Numerical Weather Prediction using Nonlinear Auto Regressive Network for the Manaus Region, Brazil

Geetali Saha
Faculty, Dept. of Electronics and Communication Engg.
G. H. Patel college of Engineering & Technology,
Vallabh Vidyanagar, Gujarat, India
geetalisaha@gcet.ac.in

Narendra Chauhan
Faculty, Dept of Information Technology,
A D Patel Institute of Technology,
New Vallabh Vidyanagar, Gujarat, India
narendracchauhan@gmail.com

Abstract— Weather prediction plays a very significant role in defining the various atmospheric changes that has been occurring from time immemorial over the surface of the earth. These changes, though gradual; their impacts are sudden. Amongst the parameters, temperature, wind speed and humidity are few of the significant ones. A model based on Non-linear Auto Regressive networks using Back Propagation Algorithm is trained using past data to predict numerical values for the future. The proposed algorithm tries to perform Temporal Regression using input weather data of the Manaus city over the years- 1970-2015. To check the performance, MSE, RMSE and NRMSE are tabulated. Results are also tabulated against variation in the ratio of training data: validating data: test data. Performance is found to improve when a larger ratio is used to train and validate the network.

Keywords—Weather Prediction; Temporal Regression; Non linear Auto Regressive Network; NRMSE.

I. INTRODUCTION

Weather forecasting has got a varied range of applications ranging from basic day to day needs, to agricultural plantation, reaping, harvesting, storing and even severe weather notifications. However, given the variability of the terrains, it is a challenge to predict its state for a given location for a given time. But, with the progress of science and technology, the availability of sensors and processing capabilities, significant progress has been observed in the domain of forecasting. The general procedure involves analysis of the past data and based on it; with scientific understanding the future numerical values can be predicted.

There are a variety of parameters that comprise the weather, the most significant ones being Temperature-both high and low, Relative Humidity, Wind Speed, Rainfall, etc. Weather predictions help to protect life and property by timely updating against severe changes. Many weather stations are installed at different locations to record the values of different parameters. These records are flow series data; in the sense that it is a measure of activity over a given period of time - Measurements taken multiple number of times in a day of the month are recorded and averaged over the duration of one day. This is how daily data is generated for every day.

Ratnadip and Agrawal [1] contain the dissertation work about time series data. Chapters ranging from basic time series models, their mathematical understanding, using preliminary methods like Moving Average, Auto Regressive Analysis to Artificial Neural Network based and Support Vector Machines are discussed in details. It also contains details of Forecast Performance Measures and web links for standard Databases, including the ones by Rob J Hyndman are provided.

In Mahalakshmi [2], a survey of most prevailing techniques involved in Time Series Analysis is discussed for Temporal Data Mining. Various applications are highlighted along with the corresponding favorable techniques. Forecasting techniques ranging from basic mathematical approaches like regression to soft computing techniques are all summarized for reference.

In Yassin [3], Non linear Auto Regressive Networks are explored to establish a relation between Tourism and Internet Searches. A model that works on optimizing resource allocation in face of Tourism is proposed and tested.

In Patrick [4], rainfall prediction in Manaus region of Brazil is done using multilayer artificial neural network. Out of the total available data, 70% is used for training neural network, 15% for validation and remaining 15% is used for testing. By testing different architecture of network and different transfer functions, normalized mean square error is calculated. The result shows that with 2 hidden layers having 10 neurons each and tangent sigmoidal transfer function, lowest error is obtained.

With reference to the database used in [5], a comparative of Basic Time Series Analysis methods is executed using MATLAB tool, version (R2015b). In this paper, we used meteorological data such as average high temperature, average wind speed and relative humidity collected from an automatic meteorological station from 1970 to 2015.

The present paper is organized as follows: Section II gives an overview of the time series formed using the NAR model for prediction and the equations for error calculations. Section III

978-1-5090-5682-8 /17/$31.00 ©2017 IEEE

shows the information about the data used for proposed work. Section IV is about the proposed algorithm. Section V is about the result obtained using the time series methods. Section VI gives the conclusion and suggestion for future work.

II. Time Series and the NAR model

A time series is essentially composed of a set of time stamped data, usually collected at regular interval of time. A time series is formed when successive observations are made at fixed intervals of time. Many time series data, ranging over many applications and various lengths is available for researchers and indexed in [4] for analysis. The Time Series Data Library (TSDL) was created by Rob J Hyndman, Professor of Statistics at Monash University, Australia.

For most of the Temporal Data, it is an average value, either on an hourly basis, daily basis or monthly basis- but whatever the duration, it remains the same over the entire series. Over long duration of time, by analysis, the hidden patterns of the series can be harnessed to generate new predicted values. Time series data occur naturally in many application areas. The aim of time series analysis is to describe and summarize time series data and make a forecast. Here, in site data is fed and simulated using software for data analysis. Later, the forecasted values are compared with actual values from the site to judge its accuracy.

A nonlinear auto regressive network is composed of D number of Delay elements apart from the regular Input, Hidden and Output layer associated with Neural Networks. Based on this D, a time series is constructed from the input fed. Variation in the number of D may lead to error variation and in most of the practical cases; a tradeoff is achieved between calculation time and overhead.

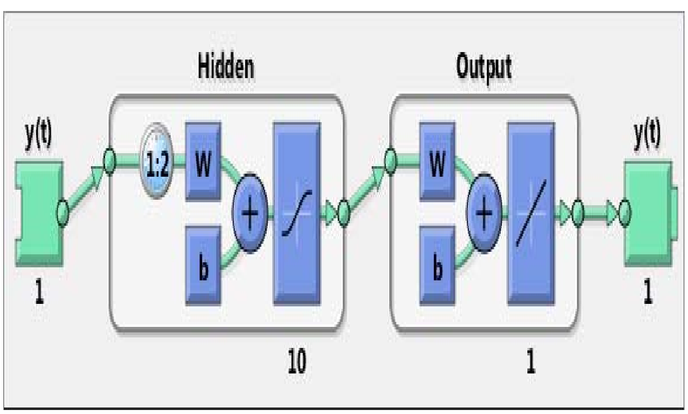

Figure 1**- A sample NAR model; composed of 10 hidden neurons and 2 delay elements [7].**

Error Measures

Whenever prediction is made, the actual value ought to be compared with the predicted value. Various error measures can be used for this. Following is the list of error parameters along with their equations for reference: Mean Square Error (MSE), Root Mean Square Error (RMSE) and Normalized Root Mean Square Error (NRMSE) all of which are individually defined as follows:

Mean Square Error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_t - F_t\right)^2 \qquad (1)$$

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_t - F_t\right)^2} \qquad (2)$$

Normalized Root Mean Square Error (NRMSE):

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_t - F_t\right)^2}}{Y_{\max} - Y_{\min}} \qquad (3)$$

$Y_{\max}$ = Actual Maximum value
$Y_{\min}$ = Actual Minimum value
Yt=Actual value of the variable
Ft= Forecasted value of the variable

III. Dataset Used:

The data under consideration is obtained from an automatic weather station which is composed by a central memory, called "data logger'', connected to several sensors of meteorological parameters. The data is connected every minute and is available automatically within every hour. Such station is hosted and maintained by the *Instituto Nacional de Meteorologia* (INMET) located at 3°6'13.2552"S, 60°0'6596"W, 61.25 m above sea level [6] placed in Manaus, Amazonas, Brazil.

From INMET database, we obtained data from Manaus from January 1st 1970 to January 1st 2015. In this period, total 16436 days of data is obtained but 3227 entries are discarded for being corrupted. So, total 13179 days of data is available for experiment. The parameters observed are as follows:

1) Average High Temperature
2) Average Low Temperature
3) Average Wind speed
4) Average Relative humidity

Figure 2**- The Manaus Region on the map of Brazil.**

IV. Proposed Algorithm

The proposed algorithm has user defined data divide algorithm

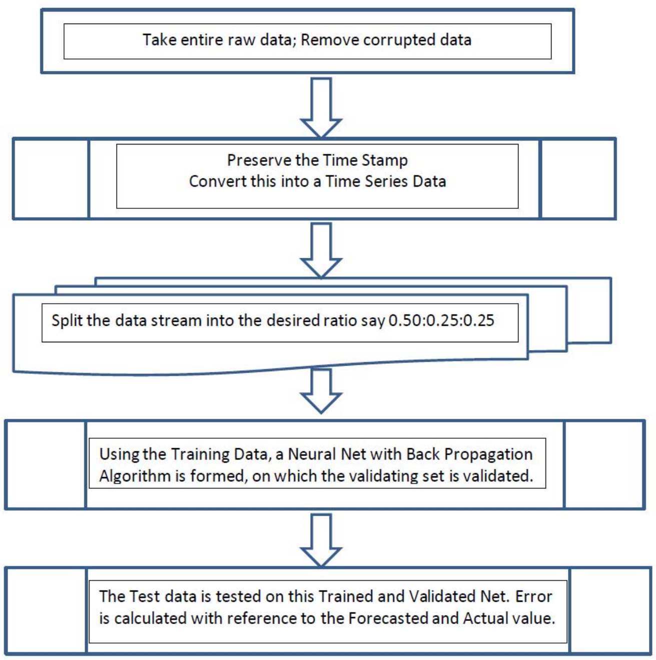

Figure 3**- The proposed algorithm**

V. Results

All the four parameters, whose observations are made are taken one at a time, and used to train and validate a net. Thereafter, the performance of a set of Test data is evaluated. For analysis, at times, keeping the value of Hidden Neurons constant, the number of Delay is varied and vice versa for a predefined ratio of data division. The cell with minimum error is highlighted.

Table-1 **Error comparative for average low temperature; Delay=1 neuron; Error parameters: MSE, RMSE, NRMSE; Data Division Ratio is 0.5:0.25:0.25**

Hidden neuron	Perf @ Epoch	MSE	RMSE	NRMSE
1	0.9143@ 6th	1.0231	1.0115	0.0472
2	0.9142 @ 6th	1.0225	1.0112	0.0473
4	0.9081@ 6th	1.0207	1.0103	0.0472
5	0.9063 @ 6th	1.0195	1.0097	0.0472
8	0.9054 @ 6th	1.0213	1.0106	0.0472
10	0.9054 @ 6th	1.0218	1.0109	0.0472

Table-2 **Result comparative for average high temperature; Error parameter: NRMSE; Data Division Ration is 0.5:0.25:0.25**

Hidden neurons	1	2	4	5	8	10
Delay of 1 neuron	0.1156	0.1155	0.1153	0.1157	0.1156	0.1155
Delay of 2 neurons	0.1296	0.1274	0.1276	0.1269	0.1268	0.1273
Delay of 4 neurons	0.1299	0.1285	0.1274	0.1275	0.1278	0.1275
Delay of 5 neurons	0.1302	0.1303	0.1289	0.1288	0.1283	0.1284
Delay of 8 neurons	0.1312	0.1312	0.1308	0.1309	0.1308	0.1305
Delay of 10 neurons	0.1324	0.1324	0.1322	0.1321	0.1314	0.1315

Table-3 **Result ccomparative for average relative humidity; Error parameter: NRMSE; Data Division Ration is 0.5:0.25:0.25**

Hidden neurons	1	2	4	5	8	10
Delay of 1 neuron	0.1269	0.127	0.1271	0.1269	0.1271	0.127
Delay of 2 neurons	0.1467	0.1469	0.146	0.1456	0.146	0.1456
Delay of 4 neurons	0.1484	0.1485	0.1476	0.479	0.1474	0.1466
Delay of 5 neurons	0.1474	0.1472	0.1471	0.1471	0.1466	0.1468
Delay of 8 neurons	0.1515	0.1517	0.1509	0.1503	0.1515	0.1511
Delay of 10 neurons	0.151	0.1505	0.1508	0.1515	0.1513	0.1512

Table-4 **Error comparative for average wind speed; Delay=8 neuron; Error parameters: MSE, RMSE, NRMSE; Data Division Ration is 0.5:0.25:0.25**

Hidden neuron	Perf @ Epoch	MSE	RMSE	NRMSE
1	1.0325@ 6th	0.3470	0.5891	0.0512
2	1.0292@ 6th	0.3468	0.5889	0.0512
4	1.0232 @ 6th	0.3486	0.5904	0.0513
5	1.0203 @ 6th	0.3491	0.5908	0.0514
8	1.0096 @ 6th	0.3523	0.5936	0.0516
10	1.0186 @ 9th	0.3510	0.5925	0.0515

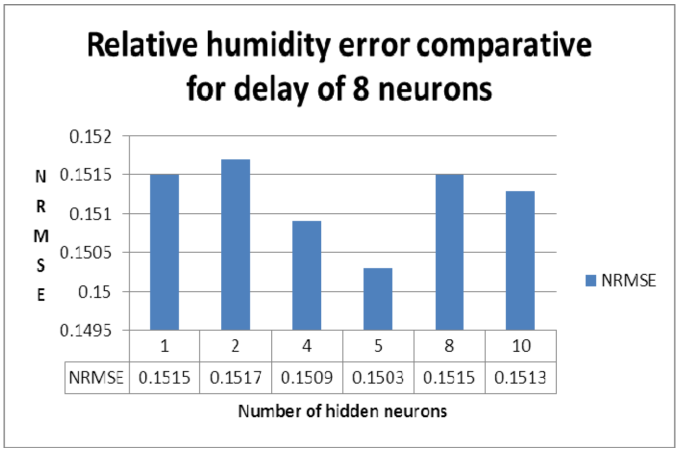

Figure 4: **Error comparative for relative humidity against variation in the number of hidden neurons; Delay=8 neurons; Error parameter: NRMSE; Data Division Ration is 0.5:0.25:0.25.**

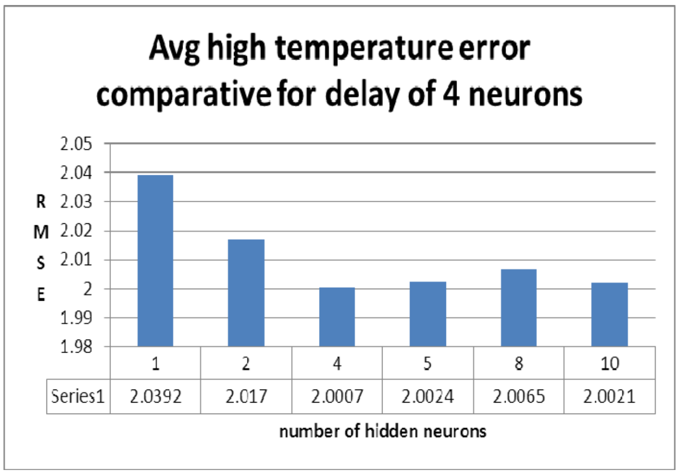

Figure 5: **Error comparative for average high temperature against variation in the number of hidden neurons; Delay=4 neurons; Error parameters: RMSE; Data Division Ration is 0.5:0.25:0.25.**

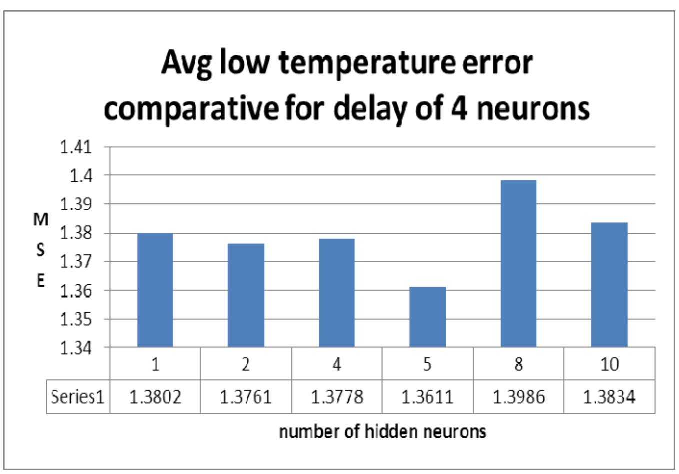

Figure 6: **Error comparative for average low temperature against variation in the number of hidden neurons; Delay=4 neurons; Error parameters: MSE; Data Division Ration is 0.5:0.25:0.25.**

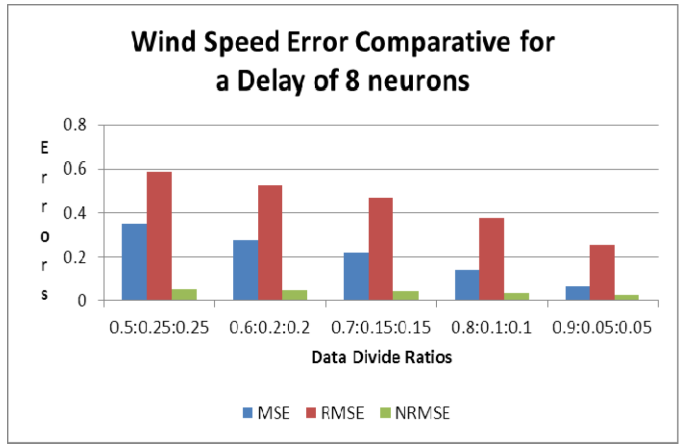

Figure 7: **Error comparative for average wind speed against variable data division ratios; Delay=8 neurons; Hidden Neuron=1; Error parameters: MSE, RMSE, NRMSE.**

VI. Conclusion

In this paper, different weather parameters such as average high temperature, average low temperature, humidity and wind speed are forecasted using Neural Auto Regressive Neural Network. It is observed that in general; 3, 4 and/or 5 hidden neurons are associated with minimum error.

In future, we intend to compare the results with other soft computing methods of Time Series Analysis.

ACKNOWLEDGMENT

The authors gratefully acknowledge the contributions of Patrick M. de Lima, Escola Superior de Tecnologia, Universidade Estadual do Amazonas, Manaus, Brazil for sharing the data with us.

REFERENCES

[1] R. Adhikari, R. Agrawal, “An introductory study on time series modeling and forecasting”, *arXiv:1302. 6613*, 2013.

[2] G. Mahalakshmi, Dr. S Sridevi and Dr. S. Rajaram.. “Development A survey on forecasting of time series data” ICCTIDE’16 , IEEE 7-9 Jan 2016.

[3] I. M.Yassin, A. Zabidi, M. K. M. Salleh, “Malaysian Tourism Interest Forecasting using Nonlinear Auto-Regressive (NAR) Model ”, IEEE 3rd International Conference on System Engineering and Technology, 19 - 20 Aug. 2013, Shah Alam, Malaysia.

[4] http://data.is/TSDLdemo last assessed on 11th Feb, 2017.

[5] de Lima, Patrick M. “Rainfall prediction for Manaus, Amazonas with artificial neural networks.” IEEE. *2015 Latin America Congress on Computational Intelligence (LA-CCI)*, 13-16 Oct, 2015.

[6] INMET, “Instituto nacional de meteorologia”, https://www.inmet.gov.br last accessed on 5th January, 2017

[7] www.mathworks.com last accessed on 5th February, 2017