Chapter 1

Literature Review

1-1 Introduction

Science has evolved from trying to understand and predict the behavior of the universe and systems within it. Much of this is based on finding suitable models, which agree with observations, and analyzing the results. These models can come in many different forms such as regression models, Artificial Neural Networks (ANN) and Fuzzy systems.

Forecasting is a systematic effort to anticipate future events or conditions. The most well known type of forecast may be that of the meteorologist who prepares daily weather forecasts that help us decide how to dress each day and whether to take an umbrella when we leave for work in the morning. Other common forecasts are those that anticipate future economic conditions, traffic patterns, and even the size and number of classrooms that will be needed in local schools.

In a prediction framework, the results of a statistical analysis based on data about the past are used to speculate about the future and to make decisions. In other way, forecasting and decision-making are very closely related. In a prediction context, researchers use data about the
past with the newest data about actual to speculate about the future and they encourage policy
makers to act on that statistical vision of the future.

1-2 Identification

Forecasting and identification have very close relationship with each other. Hence, better
identification model results the high precision forecasting. Identification is a process through
which one ascertains the identity of another person or entity.

Simulations (and models, too) are abstractions of reality. Often they deliberately
emphasize one part of reality at the expense of other parts. Whereas models are mathematical,
logical, or some other structured representation of reality, simulations are the specific application
of models to arrive at some outcome.

In order to achieve the mission and goals, more industrial specific properties should be
needed to enable the sharing and the reusing of semantics of models among different domains,
territories or countries.

1-3 Soft Computing

Soft computing refers to a collection of computational techniques in computer science,
artificial intelligence, machine learning and some engineering disciplines, which attempt to
study, model, and analyze very complex phenomena: those for which more conventional
methods have not yielded low cost, analytic, and complete solutions. Earlier computational
approaches could model and precisely analyze only relatively simple systems. More complex
systems arising in biology, medicine, the humanities, management sciences, and similar fields
often remained intractable to conventional mathematical and analytical methods. That said, it
should be pointed out that simplicity and complexity of systems are relative, and many conventional mathematical models have been both challenging and very productive.

Unlike hard computing schemes, which strive for exactness and full truth, soft computing techniques exploit the given tolerance of imprecision, partial truth, and uncertainty for a particular problem. Another common contrast comes from the observation that inductive reasoning plays a larger role in soft computing than in hard computing.

In effect, the role model for soft computing is the human mind. The guiding principle of soft computing is: Exploit the tolerance for imprecision, uncertainty, partial truth, and approximation to achieve tractability, robustness and low solution cost. The basic ideas underlying soft computing in its current incarnation have links to many earlier influences, among them Zadeh's 1965 paper on fuzzy sets; the 1973 paper on the analysis of complex systems and decision processes; and the 1979 report (1981 paper) on possibility theory and soft data analysis. The inclusion of neural computing and genetic computing in soft computing came at a later point.

1-3.1 What is absorbed in Soft Computing?

Now, the principal constituents of Soft Computing (SC) are Fuzzy Logic (FL), Neural Computing (NC), Evolutionary Computation (EC) Machine Learning (ML) and Probabilistic Reasoning (PR), with the latter subsuming belief networks, chaos theory and parts of learning theory. What is important to note is that soft computing is not a melange. Rather, it is a partnership in which each of the partners contributes a distinct methodology for addressing problems in its domain. In this perspective, the principal constituent methodologies in SC are complementary rather than competitive. Furthermore, soft computing may be viewed as a foundation component for the emerging field of conceptual intelligence.
1-3.2 Importance of Soft Computing

The complementarities of FL, NC, GC, and PR have an important consequence: in many cases a problem can be solved most effectively by using FL, NC, GC and PR in combination rather than exclusively. A striking example of a particularly effective combination is what has come to be known as "neuro-fuzzy systems". Such systems are becoming increasingly visible as consumer products ranging from air conditioners and washing machines to photocopiers and camcorders. Less visible but perhaps even more important are neuro-fuzzy systems in industrial applications. What is particularly significant is that in both consumer products and industrial systems, the employment of soft computing techniques leads to systems, which have high MIQ (Machine Intelligence Quotient). In large measure, the high MIQ of SC-based systems account for the rapid growth in the number and variety of applications of soft computing.

In many ways, soft computing represents a significant paradigm shift in the aims of computing - a shift which reflects the fact that the human mind, unlike present day computers, possesses a remarkable ability to store and process information which is pervasively imprecise, uncertain and lacking in categoricity.

1-4 Wavelet

Wavelet analysis is a new development in the area of applied mathematics. They were first introduced in seismology to provide a time dimension to seismic analysis that Fourier analysis lacked. Wavelet analysis allows researchers to isolate and manipulate specific types of patterns hidden in masses of data [Soman'05].

Wavelets are mathematical functions that cut up data into different frequency components, and then study each component with a resolution matched to its 'scale'. They have advantages over traditional Fourier methods in analyzing physical situations where the signal
contains discontinuities and sharp spikes. Wavelets were developed independently in the fields of mathematics, quantum physics, electrical engineering, and seismic geology. Historical perspective of wavelets is as follows:

**Historical Perspective:** In the history of mathematics, wavelet analysis shows many different origins [Meyer'93]. Much of the work was performed in the 1930’s, and at that time, the separate efforts did not appear to be parts of a coherent theory.

**Pre-1930:** Before 1930, the main branch of mathematics leading to wavelet began with Joseph Fourier (1807) with his theories of frequency analysis, now often referred to as Fourier synthesis. He asserted that any $2\pi$-periodic function $f(t)$ is the sum of its Fourier series.

$$a_0 + \sum_{k=1}^{\infty} \left( a_k \cos kt + b_k \sin kt \right)$$

The coefficients $a_0$, $a_k$ and $b_k$ are calculated by

$$a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(t) dt, \quad a_k = \frac{1}{\pi} \int_0^{2\pi} f(t) \cos kt dt, \quad b_k = \frac{1}{\pi} \int_0^{2\pi} f(t) \sin kt dt$$

Fourier’s assertion played an essential role in the evolution of the ideas mathematicians had about the functions. He opened up the door to a new functional universe.

After 1807, by exploring the meaning of functions, Fourier series convergence, and orthogonal systems, mathematicians gradually were led from their previous notion of frequency analysis to the notion of scale analysis. That is, analyzing $f(x)$ by creating mathematical structures that vary in scale. How? Construct a function, shift it by some amount, and change its scale. Apply that structure in approximating a signal. Now repeat the procedure. Take that basic structure, shift it, and scale it again. Apply it to the same
signal to get a new approximation and so on. It turns out that this sort of scale analysis is less sensitive to noise because it measures the average fluctuations of the signal at different scales.

**Wavelet multi resolution analysis**

*The 1930s:* In the 1930s, several groups, working independently, researched the representation of the functions using scale-varying basis functions. By using scale varying basis function, called the Haar basis function, Paul Levy a 1930s physicist, investigated Brownian motion, a type of random signal [Meyer'93]. He found that the Haar basis function is superior to the Fourier basis functions for studying small-complicated details in the Brownian motion.

Another 1930s research effort by Littlewood, Paley, and Stein involved computing the energy of the function $f(x)$:

$$\text{energy} = \frac{1}{2} \int_0^{2\pi} |f(t)|^2 \, dt \quad (1.3)$$

The computation produced different results if the energy was concentrated around a few points or distributed over a larger interval. This result disturbed the scientists because it indicated that energy might not be conserved. The researchers discovered a function that can vary in scale and can conserve energy when computing the functional energy. Their work provided Devid Marr with an effective algorithm for numerical image processing using wavelets in the early 1980s.

*1960-1980s:* During these years a lot of work has been done. Some of the pioneering works done by Coifman and Morlet are given below:
- **Guido Weiss and Ronal R. Coifman (1960-1980):** These two mathematicians studied the simplest element of a function space, called atoms, with the goal of finding the atoms for a common function and finding the "assembly rules" that allows the reconstruction of all elements of the function space using these atoms.

- **Grossman and Morlet (1980):** A physicist and an engineer, broadly defined wavelets in contest of quantum physics. These two researchers provided a way of thinking for wavelets based on physical intuition.

- **1980-1990s:** In these years, the pioneering work of the Stephane Mallat (1985) on pyramidal algorithm or multi-resolution theory gave the new apex in wavelet era.

- **Stephane Mallat (1985):** In 1985, Stephane Mallat gave wavelets an additional jump-start through his work in digital signal processing. He discovered some relationship between quadrature mirror filters, pyramidal algorithms, and orthonormal wavelet bases. Y. Meyer constructed the first non-trivial wavelets. Ingrid Daubechies used Mallat's work to construct a set of wavelet orthonormal basis functions that are perhaps the most elegant, and have become the corner stone of wavelet applications today.

- **Post-1990s:** During this decade application of wavelets, develop in many branch of science, same as signal processing, identifications, numerical analysis and networks.

### 1-5 Motivation

During the nineteen century Fourier transform, solved many problems in physics and engineering. This prominence led scientists and engineers to think of them as the preferred way to analyze phenomena of all kinds. This ubiquity forced a close examination of the method. As a result, through the twentieth century, mathematicians, physicists, and engineers came to realize a
drawback of the Fourier transforms: they have trouble reproducing transient signals or signals with abrupt changes, such as the spoken word or the rap of a snare drum [Soman'05].

At the present scenario, wavelet decomposition emerges as a new powerful tool for function approximation due to its multi-resolution property. Recent advances have shown the existence of orthonormal wavelet bases, from which follows the availability of rates of convergence for approximation by wavelet based networks.

Several works has been done and so many works are going on for wavelets. Its application in neural network and neuro-fuzzy model gives tremendous performance for function approximation. However, until this time, selection of parameters and support of wavelet properties are mystery. Due to these discrepancies and multi resolution property of the wavelets, we have motivated to work with wavelet for forecasting and modeling applications of dynamic systems.

1-6 Scope of the Thesis

In recent years, wavelets have become a very active subject in many scientific and engineering research areas. Especially, Wavelet Neural Networks (WNN), inspired by both the feed forward neural networks and wavelet decompositions, have received considerable attention [Q. Zhang'92, Q. Zhang'97, J. Zhang'95] and become a popular tool for function approximation. The main characteristic of WNN is that wavelet functions are used as the nonlinear transformation function in the hidden layer, instead of the usual sigmoid function. Incorporating the time-frequency localization properties of wavelets and the learning ability of the Neural Network (NN), WNN has shown its advantages over the regular methods such as NN for complex nonlinear system modeling.
At present, there are two different kinds of WNN structure. One is with fixed wavelet bases, where the dilation and translation parameters of wavelet basis are fixed, and only the output layer weights are adjustable. Another type is the variable wavelet bases, where the dilation parameters, translation parameters and the output layer weights are adjustable [Billings'05]. For the WNN with fixed wavelets, the main problem is the selection of wavelet bases/frames. The wavelet bases have to be selected appropriately since the choice of the wavelet basis can be critical to approximation performance. Obviously, to improve the approximation accuracy, a large number of wavelet neurons are required for WNN with fixed wavelet bases. This will result in a large complex network structure and cause over-fitting problem.

Since the dilation parameter has explicit physical concept, i.e., resolution, it plays a significant role in wavelet analysis and approximation of a given function. In this thesis, for selection of the wavelet bases/frames, we have used variable wavelet bases for the better accuracy of function approximation though its complexity is increased. In addition, we have presented a comparative study for different types of the wavelet functions. To used approximation of inputs by Sigmoid Activation Function (SAF) and Wavelet Activation Functions (WAF), separately, we have proposed two neuron models to combine them.

In dealing with the modeling of dynamic systems recurrent network have better performance as compared to static behavior of feed-forward network based on proposed sigmoid-wavelet neuron models different types of recurrent neuron models are introduced. These recurrent neurons give us opportunity of comparative study of recurrent neuron models consist of SAF and WAF in feed-forward neural network architecture.

In many complex and ill-defined systems especially with the uncertainty of the systems, the fuzzy models have shown high performance. Motivated by both the theory of multi-
resolution analysis of WNN and the traditional Neuro-fuzzy model, Wavelet Neuro- Fuzzy (WNF) model is introduced. The goal of introducing the WNN in the fuzzy model is improving function approximation accuracy in terms of the dilation and translation parameters of wavelets, meanwhile not increasing the number of wavelet bases. In general, the Takagi-Sugeno-Kang (TSK) fuzzy models consist of a set of rules, and the consequent of each rule acts like a “local model” by using fuzzy set to partition the input space into local fuzzy regions. The consequents of these rules are represented by either a constant or a linear equation. In this work, the consequent part of each fuzzy rule corresponds to sub-WNNs at different resolution levels and used to capture different behaviors (global or local) of the approximated function. Here, the role of the fuzzy set is to determine the region for the contribution of the sub-WNNs to the output of the WNF. As a result, wavelets with different dilation values under these fuzzy rules are fully utilized to capture various essential components of the system.

In addition, in this work, the series-parallel and parallel configurations, which are used in parameter identification of networks\models, are exploited simultaneously for learning the parameters of the premise and the consequent part of the neuro-fuzzy model.

1-7 Organization of Thesis

The organization of the thesis is as follows:

Chapter 2: Wavelet Networks

A comparative study of two existing wavelet networks namely Wavelet Synapses Neural Network (WSNN) and Wavelet Activation Neural Network (WANN), based on three different
wavelet function is presented in this chapter. Feed-forward neural networks show the ability to deal with complex problems and especially in input-output data systems. In addition, wavelet transformation has the ability of representing a function and revealing the properties of the function in the localized regions of the joint time frequency space. The chapter covers some basic concept of wavelet same as wavelet transform, Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT). Three types of non-orthogonal wavelet are introduced in this section. These wavelets when used in feed-forward network give wavelet network.

**Chapter 3: Generalized Wavelet Networks**

The main objective of this thesis is to improve existing one layer feed-forward network with SAF and WAF. Feed-forward neural networks show the ability to deal with complex problems and especially in input-output data systems. In addition, wavelet transformation has the ability of representing a function and revealing the properties of the function in the localized regions of the joint time frequency space. Due to above ability, in this chapter, combination of sigmoid and wavelet activation function is proposed. It has shown that a smart combination of these not only decreases the size of the network, it also increases the accuracy of the network. Two proposed wavelet neural network namely Summation Sigmoid-Wavelet (SS-W) and Multiplication Sigmoid-Wavelet (MS-W) neuron model are discussed in details. One method for structure identification of the model is introduced. General approximation capability of the network has also been presented in this chapter with different theorems.
Chapter 4: Neuro-Fuzzy Model

This chapter serves as an introduction into the basic concept of parameter identification for neuro-fuzzy models. Two parameter identification schemes, namely Parallel (P) and the Series-Parallel (S-P) configurations, are described in this chapter. A combination of these two configurations is proposed for neuro-fuzzy models. Modified mountain clustering is applied to neuro-fuzzy models for structure determination and initialization of the neuro-fuzzy models. An algorithm with adaptive learning rate is used to learn learning parameters of the model. Convergence of the learning procedure is guaranteed by Lyapunov stability theorem.

Chapter 5: Wavelet Neuro-Fuzzy Model

This chapter discusses about the wavelet neuro-fuzzy model. The proposed network in chapter 3 with better performance is used in the consequent part of each fuzzy rule in TSK neuro-fuzzy model that results WNF model. A hybrid of Genetic Algorithm and Gradient Descent has been employed to learn the model parameters.

Chapter 6: Recurrent Wavelet Networks

In this chapter, recurrent neuron models are introduced. Due to the dynamic behavior of recurrent networks, they are suitable in dealing with the modeling of dynamic systems as compared to static behavior of feed-forward network. The quantitative behavior of the sigmoid and wavelet activation functions for dealing with and saving the dynamic of systems are considered. The general approximation properties of the recurrent neuron models are also evaluated. Since the convergence analysis plays an important role in the recurrent networks, the Lyapunov stability approach is employed to guarantee the convergence of network.
Chapter 7: Case study, Indian Monsoon Rain-Fall

The agricultural economy of India is closely linked to the performance of summer monsoon rainfall all over India. The ability to understand and predict circulation and rainfall during the Asian summer monsoon on various time-scales is of prime importance to the economy of several nations of this region because of its affect on agriculture, drinking water, transportation, health, power, and the very livelihood of billions people living in the monsoon region. Due to these reason, in this chapter, all the proposed networks are tested on rainfall data.

Chapter 8: Conclusion

Finally, conclusions of the thesis and suggestions for the future work have been covered in chapter 8.

1-8 Description of Some Dynamic Systems

Six different classes of dynamic systems are described in the following examples for validation of the proposed work. Among them the selected four dynamic examples are different nonlinear differential equations with different order [Narendra'90]. Example five is a general benchmark problem of gas furnace data [Box'70], whereas, example six is an action performed by the operator at chemical plant [Sugeno'93].

Example 1: Linear regression with nonlinear input

The system is a non-linear second order dynamical model [Narendra'90]. The function $f$ is a polynomial of current input $u(k)$ of degree three whereas the input $u(k)$ is a sum of two sinusoids given in (1.6).
\[ y(k+1) = 0.3y(k) + 0.6y(k-1) + f[u(k)] \]  

(1.4)

Where

\[ f[u(k)] = [u(k)]^3 + 0.3[u(k)]^2 - 0.4u(k) \]  

(1.5)

\[ u(k) = \sin(2\pi k/250) + \sin(2\pi k/25) \]  

(1.6)

In this example, 500 input-output data are generated. First three hundred data are used for learning procedure and remaining 200 data are for prediction.

**Example 2: Non-linear regression with random input**

This system expressed as second order nonlinear function that is presented by (1.7). The input \( u(k) \) is a random variable uniformly distributed in the interval \([-1, 1]\). Five hundred input-output data are generated by the second order difference equation [Narendra'90]: three hundred data are used to train the model and remaining two hundred data are used for validation of the model.

\[ y(k+1) = f[y(k), y(k-1)] + u(k) \]  

(1.7)

where function \( f \) is:

\[ f[y(k), y(k-1)] = \frac{y(k)y(k-1)[y(k)+2.5]}{1+y^2(k)+y^2(k-1)} \]  

(1.8)

**Example 3: Non-Linear Regression with Non-Linear Input**

A system described by difference equation [Narendra'90] and expressed as (1.9).
This system is having first order nonlinear dynamic. Hundred input-output data are produced by input \( u(k) \) as given in (1.10). Eighty data is used for training and 20 remaining data are used for testing and validation.

**Example 4: Non-linear Regression of Input and output**

In this example, a nonlinear plant with third delay in output and with two delays in inputs has been taken from [Narendra'90, Lee'00] and describes as:

\[
y(k + 1) = f(y(k), y(k-1), y(k-2), u(k), u(k-1))
\]

Where \( f \) is:

\[
f(x_1, x_2, x_3, x_4, x_5) = \frac{x_1 \cdot x_2 \cdot x_3 \cdot x_5 (x_3 - 1) + x_4}{1 + x_3^2 + x_2^2}
\]

The reference [Narendra'90] has used five input to predict next output but [Lee'00] used only \( u(k) \) and \( y(k) \) to predict next output \( y(k+1) \). Here we also used these two variables to predict the output \( y(k+1) \). One thousand input-output data are produced by using the input expressed by (1.12) to identify the models. The input \( u(k) \) is selected same as equation (1.12) for data 1 to 1000.
\[
    u(k) = \begin{cases} 
    \sin \left( \frac{\pi k}{25} \right), & k < 250 \\
    1.0 & 250 \leq k < 500 \\
    -1.0 & 500 \leq k < 750 \\
    0.3 \sin \left( \frac{\pi k}{25} \right) + 0.1 \sin \left( \frac{\pi k}{32} \right) & 750 \leq k < 1000 \\
    + 0.6 \sin \left( \frac{\pi k}{10} \right) & 
    \end{cases}
\]

Example 5: Gas Furnace data

A benchmark problem of system identification is considered [Box'70]. The process in this example is a gas furnace with single input \( u(t) \), i.e., gas flow rate, and single output \( y(t) \), i.e., \( CO_2 \) concentration. Here we supposed there are three inputs: \( y(t-1), u(t-3) \) and \( u(t-4) \) to the model [Sugeno'93]. Total 290 data are utilized which can be found in [Box'70]. First 250 data are used to train the models and remaining 40 data are used for testing and validation of the model.

Example 6: Human Operation at a Chemical Plant

We deal with a model of an operator's control action of a chemical plant [Sugeno'93]. The plant is for producing a polymer by the polymerization of some monomers. Since the start-up of the plant is very complicated, a man has to make the manual operation at the plant. As shown in Fig. 1.1 there are five input candidates \( (u_1, u_2, \cdots, u_5) \) whom a human operator might refer to at the start-up of the chemical plant to take control action \( y \), for production of polymer.
\[ u_1 : \text{monomer concentration.} \]
\[ u_2 : \text{change in monomer concentration.} \]
\[ u_3 : \text{monomer flow rate.} \]
\[ u_4, u_5 : \text{local temperatures inside the plant.} \]
\[ y : \text{set point for monomer flow rate.} \]

Here \( u_1 \) and \( u_3 \) are employed to model the control action [Azeen'03]. Out of 70 data points of the above six variable from the actual plant, first 60 data are used for training the model and remaining 10 data are used for prediction.

Fig. 1.1. Control action of an operator