การทำนายอนุกรมเวลาด้วยการแปลงเวฟเลตและโครงข่ายประสาท

นางสาวปิยะมาศ กานวล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคณนา ภาควิชาคณิตศาสตร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2550

PREDICTION OF TIME SERIES USING WAVELET TRANSFORM AND NEURAL NETWORK

Miss Piyamas Kanuan

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Computational Science

Department of Mathematics

Faculty of Science

Chulalongkorn University

Academic Year 2007

Thesis Title	PREDICTION OF TIME SERIES USING WAVELET TRANSFORM AND NEURAL NETWORK
By	Miss Piyamas Kanuan
Field of Study	Computational Science
Thesis Advisor	Associate Professor Paisan Nakmahachalasint, Ph.D.
Thesis Co-advisor	Chatchai Srinitiworawong, Ph.D.
	Assistant Professor Sirisap Laohakiat

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

..... Dean of the Faculty of Science

(Professor Supot Hannongbua, Ph.D.)

THESIS COMMITTEE

..... Chairman

(Professor Chidchanok Lursinsap, Ph.D.)

..... Thesis Advisor

(Associate Professor Paisan Nakmahachalasint, Ph.D.)

..... Thesis Co-advisor

(Chatchai Srinitiworawong, Ph.D.)

..... Thesis Co-advisor

(Assistant Professor Sirisap Laohakiat)

..... Member

(Associate Professor Pornchai Satravaha, Ph.D.)

ปิยะมาศ กานวล: การทำนายอนุกรมเวลาด้วยการแปลงเวฟเลตและโครงข่ายประสาท. (PREDICTION OF TIME SERIES USING WAVELET TRANSFORM AND NEURAL NETWORK) อ.ที่ปรึกษา: รศ.ดร. ไพศาล นาคมหาชลาสินธุ์, อ.ที่ปรึกษาร่วม: ดร. ฉัตรชัย ศรีนิติวรวงศ์, ผศ. ศิริสรรพ เหล่าหะเกียรติ, 34 หน้า.

วิทยานิพนธ์นี้ นำเสนอการทำนายอนุกรมเวลาด้วยการแปลงเวฟเลตและโครงข่ายประสาท ซึ่งประกอบด้วย 3 ขั้นตอนหลัก คือ การแยกข้อมูลอนุกรมเวลาให้เป็นข้อมูลอนุกรมเวลาย่อยด้วยเวฟเลต จากนั้นทำนายอนุกรมเวลาย่อยแต่ละชุดด้วยโครงข่ายประสาท และการรวมกันของอนุกรมเวลาย่อยที่ทำนายได้ของชุดต้นแบบเป็นขั้นตอนที่สาม ข้อมูลที่ใช้ในงานวิจัยนี้ประกอบไปด้วยข้อมูล 3 ชุด คือ ข้อมูลราคาน้ำมันดิบ, Sunspot series และ Mackey-Glass time series โดยวัดประสิทธิภาพของผลการทดลองโดยค่าคลาดเคลื่อนกำลังสองเฉลี่ยที่ถูกทำให้เป็นบรรทัดฐาน (Normalized Mean Squared Error) โดยให้ผลของการทำนายสำหรับข้อมูลแต่ละชุด คือ 0.08936, 0.10152 และ 0.00045 ตามลำดับ ซึ่งผลการทำนายที่ได้จะนำมาเปรียบเทียบกับผลที่ได้จากการทำนายโดยใช้เพียงโครงข่ายประสาท คือ 0.10192, 0.13059 และ 0.00092 ตามลำดับ นอกจากนี้ เรายังพบด้วยว่าค่าสัมประสิทธิ์สหสัมพันธ์ของวิธีการนี้อยู่ในช่วง 0.90 ถึง 0.99

ภาควิชา คณิตศาสตร์ ลายมือชื่อนิสิต.....
สาขาวิชา วิทยาการคณนา ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา 2550 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

4772380023 : MAJOR COMPUTATIONAL SCIENCE
KEY WORD: PREDICTION / TIME SERIES / WAVELET TRANSFORM / NEURAL NETWORK

PIYAMAS KANUAN: PREDICTION OF TIME SERIES USING WAVELET TRANSFORM AND NEURAL NETWORK. THESIS ADVISOR: ASSOC. PROF. PAISAN NAKMAHACHALASINT, Ph.D., THESIS COADVISOR: CHATCHAI SRINITIWORAWONG, Ph.D., ASST. PROF. SIRISAP LAOHAKIAT, 34 pp.

In this thesis, we propose a combination of the wavelet transform and the neural network for time series prediction. Three main computational steps were applied: the wavelet decomposition, prediction by neural networks, the wavelet reconstruction. The method was applied to three sets of data: the Crude Oil Prices, the Sunspot time series, and the Mackey-Glass time series, and the normalized mean square errors of the prediction are 0.08936, 0.10152, and 0.00045, respectively. For a comparison purpose, we predicted the same data sets using only a neural network, and the normalized mean square error were 0.10192, 0.10828, and 0.00092, respectively. In addition, we found that the correlation coefficients of the scatter plot of the prediction from our method range from 0.90 to 0.99.

Department: Mathematics
Field of Study: Computational Science
Academic Year: 2007

Student's Signature:.....
Advisor's Signature:.....
Co-advisor's Signature:.....
Co-advisor's Signature:.....

ACKNOWLEDGEMENTS

First of all, I am deeply grateful to my advisor, Associate Professor Dr. Paisan Nakmahachalasint, for his generous advice, emotional support, encouragement, and patience on the work of this thesis. I am particularly grateful to my co-advisors: Assistant Professor Sirisap Laohakiat and Dr. Chatchai Srinitiworawong, with his kindness, understanding and valuable advice enable me to make this thesis possible.

I sincerely thank to the chairman and member of the committee of this thesis, Professor Dr. Chidchanok Lursinsup and Associate Professor Dr. Pornchai Satravaha, for their valuable comments and personal efforts in reviewing my work.

I am grateful to the Advanced Virtual and Intelligence Computing Research Center (AVIC) for material support in enabling me to accomplish this research.

Many thanks to Miss Kodchakorn Na Nakornphanom, Mr. Sasikesem Suttumsakul, and Miss Nuttita Pophet for supporting and encouraging me in research work. Also thanks to my friends and everyone who gave such a nice cheer up.

Finally, I would not be able to accomplish this work without the loves and supports from my beloved mother and older sister that becomes an invaluable source of strength along the duration of this study.

CONTENTS

Page

ABSTRACT (THAI)……..iv

ABSTRACT (ENGLISH)……v

ACKNOWLEDGMENTS……vi

CONTENTS……vii

LIST OF TABLES……..viii

LIST OF FIGURES……..ix

CHAPTER

I INTRODUCTION……1

II BACKGROUND KNOWLEDGE……...3

2.1 Wavelet transform……..3

2.2 Neural network……7

III THE PREDICTION……...11

3.1 The Method……11

3.1.1 Wavelet decomposition……..14

3.1.2 Neural network prediction……..17

3.1.3 Wavelet reconstruction……...18

3.1.4 Error measurement……..18

3.2 Time Series Prediction……19

3.2.1 The Crude oil price……..19

3.3.2 The Sunspot time series……23

3.4.3 The Mackey-Glass time series……..27

IV CONCLUSION AND DISCUSSION……32

References……33

VITAE……..34

LIST OF TABLES

Table Page

3.1 Result of Prediction by the NN of Oil Price data……..……20

3.2 Result of Prediction by the NN of Sun Spot data……..……..……24

3.3 Result of Prediction by the NN of Mackey-Glass time series……....28

3.4 Comparison of the correlation coefficient between the NN and the WTNN……....30

LIST OF FIGURES

Figures Page

2.1 Haar scaling function and Haar wavelet function……..…4

2.2 Nested Vector Space Spanned by the scaling Function……...5

2.3 The Recursive Decomposition……....6

2.4 The Recursive Reconstruction……....6

2.5 A biological neuron……...7

2.6 A basic artificial neuron……...8

2.7 Three-layer multilayer perceptrons……9

3.1 Two time step of input and output of neural network……....12

3.2 Diagram of our process……13

3.3 Diagram of decomposition with à trous algorithm……...15

3.4 Level decomposition of 4 with Haar wavelet of the Oil Price data……..……..…16

3.5 Illustration a network of wavelet coefficient c_2 by 2 level of decomposition……17

3.6 Illustration an idea of first three steps……18

3.7 The Crude Oil Price data……..19

3.8 Result of Prediction by the NN of Oil Price data……..…20

3.9 Comparison of NMSE between the NN and the WTNN with the number of neuron is 16……..21

3.10 Comparison between observed value and predicted value of the WTNN…..…22

3.11 The Sunspot data……..23

3.12 Result of Prediction by the NN of Sunspot data…..…24

3.13 Comparison of NMSE between the NN and the WTNN with the number of neuron is 32……25

3.14 Comparison between observed value and predicted value of the WTNN…..…26

3.15 The Mackey-Glass Time series……27

3.16 Comparison of NMSE between the NN and the WTNN with the number of neuron is 32……...28

3.17 Comparison between observed value and predicted value of the WTNN…..…29

3.18 The scatter plot of the Crude Oil Prices data……30

LIST OF FIGURES

Figures Page

3.19 The scatter plot of the Sunspot data……31

3.20 The scatter plot of the Mackey-Glass time series……31

CHAPTER I

INTRODUCTION

Time series prediction is an application of the estimation of the future values according to historical data [1]. Examples of data series include physically observed data series (sunspots, weather, etc.), mathematical data series (integrals of differential equations, Fibonacci sequence, etc.), and financial data series (prices, stock, rates, etc.). Prediction is necessary for preventing undesirable events by forecasting the event, identifying the circumstances preceding the event, and taking corrective action so the event can be avoided. In financial market, it is profitable investment. Financial prediction is a particular application in time series prediction. Since the nature of the financial data is highly non-stationary, financial forecasting is considered as a difficult problem, and is a challenging application in time series prediction.

In the past, various approaches have been developed for time series prediction. Among them, linear regressive method such as autoregressive (AR) and autoregressive moving average (ARMA) models have been the most used method in practice [2]. The linear models are usually inadequate for financial time series. The nonlinear methods such as neural networks have become very popular. Many different types of neural networks such as multilayer perceptrons (MLP) and radial basis function (RBS) have been proven to be universal function approximations, which make neural networks attractive for time series modeling [2]. Artificial neural networks (ANNs) have shown great ability in the modeling and forecasting of nonlinear and non-stationary time series in various fields of study due to their innate nonlinear property and flexibility for modeling [3].

The wavelet transform has been introduced as a powerful tool for the analysis of time series due to its ability to separate a real signal into sub-signals at different levels of resolution, which improve the ability of artificial neural networks to capture valuable information in sub-signal and successfully forecast the original signal [4]. There are several approaches proposed for time series prediction by wavelet transform. They are

based on autoregressive (AR) model [5], threshold autoregressive (TAR) model [4], autocorrelation shell representation [2], multiscale autoregressive (MAR) model [7], or neural network [2], [6], [8]. In Soltani et al. [5] and Zheng et al. [9], the Haar wavelet transform was used. This choice of the Haar transform was motivated by the fact that the wavelet coefficients are calculated only from data previously obtained in time.

A number of techniques have been implemented to perform time series forecasting. In this work, the undecimated wavelet transforms and the neural network will be applied to predict the time series.

The goal of this thesis is to model the historical data in order to accurately forecast future values. In this study, three main sets of data were analyzed: the Crude oil prices, the Sunspot series, and the Mackey-Glass time series. We applied a combination of wavelet transform and neural network to predict one step ahead of the time series. The prediction task is simplified by decomposing data into different scales of resolution and predicting the coefficients of each scale by a multilayer perceptrons neural network. Experimentally, the level of decomposition was determined from 1 to 5 and there were five types of wavelets: Haar, Daubechies4, Daubechies6, Daubechies8, and Daubechies10 considered.

This thesis is organized as follows: Chapter II describes the background knowledge of wavelet transform and neural networks. Implementation of prediction is elucidated in Chapter III. The result of prediction of three data is then summarized in Chapter III. Finally Chapter IV presents the conclusion as well as the discussion for future work.

CHAPTER II

BACKGROUND KNOWLEDGE

This chapter provides a summary of important theoretical backgrounds that are required in this research. It contains two main sections: wavelet transform (WT) and neural network (NN).

2.1 Wavelet Transform

A wavelet transform is the representation of a function by wavelets [10]. Wavelets are a set of basis functions in Hilbert space $L_2(\mathbb{R})$. An important property of the wavelet basis is to provide a multi-scale analysis, involving a hierarchical sequence of nested subspaces V_j of the function space V (i.e. $\cdots \subset V_{j-1} \subset V_j \subset V_{j+1} \subset \cdots$), where the spaces $V_j's$ satisfy a natural scaling condition

$$f(t) \in V_j \Leftrightarrow f(2t) \in V_{j+1} \tag{2.1}$$

In order to use the idea of multi-scaling, it needs two basis functions: the scaling function φ (the father wavelet) and the wavelet function ψ (the mother wavelet). There are many wavelet functions that can be used in wavelet transform such as Haar wavelets, Daubechies wavelets, Morlet wavelets. Haar wavelets may be a good example to help us understand a wavelet transform.

Haar wavelets was introduced in 1910 by Alfred Haar is defined by

$$\psi(t) = \begin{cases} 1, & 0 \le t < \frac{1}{2} \\ -1, & \frac{1}{2} \le t < 1 \\ 0, & otherwise. \end{cases} \tag{2.2}$$

The function ψ defined above can be obtained from the Haar scaling function:

$$\varphi(t) = \begin{cases} 1, & 0 \le t \le 1 \\ 0, & otherwise. \end{cases} \tag{2.3}$$

The natural scaling condition (2.1) tell us that $\varphi(t)$ can be expressed in terms of a weighted sum of shifted $\varphi(2t)$ as

$$\varphi(t) = \sum_{k \in Z} a_k \varphi(2t - k) \tag{2.4}$$

with appropriate coefficients $a_k, k \in \mathbb{Z}$. Then, a family of functions can be generated from the basis functions by scaling and translation by

$$\varphi_{j,k}(t) = 2^{j/2} \varphi(2^j t - k) \tag{2.5}$$

for all $j, k \in \mathbb{Z}$.

The complement of V_j in V_{j+1} is defined as W_j, which means that $V_1 = V_0 \cup W_0$. It consequently implies that $W_0 \subset V_1$. Thus wavelet function $\psi(t)$ can be expressed by a weighted sum of shifted $\varphi(2t)$ as

$$\psi(t) = \sum_{k \in Z} b_k \varphi(2t - k) \tag{2.6}$$

with appropriate coefficients $b_k, k \in \mathbb{Z}$. Similarly, a family of functions can be generated from the basis functions by scaling and translation by

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k) \tag{2.7}$$

for all $j, k \in \mathbb{Z}$

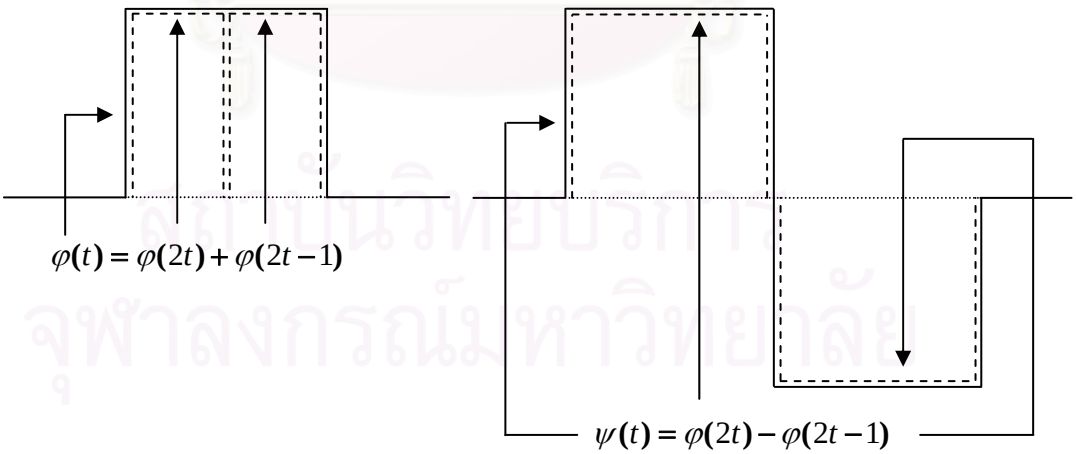

Figure 2.1: The Haar Scaling function and the Haar wavelet function.

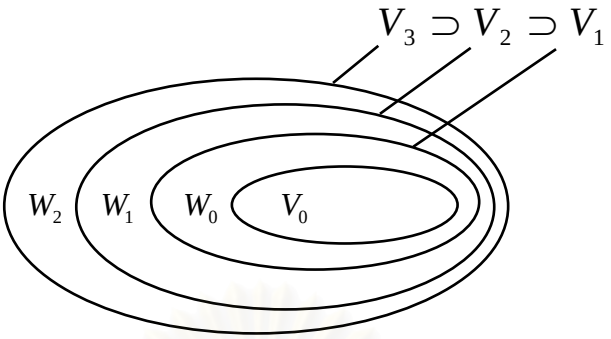

Figure 2.2: Nested Vector Spaces Spanned by the scaling Function.

Let $f(t)$ be the function

$$f(t) = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} c_{j,k}\psi_{j,k}(t) = \sum_{k\in\mathbb{Z}} a_{J,k}\varphi_{J,k}(t) + \sum_{j\le J}\sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(t) \qquad (2.8)$$

where J is a pre-determined level of decomposition. This means that a function f is decomposed into difference scales of resolution. The discrete wavelet transform (DWT) corresponds to represent function or data in terms of coefficients that are associated with difference scale of resolution. These coefficients are determined into two components, the *approximation* of the data is the first component and the second component is *detail*.

A function f can be decomposed into a low frequency part a_1 and a high frequency part d_1. Then, the decomposition of f is represented by $f = a_1 + d_1$. The decomposition is performed on a_1 in order to obtain finer scales. That is $a_1 = a_2 + d_2$. Thus, recursive decomposition for the low frequency parts follows the directions that are illustrated in Fig. 2.3 and 2.4.

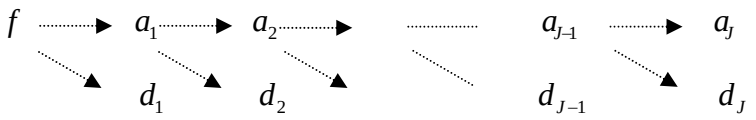

Figure 2.3: The recursive decomposition.

Figure 2.4: The recursive reconstruction.

From the result of recursive decomposition, low frequency parts $a_1,\ldots,a_J$ are the approximation of f, and high frequency parts $d_1,\ldots,d_J$ are the details of f. Fig. 2.3 and 2.4 illustrate a wavelet decomposition and reconstruction of J levels which corresponds to

$$f = a_J + d_J + d_{J-1} + \cdots + d_2 + d_1 \qquad (2.9)$$

2.2 Neural Network

A neural network [11] is a mathematical model that is able to capture and represent complex input/output relationship. The architect of artificial neural network is inspired by the observation of biological nervous system which is a part of a human brain. Our brains contain many neurons. Each neuron has three principal components: a cell body, an axon, and dendrites as shown in Fig. 2.5. The dendrites act as a neuron's input receptors for signal incoming from other neurons. The cell body effectively evaluates these incoming signals and determines the output. The axon is a signal long fiber that carries the signal from the call body out to other neurons as the neuron's output channel. The information is transferred across a synapse by electrochemical voltage.

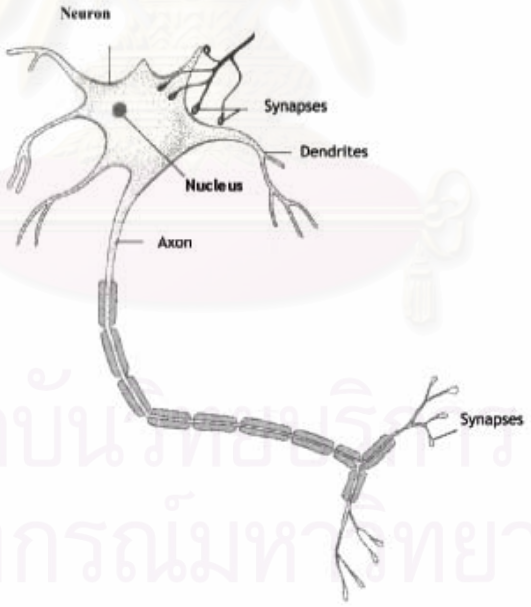

Figure 2.5: A biological neuron.

The basic artificial neuron is shown in Fig. 2.6.

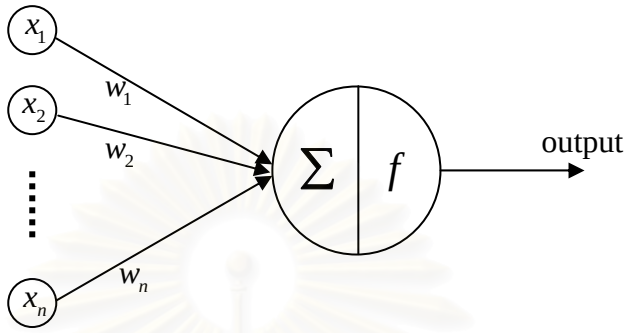

Figure 2.6: A basic artificial neuron.

We see that the connections, w_i, transfer the signals, x_i, into the neuron by computing the weighted sum of input and fed through an activation function to produce an output. Its output is equal to

$$output = f\left(\sum_{i=1}^{n} x_i w_i\right) \tag{2.9}$$

The values w_i can be interpreted as a weight representing the importance of that specific input x_i and f is the activation function, which is sigmoid function, linear function or other function. The neural network model is created by interconnecting many of basic neurons. It captures the information from data by learning and stores the information among its weights.

2.2.1 Multilayer Perceptrons

In this part, we focus on the multilayer perceptrons (MLPs). This model consists of a set of sensory units that constitute the input layer, one or more hidden layer of computational nodes, and one output layer of computational nodes as shown in Fig. 2.7.

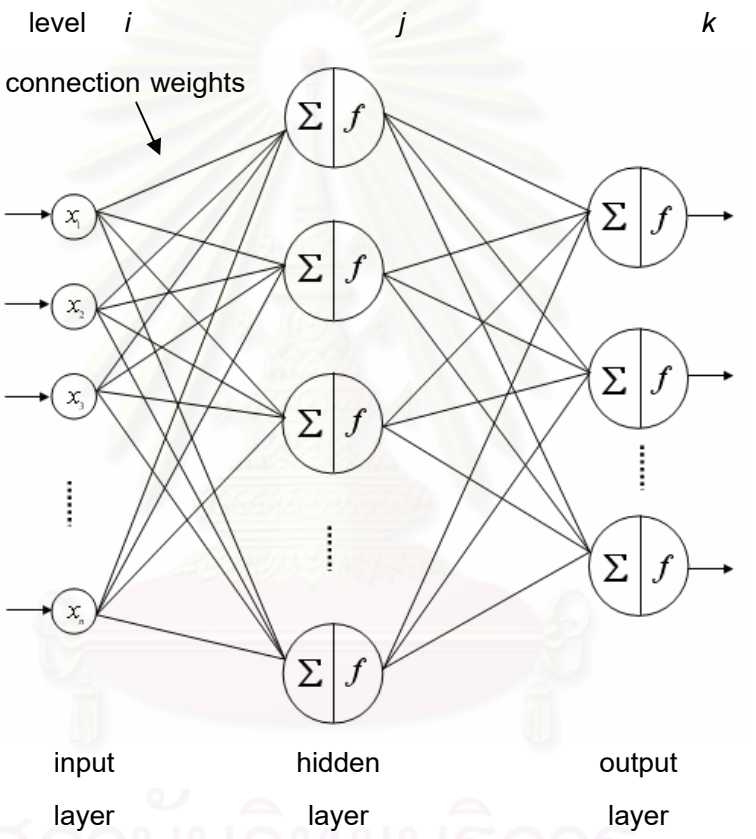

Figure 2.7: Three-layer multilayer perceptrons.

2.2.2 Learning Algorithm

Backpropagation algorithm is one of the most popular used for training neural network [12]. It used a set of pairs of input and output values (called patterns). An input pattern is fed into the network to produce an output, which is then compares with the actual output pattern. After all the pairs of patterns in the training set are treated, the weights are adjusted to minimize the difference between the actual value and the output of the neural network. The gradient descent is the original method to adjust the weights toward convergence. The goal of training is to search an optimal set of connection weights in the manner that the performance function (E) of the network output can be minimized.

$$E = \frac{1}{2}\sum_{k}(t_k - o_k)^2 \tag{2.10}$$

where o_k is the output of k^{th} training pattern and t_k is the target correspond to k^{th} training pattern. Since the training makes use of the actual output, the backpropagation method is referred as a supervised learning.

In this study, we focus on the neural network, which is trained by the backpropagation algorithm using the conjugate gradient method to adjust the weights. The backpropagation algorithm adjusts the weights using gradient descent often converges very slowly. The conjugate gradient method produces generally faster convergence than using gradient descent.

CHAPTER III

THE PREDICTION

3.1 The method

This chapter shows how to predict the time series. In this work, we will compare the result of the prediction by the sole neural network (NN) to that of the combination between the wavelet transform and the neural network (WTNN). The prediction by the NN is as follows. The determination of network structure is a user-defined process, where the number of input nodes and the number of the hidden neurons were pre-determined. Although the MLP is a universal functional approximation, no universal guideline exists in choosing the appropriate network structure for forecasting application. The number of input node corresponds to the number of past data. Since there is no theoretical result suggesting the best number of past data for prediction problem, we will experimentally vary this parameter on each data. However, the number of hidden node is similarly analyzed in the same way.

In this work, the training of the neural network, the feed forward neural network known as the multilayer perceptrons with one hidden layer is simulated to perform one-step ahead prediction. The backpropagation based on the conjugate gradient descent is utilized to perform weight adjustment. The value of initial weights is randomized. In order to remove the effects of initial values, 50 experiments (runs) were performed. We used the window size in the past as input to predict the next value. If the window size is too narrow, main point may fall aside, while if the window size is too wide, useless inputs may act as noise. In order to predict the future value, $x(t+1)$, the past information of n preceding values, $(x(t), x(t-1), ..., x(t-n+1))$, is fed in the input layer. The input and output of neural network is shown in Fig. 3.1.

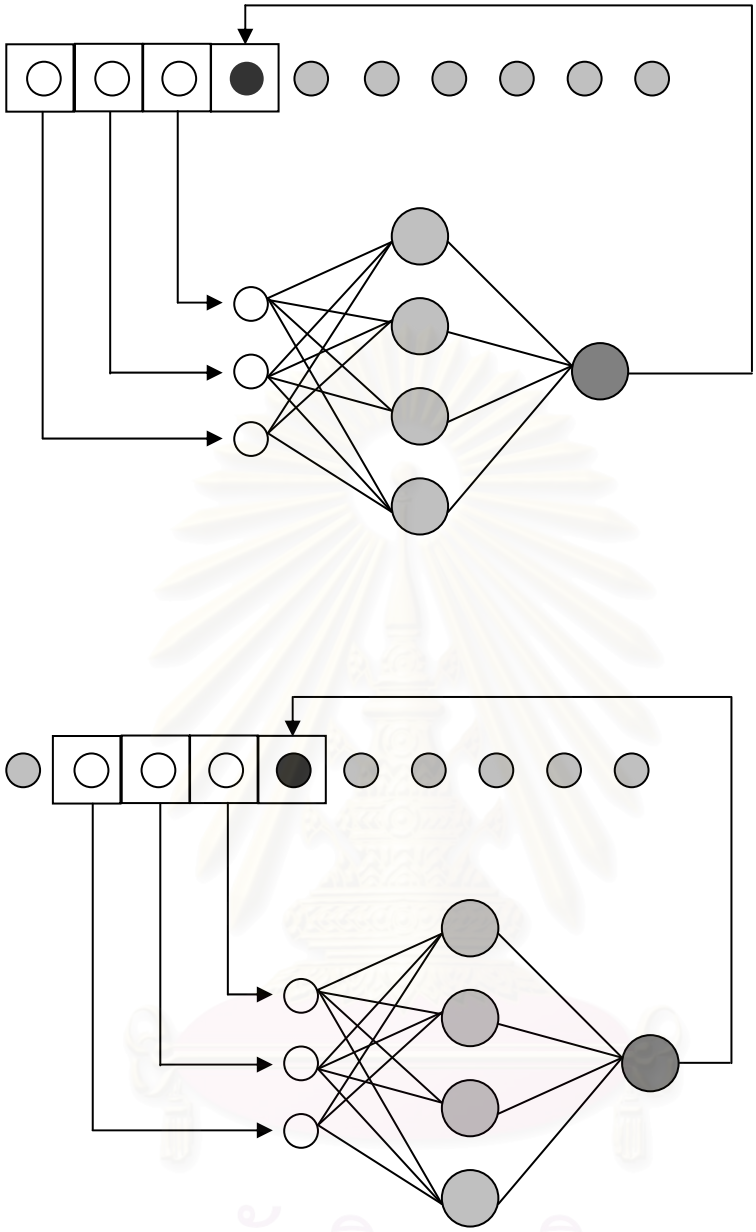

Figure 3.1: Two time steps of input and output of the neural network.

After that, the prediction by the WTNN is determined. There are four main stages of the process. A diagram of our process is shown in Fig. 3.2.

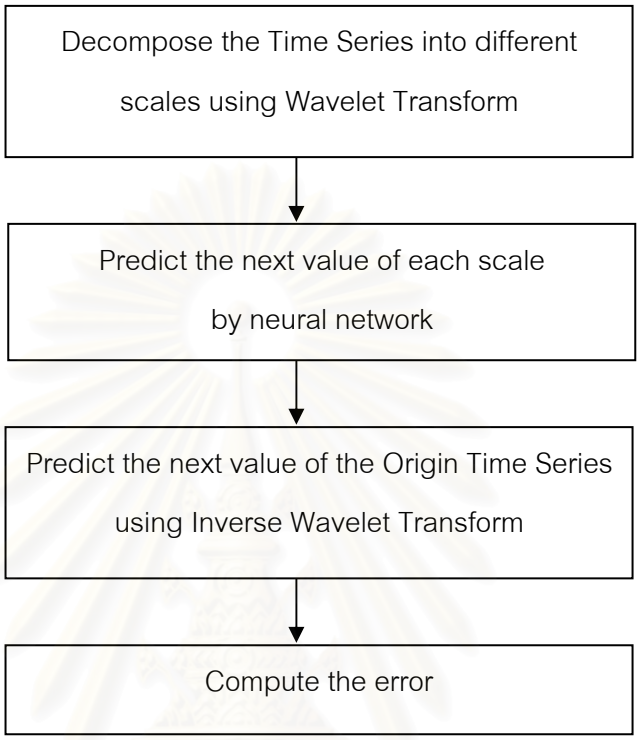

Figure 3.2: Diagram of our process.

3.1.1 Wavelet decomposition

In this work, the time series are discrete. The wavelet transform for discrete data is provided by particular version known as the “à trous” algorithm. The à trous wavelet transform is a redundant transform, i.e. decimation is not carried out. The redundant transform based on an *n*-length input time series has an *n*-length coefficients for each of the resolution levels. Hence, information at each resolution scale is directly related at each time point. It decomposes the original time series into *detail* and *approximation* components. The main motivation of using wavelet decomposition is (give more useful information about the signal than is directly obvious form the signal itself) to decompose signal into various simpler components which are easier to be modeled (predicted). The original time series can be expressed as an additive combination of the wavelet coefficients at the different resolution levels.

Let $c_0(t) = x(t)$ be the original time series. We compute the approximation wavelet coefficients using the following convolution with the low-pass filter *h*:

$$c_{j+1}(k) = \sum_{l=-\infty}^{\infty} h(l) c_j(k + 2^j l) \tag{3.1}$$

After the recursive of the approximation wavelet coefficients, we compute the detail wavelet coefficients by taking the difference between consecutive resolutions:

$$w_{j+1}(t) = c_j(t) - c_{j+1}(t) \tag{3.2}$$

Then, the wavelet expansion of the original time series in term of wavelet coefficients up to the resolution level J is given by

$$x(t) = c_J(t) + \sum_{j=1}^{J} w_j(t) \tag{3.3}$$

This equation is also provided a reconstruction formula for the original time series.

The low-pass filters, h, which we used was a Haar, define as $h = (\frac{1}{2}, \frac{1}{2})$. Then,

$$c_{j+1}(t) = \frac{1}{2} c_j(t - 2^j) + \frac{1}{2} c_j(t) \tag{3.4}$$

$$w_{j+1}(t) = c_j(t) - c_{j+1}(t) \tag{3.5}$$

The choice of the à trous wavelet transform was motivated by two reasons. Firstly, the wavelet coefficients are calculated only form data obtained previously in time. We see that, at any point time, t, we never use information after t in calculating the wavelet coefficients. And the second point, if the data is updated, we do not have to recompute the wavelet of the full signal. Diagram of decomposition with à trous algorithm is shown in Fig. 3.3.

In this study, the time series is decomposed into the *approximation* and *detail* coefficients by decomposition level, l, from 1 up to 5 and five types of wavelets: Haar, Daubechies4, Daubechies6, Daubechies8 and Daubechies10.

The wavelet transform is employed to preprocess the data to be input of the network. Fig. 3.4 is an example of decomposition by Haar of the Oil price data.

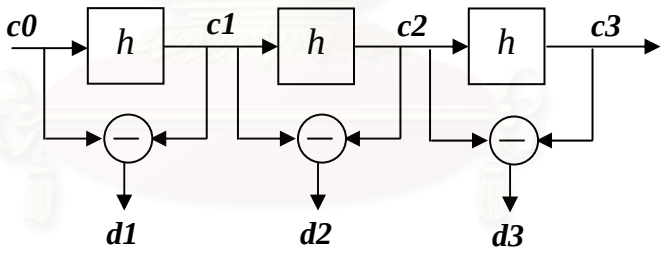

Figure 3.3: Diagram of decomposition with à trous algorithm.

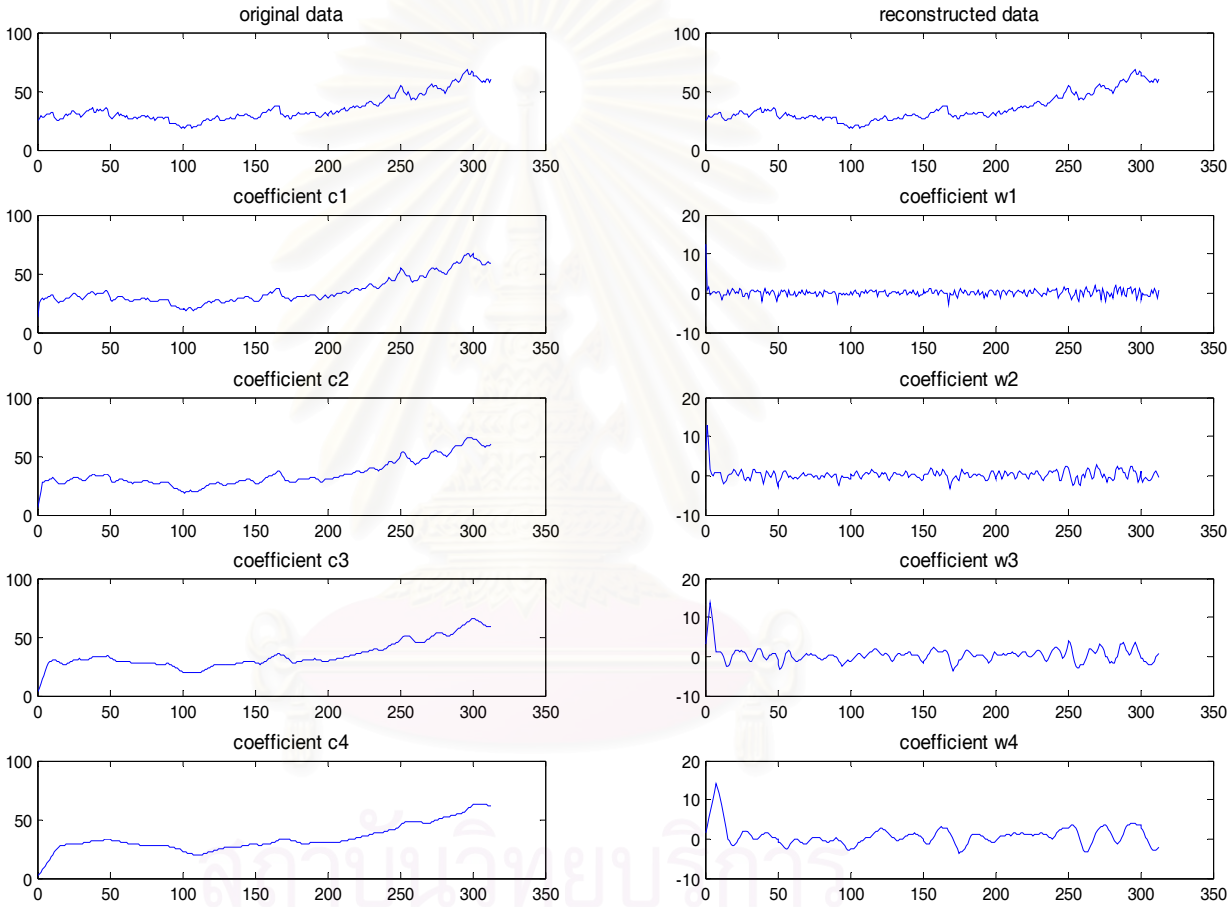

Figure 3.4: Level decomposition of 4 with Haar wavelet of the Oil Price data.

3.1.2 Neural network prediction

After the decomposition of the original time series, we got the set of the coefficients, $\{c_J, d_J, d_{J-1}, \ldots, d_1\}$. Soltani et al. [13] have proposed the prediction of each level of the coefficient by

$$c_J(t+1) = f(c_J(t), \ldots, c_J(t-n+1), d_j(t), \ldots, d_j(t-n+1)). \quad (3.6)$$

$$d_j(t+1) = f(c_J(t), \ldots, c_J(t-n+1), d_j(t), \ldots, d_j(t-n+1)). \quad j = 1, \ldots, J. \quad (3.7)$$

where J, n is the number of decomposition levels and the window size, respectively. The window size is defined according to the structure of the NN. So, there are $J+1$ of neural networks and the number of input nodes of each network is equal to $n \times (J+1)$. Fig. 3.5 illustrates this step.

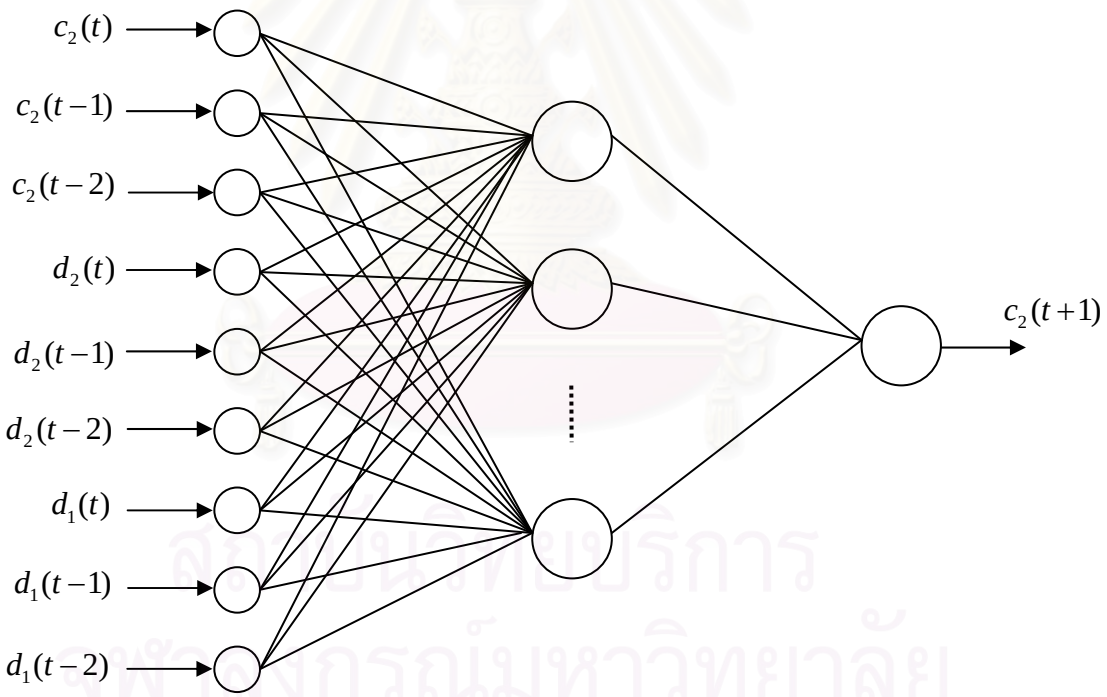

Figure 3.5: Illustration a network of wavelet coefficient c_2 by 2 level of decomposition.

3.1.3 Wavelet reconstruction

In this stage, after we obtained the coefficients of each level. The wavelet reconstruction of the original time series in term of wavelet coefficients is given by

$$x(t) = c_J(t) + \sum_{j=1}^{J} w_j(t), \tag{3.8}$$

which is the à trous reconstructing algorithm. Fig. 3.6 shows idea of first three steps.

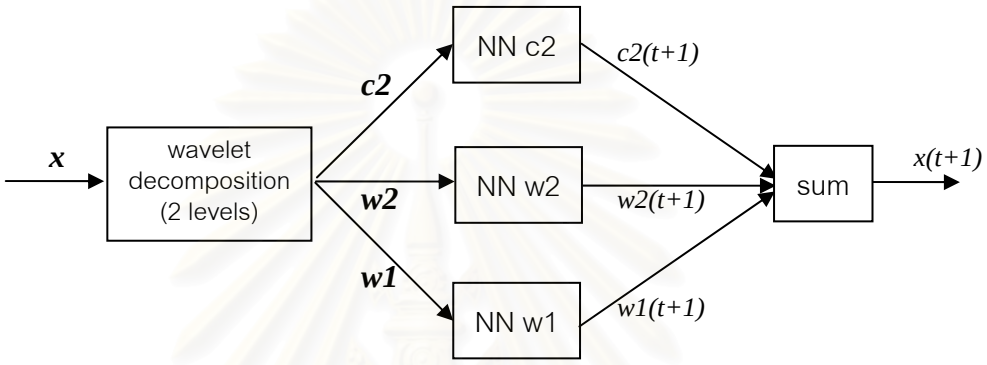

Figure 3.6: Illustration an idea of first three steps.

3.1.4 Error measurement

Prediction quality is measured by means of the average relative variance, which is also known as *normalized mean squared error* (NMSE).

$$NMSE = \frac{1}{\sigma^2}\frac{1}{N}\sum_{k=1}^{N}(x_k - \hat{x}_k)^2 = \frac{\sum_{k=1}^{N}(x_k - \hat{x}_k)^2}{\sum_{k=1}^{N}(x_k - \bar{x})^2} \tag{3.9}$$

where $\bar{x} = \frac{1}{N}\sum_{k=1}^{N} x_k$ is the average and x_k and $\hat{x}_k$ are the exact value and the predicted value, respectively. A value of $NMSE = 0$ indicates prefect prediction while a value of 1 simply corresponds to predicting the average [2].

3.2 Time series Prediction

In this section, we present the results of the experiments. Three benchmarks: the Crude Oil Price, the Sunspot time series, and the Mackey-Glass time series were chosen to validate our proposed method. Each of the data sets was divided into two parts: the training set and the testing set.

3.2.1 The Crude Oil Price

The Crude Oil Price data is available from the website of US Department of Energy, EIA's web site (http://tonto.eia.doe.gov/dnav/pet/pet_pri_spt_s1_w.htm). In the experiments performed in this work, the data consists of averaged weekly WTI spot price (US$/bbl) that cover the period from 7 January 2000 to 30 December. The total number of data points is 313 were use for our implementation, 264 total points were used for the training and the remaining 49 were used for testing the accuracy of the prediction. The period of the Crude Oil Price data is shown in Fig. 3.7.

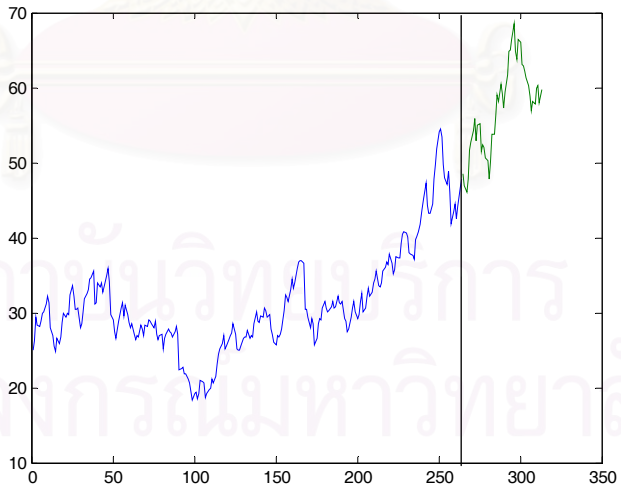

Figure 3.7. The Crude Oil Price data.

The prediction by the NN was carried out in the first step. The appropriate number of input nodes and the number of hidden nodes were determined by mere experiments. For the number of input node, it varies from 2 to 9, which is shown in Fig. 3.8.

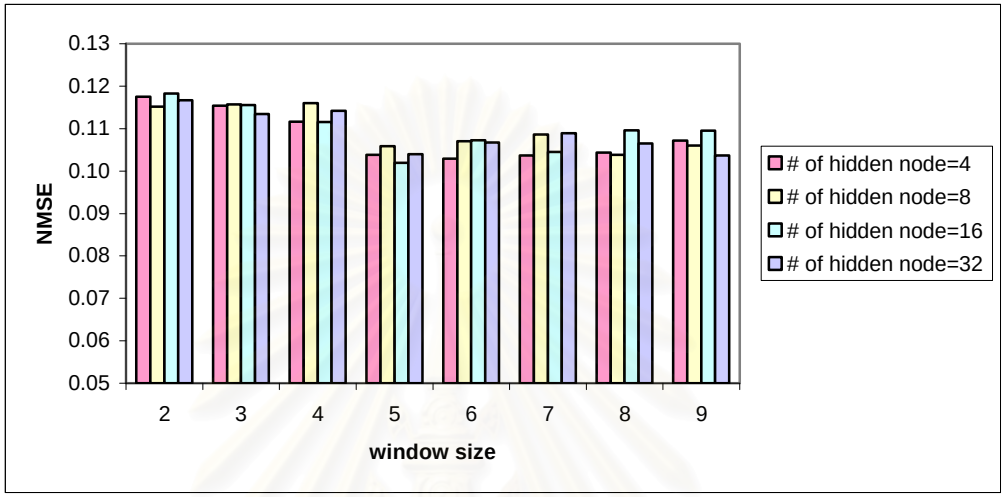

Figure 3.8: Result of Prediction by the NN of Oil Price data.

Table 3.1 shows the NMSE result of the prediction by the NN. It shows that the window sizes of more than 5 do not yield better prediction result. So we choose the window size of 5 as reference for the window size of the WTNN. The optimal NMSE is 0.10192, occurring at 16 nodes of neuron, as shown in table 3.1.

Number of neurons	window size							
	2	3	4	5	6	7	8	9
4	0.11749	0.11539	0.11167	0.10380	0.10292	0.10367	0.10430	0.10717
8	0.11514	0.11573	0.11598	0.10587	0.10705	0.10865	0.10379	0.10599
16	0.11828	0.11552	0.11153	**0.10192**	0.10726	0.10447	0.10961	0.10953
32	0.11665	0.11346	0.11418	0.10398	0.10674	0.10892	0.10655	0.10364

Table 3.1: Result of Prediction by the NN of Oil Price data

After that, the method the WTNN is applied for prediction. Start at wavelet decomposition of data into different scale. Here we considered five types of wavelets with five decomposition levels. Architecture of neural network in this stage, the number of neuron node is same as above, i.e. 16 neurons, and the number of output node is one. In case of the number of input node, it depends on levels of decomposition. For example if number of decomposition level is 3, there are 4 sub-series of c_3, w_3, w_2, w_1. Then the number of input node is 4x5. Fig. 3.9 shows the NMSE of each wavelets and level of decomposition comparing to the NMSE of the NN shown in the solid line.

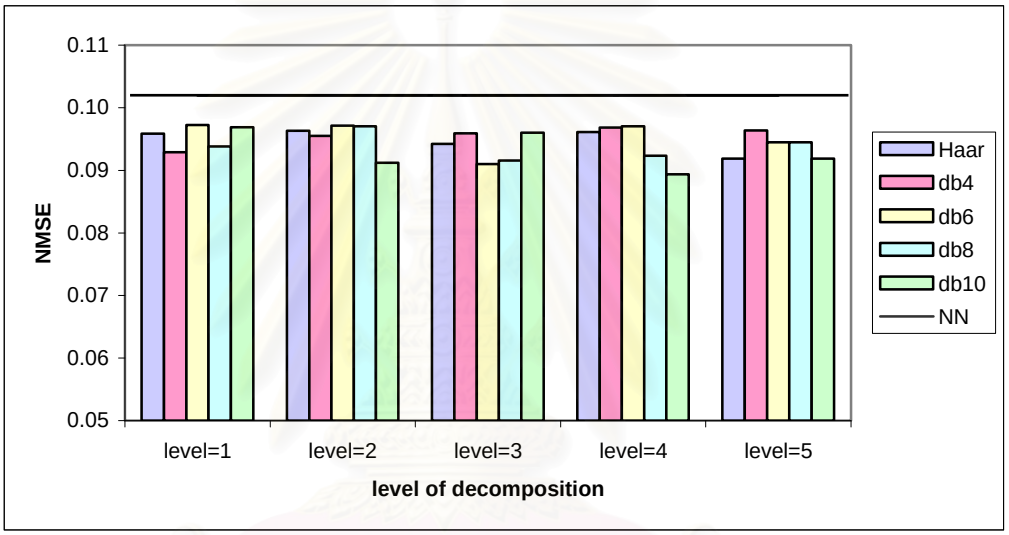

Figure 3.9: Comparison of NMSE between the NN and the WTNN with the number of neuron is 16.

Fig. 3.9 indicates clearly that the prediction by using the WTNN is better than that of the NN. Fig. 3.10 shows the comparison between observed value and predicted value of the WTNN of Daubechies10 with 4 level of decomposition that obtained the optimal result of the NMSE is 0.08936. From this result it is shown that the error of prediction decreased by 12.33%.

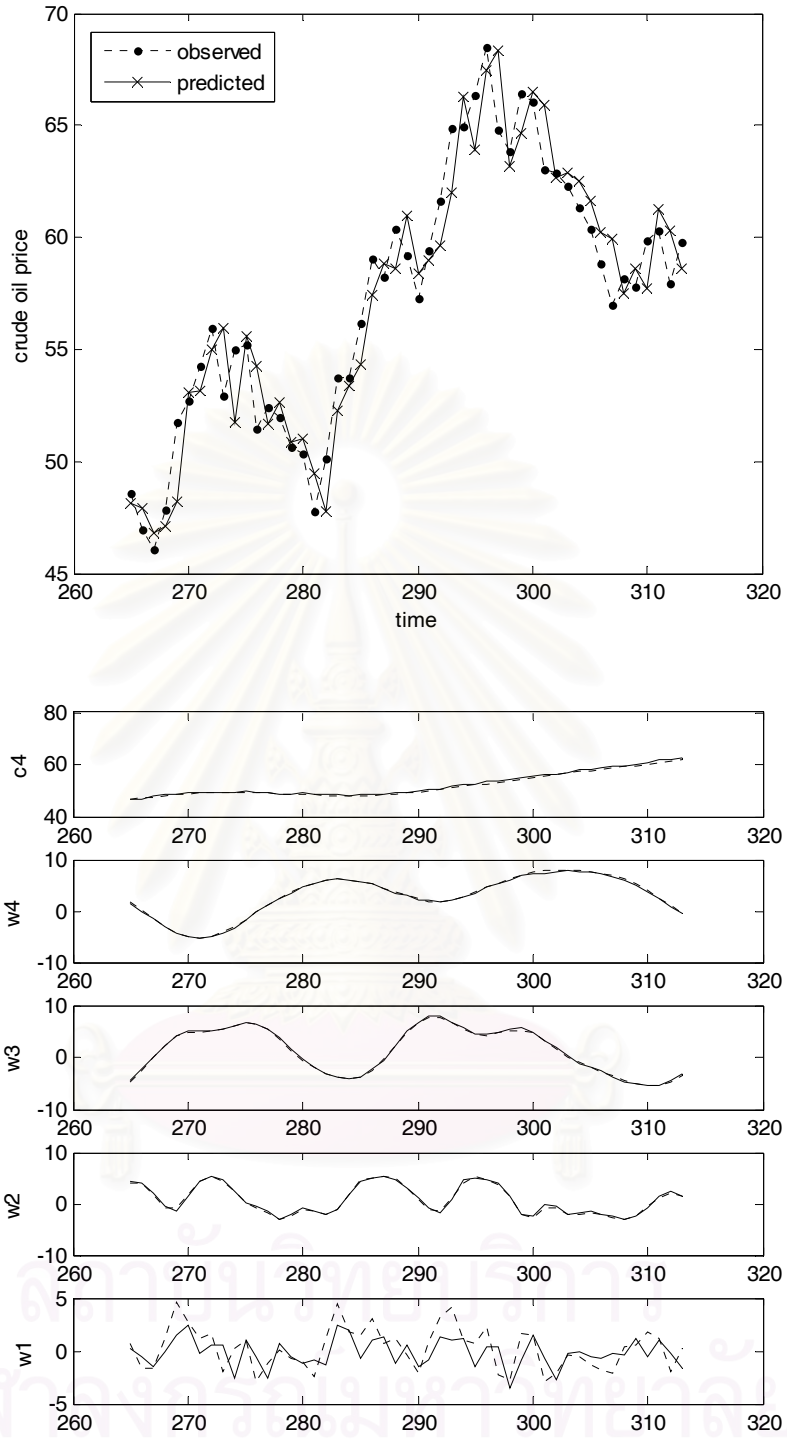

Figure 3.10: comparison between observed value and predicted value of the WTNN (Daubechies10 with 4 level of decomposition).

3.2.2 The Sunspot time series

The annual average number of sunspots during the period from 1700 to 1979 was studied here [1]. Fig. 3.11 shows the sunspot time series from 1700 to 1979. The sunspot data is a classical example of a combination of periodic and chaotic phenomena and activity is cyclical about every 11 years. The sunspots of years 1700 through 1920 were chosen to be the training set and single step prediction of the sunspots of year 1921 through 1979 was predicted.

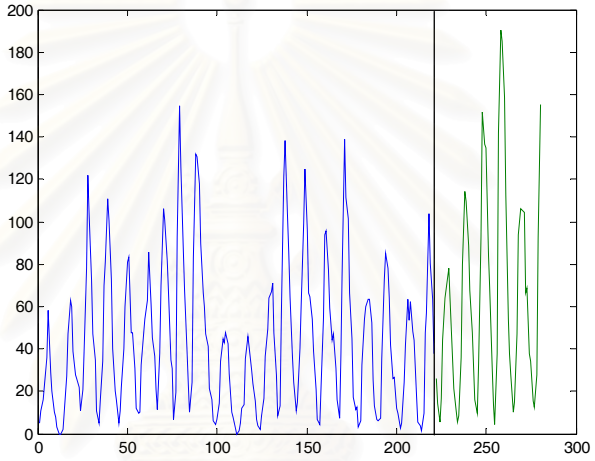

Figure 3.11: The Sunspot data.

Data has periodicity about every 8 to 9 years. The appropriate number of input nodes is determined mainly by trial and error that varied from 6 up to 12. The result of prediction by the NN is shown as Fig. 3.12. It shows that the window size is equal to 9 give the better prediction than the other. The optimal NMSE is 0.13059, occurring at the number of neuron node is 32 as shown in table 3.2. Fig. 3.13 shows the NMSE of each wavelets and level of comparing to the NMSE of the NN shown in the solid line.

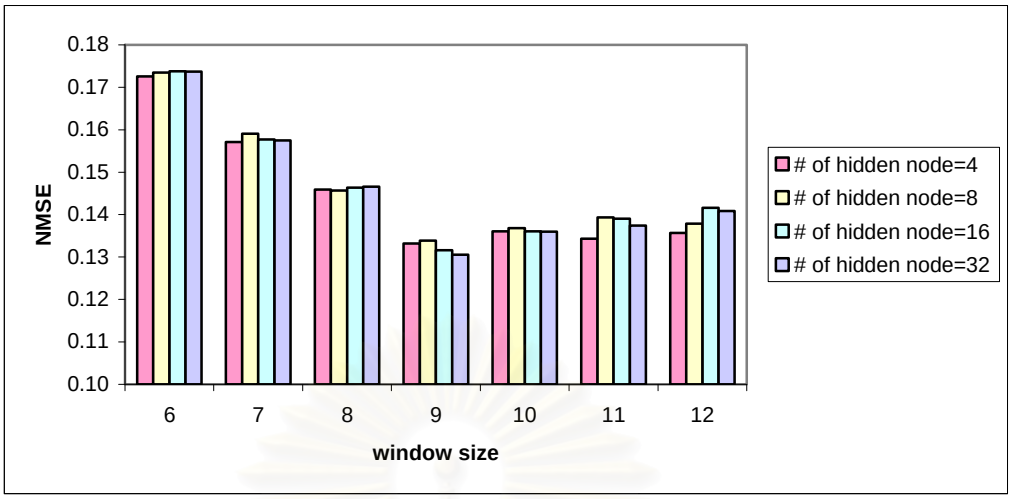

Figure 3.12: Result of Prediction with only the neural network

number of neurons	window size						
	6	7	8	9	10	11	12
4	0.17257	0.15714	0.14594	0.13319	0.13601	0.13434	0.13567
8	0.17342	0.15904	0.14568	0.13387	0.13684	0.13933	0.13784
16	0.17373	0.15775	0.14636	0.13159	0.13602	0.13909	0.14164
32	0.17370	0.15750	0.14660	**0.13059**	0.13600	0.13740	0.14090

Table 3.2: Result of Prediction by the NN of the Sunspot data

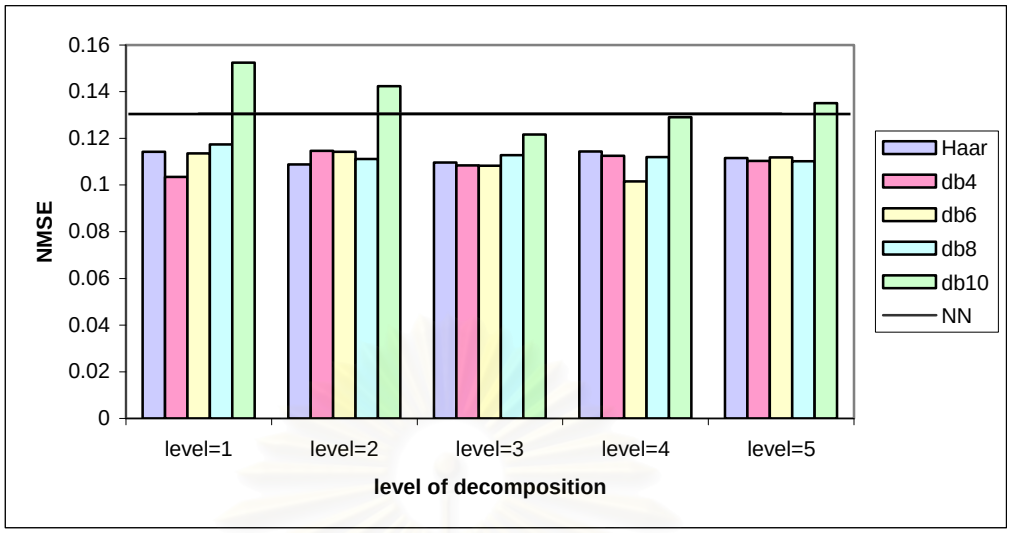

Figure 3.13: Comparison of NMSE between the NN and the WTNN with the number of neuron is 32.

Fig. 3.13 indicates that the prediction result by using the WTNN is better than that of the NN. Fig. 3.14 shows the comparison between observed value and predicted value of the WTNN of Daubechies6 with 4 level of decomposition that obtained the optimal result of the NMSE is 0.101522. From this result, it is shown that the error of prediction decreased by 22.26%.

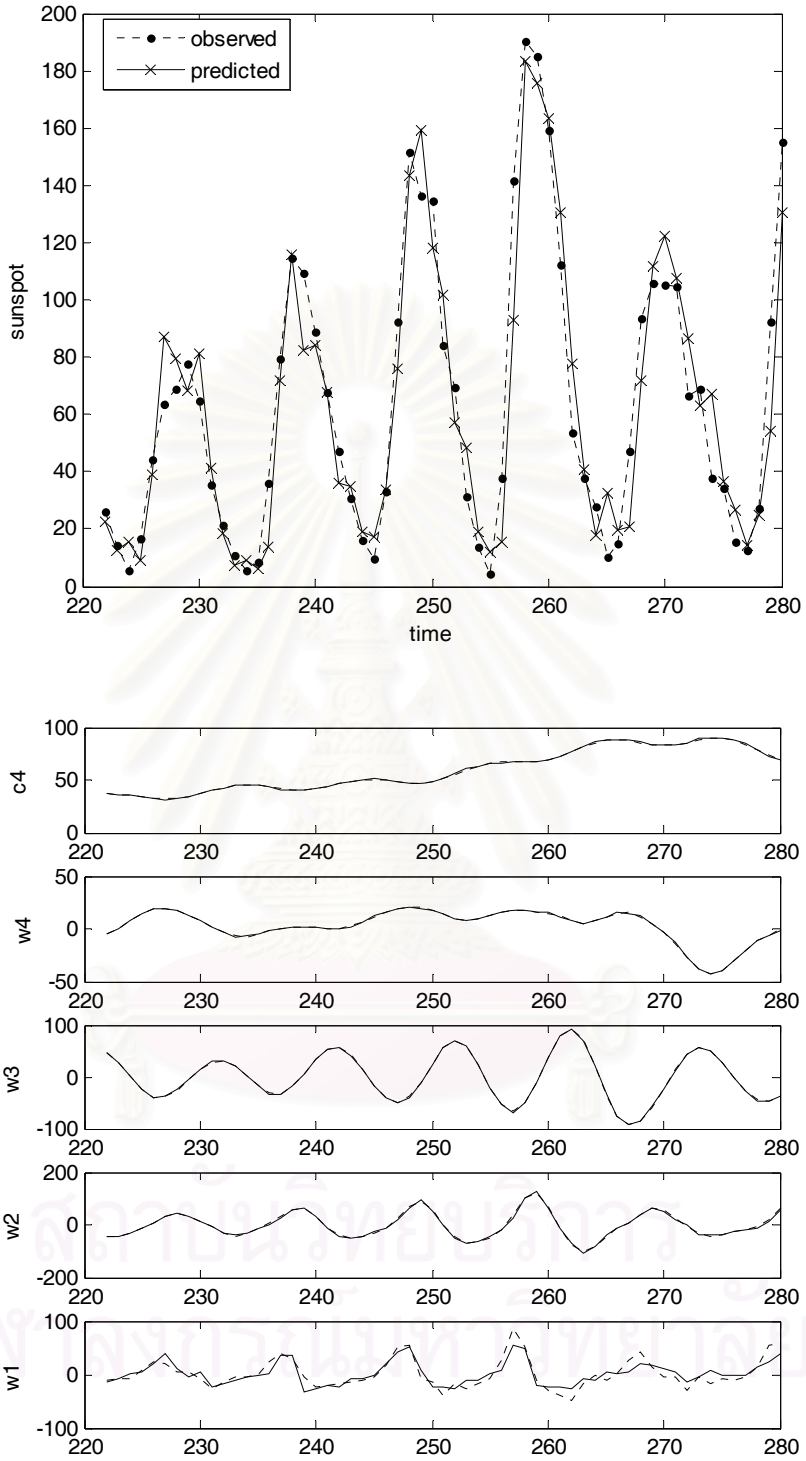

Figure 3.14: comparison between observed value and predicted value of the WTNN (Daubechies6 with 4 level of decomposition).

3.2.3 The Mackey-Glass time series

This data set is a solution of the well-known Mackey-Glass [14] time-delay chaotic differential equation defined as follows:

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t)$$

We consider here τ =17 (MG17). Fig. 3.15 shows the Mackey-Glass time series. The data was sampled every 6 points, as usually recommended for the Mackey-Glass time series. Then, the task of prediction is to predict the value $x(t+6)$ from the values $x(t), x(t-6), x(t-12), x(t-18)$. We generate a 1000 input/output data patterns and the first 500 patterns are used for training set, followed by testing data set of 500 patterns. It is clearly seen that the number of input node is four. So, we only try to find the number of neuron node. Table 3.3 shows the NMSE using different number of neuron node. It shows that the NMSE yields an optimal value of 0.00092 when the number of neuron nodes is 32.

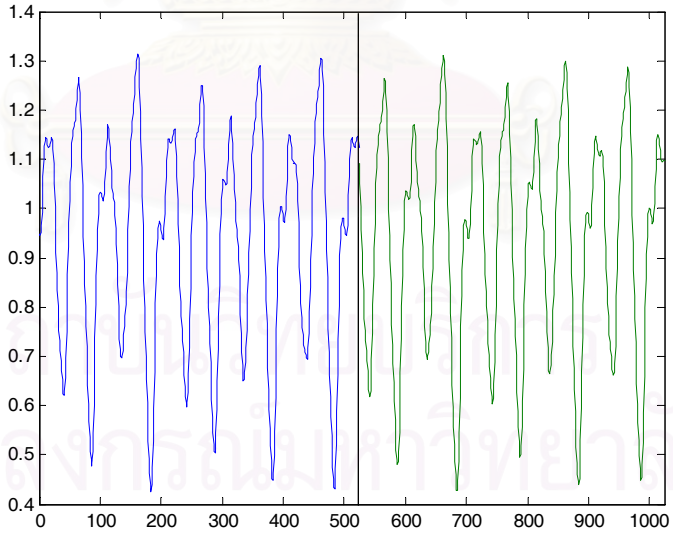

Figure 3.15: The Mackey-Glass Time series.

number of neuron node	NMSE
4	0.00294
8	0.00156
16	0.00119
32	**0.00092**

Table 3.3: Result of Prediction by the NN of Mackey-Glass

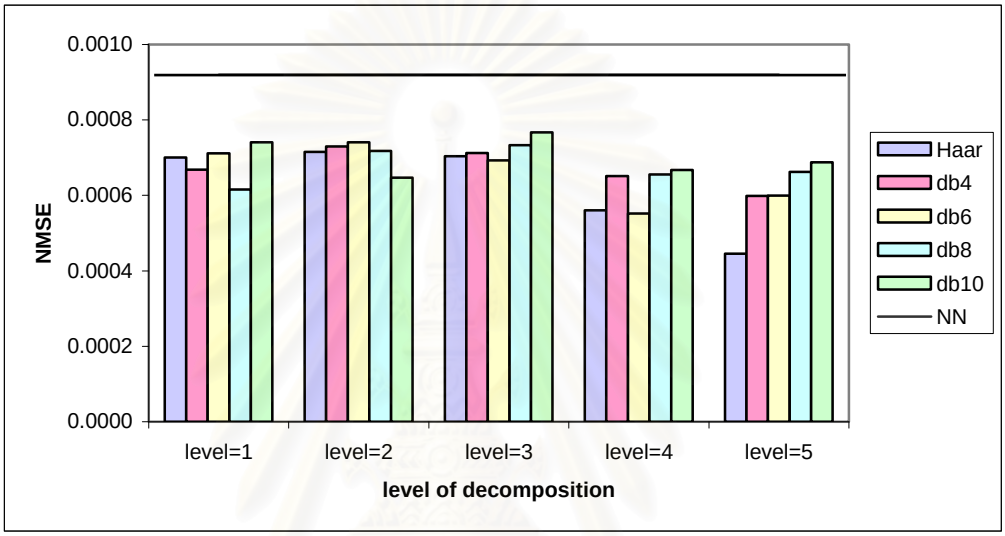

Figure 3.16: Comparison of NMSE between the NN and the WTNN with the number of neuron is 32.

Fig. 3.16 shows the comparison of NMSE between the NN and the WTNN of the Mackey-Glass time series. It indicates that the prediction by using the WTNN is better than the NN. Fig. 3.17 shows the comparison between observed value and predicted value. It is found that the optimal NMSE, 0.00045, is achieved with Haar wavelets and 5 levels of decomposition. It shows that the error of prediction decreased by 51.85%.

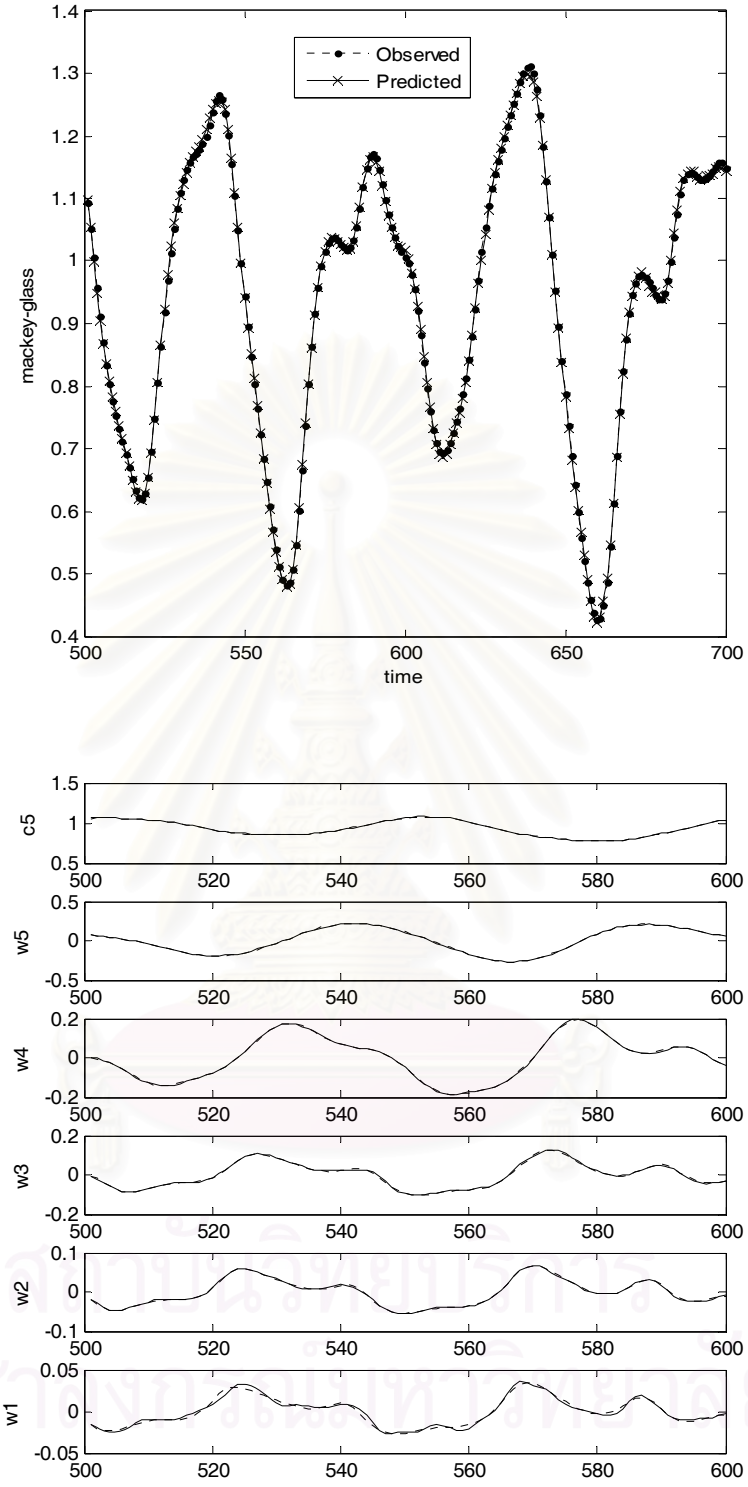

Figure 3.17: comparison between observed value and predicted value of the WTNN (Daubechies6 with 4 level of decomposition).

Besides, the scatter plot of propose method is closer to the ideal line than that of the NN as shown in Fig. 3.18 for the Crude Oil Prices, Fig. 3.19 for the Sunspot, and Fig. 3.20 for the Mackey-Glass time series. Table 3.4 shows comparison of the correlation coefficient of the NN to the WTNN that is consistent with the NMSE. Therefore, it supports that the proposed method yield better result than the NN.

Data	correlation coefficient	
	NN	WTNN
Curde Oil Prices	0.95196	0.95684
Sunspot	0.93479	0.94825
Mackey-Glass	0.99914	0.99978

Table 3.4: Comparison of the correlation coefficient between the NN and the WTNN.

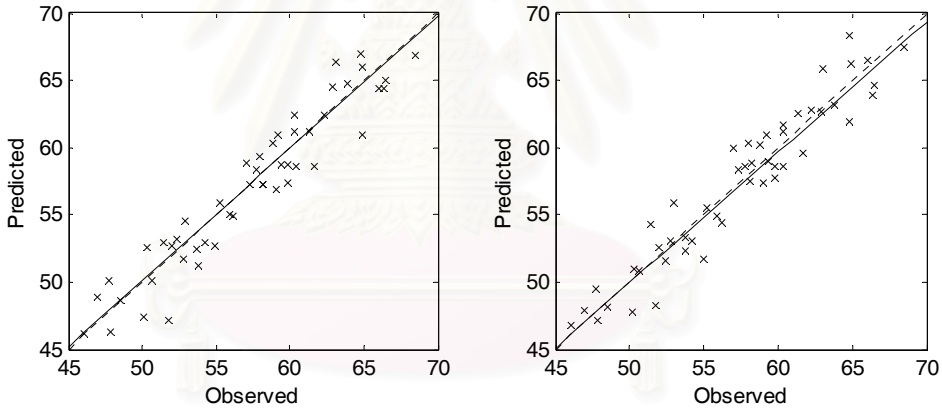

Figure 3.18: The scatter plot of the Crude Oil Prices data (a) the predicted by the WTNN versus the observed (b) the predicted by the NN versus the observed.

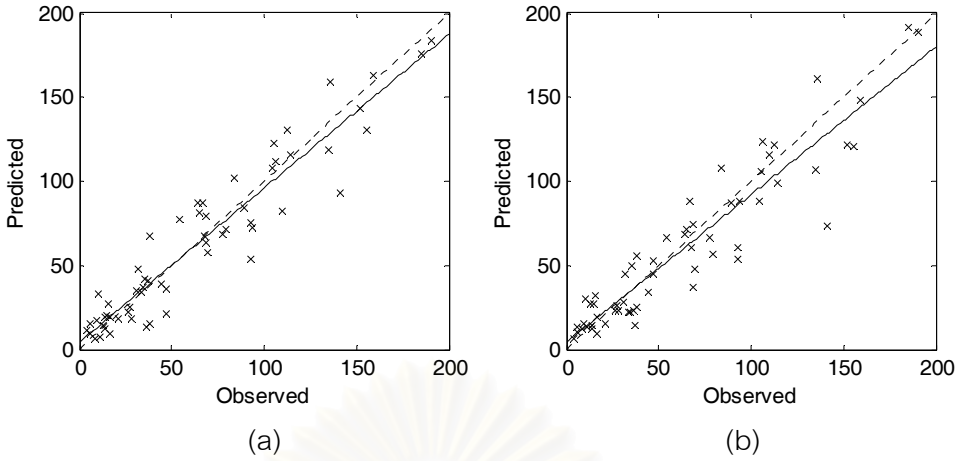

Figure 3.19: The scatter plot of the sunspot data (a) the predicted by the WTNN versus the observed (b) the predicted by the NN versus the observed.

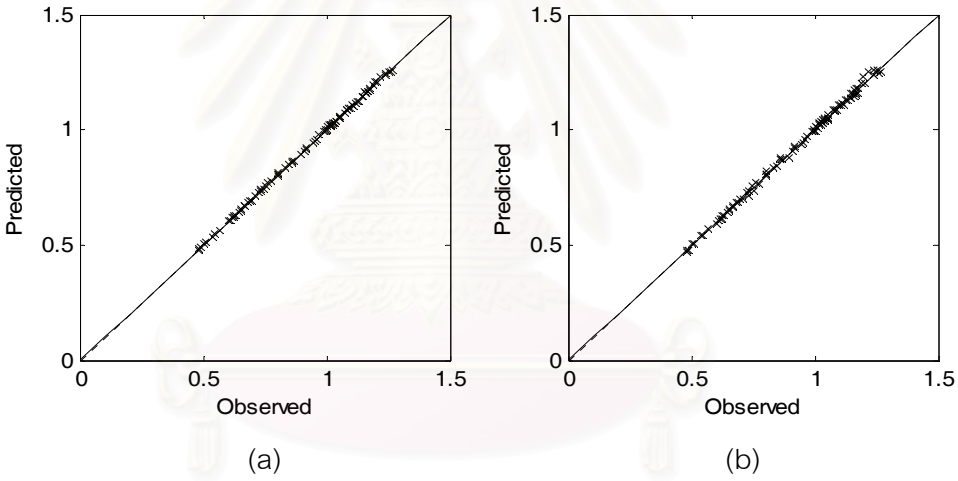

Figure 3.20: The scatter plot of the Mackey-glass time series (a) the predicted by the WTNN versus the observed (b) the predicted by the NN versus the observed.

CHAPTER IV

CONCLUSION AND DISCUSSION

This thesis focuses on the one step ahead prediction of time series using the wavelet transform and the neural network (WTNN). Three data sets were tested: the Sunspot (SS) series, the Mackey-Glass (MG) time series and the Crude Oil Price (COP). Experimentally, the level of wavelet decomposition ranges from 1 to 5 and we consider wavelets: Haar, Daubechies4, Daubechies6, Daubechies8, and Daubechies10.

The Normalized Mean Square Error (NMSE) of prediction from WTNN was compared to the NMSE of the prediction from the Neural Network (NN). In general, the WTNN yields less error (0.13059 to 0.101522 for the SS, 0.00092 to 0.00045 for the MG, and 0.10192 to 0.08936 for the COP) which signifies the success of our method. Nevertheless, the running time of the WTNN is more than the NN. Since, there are $J+1$ networks to be trained, where J is level of decomposition, where the NN has to train a single network.

It is worth noting that the error of prediction varies with the levels of wavelet decomposition as well as the choice of the wavelets. From the three data sets that we analyzed, it is found that only 3 levels of wavelet decomposition gave satisfactory results and Daubechies6 was a good candidate for the choice of the wavelets.

During the study, we found that there are some suggestions that should adjust to improve the result of prediction. Since, a wavelet transform is decomposition of the high frequency component and the low frequency component of the input signal at various scales. In particular at the coarsest scales of decomposition, we will get the highest frequency component. Experimentally, we found that the prediction of this component is not good compare with the other component. Thus, if the prediction of the highest component was applied by the other methods that more appropriate, its result of prediction may be more accurate. Moreover, our research applied the wavelet decomposition with only the wavelets family of Daubechies. However, there are many wavelets families that are interesting to be used in our method.

References

[1] H. Li and R. Kozma. 2003. A dynamic neural network method for time series prediction using the KIII model. *IEEE Transaction on Neural Networks*: 347-352.

[2] B. L. Zhang, R. Coggins, M. A. Jabri, D. Dersch, and B. Flower. 2001. Multiresolution forecasting for future trading using wavelet decompositions. *IEEE Transaction on Neural Networks* 12: 765-775.

[3] G. Zhang, B. Patuwo, and M. Hu. 1998. Forecasting with artificial neural network: The state of the art. *International Journal of Forecasting* 14: 36-62.

[4] A. B. Geva. 1998. ScaleNet—Multiscale neural network architecture for time series Prediction. *IEEE Transaction on Neural Network* 9: 1471-1482.

[5] S. Soltani, D. Boichu, P. Simard, and S. Canu. 2000. The long-term memory prediction by multi-scale decomposition. *Signal processing* 80: 2195-2205.

[6] A. Aussem and F. Murtagh. 1997. Combining neural forecasts on wavelet-transformed time series. *Connection Science* 9: 113-121.

[7] O. Renaud, J.-L. Starck, and F. Murtagh. 2005. Wavelet-Based Combined Signal Filtering and Prediction. *IEEE Transaction on System, Man, and Cybernetics* 35(6).

[8] Y. Chen, B. Yang, and J. Dong. 2006. Time series prediction using a local linear wavelet neural network. *Neurocomputing* 69: 449-465.

[9] G. Zheng, J.-L. Starck, J. Campbell, and F. Murtagh. 1999. The wavelet transform for filtering financial data streams. *Journal of Computer Intelligent Finance* 7(3): 18-35.

[10] C. S. Burrus, R. A. Gopinath, and H Guo. 1997. *Introduction to Wavelets and Wavelet Transforms: A Primer*. New Jersey: Prentice Hall.

[11] S. Haykin. 1994. *Neural Network: A comprehensive foundation*. 2nd ed. New York: Prentice Hall.

[12] M. Chester. 1993. *Neural network: A tutorial*. New Jersey: Prentice Hall.

[13] S. Soltani. 2002. On the use of the wavelet decomposition for time series prediction. *Neurocomputing* 48: 267-277.

[14] M. Mackey and J. Glass. 1977. Oscillation and chaos in physiological control systems. *Science* 197(4300): 287-289.

VITAE

Miss Piyamas Kanuan was born in October 23, 1981, in Chanthaburi, Thailand. She received a bachelor degree in Mathematics from the Department of Mathematics, Faculty of Science, Srinakharinwirot University, Thailand in 2003.