

# Distance Prediction for Commercial Serial Crime Cases Using Time Delay Neural Networks

Anahita Ghazvini

Faculty of Information Science and Technology  
Universiti Kebangsaan Malaysia (UKM)  
43600 Bangi, Selangor, MALAYSIA

Siti Norul Huda Sheikh Abdullah

Faculty of Information Science and Technology  
Universiti Kebangsaan Malaysia (UKM)  
43600 Bangi, Selangor, MALAYSIA

Mohammed Ariff Abdullah

Faculty of Information Science and Technology  
Universiti Kebangsaan Malaysia (UKM)  
43600 Bangi, Selangor, MALAYSIA

Md Nawawi Junoh

upt. at Royal Malaysia Police (RMP), Dept. of  
Crime Prevention & Community Safety  
Royal Malaysia Police Headquarters, 50560  
Bukit Aman, Kuala Lumpur, MALAYSIA

Zainal Abidin bin Kasim

SAC at Royal Malaysia Police,  
Inspector General Secretariat, R&D  
Royal Malaysia Police Headquarters, 50560  
Bukit Aman, Kuala Lumpur, MALAYSIA

**Abstract**—The prediction of the next serial criminal time is important in the field of criminology for preventing the recurring actions of serial criminals. In the associated dynamic systems, one of the main sources of instability and poor performances is the time delay, which is commonly predicted based on nonlinear methods. The aim of this study is to introduce a dynamic neural network model by using nonlinear autoregressive time series with exogenous (external) input (NARX) and Back Propagation Through Time (BPTT), which is verified intensively with MATLAB to predict and model the crime times for the next distance of serial cases. Recurrent neural networks have been extensively used for modeling of nonlinear dynamic systems. There are different types of recurrent neural networks such as Time Delay Neural Networks (TDNN), layer recurrent networks, NARX, and BPTT. The NARX model for the two cases of input-output modeling of dynamic systems and time series prediction draw more attention. In this study, a comparison of two models of NARX and BPTT used for the prediction of the next serial criminal time illustrates that the NARX model exhibits better performance for the prediction of serial cases than the BPTT model. Our future work aims to improve the NARX model by combining objective functions.

**Keywords**—Criminology and Computational Criminology; Neural Network; modeling; NARX; BPTT; Quantum GIS

## I. INTRODUCTION

Crime is a long-standing concern of everyday people, governments, and researchers, and an education on crime and criminals has become a requirement for many careers. Criminology, the study of crime, is a complicated field that examines and combines all the connections and disagreements of law, social science and psychology to solve criminal cases. The implementation of criminal behavioral analysis on a

database of behavioral patterns can be used to recognize crime patterns; however, a combined modeling technique can be applied to lead to a viewpoint that can state the criminal behavior analysis in a precise, official and computer-supported manner. Computational Criminology uses the power of computers to recognize different types of crimes, such as patterns and developing patterns, crime producers and attractors, and cybercrime. For development of algorithms, Computational topology and dynamic information systems analysis are used. Crime does not appear randomly; it follows rules that are similar to those in various types of non-criminal behavior. Thus, improvement of the understanding of these rules for the development of suitable algorithms for identifying risky parts is the continuing focus of computational criminology [1]. According to computational criminology research, the next crime time can be predicted based on previous data that has been interconnected to serial criminals. Thus, in this study, an appropriate improved method for the future values of time series is selected to create modeling distance crime predictions for serial cases using a time delay neural network, which could help the police to reduce their workload in the investigation of serial criminals.

A neural network is a structure that represents the biological actions of neurons, and it can integrate with a learning or training process such that the known data can be used to alter the coefficients of the neural network. The neural network can function as a way to present the actual output for all of the known inputs in the case of a successful training process. If the units' connections to one another and their weights are set correctly, artificial neural network can perform a particular task. The connections show one unit can influence another unit and the strength of this influence is

being identified by weight. In many fields of science and engineering, artificial neural networks are widely used for extensive classification and regression tasks [2]. The time delay in process control can destabilize or degrade the performance of the system. The existent control action component cannot alter the current output unless the time delay elapsed. Therefore, for large time delay the predictor is used. To design the predictor, identification of the time delay parameter that has already been used by the neural network is crucial. Static neural networks are widely used to process modeling and control because of their approximation capabilities, but they usually cannot give satisfactory results in the case of dynamical or time delay systems [3], [4].

Quantitative forecasting methods, based on mathematical modeling, are classified into two models: causal and time series. A linear relationship is used in a casual model based on predicted and independent variables. The time series models have been categorized into two models, namely linear and nonlinear, and are used for the estimation of historical data. The linear models are mainly statistical models; however, the nonlinear models are comprised of both statistical and non-statistical models. Artificial Neural Networks (ANN) and Support Vector Regression (SVR) are worthy instances of nonlinear models. Crime forecasting is one of the principal areas of criminology. Regression and econometric models can explain the fundamental connection of crime variables and other descriptive variable factors to forecast the rate of crime. Time series models are used as an alternative method to forecast crime rates, as some theoretical rules are needed to describe the connectivity of descriptive variables. The normal time series models need a considerable number of observations, but, due to the lack of information in the crime area, these models are not considered appropriate to forecast the rate of crime. Thus, a novel model that outfits a small dataset is required to improve the performance of crime rate forecasting. Time series models are divided into two groups, linear and non-linear models, in which linear and nonlinear relationships can be modeled, respectively.

The effects of three primary factors on the forecasting of nonlinear time series have been investigated by computer simulation-based tests, and the findings showed that neural networks are suitable for both modeling and forecasting of nonlinear time series, whereas traditional linear methods are not [5]. Artificial neural network (ANN) can be simulated via MATLAB to predict maximum and minimum temperatures (two important weather parameters). Based on findings of this research, multilayer neural network predicts weather as a powerful method that correctly provides mapping between input and output by using historical data [6].

Geographical information systems (GIS) can be used to interpret physical components in the real world and their arrangements is shown, operated and analyzed. The two types of data models that can be used by GIS for the interpretation of information are vector and raster information. The basic components of spatial data, such as points, lines, and polygons, can be signified by the vector data, whereas raster

data are images, i.e., a grid formed by pixels, and can be a photograph provided by aerial means, satellite or remote sensing. Fig.1 (a) and Fig.1 (b) shows a Malaysian map of crime hot spots.

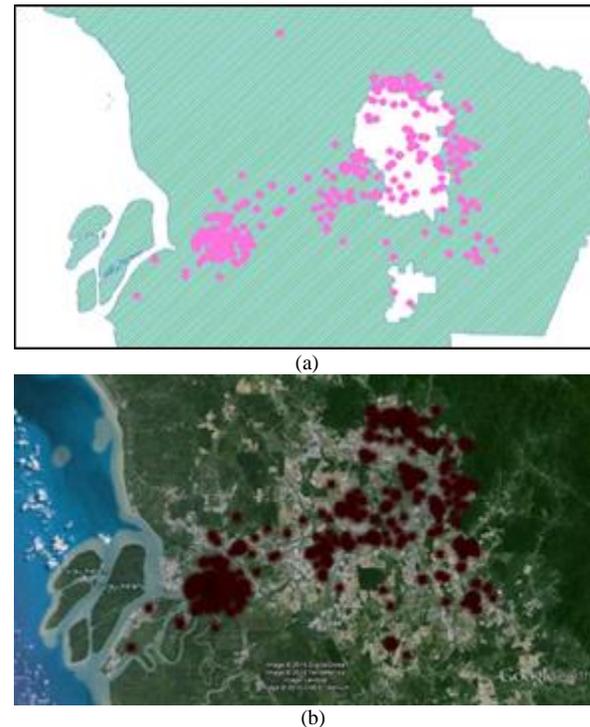


Fig. 1. Malaysian Map for serial crime hotspots in Selangor and Federal District

A raster-type image can be larger or smaller depending on its resolution. GIS analysis is improved by using these two data models. Collection, editing and validation of data can be carried out via spatial analysis to assess the Relations and configuration between them, as well as those with the space surrounding them [7]. A massive number of studies, including criminology ones, used Geographical Information Systems (GIS) and considered a significant development of GIS to have been observed for the analysis of crime, police precinct repositioning, and schemes for the decrease of crime. When police records become computerized, this material can be useful for the recognition of crime and intelligence analyses. [8]. Predictive models, which are geo-statistics tools, can predict the occurrence of incidents by allowance of simple descriptions of phenomena.

This study includes Dynamic Neural Networks using nonlinear autoregressive time series with exogenous (external) input (NARX) and Back Propagation Through Time (BPTT), which is verified intensively with MATLAB to predict and model crime time for serial cases.

The present study is structured as follows: a literature review in Section II, materials and methods in Section III, experimental results in Section IV, the conclusion in Section V, And, finally, discussion and future work in Section VI.

## II. LITERATURE REVIEW

### A. Neural Network Modelling

Neural Networks are an influential data modeling method that can signify complex input/output relationships to learn to identify patterns in any data, which is a division of the field recognized as "Artificial Intelligence." The primary advantage of Neural Networks is to solve complex problems for conventional technologies that cannot be solved by algorithm, or for which an algorithmic solution is too difficult to define for it. Neural networks are able to solve problems that can be easily solved by people but not by computers, such as pattern recognition and forecasting that the tendency of the data are essential.

Artificial Neural Network implementations have three different neurons, input, hidden and output. Neural networks can use linear and nonlinear relationships directly from the data that is modeled, but, for modeled data with nonlinear characteristics, the traditional linear models are not appropriate. The multi-layer perceptron (MLP), which is known as a superintended network because it requires a specific output to learn, uses the historical data to map the input and output correctly to produce the output when the anticipated output is unknown. An MLP contains nodes in various layers in a fixed graph, for which each layer is totally connected to the following layer. A superintended learning method, error back propagation, was used to train the network [9], [10]. The multilayer perceptron (MLP) mainly uses a feed-forward topology that has an architecture of a layered type with entirely interrelated layers. The back propagation through time (BPTT), a feed-forward time-delay network that is used instead in the case of time-related problems, mostly uses the back propagation variant in the case of recurrent topologies, which is a suitable algorithm for recurrent topologies. The design of time delay networks is similar to MLP design, which is a particular topology of back propagation algorithms.

### B. Time Series Neural Network

The precision of pure time delay measurements by traditional methods is not accurate or reliable. For a linear system, the weight can be distributed in place of sampling to place the correct value of the time delay by the neural network, but identification of the time delay of the nonlinear system still remains a difficult problem [3]. A new branch of computing systems is inspired by the human brain as a biological structure for the identification of parameters to provide a new identification method for the time delay. The nonlinear system problems can be identified for theorem of approximation by NN methods due to their capabilities for process modeling and control [11].

For system modeling and forecasting, a recurrent neural network with internal feedback was used [12], [13]. The weakness of static networks can be overcome with dynamic recurrent networks, which have boundaries to recognize the dynamical systems. The nonlinear dynamics method and chaos theory, which are often related with continuous differential equations (but not for discrete variables), are, together, one of the effective techniques for the prediction of

difficult natural procedures and variables. The nonlinear dynamics and chaos theory are data driven methods, as they are totally based on the observation data analysis [14].

## III. MATERIALS AND METHODOLOGY

Dataset descriptions and attribute information are followed by pre-processing on the attributes of datasets, pseudo code, pre-processing steps, the NARX method and the BPTT Neural Network.

### A. Dataset Descriptions

In this study, two types of datasets were used: (1) Raw Commercial Criminal (SCC) datasets and (2) Crime Plotting (CP) datasets using the SPBS system [15]. SCC provides information about 'Suspect & Capture,' while CP shares information about Longitude and Latitude for each matching criminal case from SCC. We obtained ethical approval and used SCC datasets from "PDRM" for the year of 2013 covering only Selangor and Wilayah Persekutuan states to map with CP datasets obtained from "PDRM." The "Suspect & Capture" information is one of the large datasets, and we combined its plotting from CP datasets to predict the distance for commercial serial criminals. The descriptions of the "Suspect & Capture" and "Crime Plotting" datasets are shown in Table 1 and Table 2, respectively.

TABLE I. "SUSPECT & CAPTURE" DATASET DESCRIPTION

<b>Name of Dataset</b>	Suspect & Capture
<b>Data set characteristics</b>	Multivariate
<b>Attribute characteristics</b>	Categorical, Integer
<b>Number of instances</b>	105934
<b>Number of attributes</b>	14
<b>Missing values</b>	Yes

TABLE II. "CRIME PLOTTING" DATASET DESCRIPTION

<b>Name of Dataset</b>	Crime Plotting
<b>Data set characteristics</b>	Multivariate
<b>Attribute characteristics</b>	Categorical, Integer
<b>Number of instances</b>	136407
<b>Number of attributes</b>	14
<b>Missing values</b>	Yes

In this study, the total number of instances for the "Suspect & Capture" and "Crime Plotting" datasets are 105,934 and 136,407, respectively. The "Suspect & Capture" and "Crime Plotting" datasets both contain fourteen attributes; the Name, No ID and NO Case attributes from "Suspect & Capture" and Latitude & Longitude and NO Case from "Crime Plotting" were used, while No of Case in both datasets plays a key role in combining Latitude & Longitude from one dataset with another for use in the prediction of distance crimes for serial cases.

Based on the number of duplicates, the sum of data for repetition of three and above after the combination of the two above-mentioned datasets comes to 8256. The serial criminal cases of repetitions three and four at the hidden layers of 10, 11, 12, 13, 14 and 15 at the default epoch of 1000 after removing the outliers were examined.

B. Merging into the Serial Commercial Crime & Plotting (SCCP) Dataset

The specific information is needed for every attribute instead of the total description of the dataset. Continues values and String values are two main groups of attributes, which are shown in Table 3.

TABLE III. ATTRIBUTE INFORMATION

NO	Attribute
1	Name
2	No ID
3	No Case
4	Latitude
5	Longitude

C. Pseudo code

Pseudo code of the Pre-Processing Steps

Start

- a. **Step 1:** Input attributes.
- b. **Step 2:** Sort the suspect names from the “Suspect & Capture” and “Crime Plotting” datasets in ascending order.
- c. **Step 3:** Sort the numbers of IDs.
- d. **Step 4:** Find the duplicate values from among the suspect names and numbers of IDs.
- e. **Step 5:** Find the frequency of each duplicate.
- f. **Step 6:** Keep the duplicates with frequencies of three or higher.
- g. **Step 7:** Match the “Suspect & Capture” dataset entries to those of “Crime Plotting” by the Number of Cases.
- h. **Step 8:** Find the distance between the serial events of A (Latitude1, Longitude2)-B (Latitude1, Longitude2) by:

$$Distance(i, j) = \sum_{i,j}^{n,m} \cos^{-1}((\sin(\pi - \theta^{La1}) \times \sin(\pi - \theta^{La2}) + \cos(\pi - \theta^{La1}) \times \cos(\pi - \theta^{La2})) \times \cos(\theta^{Lo2} - \theta^{Lo1})) \times R, [16].$$

where i = the number of SCCP criminal records between [1, n], n = the number of records, j = the number of frequencies between [1, m], m = the number of repetitions, La and Lo are the latitude and longitude of two serial criminal events, and R = 6371 Km (the Earth’s radius).

End

D. Pre-Processing Steps

The name and the No ID attributes in the “Suspect & Capture” dataset were sorted to find the duplicate values of the above-mentioned attributes. Table 4 illustrates the duplicate values of the name (Sample Name) and No ID (Sample No ID) attributes.

TABLE IV. DATA SORTING AND OBSERVATION OF DUPLICATE VALUES

Name	No ID
ABCD	123456789
ABCD	123456789
ABCD	123456789
XDFT	908765443

Combination results of two “Suspect & Capture” and “Crime Plotting” datasets by using a primary key (No Case) are illustrated in Table 5.

TABLE V. COMBINATION RESULTS OF “SUSPECT & CAPTURE” AND “CRIME PLOTTING” DATASETS BASED ON A PRIMARY KEY (NO CASE)

Name	No ID	No Case	Latitude	Longitude
ABCD	123456789	Pk188888009	3.106197	101.5467
ABCD	123456789	Pk188888009	3.206071843 2205	101.6068
ABCD	123456789	Pk188888009	3.052487530 6253	101.5377
XDFT	908765443	Lkmp777885 433	3.148755792	101.7566
XDFT	908765443	Lkmp777885 433	3.148755792	101.7566
XDFT	908765443	Lkmp777885 433	3.148755792	101.7566
XDFT	908765443	Lkmp777885 433	3.148755792	101.7566
.....	.....	.....	.....	.....

The distances were calculated based on step eight of the pseudo code of the pre-processing steps for the attributes of Name and No ID with three or more repetitions. Table 6. Illustrates the Distance Values.

TABLE VI. DISTANCE VALUES

Name	Number of ID	No Case	Latitude	Longitude	Distance
ABCD	123456789	Pk188888009	3.106197	101.546686	11.11168
ABCD	123456789	Pk188888009	3.20607184 3	101.606829 8	17.08294
ABCD	123456789	Pk188888009	3.05248753 1	101.537728 7	Will be predicted by best model
XDFT	908765443	Lkmp77788543 3	3.1487558	101.789299	0.071799
XDFT	908765443	Lkmp77788543 3	3.1487568	101.777543	0.127581
XDFT	908765443	Lkmp77788543 3	3.14875679	101.756655	9.49E-05
XDFT	908765443	Lkmp77788543 3	3.14875578	101.756656	Will be predicted by best model
.....	.....	.....	.....	.....	.....

Normalization is mainly suitable for the classification of algorithms concerning neural networks. In this study, because the data are serial and the min and max values of the data in repetitions three and four are unknown, the Z-score normalization method was used. In z-score normalization, the values for the attribute D<sub>v</sub> are normalized based on the mean and standard deviation of D<sub>v</sub>. The data value of D<sub>v</sub> is normalized to D<sub>v</sub>' by the computation of (1):

$$D'_v = \frac{D_v - M}{STN\_D} \tag{1}$$

where M and STN\_D are the mean and the standard deviation, respectively, of the attribute D<sub>v</sub>. This method of normalization is useful when the actual minimum and maximum of the attribute D<sub>v</sub> are unknown.

The calculated values of the distance based on the Latitude and Longitude showed that some values of distances are outliers, which are defined as data points that are statistically inconsistent with the rest of the data.

Therefore, to determine the outliers, the percentiles function in Microsoft Excel was used. By using this statistical method, the percentile function calculated the 30th percentile and the 90th percentile of the data to remove the outliers. Fig.2 and Fig.3 show the data of repetitions of three and four (a) with and (b) without the outliers.

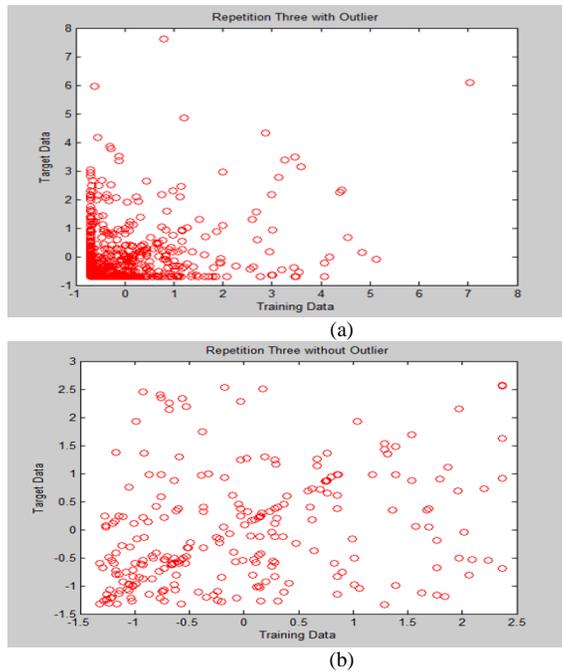


Fig. 2. Data of repetitions of three (a) with and (b) without outliers

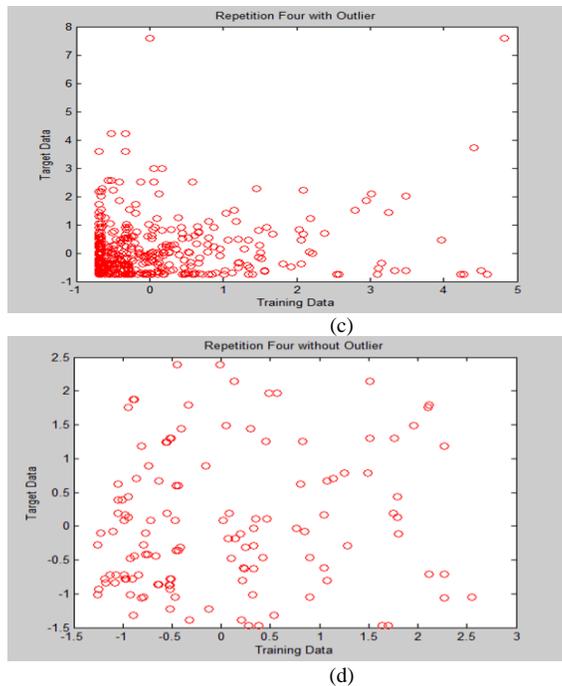


Fig. 3. Data of repetitions of four (c) with and (d) without outliers

**E. NARX Neural Network**

Recurrent NNs have been widely used for the modeling of nonlinear dynamical systems [17], [18]. The NARX model,

which is based on the model of linear ARX, is used to predict the serial criminal time, and, for the classification stage, the back propagation algorithm is used to find the best repetition experimental results. NARX consists of input, hidden, and output layers, where the current and historical inputs and outputs that are provided for the hidden layer are held by the input.

One or more neurons in the hidden layer enable it to produce a nonlinear charting of the affine weighted mixture of the values from the input, whereas the output contains an affine combination of the values from the hidden layer. The dynamical order of inputs and outputs, as well as the number of neurons in each layer of this network, are specified. For determination of a valid training set, some pre- and post-processing should be performed on the input and target values [19]. These procedures record the input and target data in the range of [-1, 1], and normalization of the inputs and targets will be performed to achieve zero mean and variance. Finally, constant inputs and outputs are eliminated for the processing of unknown inputs. (2), which describes the input-output recurrent model of NARX, is as follows:

$$y(t + 1) = F(y(t), \dots, y(t - 1), D(n), \dots, D(t - 1)) \quad (2)$$

This model contains one single input and one single output for repetitions of three and two inputs and one output for repetitions of four. The single input is applied to a tapped delay line memory, whereas the single output is returned to the input through another tapped delay line memory for the purpose of storing the previous values. The input layer of the multilayer perceptron is fed with the content of the two above-mentioned tapped delay line memories, whereas the value of the input of the model is shown as  $D(n)$  and the value of the output of the model is presented as  $y(t+1)$ .

The signal vector provided for the input layer of the multilayer perceptron contains current and past values of the input (exogenous inputs) and delayed values of the output (regressed). Fig.4 shows the NARX model.

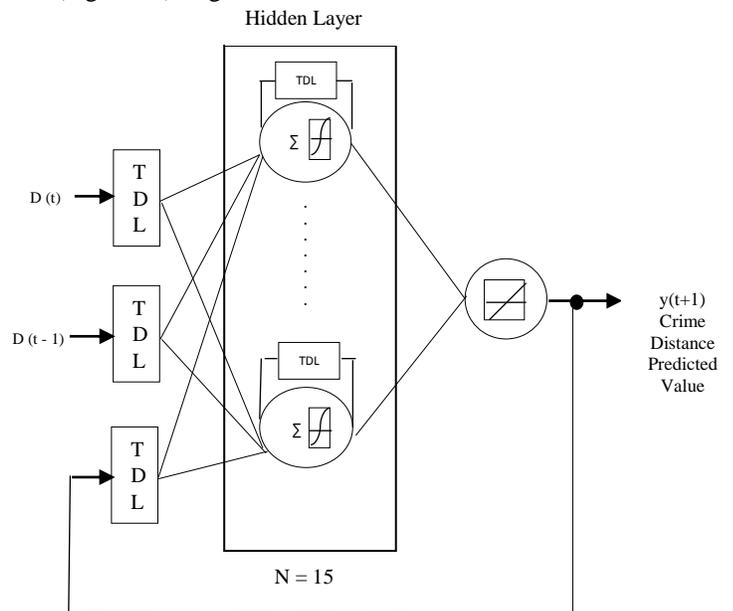


Fig. 4. Nonlinear autoregressive NARX model with exogenous inputs

### F. Back Propagation Through Time Neural Network

The back propagation through time (BPTT) algorithm is an extension of the standard Back propagation algorithm for recurrent neural networks [20]. The first step of the BPTT algorithm contains the unfolding process of the network which involves replicating times of the recurrent network to obtain a corresponding feed-forward network. In this procedure, each replicated connection shares its value in all times, which can be trained by using the Back propagation algorithm. In the forward process of Back propagation, the calculation of output  $y_i$  at time  $t$  of each neuron can be defined in (3) to (7) as follows:

$$Y_{i(t)} = f(D_i(t)) \quad (3)$$

$$D_i(t) = \sum_{j \in H} Y_j(t)W_{ij} + \sum_{j \in i} D_j^{in}(t)W_{ij} + \sum_{j \in M} y_j(t - T_{ij})W_{ij} \quad (4)$$

Where  $f$  = the neuron activation function,  $H$  = the hidden layer indexes,  $i$  = the input neuron indexes, and  $D_j^{in}$  is the  $j^{\text{th}}$  input neuron; however,  $M$  indicates the indexes of neurons that supply information regarding the previous network stages, and  $T_{ij}$  is an integer value representing the displacement in recurrent connections through time. Regarding the error of the back propagation process, each neuron has an error size, and, for the output layer, this error value is considered as the variation between the predictable and the obtained value (4), while, for the hidden layers, it can be gained by taking into account the error in the successor layers (5).

$$\delta_j(t) = (d_j - y_j)y_j(1 - y_j) \quad (5)$$

$$\delta_j(t) = y_j(1 - y_j)\sum_{i \in suc(j)} W_{ij}\delta_i \quad (6)$$

The weights are calculated with regard to (6), where the necessary change in weights is computed based on the learning rate  $\alpha$ :

$$\Delta W_{ij}^{e+1} = \alpha\delta_j y \quad (7)$$

where  $e$  = the order of the weight updates in the learning process. The second phase of the BPTT algorithm continues to process (the folding process) when the back propagation algorithm is used to obtain the FFN. The original recurrent network is obtained in this stage, where the connection weights are the combination of the corresponding connections on each fold. "Fig.4" shows the BPTT with its Architecture Graph of two neurons, and the Single flow graph of network 15 unfolds in time where  $D$  is the actual distances value. Fig.5 shows the BPTT model.

### G. Evaluation Metrics

There are various statistical metrics to validate the performance of the models, but a more commonly used benchmark metric is to use the regression coefficient (R) and mean square error (MSE), which are used in this paper. The R criterion is given by (8) as follows:

$$R = \frac{R_{yt}}{std(y)std(t)} \quad (8)$$

where  $y$  and  $t$  are the predicted and actual distance of the serial criminal,  $R_{yt}$  = the covariance between  $y$  and  $t$ , and  $std(y)$  and  $std(t)$  represent the standard deviation of  $y$  and  $t$ , respectively. The value of the regression coefficient varies between 0 and 1.

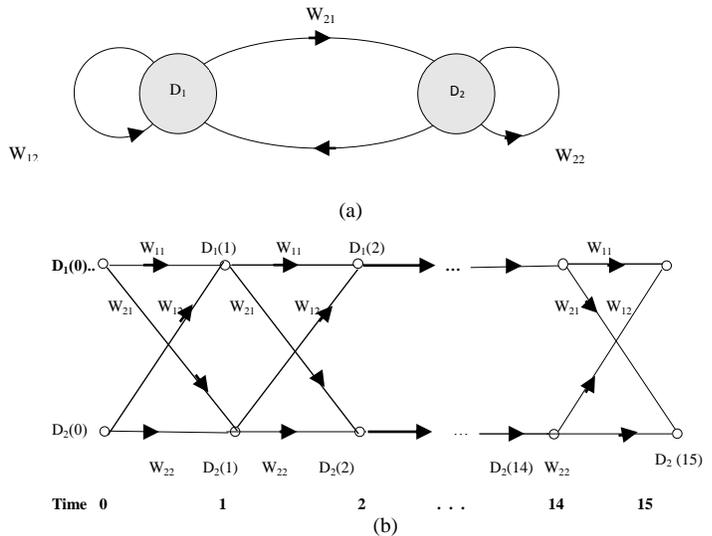


Fig. 5. Back Propagation Through Time model (a) Architecture Graph of two neurons and (b) Single flow graph of network 15 unfolded in time

The value being near to 0 indicates that the output is deviating and near to 1 that the output matches the target.

The mean square error (MSE) criterion is widely used to find the difference between the values gained by the given model and the actual values that are expected. As the MSE is less, the model is more accurate and reliable.

The MSE criterion is given by (9) as follows:

$$MSE = \frac{1}{N} \sum_{i=1}^N (t_i - y_i)^2 \quad (9)$$

where  $N$  is the number of predicted sample.

## IV. EXPERIMENTAL RESULTS

This research aims to solve the problem of next crime time prediction for serial cases based on time delay neural networks. To predict the time series data, the NARX and BPTT methods were used to create a model for Distance Crime Prediction for serial cases using Time Delay Neural Networks. The NARX and BPTT models are examined and compared for each dataset for repetitions of three and four with 558 and 198 instances (after removing the outliers with the percentile method) of time series criminal data, respectively. The performances of serial criminal cases of repetitions of three and four were examined at the default epoch of 1000. The NARX network is a feed-forward network, for which its hidden layer contains a tan-sigmoid transfer function and its output layer contains a linear transfer function as its default. NARX has two inputs, an external input and a feedback connection from the network output that can be closed after training of the network. There is a tapped delay line to stock previous values of each of the above mentioned inputs to allocate the network architecture for a NARX network, and it must choose the delays related with each tapped delay line. The input delays and the feedback delays can be set in a range from 1 to 4, and, as the delay number increases, more calculation is required and over-fitting of the data can occur. However, an overly high delay number can allow the network to solve complex problems. In this study, the delay number is set to 2, and the number of hidden

layers of 10, 11, 12, 13, 14 and 15 were chosen to examine the performance of the NARX and BPTT models.

This study is considered a one-step-ahead prediction for the performance of all the experiments. Tables 7 and 8 show the performance measures of the NARX model, and Tables 9 and 10 present the performance of the BPTT model with various hidden layer selections for repetitions of three and four. It was decided to vary the hidden layers as it has been proven that, a single hidden layer with sufficient neurons in a neural network will give the desired accuracy for any nonlinear function [21], [22], [23]. Fig.6 and Fig.7 show comparisons of the accuracy versus hidden layers for the NARX and BPTT Models for repetitions of three and four.

TABLE VII. PERFORMANCE METRIC COMPARATIVE RESULTS FOR REPETITION OF THREE WITH THE NARX MODEL

Hidden Layer	R	MSE	RMSE	R <sup>2</sup>
10	0.4334	0.8192	0.905	0.1878
11	0.4186	0.8479	0.9208	0.1752
12	0.4272	0.8223	0.9068	0.1825
13	0.4227	0.8238	0.9086	0.1787
14	0.4008	0.8532	0.9237	0.1606
15	<b>0.4570</b>	<b>0.8149</b>	<b>0.9027</b>	<b>0.2088</b>

TABLE VIII. PERFORMANCE METRIC COMPARATIVE RESULTS FOR REPETITION OF FOUR WITH THE NARX MODEL

Hidden Layer	R	MSE	RMSE	R <sup>2</sup>
10	0.4571	0.8598	0.9273	0.2089
11	<b>0.7441</b>	<b>0.5154</b>	<b>0.7179</b>	<b>0.5537</b>
12	0.7420	0.5292	0.7274	0.5506
13	0.6344	0.6114	0.7819	0.4025
14	0.5013	0.7494	0.8657	0.2513
15	0.7340	0.5701	0.7551	0.5388

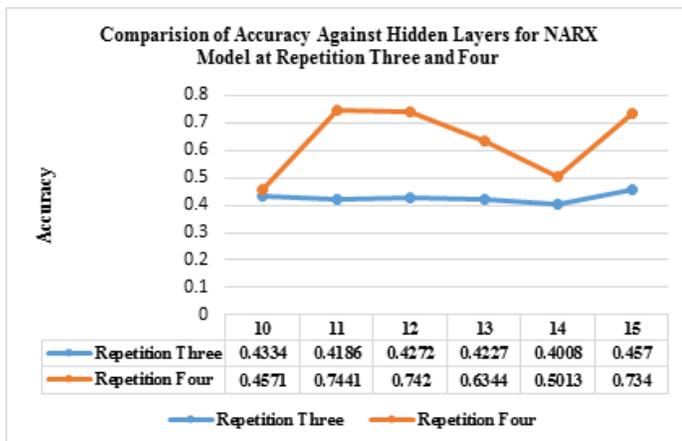


Fig. 6. Comparison of the accuracy versus the hidden layers for the NARX model with repetitions of three and four

Tables 9 and 10 present the NARX model performance measurements with different hidden layers for repetitions of three and four. The observation of results showed that the NARX model with 15 hidden layers for repetition of three and 11 hidden layers for repetition of four yields good accuracy for the prediction with regard to lower error measures (0.8149 and 0.5154) and a high accuracy coefficient (0.4570 and 0.7441) for repetitions of three and four, comparatively.

TABLE IX. PERFORMANCE METRIC COMPARATIVE RESULTS FOR REPETITION OF THREE WITH THE BPTT MODEL

Hidden Layer	R	MSE	RMSE	R <sup>2</sup>
10	<b>0.4462</b>	<b>0.8151</b>	<b>0.9029</b>	<b>0.1991</b>
11	0.4188	0.823	0.9072	0.1754
12	0.4108	0.8504	0.9222	0.1688
13	0.4057	0.8512	0.9387	0.1646
14	0.4398	0.8196	0.9053	0.1934
15	0.4340	0.8240	0.9078	0.1884

TABLE X. PERFORMANCE METRIC COMPARATIVE RESULTS FOR REPETITION OF FOUR WITH THE BPTT MODEL

Hidden Layer	R	MSE	RMSE	R <sup>2</sup>
10	0.4533	0.8867	0.9416	0.2055
11	0.4853	0.8165	0.9036	0.2355
12	0.5042	0.783	0.8849	0.2542
13	0.4347	0.9019	0.9497	0.1890
14	0.5600	0.7378	0.8595	0.3136
15	<b>0.5604</b>	<b>0.7005</b>	<b>0.8370</b>	<b>0.3140</b>

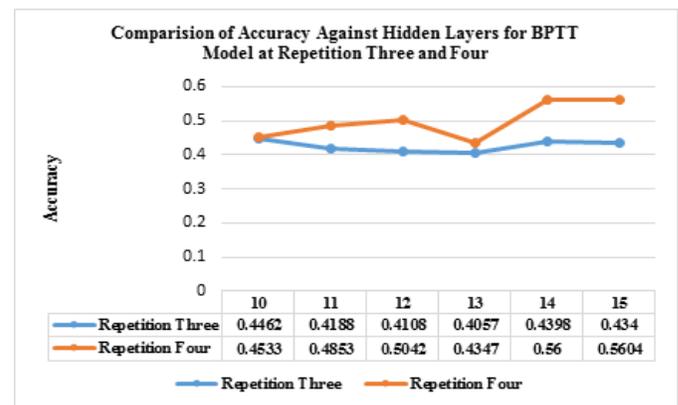


Fig. 7. Comparison of the accuracy versus the hidden layers for the BPTT model with repetitions of three and four

Tables 9 and 10 present the BPTT model performance measurements with different numbers of hidden layers for repetitions of three and four. The observation of results showed that the BPTT model with 10 hidden layers for repetitions of three and 15 hidden layers for repetitions of four yields good accuracy for prediction with regard to lower error measures (0.8151 and 0.7005) and a high regression coefficient (0.4462 and 0.5604) for repetitions of three and four, comparatively.

The comparison of the NARX and BPTT models indicates that the NARX model yields more accurate results than the BPTT model regarding the mean and standard deviation; thus, to evaluate the performance of the NARX and BPTT models, two repetitions were used for the training, validation and testing phases of both the NARX and BPTT models; the performance metrics with the mean and STDV are presented in Table 11. Fig.8 shows a comparison of the accuracy values of the NARX and BPTT models.

TABLE XI. PERFORMANCE COMPARISON FOR THE NARX AND BPTT MODELS

Repetition n	NARX MODEL				BPTT MODEL			
	R	MSE	RMS E	R <sup>2</sup>	R	MSE	RMS E	R <sup>2</sup>
3	0.4570	0.8149	0.9027	0.2089	0.4462	0.8151	0.9029	0.1991
4	0.7441	0.5154	0.7179	0.5537	0.5604	0.7005	0.8370	0.3140
Mean	0.6006				0.5033			
STDV	0.2030				0.0808			

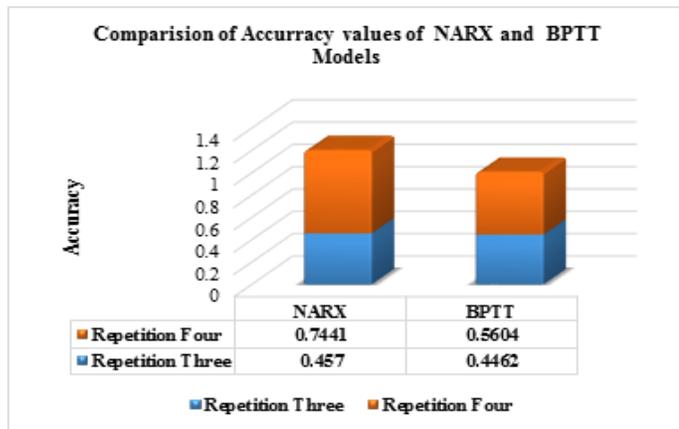


Fig. 8. Comparison of the Accuracy values of the NARX and BPTT Models

H. Quantum Geographical Information System (QGIS)

In this study, the actual and predicted values have been demonstrated by Google Earth. From the results of this study, the NARX model showed better performance, so the predicted value from this model was used. To plot the differences of distance between the actual and predicted values for repetitions of three and four, one case of a repetition of three and one case of four were selected randomly. The repetitions of three and four contain three and four distance values, respectively; the first two distance values in the case of the repetition of three and the first three for the repetition of four were calculated based on the formula mentioned in step eight of the pseudo code from the pre-processing steps, and then the third and fourth distance values of the repetitions of three and four are predicted by the NARX model. To present the

differences between the first two distance ( $D_1, D_2$ ) values with the predicted distance ( $D_{P3}$ ) value in the repetition of three and the first three distance ( $D_1, D_2, D_3$ ) values with the predicted distance ( $D_{P4}$ ) value in the repetition of four, buffering was created around the latitude and longitude of each serial event for the repetitions of three and four by a Quantum Geographical Information System (QGIS). The QGIS was used for the buffering of each point that was calculated from the distance values for the repetitions of three and four to be used by Google Earth to show the exact area of each serial event and the differences of the actual from the predicted distance to prevent the next serial event. Fig.9 shows buffering for the repetitions of three and four to present the differences of distance between the actual and predicted values by the QGIS, and Fig.10 shows the differences between the first two distance values with the predicted distance value for (a) the repetition of three and (b) the repetition of four serial criminal events at Selangor and Wilayah Persekutuan by Google Earth.

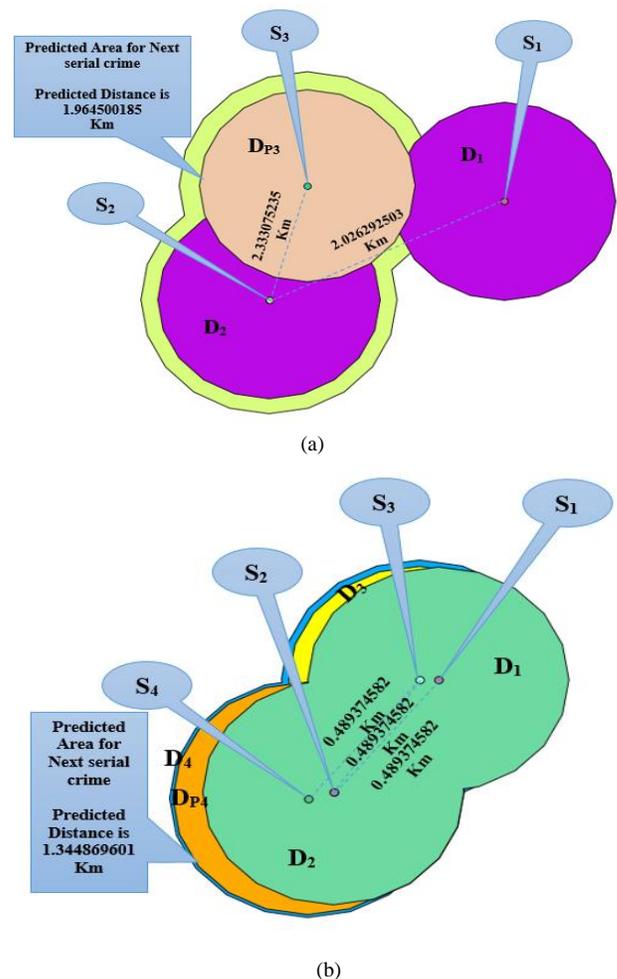


Fig. 9. Differences between the first two distance values with the predicted distance value for (a) the repetition of three and (b) the repetition of four serial criminal event at Selangor and Wilayah Persekutuan by the QGIS



First distance ( $D_1$ ) between  $S_1$  and  $S_2$  of the serial events for the repetition of three at a zoom of 11.29 mi  
(a)



First distance ( $D_1$ ) between  $S_1$  and  $S_2$  of the serial events for the repetition of four at a zoom of 17101 ft  
(b)



Second distance ( $D_2$ ) between  $S_2$  and  $S_3$  of the serial events for the repetition of three at a zoom of 11.29 mi  
(a)



Second distance ( $D_2$ ) between  $S_2$  and  $S_3$  of the serial events for the repetition of four at a zoom of 17101 ft  
(b)



Third distance ( $D_3$ ) between  $S_3$  and  $S_{P4}$  of the serial events for the repetition of three at a zoom of 11.29 mi  
(a)



Third distance ( $D_2$ ) between  $S_3$  and  $S_4$  of the serial events for the repetition of four at a zoom of 17101 ft  
(b)



Third distance ( $D_2$ ) between  $S_3$  and  $S_{P4}$  of the serial events for the repetition of four at a zoom of 17101 ft  
(b)

Fig. 10. Differences between the first two distance values with the predicted distance value for (a) the repetition of three and (b) the repetition of four serial criminal events at Selangor and Wilayah Persekutuan

### V. CONCLUSION

For prediction of next crime distances NARX and BPTT models were chosen due to their capabilities. In this research, the two models were tested for modeling distance crime prediction of serial cases by using time delay neural networks for repetitions of three and four serial cases after removing the outliers. Serial criminal cases of repetitions three and four were examined with hidden layers of 10, 11, 12, 13, 14 and 15 for both models. The best performances of NARX is observed with hidden layers of 15 for repetition three and 11 for repetition four, while this is at hidden layers of 10 and 15 for BPTT model. The results of this study have illustrated that the accuracy of the NARX model for repetitions three and four are 0.4570 and 0.7441, respectively, whereas, for the BPTT model, they are 0.4462 and 0.5604, respectively. Based on mean and standard deviation measurements, the results stated that NARX model has better performance than the BPTT model.

### VI. DISCUSSION AND FUTURE WORK

NARX is a well-known model for dynamic systems, not only for being an accurate model but also for its simplicity of training. The use of the Lavenberg algorithm to create the NARX model has some advantages and disadvantages. As a disadvantage, there are various complications in the training process in NARX relating to weights or the number of connections, which can cause overtraining of the data and the wrong fit, which then cannot yield a better prediction. According to Levenberg-Marquardt optimization, modification of the training functions of NARX can update the weight and bias, including a regularization technique to minimize the squared errors and weights to present a correct combination for a network that is well generalized. Additionally, for a network with a few hundred weights, faster convergence occurs to solve the problem of function approximation, and this advantage in the case of an accurate training requirement is more obvious in the NARX model. Regarding the disadvantages of this network, it can be noted that, as the weights in the network rise, the advantage of the Levenberg algorithm will be reduced [24]. BPTT is, in fact, a back propagation algorithm with memory of the previous

outputs, which is considerably faster for the training of recurrent neural networks than general purpose optimization techniques. A low speed of convergence and the possibility of becoming trapped in local minima of the error surface are disadvantages of BPTT [25].

The results illustrate that the NARX network shows better results for distance crime time prediction than Back Propagation Through Time (BPTT). Additionally, using the Quantum Geographical Information System (QGIS) can show the difference between the predicted values of the next crime distance and our actual values in the NARX model and give a better understanding of the entire concept. Because the NARX approach is a more suitable model for the prediction of the next crime distance, in our future work, the NARX model will be combined with objective functions that are directly minimized during training to improve the NARX model by getting more accurate results of prediction with a lower mean square error, causing the predicted value of the next crime distance to be much closer to the actual value. As a function of the weights, the total error or the average error for the training set is usually minimized by the NN training performance, which can lead to overfitting and poor generalization when the number of training cases is small in comparison to the complexity of the network, so regularization (a function of the weights) can be used as a method to improve generalization in a way that minimizes an objective function by adding the total error function and a regularization function.

### ACKNOWLEDGEMENT

This research is based on two fundamental research grants from Exploration Research Grant (ERGS) “ERGS/1/2013/ICT02/UKM/02/4 from Ministry of Education, entitled “GEO-TEMPORAL CRIME NAVIGATION BASED ON MULTI-OBJECTIVE TIME DELAY NEURAL NETWORK”. I also would like to thank previous the CAIT researchers such as Associate Prof. Dr. Siti Norul Huda Sheikh Abdullah, PHD Candidate Mohd Ariff bin Abdullah, Datuk’ Md Nawawi Junoh. Supt police, Dato’ Zainal Abidin bin Kasim.

### REFERENCES

- [1] P. L. Brantingham, and S. Fraser, “Computational Criminology”, European Intelligence and Security Informatics Conference, pp. 3-5, IEEE Computer Society, DOI: 10.1109/EISIC.2011.79, 2011.
- [2] S. Haykin, “A Comprehensive Foundation”, 2<sup>nd</sup> edition, Prentice Hall PTR Upper Saddle River, NJ, USA, 1999.
- [3] A. Yazdizadeh, K. Khorasani, and R. V. Patel “Identification of a Two-Link Flexible Manipulator Using Adaptive Time Delay Neural Network” in IEEE Trans. on System, Man and Cybernetics Part B: Cybernetics. Vol.30, No. 1, pp. 165-172, 2000.
- [4] H. Karlene, A. Sinzinger, D. Eric, and M. J., “Improvements in the Predictive Capability of Neural Networks” In Journal of Process Control, Vol. 12, pp. 193-202, 2002.
- [5] G. P. Zhang, B. E. Patuwo, M. Y. Hu, “A simulation study of artificial neural networks for nonlinear time-Research 28, 381-396, 2001.
- [6] N. Kumar, G. Kumar Jha, “A Time Series ANN Approach for Weather Forecasting” in International Journal Of Control Theory and Computer Modeling (IJCTCM) Vol.3, No.1, January 2013.
- [7] J.H., Ratcliffe, “Strategic Thinking in Criminal Intelligence”, Sydney: Federation Press.2004.

- [8] J. Ferreira, P. João, and J. Martins, "GIS for Crime Analysis: Geography for Predictive Models" in *The Electronic Journal Information Systems Evaluation* Volume 15 Issue 1, pp36-49, 2012.
- [9] Rosenblatt, Frank. X., "Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms" in Spartan Books, Washington DC, 1961.
- [10] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Error Propagation", D. E. Rumelhart, J. L. McClelland, and the PDP research group. (Editors), *Parallel distributed Processing: Explorations in the microstructure of cognition*, Volume 1: Foundations. MIT Press, 1986.
- [11] P. J. Werbos, "the Roots of Backpropagation: From Ordered Derivation to Neural Networks and Political Forecasting, February 1994.
- [12] A. Balestrino, F. B., and A. Landi, "On-line Process estimation by ANNs and Smith Controller design" in *IEE Proc.-Control Theory Appl.*, Vol.145, No.2, 1998.
- [13] K. S. Narendra, and K. Parthasarathy, "Identification and Control of Dynamical Systems Using Neural Networks" in *IEEE Trans. Neural Networks*, Vol. 1, No. 1, pp. 4-27, 1990.
- [14] Z. Hongyu, Z. jian, and C. LiMin, "Elman Network-Based Implicit Generalized Predictive Control of Superheated Steam Temperature", *Control Theory and Applications*, Vol.22, No.7, pp 13-16, 2003.
- [15] Siow Suan Neo, "Safe City Monitoring System", GIS Web Based Applications in Crime Monitoring in Malaysia, Geospatial World Forum, Geneva, Switzerland, May 2014.
- [16] Deriving the haversine formula, Ask Dr. Math, Apr. 20–21, 1999.
- [17] Siek, M. and D. P. Solomatine, "Nonlinear chaotic model for predicting storm surges", *Nonlin. Processes Geophys.* 17, 405–420, 2010.
- [18] A. Waibel, T. Hanazawa, G. Hinton, G., and S. Kiyohiro, "Phoneme Recognition Using Time-Delay Neural Networks", in *IEEE Transactions on Acoustics, Speech, and Processing*, Vol. . 37, NO. 3, March 1989.
- [19] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Prentice Hall PTR, 1998.
- [20] P. J. Werbos, "Backpropagation through Time: What it does and How to do it", *Proceedings of the IEEE*, Vol. 78 NO. 10, pp. 1550–1560, 1990.
- [21] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feed forward network are universal approximators", *Neural Networks*, Vol. 2 NO. 5, pp. 359–366, 1989.
- [22] G. Cybenko, "Approximations by Superpositions of a Sigmoidal Function", *Mathematics of Control, Signals and Systems*, Vol. 2, pp. 303–314, 1989.
- [23] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art", *International Journal of Forecasting*, Vol. 14, pp. 35–62, 1998.
- [24] Diaconescu, E., "The use of narx neural networks to predict chaotic time series, WEEAS Transactions on Computer Research", issue 3, volume 3, March 2008.
- [25] Mastorocostas, P., Varsamis, D., Mastorocostas, C., Rekanos, I., "An Accelerating Learning Algorithm for Block-Diagonal Recurrent Neural Networks", *IEEE Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation, and International Conference Intelligent Agents*, 2005.