# Embedding dimension estimation of high dimensional chaotic time series using distributed time delay neural network

MARYAM PARIZANGENEH[1], MOHAMMAD ATAEI [2], PEYMAN MOALLEM [2]
[1]Control engineering department, [2]Electrical engineering department
[1]Azad university najafabad branch, [2]Isfahan university
[1]University of Najafabad, Bolvar daneshgah st., Najafabad ,Isfahan, Iran, Zipcode.517,
[2]University of Isfahan, Hezar jerib st., Isfahan, Iran, Postal code.81746-73441
IRAN

*Abstract:* - In the face of a practical chaotic system whose mathematical model is not available, because of unknown input factors and unavailable dynamical equations, using time series approach can be useful. Therefore, space state reconstruction of a chaotic system by using a scalar time series from its output observations is considered for obtaining information on this system from its one-dimensional signal. In this paper a method for estimation of an appropriate embedding dimension for phase space reconstruction of underlying high dimensional system from the observed chaotic time series by a Distributed Time Delay Neural Network (DTDNN) is proposed. Various methods for embedding dimension estimation have been previously studied from which False Nearest Neighbours (FNN) is the most conventional method, however the performance of this method for the high dimensional chaotic systems is not acceptable. The proposed method is applied to high dimensional chaotic systems such as approximated Mackey-Glass time series with dimensions 7, 13 and data set $D_1$ from the Santa Fe institute. Our method for embedding dimension estimation has been compared with the conventional estimation method, and their comparisons showed the effectiveness of the proposed methodology. The results show that this method is feasible and fit for the embedding dimension estimation of high-dimensional chaotic systems.

*Key-Words:* - Embedding dimension, High dimensional chaotic time series, False nearest neighbors, Distributed time delay neural network.

## 1 Introduction

Observation of chaos has been reported in numerous fields such as in physics, laser technology, chemistry, biology, and etc. Hence, in last decade there have been successful attempts by many researchers to apply a branch of nonlinear analysis, chaos theory, to model and study highly irregular signals arising in different fields of natural sciences. Nonlinear time series analysis is a powerful tool for obtaining information on the nonlinear dynamical system from their one-dimensional signals. The basic idea of nonlinear time series analysis is that a complex system can be described with a strange attractor. This idea is called embedding, which has been proved by Takens [1]. The embedding theorem of Takens offers the concept of embedding a single variable series in a multi dimensional phase space to represent the underlining dynamics. By examining a single as a geometric object in its state space, new types of information about the underlining system become available to the practitioner. The embedding theorem guarantees that the space of time delayed vectors with sufficiently large dimension will capture the structure of original phase space. However, the embedding theorem did not directly answer how to choose embedding dimension $m$ and delay time $\tau$ .Therefore determination of the optimal embedding parameters for phase space reconstruction of nonlinear dynamics has been studied as an important problem. There have been many discussions on how to estimate the appropriate embedding dimension from a scalar time series. Three basic methods are used to choose the optimal embedding dimension:

1) Correlation theorem is employed to estimate appropriate dimension $m$. By increasing the embedding dimension used for the computation one notes an appropriate dimension $m$ when the value of the correlation dimension stops changing, in other words, the concept of the correlation function is that a seemingly irregular phenomenon arising from deterministic dynamics will have a limited number of degrees of freedom equal to the smallest number of first order differential equations that capture the most important features of dynamics. This method is often very data intensive, certainly subjective, and time-consuming for computation.

2) The idea of singular system analysis that determines an appropriate embedding dimension $m$ directly from the raw time series. It provides its convenience for the further analysis of the given system. Numerical experience, however, led several authors [2-3] to express some doubts about reliability of singular system analysis in the attractor reconstruction. Singular value decomposition (SVD), the heart of singular system analysis and by nature a liner method may cause to misleading technique when it is used in nonlinear dynamics studies that reconstruction parameters are time-delay and embedding dimension.

3) The method of false nearest neighbours (FNN) that is a geometric approach for the estimation of the sufficient embedding dimension. It is based on the property of chaotic attractors in that their orbits should not intersect or overlap with each other. Such an intersection or overlap may result when the attractor is embedded in a dimension lower than the sufficient one stated by the delay embedding theorem [4].

Disability of these methods for determining embedding dimension using measurements taken from a high dimensional chaotic time series encouraged us to use alternative models. One that has become popular over the last decades is the artificial neural network. One popular choice of neural network for some practical engineering problem such as recognition and nonlinear prediction is the time delay dynamic neural network (TDNN). Stability analysis of TDNNs has been analyzed in several studies [5-7].

In this work, a DTDNN, called Distributed Time Delay Neural Network is introduced for estimating embedding dimension, with emphasis on utilizing temporal characteristics, which has special structures that store temporal information explicitly using time-delayed structures at the first layer (hidden layer), and also implicitly time-delayed structures are connected to the input of other layers. In our method number of the taps in the first Tapped Delay Line (TDL) changes and for each $m$ value, the network is trained to predict the current measurement $y(t+1)$, after training process, mean absolute prediction errors for training, validation and test sets are recorded as a function of $m$.

In section 2, the most popular conventional phase space reconstruction method of a chaotic time series will be given. Proposed method to accurately determine the embedding dimension of high dimensional nonlinear time series will be presented in section 3. Introduction of approximated Mackey-Glass chaotic systems, real data set $D_1$ and performance of DTDNN on them over widely used conventional method (FNN) is evaluated in section 4. Finally in section 5, the conclusion of this work will presented.

# 2 phase space reconstruction of chaotic time series

In real life, the system under study gives usually one observable, i.e., the only information about the system, is one-dimensional signal. The monitoring of a single scalar observable is sufficient for characterizing and understanding the dynamics on a finite-dimensional attractor, although we have any knowledge neither of system equations nor of geometry of its phase portraits**.** For a scalar time series $y(1), y(2),..., y(N)$ the phase space can be reconstructed using method of delays. The basic idea in the method of delays is that the evolution of any single variable of a system is determined by the other variables with which it interacts. Information about the relevant variable is thus implicitly contained in the history of each single variable. Equivalent phase space can be reconstructed by assigning an element of the time series $y(t)$ and its successive delays as coordinates of a new vector called time-delay vector , and so according to Tankens [1] and Packard [8], the method of delays can be used to embed a scalar time series into an m-dimensional space as follows:

$$Y_i(t) = [y(t_i), y(t_i-\tau), y(t_i-2\tau),..., y(t_i-(m-1)\tau)]^T$$
$$i = 1,2,..., N-(m-1)\tau \qquad (1)$$

where $m$ is the embedding dimension and $\tau$ is the time delay. Note that $\mathrm{Y}_i(t)$ means ith reconstructed vector whit embedding dimension $m$. It is clear that for correct reconstruction, a fine estimation of the parameters ( $m, \tau$ ) is needed.

## 2.1 Determination of the delay time

Takens`s theorem assumes that we have an infinite noise-free data set , in which case , we can choose the delay time almost arbitrarily. Of course, infinite noise-free data simply do not occur, therefore, some thought must be given to choosing a delay such that the underlying dynamical attractor is faithfully reconstructed. If the time delay is too small, the reconstructed attractor is compressed along the identity line (the 45° line in phase space), i.e. the resulting vectors will be very nearly the same and each will be carrying a great deal of redundant. If the time delay is too large, the attractor dynamics may become causally disconnected and this is called irrelevance. For chaotic systems, tiny errors in data become exponentially magnified in time so that too large a delay will tend to decorrelate the signal from itself $\tau$ time steps into the future [9]. There are several heuristic methods for determining parameter time delay $\tau$ such as [10]. The most representative way for choosing the time delay is the probabilistic method advocated by Fraser [11] known as Average Mutual Information (AMI). The idea is to extract the amount of information that a vector at discrete time t, conveys about itself some lag $\tau$, later by computing:

$$I(\tau) = \sum_{y(t),y(t+\tau)} P(y(t), y(t+\tau)).\log_2\left[\frac{P(y(t), y(t+\tau))}{P(y(t)).P(y(t+\tau))}\right] \quad (2)$$

where $p(y(t))$, $p(y(t+\tau))$ and $p(y(t), y(t+\tau))$ are the marginal and joint probability densities for observations $y(t)$, $y(t+\tau)$ and the sum is taken over all non-zero probabilities. According to [11] the first minimum in the AMI graph is considered as the most suitable choice for $\tau$, since this is the time when $y(t+\tau)$ adds maximum information to the knowledge we have from $y(t)$.

## 2.2 Determination of the embedding dimension

The most popular conventional method used for the estimation of sufficient embedding dimension is based on the fact that choosing too low an embedding dimension results in points that are far apart in the original phase space being moved closer together in the reconstruction space.In order to apply false nearest neighbours (FNN) method, we do following steps. First we should make delay vectors in dimension $m$ according to Eq. (1), using the time delay $\tau$ suggested by average mutual information .Then we examine rth near neighbour in phase space of each delay vectors:

$$Y_r^{NN}(t) = [y(t_r), y(t_r-\tau), y(t_r-2\tau),..., y(t_r-(m-1)\tau)]^T$$
$$r = 1,...,5 \quad (3)$$

that can be in proximity in the phase space because of the dynamic evolution of the orbits or due to an overlap resulting from the projection of the attractor to a lower dimension.

If the vector $Y_r^{NN}(t)$ is a false neighbour of $Y_i(t)$ having arrived in its neighbourhood by projecting from a higher dimension because the present dimension $m$ does not unfold the attractor, then by going to the next dimension $m+1$ we may move this false neighbour out of the neighbourhood of $Y_i(t)$. By looking at every delay vectors $Y_i(t)$ and asking at what dimension we remove all false neighbours, we will sequentially remove intersections of orbits of lower and lower dimension until at last we remove point intersections. At that juncture we will have identified that $m$ where the attractor is unfolded.

The square of the Euclidian distance between neighbours points in $m$ dimension is:

$$R_m^2 = \sum_{i=0}^{m-1}\left[y(t-i\tau) - y(t_r-i\tau)\right]^2 \quad (4)$$

Moving from dimension $m$ to $m+1$ means that a new coordinate equal to $y(t-m\tau)$ is being added in each delay vectors, so the Euclidean distance of two points in dimension $m+1$ will be:

$$R_{m+1}^2 = \sum_{i=0}^{m}\left[y(t-i\tau) - y(t_r-i\tau)\right]^2$$
$$= R_m^2 + \left[y(t-m\tau) - y(t_r-m\tau)\right]^2 \quad (5)$$

The relative distance between the two points in two consecutive embedding dimension $m$ and $m+1$ will be the ratio:

$$\sqrt{\frac{R_{m+1}^2 - R_m^2}{R_m^2}} = \frac{\left|y(t-m\tau) - y(t_r-m\tau)\right|}{R_m} \quad (6)$$

when this distance ratio is greater than a predefined threshold, we have a false neighbour (being in the same neighbourhood because of the projection and not because of the dynamics). It should be mentioned that the determination of which neighbour is true and which is false is quite insensitive to the threshold we use. A scalar value of $\gamma = 15$ are recommended by Abarbanel [9] is usually used although values of 10 to 20 will produce little variability in the results.

Although FNN is one of the most conventional methods for determining embedding dimension, the performance of this method for the high dimensional chaotic systems is usually not acceptable and it gives poor estimate of $m$ as we will show in the sec. 4.2.1. In order to remove this defect, we present a method that will improve the embedding dimension estimation for the high dimensional chaotic systems.

# 3 Proposed method for embedding dimension estimation

When the system or physical mechanism is nonlinear time-varying, we have a more difficult task in the context of neural network, usually time lagged feed forward networks are good candidate for temporal processing. Temporal tasks and static tasks have fundamentally different characteristics. The coordinates of a static vector in an N-dimensional feature space can be randomly permuted with out affecting the classification. In contrast, the temporal sequence is order sensitive. The results will strongly depend on the order of each feature.

In practice, time is important in many of cognitive tasks such as signal processing. The memoryless networks are inadequate for temporal tasks. For a neural model to deal with temporal tasks, the network should have dynamic properties which make the network responsive to time-varying signals. For a memoryless network to be dynamics, it must be given the memory. For example, a multilayer perceptron network may be made dynamic by introducing time-delayed structures to the hidden, and/or output layers. Another way is to route delay connections from one layer to another. The memory elements can be either fully or sparsely interconnected.

In order to estimate optimal embedding dimension, we will use a class of dynamic neural network, which consist of a feed forward network with a tapped delay line (TDL) at the first layer (hidden layer) and also distributed tapped delay lines at other layers throughout the network. This is called the Distributed Time Delay Neural Network (DTDNN), in which the dynamics appear at the layers of a static multi layer feed forward network. In the context of neural network, time delay feed forward networks are appropriate candidate for temporal processing because in order to understand complicate chaotic dynamical system, a layered feed forward neural network need that have multiple layers and sufficient interconnections between units in each of these layer, this is to ensure that the network will have ability to learn complex nonlinear decision surfaces [12], another necessary property that persuades us to use time delay feed forward networks, is the ability to represent relationships between events in time [13]. For a neural network to be dynamic, it must be given memory. Most commonly used form of short-time memory is called a tapped delay line (TDL) memory. In our distributed time-delay dynamic neural network, historical information is stored in its first layer (hidden layer) explicitly using time-delay structures of order $m$. This memory element is the $\tau$ time delay, which has the transfer function $H(Z) = Z^{-\tau}$. Although such a memory give 100% accurate historical information, its size is always limited because an increase in memory size increases significantly the complexity of the structure and makes learning much more difficult. A hidden neuron with TDL at its input is shown in Fig. 1.

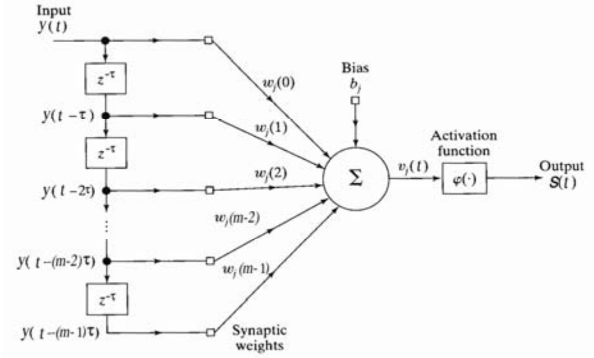


Fig. 1. A hidden neuron (processing unit) with TDL at its input.

The output, in response to the input $y(t)$ and it's past values, $y(t-\tau), y(t-2\tau),..., y(t-(m-1)\tau)$ given by:

$$s_j = \varphi(\sum_{i=0}^{m-1} w_j(i) y(t - i\tau) + b_j) \qquad (7)$$

Where $\varphi$ is the transfer function of jth neuron, the $w_j(i)$ is its synaptic weights, and $b_j$ is the bias. The model of Fig. 1 is referred to as a multiple input to neuronal filter. It is considered as a distributed neuronal filter, in the sense that the filtering action is distributed across different points in space.

Since, we have a chaotic time series generated by a deterministic dynamic system, what has been learned from historical behaviour always apply future. Our measurements arising from a chaotic system are produced from a map $h: R^n \to R^1$ where $R^n$ is the space of the original unknown attractor of the system. The evolution of original state can be expressed as $Y(t+1) = F(Y(t))$, where $Y(t)$ and $Y(t+1)$ are n dimensional vectors describing the state of the system at times $t$ and $t+1$, the problem is to find a good approximation to $F$. The original system dynamics can be reconstructed in space of dimension $m$, using the delay-coordinate map. The states in this space are the time-delay vectors $Y_i(t) \in R^m$ as shown in Eq. (1) and evolution of reconstructed states expressed as $Y_i(t+1) = F_i(Y_i(t))$. However, as it often happens, we are only interested in forecasting the first component ( $y(t+1)$ ) of next time delay vector $Y_i(t+1)$, the search is limited to a map $g: R^m \to R^1$,which interpolates the pairs

($Y_i(t), y(t+1)$) instead of a function $F_i : R^m \to R^m$ [14].

If the reconstructed system is equivalent to the original one, the function $g : R^m \to R^1$ can be used to recreate the original measurements $\{y(t)\}_{t=1}^N$. The learning of nonlinear predictive mapping is the most important task for us. A neural network should be used with $m$ dimensional phase space vectors $Y_i(t)$ as the inputs and the scalars $y(t+1)$ as the outputs to approximate a map $g : R^m \to R^1$.

We used a distributed time-delay neural network which consists of two Layers. The first layer has 10 neurons with tangent hyperbolic transfer function and a tapped delay line (TDL) at its input. The second layer has one neuron with a linear transfer function which is the customary practice in the design of nonlinear regression models and also a TDL at its input. Our distributed time delay dynamic neural network is shown at Fig. 2, that is a more powerful nonlinear filter consisting of a tapped delay line memory of order $m$ and delay time $\tau$ at the input of its hidden layer and also another tapped delay line of order 2 with zero and unit delay time at the input of its output layer. Some other different structures were also tried but the best results were obtained with this structure.

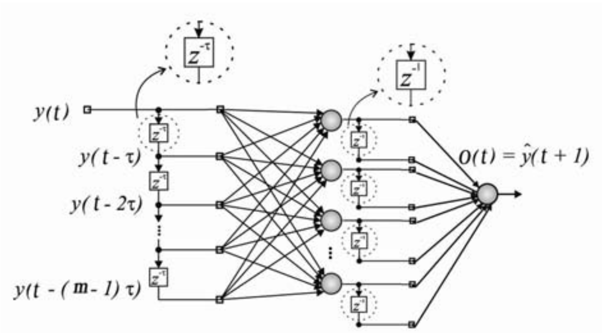


Fig. 2. Distributed time delay neural network

In order to pre-processing of data, desired time series is normalized by mapping minimum and maximum values to [-1 1], then the TDL is fed with normalized $\{y(t)\}_{t=1}^N$ to produce delay vectors $Y_i(t) \in R^m$ according to Eq. (1).

We used LM back propagation that is a network training that updates weight and bias values according to Levenberg-Marquardt optimization [15].The DTDNN requires dynamic back propagation to compute the network gradient. This is because the tapped delay line appears at the input of every layer in the network.

The first TDL as an SIMO structure (Single Input Multiple Output) takes $y(t)$ as input and according to selected $\tau$ (from sec. 2.1) and different $m$ produces time delay vectors $Y_i(t)$ with various dimensions to predict the current measurement $y(t+1)$ where $m$ is the number of taps in TDL.As $m$ is changed, each time the network is trained to predict the current measurement $y(t+1)$ and the prediction error between the output of network $\hat{y}(t+1)$ and $y(t+1)$ is computed. When we change the number of taps in TDL, mean absolute error measurement is recorded as a function of $m$. The mean absolute error (MAE) is given as:

$$MAE = \frac{\sum_{i=1}^{N} \| y_i - \hat{y}_i \|}{N} \qquad (8)$$

where $y_i$ is the desired value and $\hat{y}_i$ is the predicted value.

As the number of taps in the tapped delay line increases, the MAE decreases. At one point, further increase of taps does not improve absolute error significantly. This point is considered as the optimal embedding dimension if we have no improvement of error for several later dimensions, so we consider the number of taps in the TDL equals $m$ of the system.

## 4 Simulation Experiment

In order to evaluate performance of our proposed method, data generated from approximated Mackey-Glass chaotic system has been used for simulation. Furthermore, , we have used one practical system which is $D_1$ data set from the Santa Fe institute for simulation one real data set.

### 4.1 Description of data set

We introduce three high dimensional chaotic time series in this section.

#### 4.1.1 Real world data sets $D_1$

Data set $D_1$ were recorded by Santa Fe institute, the details can be found in [16]. The 6400 sequential data is used in our study. The first 4000 points uses to feed the net and perform the training. The validation process is carried out for the next 1200 points, and the last 1200 points is considered as the test set.

#### 4.1.2 Mackey -Glass time series

MG system is represented by the following differential equation:

$$\dot{x}(t) = \frac{0.2x(t-\Delta)}{1+x(t-\Delta)^{10}} - 0.1x(t), \qquad (9)$$

where $\Delta$ is a lag-time, above equation can be approximated by the following difference equation:

$$x(n+1) = \frac{1}{2m+bt_f} \qquad (10)$$

$$\left((2m-bt_f)x(n) + at_f\left(\frac{x(n-m)}{1+x(n-m)^{10}} + \frac{x(n-m+1)}{1+x(n-m+1)^{10}}\right)\right)$$

which can be used to produce chaotic systems of dimension $m+1$ [17]. The time series is generated with the parameters $a=0.2$, $b=0.1$ and initial values are generated randomly. 5000 values of this $x(n+1)$ time series is used in this study. We consider the first 3000 points as the training set, next 1000 points as the validation set and the last 1000 points as the test set. We fixed $t_f$ to be 23 and tested our proposed method with two examples from the MG approximation of dimension 7,13.

## 4.2 Simulation Results

An investigation on the performance of FNN and the proposed approach in embedding dimension estimation is conducted in this section.

### 4.2.1 Simulation results by conventional methods

The simulation results of the conventional methods are shown in Figs. 3 and 4. From the Fig. 3 the delay time $\tau$ are given for all three time series. In order to determine the time delay $\tau$ in the memory structure (TDL) at the first layer (hidden layer), we need to know $\tau$. AMI algorithm gives $\tau=4$ (MG7), $\tau=7$ (MG13), $\tau=7$ (data set $D_1$) respectively.

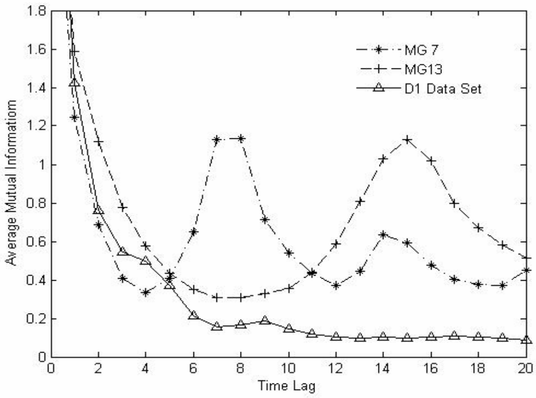


Fig. 3. Average mutual information.

The plots in Fig. 4 show the results of the estimated $m$ for the three systems using FNN method. As we can see from the Fig. 4, FNN gave poor estimates of $m$ for the MG systems of dimension 7, 13 and data set $D_1$ which has 9 degrees of freedom [16]. FNN gives $m=4$ (MG7), $m=5$ (MG13), $m=5$ (data set $D_1$) respectively.

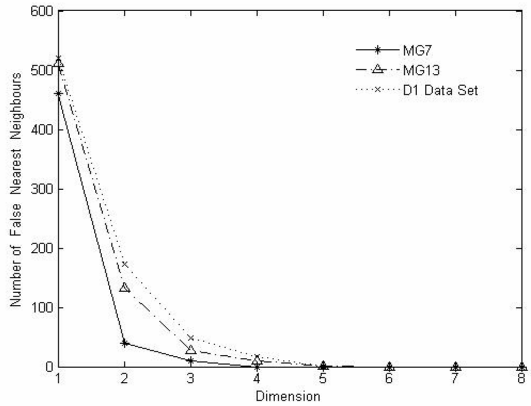


Fig. 4. False nearest neighbors for high dimensional chaotic systems.

### 4.2.2 Simulation results by proposed methods

The resulting plots for training, validation and test sets when using our method to estimate $m$ for the high dimensional systems from the observed chaotic time series are shown in Figs. 5,6 and 7.As we can see from these figures the MAEs have dropped significantly when the dimension $m$ has reached the minimum embedding dimension which the MAE of the prediction error was not changing significantly after it. The number of tap that after it, there is no significant improvement in mean absolute errors are $m=7$ (MG7), $m=13$ (MG13) and $m=9$ (data set $D_1$) respectively. As seen from these figures, our proposed method gave good estimates of $m$ for the high dimensional cases. These simulation results show that the estimated embedding dimension is good enough for reconstruction of the phase space as is evident from comparison with the results of conventional method.

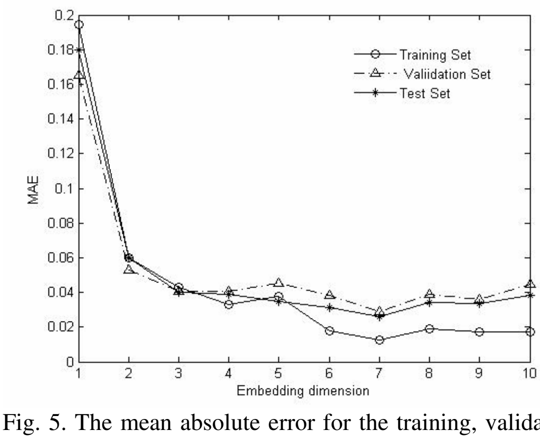


Fig. 5. The mean absolute error for the training, validation and test set of Mackey-Glass data with dimension 7.

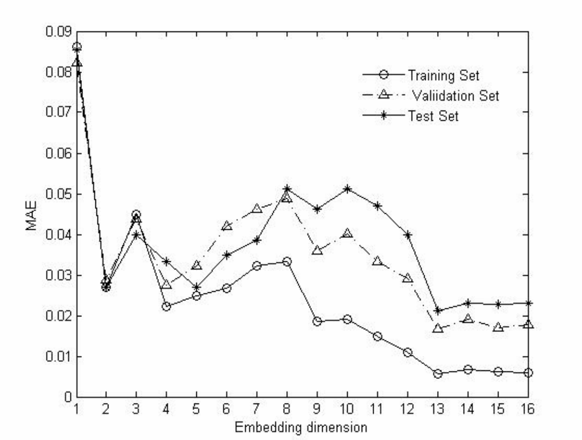


Fig. 6. The mean absolute error for the training, validation and test set of Mackey-Glass data with dimension 13.

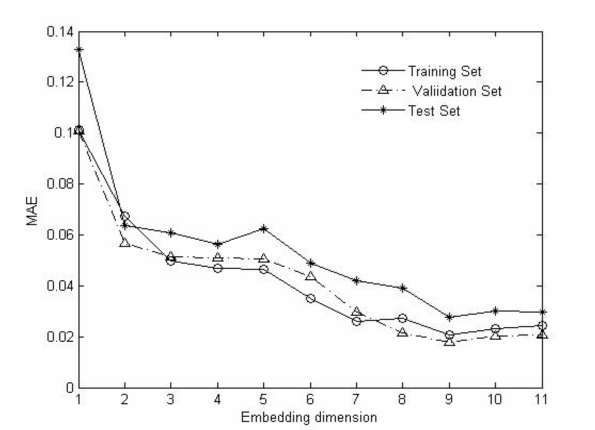


Fig. 7. The mean absolute error for the training, validation and test set of $D_1$ data set.

## 5 conclusion

In a typical chaotic time series analysis, all we can observe is a set of scalar measurements $\{y(t)\}_{t=1}^{N}$ taken from the system which can be low or high dimensional. Estimation of an appropriate embedding dimension for phase space reconstruction of the underlying dynamical system from the observed chaotic time series is greatly considered for the purpose of its understanding and correct prediction of future values. It has been shown in this work that distributed time delay neural network as a dynamic neural network can be successfully used to determine optimal embedding dimension. The prediction error approach used herein has shown to have considerable advantages in terms of quality of the result especially for high dimensional systems.

*References:*

[1] F. Takens, Detecting strange attractors in turbulence, in: *lecture notes in mathematics*, Vol.898, 1981, pp.366-381.

[2] A.I. Mees, P.E. Rapp, L.S. Jennings, Singular value decomposition and embedding dimension. *Phys, Rev.A*, Vol.36, No.1, 1987,pp.340-346.

[3]A.M. Fraser, Reconstructing attractors from scalar time series: A comparison of singular system and redundancy criteria, *Physica D* Vol.34, 1989,pp.391-404.

[4] M. Kennel, R. Brown, H. Abarbanel, Determining embedding dimension for phase space reconstruction using a geometrical construction,*Phys. Rev. A*,Vol.45,1992, pp.3404-3411.

[5] H. Lu, On stability of nonlinear continuous-time neural networks with delays, *Neural Network* Vol.13, No. 10, 2000, pp.1135-1143.

[6] Y.J. Cao and Q.H. Wu, A note on stability of analog neural network with time delays, *IEEE Trans.Neural Network 7*, 1996, pp.1533-1535.

[7] J. Peng and H. Qiao, A new approach to stability of neural networks with time-varing delays, *Neural Network*, Vol.13, No.10, 2002, pp.95-103.

[8] N.H. Packard, J.P. Crutchfield, J.D. Farmer, R.S. Shaw, Geometry from a time series,*Phys. Rev. Lett*, Vol.45, No.9,1980,pp.712-716.

[9] H. Abarbanel, *Analysis of observed chaotic data*, Springer, 1996.

[10]A.A Tsonis, *Chaos From Theory to Applications*, plenum press, New York, 1992 .

[11] A.M. Fraser and H.L .Swinney, Independent coordinates for strange attractors from mutual information,*Physical Review A*,Vol.33,No.2, 1986, pp. 1134-1140.

[12] R.P. Lippmann, An introduction to computing with neural nets, *IEEE ASSP Mag*, Vol.4, Apr.1987.

[13] A. Waibel and T. Hanzawa, Phoneme recognition using time-delay neural network *IEEE Trans on ASS*, Vol. 37, No.3, 1989 pp.328-339.

[14] A. Porporato, L. Ridolfi, Clues to the existence of deterministic chaos in river flow, *Int.J.Mod.Phy.B*, Vol.10 No.15, 1996 pp.1821-1862.

[15] M.T. Hagan and M. Menhaj, Training feed-forward networks with the Marquardt algorithm, *IEEE Trans on Neural Networks*, Vol.5, No.6, 1999, pp.989-993.

[16]The Santa Fe Time Series Competition Data (URL:http://www.psychstanford.edu/~andreas/Time Series/SantaFe.html).

[17] H. Kantz and T. Scheriber, *Nonlinear Time Series Analysis,* Cambridge University Press, 1997.