International Journal of Environment and Pollution, Vol. 28, No. 3-4, 2006, pp. 364-381

A hybrid adaptive time-delay neural network model for multi-step-ahead prediction of sunspot activity

Jing-Xin Xie[1,3], Chun-Tian Cheng*[,1], Kwok-Wing Chau[2], Yong-Zhen Pei[4]

[1]Department of Civil Engineering, Dalian University of Technology, Dalian, 116024, P.R. China

[2]Department of Civil and Structural Engineering, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

[3]College of Mechanical and Electronic Engineering, Hebie Agricultural University, Baoding 071001, China

[4]Department of Engineering Mathematics, Dalian University of Technology, Dalian 116024, China

Abstract: The availability of accurate empirical models for multi-step-ahead (MS) prediction is desirable in many areas. Some ANN technologies, such as multiple-neural-network, time-delay neural network (TDNN), and adaptive time-delay neural network (ATNN), have proven successful in addressing various complicated problems. The purpose of this study was to investigate the applicability of neural network MS predictive model. Motivated by the above mentioned technologies, we proposed a hybrid neural network model which integrated characteristics decomposition units, and a dynamic spline interpolation unit into the multiple ATNNs. Inside the net, the regular and certain information were extracted to ATNN, while both time delays and weights were dynamically adapted. The yearly average of the sunspots, which has been considered by geophysicists, environment scientists, and climatologists as a complicated nonlinear system, was selected to test hybrid model. Comparative results were presented between traditional MS predictive model based on TDNN and the proposed model. Validation studies indicated that the proposed model is quite effective in MS prediction, especially for single factor time series.

[1,]*Chun-Tian Cheng, Corresponding author. Professor, Department of Civil Engineering, Dalian University of Technology, Dalian, 116024, P.R. China. Tel:+86-411-84708768. Fax:+86-411-84674141 Email: ctcheng@dlut.edu.cn

[1,3]Jing-Xin Xie, Ph.D. Candidate, Department of Civil Engineering, Dalian University of Technology, Dalian, 116024, P.R. China. Email: xjxie@student.dlut.edu.cn

[2]Kwok-Wing Chau. Associate Professor, Department of Civil and Structural Engineering, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. Email: cekwchau@polyu.edu.hk

[4]Yong-Zhen Pei, Ph.D. Candidate, Department of Engineering Mathematics, Dalian University of Technology, Dalian, 116024, P.R. China.

Keywords: time-delay neural network, adaptive time-delay neural network, multiple-neural-network, multi-step-ahead prediction, single step iteration, characteristics decomposition, spline interpolation

Biographical notes

Jing-Xin Xie is currently Ph.D. Candidate in Department of Civil Engineering of Dalian University of Technology, China.

Professor Chun-Tian Cheng is currently Professor in Department of Civil Engineering of Dalian University of Technology, China.

Dr. Kwok-Wing Chau is currently Associate Professor in Department of Civil and Structural Engineering of The Hong Kong Polytechnic University.

Yong-Zhen Pei is currently Ph.D. Candidate in Department of Engineering Mathematics of Dalian University of Technology, China.

1 Introduction

Multi-step-ahead (MS) is a classical model predictive algorithm with which at any given time the process outputs can predict time series values of many time-steps into the future. In most of the published literature, single-step (SS) prediction was used although reliable MS prediction has important applications ranging from system identification to ecological modeling. This can be attributed to several reasons (Parlos *et al.,* 2000). The lack of measurements in the prediction horizon necessitates the recursive use of SS predictors for reaching the end-point in the horizon. Even small SS prediction errors at the beginning of the horizon accumulate and propagate, often resulting in poor prediction accuracy. The situation is even worse for complex systems which are characterized by poorly understandable, noisy and often nonlinear dynamics. Furthermore, the presence of one or more independent process inputs increases the dimensionality of the input space, resulting in a challenging MS prediction problem. The use of linear model structures for MS prediction has been proven unsatisfactory in real-world applications.

One of the earliest attempts in using neural networks for long-range (or multi-step) prediction was reported by Su *et al.* (1992). Schenker proposed a method for training two distinct networks in order to perform long-range prediction (Schenker *et al.*, 1995), in which feedforward networks were used in a feedback configuration. A neural network in a nonrecursive form based on MS prediction was presented by Yang *et al.* (1997). Prasad *et al.* (1998) employed a multivariable long-range predictive control strategy based on a neural network and detailed its application for power plant control. Atiya *et al.* (1999) presented the comparison of various neural methods for MS prediction in time-series. Two approaches to MS prediction were found to attain promising performance. The recurrent neural network was proven to be able to improve MS-based prediction. (Khotanzad *et al.*,1994; Parlos *et al.*, 2000; Bone *et al.*2002) whilst Zhang *et al.* (1998) and Duhoux *et al.* (2001) improved long-term (which is equivalent to MS) prediction by utilizing a combined neural networks. Ahmad *et al.* (2002) also illustrated that combining multiple-neural-network can improve long range (MS) prediction for nonlinear process modeling. In summary, accurate long range predictions can be obtained from recurrent neural networks or combined neural networks. However, training of a recurrent neural network is usually very time consuming and a single recurrent neural network might lack in robustness (Ahmad *et al.*, 2002). In view of these, a combined multiple-neural-network was selected as model primary architecture in this study.

In general, there is no assurance that any individual model has extracted all the relevant information from the data set. Wolpert (1992) proposed the idea of stacked generalization to combine multiple models. Sridhar *et al.* (1996) implemented the stacked generalization for neural network models by integrating multiple neural networks into an architecture known as stacked neural networks (SNNs). Among different SNNs-based approaches, the combination of multiple-neural-networks shows some encouraging results by improving the overall prediction properties (Hashem, 1997; Sharkey, 1999; Sridhar *et al.*1999; Zhang, J., 1999; Ahmad *et al.*, 2002).

Time-Delay neural network (TDNN), introduced by Waibel (1989) and employed time delays on connections in feedforward networks, has been successfully applied in many areas (Haffner *et al.*, 1992; Yamashita, 1997; Ng *et al.*, 1998; Tan *et al.*, 1999, Luk *et al.*, 2000; Shi *et al.*, 2003). An adaptive version of TDNN, adaptive time-delay neural network (ADNN), which was originally proposed by Day *et al.* (1991), adapts its time-delay values and its weights to better accommodate to changing temporal patterns, and to provide more flexibility for optimization tasks. It was successfully utilized in nonlinear system identification (Lin *et al.*, 1995; Yazdizadeh *et al.*, 2000; Yazdizadeh *et al.*, 2002).

In this paper we propose a MS prediction model dynamic spline ATNN (DSTNN) in the form of combined multiple-neural-networks. The model utilizes multiple ATNNs combined with characteristics decomposition units and dynamic spline interpolation units. Inside the net, via features extraction and dynamical spline interpolation units, current and delayed (or past) observations of the measured system input and output are supplied as inputs to the ATNN. Moreover, we employ other numerical algorithms including characteristics decomposition and dynamic spline interpolation as technologies of the above special units. In this paper, the effectiveness of the MS prediction is demonstrated by the application to real-world case studies, and comparison is made with a traditional MS ANN model based on TDNN (simply termed TDNN).

Sunspots time series were selected as the empirical data sets. Nowadays, the study of sunspots activity has practical significance to geophysicists, environment scientists, and climatologists. It is a well-known benchmark time series which is often regarded as nonlinear and non-Gaussian, and is often used to evaluate the effectiveness of nonlinear models (Zhang, 2003). By performing prediction on the sunspots time series, comparative experiments with both TDNN and DSTNN were obtained. All the error analysis demonstrated the effectiveness of the proposed model. By taking advantages of the technology of every unit embodied in the combined network, the model presented can be an effective way to improve forecasting accuracy for MS prediction.

The remainder of this paper is organized as follows. Section 2 reviews the mechanism of a general ATNN neuron and multiple-neural-network architecture for MS prediction. Section 3 presents the hybrid ATNN model as the resolution for multi-step ahead forecasting, and then gives the model structure and algorithms. Section 4 presents prediction results and discussion and also shows the error analysis with linear regression. Section 5 presents the conclusion.

2 ATNN and Multiple-neural-network architecture

2.1 ATNN

Insert Figure 1

The structure of the corresponding dynamic neuron of ATNN is shown in Figure 1, in which $q-\tau$ is the shift operator. The input-output mapping is then governed by

$$y(t)=\sigma\left(\sum_{i=1}^{M}\omega_i x_i\left(t-\tau_i\right)\right) \quad (1)$$

where ω_i's are the neuron weights, τ_i's are the delays, and $\sigma(\cdot)$ is a nonlinear activation function. It has been shown that, even by making the above simplifying assumption, the resulting input-output map is still capable of representing the nonlinear system (Waibel, 1989). It should be noted that the output of the neuron at time *t* depends on the previous values of the inputs resulting in a dynamic behavior. This dynamics will be modified subsequently for representing the nonlinear system.

2.2 Multiple-neural-network for MS prediction

In this paper, we only consider the one-stage p-step ahead MS prediction, which is termed iterative prediction. It is a standard MS iterative approach that is widely utilized in most ANN predictions, and also is a kind of challenging algorithm for testing the adaptability and the generalization of neural networks.

There are many theoretical approaches to network combination such as stacked network and bootstrap aggregation network where multiple networks are created on bootstrap re-sample of the original training data (Sridhar, 1996). For multiple-neural-network, an iterative MS approach as shown in Figure 2 could be followed.

Insert Figure 2

Figure 2 shows a schematic view of general multiple-neural-network structure, in which *X* is the input of combined network, and *Y* is the output. Inside the architecture of neural network, there are more than one subnets conjoined together which can be simulative or not. For single factor MS prediction, the procedure of input-output can be described as Eq. (2). The net shifts new estimates by the input vector, and thus needs a single one-step ahead predictor. Given the time series data set $\{X\ |(x_1,x_2,\dots,x_n)\}$, the new predictions are based on observations, parts of *X*, and a group of previous ones, parts of $\{Y\ |(\hat{x}_{t+1},\hat{x}_{t+2},\dots\ \hat{x}_{t+p})\}$.

$$\begin{cases}\hat{x}_{t+p}=F(\hat{x}_{t+p-1},\hat{x}_{t+p-2},\dots,\hat{x}_{t+1},x_t,\dots,x_{t+p-m}) & p<m\\ \hat{x}_{t+p}=F(\hat{x}_{t+p-1},\hat{x}_{t+p-2},\dots,\hat{x}_{t+p-m}) & p\geq m\end{cases} \quad (2)$$

where the quantities with a "hat" such as $\hat{x}_{t+p},\hat{x}_{t+p-1},\hat{x}_{t+p-m}$ represent estimates of the actual states and outputs, and the others without a "hat" represent observations, i.e. x_t and x_{t+p-m}. p is the number of steps ahead of *p*-step-ahead prediction model *F(.)*. m is defined as the number of inputs of the model which implies the size of moving window of prediction horizon. It is easily seen that, if $p < m$, the model input consists of observations and prediction values, and if $p \geq m$, it consists of all prediction values. It should be noticed that the variable to be predicted uses a constant number of previous values, namely q. That is the reason why the network $F(\cdot)$ can be used in all the steps. In fact, it is trained in a feedforward manner and used as a recurrent model to generate the prediction (Duhoux *et al.*,2001).

3.DSTNN architecture and algorithm

3.1 DSTNN architecture

Insert Figure 3

Integrating dynamic neurons of ATNN, special units and multiple-neural-network, we proposed a combined neural network DSTNN. Its schematic view is shown in Figure 3. DSTNN embodies characteristics decomposition units T_e and P_e, dynamic spline interpolation unit S_I, and several ATNNs. The decomposition units T_e and P_e extract the trend and period factors from observations in order to offer more information to the net. S_I dynamically develops a group of derivative data between two sampling points with cubic polynomial spline interpolation. $ATNN$ is a general adaptive time-delay neural network based on feedforword neural network. It is worth observing that all the inputs and neural network components of the model, i.e. ATNN, are preprocessed by the moving time window with parameter m ($m \leq n$) that can be dynamically adjusted inside the ATNN. In all, a single-factor time series X is supplied to the network. Through the cooperation of the abovementioned units, the prediction results of X, Y, can be obtained. The input of net X can be defined as follows:

$$\{X \mid (x_1, x_2, \ldots, x_m), P_t\} \tag{3}$$

where $(x_1, x_2, \ldots, x_m)$ is input of net, which is the portion of time sequences, m is the number of input knots, and P_t is the step number of MS method at which the model can terminate.

3.2 Algorithms

3.2.1 Multilayer ATNN with spline interpolation

In this study, ATNN with spline interpolation is a dynamic multilayer feedforward network, which is constructed by utilizing the dynamic neuron described in the previous section. The network consists of L layers with N_L neurons in the Lth layer. Since the structures proposed here are intended for identification of single-input single-output nonlinear systems, therefore, the networks have only one input neuron and one output neuron. The bipolar sigmoid function is applied as the activation function. By using the spline interpolation, the typical neuron governing equations are developed as follows

$$
\begin{cases}
net_j^l(t) = \sum_{i=1}^{N^{l-1}} w_{ji}^l o_i^{l-1}(t-\tau_{ji}^l) \\
net_j^{l_s}(t) = \sum_{i=1}^{N^{l-1}} w_{jis_{ji}^l}^l S^l(o_i^{l-1}(t-\tau_{ji}^l), s_{ji}^l) \qquad 0 < \tau_{ji}^l \le \tau_{\max}^l \\
o_j^l(t) = \delta(net_j^l(t) + net_j^{l_s}(t))
\end{cases} \tag{4}
$$

The output of the j th neuron in the lth layer at time t is denoted by $O_j^l(t)$. The first equation depicts the governing algorithm of original typical multilayer adaptive time-delay, in which the weight and associated delay connecting the jth neuron in the *lth* layer to the i th neuron in the (l-1)th layer are denoted by w_{ji}^l and τ_{ji}^l, respectively. It should be noted that j varies from 1 to N^l, i varies from 1 to N^{l-1}, and τ_{ji}^l values form 0 to $\tau_{\max}$. The other neuron governing algorithm related to the spline interpolation is described by the second equation, in which the spline interpolation function $S^l(\cdot)$ corresponds to the l th layer with $o_i^{l-1}(t-\tau_{ji}^l)$ and s_{ji}^l as two parameters. $S^l(o_i^{l-1}(t-\tau_{ji}^l), s_{ji}^l)$ returns a group of the interpolation values between $o_i^{l-1}(t-\tau_{ji}^l)$ and $o_i^{l-1}(t-\tau_{ji}^l+1)$ with the number of fitted data depending on s_{ji}^l. For simplicity, in the following, we use a value of $(\tau_{ji}^l - 2)$, implying that there are $(\tau_{ji}^l - 2)$ spline interpolation data are simulated and inserted into region between $o_i^{l-1}(t-\tau_{ji}^l)$ and $o_i^{l-1}(t-\tau_{ji}^l+1)$. The weight of the corresponding spline interpolations is expressed by $w_{jis_{ji}^l}^l$ with s_{ji}^l denoting the order of interpolation data. $net_j^l(t)$ is the weighted input of the jth neuron in the lth layer at time t, while $net_j^{l_s}(t)$ is the weighted input of the jth neuron in the lth layer at time t related to all the spline interpolation data. Besides, the jth output of lth $o_j^l(t)$ collects all the inputs including both $net_j^l(t)$ and $net_j^{l_s}(t)$ by a nonlinear activation function $\delta(\cdot)$. Thus, when the distance between the sampling points and forecasting point increases, the interpolation number inserted between the current point and the next will also increase. Assume that several prediction values are needed for iteration into inputs node for MS forecasting, we can enrich net inputs with spline interpolation data to reduce the effect of errors iteration.

Thus, the prediction model Eq. (2) can be rewritten as shown

$$\begin{cases} \hat{x}_{t+p} = F(\hat{x}_{t+p-1}, \hat{x}_{t+p-2}, S(\hat{x}_{(t+p-3)}, 1), \hat{x}_{t+p-3}, S(\hat{x}_{(t+p-4)}, 2)\ldots, \\ \qquad \hat{x}_{t+1}, S(x_t, (t+p-2)), x_t, \ldots, \\ \qquad x_{(t+p-\tau_{\max}+1)}, S(x_{(t+p-\tau_{\max})}, (\tau_{\max}-2)), x_{t+p-\tau_{\max}}) & p < \tau_{\max} \\ \hat{x}_{t+p} = F(\hat{x}_{t+p-1}, \hat{x}_{t+p-2}, S(\hat{x}_{(t+p-3)}, 1), \hat{x}_{t+p-3}, S(\hat{x}_{(t+p-4)}, 2), \ldots, \\ \qquad \hat{x}_{t+p-\tau_{\max}+1}, S(\hat{x}_{(t+p-\tau_{\max})}, (\tau_{\max}-2)), \hat{x}_{t+p-\tau_{\max}}) & p \geq \tau_{\max} \end{cases} \tag{5}$$

where two groups of derivative data set, $\{S(\hat{x}_{(t+p-3)}, 1), S(\hat{x}_{(t+p-4)}, 2), \ldots, S(x_t, (t+p-2)), \ldots, S(x_{(t+p-\tau_{\max})}, (\tau_{\max}-2)\}$ (if $p < \tau_{\max}$) and $\{S(\hat{x}_{(t+p-3)}, 1), S(\hat{x}_{(t+p-4)}, 2), \ldots, S(\hat{x}_{(t+p-\tau_{\max})}, (\tau_{\max}-2)\}$ (if $p \geq \tau_{\max}$), are given by the spline interpolation function $S(\cdot)$. Regarding the above inputs rule, the input knots will be increased on a pro rata basis by spline interpolation.

The basic idea of the net is to use higher temporal resolution (i.e., a higher sampling rate and higher frequencies) for the long-term history while using lower temporal resolution for the short-term history (we can compare this approach to the "detailed" certain-memory versus the "general" uncertain-memory that are combined by the human brain when predicting future events). By this means, we can use more essential information on the "detailed" and "general" history of the time series with a relative small number of inputs to the forecasting system. When we make multi-step ahead prediction, the simulated data are uncertain, whereas observations are certain data.

The accumulation of the errors in the recursive predictions renders higher difficulty to achieve accurate long range predictions than accurate one-step-ahead predictions. By using the spline interpolation technology, however, we can reduce the stimulation of previous prediction error on the model, and thus enforce the robustness of net.

3.2.2 Decomposition unit

Single factor prediction systems exist widely in real-world, or to some extent, whilst other factors are not easy to be obtained. For example, in hydrological forecasting area, how to make use of single factor information sufficiently has become more and more vital. Grassberger (1983) proposed a method of time series analysis based on expanding the one-dimension into multi-dimension with phase-space expansion, which can efficiently gather more potential information from single factor observations. He has successfully built a dynamic model and accomplished analysis on climate changing.

Time series is a random variable series that is sorted by time. Almost all measured samples of nonlinear time series have some of the following features such as tendency, period and random. We consider that most time series can be extracted into components by function transforming. The process is shown as follows:

$$X_t = Te_t + Pe_t + Ra_t \qquad t = 1, 2, \ldots n \tag{6}$$

where Te_t is the tendency component; Pe_t is the periodic component; Ra_t is the random component.

In view of the above, we attempt to expand one-dimension time series to multi-dimension for neural network by determining the tendency component and periodic component time series. The method can supply more regular information for training net.

Period extraction

Identifying and extracting potential period components in nonlinear time series is very significant yet difficult. Currently spectra analysis is a relatively fundamental and efficient technology. Entropy in classical thermodynamics in the 19th century is a vital physical quantum. Janes (1957) proposed Principle of Maximum Entropy (POME) on the basis of Entropy and developed the Maximum Entropy Method 1 (MEM1) spectra analysis. Because of its smooth spectrum curve and high resolution ratio, many application methods were later developed by different researchers. The detailed process of period extraction comprises the following steps:

- Step1 Make spectral estimate with MEM1,
- Step2 Obtain remarkable frequencies.
- Step3 Build filters with above frequencies as parameters.
- Step4 Gain period components.
- Step5 Repeat the above four steps until there is no remarkable periods.

According to the entropy analysis theory and its physical background (Press, 1991; Gliebin, 1991), not all the frequencies can be useful. The general rule is a higher frequency together with a lower power (Ma *et al.*, 1996). So the remarkable periods with high frequency can be ignored. On the other hand, the less the phase difference, the more will be errors after filtering. So we also ignore those frequencies with small phase difference.

Tendency extraction

We adapt moving average (MA) and polynomial fit to accomplish tendency extraction. Moving Average (MA) is a primary technology to analyze the tendency of nonlinear system. In the study, we first apply MA to calculate evolution trend of observations, then fit the results with polynomial equation, and lastly extend the polynomial to develop tendency components to enrich prediction model parameters. On the other hand, polynomial fit is based on a polynomial. Different polynomials have different shapes with larger powers (and therefore larger numbers of terms) having steadily more eccentric shapes. Given a set of data, we may want to fit a polynomial curve (i.e., a model) to explain the data. The data is probably noisy, so it is not necessarily to expect the best model to pass exactly through all the points. A low-order polynomial may not be sufficiently flexible to fit closely to the points, whereas a high-order polynomial is actually too flexible. In this study, a least-square algorithm is selected to perform the fit.

Prediction with expanded –dimension

We append the tendency and period components into the prediction model, termed dimension expansion. Then the prediction equation of the net is given by

$$Y = P[F_{SATNN}(\hat{X}), F_{ATNN}(\hat{X}_{Te}), F_{ATNN}(\hat{X}_{Pe})] \tag{7}$$

where F_{SATNN} is the forecasting model based both on ATNN and spline interpolation; F_{ATNN} is the model based on ATNN only; $\hat{X}$ is the input data including previous prediction values; $\hat{X}_{Te}$ and $\hat{X}_{Pe}$ are all estimations of tendency and periodic components. The prediction procedure of the model can be clearly illustrated by schematic view of the DSTNN (Figure 3). The original input data set

X is transformed into two kinds of characteristic data sets, tendency component $\hat{X}_{Te}$ and period component $\hat{X}_{Pe}$, and the number of each set is determined by X. Meanwhile, via dynamic spline interpolation unit, X can also be changed into $\hat{X}$ in which many spline interpolation values are dynamically inserted into the original time sequence at a certain position. Then, semi-finished data sets are converged into the integration model *P(.)* which is usually a simple linear function as the prediction result *Y* arises.

4. Empirical results

4.1 Data sets

Sunspots are dark blotches on the sun which are caused by magnetic storms on the surface of the sun. The underlying mechanism for sunspot appearances is not exactly known. Geller (1988) proposed that increased solar activity during active sunspot periods can influence the atmospheric temperature. The study of sunspot activity has practical importance to geophysicists, environment scientists, and climatologists. The data series is regarded as nonlinear and non-Gaussian, and is often used to evaluate the effectiveness of nonlinear models (Zhang, 2003). The yearly average of the sunspots area has been recorded since 1700, which is a classical example of a combination of periodic and chaotic phenomena and has been served as a benchmark in the statistics literature of time series. Much work has been done in trying to analyze the sunspots data using linear and nonlinear methods. Amir (1998) employed ScaleNet to predict the sunspots from 1700 through 1920 with single step method.

In our study, as a single factor nonlinear sequence, the sunspots average of years 1810 through 1974 is chosen to train and test model for multi-step-ahead forecasting. Two training sets and two testing sets are selected form it. The two testing sets: set 1 (set1)—years 1865 through 1879 and set 2 (set2)—years 1960 through 1974, while the remaining are training sets. The latter are used exclusively for net development, and the former for evaluation.

4.2 Results and discussion

Two neural networks, DSTNN and TDNN, are implemented over the sets mentioned earlier. The 15-step-ahead forecasting is considered. The root-square error (RMSE) and mean absolute error (MAE) are employed as the forecasting accuracy measures. Figure 4 gives the result of entropy analysis with MEM1. Two remarkable frequencies with high amplitude are found, which is 0.0957Hz and 0.01807Hz (the equalities of 11 and 6 years for period depiction) . By tendency extraction, the fit equation is attained as $Y = 21.42 + 1.24x - 0.02x^2 + 7.29E - 5x^3$. The tendency is plotted in Figure 5. Appling filters with the above frequencies as parameters, we get period data sets shown in Figures 6 and 7.

Inset Figure 4

Inset Figure 5

Inset Figure 6

Inset Figure 7

DSTNN and TDNN are all trained and tested using time-delay technique. The comparison between them for 15-step-ahead forecasting is given in Figures 8 and 9.

Though at some data points, the DSTNN model gives worse predictions than TDNN, its forecasting capability is improved in all. For the first six or seven points, the prediction results are quite similar. But with time moving forward, the forecasting values of DSTNN show better accuracy than TDNN, especially in the end of process. More errors occur at 14th and 15th points related to TDNN. The overall forecasting results for the two sets are summarized in Table 1.

Inset Figure 8

Inset Figure 9

Inset Table 1

Error analyses are shown in Figures 10 to 13. In the graph, parameters of linear regressions (correlation coefficients and slope of the best fit lines) for relationships between measured and predicted values by DSTNN and TDNN are presented. The former model predicted better than the latter for two data sets. The difference is apparent for the first data set (see Figures 10 and 11) where DSTNN predicted reasonably well (the correlation coefficient and slope of the best fit lines are 0.9713 and 1.11031, respectively), whereas TDNN predicted poorly (the correlation coefficient and slope of the best fit lines are 0.87755 and 0.92052, respectively). In general, DSTNN can give more focused results on the measured values, although its predicted values are most likely larger than measured values compared with TDNN. The same tendency for set2 of DSTNN predicting better than TDNN is observed in Figures 12 and 13. (refer to the correlation coefficient), but difference in slope of the best fit. In DSTNN forecasting, the correlation coefficient and a slope of best fit line are 0.98988 and 1.29151, respectively, whereas its counterparts of TDNN are 0.96926 and 1.42496. Thus, DSTNN can predict in a more concentrative manner and closer to observations than TDNN for set2.

Inset Figure 10

Inset Figure 11

Inset Figure 12

Inset Figure 13

5.Conclusion

Time series analysis and forecasting is an active research area over the past few decades. The objective of this research study is to present a method for improving MS prediction model for nonlinear systems, which is capable of performing accurate MS, especially in the single factor system. Among all the ANN methods, most predictive models have some drawbacks, which include potential inaccuracies to drifts, inability to incorporate aging, and wear-and-tear effects, development and execution cost (Parlos et al., 2000). On the contrary, several approaches to neural network prediction, such as multiple-neural-network, TDNN and ATNN, have provided satisfactory results.

Inspired by theories and applications of the above mentioned and other technologies, we propose a hybrid model for MS forecasting on single factor time series. The model integrates characteristics discomposing unit, spline interpolation unit, and several ATNNs together. The comparison of empirical results for classical time series of sunspots and a benchmark model based on TDNNs indicate that the multi-function combined model is capable of capturing potential information and relationship in the time series, and takes advantage of the unique strength of every unit. For MS forecasting based on time-delay problems, the net have both dynamic and correlation structures, and can be extended to other professional areas as well.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (No. 50479055) and the Central Research Grant of Hong Kong Polytechnic University (G-T592).

References

Ahmad, Z., and Zhang, J. (2002) 'Improving long range prediction for nonlinear process modeling through combining multiple neural network', *Proceeding of the 2002 IEEE Interational Conference on Control Applications*, pp. 966-971.

Amir, B. G. (1998) 'ScaleNet—Multiscale Neural-Network Architecture for Time Series Prediction', *IEEE Transactions on Neural NetWorks*, Vol. 9, No. 5, pp. 1471-1482.

Atiya, A. F., El-Shoura, S. M., Shaheen, S. I., and El-Sherif, M. S. (1999) 'A comparison between neural network forecasting techniques-case study: river flow forecasting', *IEEE Transactions on Neural Networks*, Vol.10, pp. 402-409.

Barai, S. V., and Pandey, P. C. (1997) 'Time-delay neural networks in damage detection of railway bridges', *Advances in Engineering Software*, Vol.28, pp. 1-10.

Bone, R., Crucianu, M.(2002)'An evaluation of constructive algorithms for recurrent networks on multi-step-ahead prediction', *ICONIP'02 Proceedings of the 9th International Conference on Neural Information Processing,* Vol. 2, pp. 547-551.

Day, S. P., and Davenport, M. R.(1991)'Continuous-time temporal back-propagation with adaptive time delays', *Neuroprose. Archive.*, Ohio State University, Accessible on Internet.

Duhoux, M., Suykens, J., Moor, B. D., and Vandewalle, J. (2001) 'Improved long-term temperature prediction by chaining of neural networks', *International Journal of Neural Systems*, Vol.11, No.1, pp. 1-10.

Geller, M. A. (1988) 'Solar cycles and the atmosphere', *Nature,* Vol.332, pp. 584-585.

Gliebin, J. (1991) Climate change, *Ocean Press*, pp. 131-140, Beijing.

Grassberger P. (1986) Discovery on complexity, *SiChuan Education Press*, ChengDu.

Haffner, P., Waibel, A. (1992) 'Multi-state time delay neural net works for continuous speech recognition'. *Advances in Neural Information Processing Systems*, Vol. 4, pp.135-142.

Hashem, S. (1997)'Optimal linear combination', *Neural Networks*, Vol.10, No.4, pp. 599-614.

Janes, E. T. (1957) 'Information theory and statistical mathematics', *Phys. Rev.,* Vol.106, pp. 620-630.

Khotanzad, A., Abaye, A., and Maratukulam, D. (1994) 'An adaptive recurrent neural network system for multi-step-ahead hourly prediction of power system loads', *1994 IEEE International Conference on Neural Networks,1994 IEEE World Congress on Computational Intelligence,* Vol.5, pp. 3393-3397.

Lin, D.-T., Dayhoff, J. E., and Ligomenides, P. A. (1995) 'Trajectory Production with the adaptive time-delay neural network'. *Neural Network*, Vol.8, No.3, pp. 447-461.

Luk, K. C., Ball, J. E., Sharma, A. (2000) 'A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting', *Journal of Hydrology,* Vol.227, pp. 56-65.

Ma, M. J., Tian, S. Z., Zheng, W. Z., and Chai, X. M.(1996) 'Spectral analysis and check method on Monthly mean water level periodic signal', *Acta Oceanologica Sinica*, Vol.18, No.3, pp. 5-12.

Ng, G. W., and Cook, P. A. (1998) 'Real-time control of systems with unknown and varying time-delays using neural networks', *Eng. Applicat. Artif. Intell.*, Vol. 11, pp. 401-409.

Parlos, A. G., Rais, O. T., Atiya, A. F. (2000) 'Multi-step-ahead prediction using dynamic recurrent neural networks', *Neural Network*, Vol.13, pp. 765-786.

Prasad, G., Swidenbank, E., and Hogg, B. W. (1998) 'A neural net model-based multivariable long-range predictive control strategy applied in thermal power plant control', *IEEE Transactions on Energy Conversion,* Vol.13, pp. 176-182.

Press, W. H. (1991) Entire numerical methods, the art of scientific numeration, Lanzhou Uuniversity Press, pp. 622-627, LanZhou.

Schenker, B., and Agarwal, M. (1995) 'Long-range prediction for poorly-known systems', *International Journal of Control*, Vol.62, pp. 227-238.

Sharkey, A. J. C. (1999) 'Multi nets system', *Combining Artificial Neural Nets Ensemble and Modular*, Amanda J.C Sharkey(Ed), Springer Publication, London.

Shi, D., Zhang,H. J., and Yang, L.M. (2003) 'Time-Delay Neural Network for the Prediction of Carbonation Tower's Temperature', *IEEE Transactions on Instrumentation and Measurement,* Vol. 52, No. 4, pp. 1125-1128.

Sridhar, D. V., Bartlett, E. B., and Seagrave, R. C. (1996) 'Process modeling using stacked neural networks', *AIChE Journal*, Vol.42, No.9, pp. 2529-2539.

Sridhar, D. V., Bartlett, E. B., and Seagrave, R. C. (1999) 'An information theoretic approach for combining neural network process models', *Neural Networks,* Vol. 12, pp. 915-926.

Su, H. T., McAvoy, T. J., & Werbos, P. J. (1992) 'Long-term predictions of chemical processes using recurrent neural networks: a parallel training approach', *Industrial Applications of Chemical Engineering Research*, Vol.31, pp. 1338-1352.

Tan,Y. , and Cauwenberghe, A. V. (1999) 'Neural-network-based d-step-ahead predictors for nonlinear systems with time delay', *Eng. Applicat. Artif. Intell.*, Vol.12, pp. 21-35.

Waibel,A., Hanazawa, T., Hinton, G., Shikano, K., and Lang, K. J. (1989) 'Phoneme recognition using time-delay neural networks ', *IEEE Transactions on Acoustics, Speech, Signal Procesings*, Vol.37, No.3, pp. 328-339.

Wolpert, D. H. (1992) 'Stacked generalization'. *Neural Networks*, Vol.5, No.2, pp. 241-259.

Yamashita, Y. (1997) 'Time delay neural networks for the classification of flow regimes', *Comput. Chem. Eng.*, Vol. 21, pp. S367-S371.

Yang, Y. and Chai, T. (1997) 'Soft sensing based on artificial neural network', *Proceedings of the 1997 American Control Conference*, pp. 674-678, Boston, MA: USA.

Yazdizadeh, A., and Khorasani, K., and Patel, R. V. (2000) 'Identification of a two-link flexible manipulator using adaptive time delay neural networks'**,** *IEEE Transactions on Systems, Man and Cybernetics, Part B,* Vol.30 , No. 1, pp. 165-172.

Yazdizadeh , A., and Khorasani , K. (2002) 'Adaptive time delay neural network structures for nonlinear system identication', *Neurocomputing* , Vol.47, pp. 207-240.

Zhang, P. G. (2003) 'Time series forecasting using a hybrid ARIMA and neural network model', *Neurocomputing*, Vol.50, pp. 159-175.

Zhang, J. (1999) 'Inferential estimation of polymer quality using bootstrap aggregated neural networks', *Neural Networks*, Vol.12, pp. 927-938.

Zhang, J., Martin, E. B., and Morris, A. J. (1998) 'Long-term prediction models based on mixed order locally recurrent neural networks', *Computers & Chemical Engineering*, Vol.22, pp. 1051-1063.

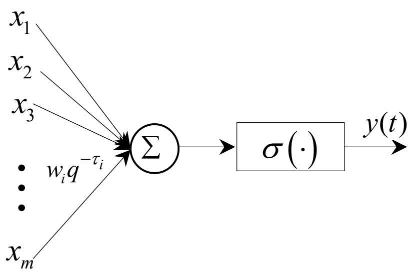

Figure 1 Dynamic neuron in ATNN. $q^{-\tau}$, the shift operator. $\sigma(\cdot)$, activation function. Reprinted from Lin (1993).

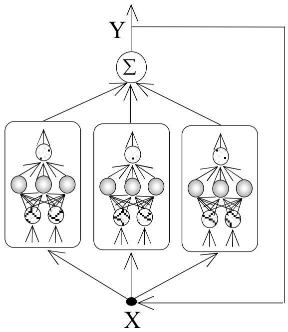

Figure 2 MS prediction based on multiple-neural-networks. Input vector set *X*, contains all the observations. Output vector set *Y*, consists of prediction values. In iteration prediction, a previous prediction of the model iterates into the input knots.

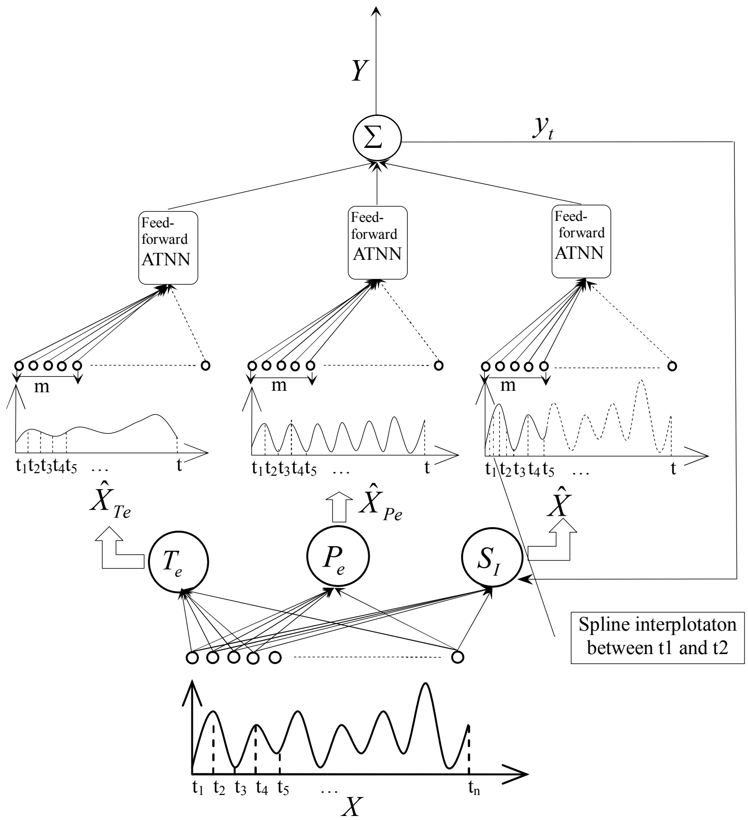

Figure 3 A schematic view of the DSTNN. Characteristics decomposition units, T_e and P_e. Dynamic spline interpolation unit, S_I .Net input, X . Prediction result, Y . General feedforward adaptive time-delay neural network, $ATNN$. The values corresponding to dashed dotted curves between t_1 and t_2 can be obtained by dynamic spline interpolation unit.

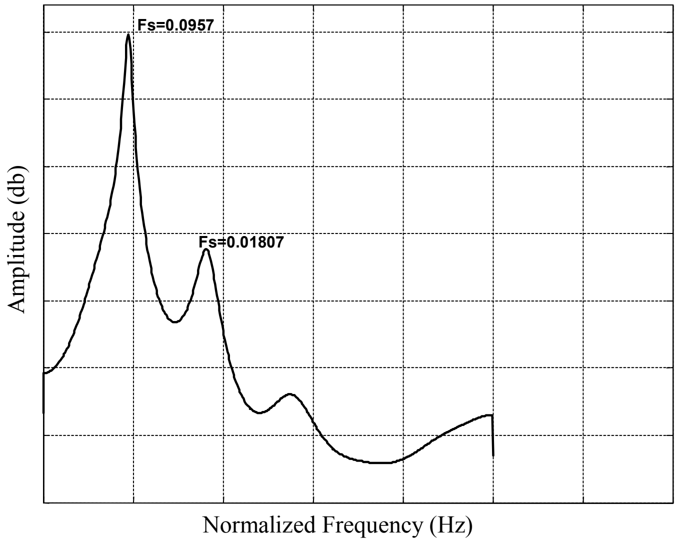

Figure 4 The result of entropy analysis with MEM1. Two remarkable amplitudes (db) can be obtained. The corresponding frequencies are 0.0957 and 0.01807.

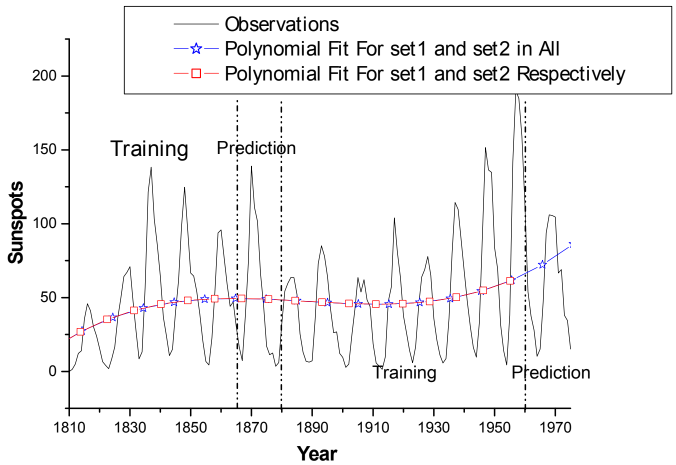

Figure 5 Polynomial fit results for three sets. Dashed pentacle for testing set1 and set2. Dashed square for all the training set.

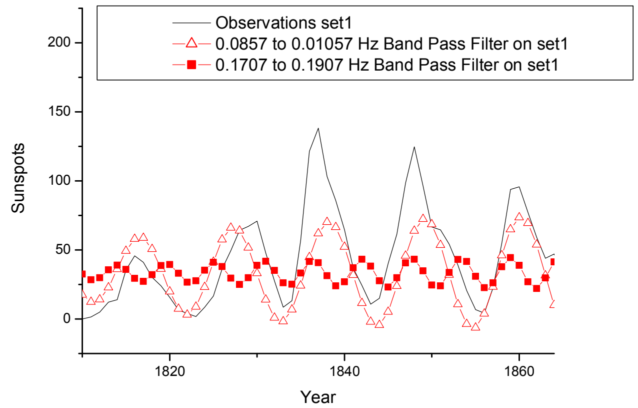

Figure 6 Two curves obtained by using band pass filters on set 1with different periods: –△–, curve with period of 11-year ; –■–, curve with period of 6-year.

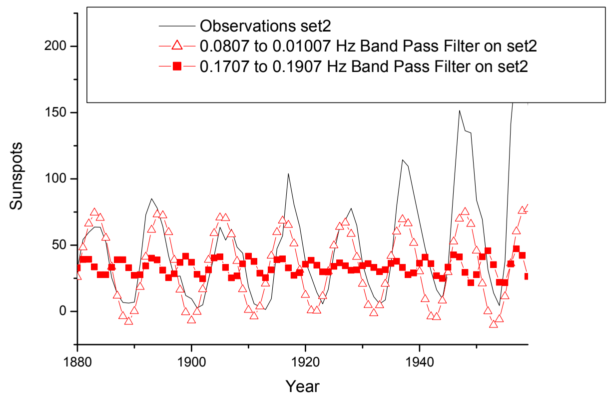

Figure 7 Two curves obtained by using band pass filters on set 2 with different periods: –△–, curve with period of 11-year ; –■–, curve with period of 6-year.

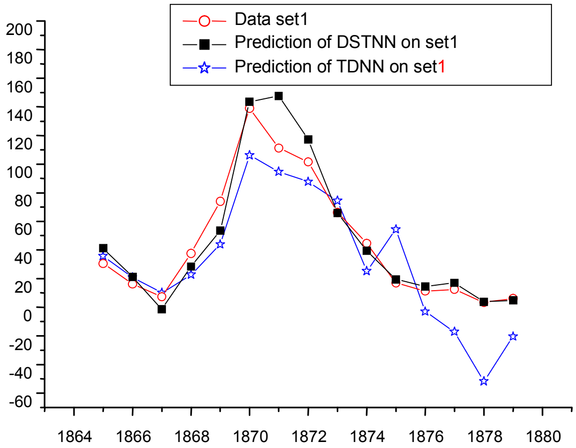

Figure 8 Prediction results of two neural networks for set 1. First 5 points are similar, while from 1870 the DSTNN obtains more accurate values than TDNN, especially in the end of multi-step-ahead.

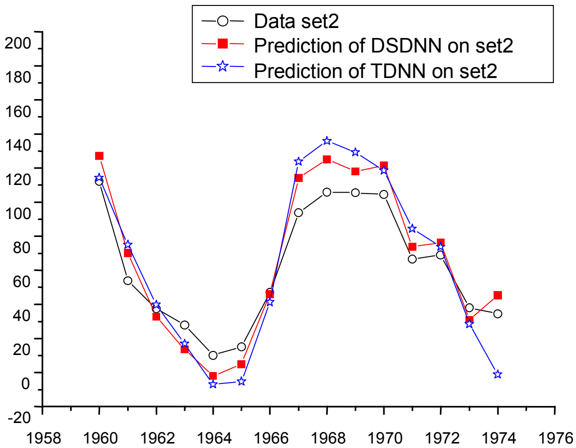

Figure 9 Prediction of two neural networks for set 2. First 7 points are similar, while from 1967 the DSTNN obtains more accurate values than TDNN, especially in the end of multi-step-ahead.

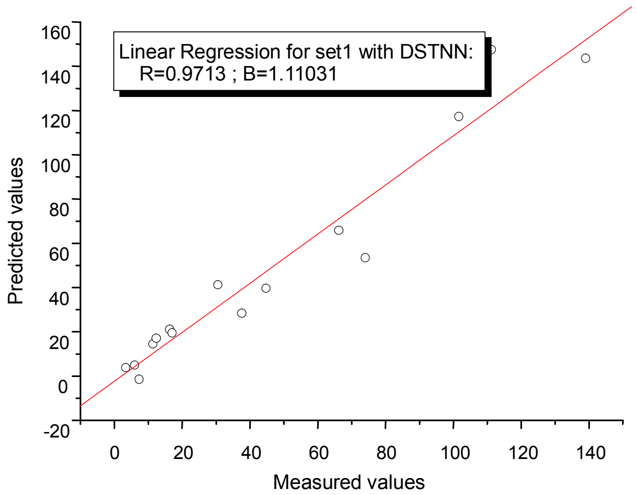

Figure 10 Correlation between the prediction results for set1 with DSTNN and the observations. The correlation coefficient and slope are 0.9713 and 1.11031, respectively.

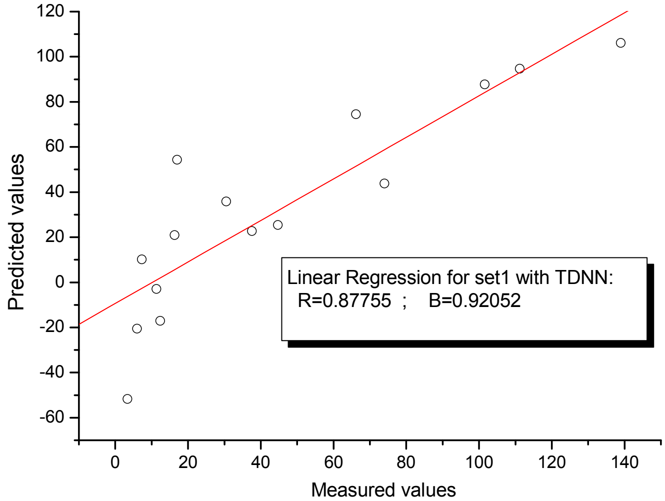

Figure 11 Correlation between the prediction results for set1 with TDNN and the observations. The correlation coefficient and slope are 0.87755 and 0.92052, respectively.

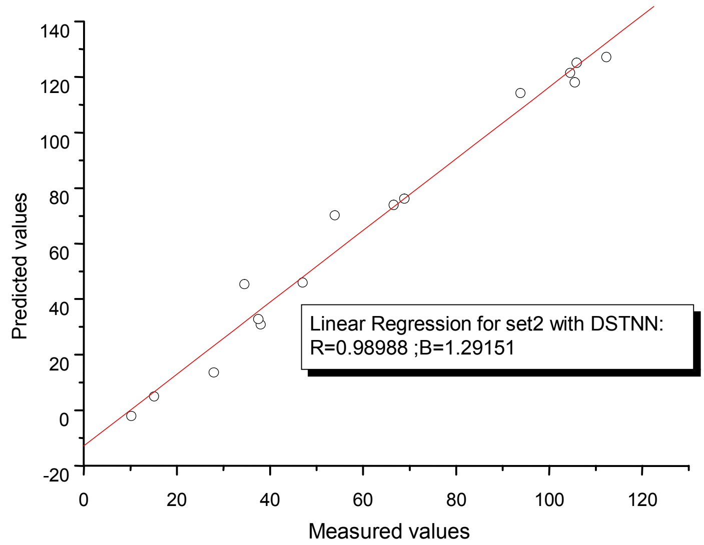

Figure 12 Correlation between the prediction results for set2 with DSTNN and the observations. The correlation coefficient and slope are 0.98988 and 1.29151, respectively.

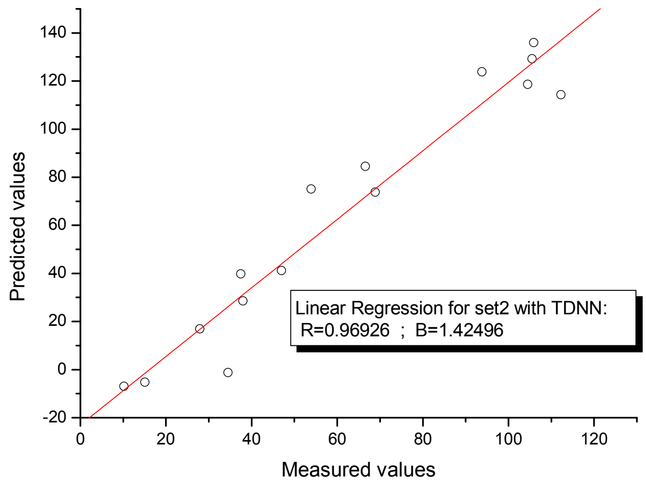

Figure 13 Correlation between the prediction results for set2 with TDNN and the observations. The correlation coefficient and slope are 0.96926 and 1.42496, respectively.

Table 1 The RMSE and MAE results for forecasting accuracy measures. Both RMSE and MAE indicate that DSTNN can offer more accurate results.

	set 1		set 2	
	DSTNN	TDNN	DSTNN	TDNN
RMSE	0.26	0.78	0.19	0.3
MAE	8.52	20.76	11.69	16.37