Artificial Neural Network Non-linear Auto Regressive Moving Average (NARMA) Model for Internet Traffic Prediction

Mohd Naqiuddin Sahrani, Md Mahfudz Mat Zan, Ihsan Mohd Yassin, Azlee Zabidi, Megat Syahirul Amin Megat Ali
Faculty of Electrical Engineering, Universiti Teknologi Mara, Shah Alam, Malaysia.
ihsan.yassin@email.com

***Abstract*—The technology of computing and network communication is undergoing rapid development, leading to increasing number of applications and services being available online. As more applications are available online, network traffic becomes a significant problem as high network loads may limit access to users. In this paper, we propose an internet traffic Nonlinear Auto-Regressive Moving Average model (NARMA) prediction model to assist network managers in forecasting internet traffic and planning their resources accordingly. The Multi-Layer Perceptron (MLP) estimator was used in this paper. The performance of the model were evaluated using Mean Squared Error (MSE), correlation tests, and residual histogram tests with good agreement between the model and actual outputs.**

***Index Terms*—Network Traffic Prediction; Nonlinear Auto-Regressive Moving Average (NARMA); System Identification; Forecasting.**

I. INTRODUCTION

Computer networks are an essential part of modern computing with the ability to share data and information across geographical borders [1]. Due to the explosive demand for internet-based applications [2, 3], computer networks may experience congestion due to high traffic [4, 5]. When this occurs, the networks users may experience reduced quality of service, i.e. higher latency, delay and packet loss [4]. Therefore, there is a need for an effective network traffic prediction model prediction model for network administrators to plan and manage their resources [6], detection of security threats [7, 8], and early discovery of potential network errors [6].

In this paper, we present two Nonlinear Auto-Regressive Moving Average (NARMA) models for prediction of network traffic: Multi-Layer Perceptron (MLP) and polynomial. We construct the model based on data collected from the Internet Traffic Archive [9].

The rest of this paper is organized as follows: Section 2.0 presents the literature review, followed by the methodology in Section 3.0. Results and discussions are presented in Section 4.0. Finally, concluding remarks are presented in Section 5.0.

II. LITERATURE REVIEW

A. Internet Traffic Prediction

Prediction of internet traffic is an active research field [6]. Among the advantages of internet traffic prediction is the ability to effectively allocate bandwidth, predicting unusual access patterns, and security reasons [6-8, 10]. Several works on internet traffic prediction are using traditional linear prediction models [11], such as Auto-Regressive (AR), Moving Average (MA), and Auto-Regressive Moving Average (ARMA) models. However, references [6, 11, 12] reported that these prediction models have low prediction accuracy, as the models are linear but the system is actually nonlinear [11, 12].

Nonlinear predictors, such as Artificial Neural Networks (ANNs) [10], have been used to address this issue. The nonlinear nature of ANNs, and the ability of the ANNs to theoretically approximate any input/output relationship given sufficient hidden units [11, 13].

B. Artificial Neural Network

ANNs are learning adaptive models inspired by the human neural system that is able to perform complicated tasks such as classification, clustering, prediction and association [7, 14-17].

Several learning paradigms exist for ANNs: supervised, unsupervised and reinforcement learning. Among them, supervised learning is most practical for various problems. The paradigm presents inputs and expected outputs for the ANN to learn from. The learning process associates the inputs and outputs by adjusting the weights of the ANN according to a training algorithm. The performance of the ANN is monitored via error measures such as Mean Squared Error (MSE). Supervised learning is typically used in classification and regression problems [7].

Among the most popular ANN is the Multi-Layer Perceptron (MLP) [13]. The MLP arranges its neurons (computation units) into three or more layers – one input layer, one or more hidden layer(s), and one output layer. The hidden layer(s) enable the MLP to learn complex functions. According to [18], the MLP has several distinguishing characteristics that separate it from other ANN subtypes [18]:

i. The MLP contains one or more layers of hidden units that are not part the input/output of the network. The hidden units enable the MLP to learn complex tasks and meaningful features from the input/output relationship.
ii. Presence of nonlinear activation functions in its hidden layer.
iii. High degree of connectivity between the MLP layers, determined by the weights of the network.

C. The NARMA model

The NARMA model is a nonlinear non-parametric system identification model. Together with its parent model, Nonlinear Auto-Regressive Moving Average model with Exogenous Inputs (NARMAX), they are powerful, efficient and unified representations of a variety of nonlinear systems [19-29]. Rich literature is available regarding its success in various electrical, mechanical, medical and natural applications [27, 30-32].

The identification method for NARMA and its class are performed in three steps [33]. Structure selection is performed to detect the underlying structure of a dataset. This is followed by parameter estimation to optimize some objective function (typically involving the difference between the identified model and the actual dataset) [30]. The structure selection and parameter estimation steps are typically performed using the Error Reduction Ratio (ERR) method [33]. Recursive models such as NARMAX and NARMA recursively adds residual terms to the NAR/NARX model to eliminate the bias and improve the model prediction capability [34-37]. Structure selection and parameter estimation for NARX are done twice as the steps are recursively repeated on the residual set until a satisfactory model is obtained [19]. Finally, the model is validated to ensure that it is acceptable.

III. METHODOLOGY

A. Hardware Description

All of the experiments were done on an ASUSTM laptop with Intel® core™ i5 Central Processing Unit (CPU) running at 2.40GHz, with 6.00GB of Random Access Memory (RAM), Microsoft Windows 7 Home Premium (64-bit) was installed as the operating systems. All programs were implemented using MATLAB R2015a.

B. Dataset Description

The dataset was taken from open source data from The Internet Traffic Archive website [9]. The dataset was sponsored by ACM SIGCOMM where BC-Oct89Ext was taken as dataset. The trace of dataset BC-Oct89Ext began at 23:46 on October 3, 1989, where the first 1 million external arrivals were captured. The traces was recorded in Eastern Daylight Time (EDT).

During the measurements, the packet captured excludes corrupted ones. Approximately 99.5% data taken is from the IP encapsulated packets. Timestamps that used are to six decimal point with four microsecond precision and with the accuracy of ten microseconds. The tracing was done at the Bellcore Morristown Research and Engineering facility, where all traffic between Bellcore and the Internet was captured.

The original data needs to be filtered prior to identification as the acquired traces timestamps are inconsistently spaced. After being filtered, the data consisted of 394 data points and divided into 3 sets where 276 data point was used for training set, 59 data point was used for the validation set and 59 data points were used for the testing set. The dataset for the input was plotted into graph as shown in Figure 1.

C. Experimental Overview

An overview of the experiments performed is shown in Figure 2. First, after the required dataset is obtained, the dataset was divided into three set which was training,

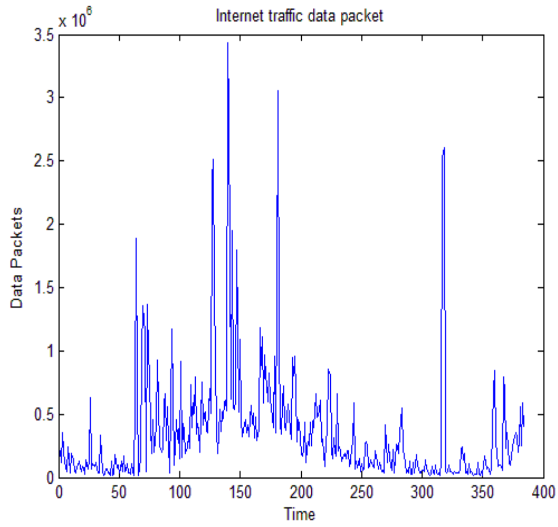

Figure 1: Dataset used in the experiments (after filtering)

validation and testing. The training set was used to adjust the weights of the MLPs, while the validation and testing sets were used to avoid overfitting and measure the performance of the MLPs with an unbiased dataset, respectively.

The first MLP (MLP1) is responsible to estimate the Nonlinear Auto-Regressive (NAR) model, which is the first step in estimating the full NARMA model. Parameters such as lag space and number of hidden units were varied and the MLP1 with the optimal performance was used to estimate the output of the NAR model. A summary of the proposed method is presented in Figure 3.

The residuals of MLP1 is produced when the prediction of the NAR model is compared with the actual output. The residuals were then used to construct the Moving Average (MA) model using the second MLP (MLP2).

The final output of the NARMA model was constructed by combining the outputs from the optimal output of MLP1 and MLP2. The NARMA model was validated using Mean Squared Error (MSE), correlation tests, and residual histogram analysis, all of which are described in Section 3.4.

D. Testing Methods

This section describes the testing methods used in the performance analysis and model validation. The methods used were MSE, correlation tests and residual histogram analysis.

A common method to measure model fit is by using MSE. MSE is defined as below [11, 38, 39]:

$$MSE = \frac{1}{n}\sum_{n=1}^{n}\left(\hat{y}(n) - y(n)\right)^2 \tag{1}$$

A low MSE score indicates a better model fit as this means the cumulative residuals are minimal. The residual produced by the model is in the form of:

$$\varepsilon(t) = y(t) - \hat{y}(t) \tag{2}$$

In system identification research, in addition to small magnitude, the residuals need to exhibit properties similar to white noise in order for it to be accepted as a valid and unbiased representation of the system. To test this, we used the correlation test and residual histogram analysis method.

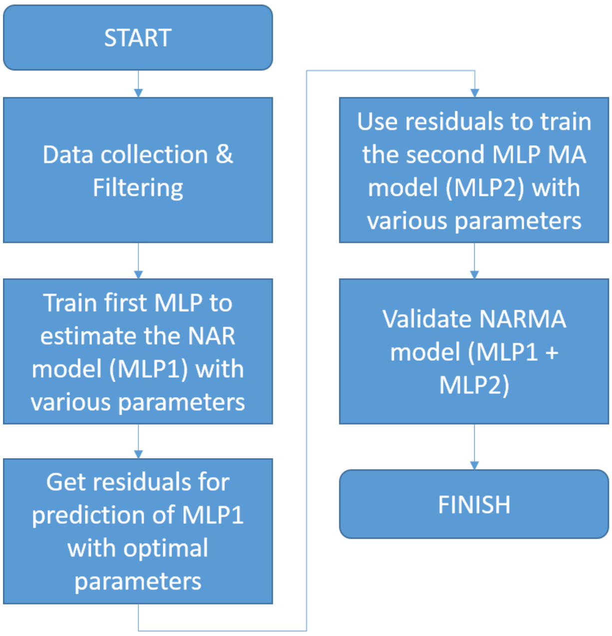

Figure 2: Experiment flowchart

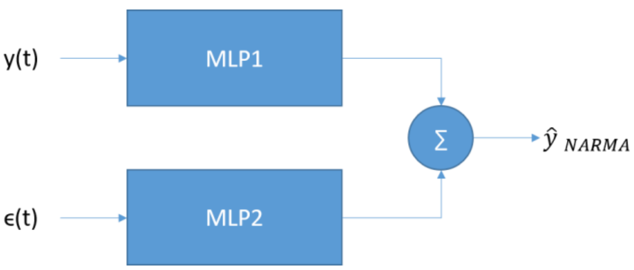

Figure 3: Block diagram of the identification process. MLP1 is responsible to represent the NAR part, and MLP2 is responsible to represent the MA part of the NARMA model

Correlation tests measure the correlation between two signals at different lags. In system identification, the residuals are tested against itself and other output signals. The model can be considered sufficiently acceptable if the correlation results are within the 95% confidence limits [38, 40]. Additionally, the distribution of the residuals can be tested using histogram analysis. The distribution of a white noise signal is similar to the Gaussian distribution (symmetric bell-shaped distribution).

IV. Results & Discussions

Both MLPs were tested with different combinations of lags and hidden units. A summary of the best fitting results are shown in Table 1. There were some notable variations of MSE results when the two parameters (lag and hidden units) were adjusted. This was because the lag space variations includes more terms as inputs to the MLPs. Additionally, the hidden units variations effect the learning ability of the MLPs to represent more complex patterns. The results indicate that the addition of MA terms to MLP1 and MLP2 had helped improve the model fit, thus justifying the used of the NARMA model.

The validation results are shown in Figure 4 to Figure 11. The residual plots show the residuals with small magnitudes, indicating small differences between the actual and predicted results. Additionally, the histogram tests shows a Gaussian

Table 1
MSE results for NAR and NARMA models

Params.		TRAINING MSE			TESTING MSE		
Lag	Hidden Units	NAR	NARMA	% Improvement	NAR	NARMA	% Improvement
5	25	0.0290	0.0224	22.7	0.1252	0.0768	38.7
10	5	0.2821	0.1856	34.2	0.2787	0.2687	3.6
	10	0.0626	0.0511	18.5	0.0659	0.0551	16.3
15	5	0.0306	0.0263	14.2	0.0384	0.0356	7.2
	20	0.0386	0.0361	6.4	0.1161	0.1090	6.1
25	5	0.0686	0.0435	36.5	0.1338	0.1148	14.3
	15	0.0161	0.0131	18.5	0.1321	0.1288	2.5

curve, and the correlation tests show that the correlation coefficients were within the 95% confidence limits (with a small number of violations). These two observations indicate that the residuals were sufficiently random and that the model is valid and acceptable.

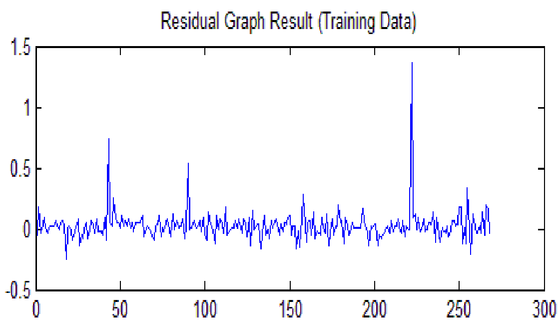

Figure 4: Residuals for training set (NARMA model)

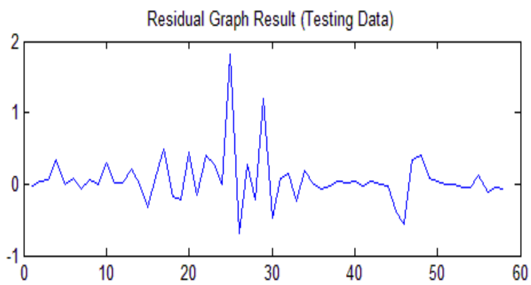

Figure 5: Residuals for testing set (NARMA model)

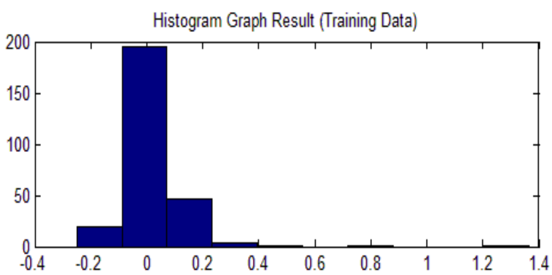

Figure 6: Histogram of residuals (training set, NARMA model)

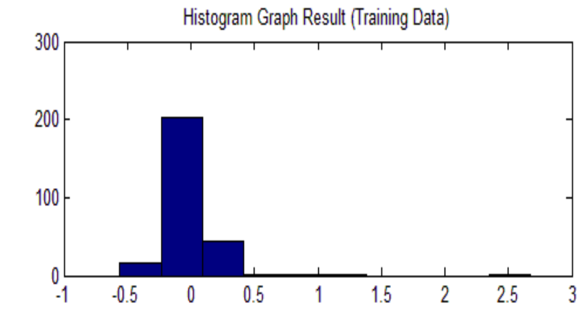

Figure 7: Histogram of residuals (training set, NARMA model)

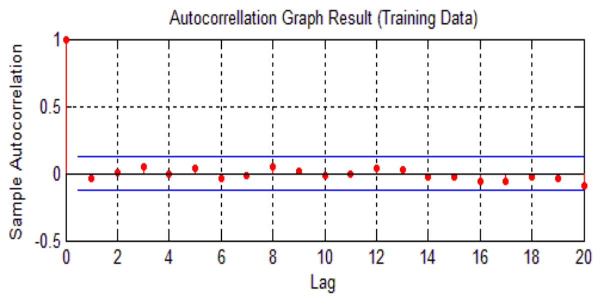

Figure 8: Autocorrelation of residuals (training set, NARMA)

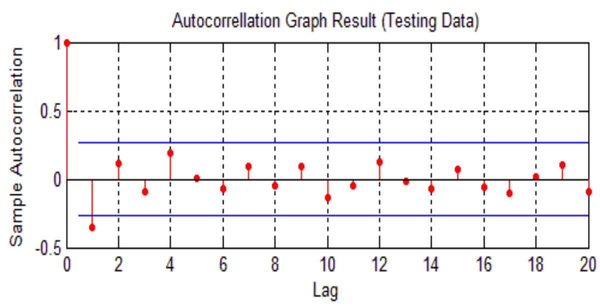

Figure 9: Autocorrelation of residuals (testing set, NARMA)

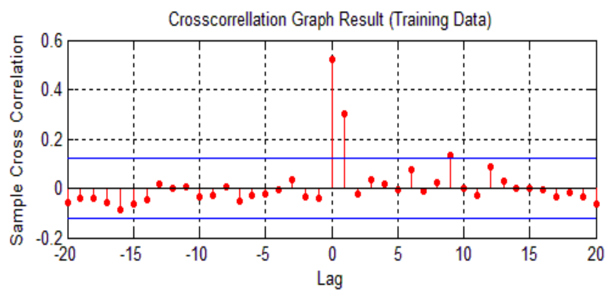

Figure 10: Crosscorrelation between residuals and output (training set, NARMA)

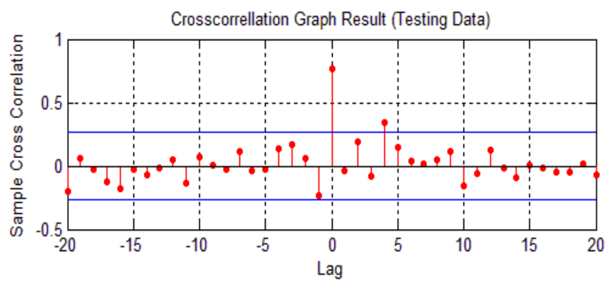

Figure 11: Crosscorrelation between residuals and output (testing set, NARMA)

V. CONCLUSION

This paper has presented a MLP-based NARMA model for internet traffic prediction. The results indicate a good agreement between the model predictions with the actual output. Additionally, the histogram and correlation tests indicate that the model was sufficiently random and acceptable.

ACKNOWLEDGMENT

We gratefully acknowledge the funding from the Faculty of Electrical Engineering, Universiti Teknologi Mara, Shah Alam for the presentation of this paper.

REFERENCES

[1] S., Koyano, S., Ata ,I., Oka, and K., Inoue, 2012. A High-Grained Traffic Prediction for Microseconds Power Control in Energy-Aware Routers. in Utility and Cloud Computing (UCC), 2012 IEEE Fifth International Conference on. 347-352.

[2] B. Tao, G. Shanqing, Z. Zonghua, R. Ando, and Y. Kadobayashi. 2011. Practical network traffic analysis in P2P environment. in Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International. 1801-1807.

[3] M.J.O., Gavade and M.P.K., Kharat, 2011. Neural network based approach for mpeg video traffic prediction . 6

[4] N., Zhenyu and G., Qing , 2011. An improved AQM scheme with adaptive reference queue threshold. in Communications and Networking in China (CHINACOM), 2011 6th International ICST Conference on. 589-593.

[5] G., Terdik and T., Gyires, 2008. Internet Traffic Modeling with Levy Flights. in Networking, 2008. ICN 2008. Seventh International Conference on. 468-473.

[6] K., Jian, Z., Dongwei, W., Xinyu ,W., Yanwen, and M., Qinghua, 2013. A network traffic prediction method using two-dimensional correlation and Single exponential smoothing. in Communication Technology (ICCT), 2013 15th IEEE International Conference on. 403-406.

[7] R.C., Jaiswal and Lokhande S. D., 2013. Machine learning based internet traffic recognition with statistical approach. in India Conference (INDICON), 2013 Annual IEEE. 1-6.

[8] Z.,Jun, C.,Chao, X.,Yang , Z.,Wanlei, and X.,Yong, 2013. Internet Traffic Classification by Aggregating Correlated Naive Bayes Predictions," Information Forensics and Security, IEEE Transactions on. 8:5-15.

[9] M.H.F. Rahiman., 2008. System Identification of Essential Oil Extraction System. Ph. D. Dissertation. Faculty of Electrical Engineering, Universiti Teknologi MARA. Shah Alam, Malaysia.

[10] S., Su, W., Zhang, and S.,Zhao. 2014. Online fault prediction for nonlinear system based on sliding ARMA combined with online LS-SVR. in Control Conference (CCC). 2014 33rd Chinese. 3287-3291.

[11] V.K., Dabhi and S., Chaudhary, 2014. Time Series Modeling and Prediction Using Postfix Genetic Programming," in Advanced Computing & Communication Technologies (ACCT). 2014 Fourth International Conference on. 307-314.

[12] Z., Run, C., Yinping, and F., Xing-an, 2012. A network traffic prediction model based on recurrent wavelet neural network. in Computer Science and Network Technology (ICCSNT). 2012 2nd International Conference on. 1630-1633.

[13] I., Yassin , 2008. Face detection using artificial neural network trained on compact features and optimized using particle swarm optimization. M. S. Thesis, Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam.

[14] T., Liu. 2012. Optimizing mining association rules based on Artificial Neural Network," in World Automation Congress (WAC), 2012. 1-4.

[15] F., Pilka and M., Oravec. 2011. Multi-step ahead prediction using neural networks," in ELMAR, 2011 Proceedings. 269-272.

[16] A., RajaRajan, 2011. Brain disorder detection using artificial neural network," in Electronics Computer Technology (ICECT). 2011 3rd International Conference on. 268-272.

[17] H., Chih-Chien, C., Wei-Chih , S., Chi-Yen, C., Yu-Ju, C., Chuo-Yean and H., Rey-Chue. 2010. Signal Processing by Polynomial NN and Equivalent Polynomial Function," in Pervasive Computing Signal Processing and Applications (PCSPA). 2010 First International Conference on. 460-463.

[18] S, Haykin. 1999. Neural Networks: A Comprehensive Foundation. Delhi, India: Pearson Education (Singapore) Pte Ltd.

[19] E.M.A.M., Mendes and S. A., Billings. 2001. An Alternative Solution to the Model Structure Selection Problem. IEEE Trans. Systems, Man & Cybernetics - Part A: Systems and Humans. 31(6):597-608.

[20] X., Hong , R.J., Mitchell , S.., Chen, C.,.J., Harris, K., Li and G.,Irwin. W., 2008. Model selection approaches for non-linear system identification: a review. Int. J. of Systems Science. 39(10):925-946.

[21] E. M. A. M , Mendes and S.A., Billings . 2001. An alternative solution to the model structure selection problem," IEEE Trans. Systems, Man and Cybernetics - Part A: Systems and Humans . 31(6):597-608.

[22] N., Chiras, C., Evans and D., Rees . 2001. Nonlinear gas turbine modeling using NARMAX structures. IEEE Trans. Instrumentation and Measurement. 50(4): 893-898,

[23] M.,Vallverdu, M.J., Korenberg and P., Caminal. 1991. Model identification of the neural control of the cardiovascular system using NARMAX models. presented at the Proc. Computer in Cardiology.

[24] S.L, Kukreja, H.,L., Galiana and R.E., Kearney, 2003. NARMAX representation and identification of ankle dynamics. IEEE Trans. Biomedical Eng. 50(1):70-81.

[25] S. L., Kukreja, R. E., Kearney and H. L Galiana., 2005. A least squares parameter estimation algorithm for switched hammerstein systems with applications to the VOR. IEEE Trans. Biomedical Eng. 52(3):431-444.

[26] N. A Rahim., M. N., Taib., and M. I., Yusof, 2003. Nonlinear system identification for a DC motor using NARMAX Approach. in Asian Conference on Sensors (AsiaSense). 305 - 311.

[27] K. K Ahn. and H. P. H., Anh. 2010. Inverse Double NARX Fuzzy Modeling for System Identification. IEEE/ASME Trans. on Mechatronics. 15(1):136-148.

[28] Cheng Y., Wang L., Yu M., and Hu J., 2011. An efficient identification scheme for a nonlinear polynomial NARX model," Artificial Life Robotics. 16:70-73.

[29] M., Shafiq and N.R., Butt . 2011. Utilizing Higher-Order Neural Networks in U-model Based Controllers for Stable Nonlinear Plants. Int. J. Control, Automation & Systems. 9(3):489-496.

[30] Z. H. Chen and Y. Q. Ni, "On-board Identification and Control Performance Verification of an MR Damper Incorporated with Structure," J. Intelligent Material Systems and Structures, vol. 22, pp. 1551-1565, 2011.

[31] B. H. G. Barbosa, L. A. Aguirre, C. B. Martinez, and A. P. Braga, "Black and Gray-Box Identification of a Hydraulic Pumping System," IEEE Trans. Control Systems Technology, vol. 19(2), pp. 398-406, 2011.

[32] B. Cosenza, "Development of a neural network for glucose concentration prevision in patients affected by type 1 diabetes," Bioprocess & Biosystems Engineering, pp. 1-9, 2012.

[33] M. A., Balikhin , R.J.,Boynton, S. N., Walker, J. E., Borovsky , S.A.,Billings and H.L.,Wei. 2011. Using the NARMAX approach to model the evolution of energetic electrons fluxes at geostationary orbit," Geophysical Research Letters. 38(L18105):1-5.

[34] H.L.,Wei and S.A.,Billings . 2008. An adaptive orthogonal search algorithm for model subset selection and non-linear system identification. Int. J. Control. 81(5):714-724.

[35] H.L., Wei and S.A., Billings .2008. Model structure selection using an integrated forward orthogonal search algorithm assisted by squared correlation and mutual information. Int. J. Modeling, Identification & Control. 3(4):341-356.

[36] L.A., Aguirre and C., Letellier . 2009. Modeling Nonlinear Dynamics and Chaos: A Review. Mathematical Problems in Engineering. 1-35.

[37] L., Piroddi and M.,Lovera . 2008. NARX model identification with error filtering. in Proc. 17th World Congress Int. Federation of Automatic Control. Seoul, Korea. 2726-2731.

[38] N. A. Jalil andI.Z.M., Darus . 2013. NARMA-L2 Vibration Controller for Flexible Structure with Non-Collocated Sensor-Actuator. presented at the Fifth International Conference on Computational Intelligence, Modelling and Simulation.

[39] Z., Muhammad , Z.M.., Yusoff , M.H.F, Rahiman and M.N.,Taib. 2012. Modeling of steam distillation pot with ARX model. in Signal Processing and its Applications (CSPA). 2012 IEEE 8th International Colloquium on. 194-198.

[40] S., Nomm, K., Vassiljeva, J., Belikov and E., Petlenkov. 2011. Structure identification of NN-ANARX model by genetic algorithm with combined cross-correlation-test based evaluation function. in Control and Automation (ICCA). 2011 9th IEEE International Conference on. 65-70.