國立政治大學理學院應用物理研究所

碩士論文

Graduate Institute of Applied Physics, College of Science

National ChengChi University

Master Thesis

基於 EEMD 之倒傳遞類神經網路方法對用電量及黃金價格之預測

Forecasting electricity consumption as well as gold price by using an EEMD-based Back-propagation Neural Network Learning Paradigm

國立政治大學
National Chengchi University

蔡羽青
Tsai Yu Ching

指導教授：蕭又新博士
Advisor：Shiau Yuo Hsien

中華民國一〇〇年九月
September, 2011

摘　　要

本研究主要應用基於總體經驗模態分解法(EEMD)之倒傳遞類神經網路(BPNN)預測兩種不同的非線性時間序列數據，包括政大逐時用電量以及逐日歷史黃金價格。透過 EEMD，這兩種資料會分別被拆解為數條具有不同物理意義的本徵模態函數(IMF)，而這讓我們可以將這些 IMF 視為各種影響資料的重要因子，並且可將拆解過後的 IMF 放入倒傳遞類神經網路中做訓練。

另外在本文中，我們也採用移動視窗法作為預測過程中的策略，另外也應用內插法和外插法於逐時用電量的預測。內插法主要是用於補點以及讓我們的數據變平滑，外插法則可以在某個範圍內準確預測後續的趨勢，此兩種方法皆對提升預測準確度占有重要的影響。

利用本文的方法，可在預測的結果上得到不錯的準確性，但為了進一步提升精確度，我們利用多次預測的結果加總平均，然後和只做一次預測的結果比較，結果發現多次加總平均後的精確度的確大幅提升，這是因為倒傳遞類神經網路訓練過程中其目標為尋找最小誤差函數的關係所致。

國立政治大學
National Chengchi University

關鍵字：總體經驗模態分解法、倒傳遞類神經網路、用電量預測、黃金價格預測、超短時間負荷預測

Abstract

In this paper, we applied the Ensemble Empirical Mode Decomposition (EEMD) based Back-propagation Neural Network (BPNN) learning paradigm to two different topics for forecasting: the hourly electricity consumption in NCCU and the historical daily gold price. The two data series are both non-linear and non-stationary. By applying EEMD, they were decomposed into a finite, small number of meaningful Intrinsic Mode Functions (IMFs). Depending on the physical meaning of IMFs, they can be regarded as important variables which are input into BPNN for training.

We also use moving-window method in the prediction process. In addition, cubic spline interpolation as well as extrapolation as our strategy is applied to electricity consumption forecasting, these two methods are used for smoothing the data and finding local trend to improve accuracy of results.

The prediction results using our methods and strategy resulted in good accuracy. However, for further accuracy, we used the ensemble average method, and compared the results with the data produced without applying the ensemble average method. By using the ensemble average, the outcome was more precise with a smaller error, it results from the procedure of finding minimum error function in the BPNN training.

國立政治大學
National Chengchi University

Keywords: *Ensemble Empirical Mode Decomposition, Back-propagation Neural Network, electricity consumption forecasting, gold price forecasting, very-short term load forecasting*

Contents

國 立 政 治 大 學 National Chengchi University

Tables

國立政治大學
National Chengchi University

Figures

國 立 政 治 大 學
National Chengchi University

國 立 政 治 大 學
National Chengchi University

1. Introduction

Empirical Mode Decomposition (EMD), first introduced by Huang et al. (1998) was developed for dealing with nonlinear and non-stationary data. The method is empirical, intuitive, direct and adaptive. By using EMD, the time-series data is decomposed into several intrinsic modes which are nearly periodic and independent with each other. The intrinsic modes may have its own physical meaning based on local characteristic time scale. For instance, if an intrinsic mode is periodic with a time scale of one month, it can be recognized as the monthly component. Similarly, an intrinsic mode with scale of three months means the seasonal component. Generally, complex time-series data is often mixed up with many different signal sources thus difficult to understand their meaning. But, through the EMD decomposition method we can divide them into several meaningful intrinsic modes, allowing us to closely analyze their characteristics.

EMD has been successfully applied in different fields such as ocean waves (Hwang et al., 2003), earthquake engineering (RR Zhang et al., 2003), wind engineering (Li and Wu, 2007), biomedical engineering (Liang et al., 2005) and structured health monitoring (R Yan, 2006). The above applications are all related to natural science and engineering. However, in the recent years, there have been more and more applications in social sciences. For example, financial time series analysis (Huang et al., 2003b), transport geography (MC Chen, 2010), disease transmission (Cummings et al., 2004), as well as combined with artificial neural networks (ANNs) to forecast crude oil price (Lean Yu et al., 2008).

In addition to the EMD originally developed, there is an improved EMD, known as Ensemble EMD (EEMD, Wu and Huang, 2009), which was proposed to solve the mode mixing problem of EMD. In this study, we first applied the EEMD method and

國 立 政 治 大 學 National Chengchi University

then combined it with ANNs to electricity consumption forecasting and gold price forecasting.

Electricity consumption forecasting (i.e. load forecasting), is commonly classified into four categories: long-term load forecasting, medium-term load forecasting, short-term load forecasting and very short-term load forecasting. Long-term load forecasting (5, 10 and 20 years ahead) is used for system planning, scheduling construction of new generation capacity and the purchasing of generation units (Jia et al., 2001). Medium-term load forecasting (a few months to 5 years ahead) is applied to maintenance scheduling, coordination of load dispatching and the setting of prices (Jia et al., 2001). Short-term load forecasting (hourly forecasting from one day to one week ahead) is usually used for optimal generator unit commitment, fuel scheduling, hydro-thermal co-ordination, economic dispatch, generator maintenance scheduling, the buying and selling of power and security analysis. Very short term load forecasting (few minutes to an hour ahead in the future) is often used for security assessments and economic dispatching, real-time control and real-time security evaluation (Jia et al., 2001). In this study we focused on very short-term load forecasting and forecast load of one hour ahead.

國立政治大學 National Chengchi University

Surprisingly, there has been very little research on very short-term load forecasting. Yang et al. (2005) used a method based on the chaotic dynamics reconstruction technique and fuzzy logic theory on the load data of Shandong Heze Electric Utility, (China). Their results demonstrated that the proposed approach could calculate 15 minutes ahead load forecasting with accurate results. James W. Taylor (2008) used minute-by-minute British electricity demand observations to evaluate different forecasting methods, including ARIMA modeling and Holt-Winters' exponential smoothing method for prediction between 10 and 30 minutes ahead. Liu et al. (1996) applied the fuzzy logic method and ANNs to the previous 30 minute-by-minute observations as input for the online forecasting process to show that the methods could outperform the simplistic non-seasonal AR model.

Unlike the various aforementioned methods, we used the EEMD-based ANN algorithm for one hour ahead forecasting. ANNs are massively parallel and robust. They contain complicated architectures of interconnected processing elements. They can learn complex linear or non-linear input-output mappings between data sets of measurements and future demand values. Based on the features presented in the data they can be designed adaptively for learning and responding with high-speed computation. The ANNs have been applied to many areas especially when dealing with the issues of forecasting.

國立政治大學 National Chengchi University

Hamid et al. (2004) applied ANNs to financial forecasting. Their goal was to forecast the volatility of S&P 500 Index future prices, and compare volatility forecasts from ANNs with the implied volatility from S&P 500 Index futures options using the Barone-Adesi and Whaley (BAW) model for pricing American options on futures. Lean Yu et al. (2008) used an EMD-based neural network ensemble learning paradigm for world crude oil spot price forecasting. In Yu’s study, West Texas Intermediate (WTI) crude oil spot price and Brent crude oil spot price were used to test the effectiveness of the method. Their results showed that the EMD-based neural network ensemble-learning model outperformed the other forecasting models in terms of criteria. Lean Yu et al. (2010) proposed an EMD-based multi-scale neural network learning paradigm to predict financial crisis events for early-warning purposes. They took the currency exchange rate series of the South Korean Won (KRW) and Thai Baht (THB) as training targets. Their tests showed that the EMD-based multi-scale neural network learning paradigm was superior to other classification methods and single-scale neural network learning paradigm when formulating currency crisis forecasting. Feng Ping et al. (2009) applied EMD-based ANNs to precipitation-runoff forecasting. They took the annual precipitation series from 1956 to 2000 from the sub-water resource regions of upper Lanzhou, China as historical training data, and showed that the EMD could decompose the data into a multi-time scale sub-series for finding their local change rule. The results demonstrated that the EMD-based ANNs model presented higher accuracy than any other models.

The ANNs are also widely used in short-term load forecasting. Ruey-Hsun Liang and Ching-Chi Cheng (2002) used an approach based on combing ANNs with the fuzzy system and applying it to data from the Taiwan Power Company. Nahi Kandil et al. (2006) applied multi-layered feed-forward ANNs by using data from Hydro-Quebec databases for forecasting. They demonstrated ANNs' capabilities without using load history as an input, instead final temperatures were the only data considered in their load forecasting procedure. Mohsen Hayati and Yazdan Shirvany (2007) used Multi-Layer Perceptron (MLP), a kind of architecture of ANNs, on data from a three year time period (2004-2006) from the Illam (Middle Eastern country, west of Iran) region, while G.A. Adepoju et al. (2007) applied ANNs to the Nigeria Electric power system.

Gold has been mined since ancient times. With recent growth in production, more than a third of the world's gold that has ever been mined, in just the last thirty years. The consumption of gold differs by application type: industrial, dental technology, jewelry products and inventory. Jewelry consistently accounts for over two-thirds of the gold demand, but in markets with poorly developed financial systems or markets experiencing crisis, gold is an attractive investment. The demand-supply equilibrium and inflation cause gold price to fluctuate. Gold is commonly a popular hedge instrument for investors against devaluation of the US dollar. In recent years as the value of US dollar has decreased relative to other major currencies, the price of gold

國立政治大學 National Chengchi University

has experienced a secular increase. The dramatic rises in gold price since the start of 2009 may have resulted from investors looking to preserve their wealth.

There have been several studies on gold price analysis: Baker and Tassel (1985) used regression results to support the theoretical analysis leading to the prediction. Akgiray, Booth, Hatem and Mustafa (1991) used the GARCH model to verify time decency of gold price. In comparison with statistical techniques, engineering- based systems, such as neural networks, make less restrictive assumptions on the underlying distribution. Mirmirani and Li (2004) used neural networks and genetic algorithm to analyze gold price. Shahriar Shafiee and ErkanTopal (2010) used long-term trend reverting jump and dip diffusion model and took monthly historical gold price data from January 1968 to December 2008, to forecast gold price for the next ten years. Yen-Rue Chang (2011) used EEMD to decompose monthly gold price data into several IMFs to observe their important properties. Following the work of Yen-Rue Chang (2011), we used the decomposed IMFs as input factors for gold price forecasting.

This research aimed to forecast using EEMD-based ANN algorithm, more specifically; we used a back-propagation neural network (BPNN) which is a kind of ANN architecture. Our testing targets were electricity load data and gold price data. We performed one hour ahead load forecasting and gold price forecasting for 2011. Section 2 gives brief introduction to the basic concept of EEMD and BPNN. In section 3, we describe the subject data, and introduce our experiment strategy, and some important measures used. Experiment results and forecast performance are discussed in Section 4, along with the results of improvement by ensemble average. In Section 5, we will present our conclusion and outlook.

國立政治大學 National Chengchi University

2. Methodology

2.1. Empirical mode decomposition

In the past, we usually used a spectral analysis method called the Fast-Fourier-Transform (FFT) to analyze time-series data; however FFT posed a few problems:
If the nonlinear and non-stationary degrees of the time-series data were to increase, the results of the FFT would produce large sets of physically meaningless harmonics.

Nowadays, we utilize a new spectral analysis method called Hilbert-Huang-Transform (HHT). While the HHT can solve the problems with the FFT, it is still limited in that it can only be used for data which are symmetric in relation to the local zero mean. We would thus need to first use the empirical mode decomposition (EMD) which was proposed by Huang et al. (1998) to decompose data into several intrinsic mode functions (IMFs). The IMFs are all symmetric in relation to the local zero mean so that we can use HHT on them.

The EMD is a data analysis method which can be used in dealing with non-linear and non-stationary time series data. It assumes that all time series data can be decomposed into a sum of oscillatory functions known as IMFs. The IMFs, based on the local characteristic scale by itself, have to satisfy the two following conditions:

(1) IMFs should have the same numbers of extrema (including maxima and minima) and zero-crossings, or differ at most by one;

(2) At any point, the mean value of the envelope defined by local maxima and the envelope defined by local minima is zero, meaning the IMFs should be symmetric in relation to the local zero mean.

國 立 政 治 大 學 National Chengchi University

We then pose the question, ‘how can we extract the IMFs from original data’? We can use the following sifting process:

(1) Identify all the local maxima and minima of the time series x(t)

(2) Connect all the local maxima and minima by cubic spline interpolation to generate its upper and lower envelopes $e_{max}(t)$ and $e_{min}(t)$

(3) Calculate mean m(t) from upper and lower envelopes point-by-point as

$$m_1(t) = (e_{max}(t)+ e_{min}(t))/2$$

(4) Calculate the difference between the time series data x(t) and the mean value $m_1(t)$, than get $h_1(t)$ as : $h_1(t) = x(t) - m_1(t)$

(5) Check the properties of $h_1(t)$:

If $h_1(t)$ doesn’t satisfy the conditions of IMF, replace x(t) with $h_1(t)$ and repeat (1)-(4) until $h_k(t)$ satisfies the stopping criterion:

$$\mathrm{SD} = \sum_{t=1}^{T} \frac{\left[h_{(k-1)}(t) - h_k(t)\right]^2}{{h_{k-1}}^2(t)}$$

A typical value for SD can be set between 0.2 and 0.3.

On the other hand, if $h_1(t)$ (or $h_k(t)$) satisfies the conditions of IMF, then it should be an IMF, and denote $h_1(t)$ (or $h_k(t)$) as the first IMF $c_1(t)$. Then we separate the IMF $c_1(t)$ from x(t) to get the residue $r_1(t)$: $x(t) - c_1(t) = r_1(t)$

(6) Now we replace x(t) with $r_1(t)$ and repeat steps (1)-(5) to get $c_2(t).c_3(t).c_4(t).c_5(t)$………..$c_n(t)$ and final residue $r_n(t)$.

The sifting process is stopped by any of the following predetermined criteria: either the component $c_n(t)$ or the residue $r_n(t)$ becomes so small that it is less than the predetermined value of the substantial consequence, or the residue $r_n(t)$ becomes a monotonic function from which no more IMFs can be extracted (Huang et al. 1998).

At the end of the sifting process, the original time series can be expressed as

國立政治大學 National Chengchi University

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t)$$

Where n is the number of IMFs, $r_n(t)$ is the final residue, also the trend of x(t), and $c_i(t)$ represents IMFs which are nearly orthogonal to each other. After the sifting process, the original data set is decomposed into these IMFs which represent high frequency to low frequency, and every IMF may have its own physical meaning. So we can regard the EMD as a filter to separate high to low frequency modes, and apply HHT on each IMF, allowing us to explore their meaning

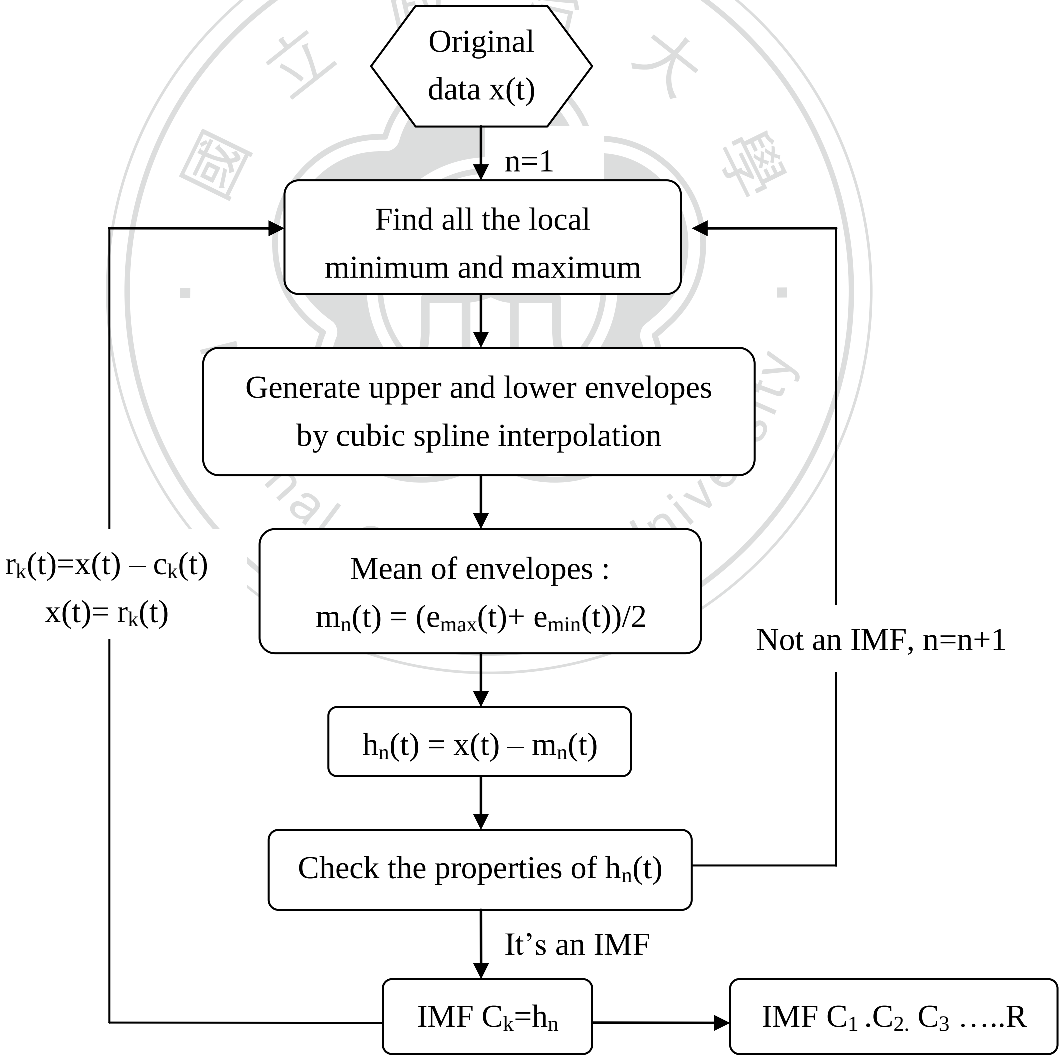

Original data x(t)
n=1
Find all the local minimum and maximum
Generate upper and lower envelopes by cubic spline interpolation
Mean of envelopes : $m_n(t) = (e_{max}(t)+ e_{min}(t))/2$
$h_n(t) = x(t) – m_n(t)$
Check the properties of $h_n(t)$
Not an IMF, n=n+1
It's an IMF
IMF $C_k=h_n$
IMF $C_1 .C_{2.} C_3$ …..R
$r_k(t)=x(t) – c_k(t)$
$x(t)= r_k(t)$

Figure 2.1 The flowchart of EMD

2.2. Ensemble EMD

EMD has proved to be a useful data analysis method for extracting signals from nonlinear and non-stationary data. However, EMD still has its defects: when the original data is intermittent, a single IMF may consist of either signals of widely disparate scales, or a signal of similar scale belonging to different IMF components; this phenomenon is called “mode mixing”. When mode mixing occurs, an IMF may cease to have physical meaning by itself. Thus, EEMD was proposed by Wu and Huang (2004) for overcoming the problem. We know that all observed data are mixed with true time series and noise. Even if data is collected by separate observations with different noise levels, the ensemble mean is close to the true time-series. This means that we can extract the true meaningful signal from data by adding some white noise. Adding white noise could provide a uniformly distributed reference scale, and help EMD to overcome the mode mixing problem.

國 立 政 治 大 學
National Chengchi University

The simple procedure of EEMD is as follows:

(1) Add a white noise series to the original data.

(2) Decompose the data with added white noise into IMFs.

(3) Repeat the previous two steps iteratively, and add different white noise each time, finally we obtain the ensemble means of corresponding IMFs of decompositions.

$$x_i(t) = x(t) + w_i(t)$$

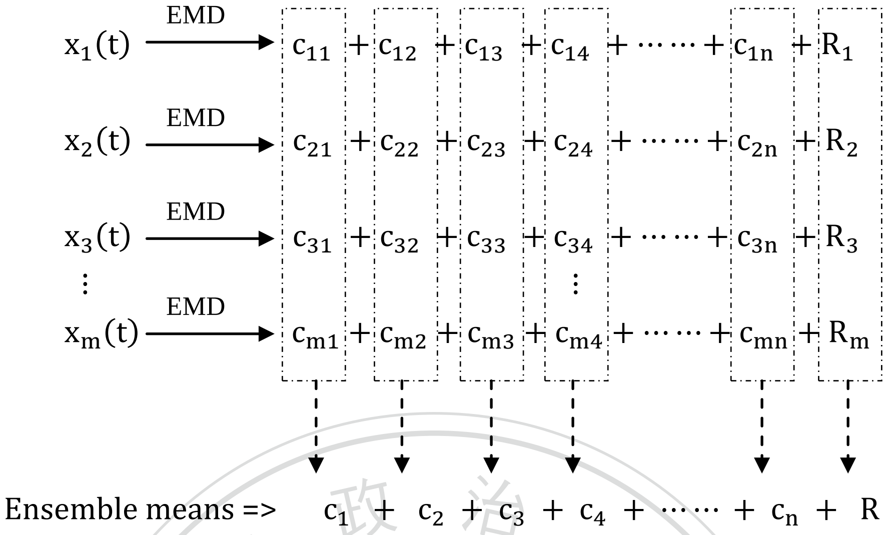

x₁(t)
EMD
c₁₁ + c₁₂ + c₁₃ + c₁₄ + ⋯⋯ + c₁ₙ + R₁
x₂(t)
EMD
c₂₁ + c₂₂ + c₂₃ + c₂₄ + ⋯⋯ + c₂ₙ + R₂
x₃(t)
EMD
c₃₁ + c₃₂ + c₃₃ + c₃₄ + ⋯⋯ + c₃ₙ + R₃
⋮
⋮
xₘ(t)
EMD
cₘ₁ + cₘ₂ + cₘ₃ + cₘ₄ + ⋯⋯ + cₘₙ + Rₘ
Ensemble means =>
c₁ + c₂ + c₃ + c₄ + ⋯⋯ + cₙ + R

Figure 2.2 Diagram to illustrate the procedure of Ensemble EMD

In figure2.2, x(t) is the original data, $w_i(t)$ is the added white noise series, $x_i(t)$ is the original data mixed with different white noise series, c_{mn} is the IMFs, and R_m represents the residues.

However, there is a well-established statistical rule proved by Wu and Huang (2004) to control the effect of adding white noise:

$$\varepsilon_n = \frac{\varepsilon}{\sqrt{N}}$$

In this formula, N represents the number of ensemble members, ε is the amplitude of the added noise and ε_n is the final standard deviation of error defined as the difference between the input signal and the corresponding IMFs. Empirically, the number of ensemble members N is always set to 100 and the ε_n is always set to 0.1 or 0.2. The procedure of adding white noise successfully makes signals of comparable scales to collate in one IMF, and then cancels itself out. Therefore, the EEMD, which

國立政治大學 National Chengchi University

can successfully reduce the chance of mode mixing, is a really good, substantial improvement over the original EMD.

2.3. Artificial neural networks

Artificial neural networks (ANNs), which have been widely used for data prediction in many application domains, are a kind of intelligent learning paradigm. They have been developed over the last 50 years, and the first, simplest model of ANNs which is called Perceptron was proposed by Frank Rosenblatt in 1957.

Nowadays, the most popular model of ANNs is feed-forward back-propagation neural network (BPNN). It adopts the Widrow-Hoff learning rule (i.e. least mean squared (LMS) rule) (Hagan et al., 1996) and different algorithms such as the steepest descent method, Newton's method and Levenberg-Marquardt (LM) algorithm to train the network. ANNs are designed to imitate the biological neural system. Typically, ANNs contain three sections: neurons, connection weights and transfer functions.

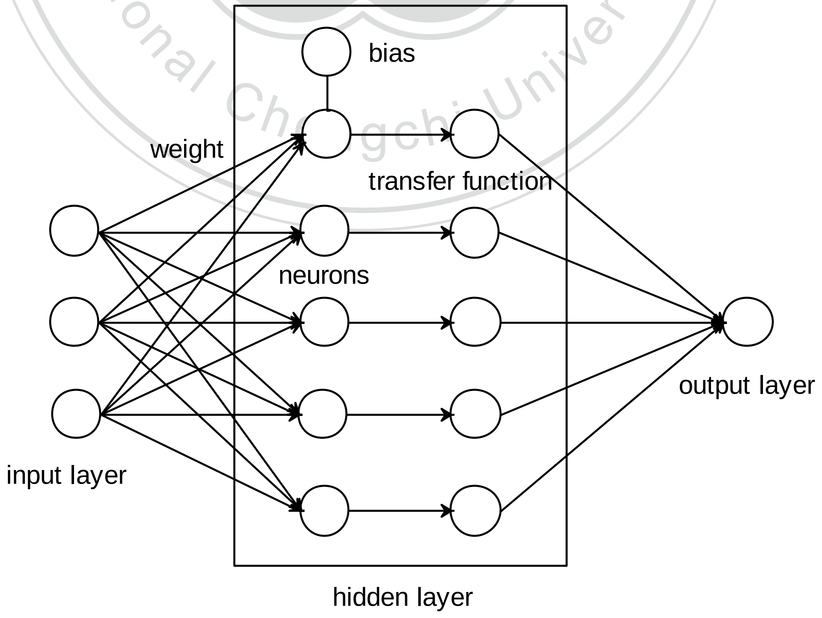


Figure 2.3 Simple structure chart of three-layers neural network

The connection weights present the strength between neurons. A larger weight means the connection is stronger while a smaller weight presents a weaker link. The neurons process input signals overlap at the neuron and are sent to the transfer function for generating output value. The transfer functions are created to restrict the output value of ANNs. Because different kinds of ANNs are used in different ways, they need different transfer functions to generate different results.

The feed-forward BPNN, which is the most popular model of ANNs for time-series data prediction, uses three kinds of transfer functions: log-sigmoid function, hyperbolic function and linear function. Log-sigmoid function and hyperbolic function are often used in the hidden layer. The log-sigmoid function takes an output value between 0 and 1 while the hyperbolic function takes an output value between 1 and -1. These two transfer functions are both differentiable so that the training algorithms may work for the network, otherwise the linear function is usually put in the output layer. It can produce values of any number.

$$Linear\ function： \quad Purelin(x) = x$$

$$Logsigmoid\ function： \quad Logsig(x) = \frac{1}{1+e^{x}}$$

$$Hyperbolic\ function： \quad Tansig(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$$

In this thesis, we used the feed-forward BPNN for modeling the decomposed IMFs and the residual component. In the BPNN, there is an important parameter called the “mean square error function”, which is a function of weights. Since our goal was to minimize the mean square error function, we had to adjust the connection weights iteratively by training the network. The mean square error function could be presented as:

國立政治大學 National Chengchi University

$$F(w) = \frac{1}{N}\sum_{i=1}^{N}[e_i(w)]^2 = \frac{1}{N}\sum_{i=1}^{N}[t_i - a_i(w)]^2$$

where a_i is the final output value, a function of weight. t_i is the target value, and e_i is the error between the values a_i and t_i. On the other hand, the input value of the *m*th layer's *i*th neuron is the nonlinear function of the output value of the *(m-1)*th layer's neurons:

$$a_i^m = f\left(\sum_{j=1}^{N} w_{ij}^m a_j^{m-1} + b_i^m\right) = f(n_i^m)$$

$$n_i^m = \sum_{j=1}^{N} w_{ij}^m a_j^{m-1} + b_i^m$$

The function $f(n_i^m)$ is the transfer function previously discussed. w_{ij}^m is the weight between the *m*th layer's *i*th neuron and *(m-1)*th layer's *j*th neuron, b_i^m is the bias of *m*th layer's *i*th neuron, a_j^{m-1} is the output value of *(m-1)*th layer, and a_i^m is the output value of *m*th layer.

In the history of BPNN, there have been several algorithms used to train the network for adjusting the weight. Here we adopted the Levenberg-Marquardt (LM) algorithm, which combines the advantages of the steepest descent method and Newton's method:

$$W^m(k+1) = W^m(k) - \left(H^m(k) + \mu_k I\right)^{-1} g^m(k)$$

$$H^m(k) = \nabla^2 F(W^m(k))$$

$$g^m(k) = \nabla F(W^m(k))$$

立 政 治 大 國 學 National Chengchi University

where $W^m(k)$ is the matrix of weights in the *m*th layer after the *k*th adjustment. H(k) is called the “Hessian matrix”, which is the second derivative of the mean square error functions, and g(k) is the first derivative of F(w). I is the identity matrix, and μ_k is the control parameter.

The control parameter μ_k changes iteratively, set with a big value in the beginning. Meanwhile, the LM algorithm will equal to the steepest descent method. The steepest descent method converges quickly when our result is still far from the optimal point, but as the result gets closer to the optimal point, the convergence speed gets slower, so the method costs much more times to find the optimal result.

The μ_k becomes a very small value in the later period of training, meanwhile the LM algorithm will become equal to Newton’s method, which converges quickly when the result is approaching to the optimal result.

The significant reason why we chose BPNN as our prediction tool was that the BPNN is usually regarded as a “universal approximator” (Hornik et al., 1989). Hornik et al. found that a three-layer BPNN can approximate any continuous function arbitrarily well with an identity transfer function (i.e. linear transfer function) in the output layer and logistic functions (i.e. log-sigmoid function and hyperbolic function) in the hidden layer.

In practice, the neural networks with one and occasionally two hidden layers are widely used and perform well. In this thesis, we utilized the four-layer feed-forward BPNN. Moreover, the number of neurons in the hidden layers were set to the same value as the IMFs or this value plus two, respectively, since the number of neurons in the hidden layer can range from one-half to two times (Mendelsohn, 1993) the sum of input and output numbers.

國立政治大學
National Chengchi University

The applied complete four-layer BPNN is shown as follows:

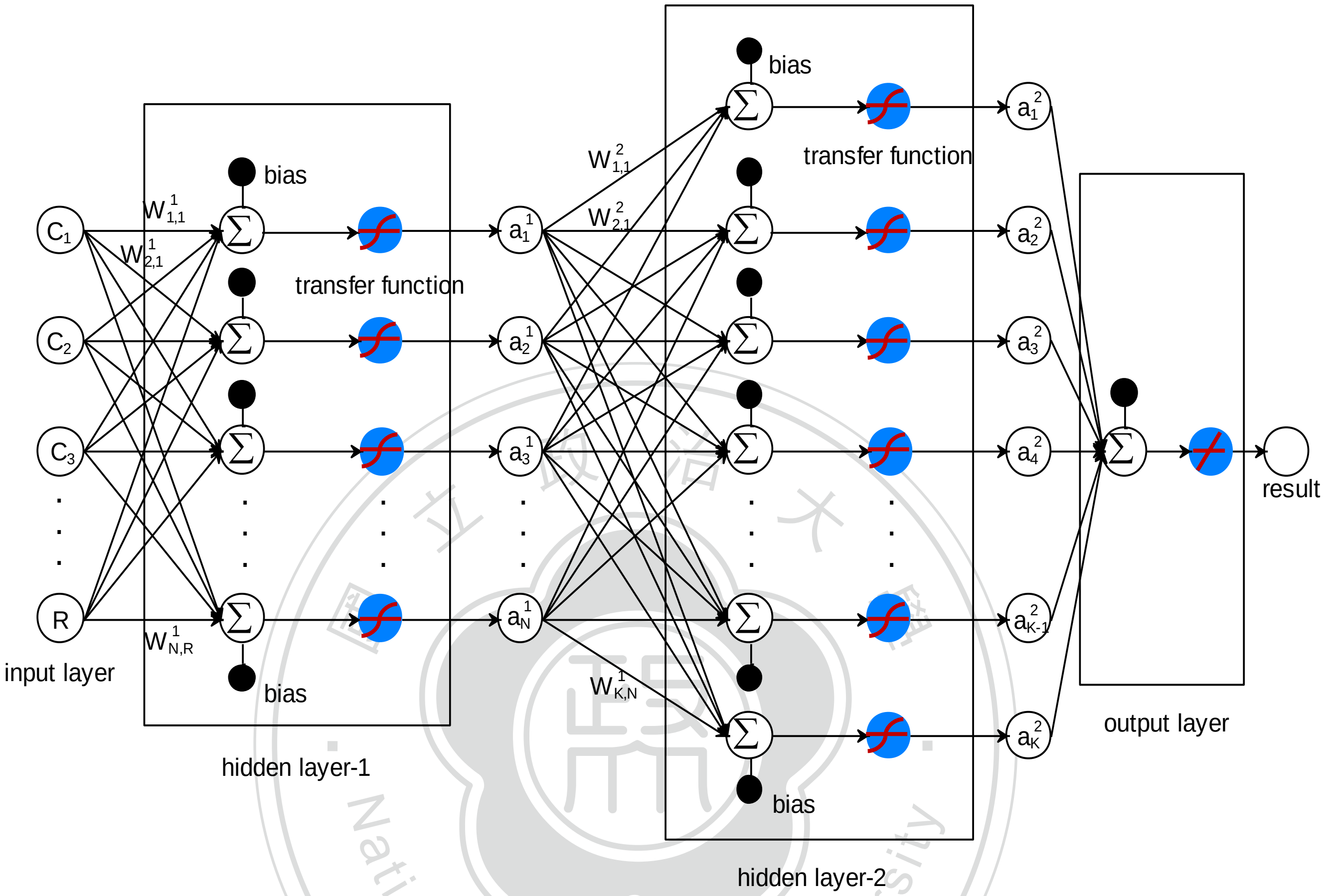
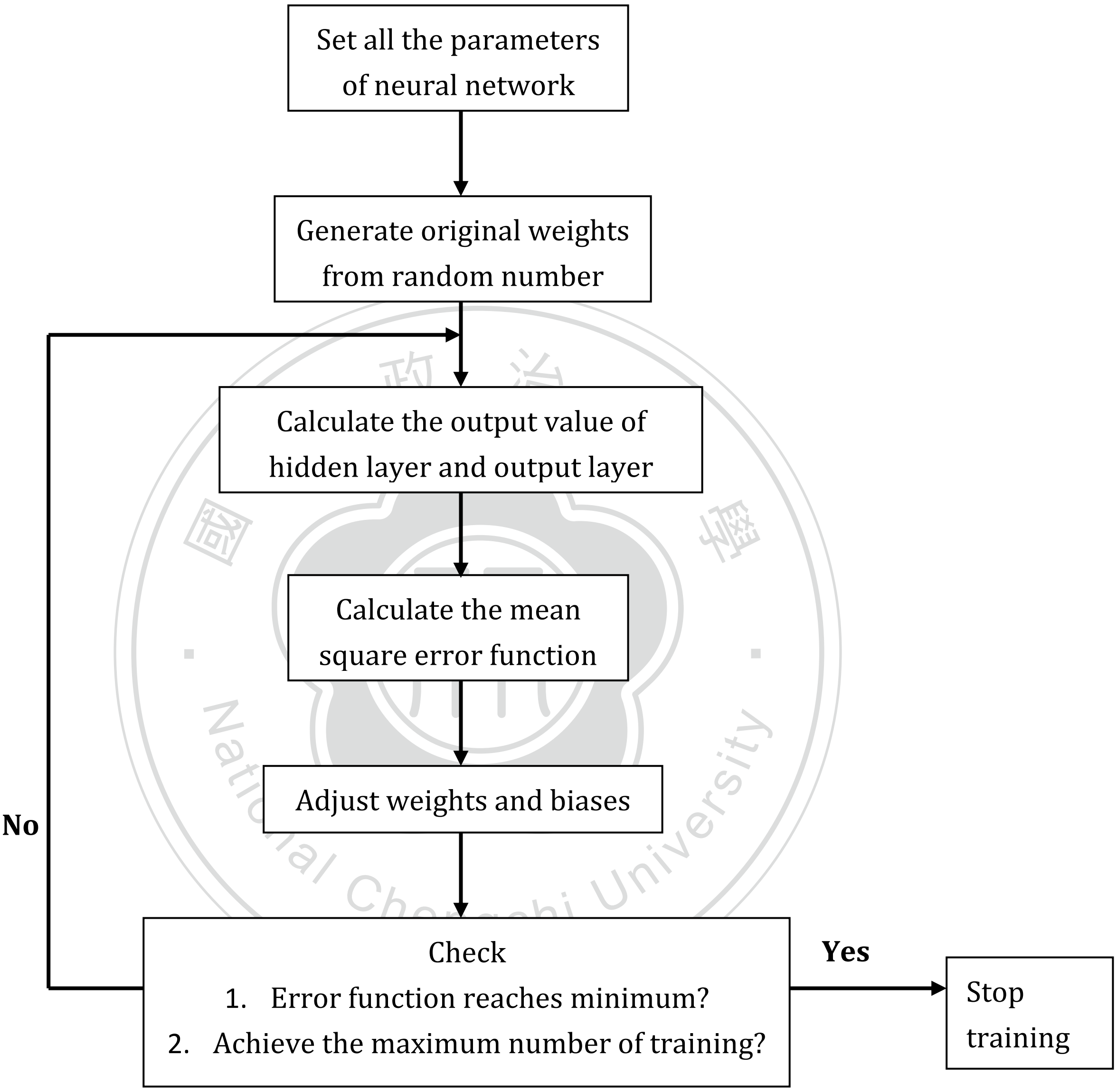


Figure 2.4 The structure chart of complete four-layer BPNN

National Chengchi University

Below is a flowchart of the BPNN learning progress which can help us to understand it more clearly.


Set all the parameters of neural network
Generate original weights from random number
Calculate the output value of hidden layer and output layer
Calculate the mean square error function
Adjust weights and biases
Check
1. Error function reaches minimum?
2. Achieve the maximum number of training?
No
Yes
Stop training

政 治 國 學 National Chengchi University

Figure 2.5 The flowchart of training progress of BPNN

2.4. Cubic spline interpolation and extrapolation

Sometimes time-series data may be missing values. In the mathematical field of numerical analysis, interpolation is a method which is applied to fill the gaps of missing values. For generating those intermediate values, there are several interpolation methods, such as piecewise constant interpolation, linear interpolation, polynomial interpolation and cubic spline interpolation. In contrast with linear interpolation, cubic spline interpolation uses low-degree polynomials in each of the intervals, and makes the intervals of data points smooth.

Extrapolation is a method which is used for forecasting outside of the known values for a given range, but the results are often less meaningful. If applied to a limited range of time-series data sets, extrapolation can extend data points, and find the variation tendency of the future of the time-series. Extrapolation uses the same methods as interpolation. Linear extrapolation uses the last two data values to create a tangent line at the end of known data, extending it beyond the limit. However, it only provides accurate results when used to extend the graph of an approximately linear function or not too far beyond the known data. Otherwise, cubic spline extrapolation, which needs more than the two values of the end of data, can create a low-degree polynomial curve which is extended beyond the known data. It is suitable for data which has curve properties, or data which we have already applied the cubic spline interpolation to.

In the use of the electricity load data of this study, the cubic spline interpolation method was applied to smooth the data points. Cubic spline extrapolation was employed to generate extra points to be included in our forecasting data before using the EEMD-based ANNs.

立 政 治 大 國 學 ‧ ‧ National Chengchi University

2.5. EEMD-based neural network learning paradigm

Artificial neural networks, which have already been applied in many practical forecasting cases, still have a lot of room for improvement. A crucial challenge nowadays is to improve the performance of the forecasting of artificial neural networks. Some researchers advocate the cross-validation technique, but it may be inadequate when the problem is complex, due to the method being based on a single series representation for entire a training process. Consequently, many scholars have done a lot of extensive research, such as Wavelet-based ANN (Ajay Shekhar Pandey et al., 2010), Fuzzy BP (Hari Seetha and R. Saravanan, 2007), PCA-based ANN (Wei Sun, 2010) and EEMD-based ANN (Lean Yu et al., 2008). In this study, the EEMD-based ANN was employed to decompose time-series into several IMFs. Each IMF had its own physical meaning so that we could regard these IMFs as the input variables of ANN. Figure 2.6 illustrates the basic structure of EEMD-based neural network.

國立政治大學 National Chengchi University

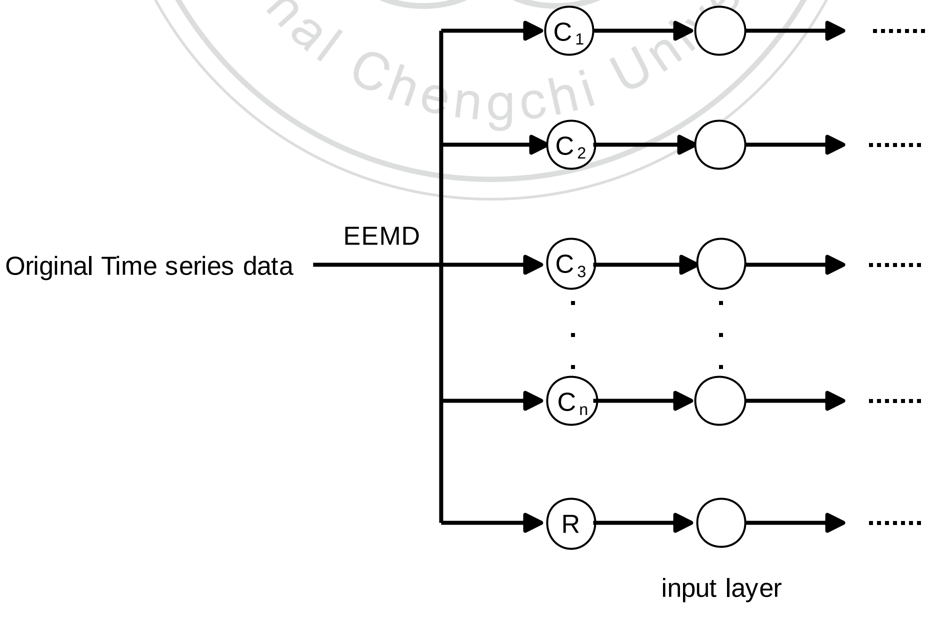


Figure 2.6 Simple illustration of EEMD to form the inputs of neural network

The EEMD-based neural network learning paradigm contains following steps:

(1) Decompose the original data into IMFs and residual components using EEMD.

(2) Put IMFs as the input variables into the four-layer BPNN.

(3) The input data set and target data set are extracted from the decomposed IMFs and original data.

(4) In the program, the input values and target time-series data will be first manipulated to a range between 1 and -1, and then divided into training , testing and validation sets. When parameters are all fixed, the four-layer BPNN structure will be created and start to train the network.

(5) In the network, the input data and target data will be compared and the optimal weights are found.

(6) The other input data will input to the trained network, and the forecasting data is generated. Then compare the forecasted data with real data.

To summarize, in this study we used EEMD to decompose the NCCU electricity load data into several meaningful IMFs, and then combined extrapolation with BPNN to forecast the electricity load of every hour. However the data used first underwent cubic spline interpolation. The forecasting results are shown in the next section.

國立政治大學 National Chengchi University

3. Forecasting experiments

3.1. Data description

3.1.1. Electricity load data from NCCU

The used electricity load data in this study was obtained from Office of General Affairs of National Chengchi University (NCCU). In NCCU, the consumption of electricity mainly comes from ten regions, labeled as GCB1 to GCB10. Among these ten regions, only three are independent buildings while the remainders contain numerous different sources. The three important buildings which contribute a majority of the electricity consumption of NCCU are the Information Building (GCB2), College of Commerce Building (GCB5) and General Building (GCB10). In this thesis, we focused on these three buildings because the electricity is controlled more easily. Generally, the electricity consumption of NCCU gradually increases after the anniversary celebrations, May 20, of every year. This is mainly due to the approach of summer and increasing temperatures prompting more use of air-conditioners and thus more consumption of electricity.

The electricity load data possesses some important properties. To illustrate these properties, we took the hourly data from May 5, 2008 to June 6, 2008 of GCB10 for an example. In Figure 3.1, we can easily observe that the data regularly changes from high to low, with peaks and valleys. Every five higher peaks follow two smaller peaks. This is illustrates the change from day to night (or peak hours to off-peak hours) and weekdays to weekends. These two important patterns are crucial factors influencing the forecasting performance.

In this study we used June 6, 2008 of GCB10 as a forecasting example. The pre-one-week data of June 6, 2008 was used as the input for historical data. We also used an ensemble average method to improve the forecasting performance, the load

國立政治大學 National Chengchi University

data of May 22, 2008 of GCB10 will be the testing sample. The results are presented in the section 4.

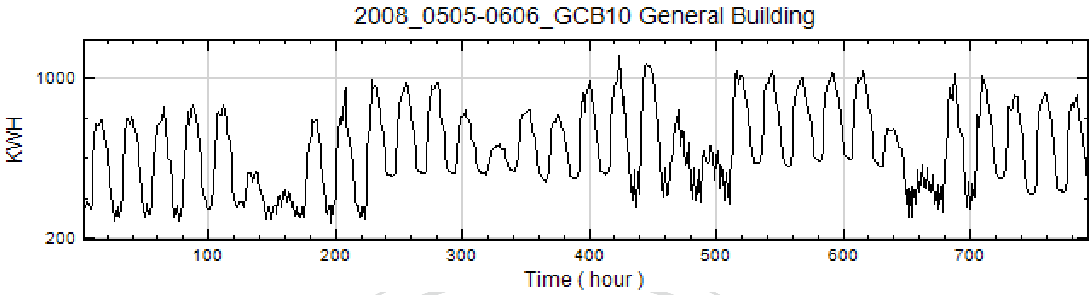
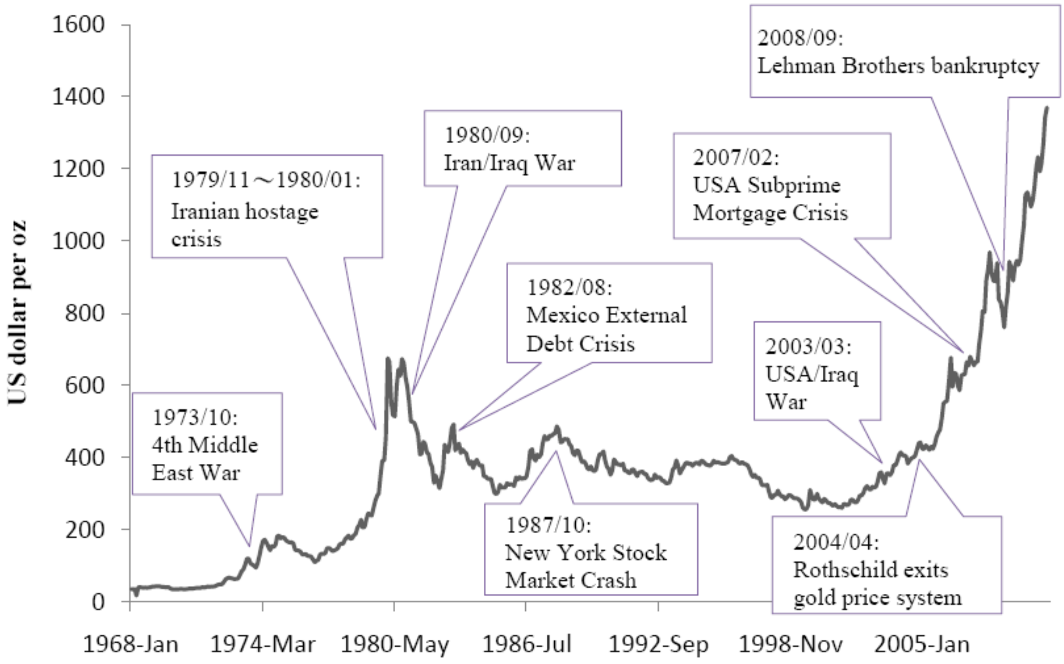


Figure 3.1 Hourly electricity load curve of GCB10 from 2008.5.5 to 2008.6.6

3.1.2. Gold price daily data

The gold price daily data used in this study was downloaded from website "Wikiposit" (http://wikiposit.org/w?filter=Finance/Commodities/). The data ranges from 1968 to the present (2011.9.1) using a unit of US dollar per oz ($/ounce). The historical gold price data was analyzed to see properties using EEMD by Yen-Rue Chang (2011). In her study the historical monthly gold price data (date from year 1968 to 2011) was decomposed into several IMFs, and the most important one was the trend. It is obviously most important because we know that gold prices were cheap in the earlier years, but in recent years, the price of gold has sky-rocketed, the reason being inflation. On the other hand, gold prices have always had a strong correlation with historical and international events. These significant events which had disturbed world gold prices, include the crude oil crisis in 1974, The Gulf War in 1980, the New York Stock market crash in 1987, the economic growth in the U.S. from 1996 to 2006, the financial crisis in 2007 and recently, the European and U.S. debt crisis of 2011; all have been important factors. In the study by Yen-Rue Chang (2011), these significant events behave in the low-frequency term (Figure 3.2).

In this study, we used the historical gold price data to forecast gold prices in 2011, and also improved the results by applying the ensemble average. These performance results are also presented in the section 4.



1973/10:
4th Middle
East War

1979/11～1980/01:
Iranian hostage
crisis

1980/09:
Iran/Iraq War

1982/08:
Mexico External
Debt Crisis

1987/10:
New York Stock
Market Crash

2003/03:
USA/Iraq
War

2004/04:
Rothschild exits
gold price system

2007/02:
USA Subprime
Mortgage Crisis

2008/09:
Lehman Brothers bankruptcy

Figure 3.2 Monthly gold price and significant world events from Jan. 1968 to Nov.2010 [Yen-Rue Chang, 2011, NCCU]

al Chengchi Uni

3.2. Experiment design

In this study, the forecasting strategies are presented as the following steps:

Step1: extract data

Based on the benchmark study (Section 4.1), we extracted data series with an interval of one week (168 points from the electricity load hourly data), and 1005 points from daily gold price data as moving window length for training. In this section we took the hourly electricity load data as a testing subject.

Step2: interpolation

The reason we used interpolation is that we found ANN could perform very well when the difference of value between one point and the next point was low while using moving window.

In this step, we applied cubic spline interpolation to the extracted data. The extracted 168 points increased to 3341 points (i.e. with interpolation of 20 points) while time scale transformed from one hour a point to 3 minutes per point. Figure 3.3 illustrates this procedure.

國立政治大學 National Chengchi University

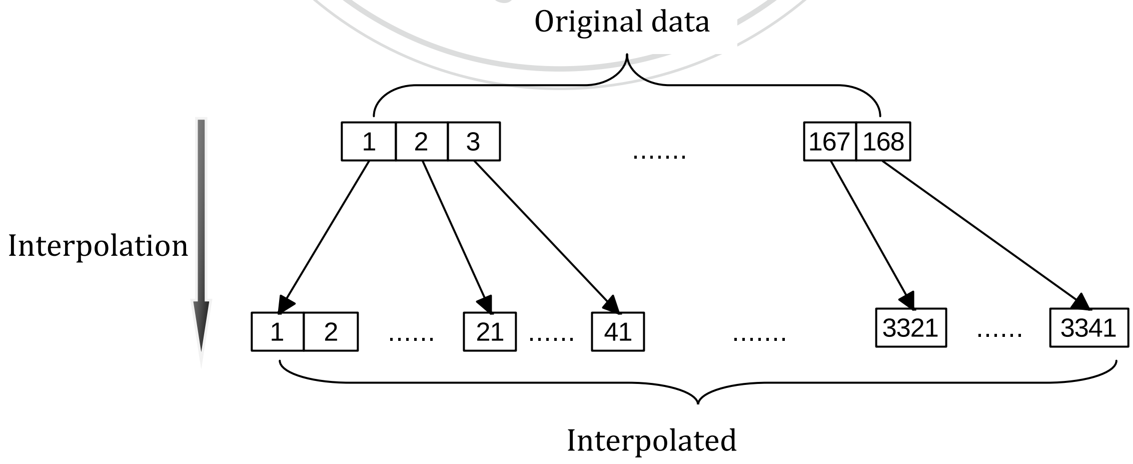


Figure 3.3 The interpolation method

Since the original data was based on hourly intervals, the interpolated points in step 2 are non-existent and unrealistic for forecasting purposes. If we wanted to forecast the value of the next hour accurately, we needed to forecast the fictitious points. We found that extrapolation could provide accurate forecasting results, but only to a certain extent. Therefore, our strategy was based on performing the extrapolation to an extent, and then forecasting residual virtual points. When a new virtual point was forecasted, we took this point and used it as real forecasted data to do the moving window forecasting procedure.

After step 2, cubic spline extrapolation was applied to the new interpolated data. The number of extrapolation was decided empirically (e.g. in this study we found that interpolation with twenty points combine extrapolation with twelve points performs very well and has good efficiency for program's working).

Step4: create a new data series for decomposition

In step 4 we combined the interpolated data produced in step 2, with the extrapolated points discarding the same amount as the extrapolated points in front of the interpolated data (i.e. maintain the length of data with one week, see Fig.3.4).

國立政治大學
National Chengchi University

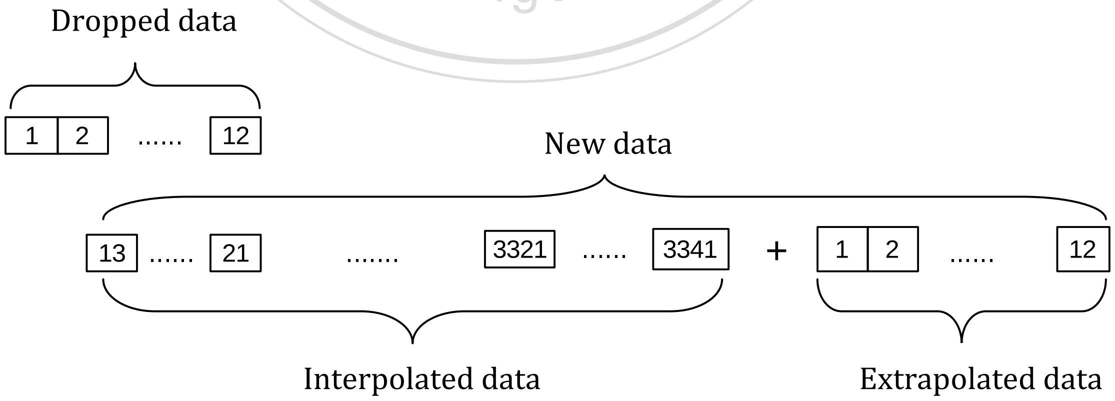


Figure 3.4 Create a new data series by interpolation and extrapolation

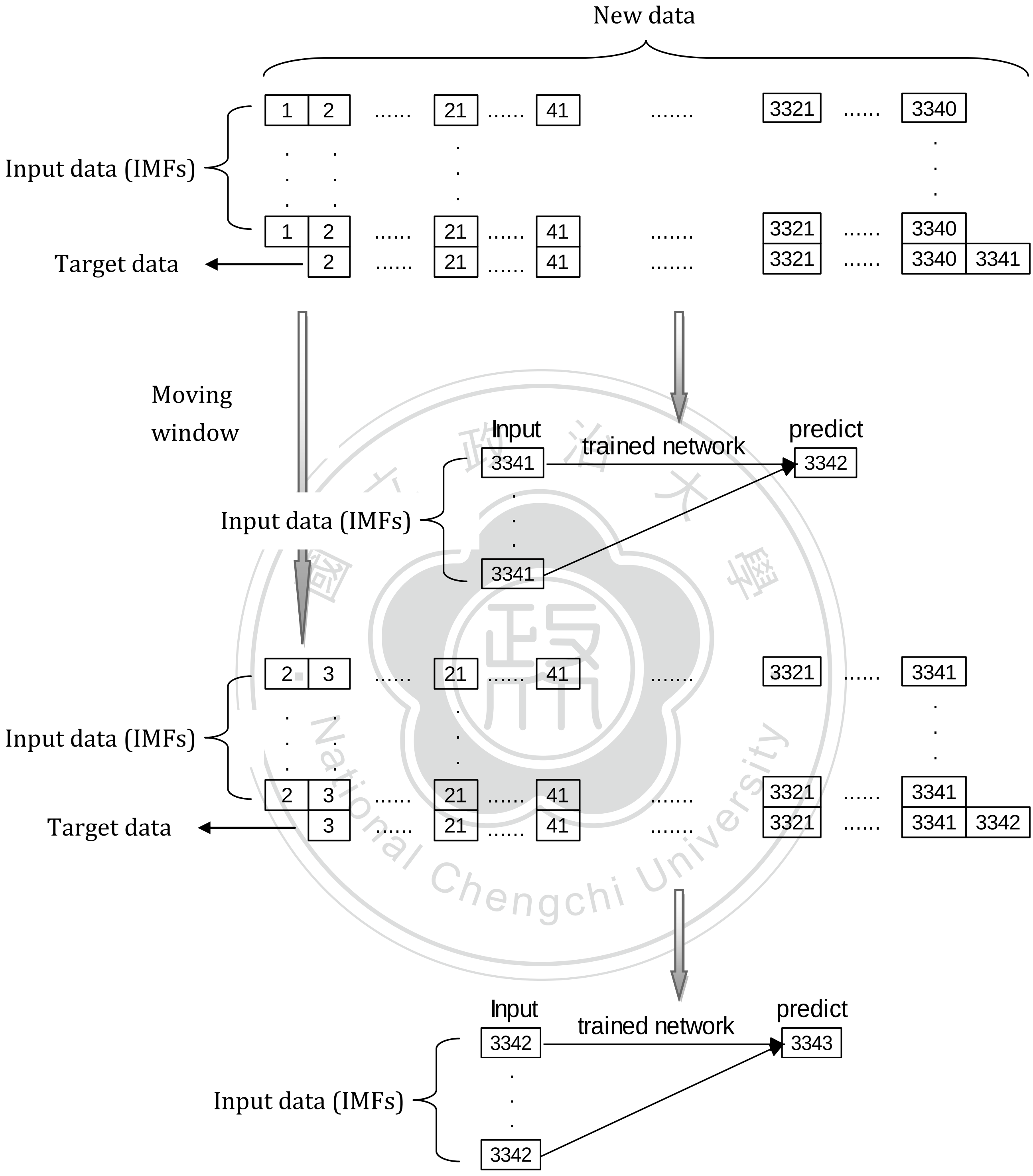
Step5: EEMD and BPNN training

The new data was created to decompose into IMFs through EEMD. After EEMD, the one week data was decomposed into meaningful IMFs which could be regarded as factors that influence the consumption of electricity. In this step, the IMFs which discarded last points were regarded as input data. The new data discarded the first point which was regarded as target data. For BPNN training, we put input and target data into a BPNN network, and the discarded the last IMFs' points which were used as input for the trained network to generate forecasting point (Fig.3.5).

Step6: Moving window and forecasting

We combined the forecast point to the new data, discarded another point in front of the new data, and update it again. After the newer data was updated, we repeated step 5 (i.e. the method is called moving window, see Fig.3.5) until the original hourly point was forecasted. Once we got the forecasted hourly data, we then compared it with the real hourly data. In order to forecast the data of next hour, we took the real data of next hour, added it into our data, and repeated step 3 to step 6.

國立政治大學 National Chengchi University



National Chengchi University

Figure 3.5 Training Network by Moving Window Process

3.3. Statistical measures

In this section we shall introduce various statistical measures that are used to analyze the IMF properties, and forecasting result performance. The measures used to analyze IMFs are mean periods, Pearson correlation coefficient, and power percentage. The measures used to see performance of forecasting results are RMSE, MAE and the standard deviation of error. These measures are presented as follows:

Mean period

The mean period which is used to see the cycle of an IMF is calculated by the inverse of mean frequency. The mean frequency is the average of "instantaneous frequency". To calculate the instantaneous frequency, we apply HHT to the extracted IMFs. In the following paragraph we will briefly introduce the HHT, proposed by Huang et al. (1998).

For any arbitrary time-series data set X(t), we can always have its Hilbert-transform Y(t) as

$$\mathrm{Y(t)} = \frac{1}{\pi}\mathrm{p.v.}\int_{-\infty}^{\infty}\frac{\mathrm{X(t^{'})}}{\mathrm{t-t'}}\mathrm{dt'}$$

Where p.v. indicates the Cauchy principal value.

Then we use X(t) and Y(t) to form the complex conjugate pair, and get an analytic signal Z(t) as

$$\mathrm{Z(t) = X(t) + iY(t) = \ a(t)e^{i\theta(t)}}$$

Where $\mathrm{a(t)} = \sqrt{\mathrm{X^2(t) + Y^2(t)}}$

and $\theta\mathrm{(t) = arctan}\ (\frac{\mathrm{Y(t)}}{\mathrm{X(t)}})$

國立政治大學 National Chengchi University

Which a(t) indicates the amplitude varied with time, and θ(t) represents the phase also a function of time. The above is the best local fit of amplitude and phase-varying trigonometric function to the original time series data X(t). Now we can define the instantaneous frequency of Hilbert transform through the phase:

$$\mathrm{f(t)} = \frac{\omega(\mathrm{t})}{2\pi} = \frac{1}{2\pi}\frac{\mathrm{d}\theta(\mathrm{t})}{\mathrm{dt}}$$

Therefore, the mean frequency F of an IMF can be presented as:

$$\mathrm{F} = \frac{1}{\mathrm{N}}\sum_{\mathrm{t}=1}^{\mathrm{N}} \mathrm{f(t)}$$

and the mean period T will be

$$\mathrm{T} = \frac{1}{\mathrm{F}}$$

Notice: We didn’t calculate the mean period of residue because it is a monotonic function, so in this study we ignored the mean period of residue.

Pearson correlation coefficient

The Pearson correlation coefficient, which’s value always ranges from -1 to +1, is the most familiar measure used to detect the dependence between two quantities. The Pearson correlation +1 means a perfect positive linear relationship while -1 means a perfect negative linear relationship. As the value approaches zero, it means there is correlation and becoming almost uncorrelated. As it approaches -1 or +1, the correlation is stronger between two variables. Here the Pearson correlation coefficient ρ_{XY} between two variables X and Y with expected values μ_X and μ_Y and standard deviations σ_X andσ_Y is shown as follows:

國立政治大學 National Chengchi University

$$\rho_{XY} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

Where E is the expected value operator and cov(X,Y) means covariance between X and Y.

Power percentage

Power percentage is a measure based on variance for detecting the weight of an IMF on the original data. A higher value of power percentage indicates a stronger weight an IMF is. The power percentage is defined as follows:

$$\text{Power percentage}(\%) = \frac{\mathrm{Var}(\mathrm{IMF})}{\mathrm{Var}(\text{original data})} \times 100$$

Root mean square error and mean absolute error

Root mean square error (RMSE) and mean absolute error (MAE) are both useful quantities to measure how closely predicted or forecasted values are to the actual data, and are good measure of accuracy. The RMSE and MAE are given by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(R(t) - P(t))^2}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{t=1}^{N}|R(t) - P(t)|$$

Where R(t) indicates the real data at time t and P(t) means predicted data value at time t.

國立政治大學 National Chengchi University

Standard deviation of error

The standard deviation of error in this study is the standard deviation calculated from the error between the predicted and real values of every hour. Using this measure, we can find the fluctuation of errors. A lower standard deviation indicates that the errors tend to be very close to the mean, whereas higher standard deviation indicates that the errors are spread out over a large range of values.

國立政治大學 · National Chengchi University ·

4. Results and discussion

4.1. Benchmark study

Before forecasting, we first did a benchmark study in order to choose the input data length. Table 4.1 shows some measures of the forecasting results of the 2011 gold price daily data which used four different data lengths as input. From this table we can see that the input data length with 1005 points and 754 points perform little better than 1255 points and 2759 points, and saved more time during training process. In addition, the data length with 1005 points was better than the 754 points because it may have contained more information. Therefore, we chose and fixed data length with 1005 points (i.e. time interval from 2007 to 2010) as our input data in this study.

國立政治大學 · National Chengchi University ·

Table 4.1 Comparison between different input data length for forecasting of 2011 gold price

	2000-2010 (2759points)	2006-2010 (1255 points)	2007-2010 (1005 points)	2008-2010 (754 points)
RMSE	27.7	34.2	26.0	26.1
MAE	19.0	21.1	19.5	19.0
Correlation coefficient	0.978	0.966	0.981	0.980
Standard deviation of error	20.2	26.9	17.2	18.0

Similarly, for judging the suitable input length of electricity load data, we tested four different data lengths. The forecasting measure results of June 6, 2008, GCB10 are displayed in Table 4.2. The table shows that the data length with one-week performed the best depending on the measures. Although the length with two-weeks and three-weeks also had good performance, the time spent training doubled and tripled. However, the measures of length which took only one-day perform so poorly due to big error, occurring from 5:00 to 6:00. The case revealed that the time interval was too short for training, and from this short training, data may have caused the occurrence probability of big error, resulting possibly from a lack of information. Based on the above reasons, the one-week data length is the most suitable choice for a training set which was used in this research.

國立政治大學 National Chengchi University

Table 4.2 Comparison between different input data length for forecasting of

June 6, 2008 of GCB10

	One day		One week		Two weeks		Three weeks	
Hour	**error**	**error%**	**error**	**error%**	**error**	**error%**	**error**	**error%**
0~1	4.3	0.81	17.6	3.88	19.4	4.28	3.2	0.71
1~2	2.8	0.53	13.8	3.11	16.7	3.76	2.5	0.56
2~3	4.6	0.87	3.9	0.89	7.5	1.71	16.4	3.74
3~4	2.6	0.48	5.4	1.24	15.7	3.63	8.1	1.88
4~5	9.8	1.86	3.3	0.76	9.8	2.24	16.9	3.86
5~6	334.1	63.63	13.4	3.12	17.1	3.97	8.2	1.89
6~7	10.0	1.86	1.8	0.40	5.8	1.29	12.6	2.82
7~8	15.2	2.48	31.0	6.34	35.7	7.32	31.9	6.54
8~9	18.7	2.52	12.5	1.85	2.4	0.36	19.9	2.94
9~10	29.4	3.73	24.6	3.06	41.2	5.13	32.3	4.02
10~11	0.7	0.08	5.3	0.62	5.8	0.68	2.8	0.33
11~12	52.4	6.51	6.3	0.72	7.9	0.90	8.3	0.94
12~13	20.5	2.49	13.7	1.59	6.0	0.69	4.2	0.48
13~14	21.8	2.61	8.6	1.00	9.4	1.09	3.3	0.39
14~15	40.7	5.01	16.4	1.83	10.8	1.21	1.9	0.21
15~16	7.6	0.95	16.4	1.80	17.0	1.86	13.6	1.49
16~17	6.5	0.82	12.3	1.40	27.9	3.17	2.9	0.33
17~18	39.0	5.12	7.6	0.92	4.3	0.52	17.6	2.13
18~19	5.5	0.78	21.0	2.80	3.6	0.48	14.0	1.87
19~20	3.9	0.57	3.3	0.48	15.9	2.27	3.6	0.52
20~21	8.3	1.22	3.6	0.54	0.1	0.02	10.9	1.64
21~22	4.4	0.69	18.5	2.86	23.9	3.69	26.6	4.10
22~23	1.2	0.20	13.3	2.43	15.2	2.78	30.8	5.63
23~24	3.3	0.58	23.8	4.89	0.0	0.00	0.9	0.19
RMSE	70.8		14.6		16.9		15.7	
MAE	27.0		12.4		13.3		12.2	
Standard deviation of error	66.9		7.8		10.6		10.1	

國立政治大學 National Chengchi University

4.2. The meaning of IMFs

Electricity load data from NCCU

The most important work in building an ANN forecasting model is the selection of input variables. In the past some researchers had collected different types of data as variables to input into ANNs for forecasting future data. For example, industrial production index, exchange rate, commodity price and interest rate are the variables used for financial stock market time series forecasting. Temperature, wind velocity, hourly and daily historical data are used for electricity load forecasting. However, as mentioned before, the EEMD method proposed by Huang et al. can decompose time-series data into several IMFs which have its own physical meaning. Consequently, we can regard IMFs as factors which influence the data for constructing an ANN model. The EEMD-Based neural network learning paradigm has been used by Yu et al. (2008) to forecast world crude oil spot price data. For the above reasons, in this study we applied the EEMD-Based neural network learning paradigm to forecast electricity load data.

Here we took an input data series for testing. The input data which length was one week from May 30 to June 5, 2008 of General Building of NCCU was used to train the network for forecasting data of June 6, 2008, Friday. The decomposed IMFs and residue are shown in Figure 4.1. The measures of these IMFs are presented in Table 4.3. In the figure and table, we can see that the IMF7, IMF8 and the residue are the most important. The IMF7 period is close to one day (24 hour means the variability of one day). As we already know, the electricity load always has a regular pattern of high consumption during the day and low consumption at night. The phenomenon is also known as peak hour and off-peak hour. We can find that the consumption always noticeably increases at 8:00, and decreases after 22:00.We can also conjecture that

國立政治大學 National Chengchi University

these fluctuations are determined by the activities and usage of the General Building by faculty and students.

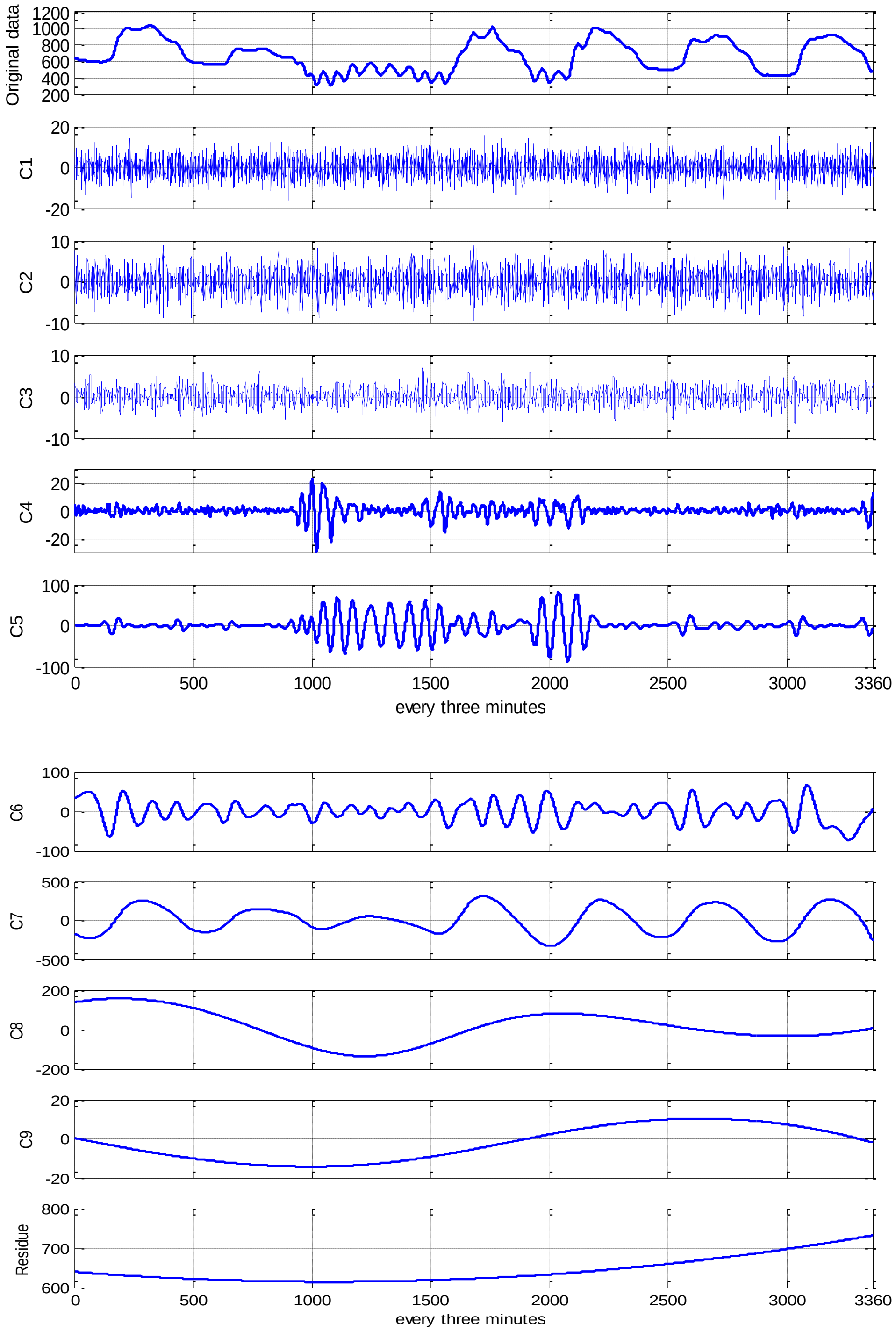


Figure 4.1 The IMFs for General Building of NCCU from May 30 to June 5, 2008

Table 4.3 Measures of IMFs for General Building of NCCU from May 30 to June 5, 2008

	Mean	**Mean**	**Power**	**Pearson**
	period(Hour)	**period(Day)**	**percentage(%)**	**correlation**
IMF1	0.158	0.007	0.054	0.017
IMF2	0.289	0.012	0.020	0.020
IMF3	0.568	0.024	0.010	0.014
IMF4	1.414	0.059	0.052	0.107
IMF5	3.117	0.130	1.431	0.152
IMF6	6.431	0.268	1.824	0.050
IMF7	23.513	0.980	74.750	0.857
IMF8	86.408	3.600	17.913	0.445
IMF9	160.699	6.696	0.204	0.230
Residue			2.829	0.191

國立政治大學
National Chengchi University

The IMF8, which is the second most important IMF, presents behavior of a half-week pattern. The half-week pattern presents the highest consumption trend of the original data. The IMF9, which period is close to 7 days, presents the one-week pattern and the property of difference between week-days and weekends. The residue, also known as trend of data, presents the mean consumption trend of the week. Compared IMF8 and residue, we can find that the highest consumption of Monday and Tuesday is higher than Wednesday and Thursday but the mean consumption of Wednesday and Thursday is more than Monday and Tuesday. The reason is that the arrangement of courses are centralized in certain sessions in Monday and Tuesday while the courses in Wednesday and Thursday are more but allocated to various sessions. Upon the IMFs' properties, the forecasting results are revealed in section 4.3.

Gold price daily data

Unlike the electricity load data, the daily gold price data didn’t reveal any regularities. Gold price data is very similar to stock market series, meaning when the moving window shifts, the property of IMFs will change. But if we could take a long enough length of gold price data, we found that the trend was always the important one due to the obvious increase of gold price in past years. Moreover, a major reason for increase was inflation as well as the global financial crisis of the recent years. We took the daily price data from Jan, 2007 to Dec, 2010 as a testing subject. The data length (1005 points) was also the length of the moving window we used. The data was decomposed into seven IMFs as well as one residue by EEMD (Figure4.2). Furthermore, from Table 4.4, we can see that the most dominant component was the residue, and the second most dominant, the IMF7. The Pearson correlation coefficient of residue was 0.942 with the power percentage reaching up to 98%. Through this case, we validated that the trend was the most important influencing factor for gold prices. With the fixed window length of 1005 points, the daily gold price was predicted. The performance is displayed in section 4.4.

國立政治大學 National Chengchi University

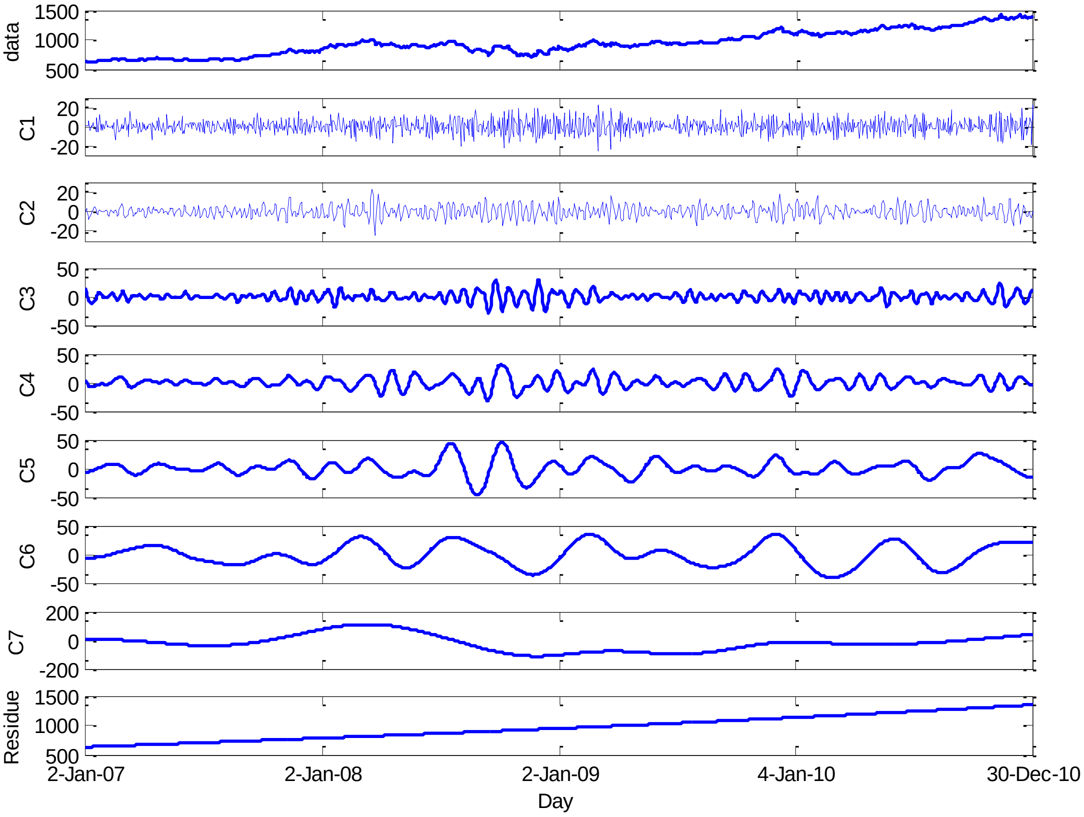


Figure 4.2 The IMFs of historical gold price daily data from 2007 to 2010

National Chengchi University

Table 4.4 Measures of IMFs for gold price daily data from 2007 to 2010

	Mean period(Day)	Power percentage(%)	Pearson correlation
IMF1	3.090	0.119	0.049
IMF2	6.592	0.080	0.032
IMF3	14.199	0.137	0.063
IMF4	26.959	0.233	0.119
IMF5	55.593	0.444	0.174
IMF6	121.516	0.857	0.158
IMF7	404.139	7.630	0.111
Residue		98.105	0.942

4.3. Forecasting performance

This section shows our forecasting results. Below is a sample of forecasting performance for June 6, 2008 of GCB10 is shown in Table 4.5:

Table 4.5 Forecasted hourly load compared with actual load for June 6, 2008 of GCB10

Hour	Actual load (KWH)	Forecasted load(KWH)	Error	Error%
0~1	453	470.6	17.6	3.88
1~2	444	457.8	13.8	3.11
2~3	439	442.9	3.9	0.89
3~4	432	426.6	5.4	1.24
4~5	437	433.7	3.3	0.76
5~6	431	417.6	13.4	3.12
6~7	446	444.2	1.8	0.40
7~8	488	457.0	31.0	6.34
8~9	676	688.5	12.5	1.85
9~10	803	827.6	24.6	3.06
10~11	861	866.3	5.3	0.62
11~12	876	882.3	6.3	0.72
12~13	862	848.3	13.7	1.59
13~14	859	850.4	8.6	1.00
14~15	894	877.6	16.4	1.83
15~16	911	927.4	16.4	1.80
16~17	879	891.3	12.3	1.40
17~18	825	832.6	7.6	0.92
18~19	751	772.0	21.0	2.80
19~20	699	695.7	3.3	0.48
20~21	667	663.4	3.6	0.54
21~22	648	666.5	18.5	2.86
22~23	546	559.3	13.3	2.43
23~24	487	463.2	23.8	4.89
RMSE			14.6	
MAE			12.4	
Standard deviation of error			7.8	

國立政治大學 National Chengchi University

In the above table, the error is the difference between forecasted value and actual value The error(%) is defined as below:

$$\mathrm{error}(\%) = \frac{|\text{Forecasted load} - \text{Actual load}|}{\text{Actual load}} \times 100$$

As we can see, the biggest error is 31 KWH. The error(%) is about 6.3% from 7:00 to 8:00, whereas the smallest error is only 2 KWH, and the error(%) is about 0.4% from 6:00 to 7:00. The RMSE, MAE and standard deviation of error are also presented in Table 4.5. The results illustrated in Figure 4.3 and Figure 4.4, and Figure 4.5 show the correlation between forecasted hourly data and actual hourly data.

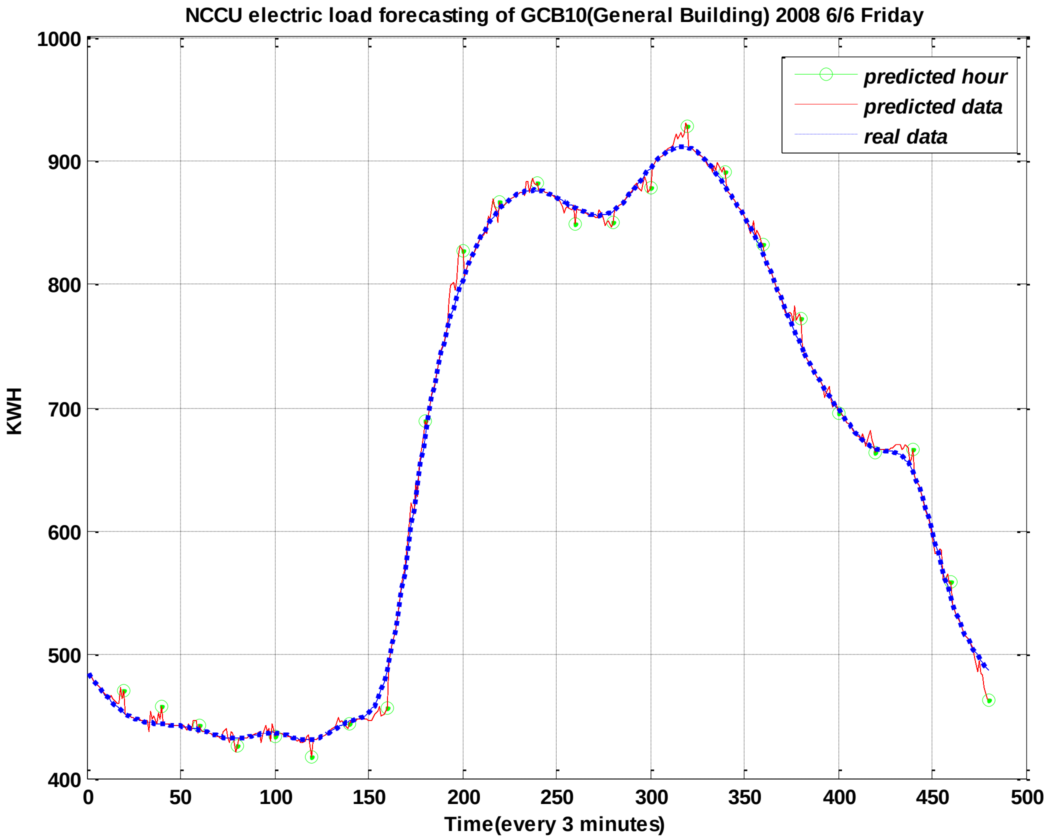


Figure 4.3 Forecasting performance for June 6, 2008 of GCB10

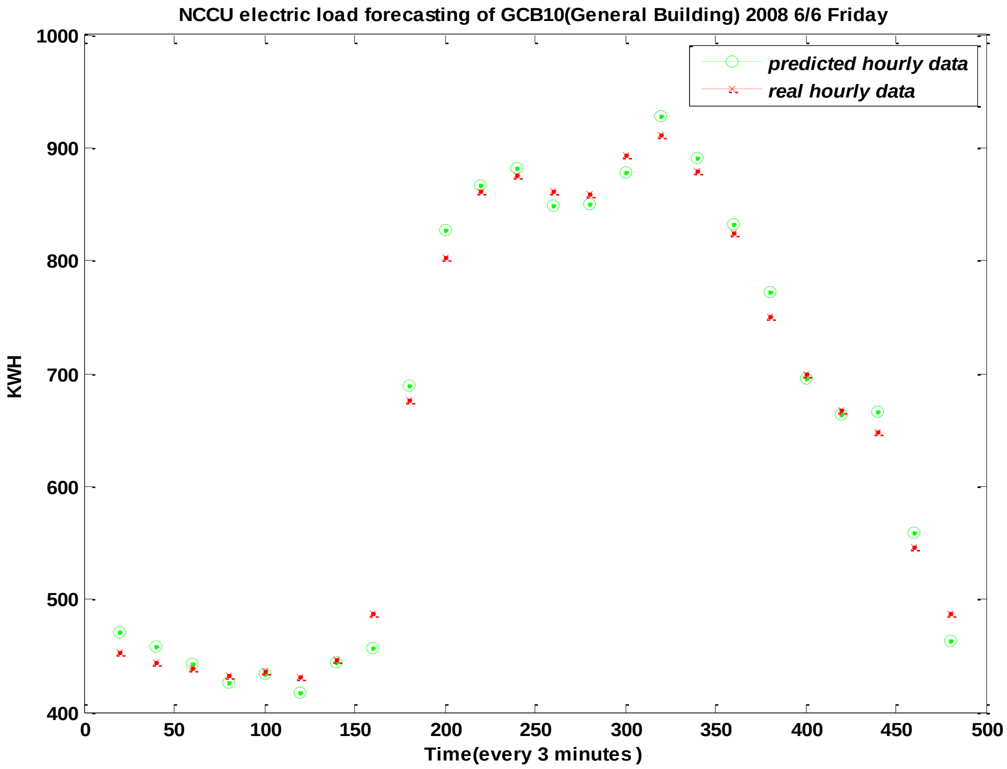


Figure 4.4 Forecasting performance of every hour for June 6, 2008 of GCB10

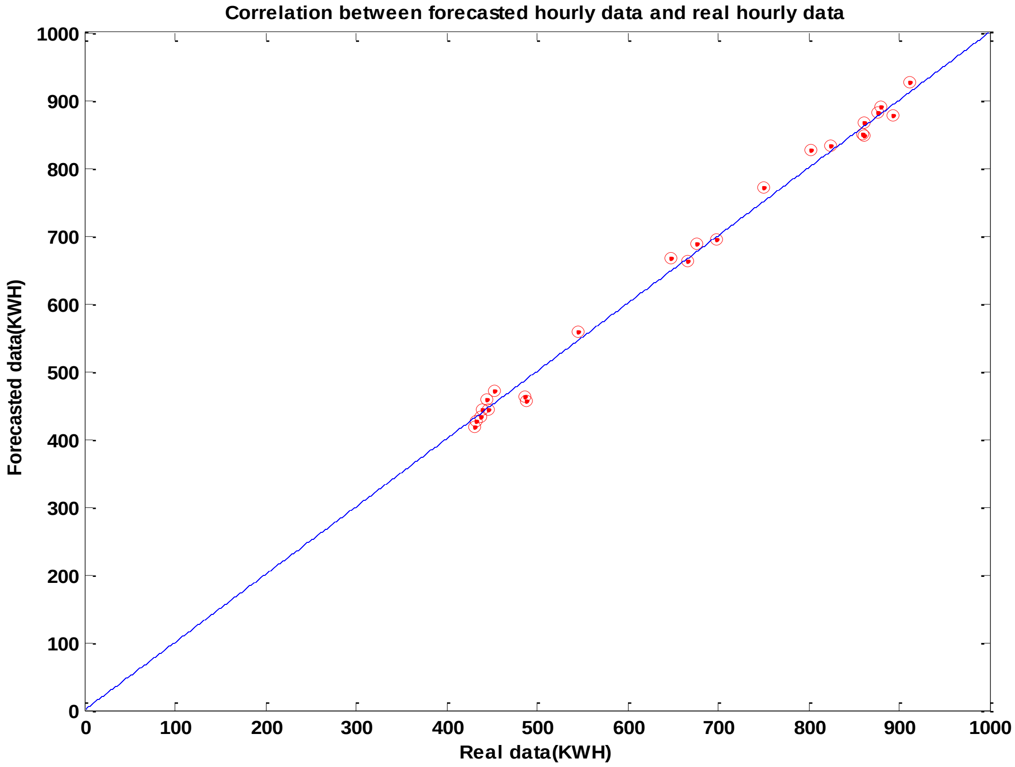


Figure 4.5 The performance of correlation between forecasted and actual hourly data

4.4. *Performance of ensemble average*

4.4.1. Electricity load data from NCCU

Since the BPNN aims to find the minimum mean square error function, the forecasting results try to approach the optimal result (i.e. global minimum). However, the program doesn't find minimum mean square error function every time; the results always deviate from the best or find the local minimum. Figure 4.6 shows the sketch of the error function surface. We can see that if the program finds a local minimum on the right of the global minimum, the error may be positive. In contrast, the local minimum on the left side may be a negative error. Therefore, we can apply the ensemble average method on to the prediction result to cancel out the error. We can see that the performance by using ensemble average was better in this section.

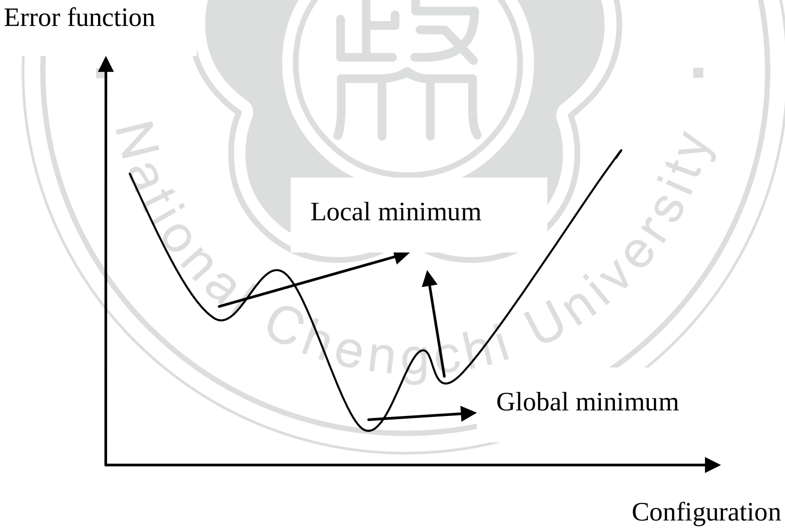


Figure 4.6 A diagrammatic sketch of the BPNN error function surface

Here we used the load data of May 22, 2008 of GCB10 for an analysis testing. We chose this data because its variation was not so regular compared with June 6, 2008 of GCB10, which had been analyzed in the proceeding text. In the Figure 4.7, we can see that the electricity load suddenly jumps over 1000KWH at the point 300 and 320 (i.e.

14:00~16:00). It decreased gradually until about 820KWH at the point 380 (i.e. 18:00~19:00). Nevertheless, it suddenly jumped to 900KWH again the next hour. Although we could find the precise future trend, we couldn't get an accurate value in the suddenly changing points. On the other hand, the smooth, stable changed points could be predicted accurately. Based on these results, we took the data to see whether or not the forecasting performance would be better after the ensemble average.

The forecasting results of the different ensemble levels are illustrated in Figure 4.7, Figure 4.8 and Figure 4.9. Figure 4.7 present forecasting results without the ensemble average. We can see that in the abruptly changed points such as 320 and 400, the large error occurs, however, in Figures 4.8 and 4.9, which apply three and five times the ensemble average respectively, we can clearly see that the forecasting results become better. In addition, Table 4.6 shows the comparison of quantity between these three different ensemble average levels. The results without ensemble average have four points at which error percentage was over 5%. The RMSE, MAE and standard deviation of error are 34.8, 21.0 and 28.4 respectively. When we apply the three times ensemble average to improve the results, the points which error percentages over 5% were three, and the RMSE, MAE and standard deviation of error become 22.8, 17.2 and 15.4 respectively, all of which are better than the results without ensemble average. Even more, from the results using five times ensemble average, the points which error percentage was over 5% were only one. The RMSE, MAE and standard deviation of error were 18.5, 13.8 and 12.5 respectively, obviously better than the other two.

In the comparison, we concluded that the forecasting error could be improved by applying the ensemble average method. However, a crucial drawback of this method is the large amount of time spent completing the process. For this reason, we aim to

國立政治大學 National Chengchi University

reduce the program processing time in the future. The outlook will be discussed in the conclusion.

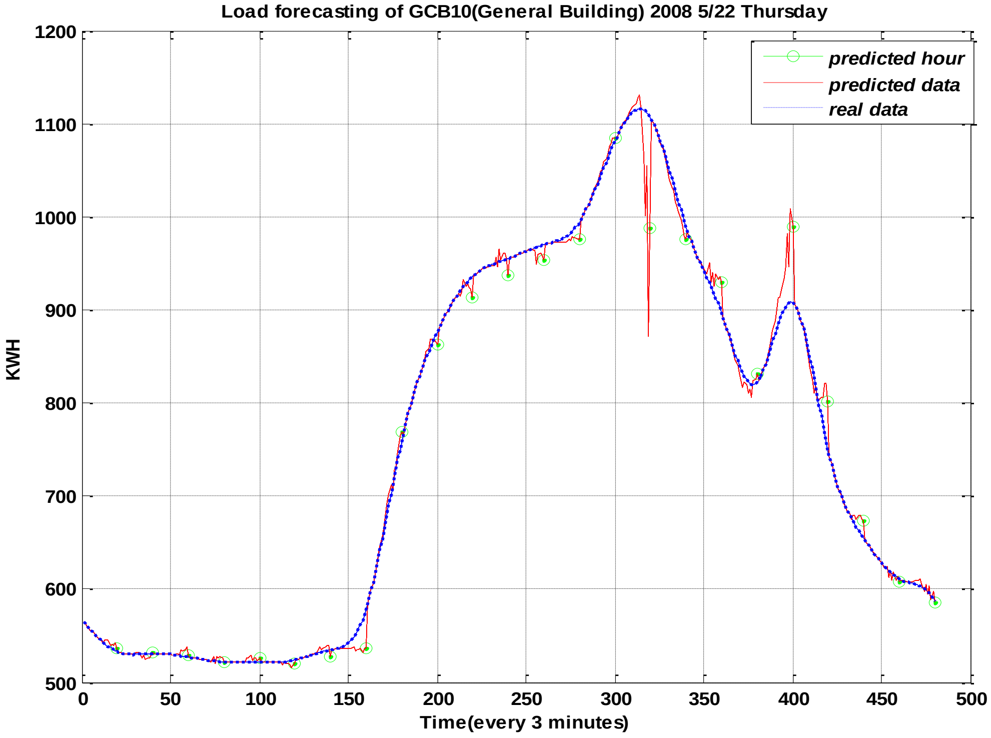


Figure 4.7 Forecasting performance for May 22, 2008 of GCB10, without ensemble average

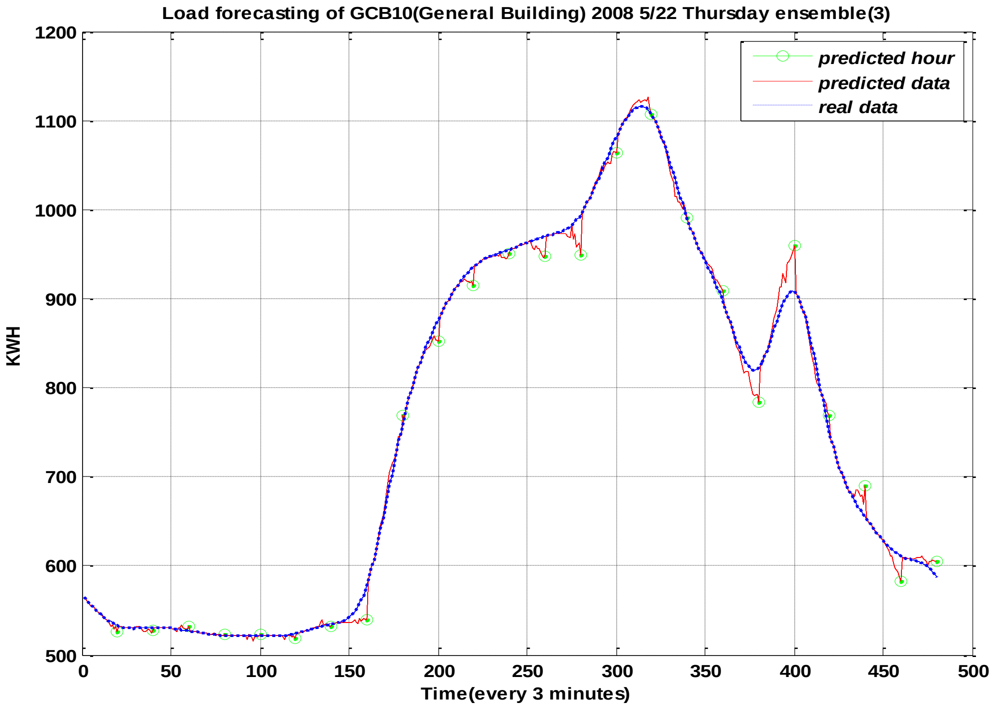


Figure 4.8 Forecasting performance for May 22, 2008 of GCB10,

with three times ensemble average

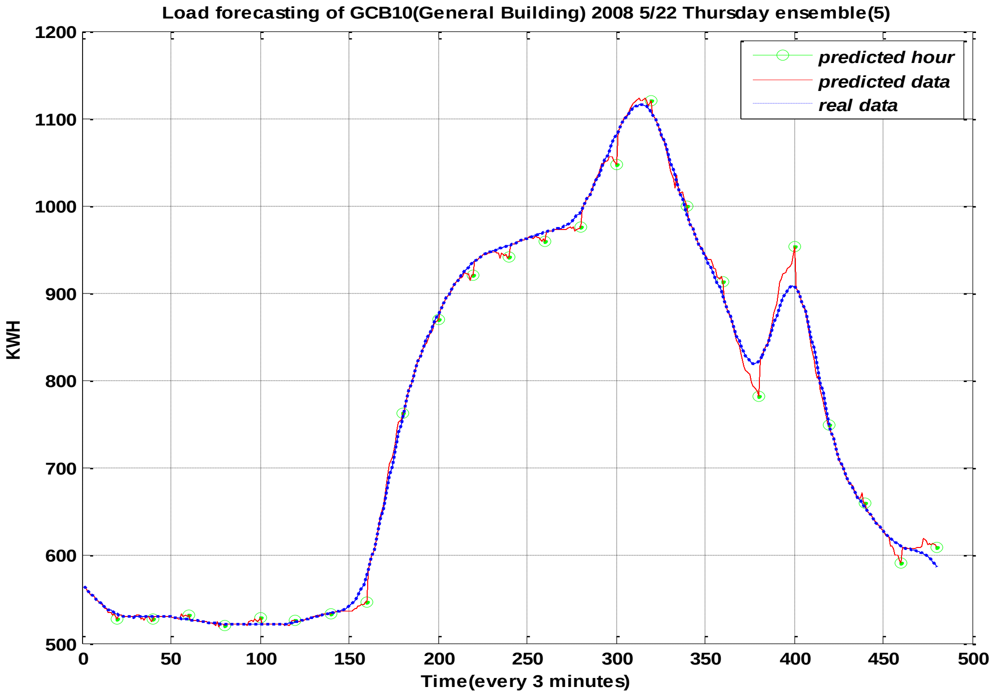


Figure 4.9 Forecasting performance for May 22, 2008 of GCB10,

with five times ensemble average

Table 4.6 The performance comparison between different ensemble average levels

	no ensemble		**ensemble(3)**		**ensemble(5)**	
Hour	**error**	**error%**	**error**	**error%**	**error**	**error%**
0~1	3.2	0.60	7.6	1.43	4.8	0.89
1~2	0.6	0.12	4.1	0.78	3.4	0.64
2~3	2.7	0.52	5.7	1.08	5.0	0.96
3~4	0.6	0.11	0.5	0.09	1.8	0.34
4~5	5.4	1.04	1.9	0.36	7.7	1.48
5~6	4.4	0.84	5.6	1.07	1.6	0.31
6~7	7.4	1.38	2.3	0.42	1.8	0.33
7~8	41.6	7.20	38.0	6.57	30.9	5.35
8~9	11.4	1.51	10.2	1.34	4.5	0.60
9~10	12.7	1.45	24.5	2.80	5.7	0.65
10~11	22.2	2.37	21.1	2.26	14.9	1.59
11~12	18.6	1.95	4.1	0.43	13.8	1.45
12~13	15.8	1.62	21.9	2.25	10.7	1.11
13~14	18.5	1.86	45.2	4.55	18.7	1.88
14~15	3.7	0.34	17.8	1.65	34.0	3.15
15~16	119.8	10.82	0.1	0.01	12.7	1.15
16~17	14.9	1.50	0.4	0.04	9.2	0.93
17~18	33.0	3.68	11.6	1.30	17.0	1.89
18~19	9.9	1.20	38.4	4.68	39.1	4.75
19~20	81.3	8.96	51.3	5.65	45.1	4.96
20~21	52.2	6.96	18.2	2.42	1.2	0.16
21~22	19.0	2.89	34.4	5.25	4.6	0.70
22~23	2.6	0.42	29.0	4.74	19.9	3.26
23~24	1.4	0.24	17.7	3.02	22.9	3.89
RMSE	34.8		22.8		18.5	
MAE	21.0		17.2		13.8	
Standard deviation of error	28.4		15.4		12.6	

國立政治大學
National Chengchi University

4.4.2. Gold price daily data

In this section the forecasting results of gold price daily data without the ensemble average and five times ensemble average are both discussed. The two cases are illustrated in Figure 4.9 and Figure 4.10, with four mark influential events on the graph. For comparing, the RMSE of result without the ensemble average is 25.95, the MAE is 19.46, the standard deviation of error is 17.2 and the correlation coefficient between predicted data and actual data is 0.9813. The RMSE of result with five times ensemble average is 19.18 and the MAE is 14.28, both better than the performance of cases without ensemble average. The standard deviation of error is 12.84, meaning that the error is more stable, and no large errors had occurred. Finally, the correlation coefficient was 0.9902, presenting a stronger correlation than the one without the ensemble average. The correlation between predicted price and real price of these two cases are presented in Figure 4.11 and 4.13.

The gold price has abruptly sky-rocketed in this past year, the main reason being the European debt crisis. The euro against the U.S. dollar came to the lowest level on 30 Jan 2011 meanwhile the price of gold also came to a local minimum. This data shows a strong correlation between currency and gold price. However, in April 23, 2011, Greece applied for assistance to EU and IMF, and the gold price suddenly increased. It could be interpreted as people were looking for a hedge by buying large amounts of gold. Similarly, due to S&P reducing Greece's credit rating to CC on July 4 and the U.S.'s later on August 5, as well as the world's stock market crashing as a result of worries about the U.S. second recession, the gold price experienced a dramatic rise.

Our results present fine performance for prediction. However, after several times ensemble average, the results became even better. Upon these accurate results, maybe we can design a trading strategy in the future to see whether we can make a profit.

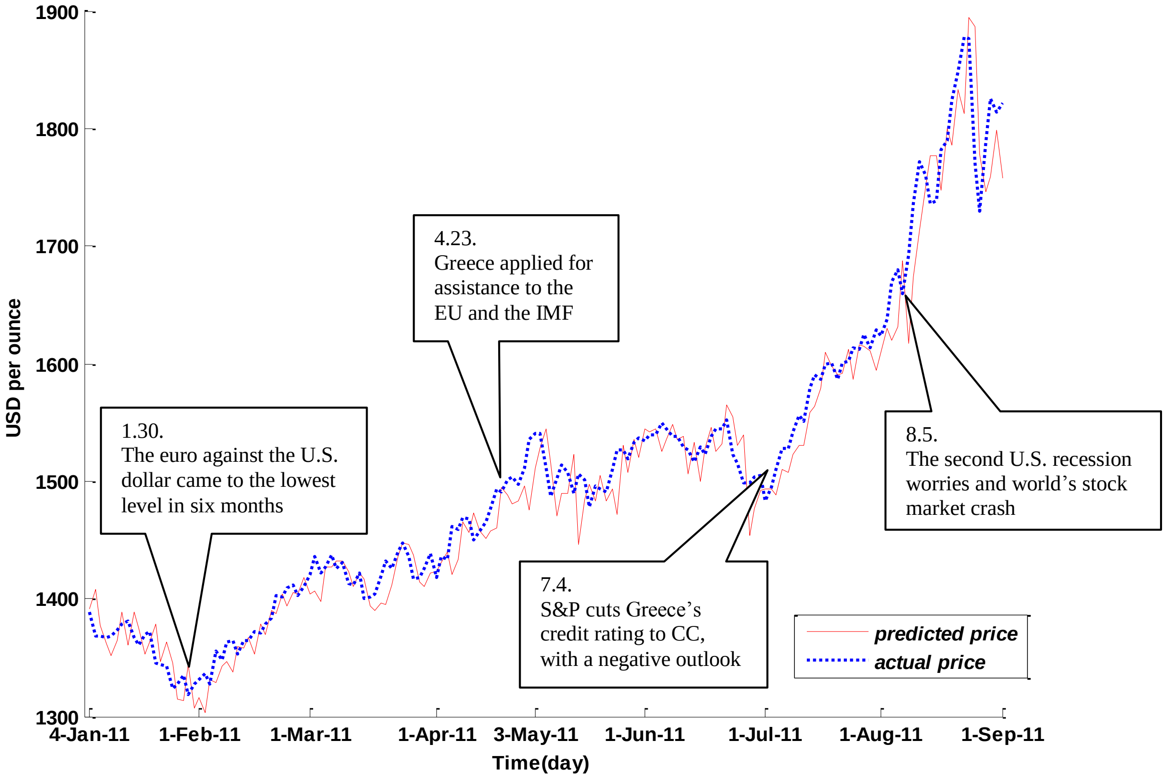


Figure 4.10 Forecasting performance for 2011 gold price daily data, without ensemble average

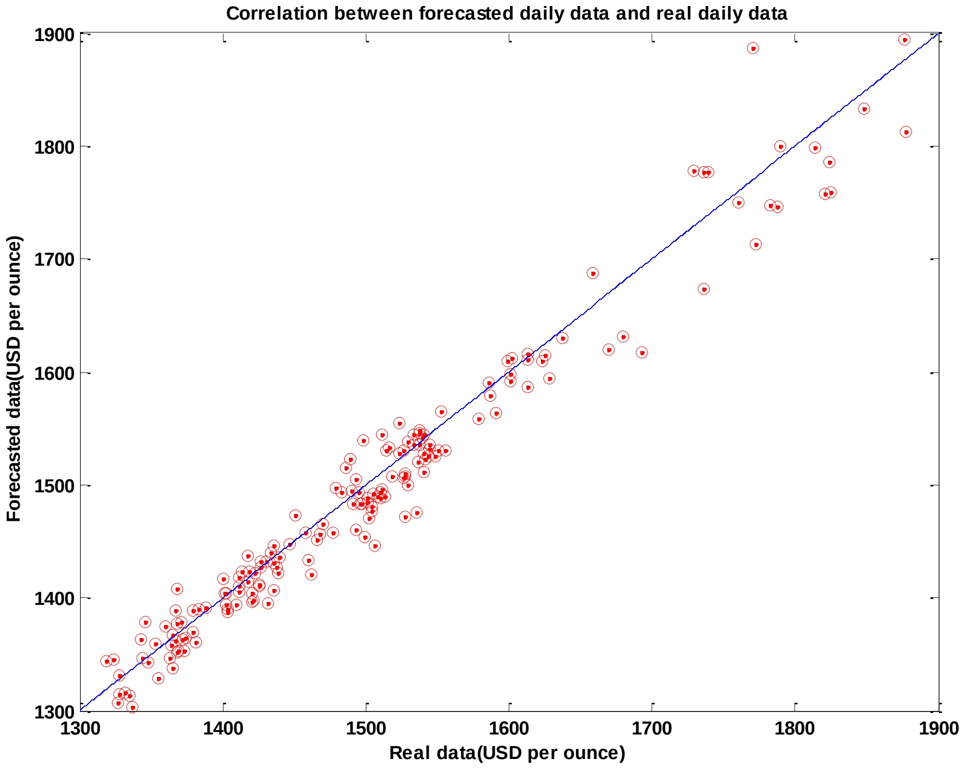


Figure 4.11 The performance of correlation between forecasted and actual daily data for gold price daily data without ensemble average

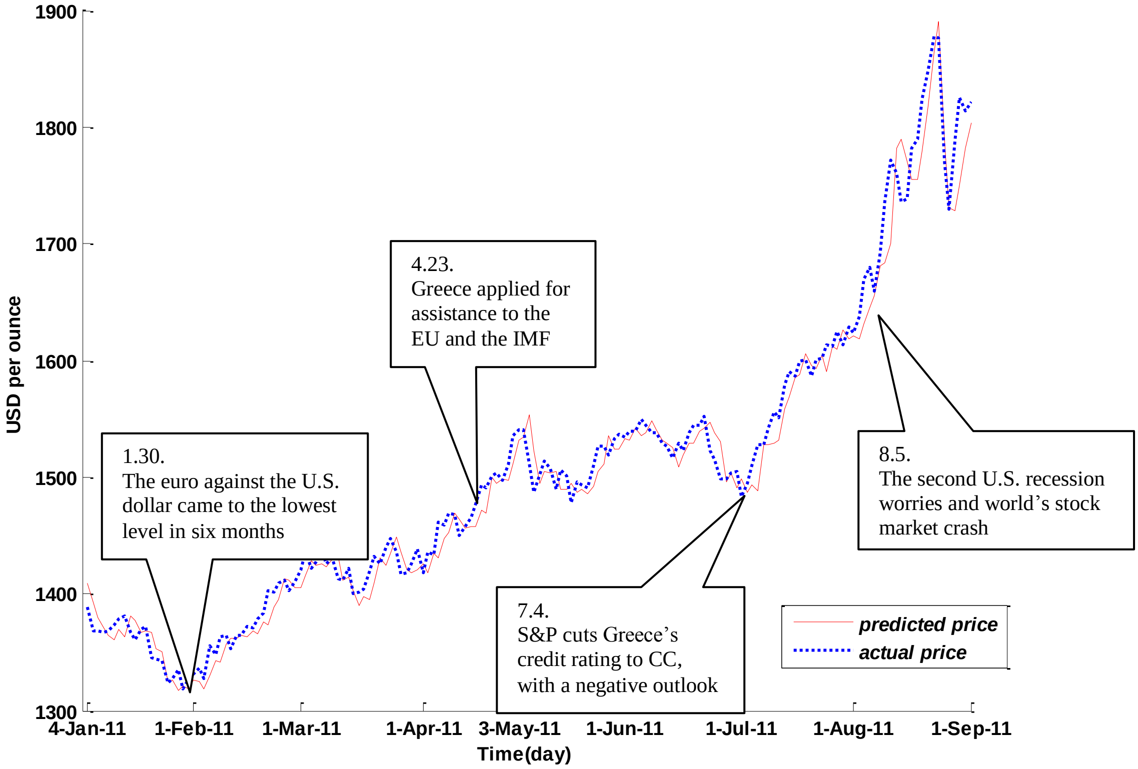


Figure 4.12 Forecasting performance for 2011 gold price daily data,

with five times ensemble average

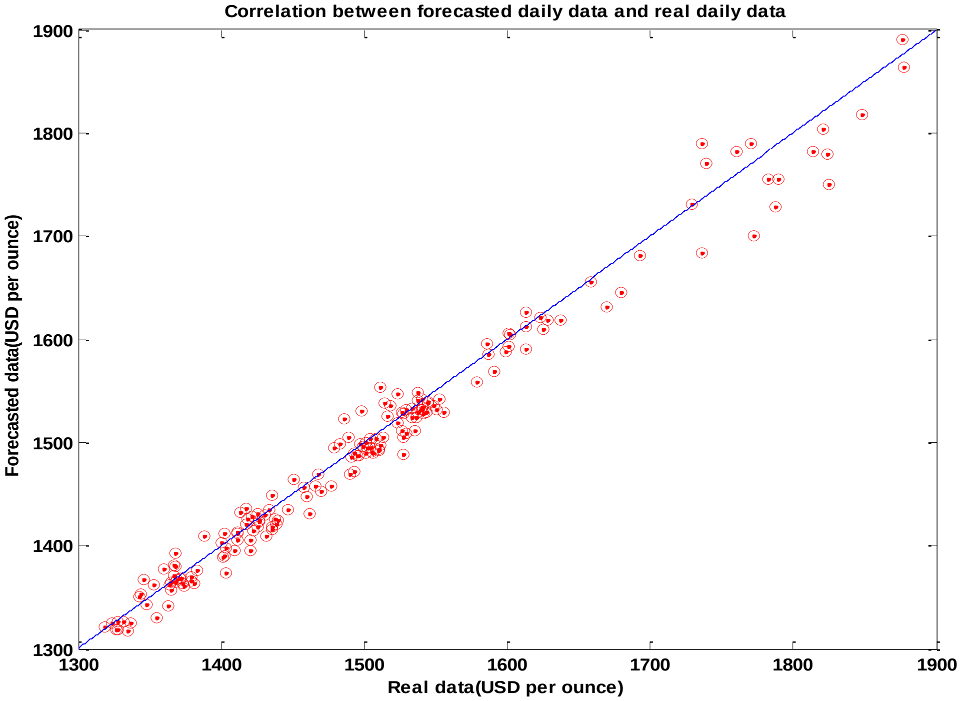


Figure 4.13 The performance of correlation between forecasted and actual daily data for

gold price daily data with five times ensemble average

5. Conclusion and outlook

An Ensemble Empirical Mode Decomposition based Back-propagation Neural Network learning paradigm has been presented for electricity load forecasting and gold price forecasting. In this research, the moving window method is applied to the prediction process. Moreover, cubic spline interpolation and extrapolation were used as our strategy for electricity load forecasting. The measures RMSE, MAE, standard deviation of error were used to judge the performance. By using the meaningful IMFs as training data, we can predict the electricity load for the next hour and the gold prices for next day, all with accurate results. However, for further improvement of the results, the ensemble average method was employed. The outcome shows that the performance was better when using the ensemble average method than if it were not in use.

For future research, there are several feasible improvements discussed as follows: First, since we used MATLAB which only can be used on normal PC to process our ANN program, in the future we can try to write an ANN program in FORTRAN which can be used on computer cluster to promote the computing speed. Second, try other combinations between the length of moving window, the numbers of interpolation and extrapolation and different parameters of network. Third, and most important, try different algorithms or architectures of neural network. For the gold price prediction, we can design some trading strategies based on our forecasting results in the future, and try to make a profit from trading gold. En Tzu Li (2011) has used this algorithm to forecast TAIEX options, and designed a moving FK indicator for algorithm trading and resulting in efficient performance. Therefore, in the future maybe we can take another financial product time-series for forecasting and with good trading strategy; we can earn lot of money.

國立政治大學
National Chengchi University

APPENDIX

In this appendix we collated the predicted results of GCB10 from May 12 to June 6, 2008 except for weekends. We selected points, error percentage are over 5%, and sketched the statistical results in Figure A.1. The 20 days' mean error and number of error-percentage points over 5% are quantized in Table A.1. We found that there are 15 points over 5% of the total 20 points from 7:00 to 8:00. Since the air-condition units are often turned on during this time, the load value jumps suddenly and the performance of prediction are usually not satisfactory. The performance of peak hour is better than off-peak hour, but in Table A.1 we can find that the mean errors between these two time periods are not too different. Due to the baseline of off-peak hour is about 400-600KWH while peak hour is about 800-1000 KWH.

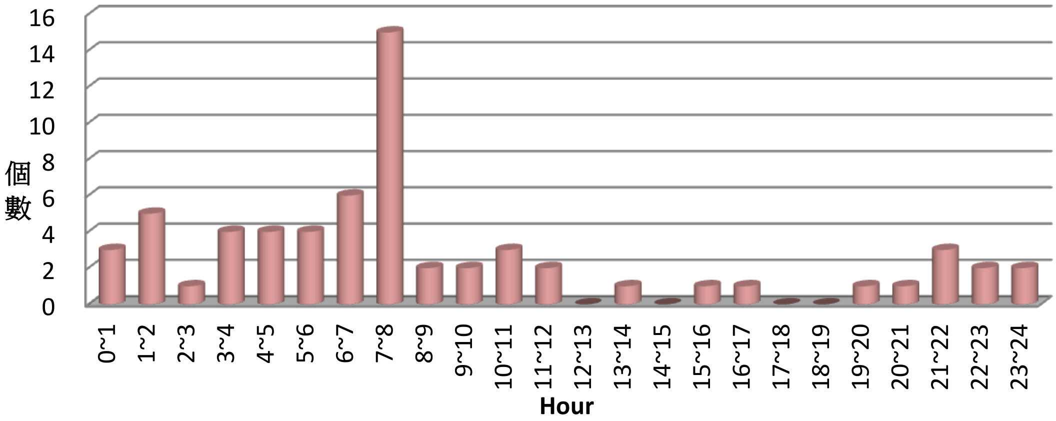


Figure A.1 Statistics of points over 5% of GCB10 from

May 12 to June 6, 2008 except for weekends

Table A.1 The mean error and collation of points over 5% of GCB10 from May 12 to June 6, 2008 except for weekends

GCB10 (2008.5.12-6.6 , weekdays)		
Hour	**Mean error**	**over 5%**
0~1	13.4	3
1~2	14.5	5
2~3	11.4	1
3~4	13.1	4
4~5	15.2	4
5~6	13.9	4
6~7	15.4	6
7~8	36.0	15
8~9	18.2	2
9~10	23.7	2
10~11	21.8	3
11~12	12.6	2
12~13	11.6	0
13~14	13.0	1
14~15	10.7	0
15~16	17.7	1
16~17	11.9	1
17~18	12.6	0
18~19	11.5	0
19~20	14.4	1
20~21	13.4	1
21~22	22.6	3
22~23	11.0	2
23~24	14.2	2
RMSE	22.5	
MAE	15.6	
Standard deviation of error	16.2	

國立政治大學
National Chengchi University

Reference

Ajay Shekhar Pandey, Devender Singh, and Sunil Kumar Sinha, 2010. Intelligent Hybrid Wavelet Models for Short-Term Load Forecasting. IEEE transactions on power systems, VOL. 25, NO. 3, August 2010.

Akgiray, V., G.G. Booth, J.J. Hatem, and C. Mustafa, 1991. Conditional Dependence in Precious Metal Prices. The Financial Review, 26, 367-386.

Chen, M.-C., Wei, Y, 2010. Exploring time variants for short-term passenger flow. J. Transp. Geogr. doi:10.1016/j.jtrangeo.2010.04.003

Cummings, D.A.T., Irizarry, R.A., Huang, N.E., Endy, T.P., Nisalak, A., Ungchusak, K., Burke, D.S., 2004. Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. Nature 427 (6972), 344–347.

En Tzu Li, 2011. TAIEX Option Trading by using EEMD-based Neural Network Learning Paradigm. Master Thesis of Graduate Institute of Applied Physics, College of Science NCCU.

FENG Ping, DING Zhi-hong, HAN Rui-guang, ZHANG Jian-wei. 2009. Precipitation-runoff forecasting ANN model based on ▪ EMD. Systems Engineering-Theory & Practice, Vol.29, No.1, Jan., 2009.

G.A. Adepoju, M.Sc., S.O.A. Ogunjuyigbe, M.Sc., and K.O. Alawode, B.Tech. Application of Neural Network to Load Forecasting in Nigerian Electrical Power System. The Pacific Journal of Science and Technology Volume 8. Number 1. May 2007 (Spring).

Hwang, P.A., Huang, N.E., Wang, D.W., 2003. A note on analyzing nonlinear and non-stationary ocean wave data. Applied Ocean Research 25 (4), 187–193.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London A 454 (1971), 903–995.

Huang, N.E., Wu, M.L., Qu, W.D., Long, S.R., Shen, S.S.P., 2003b. Applications of Hilbert–Huang transform to nonstationary financial time series analysis. Applied Stochastic Models in Business and Industry 19, 245–268.

國立政治大學 National Chengchi University

Hari Seetha and R. Saravanan, 2007. Short Term Electric Load Prediction Using Fuzzy BP. Journal of Computing and Information Technology - CIT 15, 2007, 3, 267–282.

Hong Ying Yang, Hao Ye, Guizeng Wang, Junaid Khan, Tongfu Hu, 2005. Fuzzy neural very-short-term load forecasting based on chaotic dynamics reconstruction. Chaos, Solitons & Fractals Volume 29, Issue 2, July 2006, Pages 462-469

Hamid S. A. and Iqbal Z., Using neural networks for forecasting volatility of S&P 500 Index futures prices, Journal of Business Research, 2004, 57: 1116-112

James W. Taylor. An evaluation of methods for very short-term load forecasting using minute-by-minute British data, 2008. International Journal of Forecasting, 24 (4). pp. 645-658. ISSN 0169-2070

K.Hornik, M.Stinchocombe, 1989. H.White, Multilayer feedforward networks are universal approximators,NeuralNetworks2 (1989) 359–366.

Li, Q.S., Wu, J.R., 2007. Time–frequency analysis of typhoon effects on a 79-storey tall building. Journal of Wind Engineering and Industrial Aerodynamics 95 (12), 1648–1666.

Liang, H., Lin, Q.-H., Chen, J.D.Z., 2005. Application of the empirical mode decomposition to the analysis of esophageal manometric data in gastro esophageal reflux disease. IEEE Transactions on Biomedical Engineering 52 (10), 1692–1701.

Lean Yu, Shouyang Wang, Kin Keung Lai., 2008. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. Energy Economics 30 (2008) 2623–2635.

Lean Yu, ShouyangWanga, KinKeungLai, FenghuaWenc, 2010. A multiscale neural network learning paradigm for financial crisis forecasting. Neuro computing, 73:716-725

Liu, K. Subbarayan, S. Shoults, R.R. Manry, M.T. Kwan, C. Lewis, F.I. Naccarino, J. Autom. & Robotics Res. Inst., Texas Univ., Arlington, TX, 1996. Comparison of very short-term load forecasting techniques. IEEE Transactions on Power Systems. Vol. 11. No. 2. May 1996

Mirmirani, S. and H.C. L, 2004. Gold Price, Neural Networks and Genetic Algorithm, Computational Economics, 23, 193-200.

國立政治大學 National Chengchi University

Mendelsohn L., Preprocessing data for Neural Networks, 1993. Tech Anal Stocks Commod, 1993:52-58

Mohsen Hayati and Yazdan Shirvany, 2007. Artificial Neural Network Approach for Short Term Load Forecasting for Illam Region. World Academy of Science, Engineering and Technology 28 2007

N.X. Jia, R. Yokoyamaa, Y.C. Zhoub, Z.Y. Gaoc, 2001. A flexible long-term load forecasting approach based on new dynamic simulation theory — GSIM. International Journal of Electrical Power & Energy Systems Volume 23, Issue 7, October 2001, Pages 549-556

Nahi Kandil, Rene´ Wamkeue, Maarouf Saad, Semaan Georges, 2006. An efficient approach for short term load forecasting using artificial neural networks. International Journal of Electrical Power & Energy Systems Volume 28, Issue 8, October 2006, Pages 525-530.

Ray Ruichong Zhang, M.ASCE; Shuo Ma; Erdal Safak, M.ASCE; and Stephen Hartzell., 2003. Hilbert-Huang Transform Analysis of Dynamic and earthquake motion recordings. Journal of Engineering Mechanics, Vol. 129, No. 8, pp. 861-875.

Ray Ruichong Zhang, Shuo Ma, and Stephen Hartzell., 2003. Signatures of the Seismic Source in EMD-Based Characterization of the 1994 Northridge, California, Earthquake Recordings. Bulletin of the Seismological Society of America; February 2003; v. 93; no. 1; p. 501-518.

Ruqiang Yan, Student Member, IEEE, and Robert X. Gao, Senior Member,IEEE., 2006. Hilbert–Huang Transform-Based Vibration Signal Analysis for Machine Health Monitoring. IEEE Transactions on instrumentation and measurement, Vol. 55, No. 6.

Ruey-Hsun Liang, Ching-Chi Cheng, 2002. Short-term load forecasting by a neuro-fuzzy based approach. International Journal of Electrical Power & Energy Systems Volume 24, Issue 2, February 2002, Pages 103-111

Shahriar Shafiee and ErkanTopal, 2010. An overview of global gold market and gold price forecasting. Resources Policy Volume 35, Issue 3, September 2010, Pages 178-189.

Stephen A. Baker and Roger C. van Tassel, 1985. Forecasting the price of gold: A fundamentalist approach Atlantic Economic Journal Volume 13, No. 4, 43-51

國立政治大學 National Chengchi University

Wu, Z., and N. E Huang, 2009. Ensemble Empirical Mode Decomposition: a noise-assisted data analysis method. Advances in Adaptive Data Analysis. Vol.1, No.1. 1-41.

Wei SUN, 2010. Research on GA-SVM Model for Short Term Load Forecasting Based on LDM-PCA Technique. Journal of Computational Information Systems 6:10 (2010) 3183-3189.

Yen-Rue Chang, 2011. Non-stationary time series analysis by using Hilbert-Huang transform: electricity consumption and gold price volatility. Master Thesis of Graduate Institute of Applied Physics, College of Science NCCU.

國立政治大學 · National Chengchi University · 政

國立政治大學
National Chengchi University