A Dynamic Neural Network Method for Time Series Prediction Using the KIII Model

Haizhon Li & Robert Kozma
{hli1, rkozma}@Memphis.edu
Division of Computer Science, University of Memphis
Memphis, TN 38152

***Abstract* – In this paper, the KIII dynamic neural network is introduced and it is applied to the prediction of complex temporal sequences. In our approach, KIII gives a step-by-step prediction of the direction of the currency exchange rate change. Previously, various multiplayer perceptron (MLP) networks and recurrent neural networks have been successfully implemented for this application. Results obtained by KIII compare favorably with other methods.**

I. INTRODUCTION

K models represent a family of models of increasing complexity that describe various aspects of functioning of vertebrate brains [1]. Similar to other neural network models, K models consist of neurons and connections, while its structure and behavior are more biologically plausible. The K model family includes K0, KI, KII, KIII, and KIV [2,3]. The previous works introduce K models from biological experiments and mathematical modeling perspectives. This work shows the KIII model is a promising computational methodology through its application on time series prediction.

Time series prediction takes an existing series of data X(0), …, X(t-1), X(t), and forecasts the future values X(t+1), X(t+2), …, etc. The goal is to model the history data in order to accurately forecast future unknown data values. Due to the high noise level and the non-stationary nature of the data, financial forecasting is a challenging application in time series prediction domain. Various methods are used in this application [4,5,6,7,9,10]. In this work, we use the KIII model to predict the one step direction of daily currency exchange rate. The data we used is from [4]. The experiment results show the classification capacity of the KIII model is comparable or better than alternative recurrent NN-based prediction methods.

This paper is structured as follows. In section II, an introduction on the family of K sets is given. In section III, we elaborate on the methodology used in this work. This is followed by the financial time series prediction experiment in section IV. Finally, a discussion and conclusion are given in section V.

II. A DYNAMICAL NEURAL NETWORK: K MODELS

A. *Overview of the K models*

The architecture of artificial neural networks is inspired by the biological nervous system. It captures the information from data by learning and stores the information among its weights. This computation model is especially good, considering generalization and error tolerance, compared to other symbolic computational models. In a departure from traditional neural networks, the structure and behavior of the K models are closer to the nervous system. Each K set models some biological part of the nervous system. For example, KI models dentate gyrus (DG); KII models olfactory bulb (OB); KIII models the sensory cortex; and KIV models the hemisphere-wide cooperation. [8] Another prominent difference is that the activity of the K model is oscillatory and it may involve aperiodic, instead of converging to a certain fixed point. This dynamic behavior makes K models capable of better generation and high error tolerance, just like brains.

B. *K0 and KI models*

Biological experiments indicate three major types of neurons based on their functionality: excitatory, inhibitory, and modulatory. The K0 set is a single inhibitory KOi or excitatory KOe node. It models the neuron populations with a second order ordinary differential equation (ODE):

$$(a*b)\frac{d^2P(t)}{d^2t}+(a+b)\frac{dP(t)}{dt}+P(t)=F(t). \quad (1)$$

In Eq.1, the time constants are a = 0.22, b = 0.72. Here, a and b are rise and decay constants from biological experiment; P(t) denotes the activation of the node as function of time; F(t) is the summed activation from neighbor nodes.

The K0 set has a weighted input and an asymptotic sigmoid function for the output. The sigmoid function Q(x) is given by the equation:

$$Q(x)=Q_m*\left\{1-e^{\left[-1/Q_m*(e^x-1)\right]}\right\}. \quad (2)$$

In Eq.2, Q_m = 5, is the parameter specifying the slope and maximal asymptote of the curve. This sigmoid function is modeled from experiment on biological neural activation.

Fig.1 shows our sigmoid function curve is asymmetric. For details, see [2].

Coupling either two KOe sets or two KOi sets form a KI set. The nodes in a KIe set interact by mutual excitation, while the nodes in KIi set interact by mutual inhibition. Figure 2 shows two kinds of KI sets.

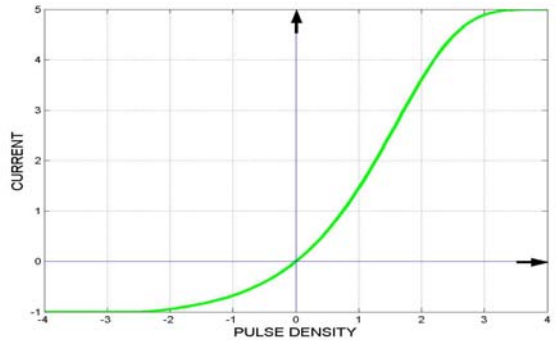

Fig. 1. Asymptotic Sigmoid Function curve.

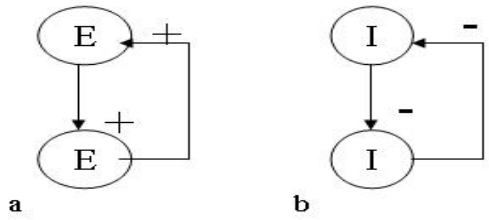

Fig. 2. KI sets: (a) excitatory KI, (b) inhibitory KI.

C. KII model

A KII model is a double layer of excitatory or inhibitory units. There are 4 nodes: two KOe represent two excitatory populations and two KOi represent two inhibitory populations; see Fig. 3.

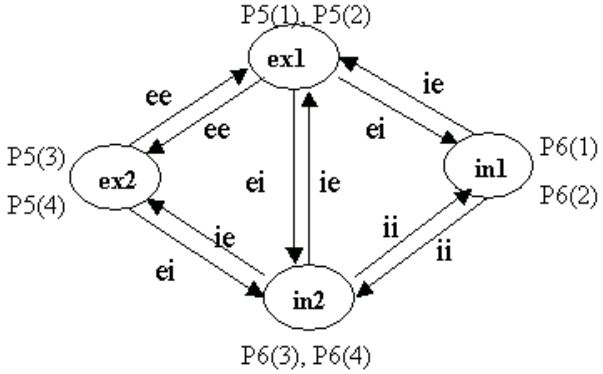

Fig. 3. Schema of the KII set with two excitatory nodes (ex1, ex2), and two inhibitory nodes (in1, in2).

In order to achieve a certain level stability on KII, we conduct experiment to search weight parameters, namely, ee, ii, ei, and ie, so that KII can sustain oscillatory activation in impulse-response tests. When a single pulse of small random input perturbation [-0.3 –1] is added to P5(1) node, all the nodes in KII will sustain oscillatory activations over a long period. Fig.4 shows such an impulse-response of a positive attractor KII when proper weights are set.

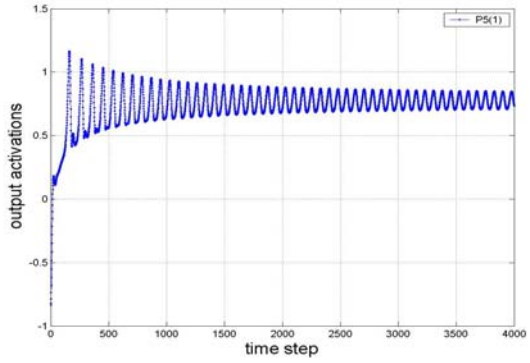

Fig. 4. Impulse-response of a positive attractor KII

KII simulates olfactory bulb (OB). From biological experiments, we know there are three types of KII based on the activation trajectory of excitatory node P5(1): positive attractor KII, zero attractor KII, and negative attractor KII. The selection of weight parameters is based on the following criteria. Firstly, the activation of P5(1) node is either positive, or zero (oscillate within small interval near zero), or negative. Secondly, a large oscillatory activation (limited cycle attractor) is considered better than low oscillatory or fixed-point attractor. In this way, when KIII is composed from KII sets, aperiodic dynamics could be achieved. Also, because of the effective range of our asymmetric sigmoid function, the oscillatory activation of each node should be in [-2 3]. At last, all the weights should be of compatible value in range [0.1 2.0]. With these four criteria, we conduct a global search on four dimension space for ee, ii, ei, and ie. Three sets of KII weight parameters are selected, as in table I.

TABLE I
KII weights parameters

	ei	ie	ee	ii
Zero attractor KII	1.0	2.0	1.8	0.8
Positive attractor KII	1.6	1.5	1.6	2.0
Negative attractor KII	1.9	0.2	1.6	1.0

ei, ie, ee, ii are the connections between excitatory node and initiatory node.

D. KIII model

The KIII model is designed to be a dynamic computational model that simulates the sensory cortex. It can perform pattern recognition and classification. Based on the structure of the cortex, KIII consists of three layers connected

by feedforward/feedback connections. Each layer has multiple KII sets, connected by lateral weights between corresponding P5(1) and P6(3) nodes, see Fig.4. Based on the structure and dynamics of the sensory cortex, KII sets in each layer should be zero attractor KII, positive attractor KII, and negative attractor KII, respectively.

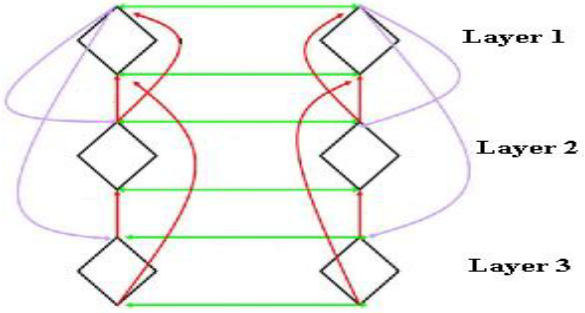

Fig. 4. Schema of a KIII set with three layers of KII units with feedforward, feedback, and lateral connections. It only shows two KII sets in each layers. In the implementation, there could be couple of dozens of KII sets in each layer.

There are feedforward and feedback interlayer connections in KIII. In order to get a homeostatic balanced state in the KIII level, we list the activation equations for all 12 nodes in a single column in KIII. For example, Eq.3 is the equation of P5(1) node in the first layer.

$$P_{15(1)} = ee1 * f(P_{15(3)}) - ie1 * f(P_{16(1)}) - ie1 * f(P_{16(3)}) + Wee * f(P_{15(1)}) + W_{P_{25(1)}P_{15(1)}} * f(P_{25(1)}) \quad (3)$$

In Eq.3, $P_{15(1)}, P_{15(3)}, P_{16(1)}, P_{16(3)}$ are the activations of node P5(1), P5(3), P6(1), and P6(3) in layer 1; $P_{25(1)}$ is the activation of P5(1) node in layer 2; ee1 and ie1 are the connections between KII sets in layer 1 (also see Fig. 3); Wee is the lateral connection between KII sets in layer 1; $W_{P_{25(1)}P_{15(1)}}$ is the feedback connection from $P_{25(1)}$ node in layer 2 to $P_{15(1)}$ node in layer 1; $f(x)$ is the asymmetric sigmoid function in Eq. 2.

After plugging in the parameters in table I, we balance 12 nonlinear equations for each node. Therefore, we get the parameters for all the lateral connections and feedforward/feedback connections between layers in Table II, III, and IV.

TABLE II
Lateral connections of three layers in KIII

	Wee	Wii
Layer 1	0.15	0.1
Layer 2	0.2	0.2
Layer 3	0.15	0.1

Wee is the lateral connections between excitatory nodes; Wii is the connections between inhibitory nodes.

TABLE III
Lateral connections of three layers in KIII

$W_{P_{25(1)}P_{15(1)}}$	0.05	Feedback connection
$W_{P_{25(1)}P_{16(1)}}$	0.25	Feedback connection
$W_{P_{36(1)}P_{16(1)}}$	0.05	Feedback connection
$W_{P_{15(1)}P_{25(1)}}$	0.15	Feedforward connection
$W_{P_{35(1)}P_{26(1)}}$	0.2	Feedback connection
$W_{P_{15(1)}P_{35(1)}}$	0.6	Feedforward connection

After we set up KIII with the parameters from table III and IV, we test impulse-response on KIII. Same as the experiment on KII, only a single pulse of negative input [-0.3 –1] is added to the P5(1) node in layer 1. (a), (b) and (c) are activations of the excitatory node P5(1) in three layers respectively. Figure 5 clearly shows that KIII sustains oscillatory aperiodic activation for a long period. KIII is in a homeostatic balanced state.

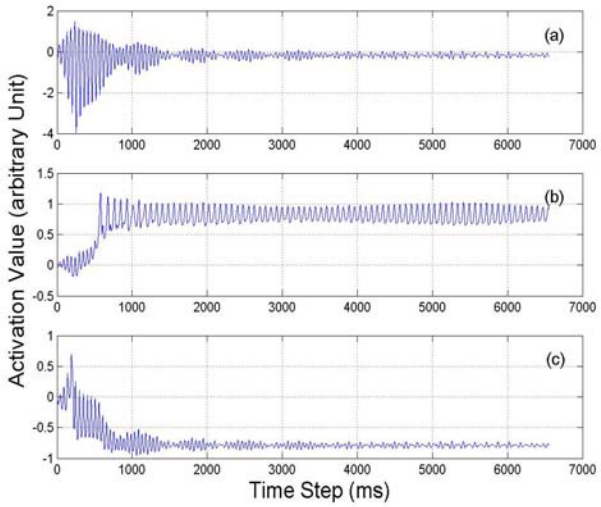

Fig. 5. Impulse-Response test on KIII: Temporal behavior of the activation of a KIII set with three layers.

III. KIII-BASED TIME SERIES PREDICTION ALGORITHM

A. Overview of prediction methods

Time series prediction is the application of estimating the future value according to the historical value. A periodic time series can be modeled by a superposition of sine and cosine waves of different frequencies, amplitudes, and phases. However, more complicated time series, such as currency exchange rate, are often dynamic, highly non-linear processes. There are several approaches to time series prediction.

- The basic methods of time series prediction use extrapolation through a global fit in the time domain or

autoregressive technique. These linear models are inadequate for dynamic data, such as the data generated by logistic equation, financial data, etc.

- The second approach is state-space reconstruction by time-delay embedding. This technique is powerful if the time series is generated by deterministic governing equations [7].
- The third method utilizes neural network. It can adaptively learn the patterns from data, generalize, and predict unknown patterns based on the information acquired from the historical data. Some neural networks, such as MLP, SOM, recurrent network, etc. have been implemented to forecast time series with certain prediction accuracy. [4, 5].

Financial data prediction is a particular application in time series prediction. Since the nature of the financial data is highly non-linear, non-stationary, and highly noisy, financial forecasting is considered a difficult problem. On the other hand, financial prediction methods are intensively researched, due to the special interest in investments.

There are two types of financial time series prediction: one-step prediction and multi-step prediction. Due to the model constraint, KIII is easier to classify patterns instead of giving real value output. In this work, we use the KIII model to predict the single-step direction of daily currency exchange rate. The data is from [4]. The comparison with the prediction results cited in that paper should be meaningful.

B. Neural Network Structure and Learning Rule

After the set up of KIII to a homeostatic balanced state, its aperiodic dynamic could sustain if the inputs are within certain range. We construct KIII with different lateral nodes, such as 30, 40, and 60. This parameter decides the dimension of the input time series. For example, in a KIII with 40 lateral nodes, the inputs for each excitatory node are X(t-40), X(t-39)… X(t-1). In this application, KIII with 40 lateral nodes performs better than other dimension choices. In other words, previous 40-day currency exchange rates are entered into the system, and our model predicts the next day rate change direction.

The learning rule in KIII is associative Hebbian learning. Since the system is always in a dynamic state, there is no single converged value. The activation standard deviation (σi) of each node in certain duration T is used.

$$\Delta Wij = L*(\sigma i - \sigma) * (\sigma j - \sigma). \quad (4)$$

In Eq. 4, L is the learning rate; σi and σj are activation standard deviation of two nodes i, j; ΔWij is the weight change between these two nodes.

$$\sigma_i = \frac{1}{T}\sqrt{\int_0^T \left(P35_i(t) - \frac{1}{T}\sum_{t=1}^{T} P35_i(t) \right)^2} \ . \quad (5)$$

In Eq.5, T = 200 ms, $P35_i(t)$ is the activation of excitatory node at instant t in layer 3; σ is defined as the mean of the σi.

$$\sigma = \frac{1}{L}\sum_{i=1}^{L} \sigma_i \ . \quad (6)$$

In Eq. 6, L = 40, is the number of lateral nodes.

Experiments show that the associative Hebbian learning rule we adopt adjusts the weights well. During the learning, the activations of all the nodes are always within the desired range [-2, 3], which is in the active region of our sigmoid function. If it is not this case, the activations will saturate the sigmoid function, and in turn cause the system to lose sensitivity. Some other variations of Hebbian learning rules, such as Hebbian with habituation are also tested in the experiments. The results show that the learning rule in Eq. 4 is better than other variations for this application.

IV. IMPLEMENTATION FOR CURRENCY EXCHANGE RATE PREDICTION

A. Data Normalization

The data we used is from [3]. It consists of more than three thousand data points of the daily closing bids of five currencies. We test our method on the Japanese Yen (JPY), the German Mark (DM), and the British Pound (GBP). In order to avoid bias, which is caused by the trend of certain segment in the time series, first order difference (FOD) is used instead of raw data [3, 4]. We test two normalization methods to FOD value. In the first method, the data is normalized to range [-1, 1] by proportionally scaling FOD value. The second approach is to use logarithm compression on FOD value, then scale to [-1 1]. The experiment results show that the first one is better. Therefore, we only use proportional scaling normalization on FOD in our work. The normalized data is the time series we put into KIII.

We conducted 30 test runs for each of the three currencies: the British Pound (BP), the German Mark (DM), and the Japanese Yen (JP). The average number of correct and incorrect predictions for each currency is used to evaluate the performance. In each test run, the normalized data is split into training, validation and testing segments. We use 150 daily exchange rates as training data, followed by 25 data as validation data, and then 25 data points for testing. In the next run, all the time series data are moved forward 50 days and so on.

As described in section III B, the KIII used in this work has 40 lateral nodes. Therefore 40 consecutive time series data consist of one input widow. All 40 values are put into the 40 corresponding excitatory nodes P5(1) in the first layer

of KIII. Due to the nature of KIII, the same input needs to be repeated for certain duration and followed by a period of relaxation (zero input). This is a biologically inspired process to mimic the activity of sniff and other sensory activities. The duration of learning and relaxation are parameters subjected to testing, since they determine the sensory frequency of KIII. Longer duration means more computation is needed. We tried different durations during the experiments. Finally, with the consideration of the balance of training efficiency and prediction accuracy, we chose learning and relaxation duration to be both 25 ms in this work.

B. Learning Phase

Prior to putting learning data into KIII, each input is labeled as UP/SAME/DOWN according to the next data in the normalized FOD time series, as in table IV.

TABLE IV
DEFINITION OF CATEGORIES

Category Label	Next data in normalized FOD time series
UP	FOD >= ε
SAME	- ε < FOD < ε
DOMN	FOD < - ε

In table IV, ε is a threshold value to separate different categories. In the following experiment, we set ε = 0. Issue related to the choice of threshold ε is discussed in section V.

As the number of data belonging to category SAME is relative small in this experiment data, and also for the sake of simplicity, we only let the KIII learn from the UP/DOWN categories. The Hebbian learning process happens in the lateral connections among the excitatory nodes in the third layer, see Fig 4. From the precious section, we know each learning phase has duration of 25 ms. In our experiment, the first 5 ms response activations are skipped since those initial activations are considered unstable when the signal is first perceived by the system. Thus, only the following 20 ms activations of excitatory nodes P35(1) in layer 3 are collected for learning purpose. Then, for each of the P35(1) node, the standard deviations of collected activations in 20 ms duration are filtered to the gamma range [20 40] Hz. This filtering process is also biologically motivated. We plug the filtered data into the Hebbian learning equation 4. The lateral connections are adjusted accordingly. We call this process a cycle. The learning cycle continues 150 times for different time series data windows. Observing the weight change indicates that some weights increase, while some decrease during learning.

C. Validation Phase

The validation phase is basically the same as the training phase, except no learning is applied. In each cycle, the filtered standard deviations of 40 excitatory nodes in the third layer are stored separately, according to the input data label. Therefore, we would have some references for each category (UP/DOWN). We use these references to classify patterns in the testing phase. From the point of view of aperiodic dynamics, the values of the spatially distributed oscillation intensities encode the input data. When a given input is applied, the high dimensional dynamics collapses to a lower dimensional subspace. In other words, these oscillatory patterns generate the clusters for different categories.

D. Testing Phase

During the testing phase, the filtered standard deviations are recorded for each testing data. It is defined as the activation point for that testing data. And then, the distances from all references to the activation point are calculated. The classification algorithm selects the N references with the smallest distances. Our classification algorithm requires that the difference between the numbers of references belonging to each pattern be greater than the T, in order to claim a "winner". The "winner" pattern decides to which pattern this testing data belongs. For example, there are 9 nearest neighbors. The threshold T is 2. In Case 1, 6 references belong to pattern UP and 3 references belonging to pattern DOWN. Since the difference 3 is greater than T, the system classifies this testing point as the "winner", UP category. In Case 2, if 5 references are pattern UP, 4 references are pattern DOWN. Since the difference 1 is less than T, the system classifies this testing data as "Undecided". In other words, if the system cannot decide the "winner" by the threshold T, the testing data is "Undecided". This introduction of the "Undecided" class increases the confidence level of the results and also improves the prediction accuracy. [4]

The size of the neighborhood N and the threshold T, which decides the "winner", are acquired from the experiment in the testing phase. We use the first one third of testing data to try some N and T values. We check the classification performance for every N and T configurations. The N and T configuration with the best performance is selected for the following two-thirds of testing data. Once N and T are decided, they will remain constant for all of the following testing data. Therefore, only two-thirds of testing data is counted as performance evaluation in each test run.

E. Results of Prediction by KIII

The currency exchange rate prediction results for the British Pound (BP), the German Mark (DM), and the Japanese Yen (JPY) are given in Table V.

TABLE V
AVERAGE PREDICTION PERFORMANCE*

	JPY	DM	BP
Correct	55.0%	54.5%	55.7%
Incorrect	33.4%	42.3%	41.0%
Undecided	11.6%	3.2%	3.3%

* 30 runs are performed for each currency

In Table V, "correct" means KIII predicts the direction of the move correctly. If the predicted direction is wrong, we say it is "Incorrect". If a certain activation point could not be classified to either category, because there is no "winner" within N nearest neighbors according to threshold T, we call this case "Undecided".

Comparing table V results to [4], where the error rate is about 40%, we can conclude that the performance of KIII is comparable (for DM and BP) or better (in the case of JPY).

V. Discussion and Conclusion

Based on the biological models, we assume that the connections within layers have a certain topology [2]. In current KIII implementation, we simplify the architecture by using complete connections between KII sets in each layer. Also, each layer has the same number of KII sets. But in realistic biological models, the convergence and divergence between layers of different sizes is crucial. Both simplifications decrease the complexity of KIII. On the other hand, it might weaken the computation power of the system in terms of the capacity of classification, generalization, and robustness. Therefore, KIII architecture and dynamics has the potential of future improvement.

As mentioned in section IV B, threshold ε is used to separate different categories. Obviously, if we increase ε, the number of time series data belonging to category SAME would increase. Therefore, three categories should be learned and classified by the system. Although this change increases the difficulty for the system to classify, it makes more practical sense. This makes a certain level of fluctuations in currency exchange rates to be classified as category SAME. The value of ε is currency dependent, since some currencies have small fluctuation at high frequency, while other currencies have large variations at low frequency. How to adjust the ε value accordingly is the topic for future studies.

The K models are promising dynamic neural networks. Through this work, we have demonstrated the classification capacity of the KIII model. In the currency exchange rate prediction application, the KIII model shows very good results, compared to other advanced recurrent neural network predictions. Testing various Hebbian learning rules, the associative Hebbian learning rule is found adequate for the KIII model. Although the prediction performance obtained in the current study might not provide high enough margin to be profitable in practice, it shows the capability of the dynamic aperiodic neural networks to predict very complex temporal sequences.

Acknowledgment

This research is supported by NSF grant EIA-0130352, and NASA grant NCC-2-1244. The support of Prashant Ankaraju in building the KIII computational model is appreciated.

References

[1] W.J. Freeman (1975), *"Mass Action in the Nervous System"* Academic Press, N.Y.

[2] R. Kozma, W.J. Freeman, "Aperiodic Resonance – Methods and Applications of Noisy and Variable Patterns*," Int. J. Bifurcation & Chaos*, 11(6), 2307-2322, 2001.

[3] R. Kozma, W.J. Freeman, P. Erdi, "The KIV Model – Nonlinear spatio-temporal dynamics of the primordaial vertebrate forebrain," *Neurocomputing,* 2003, in press.

[4] C.L. Giles, S. Lawrence, A.C. Tsoi, "Noisy Time Series Prediction using a Recurrent Neural Network and Grammatical Inference," *Machine Learning*, 44(1-2), 161-183, 2001.

[5] J. G. Carney, P. Cunningham, "Neural Networks and Currency Exchange Rate Prediction," *Foresight Business Journal,* http://www.maths.tcd.ie/pub/fbj/index.html, 1996.

[6] C. Bellgard, "Forecasting foreign exchange rates for profit," *Dissertation*, Dept. of Information Management & Marketing, 1998.

[7] A. S. Weigend, N. A. Gershengeld (Eds) "*Time Series Prediction: Forecasting the Future and Understanding the Past*," Santa Fe Institute, 1994.

[8] H.Z. Chang, W.J. Freeman, B.C. Burke (1998) "Optimization of Olfactory Model in Software to Give 1/f Spectra Reveals Numerical Instabilities in Solutions Governed by Aperiodic (chaotic) Attractors", Neural Networks, 11, 449-466.

[9] J. Moody and L. Wu, "Optimization of trading systems and portfolios," in Proc. Neural Networks Capital Markets Conf. Pasadena, CA, Nov. 1996.

[10] D. Bassi, "Stock price predictions by recurrent multiplayer neural network architectures," in Proc. Neural Networks in the Capital Markets Conf., A. Reference, Ed.. London, U.K., October 1995, London Business School, pp. 331-340.

[11] N. Toomarian and J. Barhen, "Adjoint-functions and temporal learning algorithm in neural networks," Advances in Neural Information Processing Systems, 3, pp. 113-120, 1991.

[12] A. Atiya and A. Parlos, "Identification of nonlinear dynamics using a general spatio-temporal network," Math. Comput. Modeling J., vol. 21, no. 1, pp53-71, Jan. 1995.