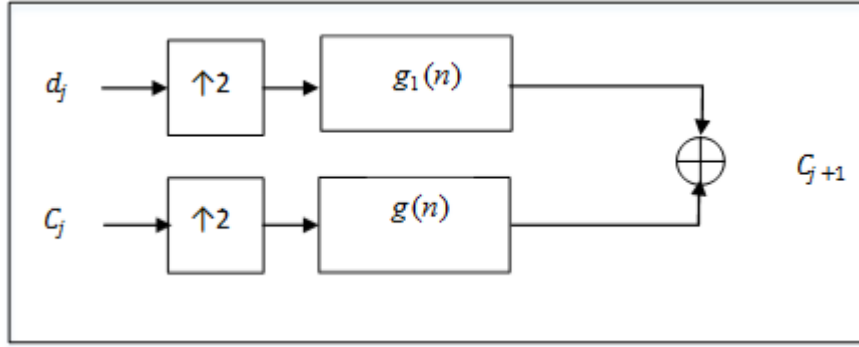


Fig. 2.4. Two Band Synthesis Bank

This set of analysis and synthesis operations is known as Mallat's algorithm. To sum up, in this algorithm, the array of analysis filters calculate the DWT coefficients by employing digital filters (high-pass and band-pass) in combination with down-samplers. In contrast, the array of synthesis filters implement the inverse DWT transform and reconstruct the initial signal from the transformed coefficients.

3 Electric Arc Furnace Modeling

3.1 Mathematical methods for electrical arc furnace modeling

Scientists and engineers have been using mathematical concepts and languages to describe a variety of systems throughout history that has led to better understanding of the phenomenon being

studied. Electrical Arc Furnaces (EAF) are among such systems, and there is an extensive body of literature on the subject of modeling them by mathematical methods. These mathematical models can be broadly categorized into black-box and white-box models, based on the amount of *a-priori* information that is available from the system. However, all available models are almost shades of grey and reside somewhere between black-boxes and white-boxes.

3.1.1 White-Box Models

When some necessary information from the system under study is available, white-box methods are particularly useful. This *a-priori* information can take the form of knowing the type of function which relates the different variables, or the subjective knowledge or intuition of an expert on the system, or a combination of both. In these models, the inner components and the logic of the system are available for inspection; therefore, they are also called glass-box.

Several conceptual frameworks have been proposed in an effort to explain the dynamics of the electric arc furnace behaviour. Most of them have focused on white-box methods. These methods include statistical approaches, probability approaches, differential equations, fast Fourier transform, Markov chain and symbolic dynamics. However, they are not limited to this short list. Most of these methods start from assuming a model structure for the electric arc furnace (which is far from the reality) and then try to calculate the parameters corresponding to the model approximately by using the empirical data. Usually, subjective information is also incorporated into these mathematical models, which cannot be obtained easily in case of complex systems and therefore is not reliable. In this section, typical examples of such methods that are frequently cited in the literature are reviewed briefly. Although each of these views has some virtues, but for the

most part, they rely on subjective and explicit knowledge of the physical process and require understanding of the underlying process.

In the work [34] of Stade et al., the electric arc was mathematically modelled for the analysis of DC EAF influence on the electrical network. The model was based on the statistical analysis of on-site measurements of instantaneous voltage and current information. These measurements were statistically analyzed, and it was shown through investigations that a statistical normal distribution was valid for these values. Consequently, the v - i characteristic was synthesized for each of the two different operation periods, namely the start of melting and quiet melting. It was proven through simulations that the DC EAF modeling was useful for power quality studies.

D. Tewari et al. in [35] applied the concept of symbolic dynamics to model highly varied electrical loads in AC arc furnaces. Here, symbolization refers to the transformation of raw signal measurements into a series of discretized symbols that can be processed to extract information about the process generating the data [36]. On the whole, in this proposed method, the discretized signal was coded into symbols and the symbols were then treated as if they were in a natural language for further processing. In a more detailed view, for the purpose of predicting future values of the signal, a symbolic dynamic dictionary was formed using the instantaneous values in the signal. Then, the forecasting was done by matching the latest measurements with items in the dictionary. Those items with high fractional occurrence were chosen as the forecast values. The researchers forecasted the AC arc furnace current and compared the results with actual data using Common Signal Index (CSI) and Kolomogorov-Smimov (KS) similarity measures. Overall, symbolic dynamics appeared to be able to predict the load current, to a reasonable degree. However, this method has some limitations that may have adversely affected the quality of the

model, as many parameters such as symbol sequence length and number of partitions are selected empirically.

Chen et al. in [37] presented a function space valued Markov model for EAF current and voltage. In this study, the state-case time series was generalised to a cycle vector- case time series in order to predict one or more cycles of EAF current or voltage. In order to reduce the computations, each cycle was then approximated by a small number of parameters using fast Fourier transform. This task was necessary for the application of the Markov method. The effectiveness of the obtained models has been demonstrated for harmonic compensation tasks.

Petersen et al. in [38] adopted a probability approach to model arc voltage and resistance for study and simulation of voltage flicker. On the whole, these models could make an accurate estimate for chaotic values of arc resistance and voltage assuming that the arc parameters were closely Gaussian. More specifically, this work was carried out based on statistical studies that had shown the distribution of arc voltage in time is approximately Gaussian, when mechanical resonance is inexistent. In a similar manner, the distribution of the furnace resistance was approximately modelled as Gaussian. This was done in despite the fact that the distribution was fairly tight around the mean and, hence, different from an exact Gaussian distribution. The authors acknowledged that these models were able to forecast the flicker accurately. However, they could not simulate the arc behaviour perfectly.

Acha et al. in [9] presented direct harmonic domain representations of dynamic nonlinear elements (such as arc furnace) in power systems that were traditionally restricted to only linear elements. Altogether, the nonlinear characteristics were represented by fitting those characteristics with a polynomial, for which special harmonic domain processing via convolutions has been developed, or by directly applying a fast Fourier transform. Subsequently, the authors derived a model for

electric arc in the form of a differential, equation grounded on the law of conservation of energy and took into account simple energy balance considerations. For example, in the power balance equation of the arc it was assumed that the cooling effect was a function of the arc radius only. While in fact, it is also a function of arc temperature. This dependence, therefore, is ignored in order to keep the model simple.

To summarize, the literature of EAF modeling is heavily dominated by methods that are based on explicit equations, oversimplified assumptions or subjective information, the validity of which are rather questionable. Besides, in these methods not enough attention has been devoted to the stochastic nature of the data. As a result, although these models are considered to be “good,” there is a tendency among researchers towards more black-box methods in pursuit of “better” models, in which the model structure is not presumed, and the behavior of the system is described entirely based on the actual input and output data.

3.1.2 Black-Box Models

In contrast, when there is no *a-priori* information available, and the peculiarities of what is going on inside the system are entirely unknown, the black-box models come into play. Using black-box methods, the system is described only in terms of its input and output, without any knowledge of the internal workings. Black-box methods have gained enormous popularity, especially in the case of modeling dynamic and highly nonlinear systems such as electric arc furnaces, where more traditional white-box methods do not seem to be able to capture all the dynamics of such a complicated system.

A. Single Scale Black-Box Methods

In this thesis, we refer to black-box methods, in which wavelet transform has not been utilized as single scale methods. These include artificial neural networks, fuzzy logic systems and neuro-fuzzy logic systems. In this section, the application of these methods to the modeling problem of electric arc furnace is investigated.

In [4] Sadeghian and Lavers applied artificial neural networks for modeling the highly non-linear and stochastic EAF systems. Neural networks are known to be especially effective and powerful for modeling these systems since all they use for describing the system under study is data. The characteristic of the EAF has been modelled by mapping from the EAF current to the EAF voltage by means of neural networks. Nevertheless, because the data proved to be insufficient and the neural network could not converge to an optimum solution, the arc current derivative with respect to time was also considered in the inputs along with past values of EAF voltage and EAF current. This added quantity could easily describe hysteresis and chaotic behaviors in the $v-i$ characteristic. The neural network-based EAF models were successfully developed using two different neural network architectures: MLP (Multi-Layer Perceptron) and RBF (Radial Basis Function). In the end, it has been shown that the computational results were strictly comparable to the existing measurements.

Chang et al. in [39] also presented an effective method, based on artificial neural networks, for modeling the highly non-linear $v-i$ characteristics in electrical arc furnaces. Multilayer perceptrons with back-propagation learning algorithms are considered to be the most popular neural networks and deliver high precision and the capacity for managing extremely nonlinear problems. However, they suffer from the disadvantage of slow convergence and quickly trapping into undesirable local minimum of the error surface [40]. Therefore, they used radial basis function neural networks instead. These networks have a much simpler structure and a set of compactly supported basis

functions that can be locally adjusted for more accurate function learning and modeling. In order to capture the highly nonlinear and time-varying $v-i$ characteristics of an EAF, they made some modifications to the traditional radial basis function networks. The problem with traditional neural networks is that they can only deal with the ordinary one-to-one or multiple-to-one function mapping problems. However, the $v-i$ characteristic of an EAF differs from these mapping definitions, typical of an ordinary function and, as the result, a conventional neural network cannot converge in all parts of the $v-i$ characteristic curve. To overcome this drawback, a lookup table method was proposed. In this method, the associated parameters of the radial basis functions used in the design of neural network at each sampled time instant were stored, over the course of the training process for all the possible function mapping relationships between the network input and output. In effect, the transfer functions were no more fixed which made it a good choice for modeling time-varying characteristics attributable to electric arc furnace. The results indicated that the proposed method was efficient and reliable and generalizable to other power engineering studies.

Fuzzy Logic Systems can be considered another nonlinear black-box modeling structures. Their key properties of universal approximation and functional equivalence with RBF, support the effectiveness of FLS as a solution to highly nonlinear problems [15].

Sadeghian and Lavers in [14, 15] proposed fuzzy logic systems as means of EAF modeling. To this end, they established two different model structures; classic and adaptive FLS. The underlying idea in both techniques was establishing an input/output structure, and employing an appropriate algorithm for learning, ensuring that the fuzzy logic system (FLS) is capable of modeling the nonlinear $v-i$ characteristics in electric arc furnace accurately.

In the classic FLS, the input/outputs of the system and their interval domains were firstly identified. Each domain interval was then divided into N regions arbitrarily based on the complexity of the problem. The Gaussian function, which covers the entire input space was used as the membership function. The input/output data pairs were used to extract the rules and the “Conflict Resolution” strategy was used to trim the rule-base. “Generalized Modus Ponens” and the “height defuzzification” methods were used as the implication and the defuzzification techniques respectively.

Their result showed that the classic fuzzy method had a faster training process compared to RBF networks. However, the trial-and-error approach in this training, sacrificed the accuracy of the models. Hence, they were inferior to RBF networks in that sense. Another disadvantage was that in this method *a-priori* or heuristic knowledge was needed in order to trim the huge number of rules in the rule base. Lastly, this technique did not consider the distribution of data in the input/output space and ignored the natural grouping of the data.

The notion of adaptive FLS was, therefore, used in order to improve fuzzy modeling. It has been demonstrated in Sadeghian’s work that by augmenting FLS with neural networks, adaptive FLSs could alleviate the main weaknesses in the classic FLSs.

The researchers used the Adaptive Neuro-Fuzzy Inference System (ANFIS) structure to develop their adaptive fuzzy network. This approach uses the Gaussian membership function for fuzzy sets, linear membership functions for the output variables, the “subtractive cluster analysis” method for extracting rules and “Sugeno’s inference mechanism” for the process of inference and reasoning. The ANFIS algorithm, which is a hybrid learning algorithm based on “gradient descent” as well as “least-squares estimate” was employed to obtain and update the parameters corresponded to the membership functions (i.e., mean and standard deviation) along with the coefficients of the output

linear functions. In effect, their adaptive FLS provided the accuracy, learning capabilities and adaptivity of neural networks along with the fast training and generalizability of FLSs at the same time [20].

In the following section, the black-box methods in which the advantageous properties of wavelet transform are also employed are reviewed. In this thesis, these methods are termed multi-scale methods. Although not all of these methods have been directly applied to electric arc furnace data, the methods can also be applied to any complex, nonlinear time series such as voltage and current of electric arc furnaces.

B. Multi-Scale Black-Box Methods

At this point, more satisfying approach seems to lie in the integration of black-box and multi-scale methods, namely wavelet transform. In the literature, several black-box methods have been proposed that take advantage of wavelet transform. However, almost all of these studies concentrated on time series prediction and not modeling task, even at a general level. It is not necessary to mention that in this case, studies on arc furnace modeling, utilizing wavelet transform are still lacking.

To begin with, in several publications, the wavelet theory and neural networks have been combined into one to yield a new form of neural network, namely Wavelet Neural Network (WNN). The main purpose of these studies was to formulate a link between theory of wavelets and artificial neural networks. In a very broad sense, in these types of networks, the ordinary activation function such as sigmoid is replaced with a multi-scale wavelet function and the corresponding wavelet coefficients are assigned to the weights. Since wavelets have already shown their excellence in many data analysis tasks, this combination may provide faster convergence, higher ability of

generalization and improved prediction, and several studies confirm this. For example, Bashir et al. in [41] discussed the application of the wavelet neural networks to the problem of short-term load forecasting and compared the superior results with that of ordinary neural networks. In this work, the Morlet wavelet was chosen as the activation function because of its simple, explicit expression. A three-layer network structure was designed, and the training was done by using the Lavenberg-Marquardt back-propagation algorithm. Weights, biases and parameters of the wavelet neural network were also initialized according to the Nguyen-Widrow algorithm and were later adjusted during the training phase.

In another study [42], Ying et al. adopted a novel wavelet neural network model with two input layers to predict the power consumption of EAFs. Input variables included the process parameters that affected the electricity consumption. These variables were not input in one layer, unlike the traditional neural networks. Instead, they were put in different layers, in compliance with their action sequence and thus, could reduce the network scale. In this work, one-dimensional Mexican hat wavelet function, which satisfies the requirements of compact support and regularity, was used for the activation function of the hidden layer; S-tangent function was chosen for the activation function in the output layer. The hidden layer was decided to have a total number of 3 and 16 neurons. In their proposed scheme, they also incorporated genetic algorithm as the learning algorithm for weight optimization in the neural network and could achieve global optimization and fast convergence. According to the results, the largest error between predicted and actual electricity consumption was only 1.69%.

Masuda et al. in [43] developed and tested the idea that wavelet transform preserves the topological structure and chaotic properties of the original time series in its frequency components, given as band-limited wavelet coefficients. Therefore, they proposed a novel method of predicting a

dynamic system by applying the inverse wavelet transform to predictees of each scale. In this work, the researchers suggested two methods based on wavelet transform and non-decimated wavelet transform, which are identical in nature and differ only in few details. First, the time series is decomposed by either wavelet transform or non-decimated wavelet transform. Then, each frequency component at different scales is predicted for the next certain amount of steps depending on the transform, using any method appropriate. Afterwards, the inverse wavelet transform is applied to specific coefficients, again depending on the transform, to obtain the forecast values in the original time series. In this study, they confirmed that the application of wavelet transform made long-term prediction of chaotic time series feasible, which was otherwise impossible because of the sensitive dependence of chaotic time series on initial conditions. They also prepared the ground for other researchers to develop novel prediction methods based on predictees of frequency components.

In a similar manner to WNNs, the theory of wavelets and Support Vector Machines (SVM) has also been combined to yield the advantages of both techniques. SVM has a rigorous theoretical and mathematical foundation, which has a powerful generalisation capability. The basic idea of using SVM is to facilitate nonlinear regression problems by transforming the data into a higher dimensional feature space, in which conducting a linear regression is then possible.

One of the examples of this integration is presented in [44]. Niaona et al. in this study introduced a multi-scale energy consumption prediction model for electrical arc furnaces based on Least Square Support Vector Machines (LS-SVM). In their three-step algorithm, wavelet transform was firstly conducted on the arc furnace power consumption sequential data. Then, the wavelet coefficients at t time instance were taken as the input of LS-SVM, while those at $t+T$ time instance were taken as the output of SVM. Finally, the predicted value was calculated by using wavelet

reconstruction. The prediction error in this combined model was only 1.98% that could meet the requirements of engineering practice.

Recently, several authors have proposed using wavelet decomposition in combination with neural networks to provide more accurate time series forecasting. However, these approaches differ from wavelet neural networks, on the way the wavelet transform is used.

Soltani in [45] used this combination to provide acceptable prediction in a measured time series. Their objective was to predict the value of the measured time series at a later time; namely at p steps ahead, using all the observations. In order to build a functional relationship that could map the time series and the prediction value accurately, they followed several steps. They firstly applied wavelet filters iteratively to construct trend series and a hierarchy of detail series at different scales. Afterwards, a neural network was trained to model every time series. Lastly, they combined these values to obtain a prediction to the original time series. In this proposed method, unlike conventional ones, in which obtained series are treated separately, the interdependencies between the different series were also considered. In other words, instead of finding estimators that mapped each series and its equivalent prediction value, they looked for ones that could find the mapping between all the different series and the predicted value at a particular scale. In effect, this sophisticated method also included information of other series for improving the prediction accuracy of each time series. The results obtained through some well-known time series substantiated that the wavelet decomposition method profoundly reduced the empirical risk.

In another study [46], Soltani et al. addressed the problem of predicting time series with long-term memories by using wavelet transform based on an Auto-Regressive (AR) model. The classical way to approach such time series was Auto Regressive fractionally Integrated Moving Average (ARfIMA) models. However, the estimation of the integration parameter in these models was not

an easy task. The way around this problem was to transform the time series into another domain such that it can be modelled with simpler model structures. In this study, the wavelet transform was applied to the time series with long memories to decompose it into a trend (an approximation to the function) and a hierarchy of detail series (each at a particular resolution) that were stationary and had short memories. As the trend series will contain no more useful information after few decomposition levels, the predicted value in the original time series can be calculated by aggregating the predicted values in just the detail series. These obtained detail series could then be modelled with less problematic ARMA models. They proceed even further by analyzing the statistical properties of the derived time series and found out since the moving average did not exist, they could simply use the straightforward AR models. In order to also consider the inter-correlation between the time series, they exploited the multi-channel version of the AR model.

Zhongliang et al. in [47] proposed another method to improve short-term load forecasting accuracy based on various RFB neural networks. The wavelet transform was firstly used to decompose the electrical load into different parts, corresponding to different influencing factors. Then a compound model with three sub-RFB neural networks was constructed, each for part of the decomposed electrical load. The model was then trained using historical data, and lastly prediction was made. According to simulation results, this multiple RBF neural network method achieved an excellent performance. The fact that the mean absolute percentage error decreased from 3.86 by ANN method to 1.7 by the proposed method, lend support to this claim.

Another study [22], also investigates the effectiveness of this setting. In this study, an accurate wavelet-GA-ANN based hybrid model was developed for short-term load forecasting in power systems. The à trous wavelet transform was firstly applied to the load data. The output was then used for the training process of the RBF neural network. The novelty of this study is that they

optimized the structure of their auto-configuring RBF neural network by a variant of genetic algorithms. The results obtained from various actual load data showed that the proposed model was efficient and could predict the load accurately for seven steps ahead. Furthermore, the proposed model proved itself to be more accurate as compared to the models that were solely based on the RBF networks.

The paper [48] utilized the synergies of three powerful techniques which could significantly enhance the power of forecasting. Bashir et al. in [48] adopted an ANN-based method to predict the load an hour ahead. They used the Particle Swarm Optimization (PSO) technique in the training phase of the neural network, instead of using a conventional back-propagation technique. As a preprocessing step, the historical load data was firstly wavelet transformed and then fed into a neural network. The obtained model could improve the accuracy by employing wavelet analysis to the inputs of the neural network. This is because the behavior of load is characterized by wavelet transforms more effectively, as confirmed by the authors.

Although each of these studies provides invaluable insight into the problem, there is a degree of ambiguity in relation to which wavelet coefficients are considered more descriptive. To our knowledge, this issue has been scarcely investigated from the theoretical point of view.

4 Methodology

In this chapter, the proposed modeling procedure, which is a combination of multi-scale filtering and nonlinear regression techniques, is presented. The proposed algorithm combines the algorithms presented by Sadeghian et al. [4] and Renaud et al. [29] and includes four main steps. These steps are:

- data acquisition,
- preprocessing,
- feature extraction, and
- hybrid neuro-wavelet modeller.

In the proposed approach, first, the voltage and current ($v-i$) information are collected from an operational electric arc furnace. Then, the preprocessing unit transforms the measured $v-i$ signals from the time domain to the joint time-frequency domain using a special form of wavelet decomposition analysis with the intention of obtaining a more detailed and deeper insight about the data. In the feature extraction phase, the relevant wavelet coefficients that are more representative of the signals are identified. These coefficients are then used to form the modellers' input vector. In the final step, several artificial neural networks are built to find the functional relationships capable of mapping the arc current to voltage with the principal concern of minimizing error. The proposed algorithm is shown in

Fig. 4.1.

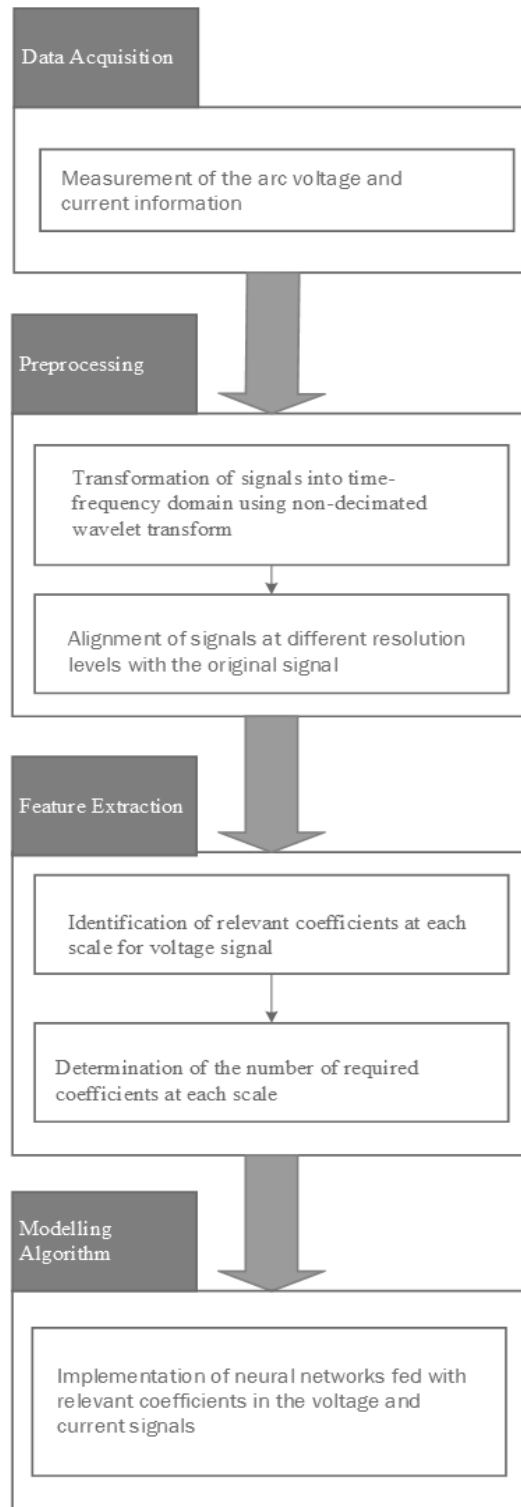


Fig. 4.1. The proposed modeling algorithm

4.1 Data Acquisition

Providing accurate and detailed functional mapping from the input space to the output space is at the heart of any modeling technique [4]. That is, the identification of the input and output components, in accordance with the objective, is necessary prior to any other tasks.

In this study, the principal objective is to model the behaviour of an electric arc furnace that operates within electrical networks. The voltage-current ($v-i$) characteristic, which signifies the relationship between these two variables, is a common way of describing the operation of electrical devices and, hence, modeling their behaviours. Consequently, the principal objective can be considered as an accurate mapping from the EAF current to the EAF voltage. In this study, the voltage and current measurements from an operational EAF was used.

4.2 Preprocessing Methods

The preprocessing tasks employed on the arc voltage and current signals include a transformation to the wavelet domain together with the alignment of the resulting coefficients with the original measurements. The description and the rationale behind each task are given in the following sections.

4.2.1 Wavelet Transform

The arc voltage and current are both non-stationary signals. That is, their statistical properties change over time [49]. Moreover, these signals have characteristics that change in both time and frequency domains. The readers are referred to Fig. 4.2, Fig. 4.3 and to Fig. 4.4, Fig. 4.5 for a close-up view of these signals. It is widely accepted that the localized wavelet transform is one of the most effective mathematical tools for describing this type of data. In a broad sense, this

transformation redistributes the embedded information in the signals, among the wavelet coefficients in a manner that is more effective and advantageous than the original signal [30]. This is the primary motivation behind analyzing these signals in the wavelet domain instead of their original one.

As has been mentioned earlier, the transfer to the wavelet domain has many advantages over time or frequency domain. To begin with, wavelets are not only time localized, but they also have the frequency localization property. For this reason, representation of the signals using wavelets can be sparse. In addition, wavelets have the capability of analyzing the signals more comprehensively due to their multi-resolution property. This means that each component of the signals can be studied with a resolution that matches its scale. In this case, higher frequency components can be studied with better time resolution and lower frequency ones with better frequency resolution. Interestingly, this is exactly how most signals, including the ones in this work need to be studied. That is because in practice, low frequencies often endure over the entire duration of the signal. In comparison, high frequencies show up as sudden bursts from time to time [24].

Put it another way, the voltage and current signals contain both discontinuities and smooth components. Analysis of such signals is highly beneficial by wavelets, as this transform offers a variety of basis functions that differ in scale and length. Consequently, short (local in the time domain and therefore, smaller compact support) and high-frequency wavelets can be used for discontinuities, and long (with larger support), low-frequency ones for smooth parts [50].

Furthermore, using wavelets, the complex multi-scale structure of the signals can be simplified. This transformation leads to the generation of signals of simpler structures with much shorter memories and less temporal dependence [51]. These signals, resulting from the transformation, are

indeed easier to model, and it is not uncommon to see that even models that are not sufficient in the time domain are accurate in the wavelet domain [25].

As a final point, wavelets are used for the purpose of learning all the EAF dynamics. This is because wavelets decompose the signals iteratively and result in a hierarchy of new signals each of which contain the EAF dynamics at a different scale.

The above mentioned reasons constitute the primary motives for using wavelet transformation prior to applying any modeling technique. However, not all forms of wavelet transform are suitable for analyzing time series for the purpose of modeling. In the following section, the wavelet transform that is the most appropriate is introduced.

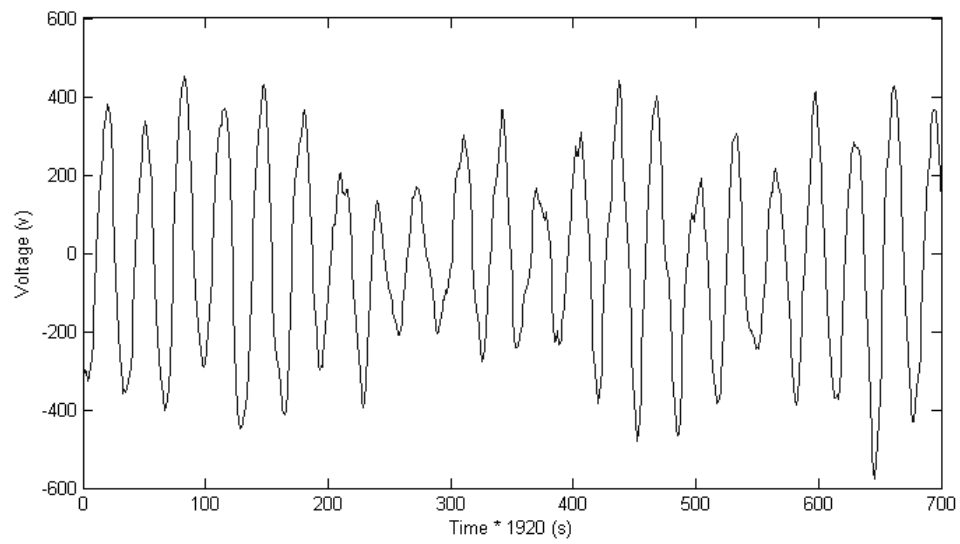


Fig. 4.2. The arc voltage signal

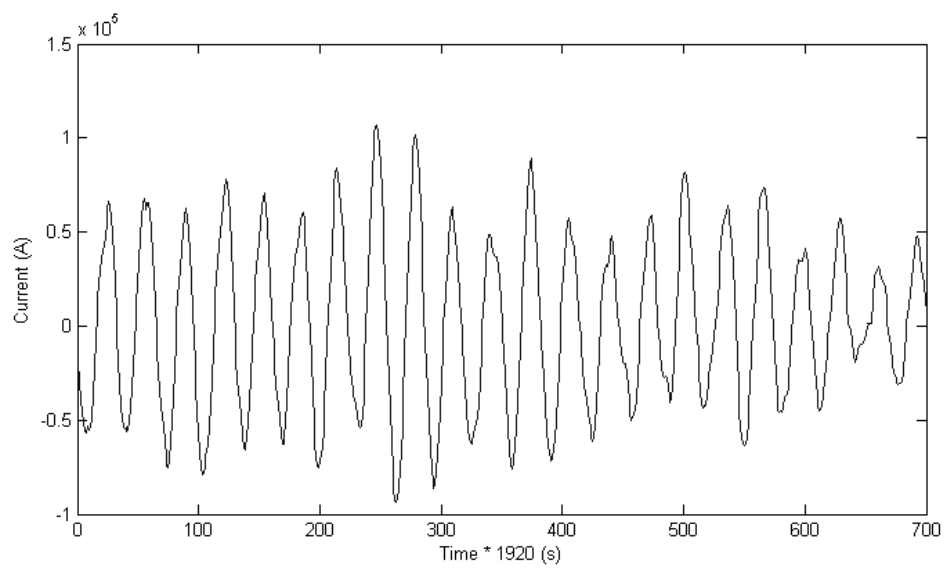


Fig. 4.3: The arc current signal

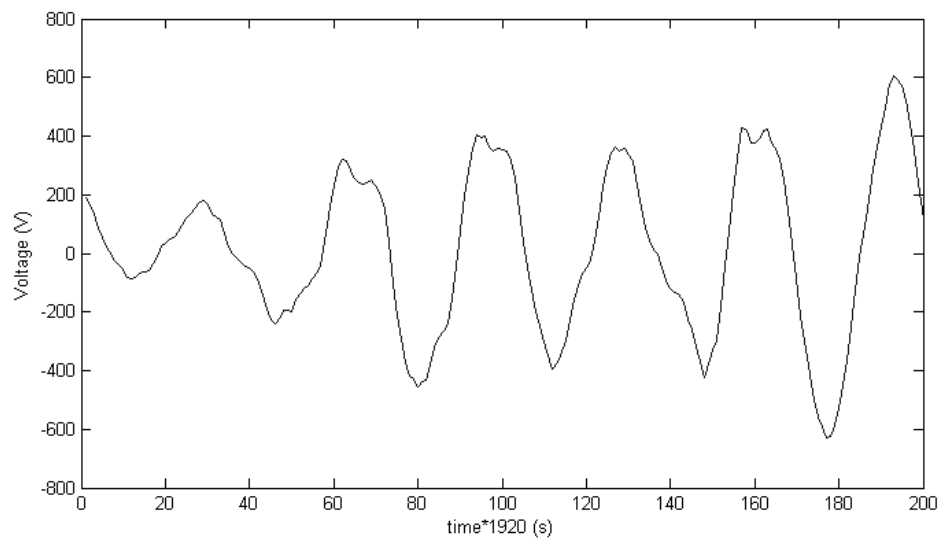


Fig. 4.4: A close-up view of the voltage signal

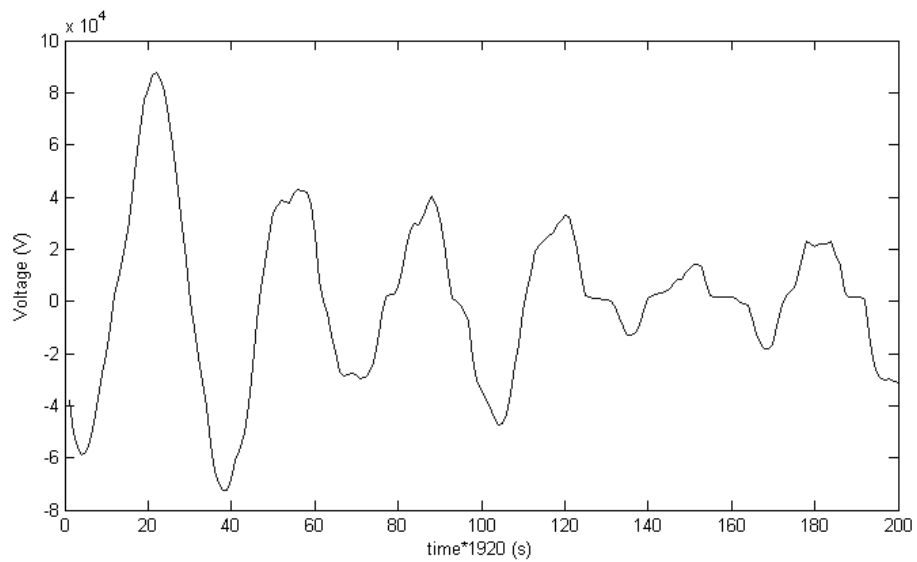


Fig. 4.5: A close-up view of the current signal

4.2.2 The Non-Decimated Discrete Wavelet Decomposition

There are a large number of wavelet transforms where each may be more suitable for a particular application. In this section, the rationale for selecting the “non-decimated discrete wavelet transform” is explained in detail.

Although the arc voltage and current are continuous variables, they are discretely sampled at a rate of 1920 Hz (Nyquist rate) during the data acquisition phase. Therefore, a discrete form of wavelet transform seems to be the most appropriate. The output of this discrete wavelet transform can be either decimated or non-decimated. Decimation in this context means retaining one sample out of every two (every other data point) in order to maintain only those information that are sufficient for perfect recovery of the input data [29]. However, under these circumstances, it is difficult to see the correspondence between information at a particular point in time at different resolution levels resulting from the wavelet transform [29]. Because this capability is essential in the modeling schemes followed in this study, a non-decimated transform is desirable. Furthermore, the decimation operation can raise another problem, namely, “aliasing.” Aliasing in this context refers to the effect that causes different signals to be indistinguishable [52]. Aliasing is the direct result of sampling a signal at a rate lower than Nyquist, or sampling at a rate which is not higher than, or equal to twice the maximum frequency in the signal [53]. As a consequence of this, higher frequencies get superimposed on lower frequency components, which adversely affect the information in both lower and higher frequencies. In sum, aliasing can corrupt the information in the signal, and hence, make the reconstruction of the original signal impossible.

Besides, a decimated discrete wavelet transform is an orthonormal transform and may not be the best representation for our signals. This is because each of the orthonormal bases contains equally important information in its corresponding coefficient [25]. Therefore, there is a possibility that

some useful information is removed during the feature extraction phase, in pursuit of retaining only the key information. The non-orthogonal bases, on the other hand, are not independent. As a result, it is less likely for some key features to be neglected.

Finally, and most importantly, transform in the decimated form suffers from the lack of shift invariance. Shift variance implies that decimated discrete wavelet transform outputs are sensitive to the choice of origin of the signals [54]. In order to clarify the reasoning behind the sensitivity to the choice of origin, the calculation of wavelet coefficients should be inspected carefully. In [28], Percival and Walden state:

“...while the wavelet coefficients for the decimated discrete wavelet transform can be interpreted as a difference between two weighted averages, the intervals over which these averages are made are rigidly fixed a priori and hence might not line up well with interesting features in a time series. A change in the starting point for a time series data can yield quite different results due to juxtaposition of the time series with the averaging intervals predefined by the decimated discrete wavelet transform.”

Given the latter points, the decimated discrete wavelet transform is not appropriate for analyzing time series with no distinctive start and end. In contrast, the non-decimated discrete wavelet transform includes all possible placements of averaging intervals and hence, cancels the undesirable effects that choosing a particular starting time might have [28], and as a result, the property of shift invariance is acquired. All in all, the non-decimated wavelet transform is the alternative solution for resolving all the afore-mentioned issues.

Using the non-decimated discrete transform, the signals are decomposed into several levels with different scales. Each level has the same length as the input signal. Consequently, locating relevant information at each level for the same point in time is now possible.

The only demerit of this transform is the extra storage requirement, which may not be very economical. This is because the non-decimated discrete wavelet transform of J levels for a time series of length N , is a highly redundant transform yielding $J+1$ new time series $\tilde{W}_1, \tilde{W}_2, \tilde{W}_3, \dots, \tilde{W}_J, \tilde{V}_J$ each of dimension N . Here, \tilde{V}_J refers to the vector of the J^{th} level scaling function coefficients and \tilde{W}_J to the J^{th} level wavelet coefficients of a signal. However, this extra storage is not too much since the value of J is a constant and also small in practice.

A. The MODWT Wavelet Transform

The non-decimated discrete wavelet transform used here is described and fully formulated in [28], and also implemented as a MATLAB toolkit in [55]. This transform is called Maximal Overlap Discrete Wavelet Transform (MODWT). A MODWT wavelet filter must satisfy the following three basic properties in (4.1) [28] having that n is a nonzero integer:

$$\sum_{l=0}^{L-1} h_l = 0, \quad \sum_{l=0}^{L-1} h_l^2 = \frac{1}{2} \quad \text{and} \quad \sum_{l=-\infty}^{\infty} h_l h_{l+2n} = 0 \quad (4.1)$$

In these formulas, h_l refers to the filter coefficients and L to the filter length.

A MODWT wavelet filter must: (i) have a zero sum, (ii) have energy of $\frac{1}{2}$, and (iii) must be orthogonal to its even shifts. The MODWT scaling filter satisfies the three conditions [28] of

$$\sum_{l=0}^{L-1} g_l = 1, \quad \sum_{l=0}^{L-1} g_l^2 = \frac{1}{2} \quad \text{and} \quad \sum_{l=-\infty}^{\infty} g_l g_{l+2n} = 0 \quad (4.2)$$

where for any arbitrary sample size n , g_l refers to the filter coefficients, l to the filter index, and L to the filter Length.

The j^{th} level MODWT wavelet and scaling coefficients are defined to be the N dimensional vectors $\tilde{W}_{j,t}$ and $\tilde{V}_{j,t}$ (same as the original signal) whose elements are:

$$\begin{aligned}\tilde{W}_{j,t} &\equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \bmod N} \quad \text{and} \\ \tilde{V}_{j,t} &\equiv \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} X_{t-l \bmod N}\end{aligned}\tag{4.3}$$

And $\{t = 0, 1, 2, \dots, N-1\}$, where $\{\tilde{h}_{j,l}\}$ and $\{\tilde{g}_{j,l}\}$ are respectively, the j^{th} level MODWT wavelet and scaling filter coefficients, which are formulated in terms of the j^{th} level wavelet and scaling filter coefficients $\{h_{j,l}\}$ and $\{g_{j,l}\}$ via $\tilde{h}_{j,l} \equiv h_{j,l}/2^{j/2}$ and $\tilde{g}_{j,l} \equiv g_{j,l}/2^{j/2}$. It is important to consider here that $L_j \equiv (2^j - 1)(L - 1) + 1$.

These j^{th} level MODWT coefficients can also be obtained from the MODWT scaling coefficients of level $j-1$ via recursions:

$$\begin{aligned}\tilde{W}_{j,t} &= \sum_{l=0}^{L-1} \tilde{h}_l \tilde{V}_{j-1,t-2^{j-1}l \bmod N} \quad \text{and} \\ \tilde{V}_{j,t} &= \sum_{l=0}^{L-1} \tilde{g}_l \tilde{V}_{j-1,t-2^{j-1}l \bmod N}\end{aligned}\tag{4.4}$$

where $\tilde{V}_{0,t} \equiv X_t$.

The j^{th} level MODWT wavelet coefficients can also be calculated by taking the difference between successive scaling function coefficients:

$$\tilde{W}_{j,t} = \tilde{V}_{j-1,t} - \tilde{V}_{j,t}\tag{4.5}$$

Although the primary focus of this study is on modeling, the created models can also be used for the prediction purpose. Modeling and forecasting are perceived as similar tasks. However, a slight notable difference exists where in any forecasting scheme, at any time instance, future measurements are indeed not available. Therefore, they cannot be part of the forecasting process [29] whereas in modeling, this limitation does not apply.

Careful attention to (4.4) reveals that in the calculation of coefficients in coarser resolutions at time t , information of other higher resolution coefficients in times after t may need to be considered. This happens in cases where non-zero filter coefficients exist for negative indexes. For this to be the case, depend on the wavelet filter being used. Therefore, in case the model is also going to be used for prediction, only “Causal” wavelet filters should be used. In other words, the nature of any forecasting technique imposes causality on filters [45]. This constraint implies:

$$\tilde{h}_l = \tilde{g}_l = 0 \quad \text{where } l < 0 \quad (4.6)$$

“Filter causality” is the essential requirement for adopting this modeling scheme for the prediction purpose. Otherwise, this limitation does not apply and, therefore, any wavelet filter (perhaps more regular) can be used. In this way, more regular wavelets can be used to adapt to the particularity of the data [56].

The filters that were tested in this study are taken to be causal and include:

- The Daubechies Extremal Phase filters,
- The Daubechies Least Asymmetric filters (or the Symlets),
- The Best-Localized filters and The Coiflets.

The wavelet and scaling function filter coefficients, belonging to the last three groups are shown in Fig. 4.6, Fig. 4.7 and Fig. 4.8 respectively.

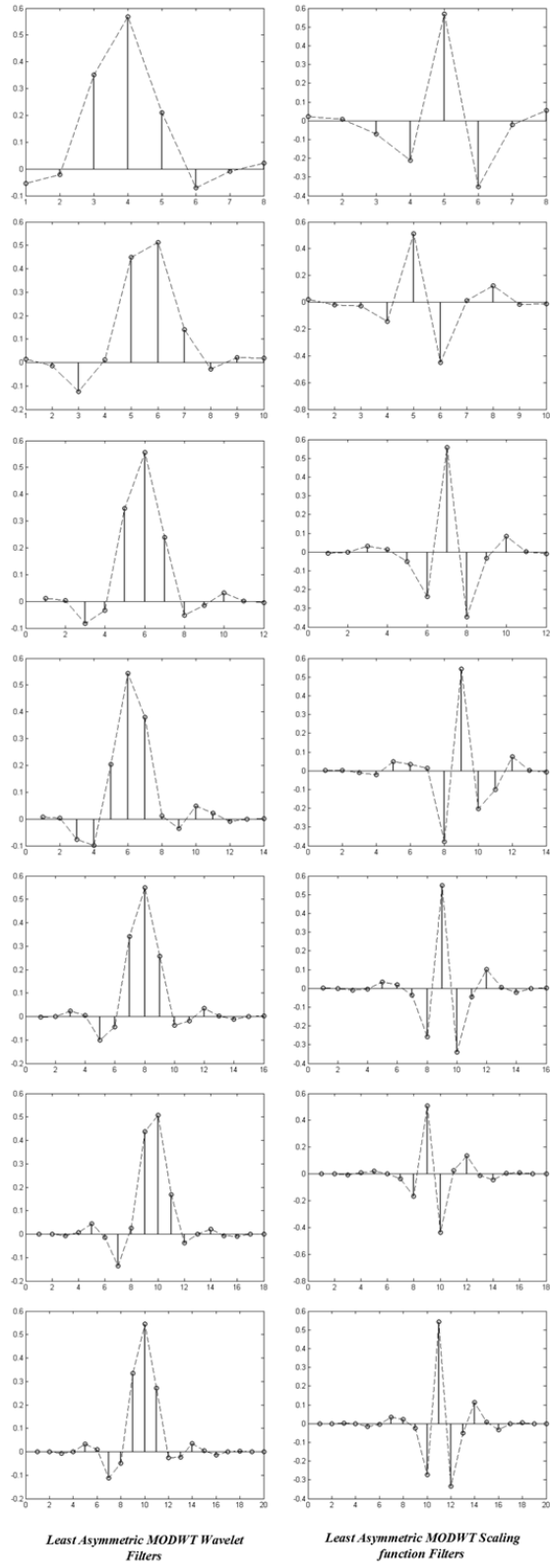
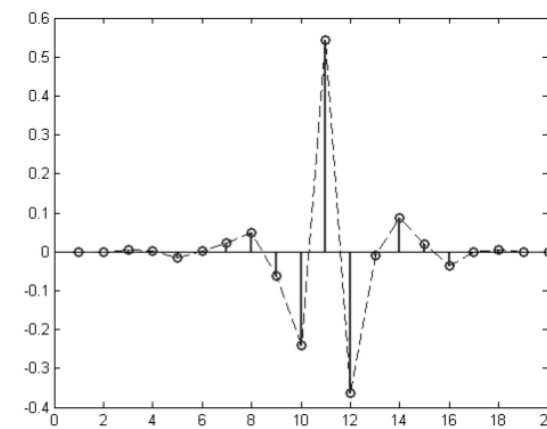
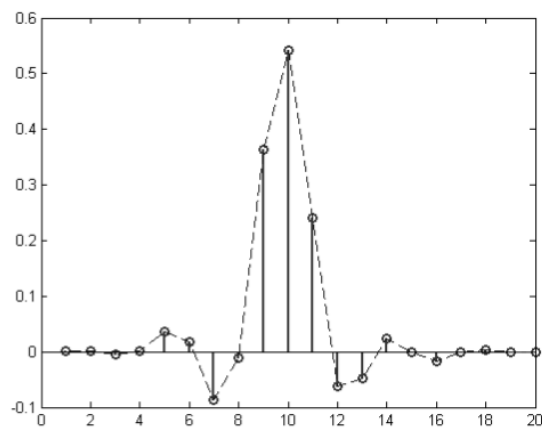
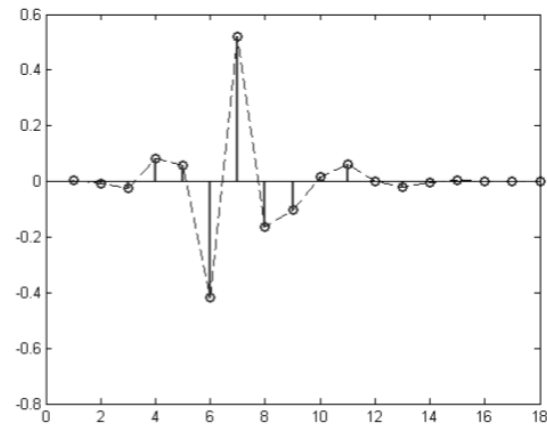
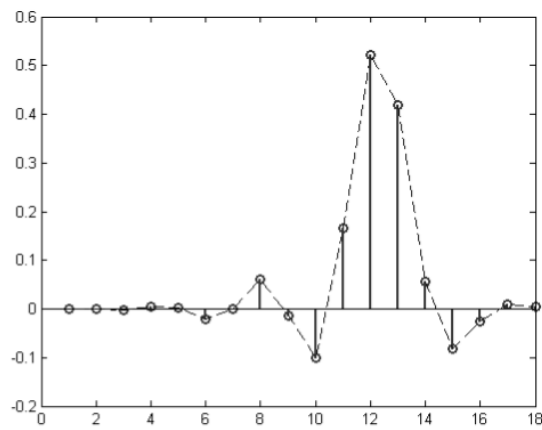
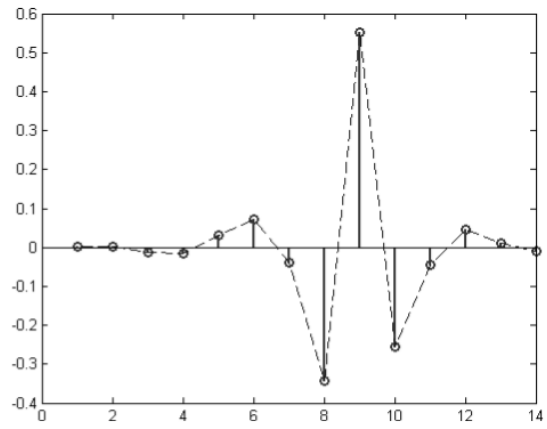
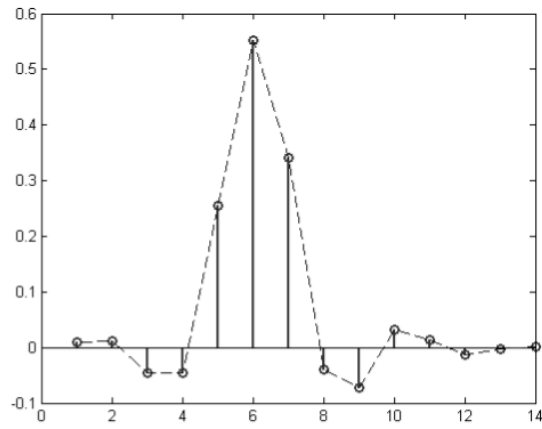


Fig. 4.6. The Least Asymmetric wavelet and scaling function filters



Best Localized MODWT Wavelet Filters

Best Localized MODWT Scaling Function Filters

Fig. 4.7. The Best-Localized wavelet and scaling function filters

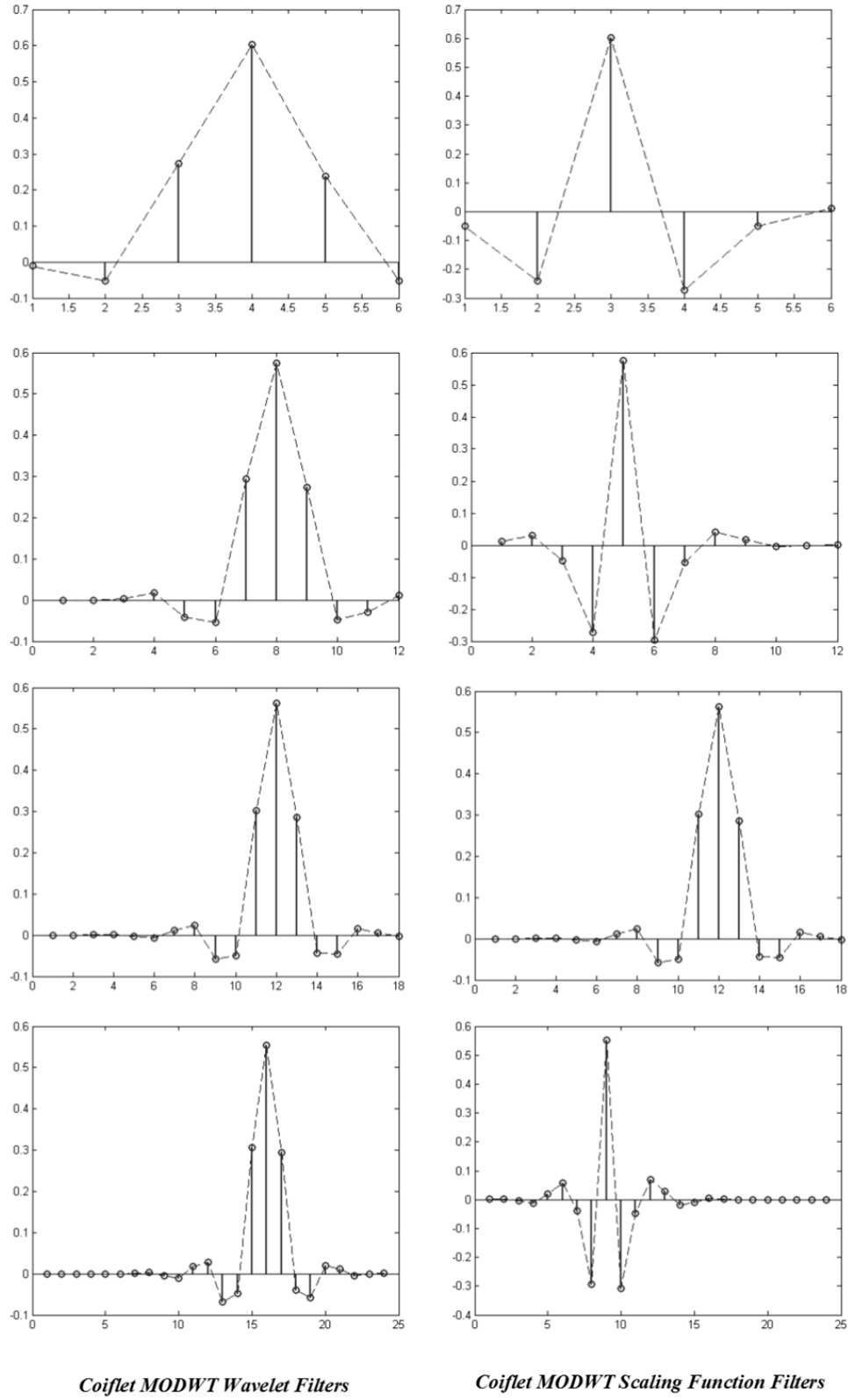


Fig. 4.8. The Coiflet wavelet and scaling function filters

4.2.3 Time series Alignment

Since neither wavelet filters, nor scaling filters at any level j are zero phase filters, the resulting output coefficients do not align with the original time series. Although this alignment is more of an implementation requirement, it is still worth elaborating on since it significantly affects the results in the feature extraction module. The corresponding phase functions of the MODWT filters are formulated as [28]

$$\begin{aligned}\theta_j^{(G)}(f) &\approx 2\pi f \nu_j^{(G)} \quad \text{and} \\ \theta_j^{(H)}(f) &\approx 2\pi f \nu_j^{(H)}\end{aligned}\tag{4.7}$$

In these formulas, $\theta_j^{(G)}(.)$ refers to the phase function for MODWT scaling function filter at level j , and $\theta_j^{(H)}$ to phase function for MODWT wavelet filter at level j . In addition, f denotes the frequency and finally, $\nu_j^{(G)}$ and $\nu_j^{(H)}$ signify “advance” for scaling function filter and wavelet filter respectively, and will be explained in continuation.

Since the phase functions in (4.7) are approximately linear, they can be changed to zero phase filters by an advancing operation [28]. The number of units $\{h_{j,l}\}$ and $\{g_{j,l}\}$ filters should be shifted at a specified level is defined as:

$$\begin{aligned}\nu_j^{(G)} &\equiv (2^j - 1)\nu \quad \text{and} \\ \nu_j^{(H)} &\equiv -[2^{j-1}(L - 1) + \nu]\end{aligned}\tag{4.8}$$

The value of constant ν depends on the wavelet basis. For example, for the least asymmetric wavelets, the value of ν is equal to:

$$\nu = \begin{cases} -\frac{L}{2} + 1, & \text{if } L = 8, 12, 16 \text{ or } 20; \\ -\frac{L}{2}, & \text{if } L = 10 \text{ or } 18; \\ -\frac{L}{2} + 2, & \text{if } L = 14. \end{cases} \quad (4.9)$$

For the Best-Localised wavelets, ν is equal to:

$$\nu = \begin{cases} -5, & \text{if } L = 14; \\ -11, & \text{if } L = 18; \\ -9, & \text{if } L = 20. \end{cases} \quad (4.10)$$

And for the Coiflet filters ν is equal to:

$$\nu = -\frac{2L}{3} + 1 \quad (4.11)$$

It should be noted that using a filter in which coefficients have been advanced circularly by ν units, corresponds to advancing the filter output by the same ν units [28]. Therefore, by doing the above calculations, how much each transformed time series at a specified level should be shifted is also known. In this study, these adjustments were made by the shifting operation, prior to commencing the feature extraction phase.

4.3 Feature Extraction Method

According to [4], for modeling the v - i characteristic of an electric arc furnace, some information about the past values of the voltage signal is necessary. This is because modeling the v - i characteristic can also be viewed as modeling the arc voltage in terms of the arc current and obviously when modeling any time series (the arc voltage being no exception) the past measurements are also required. To fulfill this requirement, recent past values of the voltage signal, at different resolution levels were used. The term “recent” here indicates that these transformed

signals were of relatively short memory. This means that the wavelet transform yielded signals of lower auto-correlation values.

Following this, it was necessary to determine how many and which coefficients at each resolution level should be considered to allow for thorough representation of the voltage signal. Renaud et al.'s approach [29] to time series forecasting was incorporated into this study, where the scaling function and wavelet coefficients are suggested to be of the form $\tilde{V}_{J,t-1-2^j(k-1)}$ and $\tilde{W}_{j,t-1-2^j(k-1)}$. It is important to consider that $k = 1..A_j$, therefore, the values of k are all positive.

Finally, the question under discussion was the value of A_j , for each of the resolution levels. If the value of A_j is determined carefully, this would be a sparse and at the same time a complete picture for all the information embedded in the series resulted from the wavelet transform [57]. That is to say, these few coefficients can be used for almost perfect recovery of the voltage signal at any point in time.

On the condition that the length of the time series is a power of two, the features of the form $\tilde{V}_{J,t-1-2^j(k-1)}$ and $\tilde{W}_{j,t-1-2^j(k-1)}$, that are selected using dyadically lagged coefficients of the non-decimated transform, are, in fact, the coefficients of the decimated form of wavelet transform [28].

As has already been mentioned, at any time t , the arc voltage not only depends on its own previous values, but also on the corresponding arc current. Given the latter point, modeling the arc voltage in terms of the arc current at time t required the following features: the past values of voltage, transformed in the wavelet domain, in the form of $\tilde{V}_{J,t-1-2^j(k-1)}$ and $\tilde{W}_{j,t-1-2^j(k-1)}$ and the current at time t . The modeling scheme, in which this feature set (see Fig. 4.9) was used, is referred to as the first modeling scheme hereafter.

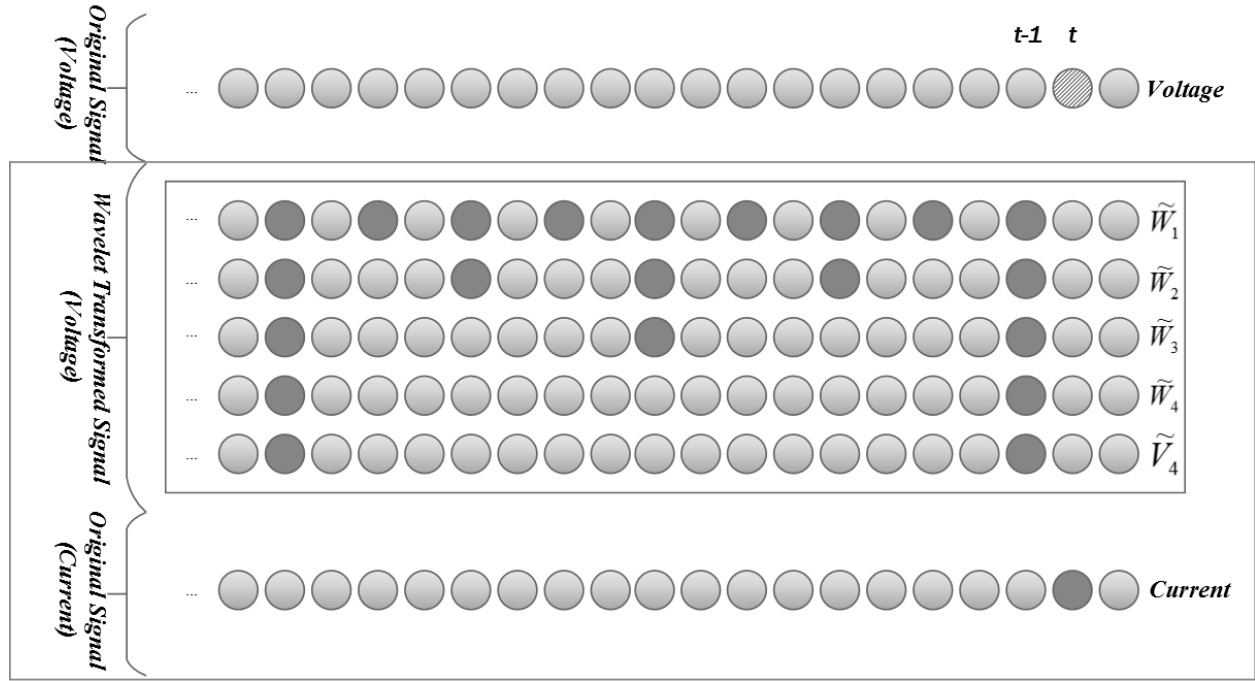


Fig. 4.9. Features used in modeling $v-i$ relationship in the first scheme

For a more comprehensive study, another feature set was created, by also transforming the current signal into the wavelet domain and extracting features based on the same formula. This can be justified on the grounds that the arc voltage also depends on previous values of the arc current. In this way, although indirectly, the present value of the current as well as its previous ones was taken into consideration. The modeling scheme, in which this feature set (see Fig. 4.10) was used, is referred to as the second modeling scheme hereafter.

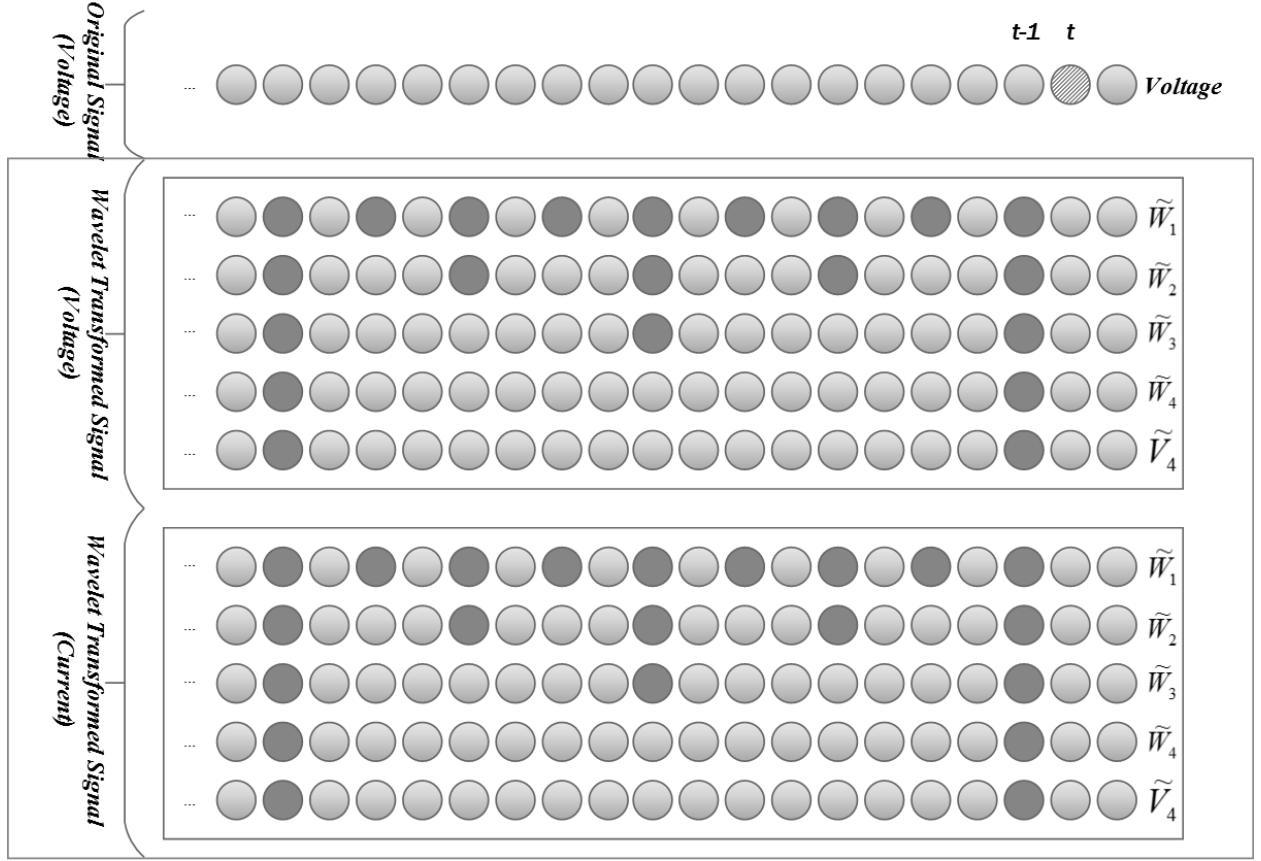


Fig. 4.10. Features used in modeling $v-i$ relationship in the second scheme

4.4 Hybrid Neuro-Wavelet Modeling Method

Artificial Intelligence (AI) methods especially neural networks have shown promising results in modeling problems. The neural networks are among supervised techniques that have long been used to deal with the nonlinearity in data and have produced satisfactory results throughout the literature. Wavelet transform, on the other hand, allows for the study of non-stationarity behaviour in data. Therefore, combining wavelet transform and neural networks seemed to be extremely appropriate in analysing the voltage and current signals in electric arc furnace, that are both

nonlinear and non-stationary and hence, gave us more power on modeling the complex, nonlinear relationships involved in arc furnace $v-i$ characteristic.

In [4], the use of nonparametric methods, such as neural networks and Radial Basis Functions (RBF), are justified to approximate any real continuous function to any particular level of accuracy. Considering this, along with the fact that the signal can be constructed from its wavelet coefficients, the foundation of our modeller used Multi-Layer Perceptron (MLP), which is the simplest form of neural networks. This MLP was fed with the selected coefficients in the previous section as inputs and the voltage at time t as the output. The complex relationships in the $v-i$ characteristics of the electric arc furnace were then modeled by learning from samples in the training data set, which consisted of the past voltage, the past current and the present current behaviour of the EAF.

For the purpose of designing such an MLP, a number of parameters were decided upon. In particular, the optimum number of hidden layer neurons was chosen. Although, the optimum number of hidden layer neurons in any neural network is highly problem dependent and a matter of experiment, a bottleneck structure was employed that is suggested by experience in [58]. According to this recommendation, there should be fewer neurons in the hidden layer than nodes in the input layer.

The architecture of such an MLP is shown in Fig. 4.11 and Fig. 4.12, for the first and second modeling scheme respectively. As can be seen from these figures, sigmoid transfer functions were used for L hidden layer neurons. Whereas, for the output layer neurons the simple linear function was employed.

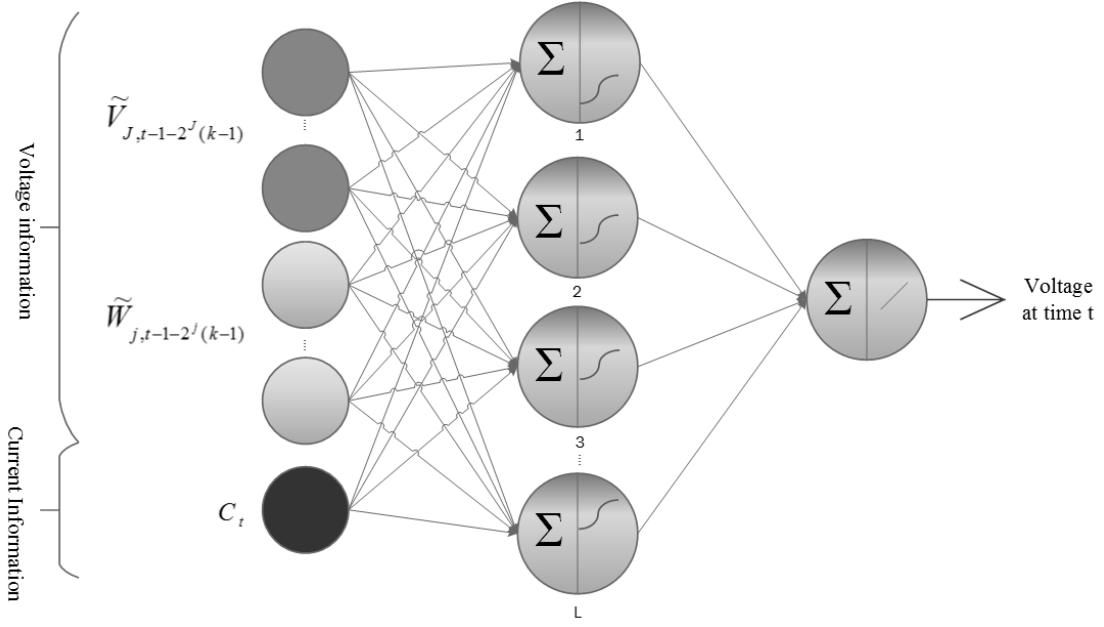


Fig. 4.11. The proposed neural network structure for modeling $v-i$ relationship in the first scheme

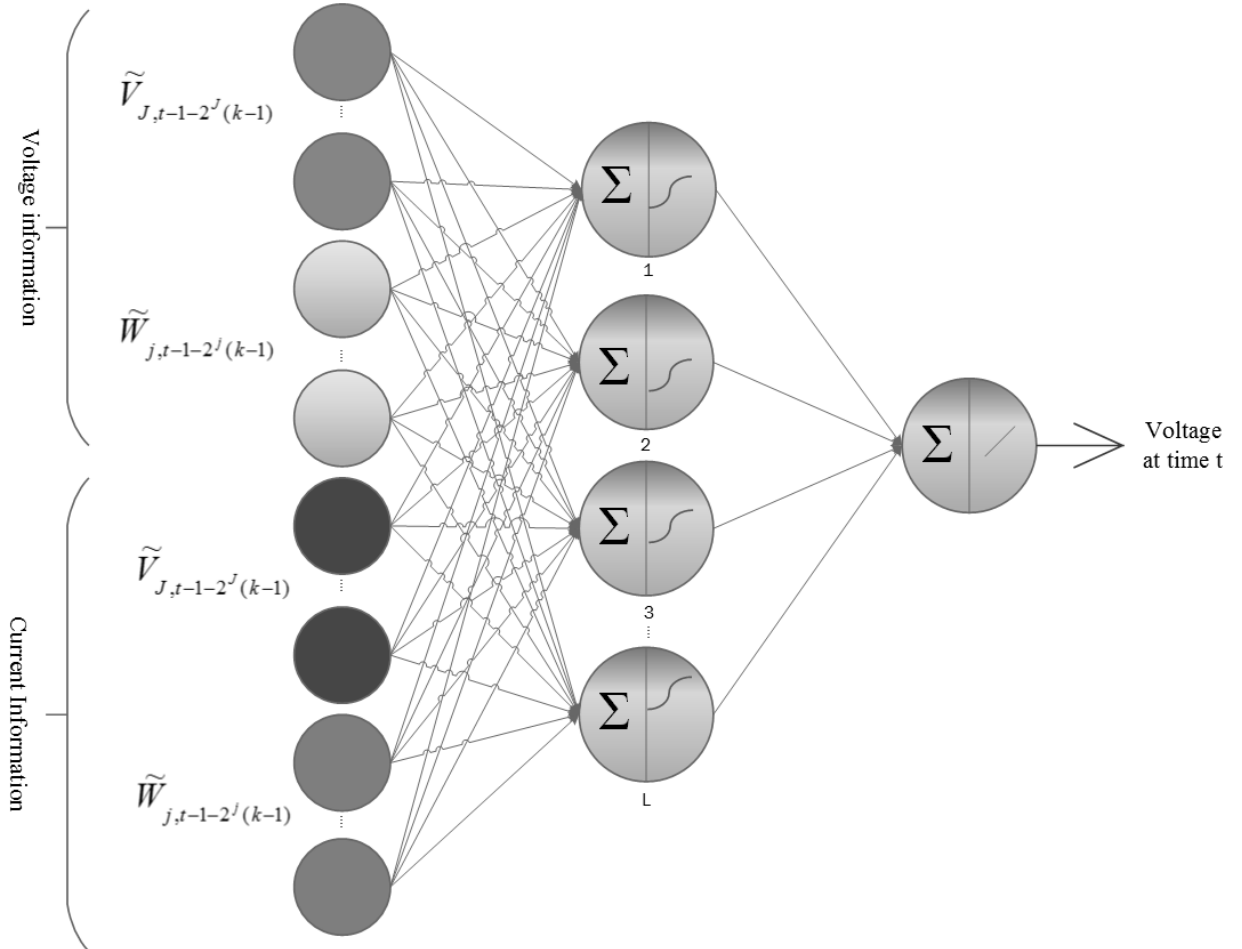


Fig. 4.12. The proposed neural network structure for modeling $v-i$ relationship in the second scheme

The selection of input layer nodes was mainly dependant on the decision about which and how many past coefficients to use for modeling. We have followed the basic guideline of Occam's Razor principle, which prefers simple models [58]. The fewer the input nodes, and, as a result, fewer weights in the network, the greater is the confidence that overtraining does not happen [58].

The general form of the relevant coefficients was formulated in the previous section. However, the values of A_j still remained unknown and had to be determined in order for the most relevant coefficients to be selected. On one hand, for large values of A_j , the selected coefficients could act as orthogonal bases capable of representing all the past values, and in a way contained all the information on the decomposed series [57]. However, this was not efficient and had the drawback of increasing the problem dimensionality. Moreover, not all these coefficients were equally relevant to the output (the voltage). As a result, the possible random correlations between these rather irrelevant coefficients and the voltage signal could make it hard for our neural network to set the weights for less useful inputs to zero. Consequently, this irrelevant data could affect the model's performance adversely [58].

On the other hand, small values for A_j could make the training of neural network simple, but the information about the past values of voltage could be insufficient to model accurately. The value of A_j was required to be optimum to result in the smallest possible number of coefficients and hence inputs for the neural network.

The preceding problem can also be looked at from another view. For the purpose of learning, the training samples were to be created using the sliding window technique. Thus, it was necessary to decide upon the size of the window for each of the series at different resolution levels.

With the intention of estimating A_j , each of the obtained time series at different resolutions levels were assumed to be an Auto Regression (AR) process. This means that every single element in these time series depends linearly on its own previous values. The number of these previous values specifies the AR model order. From this point of view, the challenge shifted from finding the appropriate value of A_j at each resolution level, to determining the model order of each time series at different resolution levels.

There are a number of metrics to measure the order of an auto-regression model. In this thesis, three well-known metrics of Akaike Information Criterion (AIC), Akaike Information Criterion, Corrected (AICc), and Bayesian Information Criterion (BIC) were considered. However, only the BIC metrics was used to estimate the value of A_j on each scale. These metrics are explained in detail in the next section.

The next thing to be taken into consideration was the total number of required training samples. A rough guideline, based on theoretical consideration of the Vapnik-Chervonenkis (VC) dimension was followed. This guideline recommends that the total number of training samples should be ten times or more the total number of weights [58]. Consequently, a large enough training data set was created in order to have the added advantage of containing the entire system dynamics and hence, the ability to model the system more accurately.

As a final point, it should be mentioned that among all the possible approaches, the Levenberg-Marquardt back propagation algorithm was used for the learning process and obtaining the optimized weights in the neural network.

4.4.1 Information Criteria

In this section, a number of information criteria, namely, AIC, AICc and BIC for comparing statistical models are introduced. Besides, the justification for the choice of BIC in this study is provided.

The first criterion is AIC, which is a measure of the relative quality of a statistical model [59], describing a particular data. Although, AIC provides a means to make selection decisions about the model, it cannot comment on the absolute quality of the model [59]. AIC is based on information theory as it relatively estimates the information loss when a particular model is used for representing the underlying data behind a system [59]. This measure is formulated as:

$$AIC = \log \hat{\sigma}_k^2 + \frac{n + 2k}{n} \quad (4.12)$$

where $\hat{\sigma}_k^2$ is the noise variance and is given by

$$\hat{\sigma}_k^2 = \frac{SSE_k}{n} \quad (4.13)$$

where SSE denotes the sum squared error, k is the number of parameters in the model, and n is the sample size. The value of k , that results in the minimum AIC specifies the best model [60].

Although aiming for a minimum value for $\hat{\sigma}_k^2$ seems to be reasonable, it decreases continually as the model order (k) increases. In other words, even though the goodness of fit of the model increases by extra parameters (increasing the number of feature coefficients), so does the complexity of the model and also chances of over-fitting of the network to which these coefficients are fed. Therefore, the noise variance needs to be penalized by a term proportional to the number of parameters. This penalty term is by no means unique, and an extensive literature on different possible penalty terms is available [60].

In view of the fact that the relative estimate of information loss through AIC is only valid asymptotically (and not for finite sample size), some correction is required when the number of data points is small [60]. This is the rationale behind introducing another metric, namely AICc.

The corrected form of AIC is defined as follows:

$$AICc = \log \hat{\sigma}_k^2 + \frac{n+k}{n-k-2} \quad (4.14)$$

Based on this formula, AICc imposes a greater penalty for extra parameters in comparison with AIC, and tries to balance model fit and parsimony properly.

A correction term based on Bayesian arguments can also be derived, which leads to the Bayesian information criterion. This criterion penalizes differently from AICc for the number of parameters and is formulated as:

$$BIC = \log \hat{\sigma}_k^2 + \frac{k \log n}{n} \quad (4.15)$$

The Bayesian information criterion also imposes heavier penalty on the number of parameters than does the AIC, and therefore favors more parsimonious models. With regard to our preference for simpler models, the BIC seemed to be a better choice in this study. Furthermore, in various simulation studies [60] it has been verified that BIC performs better in case of larger samples (which was the case with our signals), whereas AICc is superior in finding the correct order in case of smaller samples where the number of parameters is large in comparison.

5 Experimental Results

The objective of our modeling procedure, as described earlier, is to find an accurate mapping, f_r , by means of which the arc voltage can be acquired once the corresponding arc current is known. In [4] this mapping has been defined as:

$$\begin{aligned} f_r : R^3 &\rightarrow R \\ V_{EAF}(t) &= f([V^{t-1}_{EAF}, di^{t}_{EAF} / dt, i^{t}_{EAF}]) \end{aligned} \quad (5.1)$$

That is, the voltage at any time, not only depends on the current at that precise time, but also on its own previous value. In this work, keeping the last element intact, the mapping is modified to:

$$\begin{aligned} f_r : R^3 &\rightarrow R \\ V_{EAF}(t) &= f([\tilde{V}_{J,t-1-2^j(k-1)}, \tilde{W}_{j,t-1-2^j(k-1)}, i_{EAF}(t)]) \end{aligned} \quad (5.2)$$

where \tilde{V}_j refers to the vector of the J^{th} level scaling function coefficients, and \tilde{W}_j to the vector of the j^{th} level wavelet coefficients of the voltage signal. Thus, this mapping uses the first feature set as defined in section 4.3 and depicted in Fig. 5.1. Mapping (5.1) is also modified to:

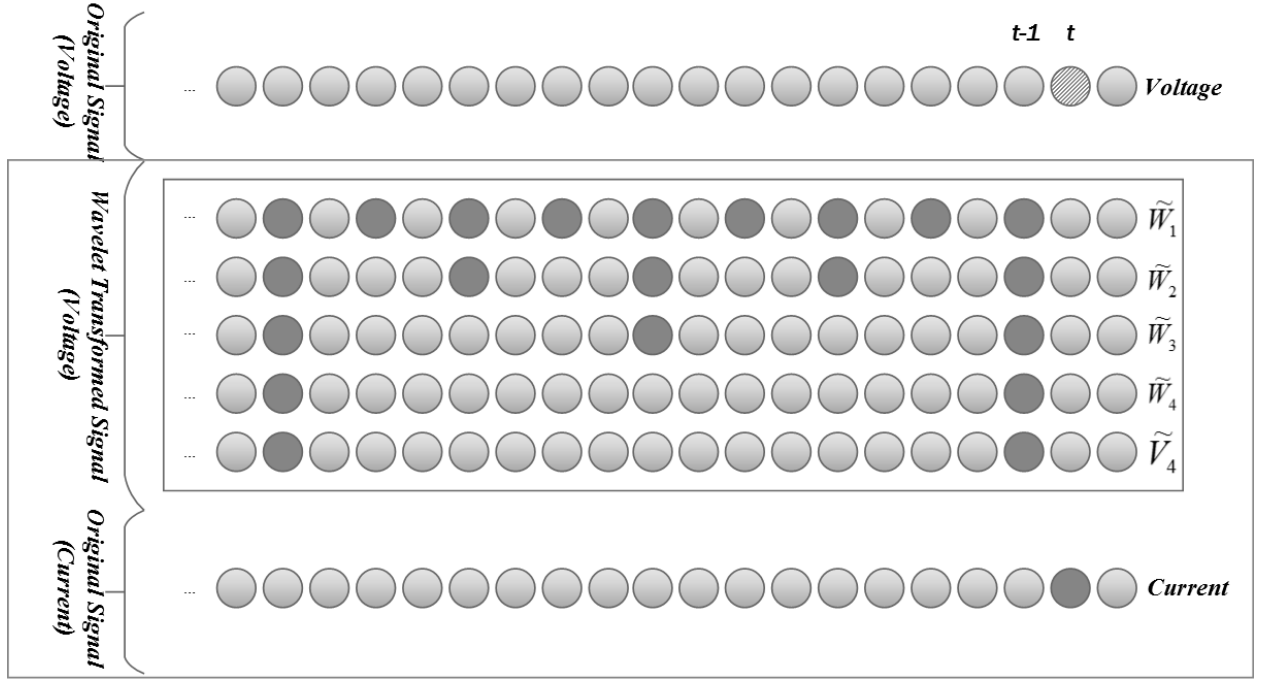


Fig. 5.1. The first feature set

$$f_r : R^4 \rightarrow R$$

$$V_{EAF}(t) = f([\tilde{V}_{J,t-1-2^j(k-1)}, \tilde{W}_{j,t-1-2^j(k-1)}, \tilde{V}'_{J',t-1-2^{j'}(k'-1)}, \tilde{W}'_{j,t-1-2^j(k'-1)}]) \quad (5.3)$$

where \tilde{V}_j and \tilde{W}_j are as defined in (5.2), and \tilde{V}'_j signifies the vector of the J'^{th} level scaling function coefficients, and \tilde{W}'_j signifies the vector of the j^{th} level wavelet coefficients of the current signal. Therefore, this mapping uses the second feature set, which is defined in section 4.3 and depicted in Fig. 5.2.

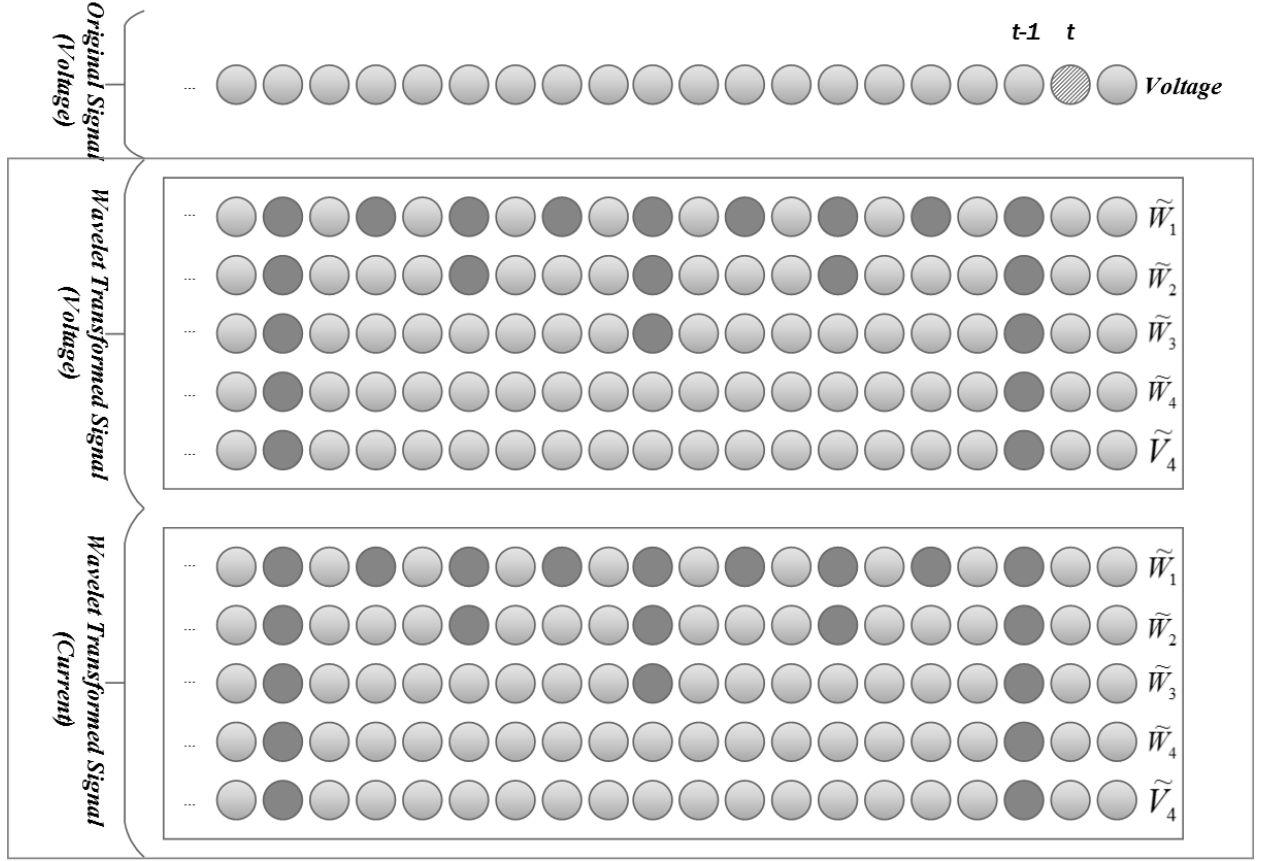


Fig. 5.2. The second feature set

The decision was made to use feed-forward neural networks as the function capable of mapping nonlinear relationships, mainly due to their universal approximation capability and also, their superiority in finding complex and nonlinear relationships among data. Consequently, in our proposed modeling procedure, vectors on the right-hand side of the equation (5.2) and (5.3) are fed into a feed-forward neural network with the intention to approximate the EAF voltage on the left-hand side.

5.1 Parameter selection

To implement neural network based models, a number of decisions on different parameters had to be made. These decisions can be summarized as follows:

- The Multi-Layered Perceptron (MLP) networks used here have two fully connected layers. Since the network has only one output, the output layer has one neuron that is taken to be linear. However, the hidden layer can have as many neurons as desired. The number of hidden neurons can be either found by trial and error or estimated by empirical guidelines. For example, it is suggested that the optimum number of neurons is estimated as [61]:

$$h_{start} = \max\left\{ \frac{(Number\ of\ Inputs + Number\ of\ Outputs)}{2}, \sqrt{Number\ of\ Inputs \times Number\ of\ Outputs} \right\}, \quad (5.4)$$

- The activation function in the hidden layer neurons is taken to be sigmoid.
- For training purpose, the minimum number of data samples was determined by the suggestions in [61]. These suggestions recommend the number of data samples created for training, which is only 70% of the total samples:

$$D_{min} = h_{start} \times (Number\ of\ Inputs + Number\ of\ Outputs) \quad (5.5)$$

where D_{min} and h_{start} respectively signify the minimum data samples required for training and the optimum number of hidden layer neurons.

- 70% of the samples were designated for training, 15% for validation and 15% for testing. In terms of the selection of training, validation and testing data, a random method has been used; which means that the training data samples were selected randomly from the available measurements in the dataset.

- The Levenberg-Marquardt backpropagation algorithm is used for the training purpose, which is fast and suitable for training of small and medium sized problems. This algorithm is a blend of steepest decent method and the Gauss-Newton algorithm and has the advantages of both.
- The training process is stopped as soon as the validation error fails to decrease for more than six epochs as the training error keeps decreasing. This is done to avoid overtraining. Overtraining refers to presenting information to a neural network too many times, which can cause the network to memorize the data that was presented to it, instead of learning the relationships among inputs and outputs. If a network is over trained, it may not respond well to new situations.
- The performance criteria for evaluating any neural network based model can be network size (in terms of number of hidden layer neurons or input size) or accuracy. In order to investigate the accuracy of the neuro-wavelet based models, the actual measured data is compared with the neural networks' output. The error metric used for the comparison purpose is the Non-Dimensional Index Error (NDIE) that is obtained by calculating the root mean squared error of EAF modeling divided by the standard deviation of target EAF voltage. This error index was selected to allow for comparisons with the previous works in [4] and [15].

5.2 Results and Discussion

In this section, to assess the validity of the proposed method, four sets of experiments are conducted. All the simulations in these experiments are run on a system with Windows server 2008 operating system, 4 Intel® Xeon® CPUs of 2.50 GHz and 32 GB of RAM.

While it is not our primary objective, these experiments also examine the effect of changes in the number of decomposition levels of both voltage and current signals and the choice of wavelet basis on the accuracy of the models.

5.2.1 The First Experiment

In the first set of experiments, the simplest external phase wavelet in the Daubechies family, Haar, is used. This wavelet is depicted in Fig. 5.3 and its main properties are summarized in Table 5.1.

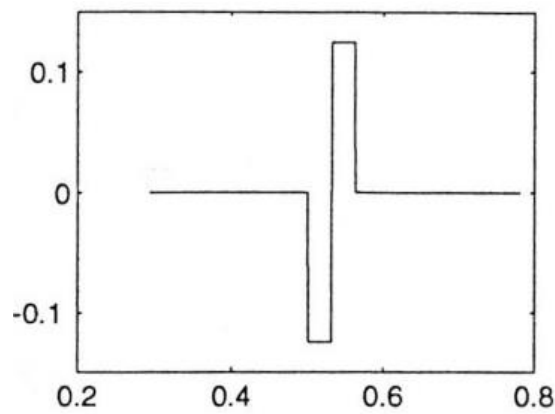


Fig. 5.3. The Haar wavelet

Table 5.1. Properties of the Haar wavelet

Property	Value
Family	Daubechies
Orthogonal	Yes
Biorthogonal	Yes
Compact Support	Yes
Support Width	1
Filter Length	2
Regularity	Haar is not continuous
Symmetry	Yes

Number of Vanishing Moments for psi (wavelet filter)	1
--	---

In order to estimate the optimum number of decomposition levels for accurate modeling, the coefficients at different resolution levels are inspected. It is observed that decomposing the voltage signal more than five levels does not add any further information. This is because the coefficients became zero after five levels. As a result, an upper bound of five decomposition levels is set for the voltage signal. This signal is then decomposed into 5, 4, 3, 2 and 1 level for further inspection. The results of these decompositions are illustrated in Fig. 5.4, Fig. 5.5, Fig. 5.6, Fig. 5.7 and Fig. 5.8. In all of these figures, the bottom most signal is the original arc furnace voltage, and others are the transformed voltage signal into the wavelet domain at different scales.

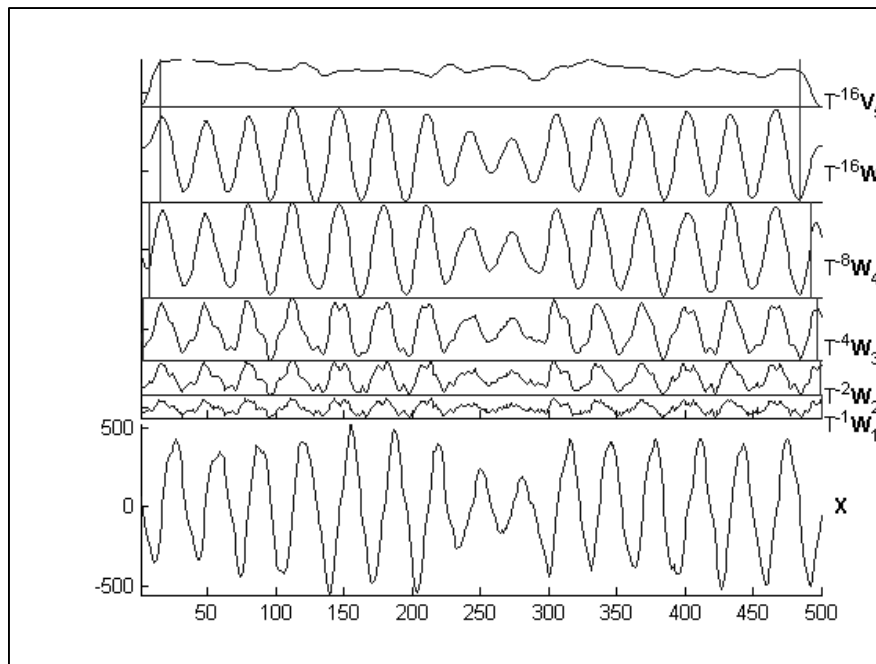


Fig. 5.4. The decomposed voltage signal into five levels of resolution

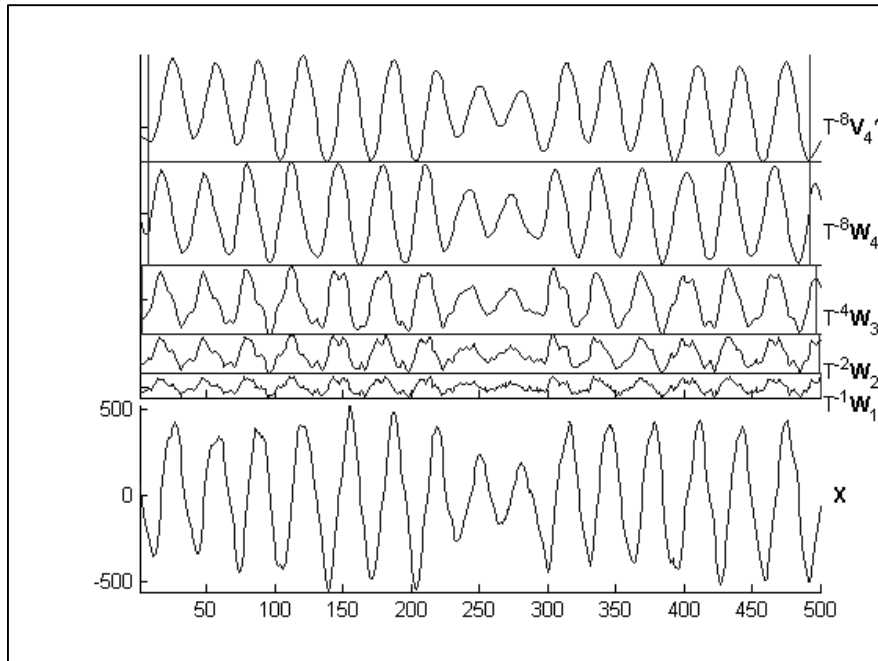


Fig. 5.5. The decomposed voltage signal into four levels of resolution

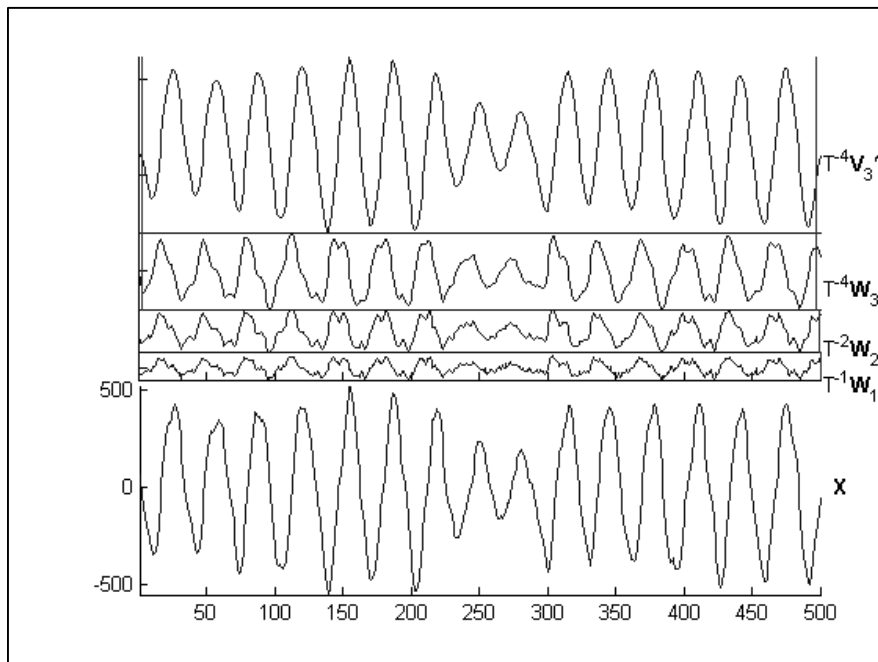


Fig. 5.6. The decomposed voltage signal into three levels of resolution

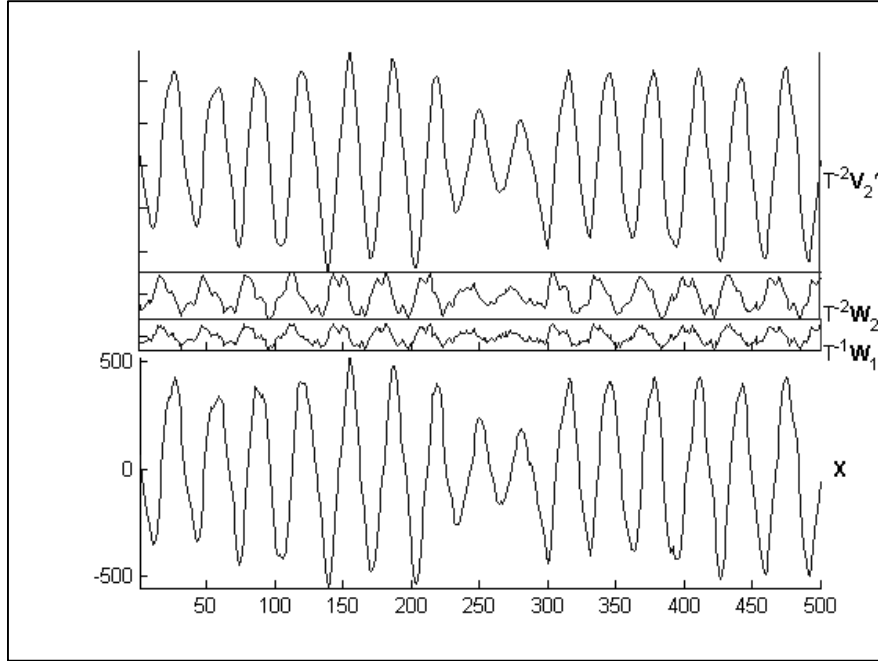


Fig. 5.7. The decomposed voltage signal into two levels of resolution

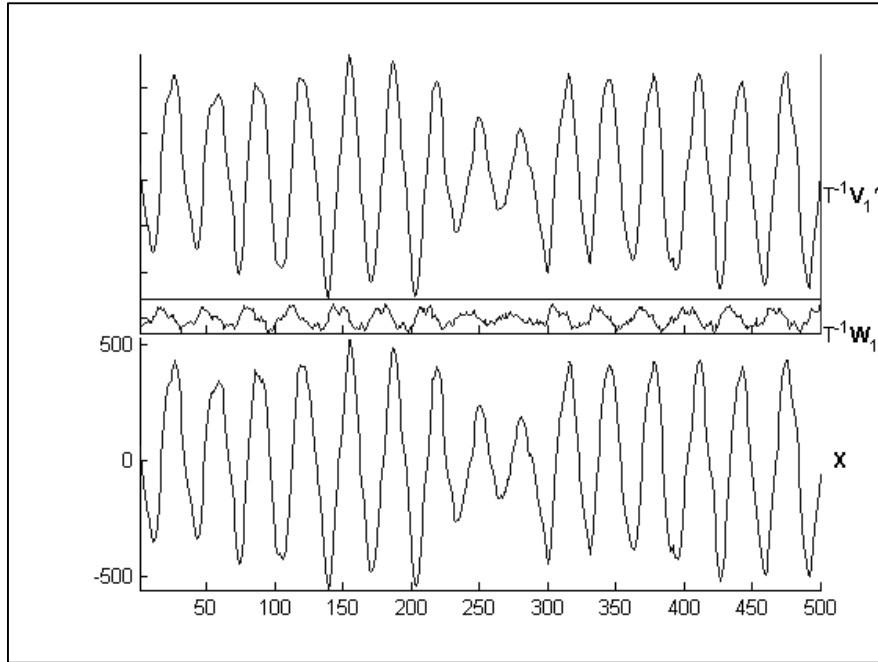


Fig. 5.8. The decomposed voltage signal into 1 level of resolution

Regardless of the decomposition level choice, the methodology dictates a number of features to be extracted from each transformed signal (or each scale). The BIC metric is used for this matter,

which can estimate the number of salient features at each resolution level. The procedure is explained in detail in continuation.

In the first place, it is important to note that the wavelet coefficients at each resolution level form an independent time series. To work towards the primary goal of determining the voltage at any time t , the wavelet coefficients at that specific time should be known. However, only coefficients at times before t are at our disposal. Given this scenario, it is important to determine the number of previous elements that affect the value of the transformed time series at any time t . In other words, the salient features at each resolution level are those on which the future coefficients are dependent. By definition, the “order” of time series specifies this number. Therefore, the order of each transformed time series, associated with a resolution level, should be calculated. However, calculation of the order of time series with an unknown nature is not a simple task and is still an open area of research in mathematics on its own. However, it has been mentioned before that the wavelet transform can simplify the structure of the original data. Therefore, it is assumed that the resulting time series, from the transformation, have noncomplex structures that can be defined by simple models or structures such as Auto Regression (AR). A standard method of determining the order of an AR process is utilizing the statistical metric of BIC. For this matter, a number of AR models are created for the data, with a presumed order. The BIC metric is then calculated for each of these models. The model that yields the smallest BIC is considered to be the best model and hence, the presumed order associated with that model is the best order for describing the underlying data.

In this study, to estimate the approximate order of each transformed time series, the BIC metric has been calculated for each order from 1 to 100. The order corresponding to the smallest BIC, which is the best order, is then identified.

Fig. 5.9 shows how BIC changes with different order values for the transformed voltage time series at the first resolution level. It can be seen that the quality of the models is improving by increasing the order from 1 to 25 as the BIC values are dropping. However, from that point on, adding extra parameters, which corresponds to increasing the model order, does not enhance the performance of the models.

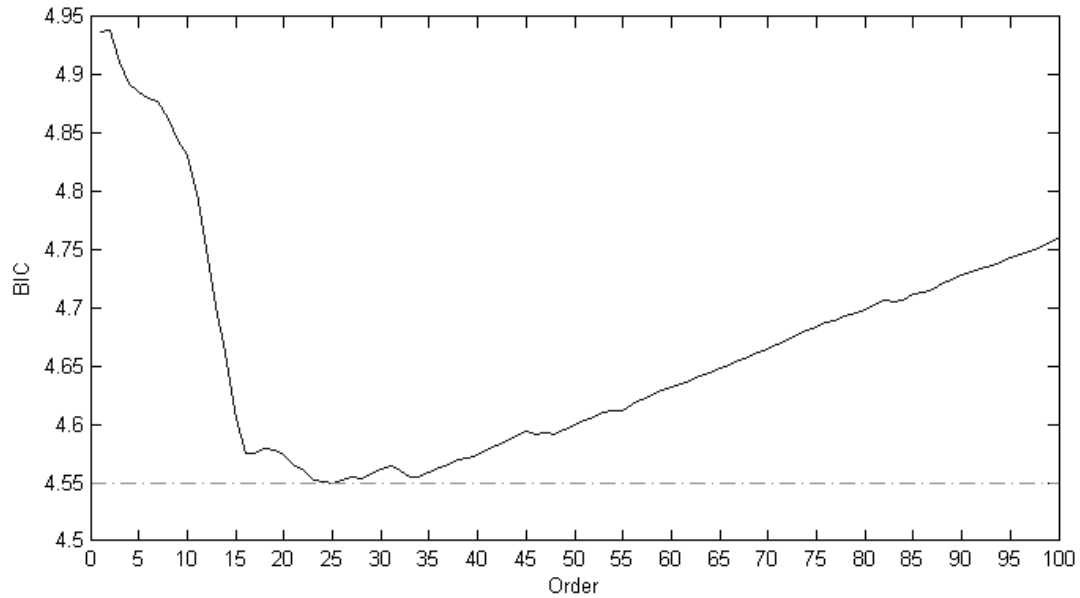


Fig. 5.9. The BIC value of the models created with different orders

According to this figure, from all the models generated for the transformed voltage time series at the first resolution level, the one of order 25 performs the best. In other words, the smallest BIC value is achieved at order of 25. Therefore, the best order for describing the first resolution level time series is 25. This means that 25 preceding samples affect the value of any element at this resolution level.

There is one more step towards determining the number of salient features, at each resolution level. It should be determined how many of the coefficients fall within this measured interval (the

windows of size 25 in the case of the first resolution level time series) according to the suggested form in section 4.3. In keeping with this form, the location (time wise) of the coefficients are formulated as $\tilde{V}_{J,t-1-2^j(k-1)}$ and $\tilde{W}_{j,t-1-2^j(k-1)}$ for the smooth and detail series respectively. More specifically, at the first resolution level, every other two coefficients within the designated windows should be selected, starting from the rightmost elements inside the windows. At the second resolution level, every other four coefficient, and for the third level every other eight coefficients should be selected and so on. The final results, according to this procedure, suggest extracting 12, 11, 6, 6, 3, 1 coefficients from $\tilde{W}_1, \tilde{W}_2, \tilde{W}_3, \tilde{W}_4$ and \tilde{W}_5 respectively.

After obtaining the required parameters, a number of different MLP networks were implemented. The results indicated that although by adding more than 5 hidden layer neurons, the training error decreased, the test error, which is a better indicator of the network's performance and its generalizability, did not. As a result, the difference between the two errors kept increasing until signs of over-fitting were observed. Over-Training is characterized by low training error and high testing error, which reduces the ability of the neural network to generalize to new situations [62]. Fig. 5.10 shows the training error versus test error in one of the experiments that employed the first modeling scheme, and in which the voltage was decomposed into five levels. The corresponding model in this experiment consisted of 40 inputs in total. This Figure suggests that over-fitting occurred by increasing the number of hidden layer neurons to 20 or more. In fact, as more and more hidden layer nodes were added, the network stopped learning the interactive relationships between inputs and outputs, and began to memorize the training data instead. Fig. 5.11 illustrates over-fitting during the training process of the neural network with 40 hidden layer neurons.

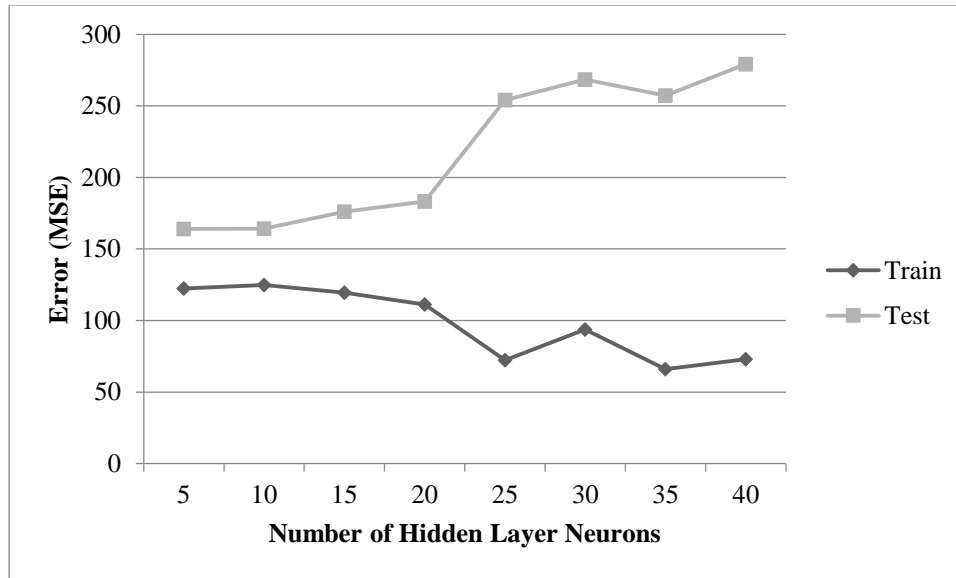


Fig. 5.10. Comparison of train and test errors having different number of neurons

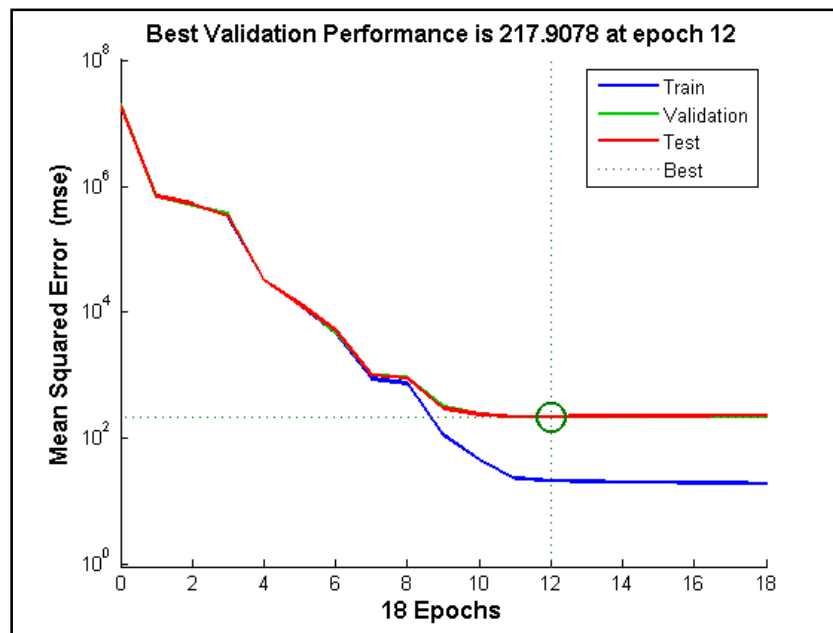


Fig. 5.11. Over-fitting occurring during the training process

The fact that a network with 40 inputs performed its best with only five neurons confirms the hypothesis that the structure of the transformed signals in the wavelet domain is simplified and is less non-linear. In other words, only five neurons in the hidden layer are sufficient for learning the training data and the underlying process behind it. This result conflicts with the suggested empirical guidelines in [61], which propose 20 neurons for such a network for convergence and generalization.

In the subsequent experiments, the first modeling scheme was used to test the effect of decomposition level of the voltage signal, on the performance. Numerous models were created using different numbers of neurons in the hidden layer (5, 10, 15, 20, 25, 30, 35 and 40), and the most accurate ones were selected for comparison. The error values for each of the selected models are shown in Table 5.2 along with other measures such as the number of hidden layer neurons involved in the creation of the model, training time and the number of epochs. The results are also illustrated in Fig. 5.12, which demonstrate that the best decomposition level for the voltage signal is 1.

Table 5.2. The results of Neuro-Wavelet models at different decomposition levels

Decomposition Level	Training/validation cases	Number of Hidden Neurons	Training Error %	Testing Error %	Time (s)	Epochs
1	2800/600	15	3.0364E-6	4.329E-6	1048	1000
2	2800/600	5	0.0076	0.0082	244	181
3	2800/600	5	0.017	0.02	15	18
4	2800/600	10	0.024	0.029	38	15
5	2800/600	5	0.042	0.048	21	15

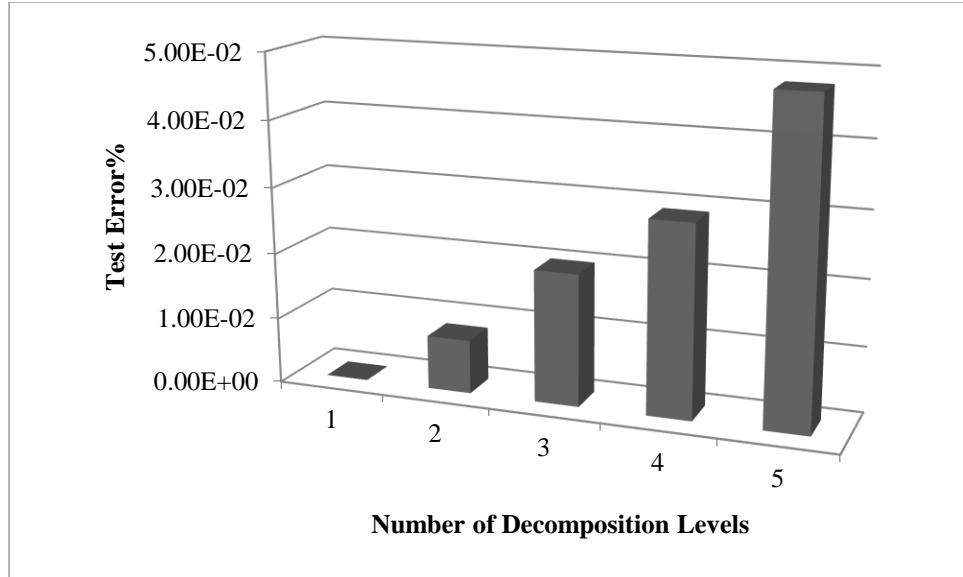


Fig. 5.12. Comparison of Neuro-Wavelet models at different decomposition levels

The actual measured EAF voltage versus the estimated voltage, using a model where the voltage is decomposed into only 1 level is depicted in Fig. 5.13. The response of this Neuro-Wavelet model to the test data and its error is also shown in Fig. 5.14.

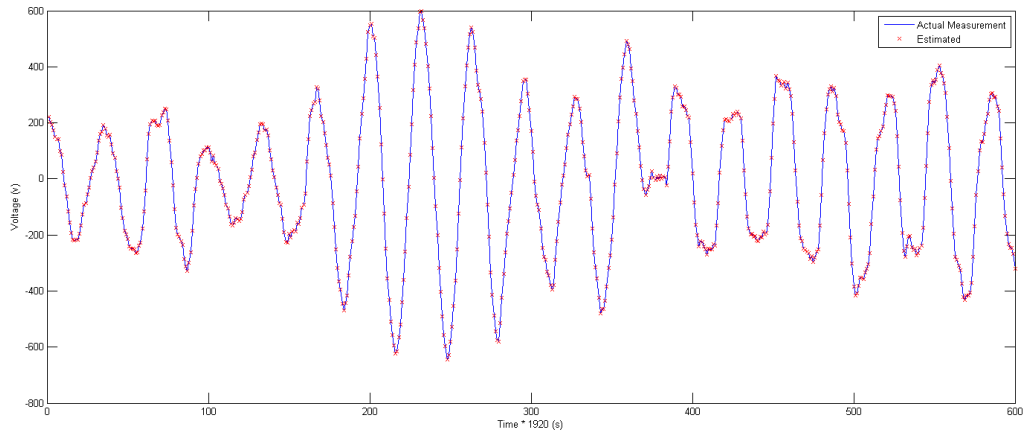


Fig. 5.13. The actual measured voltage vs. the estimated voltage

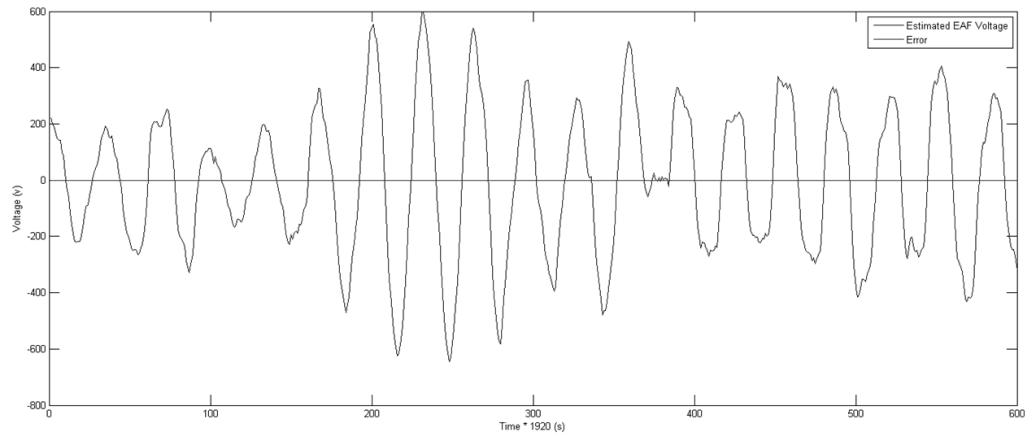
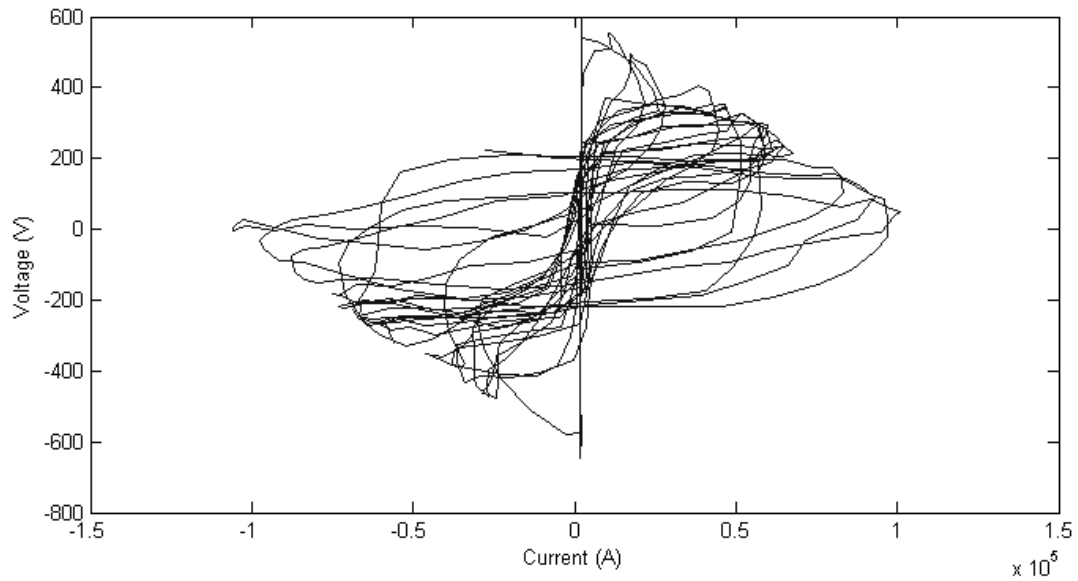


Fig. 5.14. The neuro-wavelet model response to the test data-case1: decomposition level=1

The actual and estimated EAF ($v-i$) characteristic, corresponding to this model is also shown in Fig. 5.15. The great resemblance between these two figures supports the reliability and accuracy of this model.



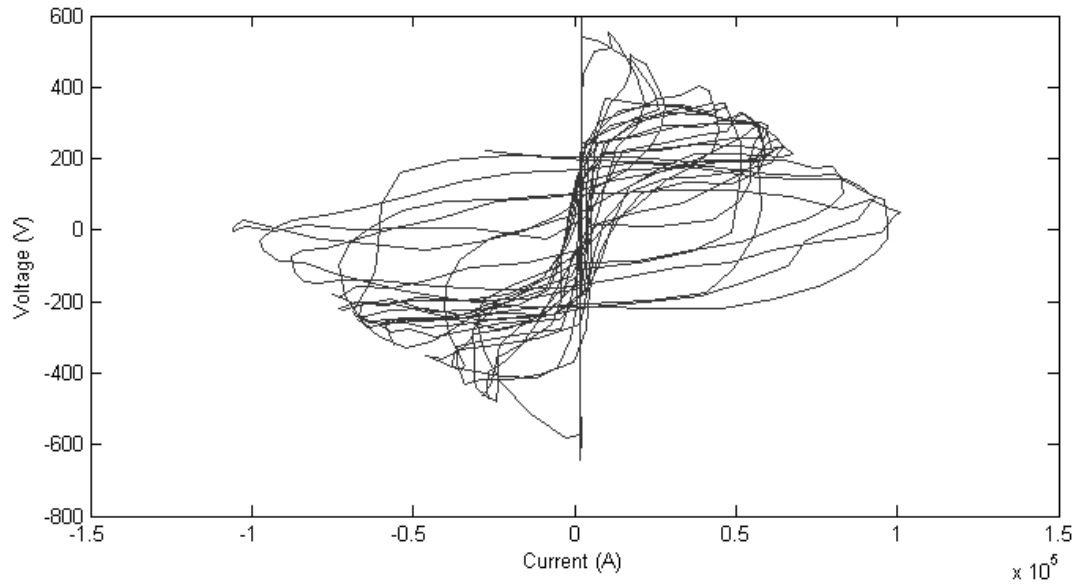


Fig. 5.15. The actual (top) and estimated (bottom) EAF ($v-i$) characteristic

Fig. 5.16 and Fig. 5.17 are the response of the models in which the voltage signal had been decomposed into 2 and five levels respectively. The difference between the error values in Fig. 5.14, Fig. 5.16 and Fig. 5.17 is not negligible and substantiates the importance of the decomposition level choice in wavelet analysis.

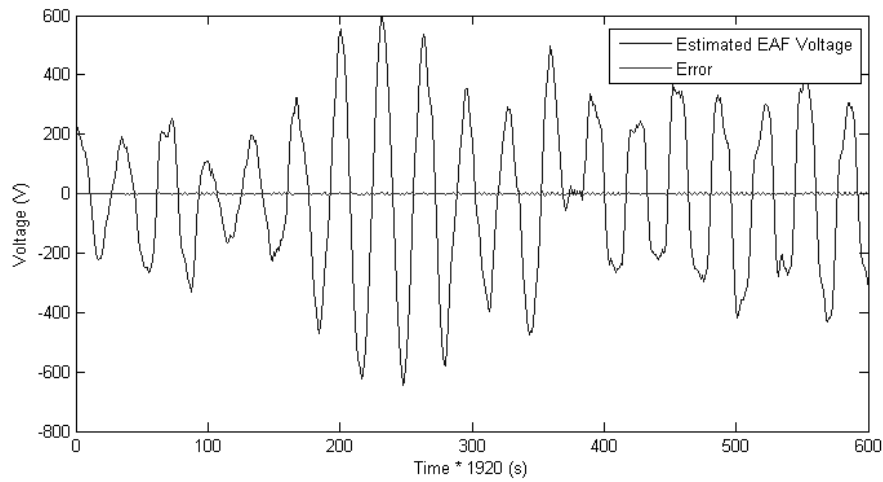


Fig. 5.16. The neuro-wavelet model response to the test data-case 2: decomposition level=2

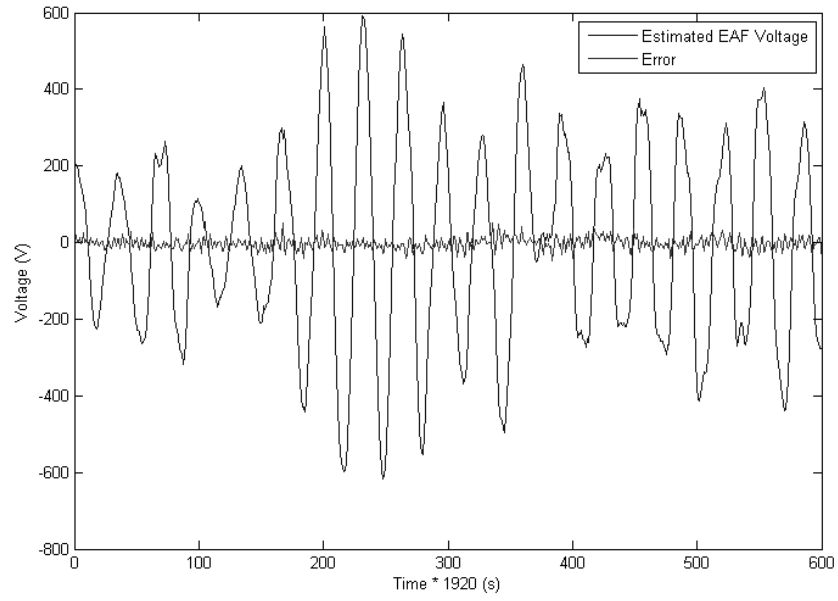


Fig. 5.17. The neuro-wavelet model response to the test data-case 3: decomposition level=5

5.2.2 The Second Experiment

The second set of experiments is designed for the purpose of evaluating the second modeling scheme, in which the feature set presented in Fig. 4.10 is used. Therefore, in these experiments, the EAF current signal has also been transferred into the wavelet domain, in accordance with the second modeling scheme. Fig. 5.18, Fig. 5.19, Fig. 5.20, Fig. 5.21 and Fig. 5.22 depict wavelet and scaling function coefficients of the current signal at different resolutions.

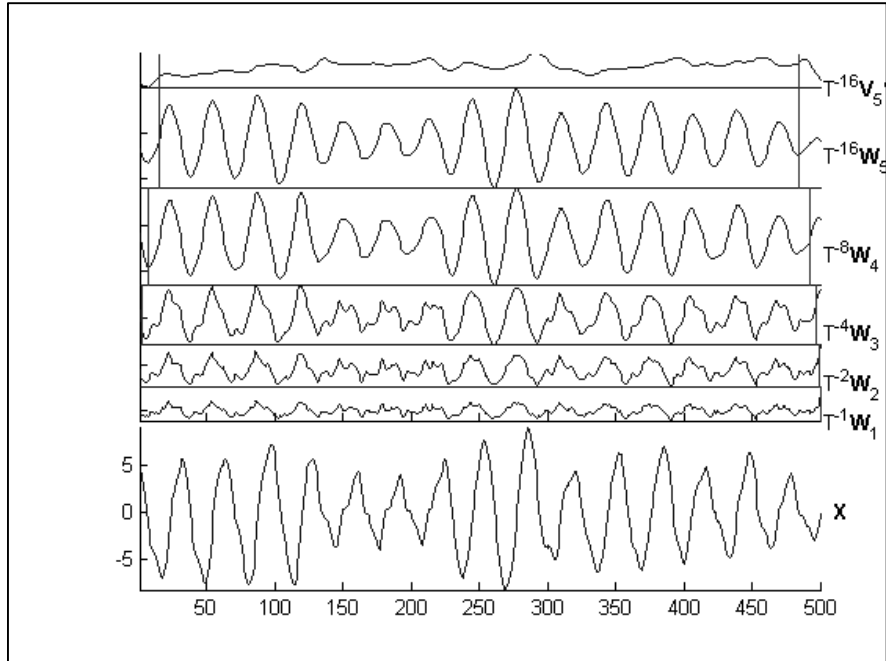


Fig. 5.18. The decomposed current signal into five levels of resolution

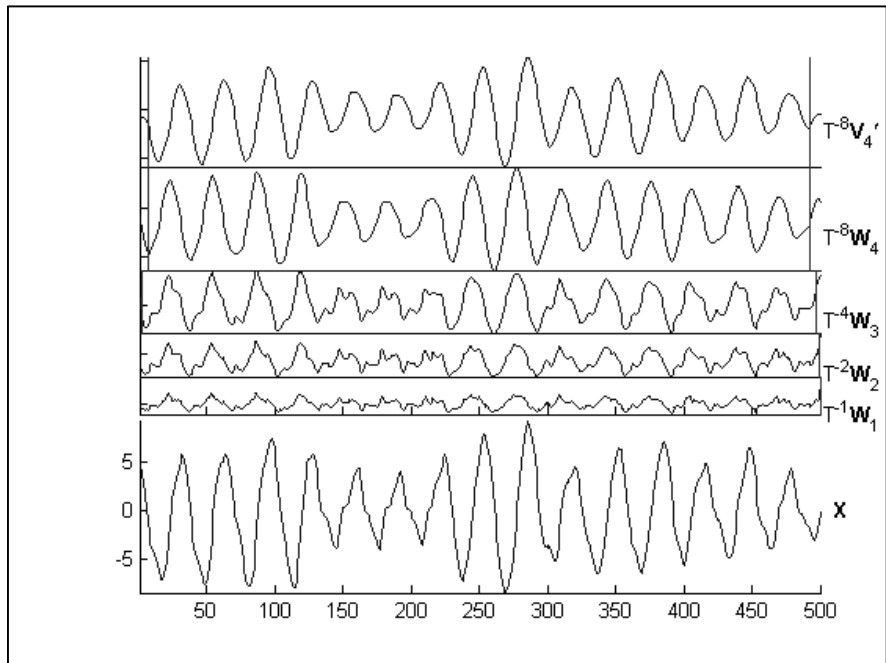


Fig. 5.19. The decomposed current signal into four levels of resolution

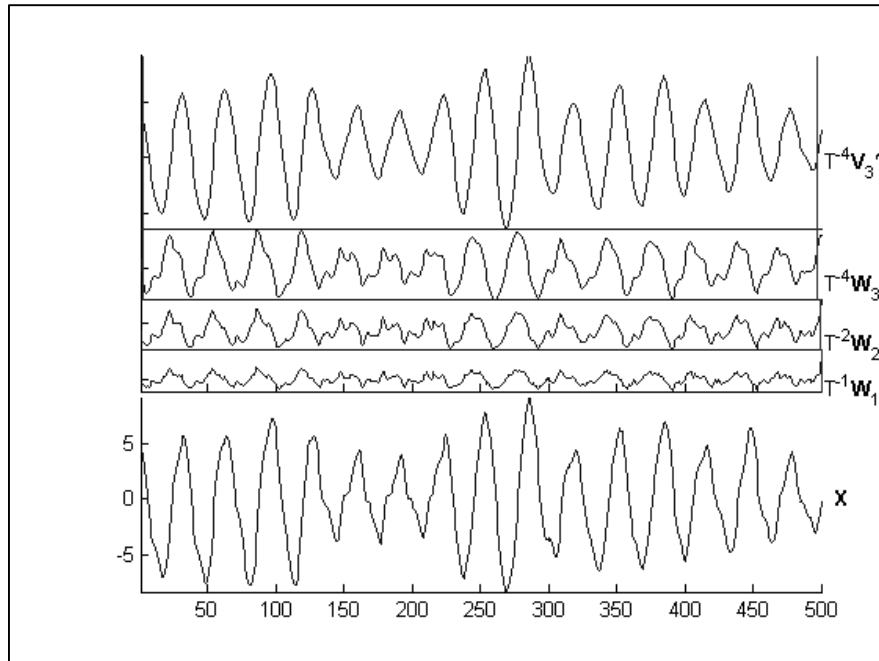


Fig. 5.20. The decomposed current signal into three levels of resolution

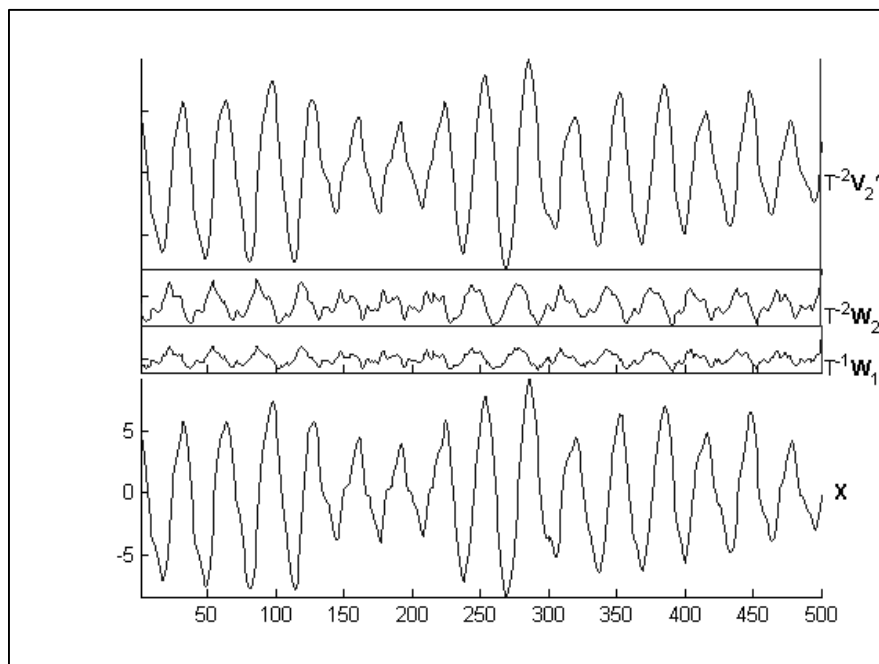


Fig. 5.21. Decomposed current signal into two levels of resolution

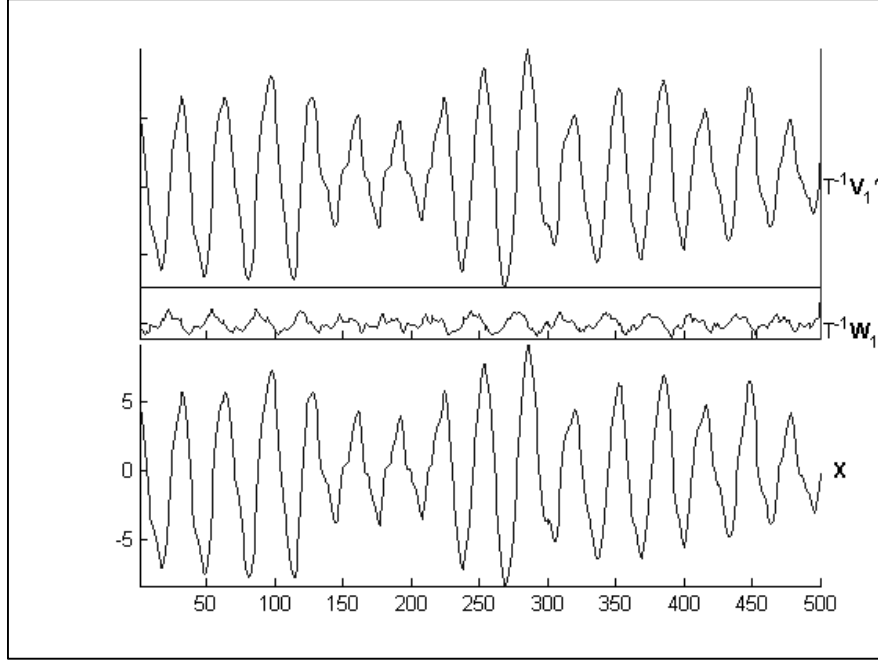


Fig. 5.22. Decomposed current signal into 1 level of resolution

The same logic underlies the sub-series of the transformed current signal. Hence, the number of salient features at each resolution level of the transformed current signal is also determined using the BIC metric, analogous to the previous experiment. This metric suggests extracting 14, 11, 6, 4 and three coefficients from $\tilde{W}_1, \tilde{W}_2, \tilde{W}_3, \tilde{W}_4$ and \tilde{W}_5 respectively.

Fig. 5.23 makes a comparison of the models in which the current signal was not decomposed (first scheme) and the models in which the current signal was decomposed (second scheme) into the same number of levels as the voltage signal. The numeric results are shown in Table 5.3.

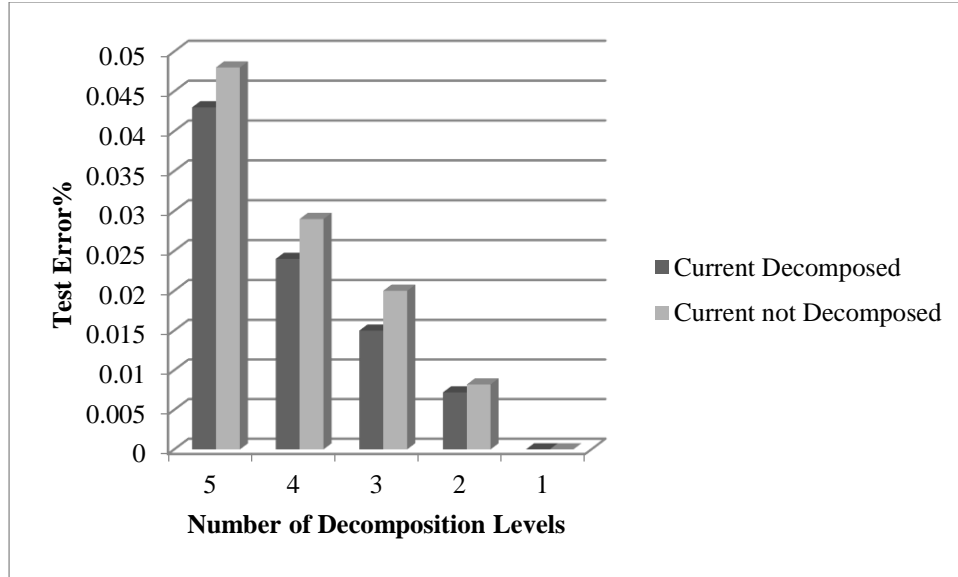


Fig. 5.23. Comparison of neuro-wavelet models under the first and second schemes

Table 5.3. The results of neuro-wavelet models under the first and second schemes

Status of Current	Level 1	Level 2	Level 3	Level 4	Level 5
Not Decomposed	4.329E-6	0.0082	0.020	0.029	0.048
Decomposed	2.98E-6	0.0072	0.015	0.024	0.043

These results reveal that in all cases, performance improved by up to 31%, by decomposing the current signal as well as the voltage signal.

5.2.3 The Third Experiment

With the knowledge that the optimum decomposition level for the voltage signal is 1, we conducted the third set of experiments. In these experiments, the performance of models with different decomposition levels of the current signals were evaluated, keeping the decomposition level of the voltage fixed at 1. A visual representation of the results is shown in Fig. 5.24, according to which the best performance was achieved when the current was decomposed into 3 levels. The numeric results are shown in Table 5.4.

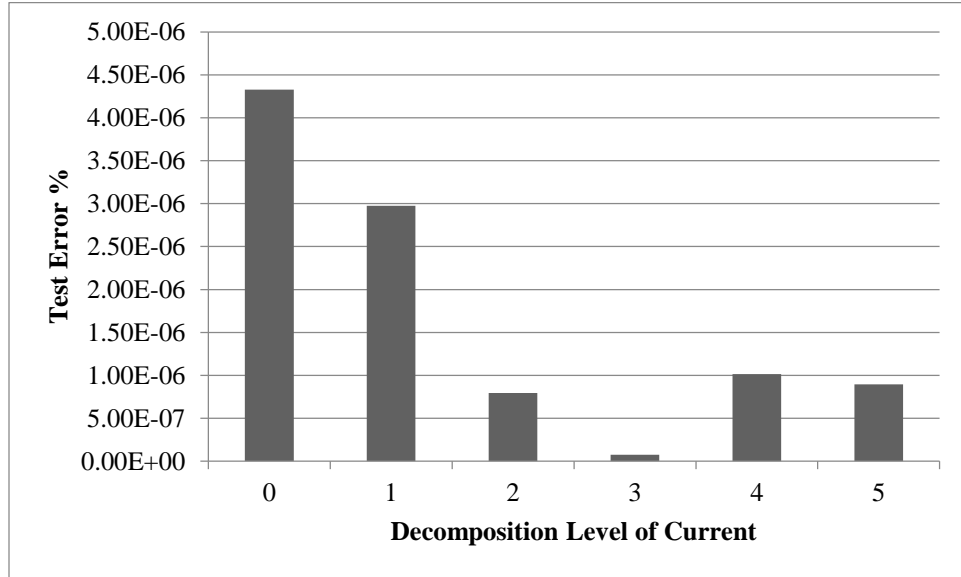


Fig. 5.24. The effect of the current decomposition level on performance

Table 5.4. The effect of the current decomposition levels on performance

Decomposition Level Current Signal	Error %
0	4.329E-6
1	2.98E-6
2	7.96E-7
3	7.68E-8
4	1.02E-6
5	8.96E-7

It can be concluded that the second modeling scheme is superior to the first scheme in yielding more accurate models. This is especially the case when an optimal decomposition level is chosen for voltage and current signals, which is not necessarily the same for both and is dependent on the intrinsic characteristics of each signal.

5.2.4 The Fourth Experiment

According to the Parseval's theorem, the energy of the original signal is preserved in its transformed coefficients [63]. The choice of wavelet basis determines whether the decomposition of energy in the transformed coefficients provides interesting information about the original signal or not [50]. That being said, it is important to decide upon a proper wavelet basis.

The last experiment was, therefore, designed to inspect the effect of the wavelet choice on the performance of models. There are an infinite number of possible wavelet bases [50]. However, the decision was made to select the first member of the Dabechies, Coiflet and Symlet family because they comply with the conditions that MODWT wavelet filters are subject to.

The Coiflets are characterized as compactly supported wavelets and therefore highly local in the time domain which have the greatest number of vanishing moments for both scaling function and wavelet function for the support width they have. Whereas, the Symlets are compactly supported wavelets that are least asymmetric and have the greatest number of vanishing moments for their support width, which is defined as the time interval, outside of which the wavelet is zero [64]. These wavelets are depicted in Fig. 5.25 and Fig. 5.26 and the more detailed properties have been summarized in Table 5.5 and Table 5.6 [64].

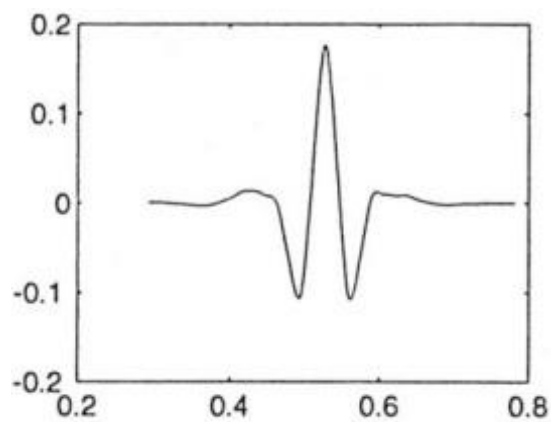


Fig. 5.25. The first member of the Coiflet family

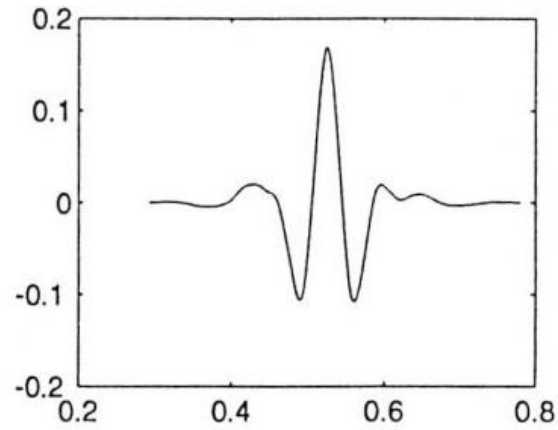


Fig. 5.26. The first member of the Symlet family

Table 5.5. Properties of the Coiflet wavelets

Property	Value
Family	Coiflets
Orthogonal	Yes
Biorthogonal	Yes
Compact Support	Yes
N values	1,2,3,4,5
Support Width	$6N-1$
Filter Length	$6N$
Symmetry	Near
Number of Vanishing Moments for ψ	$2N$
Number of Vanishing Moments for ϕ	$2N-1$

Table 5.6. Properties of the Symlet wavelets

Property	Value
Family	Symlets
Orthogonal	Yes
Biorthogonal	Yes
Compact Support	Yes
N values	2, 3, 4, ...
Support Width	$2N-1$
Filter Length	$2N$
Symmetry	Near
Number of Vanishing Moments for ψ	N

The number of features extracted from each resolution level is shown in Table 5.7 for the three different wavelet bases. Repeatedly, each number was determined using the same method explained in 5.2.1.

Table 5.7. Number of features extracted from each resolution level, using three different wavelets

Wavelet Family	Decomposition Level				
	Level 1	Level 2	Level 3	Level 4	Level 5
<i>Daubechies</i>	12	11	6	6	3
<i>Coiflet</i>	17	20	12	11	5
<i>Symlet</i>	30	36	18	9	4

In Table 5.8, the error percentage of the best models resulting from the three different wavelets, namely Haar, Symlet and Coiflet, are shown. All of these models were created under the first scheme and by decomposing the voltage signal into different number of levels, from 1 to 5. For more detailed information regarding these models, see Table 5.9 and Table 5.10.

Table 5.8. Comparison of performance of the models created by different wavelets

Wavelet Family	Level 1	Level 2	Level 3	Level 4	Level 5
Daubechies	3.95E-7	0.0082	0.02	0.029	0.048
Coiflet	0.012	0.029	0.04	0.055	0.059
Symlet	0.022	0.031	0.029	0.046	0.055

Table 5.9. The results of neuro-wavelet models at different decomposition levels, based on the Coiflet wavelet

Decomposition Level	Training/validation cases	Number of Hidden Neurons	Train Error %	Test Error %	Time (s)	Epochs
1	2800/600	10	0.011	0.012	120	71
2	2800/600	5	0.025	0.029	26	21
3	2800/600	10	0.032	0.04	83	29
4	2800/600	5	0.046	0.055	91	52
5	2800/600	5	0.047	0.059	42	14

Table 5.10. The results of neuro-wavelet models at different decomposition levels, based on the Symlet wavelet

Decomposition Level	Training/validation cases	Number of Hidden Neurons	Train Error %	Test Error %	Time (s)	Epochs
1	2800/600	5	0.019	0.022	48	71
2	2800/600	5	0.026	0.031	60	97
3	2800/600	5	0.025	0.029	38	29
4	2800/600	5	0.037	0.046	94	64
5	2800/600	5	0.042	0.055	15	15

The results suggest (see Fig. 5.27) that models created by both Haar and Coiflet wavelets follow the same pattern as the error steadily decreases by changing the decomposition level from 5 to 1. However, the error percentage has been much less for the models associated with the Haar wavelet in all the cases. On the other hand, models constructed by the Symlet wavelet indicate a different trend. Although, the model created by 1 level of decomposition had the lowest error, similar to other two wavelets, the models did not show any improvements by decreasing the number of decomposition levels from 3 to 2. This is in contrary to the expectation we built up according to the observation of models associated with other two wavelets. Additionally, it seems that the choice of the number of decomposition levels, when it is less than 4, did not affect the performance as much as the other two wavelets.

On the whole, the choice of wavelet is dictated by the signal characteristics and the nature of the application [64]; therefore, there is no straightforward way of determining which wavelet to choose other than trial and error.

However, it should be noted that picking a wavelet has the same effect as deciding on a time-frequency resolution that will be the same for all time and scaling parameter values [30]. This limits the choice of wavelet in cases where specific time-frequency resolution is needed. For example, if for any reason high-frequency resolution is needed, wavelet bases with higher orders that are smoother should be chosen.

The choice of wavelet basis and its order, which signifies its smoothness, should be in accordance with the key features in the signal. As a rule of thumb, it should match the signal as closely as possible [50].

In this study, the Haar wavelet proved to be a good choice. This can be justified intuitively by the nature of the Haar wavelet which is asymmetric and hence, similar to our sinusoid like signals. Additionally, as the Haar wavelet is very compact in time (has a very compact support), it is capable of yielding a high-resolution analysis in time. As a result, it is very well responsive to the locality of the signals in the time domain (see Fig. 4.4 and Fig. 4.5). On the contrary, the Symlets and Coiflets are nearly symmetrical and do not show any resemblance to our signal in that sense. However, they are more similar to our signals in terms of smoothness, in comparison to the Haar wavelet. Hence, these functions are probably better candidates for analysis of the smooth components in the signals.

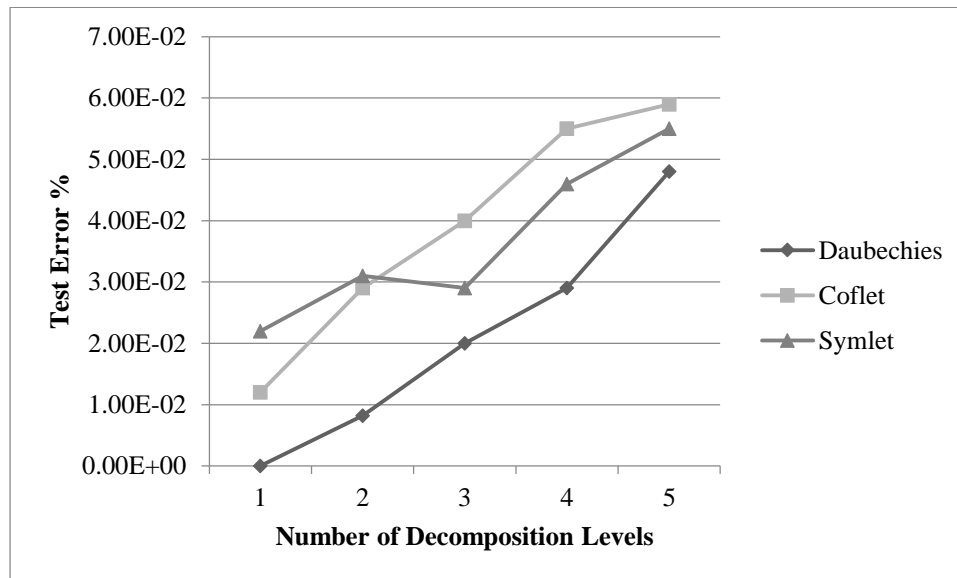


Fig. 5.27. Comparison of performance of the models created by different wavelets

These models can also be compared in terms of the size of their networks. Fig. 5.28 shows the network size in terms of number of inputs for the three types of wavelets at different decomposition levels. It was observed that the Haar wavelet resulted in models with smaller input size in comparison to the Coiflet and Symlet wavelets. Furthermore, according to Fig. 5.27, the error made by the Haar wavelet is also less than that of the Coiflet and Symlet at all the levels. Hence, The Haar wavelet appears to be the best choice among the three. It was also observed that the models created by the Symlet wavelet were more accurate than those made by the Coiflet at 3, 4 and five levels of decomposition. However, this added accuracy was achieved at the cost of many more inputs in the neural networks and hence, loss of simplicity in the models. Therefore, concluding that either wavelet basis function is more successful in this modeling task is rather unfair and can be misleading.

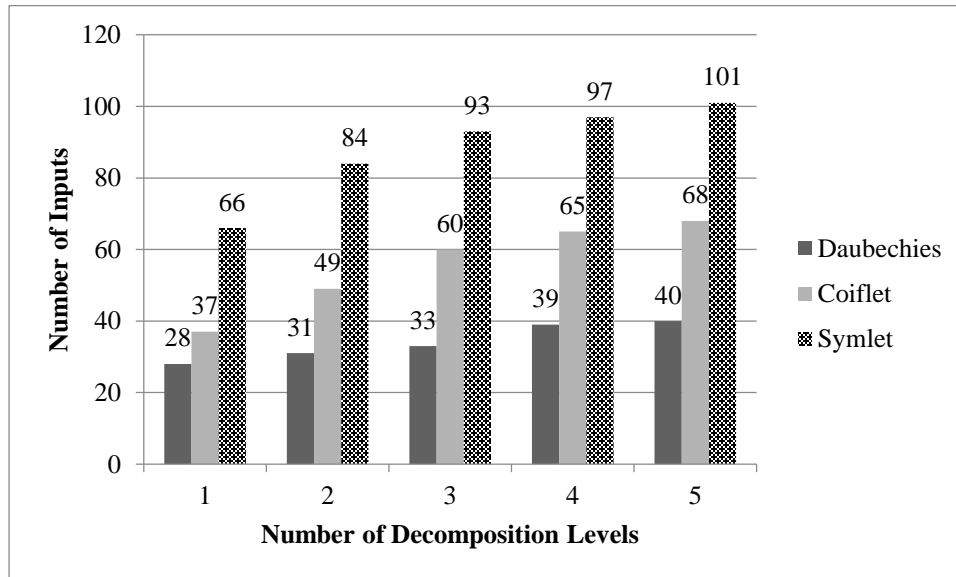


Fig. 5.28. Comparison of size of the models created by different wavelets

5.3 Comparisons to previous work

While the research study in [4] created EAF models based on radial basis function networks, the ones in [15] and [14] took another approach based on fuzzy logic systems. The results of these studies can be compared with those achieved in this work, since all of them used the same data and evaluated their results in terms of the same error index.

Based on the results obtained in our experiments, it can be concluded that the models in this study performed much better than those based on RBF in terms of accuracy. This is confirmed by the fact that even our least accurate wavelet neural network model performed better than the best RBF model (compare 0.073% to 3.25%). From another viewpoint, while the wavelet neural network models led to error percentages as small as $7.6787\text{e-}8\%$, the RBF based models delivered an error of 3.25% at their best. What is more, the neuro-wavelet models were much smaller in size (10 times on average), in terms of number of neurons in the hidden layer. This is indeed, an added advantage for any neural network-based model.

On the contrary, the results of fuzzy logic based models, namely Adaptive Fuzzy Logic Systems (AFLS) were generally very good and closely comparable to ours. To be more precise, the AFLS based models achieved an error percentage of approximately 0.013% to 0.016%, while the error associated to the wavelet neural network based models was broader, ranging from 0.077% to $7.6787\text{E-}8\%$. In other words, with proper parameter selection, our neuro-wavelet method produced much better results and yielded errors of practically zero.

As a final point, it is worth mentioning that the sole criteria used to evaluate the models goodness of fit was the non-dimensional error index. However, a comprehensive assessment of any mathematical model is only possible through a combination of several statistical analyses [65]. In addition to this, our comparisons were made only in terms of the accuracy of models and other

factors such as time and space complexity, cost of development and the level of dependency on the expert knowledge, which may be equally as important for some applications, were not taken into consideration.

6 Conclusions and Contributions

6.1 Summary

In this thesis, an accurate electrical arc furnace modeling technique was developed by incorporating a wavelet transformation into the design of artificial neural networks. By employing wavelet transform, a time-scale multi-resolution representation of the voltage and current signals was produced. These representations were then used to discern the most viable relationship for mapping the arc current to the arc voltage by means of a multi-layered perceptron.

6.2 Conclusions

The wavelet transform was demonstrated to be particularly helpful in non-stationary signal analysis and modeling, and its application in electrical arc furnace modeling was shown to be successful. It has been shown that the time-scale representation of the voltage and current signals were capable of extracting more informative features in both time and frequency domains. Consequently, the learning procedure in the neural network responsible for mapping the current signal to the voltage signal became more effective and resulted in fast convergence and negligible errors.

However, it was emphasized that the type of wavelet transform should be selected with great caution and in compliance with the desired application and the data being analyzed. Considering this, the MODWT, which is a non-decimated discrete wavelet transform and has the property of shift invariance, was used. The specific requirements for the corresponding wavelet and scaling function filters were then clarified. Furthermore, it was stated that using the models for prediction purpose imposed an extra requirement of causality on wavelet and scaling function filters.

Unlike other studies, the features were extracted from the wavelet transformed coefficients of the voltage and current signals at different resolution levels. The number of these features, residing at specific points in time, was dictated by the Bayesian information criterion.

The experimental results supported our hypothesis that neural networks could be successful in modeling the complex and nonlinear relationships involved in electric arc furnace ($v-i$) characteristics using the new, unconventional feature set. It was specifically shown that a multi-layered perceptron, which is fed with such time-scale multi-resolution features and tuned with proper design parameters could accurately model the underlying relationship among the arc furnace current and voltage data after completing the learning process.

The experimental results showed that the number of levels each of the current and voltage signals is decomposed into, particularly the latter, played an important role in the quality of the resulting models. Similarly, the choice of wavelet basis demonstrated to have a significant effect on the performance.

By comparison with similar works, it was shown that applying wavelet transform to the signals in a preprocessing step, could enhance the features extracted from the signals and, therefore, increase the performance of neural network based models fed with such features.

The developed models have the important practical implication of being used in power-quality penetration, mitigation and compensation studies. The purpose of these studies is reducing the adverse effects of EAF operation such as flicker and harmonic distortions on the power system. The proposed modeling technique is envisioned to be useful in modeling any dynamic and nonlinear load, attributable to many electrical devices, and likewise, to maintain the power quality in the network to which the load is connected.

6.3 Contributions

This paper is a modest contribution to the ongoing research on the investigation of purely data-driven models for describing the behaviour of stochastic electrical loads and in particular, electric arc furnaces. One important contribution is the establishment of a framework, in which the successful integration of wavelet theory and artificial neural networks is achieved. This framework can be considered as a “re-contextualization” of existing time series forecasting techniques to the particular case of electric arc furnace modeling, in a unique setting. Another major contribution is the provision of experimental assessments of specific aspects of the developed framework.

6.4 Future Work Directions

Based on the research that was done, the following future work is recommended:

- The experiments may be replicated using Automatic Relevance Determination (ARD) [66] technique instead of the BIC metric to estimate the number of relevant features at each resolution level. In this way, the work is no more based on the assumption that the transformed signals at different resolution levels are necessarily autoregression processes. The ARD technique, which is based on Bayesian framework, has been used in combination with neural networks in order to estimate the optimum size of the window that is used for creating training samples [58]. This technique makes no assumption about the structure of the signal being studied. Therefore, it may extract fewer and more relevant features and, as a result, develop simpler neural network models.
- A thorough examination can be considered to identify the characteristics in a signal that affect the choice of wavelet basis. These characteristics should be mathematically defined.

It is also required to match these characteristics to that of wavelet bases for automatic basis selection.

- The feasibility of a mathematical explanation for the optimum number of decomposition levels of the signal can be further investigated.
- The effect of the nature of the application on the wavelet choice can be examined.

References

- [1] P. Warrian. The importance of steel manufacturing to canada – A research study. University of Toronto. Toronto. 2010[online]. Available: http://munkschool.utoronto.ca/wp-content/uploads/2013/05/CSPA_report_web.pdf.
- [2] I. Vervenne, K. Van Reuse and R. Belmans, "Electric arc furnace modelling from a “power quality” point of view," in *Electrical Power Quality and Utilisation, 2007. EPQU 2007. 9th International Conference On*, 2007, pp. 1-6.
- [3] C. Sankaran, *Power Quality*. Taylor & Francis, 2001.
- [4] A. Sadeghian and J. D. Lavers, "Application of radial basis function networks to model electric arc furnaces," in *Neural Networks, 1999. IJCNN '99. International Joint Conference On*, 1999, pp. 3996-4001 vol.6.
- [5] J. Arrillaga and N. R. Watson, *Power System Harmonics*. Wiley, 2004.
- [6] G. W. Chang, Min-Fu Shih, Yi-Ying Chen and Yi-Jie Liang, "A Hybrid Wavelet Transform and Neural-Network-Based Approach for Modelling Dynamic Voltage-Current Characteristics of Electric Arc Furnace," *Power Delivery, IEEE Transactions On*, vol. 29, pp. 815-824, 2014.
- [7] A. S. Hauksdottir, T. Soderstrom, Y. P. Thorfinnsson and A. Gestsson, "System identification of a three-phase submerged-arc ferrosilicon furnace," *Control Systems Technology, IEEE Transactions On*, vol. 3, pp. 377-387, 1995.
- [8] R. Collantes-Bellido and T. Gomez, "Identification and modelling of a three phase arc furnace for voltage disturbance simulation," *Power Delivery, IEEE Transactions On*, vol. 12, pp. 1812-1817, 1997.
- [9] E. Acha, A. Semlyen and N. Rajakovic, "A harmonic domain computational package for nonlinear problems and its application to electric arcs," *Power Delivery, IEEE Transactions On*, vol. 5, pp. 1390-1397, 1990.
- [10] R. Balan, V. Maties, O. Hancu, S. Stan and L. Ciprian, "Modeling and control of an electric arc furnace," in *Control & Automation, 2007. MED '07. Mediterranean Conference On*, 2007, pp. 1-6.
- [11] A. Bertola, G. C. Lazaroiu, M. Roscia and D. Zaninelli, "A matlab-simulink flickermeter model for power quality studies," in *Harmonics and Quality of Power, 2004. 11th International Conference On*, 2004, pp. 734-738.
- [12] G. Chang, C. Hatziaioniu, W. Xu, P. Ribeiro, R. Burch, W. M. Grady, M. Halpin, Y. Liu, S. Ranade, D. Ruthman, N. Watson, T. Ortmeyer, J. Wikston, A. Medina, A. Testa, R. Gardinier, V. Dinavahi, F. Acram and P. Lehn, "Modeling devices with nonlinear Voltage-current

Characteristics for harmonic studies," *Power Delivery, IEEE Transactions On*, vol. 19, pp. 1802-1811, 2004.

[13] G. Chang and Cheng-I Chen, "A neural network-based method of modeling electric arc furnace load for power engineering study," in *Power and Energy Society General Meeting, 2010 IEEE*, 2010, pp. 1-1.

[14] A. Sadeghian and J. D. Lavers, "Nonlinear black-box modeling of electric arc furnace: An application of fuzzy logic systems," in *Fuzzy Systems Conference Proceedings, 1999. FUZZ-IEEE '99. 1999 IEEE International*, 1999, pp. 234-239 vol.1.

[15] A. R. Sadeghian and J. D. Lavers, "Application of adaptive fuzzy logic systems to model electric arc furnaces," in *Fuzzy Information Processing Society, 1999. NAFIPS. 18th International Conference of the North American*, 1999, pp. 854-858.

[16] A. Sadeghian and J. D. Lavers, "Dynamic reconstruction of nonlinear characteristic in electric arc furnaces using adaptive neuro-fuzzy rule-based networks," *Applied Soft Computing*, vol. 11, pp. 1448, 2011.

[17] A. Sadeghian and J. D. Lavers, "Application of feedforward neuro-fuzzy networks for current prediction in electric arc furnaces," in *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference On*, 2000, pp. 420-425 vol.4.

[18] A. Sadeghian and J. D. Lavers, "Neuro-fuzzy predictors for the approximate prediction of v-i characteristic of electric arc furnaces," in *Fuzzy Information Processing Society, 2000. NAFIPS. 19th International Conference of the North American*, 2000, pp. 183-187.

[19] A. Sadeghian and J. D. Lavers, "Recurrent neuro-fuzzy predictors for multi-step prediction of v-i characteristics of electric arc furnaces," in *Fuzzy Systems, 2000. FUZZ IEEE 2000. the Ninth IEEE International Conference On*, 2000, pp. 110-115 vol.1.

[20] A. Sadeghian, "Nonlinear neuro-fuzzy prediction: Methodology, design and applications," in *Fuzzy Systems, 2001. the 10th IEEE International Conference On*, 2001, pp. 1022-1026 vol.3.

[21] Jin Ma, R. He, Zhao-Yang Dong and D. J. Hill, "Measurement-based load modeling using genetic algorithms," in *Evolutionary Computation, 2007. CEC 2007. IEEE Congress On*, 2007, pp. 2909-2916.

[22] N. Sinha, Loi-Lei Lai, P. K. Ghosh and Yingnan Ma, "Wavelet-GA-ANN based hybrid model for accurate prediction of short-term load forecast," in *Intelligent Systems Applications to Power Systems, 2007. ISAP 2007. International Conference On*, 2007, pp. 1-8.

[23] N. Nalini and G. Raghavendra Rao. Cryptanalysis of simplified data encryption standard via optimization heuristics. *International Journal of Computer Science and Network Security* 6(1B), pp. 240-246. 2005.

- [24] R. Merry and M. Steinbuch, "Wavelet theory and applications," *A Literature Study*, Eindhoven University of Technology, 2005.
- [25] T. Li, Q. Li, S. Zhu and M. Ogihara, "A Survey on Wavelet Applications in Data Mining," *SIGKDD Explor. Newsl.*, vol. 4, pp. 49-68, dec, 2002.
- [26] S. Russell and University of Maryland, Baltimore County. Information Systems, *Resource Availability Awareness and Data Utility: The Foundation for a DSS Framework in a Pervasive Computing Environment*. University of Maryland, Baltimore County, 2008.
- [27] C. S. Burrus, R. A. Gopinath and H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer*. Prentice Hall, 1998.
- [28] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge University Press, 2006.
- [29] O. Renaud, J. Starck and F. Murtagh, "Prediction based on a multiscale decomposition," in *International Journal of Wavelets, Multiresolution and Information Processing*, pp. 217-232.
- [30] C. Gargour, M. Gabrea, V. Ramachandran and J. -. Lina, "A short introduction to wavelets and their applications," *Circuits and Systems Magazine, IEEE*, vol. 9, pp. 57-68, 2009.
- [31] D. Veitch. Wavelet neural networks and their application in the study of dynamical systems. [online]. 2005.
- [32] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions On*, vol. 11, pp. 674-693, 1989.
- [33] J. Di, *Fundamentals of Wavelets*. WIT, 2012.
- [34] H. Wu, X. Li, D. Stade and H. Schau, "Arc fault model for low-voltage AC systems," *Power Delivery, IEEE Transactions On*, vol. 20, pp. 1204-1205, 2005.
- [35] D. Tewari, and G. T. Heydt, "Load Modeling Utilizing Symbolic Dynamnics," *Power Engineering Review, IEEE*, vol. 22, pp. 53-54, 2002.
- [36] C. S. Daw, C. E. A. Finney and E. R. Tracy, "A review of symbolic analysis of experimental data," *Rev. Sci. Instrum.*, vol. 74, pp. 915-930, 2003.
- [37] F. Chen, K. B. Athreya, V. V. Sastry and S. S. Venkata, "Function space valued Markov model for electric arc furnace," *Power Systems, IEEE Transactions On*, vol. 19, pp. 826-833, 2004.
- [38] H. M. Petersen, R. G. Koch, P. H. Swart and R. van Heerden, "Modelling arc furnace flicker and investigating compensation techniques," in *Industry Applications Conference*, 1995.

Thirtieth IAS Annual Meeting, IAS '95., Conference Record of the 1995 IEEE, 1995, pp. 1733-1740 vol.2.

[39] G. W. Chang, Cheng-I Chen and Yu-Jen Liu, "A Neural-Network-Based Method of Modeling Electric Arc Furnace Load for Power Engineering Study," *Power Systems, IEEE Transactions On*, vol. 25, pp. 138-146, 2010.

[40] P. A. Castillo, J. J. Merelo, A. Prieto, V. Rivas and G. Romero, "G-Prop: Global optimization of multilayer perceptrons using \GAs\," *Neurocomputing*, vol. 35, pp. 149, 2000.

[41] Z. Bashir and M. E. El-Hawary, "Short term load forecasting by using wavelet neural networks," in *Electrical and Computer Engineering, 2000 Canadian Conference On*, 2000, pp. 163-166 vol.1.

[42] S. Ying, Z. Niaona, L. Xiuhe, Y. Hongxia and Y. Zhiyan, "Power consumption prediction of submerged arc furnace based on multi-input layer wavelet neural network," in *Mechanic Automation and Control Engineering (MACE), 2010 International Conference on*, 2010, pp. 3586-3589.

[43] N. Masuda and K. Aihara, "Prediction of chaotic time series with wavelet coefficients," *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, vol. 84, pp. 50-59, 2001.

[44] Z. Niaona, W. Zijian and Z. Dejiang, "The multi-scale forecast of submerged arc furnace energy consumption base on support vector machine," in *Computer, Mechatronics, Control and Electronic Engineering (CMCE), 2010 International Conference on*, 2010, pp. 108-111.

[45] S. Soltani, "On the use of the wavelet decomposition for time series prediction," *Neurocomputing*, vol. 48, pp. 267-277, 10, 2002.

[46] S. Soltani, D. Boichu, P. Simard and S. Canu, "The long-term memory prediction by multiscale decomposition," *Signal Process*, vol. 80, pp. 2195, 2000.

[47] L. Zhongliang, C. Zhiming, H. Xiaohong, and L. Fei, "Multiple RBF-NN model for electrical load prediction based on anti-aliasing wavelet transform," *Proceedings of the 32nd Chinese Control Conference*, pp. 3310-3313, July 2013.

[48] Z. A. Bashir and M. E. El-Hawary, "Applying Wavelets to Short-Term Load Forecasting Using PSO-Based Neural Networks," *IEEE Trans. Power Syst.*, vol. 24, pp. 20-27, 2009 FEB., 2009.

[49] G. Nason, "Stationary and non-stationary times series," *Statistics in Volcanology.Special Publications of IAVCEI*, vol. 1, pp. 129-142, 2006.

[50] Anonymous (). *Essentials in wavelet theory*. Available:
<http://www.colorado.edu/engineering/CAS/courses.d/ASEN5519.d/kaist.lecture.11.pdf>.

- [51] S. Ma and C. Ji, "Modeling heterogeneous network traffic in wavelet domain," *Networking, IEEE/ACM Transactions On*, vol. 9, pp. 634-649, 2001.
- [52] M. P. Bhogle and A. Chhangani. Content based copy detection using TIRI-DCT method. *International Journal of Engineering Sciences & Research Technology* 3(7), pp. 449-454. 2014.
- [53] T. Dubey and J. Mishra, *Krishna's Digital Signal Processing: (Principles and Applications)*. Krishna Prakashan Media, 2006.
- [54] S. Rebennack, P. Pardalos, M. V. F. Pereira and N. A. Iliadis, *Handbook of Power Systems II*. Springer, 2010.
- [55] C. Cornish. WMTSA wavelet toolkit for MATLAB. [online]. 0.2.62006. Available: <http://www.atmos.washington.edu/~wmtsa>.
- [56] M. Aminghafari and J. Poggi, "Forecasting Time Series using Wavelets," *Int. J. Wavelets Multiresolut Inf. Process.*, vol. 05, pp. 709-724, 09/01; 2014/06, 2007.
- [57] O. Renaud, J. Starck and F. Murtagh, "Wavelet-based combined signal filtering and prediction," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions On*, vol. 35, pp. 1241-1251, 2005.
- [58] B. Zhang and Z. Dong, "An adaptive neural-wavelet model for short term load forecasting," *Electr. Power Syst. Res.*, vol. 59, pp. 121-129, 9/28, 2001.
- [59] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, 2002.
- [60] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and its Applications: With R Examples*. Springer, 2010.
- [61] J. J. Rattray. Neural network predictive process modeling: Applications to food processing. [online]. 1998. Available: <http://ezproxy.lib.ryerson.ca/login?url=http://search.proquest.com/docview/304447013?accountid=13631>.
- [62] R. Rojas, *Neural Networks: A Systematic Introduction*. Springer Berlin Heidelberg, 1996.
- [63] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bull. Am. Meteorol. Soc.*, vol. 79, pp. 61-78, 1998.
- [64] M. Misiti, Y. Misiti, G. Oppenheim and J. Poggi. Wavelet toolbox user's guide. MathWorks, Inc. Natick, MA. 2014[online]. Available: http://www.mathworks.com/help/pdf_doc/wavelet/wavelet_ug.pdf.

- [65] L. O. Tedeschi, "Assessment of the adequacy of mathematical models," *Agricultural Systems*, vol. 89, pp. 225-247, 2006.
- [66] D. J. MacKay, "Bayesian nonlinear modeling for the prediction competition," *Maximum Entropy and Bayesian Methods*. Springer Netherlands, 1996. 221-234.