Comparison of Iterative and Direct Approaches for Multi-Steps Ahead Time Series Forecasting using Adaptive Hybrid-RBF Neural Network

Mazlina Mamat
Institute of Microengineering and Nanotechnology,
Universiti Kebangsaan Malaysia,
43600 UKM Bangi Selangor, Malaysia

Salina Abdul Samad
Faculty of Engineering and Built Environment,
Universiti Kebangsaan Malaysia,
43600 UKM Bangi Selangor, Malaysia

Abstract—**Most available forecasters were designed in non-adaptive approach whereby the forecasters' parameters were updated during training phase. Slightly different, this paper introduces an adaptive forecaster built from the Hybrid Radial Basis Function neural network, in which its parameters were updated continuously in real time using new data. To achieve this, two learning algorithms: Adaptive Fuzzy C-Means Clustering and Exponential Weighted Recursive Least Square were used to train the Hybrid Radial Basis Function in adaptive mode. The multi-steps ahead forecasting were achieved by using two approaches: iterative and direct. The performance of each approach is measured by the Root Mean Square Error and R^2 test of the actual and forecasted output on two time series data: Mackey-Glass and Data Series A from Santa-Fe Competition. Simulation results show that the adaptive forecaster is able to produce accurate forecasting output for several steps ahead depending on the complexity of data. Simulation results also reveal that the direct approach overcomes iterative approach in long distance forecasting.**

Keywords- ***Artificial neural networks; Adaptive learning; Iterative approach; Direct approach; Time-series forecasting.***

I. INTRODUCTION

In forecasting, not only the reliability of forecasted output is in interest but how far the forecaster is able to produce acceptable forecasting output is also important. To obtain an outstanding one step ahead (OSA) forecasting for any system is hassle free for most forecasting methods. However this is not the case in multiple steps ahead (MSA) or multi period forecasting. In general, two different ANN architectures can be used to perform MSA forecasting. The first architecture which is referred as *iterative approach* by some researchers consists of a single output ANN to obtain MSA by feeding back the forecasted output iteratively. The second architecture or renowned as *direct approach* is using *m* ANNs, that are trained independently to generate forecasting output for *m* steps ahead.

It is uncertain among the researchers to point the best approach since their findings are opposing with each other. In comparison between iterative and direct approaches, Weigend, Huberman, and Rumelhart state that the iterative approach produces better forecasting performance than direct approach in their sunspot data analysis [1]. This finding is supported by Hill, Marquez, O'Connor, and Remus in predicting the 111 M-competition time series study [2]. On the other hand, quite a number of researchers claim that the direct approach is superior than the iterative approach. This can be observed in their experiment on the temperature signal in the industrial furnace installation. They claim that the direct approach using multiple chaining of Multi Layer Perceptron neural network is superior than the standard iterative approach especially for long term forecasting [3]. Another finding which supports direct approach over iterative approach is presented by [4] in the attempt to forecast an hourly customer demand for gas at a compression station. The same observation was also presented by Atiya, El-Shoura, Shaheen, and El-Sherif which compares few methods including iterative and independent approaches to forecast the flows of rivers Nile**.** The simulation results show that the direct approach is the best method for multiple steps ahead prediction among all the proposed methods in their study [5].

The diverse findings obtained by various researchers above conclude that it is not clear which approach gives better results. One thing that can be concluded is different data need different approach to forecast well in MSA. Another thing that can be seen clearly in the above studies that most the ANN forecasters used were using Multilayer Perceptron (MLP) and Backpropagation (BP) neural network operated in non-adaptive mode. Further study also reveals that less literature can be found about the use of Radial Basis Function neural network in time series forecasting and can be said none using Hybrid Radial Basis Function (HRBF). For that reason, this paper presents the HRBF as a tool to perform MSA time series forecasting in adaptive mode. This paper also compares the performance of two approaches: iterative and direct on MSA forecasting.

The rest of the paper is organized as follows. Section 2 briefly describes the Hybrid Radial Basis Function Neural Network and the learning algorithms used. It also describes the iterative and direct approaches forecasting. The paper then presents the simulation results of multi steps ahead forecasting obtained by iterative and direct approaches on two time series data. This section also describes the data used in brief, the evaluation tests and how the analyses were performed. The last section concludes the overall findings obtained.

II. METHODOLOGY

The HRBF Neural Network is a modification of the Radial Basis Function (RBF) [6]. It differs from the RBF by the linear connections that exist between input nodes and output nodes. Similarly to RBF, the HRBF neural network has three layers: an input layer, a hidden layer with a non-linear RBF activation function and a linear output layer. Each layer has its own nodes where the nodes in input layer are connected to the nodes in hidden layer and nodes in hidden layer are connected to the nodes in output layer via linear weight. There are two parameters: the HRBF centres in hidden nodes and weights between the hidden nodes and the output nodes, that must be updated during training to produce the optimize network. Two algorithms: Adaptive Fuzzy C-Means Clustering (AFCMC) and Exponential Weighted Recursive Least Square (e-WRLS) are used to update the centre and weight of the HRBF neural network. The AFCMC algorithm is a modification of Fuzzy C-Means Clustering Algorithm to overcome the dead centre, local minima and centre redundancy: common problem in the Fuzzy C-Means Clustering Algorithm [7]. To estimate the weight between the hidden nodes and output node, the Exponential Weighted Recursive Least Square (e-WRLS) is used [8].

In iterative approach, an ANN with a single output node is used to perform OSA forecasting and MSA forecasting. In OSA, the ANN will be given input vectors of current and previous output and be trained to forecast the next output. For instance, if the ANN is given input vectors *m*(*t=1*), *m*(*t=2*), *m*(*t=3*) and *m*(*t=4*), the ANN is trained to forecast the output at time *t*=5. The forecasted output then will be compared to the real output at time *t*=5 and the statistical error between the actual and predicted output is computed. The ANN will be trained to minimize the statistical error between the forecasted and actual output for all data in the series. To perform MSA forecasting, the forecasted output at time *t-1* is included in the input vectors while removing the oldest output from the input vectors to get the forecasted output at time *t*. For example, consider an ANN receives input vectors *m*(*t=1*), *m*(*t=2*), *m*(*t=3*) and *m*(*t=4*), and produces OSA forecasting at *m(t=5)*. The MSA forecasting at *t=6*, is achieved by feeding in the OSA forecasting at *m(t=5)* as input to the ANN while removing the output *m(t=1)* from the input vector. Similarly, to obtain the MSA forecasting at *t=7*, the forecasted output at *t=6* will be used as input to the ANN while removing the oldest data from the input vector that is *m(t=2)*. This procedure is repeats to obtain the MSA forecasting for t=8, 9,10 and so on.

In direct approach, the MSA forecasting can be obtained by using an ANN which was trained to produce the particular MSA forecasting directly. To get clear view, consider the following condition: let the input vectors to the ANN are *m*(*t=1*), *m*(*t=2*), *m*(*t=3*) and *m*(*t=4*) and the ANN is needed to produce forecasting for 8 steps ahead. Using direct approach, this problem can be solved by training the ANN to minimize the error between target (actual output at t=12) and forecasted output at t=12. During testing, by feeding in the latest available input data, for instance *m(t=9)*, *m(t=10)*, *m(t=11)* and *m(t=12)*, the ANN can be used to forecast the output at t=*20*. To get forecasting for different step ahead, different ANN is required since the ANN was trained specifically for the particular step ahead.

III. SIMULATION RESULTS AND DISCUSSIONS

The simulation programs are constructed using Borland C++ Builder 6. The analysis was conducted without iteration and data was given to the forecaster as in real time application. The other parameters of HRBF such as $\mu(0)$ and q are set to 0.95 and 0.30, respectively. The Root Mean Square Error (RMSE) and Coefficients of Determination (R^2) are used as measurement of derivation between actual and forecasted values. The R^2 value denotes the close degree between actual and forecasted value and are varies, which 1 indicates that the predicted data and the actual data are identical. The RMSE and R^2 tests are defined in the following respectively:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=n_a}^{n_d}(\hat{y}(t)-y(t))^2} \tag{1}$$

$$R^2 = 1-\left[\frac{\sum_{t=n_a}^{n_d}(y(t)-\hat{y}(t))^2}{\sum_{t=n_a}^{n_d}(y(t)-\overline{y})^2}\right] \tag{2}$$

where n_a and n_d are the first and the last measurement data, n is the number of data in the measurement: $n = (n_d - n_a+1)$, $y(t)$ and $\hat{y}(t)$ are actual and forecasted value at time t and $\overline{y}$ is the average actual value, respectively.

The purpose of the experiments in this study is to compare the performance of iterative and direct approaches in multiple steps ahead forecasting. Therefore both approaches are set to produce up to 25 steps ahead forecasting. By considering the data is given without iteration, the first 500 data were not included in the measurement as the forecaster is still in learning process. Therefore the real and forecasted data from 501 to 900 (400 data) were used. The selection of input lag and the number of HRBF center have strong impact on the forecaster performance. Therefore the forecaster undergoes two analyses, first to determine the best input lag and second to determine the correct number of HRBF center. Both analyses were made by replacing other HRBF parameters with the typical values [6]. The analysis to determine the best input lags for all data is conducted by fixing the HRBF centers at 10 and varies the input lag from *(t-1)* to *(t-1)(t-2)(t-3)…(t-x)* where the maximum number of x is 10% out of the total available data. The input lags which produces the highest R^2 values for all steps ahead then undergoes the analysis to determine the best number of HRBF center. This analysis is conducted by increasing the number of HRBF center one by one until it reaches 10% out of the total available data. For each approach, the structure that produces the best forecasting performance (lowest RMSE values and highest R^2 values) for all steps ahead forecasting is selected as the best structure for that approach.

The forecasting performance of the HRBF trained with two algorithms described above is evaluated by using two time series: Mackey-Glass nonlinear time series and Data Set A from the Santa Fe Competition.

A. Mackey Glass Data

The forecasting based on time series produced by the Mackey-Glass equation is regarded as a criterion for comparing the ability of different predicting method and is used in many time series forecasting researches [9],[10]. The Mackey-Glass equation is a time-delayed differential equation proposed as a model of white blood cell production by Mackey and Glass. The Mackey-Glass equation is given by:

$$x_t = \frac{\alpha x_{t-s}}{1+(x_{t-s})^{10}} + (1-\beta)x_{t-s} \tag{3}$$

where $\alpha = 0.2$, $\beta = 0.1$, and s is delayed time. In this research, s is set to 17 in which the equation exhibits chaotic behavior with a fractal dimension.

For Mackey Glass data, the iterative approach requires *(t-1)(t-2)(t-3)…(t-41)(t-42)(t-43)* input vectors and 22 centres while direct approach requires *(t-1)(t-2)(t-3)…(t-22)(t-23)(t-24)* input vectors and 69 centres respectively to generate reliable forecasting output for all 25 steps ahead. Table 1 displays the RMSE values and the R^2 values for 1 to 10, 15, 20 and 25 steps ahead forecasting by using iterative and direct approaches for Mackey Glass Data.

For Mackey-Glass data, the iterative approach overcomes the direct approach for one to nine steps ahead forecasting. The RMSE value obtained by all these nine steps ahead forecasting are at least 9% smaller than those obtained by direct approach. The same observation was also reflected in the R^2 values obtained by iterative approach which denote higher values to compare with those obtained by direct approach. However for longer distance such as 10, 15, 20 and 25 steps ahead, the direct approach produces better forecasting to compare with iterative approach where for 15, 20 and 25 steps ahead, the performance are good with the R^2 values obtained exceed 0.9. Fig. 1 to 6 show comparison between iterative and direct approach for data 501 to 1000 of Mackey Glass data for 1, 5, 10, 15, 20 and 25 steps ahead forecasting. For 1 step ahead forecasting (Fig.1), both approaches produce very accurate forecasting outputs and are identical to the actual outputs. For 5 steps ahead, the performance of both approaches was still good and the forecasting outputs show very small variation to the actual output. However as the forecasting distance increases to 10, 15, 20 and 25 steps ahead, the iterative approach shows larger variation to compare with direct approach.

TABLE 1. COMPARISON OF ITERATIVE AND DIRECT APPROACHES FOR MACKEY GLASS DATA

	RMSE		R^2	
Approach	Iterative	Direct	Iterative	Direct
Input Lag	43	24	43	24
HRBF Center	22	69	22	69
1 Step Ahead	0.0007	0.0013	1.0000	1.0000
2 Steps Ahead	0.0031	0.0057	0.9999	0.9996
3 Steps Ahead	0.0085	0.0167	0.9990	0.9966
4 Steps Ahead	0.0175	0.0269	0.9956	0.9903
5 Steps Ahead	0.0298	0.0436	0.9872	0.9746
6 Steps Ahead	0.0446	0.0615	0.9713	0.9492
7 Steps Ahead	0.0604	0.0778	0.9473	0.9191
8 Steps Ahead	0.0759	0.0904	0.9168	0.8886
9 Steps Ahead	0.0900	0.0982	0.8831	0.8679
10 Steps Ahead	0.1018	0.1010	0.8504	0.8600
15 Steps Ahead	0.1247	0.0732	0.7757	0.9282
20 Steps Ahead	0.1228	0.0478	0.7825	0.9714
25 Steps Ahead	0.1346	0.0712	0.7385	0.9481

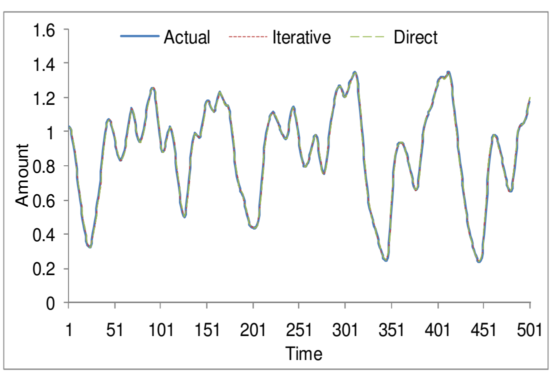

Figure 1. One step ahead forecasting using iterative and direct approaches (Mackey Glass Data)

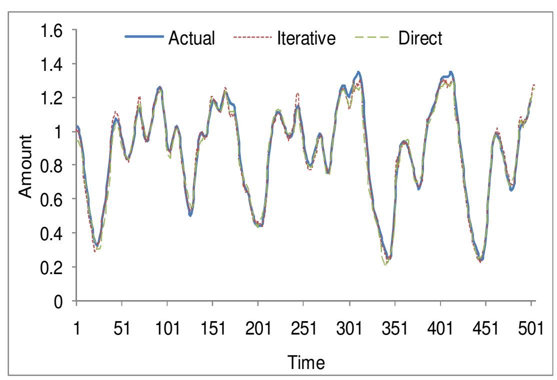

Figure 2. Five steps ahead forecasting using iterative and direct approaches (Mackey Glass Data)

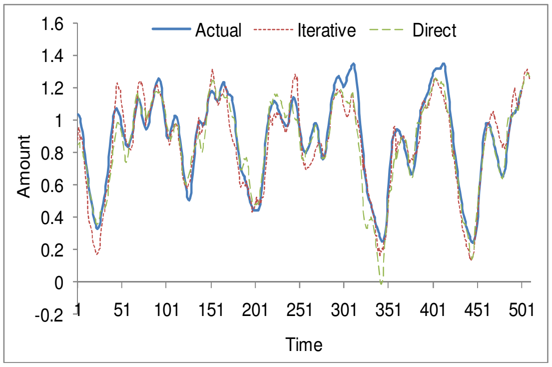

Figure 3. Ten steps ahead forecasting using iterative and direct approaches (Mackey Glass Data)

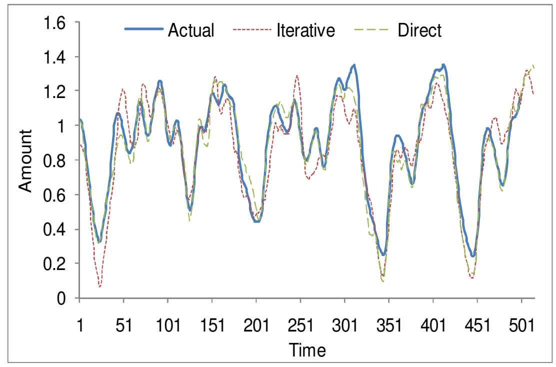

Figure 4. Fifteen steps ahead forecasting using iterative and direct approaches (Mackey Glass Data)

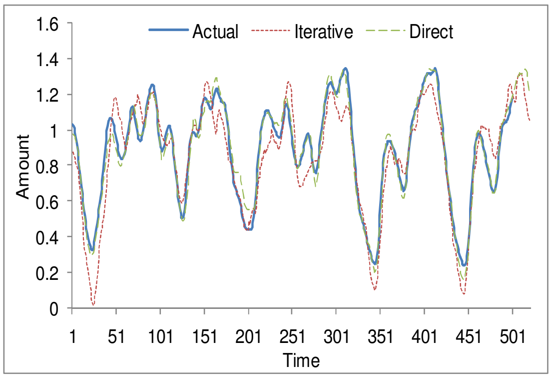

Figure 5. Twenty steps ahead forecasting using iterative and direct approaches (Mackey Glass Data)

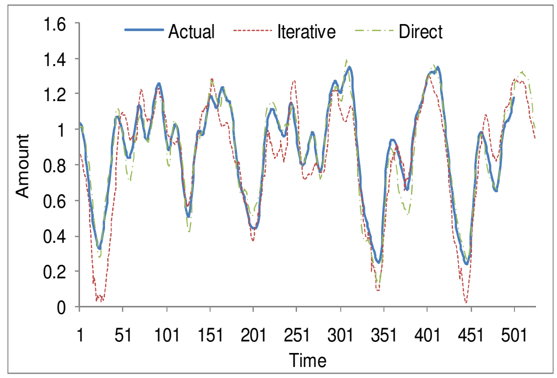

Figure 6. Twenty-five steps ahead forecasting using iterative and direct approaches (Mackey Glass Data)

B. *Data Set A from Santa-Fe Competition*

The Data set A from the Santa-Fe Competition were recorded from a Far-Infrared-Laser in a chaotic state. These data were chosen because they are a good example of the complicated behavior that can be seen in a clean, stationary, low-dimensional non-trivial physical system for which the underlying governing equations dynamics are well understood. For Santa-Fe data, the iterative approach requires *(t-1)(t-2)(t-3)(t-4)* input vectors and 30 centres while direct approach requires *(t-1)(t-2)(t-3)* input vectors and 29 centres respectively to generate reliable forecasting output for all 25 steps ahead. Table 1 displays the RMSE values and the R^2 values for 1 to 10, 15, 20 and 25 steps ahead forecasting by using iterative and direct approaches for Santa-Fe Data.

For Santa-Fe data, both approaches show good performance for 1 step ahead forecasting where the R^2 value obtained by them exceeds 0.9. For 2 to 25 steps ahead, the performance of both approaches was not encouraging with the R^2 values obtained for these steps ahead are lesser than 0.65. The iterative approach was totally failed to perform more than 2 steps ahead forecasting while the direct approach shows better performance for 5 to 10 steps ahead forecasting. Both approaches fail to forecast satisfactorily for distance more than 10 steps ahead.

Fig. 7 to 9 show comparison between iterative and direct approach for data 501 to 1000 of Santa-Fe data for 1, 5, and 10 steps ahead forecasting. For 1 step ahead forecasting (Figure 7), the performance of both approaches can be considered as good with the RMSE and R^2 obtained are lower than 12.20 and greater than 0.91 respectively. For 5 steps ahead (Fig. 8) and 10 steps ahead (Fig. 9) the performance of iterative approach was poor while the forecasting outputs generated by direct approach are acceptable.

IV. Conclusion

This paper compares the performance of iterative and direct approaches in multiple steps ahead forecasting using an adaptive HRBF. From the results, it can be noted that both iterative and direct approaches demonstrate good performance for short distance forecasting. However for long distance forecasting, direct approach produces better results. These were proved by the results obtained from the experiments on two time series data: Mackey Glass and Data Set A from Santa-Fe Competition.

TABLE 2: COMPARISON OF ITERATIVE AND DIRECT APPROACHES FOR DATA SET A FROM SANTA-FE COMPETITION

	RMSE		R^2	
Approach	Iterative	Direct	Iterative	Direct
Input Lag	4	3	4	3
HRBF Center	30	29	30	29
1 Step Ahead	10.73	12.17	0.9325	0.9132
2 Steps Ahead	24.69	30.74	0.6427	0.4464
3 Steps Ahead	33.63	36.59	0.3374	0.2152
4 Steps Ahead	36.80	28.75	0.2067	0.5157
5 Steps Ahead	42.59	25.53	< 0	0.6180
6 Steps Ahead	49.47	25.82	< 0	0.6094
7 Steps Ahead	> 50	25.62	< 0	0.6152
8 Steps Ahead	> 50	27.38	< 0	0.5609
9 Steps Ahead	> 50	24.60	< 0	0.6453
10 Steps Ahead	> 50	27.55	< 0	0.5554
15 Steps Ahead	> 50	35.37	< 0	0.2668
20 Steps Ahead	> 50	36.29	< 0	0.2283
25 Steps Ahead	> 50	38.80	< 0	0.1181

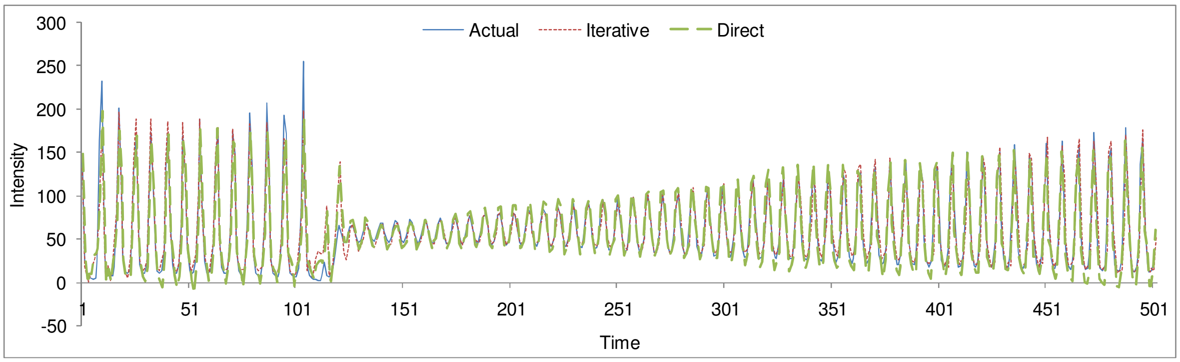

Figure 7. One step ahead forecasting using iterative and direct approaches (Data Set A)

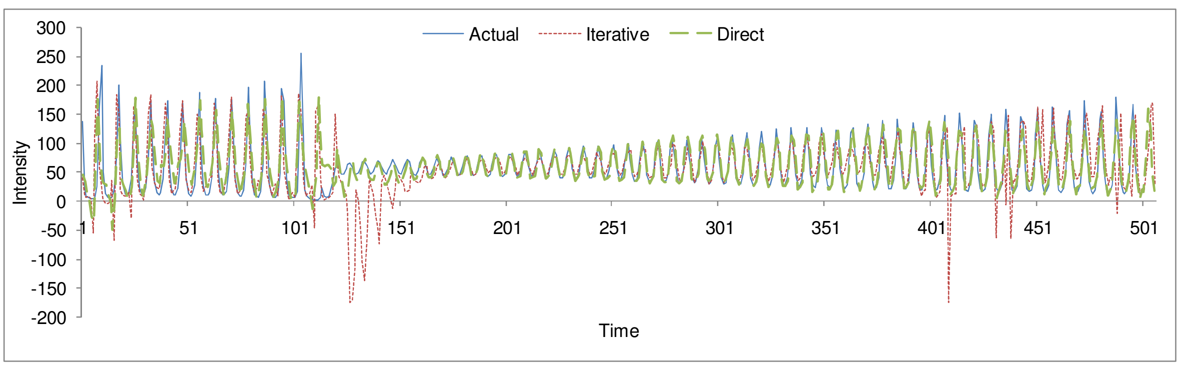

Figure 8. Five steps ahead forecasting using iterative and direct approaches (Data Set A)

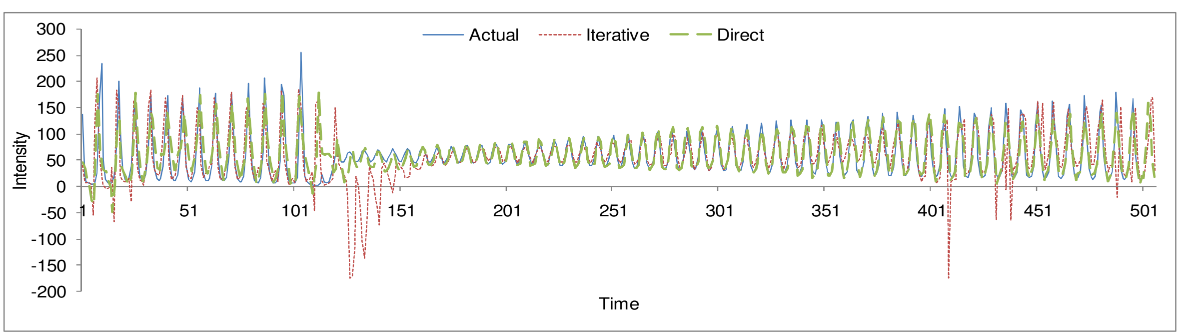

Figure 9. Fifteen steps ahead forecasting using iterative and direct approaches (Data Set A)

REFERENCES

[1] A. S. Weigend, B. A. Huberman, and D. E. Rumelhart, "Predicting sunspots and exchange rates with connectionist networks," In M. Casdagli & S. Eubank (Eds.), Nonlinear modeling and forecasting. Redwood City, CA: Addison-Wesley, 1992.

[2] T. Hill, L. Marquez, M. O'Connor, and W. Remus, "Artificial neural networks for forecasting and decision making," International Journal of Forecasting, vol. 10, pp. 5-15, 1994.

[3] M. Duhoux, J.A.K Suykens, B. De Moor and J. Vandewalle, "Improved long-term temperature prediction by chaining of neural networks," Int. J Neur. Sys., vol. 11(1), pp.1–10, 2001.

[4] N. Hanh, and C. Chan, "Multiple neural networks for a long term time series forecast," Neural Computing & Applications, vol. 13 (1), pp. 90-98, 2004.

[5] A. F. Atiya, S. M. El-Shoura, A. I. Shaheen, and M. S. El-Sherif, "A Comparison Between Neural-Network Forecasting Techniques—Case Study: River Flow Forecasting," IEEE Transactions on Neural Networks, vol. 10(2), 1999.

[6] Mashor, M.Y. "Modification of the RBF Network Architecture," ASEAN J. Sci. Technol. Dev., vol 17(1), pp 87-108, 2000.

[7] M.Y. Mashor, "Adaptive Fuzzy C-Means Clustering Algorithm for A Radial Basis Function Network," Int. J. Syst. Sci., vol. 32(1), pp. 53-63, 2001.

[8] J. A. Karl and W. Bjorm, "Computer Controlled Systems : Theory & Design," 3rd edn. New Jersey: Prentice Hall. 1997.

[9] C. F. Gao, T. L. Chen, and T. S. Nan, "Discussion of Some Problems About Nonlinear Time-Series Prediction Using v-Support vector Machine," Communication Theory Physics, vol. 48, pp. 117-124, 2007.

[10] K.R. Muller, A. J. Smola, G. Ratsch, B. Scholkopf, and J. Kohlmorgen, "Using Support Vector Machines for Time series Prediction," in Advances in kernel methods: support vector learning. MA, USA. MIT Press Cambridge, 1999.