Journal of Theoretical and Applied Information Technology
31st December 2016. Vol.94. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TIME SERIES FORECASTING FOR OUTDOOR TEMPERATURE USING NONLINEAR AUTOREGRESSIVE NEURAL NETWORK MODELS

SANAM NAREJO, EROS PASERO

Department of Electronics and Telecommunication, Politecnico Di Torino, Italy

E-mail: sanam.narejo@polito.it

ABSTRACT

Weather forecasting is a challenging time series forecasting problem because of its dynamic, continuous, data-intensive, chaotic and irregular behavior. At present, enormous time series forecasting techniques exist and are widely adapted. However, competitive research is still going on to improve the methods and techniques for accurate forecasting. This research article presents the time series forecasting of the metrological parameter, i.e., temperature with NARX (Nonlinear Autoregressive with eXogenous input) based ANN (Artificial Neural Network). In this research work, several time series dependent Recurrent NARX-ANN models are developed and trained with dynamic parameter settings to find the optimum network model according to its desired forecasting task. Network performance is analyzed on the basis of its Mean Square Error (MSE) value over training, validation and test data sets. In order to perform the forecasting for next 4,8 and 12 steps horizon, the model with less MSE is chosen to be the most accurate temperature forecaster. Unlike one step ahead prediction, multi-step ahead forecasting is more difficult and challenging problem to solve due to its underlying additional complexity. Thus, the empirical findings in this work provide valuable suggestions for the parameter settings of NARX model specifically the selection of hidden layer size and autoregressive lag terms in accordance with an appropriate multi-step ahead time series forecasting.

Keywords: *Artificial Neural network (ANN), multi-step ahead forecasting, Nonlinear Autoregressive (NARX) model, Outlier Detection, Time Series Prediction, Temperature forecasting.*

1. INTRODUCTION

During recent decades, several studies have been conducted in the field of weather forecasting providing various promising forecasting models. Weather forecasting usually depends on the models whose predictions are susceptible to chaotic dynamics. Weather forecasting is not only significant for individual’s everyday life schedule, but agriculture sector as well as several industries are also dependent on the condition of the weather. In scientific research of metrology, the weather forecasting is typically an unbiased time series forecasting problem. Time series forecasting is in fact an expanding field of interest which is playing a significant role in almost all fields of engineering and science.

A time series is an ordered sequence of data samples which are recorded over a time interval. Time series are largely used in any domain of applied science and engineering which involves temporal measurements. Time series data includes a variety of features. For instance, few of data series may posses seasonality, few reveal trends, i.e, exponential or linear and some are trendless, just fluctuating around some level. In response to this, some preprocessing needs to be done to handle these features. In order to extract meaningful statistics, other characteristics of the time series data and to predict future values based on previously observed values, time series analysis and forecasting models are implemented.

The forecasting domain for a long time has been influenced by traditional linear statistical methods for example, ARIMA models or Box-Jenkins[1-5]. However, these methods are totally inappropriate if the underlying mechanism is nonlinear. The real life process is mostly nonlinear which brings deficiency in traditional mathematical model. Later in 70’s and early 80’s, after the realization that linear models may not adapt to the real life process several useful nonlinear models were developed [6][7].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The metrological data is irregular and follows a nonlinear trend. Weather forecasting is a challenging time series forecasting problem because of its dynamic, continuous, data-intensive, chaotic and irregular behavior. Recently, several studies have been conducted in the field of weather forecasting providing numerous promising forecasting models. However, the accuracy of the predictions still remains a challenge.

Generally, one step ahead forecasting is performed by making the use of current and observed values of a particular variable to estimate its expected value for the next time step following the latest observation. Predicting two or more steps ahead is considered as a multi-step ahead prediction problem, which is often denoted by h-step ahead prediction. Where h corresponds to the predicting horizon [8]. Multi-step ahead forecasting is more challenging task than one step ahead prediction. The complexity in forecasting increases as the horizon to be predicted is expanded. Because the errors in each estimation are accumulated. Consequently, the reduced accuracy and increased uncertainty degrade the performance of forecasting model.

In fact, the formulation and preparation of a nonlinear model to a specific type of data set is a very challenging assignment as there are too many unknown nonlinear patterns and a specified nonlinear model may not be sufficient to acquire all the significant representations or features. Artificial neural networks, are actually nonlinear data-driven approaches as opposed to the statistical model-based nonlinear methods and are capable of performing nonlinear modeling without a priori knowledge about the relationships between input and output variables. Therefore, they are a more general and flexible modeling tool for forecasting.

An ANN (Artificial Neural Network) model takes input and produces one or more output. In between input and output variables, the ANN does not require any presumption on logical or analytical forms [9]. It's mapping ability is achieved through the architecture of developed network and training of its parameters with experimental data. A neural network gains the knowledge over the system dynamics by examining the patterns between input data and corresponding outputs, and becomes able to use this knowledge to predict a system's output [10].

An application of neural networks in time series forecasting is based on the ability of neural networks to approximate nonlinear functions. ANNs are able to learn the relationship between input and output through training. However, the other models are either mathematical or statistical. As mentioned earlier, these models have been found to be very accurate in calculations, but not in predictions as they cannot adapt to the irregular patterns of data which can neither be written in the form of function nor can be deduced in the from a formula.

The purpose of this research is to establish and train a network that can well predict and forecast the weather component temperature with optimization of neural network parameters. In order to achieve this, Several ANN models were trained with different parameter settings. The ANN models, trained as predictive models were converted into forecaster for multi step ahead predictions in future interval of 3 different horizons 4, 8 and 12. This indicates that the forecaster was extrapolated to predict outdoor temperature accurately for the next one hour, two hours and three hours ahead. The performance of the forecasting model is further investigated on four, separate unseen data sets. The model that outperforms other models on the basis of its performance accuracy on the test sets is finally selected as the accurate forecaster for multi-step temperature prediction. However, the adopted methodology can be considered general and applicable to different and larger sets of meteorological parameters. For example, the similar procedure can be further implemented for humidity, pressure and rain forecasting.

The paper is organized as follows. Related work for time series prediction and forecasting using Machine learning and computational intelligence tool is summarized in Section 2. Section 3 demonstrates the adopted research methodology in detail which consists of different steps to be followed. The results along with appropriate findings and discussion are reported in Section 4. Section 5 provides the conclusion of the current study and possible hints for future work.

2. RELATED WORK

This section summarizes some previous related work in the realm of time series prediction and forecasting. Machine learning approaches and several soft computing techniques have been used on high scale for various weather forecasting applications. In scientific research, weather forecasting is an unprejudiced time series forecasting problem. The literature review suggests

Journal of Theoretical and Applied Information Technology
31st December 2016. Vol.94. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the presence of various models available for time series analysis, prediction and forecasting. Time series forecasting based on Computational Intelligence techniques generally falls into two major categories: (1) based on ANN and (2) based on evolutionary computing.

The researchers in [11] provided pioneering study, as the first attempt to model nonlinear time series with ANNs. It was demonstrated in their study that the achieved accuracy of ANN models overtook the conventional methods, including the Linear Predictive Method and the Gabor-Volterra-Weiner Polynomial Method. This indicates that the ANN model outperforms the previously mentioned methods.

The researchers in [12] conducted univariate time series forecasting using feedforward neural networks for two benchmark nonlinear time series. The empirical study on multivariate time series forecasting is presented in [13]. Major development in ANN models progressed with the use of ensemble modeling and hybrid approaches as demonstrated in [14],[15],[16].

Taking into consideration, the performance of ANN models for time series forecasting, the focus here is specifically on the weather forecasting of temperature series with ANN approach. In [17],[18] temperature and humidity forecasting is inspected by proposing a "local level" approach, based on time series forecasting using Type-2 Fuzzy Systems.

The research in [2][19] present that a hybrid technique can be used to further decomposes a time series data into linear and nonlinear form for further modeling. For example, for seasonal time series, firstly the seasonal component is removed by a linear model, such as a seasonal autoregressive model and subsequently the further analysis is conducted. The authors in [20] proposed a novel approach of an ensemble neural network for weather forecasting of Saskatchewan Canada. In their research, they further provided a comparative analysis of their proposed ensemble model with Multi Layer Preceptron (MLP) Elman Recurrent Neural Networks (RNN), Radial Basis Function Networks (RBFN) and Hopfeild model. In spite the fact that their proposed ensemble neural network proved to be dominant, on the contrary, the researchers in [21] suggested that ensemble of linear and non-linear components do not necessarily exceed than constitutional models. They specify that such combinations carries danger of underestimating the relationship between the models linear and nonlinear mechanisms.

Large scale comparative study of major machine learning models for time series forecasting is conducted in [22]. The study is based on the benchmark dataset of Monthly M3 time series competition data. The findings in this study suggest first MLP and then Gaussian Process (GP) as two best models for business type time series.

3. RESEARCH METHODOLOGY

The main motivation for this research work is to compute multi-step ahead forecasting using ANN models. Therefore, in order to forecast the temperature for multi steps ahead prediction the flow of work is presented in Figure 1. Initially, the temperature series data were downloaded from the METEO weather station. Subsequently, the dataset was further preprocessed with filtering and feature extraction step. The NARX ANN models were developed and trained on the extracted feature set with different parameter settings. For multi step ahead forecasting these trained models were further extrapolated in the form of close loop networks as forecasters.

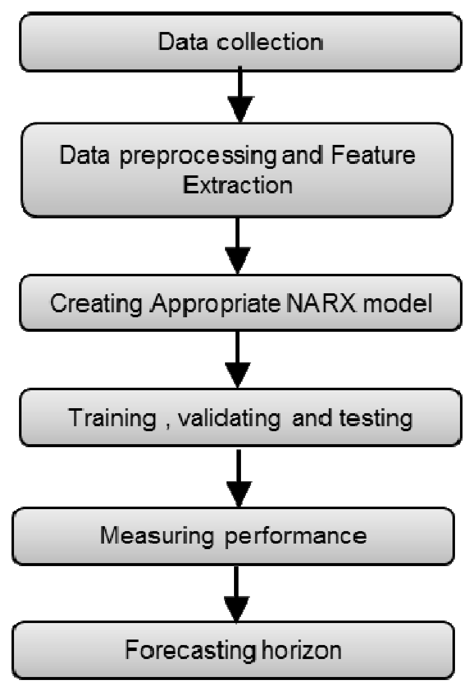

Figure 1. Flow of Methodology

Journal of Theoretical and Applied Information Technology
31st December 2016. Vol.94. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.1. Data Collection

The data for this study is collected from the weather station installed at Meteo station of Neuronica Laboratory at Politecnico Di Torino. The sensors provide a new recording after every fifteen minutes. The dataset was downloaded which contains the records from 4 October 2010 to 3 September 2015.

3.2. Data Preprocessing and Feature Extraction

The data recorded through sensors may have noise, some of missing samples and unwanted frequency fluctuations. In order to detect the outliers and to remove sensor noise, some of the pre-processing in the form of filtering has been done on the data prior to considering it as an input to train the ANN model. Figure 2 demonstrates the time series of temperature data with the missing samples and the detected outliers as identified in red cirles.

The recorded data series was filtered with a Butterworth filter of order 2. Filtering was performed in order to remove sharp fluctuations residing in data patterns and to smooth the time series in order to train the model for accurate predictions. The actual and filtered data can be seen in Figure 3 and Figure 4. It is clearly depicted from Figure 4 that sharp edges and noisy fluctuations present in the recorded data are removed and smoothened by applying a filter of an appropriate order.

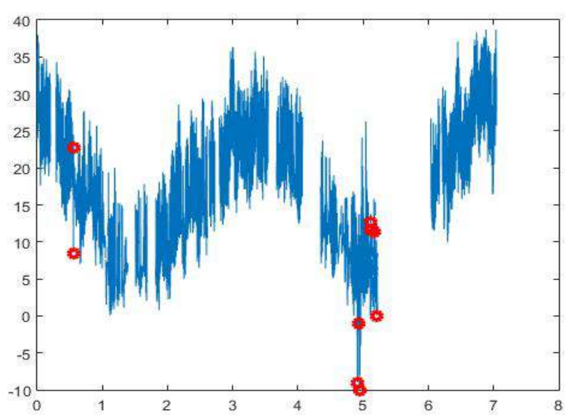

Figure 2. Temperature Time Series

Afterwards, the filtered data is normalized in the range of (-1 1). It is often useful to scale the inputs and targets so that the parameters with different unit values fall within a specified standard range. On the other hand, the use of original data as input to the network model without performing any normalization may cause a convergence problem [23]. The primary goal of normalization, is to allow the squashed activity function to work at least at the beginning of the learning phase when the weights are initialized. Therefore, the derivative of nonlinearity, i.e, the gradient will always be different from zero. The input features chosen as attributes for temperature prediction, were selected as, a month, an hour and minutes of each recorded sample.

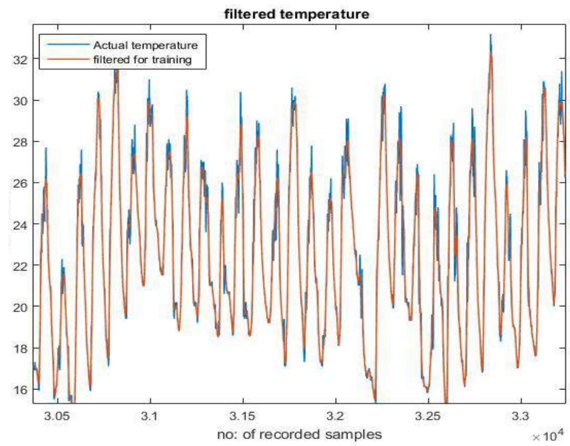

Figure 3. Actual and filtered Temperature Series

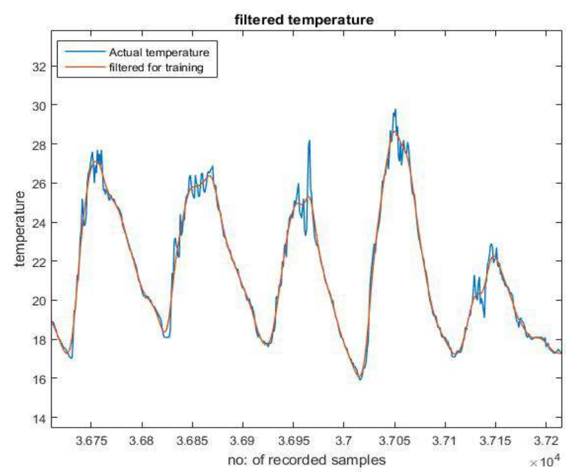

Figure 4: Close View Of Filtered Temperature Data

3.3. ANN Model

In order to predict the future values, choice of ANN opted is Non-linear Autoregressive neural network. NARX stands for Non-linear AutoRegressive with eXogenous inputs. The NARX model is derived from autoregressive models. Weather data ensembles highly nonlinear, stochastic and chaotic behavior in its nature. As it changes with time it has also an element of linear dependency with its own value on previous time

Journal of Theoretical and Applied Information Technology
31st December 2016. Vol.94. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

step, which is the property of time series data known as auto-regression.

The term auto-regression indicates that it is a regression of the variable against itself. In consideration to this, the NARX model intakes the past target values as feedback delays along with the exogenous inputs. NARX networks have a limited feedback which arises only from the output neuron, rather than from hidden states [12]. These models are formalized as expressed in (1).

$$y(t)=f(u(t-nu),\dots,u(t-1),u(t),y(t-ny)\dots,y(t-1)) \quad (1)$$

where u(t)and y(t) represent input and output of the network at time t. The variables nu and ny are the lagged terms of input and output order and the function f is the mapping performed by a Multilayer Perceptron. Literature suggests that NARX networks typically converge much faster and generalize better than rest of the recurrent architectures [24]. NARX network performs better on problems involving long term dependencies [25],[26]. Therefore, in current study this type of recurrent neural network has been chosen as justified choice of model. According to our research work objective, the problem of forecasting for D time steps ahead can be estimated as presented in (2).

$$\hat{y}(t+D)=f(y(t),\dots y(t-dy),\dots,u(t),y(t-du)\dots,y(t-1)) \quad (2)$$

Where D is 4, 8, and 12 in our study, which is based on the frequency of recorded samples collected from the sensor and the forecasting horizon. The temperature sensor is configured to record a new value after every 15 minutes. To forecast for the next one hour, the horizon consists of 4 steps. Similarly for 2 hours it is next 8 steps and for 3 hours the prediction is next 12 samples.

Backpropagation was used as a training method to train ANN models. The simple practice of backpropagation algorithm is to update the network weights and biases in which the performance function decreases most expeditiously i.e. the negative of the gradient. Normally, gradient g is defined as the first-order derivative of the total error function as presented in (3). To train the network, one complete iteration of the algorithm is indicated in (4).

$$g = \frac{\partial E(x,w)}{\partial w} = \left[\frac{\partial E}{\partial w_1} \; \frac{\partial E}{\partial w_2} \cdots\cdots \frac{\partial E}{\partial w_N}\right]^T \quad (3)$$

$$\theta_{k+1} = \theta_k - \alpha_k g_k \quad (4)$$

Where θ_k is a vector of current weights and biases, g_k is the current gradient, and $\alpha_k$is the learning rate. If the learning rate sweeps too high, then the training may oscillate and become unstable. Contrary to this, if the learning rate is too small, the algorithm will take too long to converge. Practically, it is challenging to decide the optimal setting for the learning rate before training. In fact, the optimal learning rate changes during the training process, as the algorithm moves across the performance surface [27]. The performance of the steepest descent algorithm can be further enhanced in case that the learning rate is granted some flexibility of being adaptive during the training process. An adaptive learning rate attempts to keep the learning step size as large as possible while keeping learning stable [28].

3.4. Experimental Setup

As it is explained in an earlier section that for choosing an appropriate forecasting model, different combinational settings were applied to the parameters of the models. This was performed, in order to select the accurate forecaster for 1 hour, 2 hours and 3 hours with the horizon of 4, 8 and 12 respectively. For temperature prediction input features were selected as corresponding month, an hour and minutes. A certain section of data was used for training the network models and the rest of the data was used to assess the capability of neural network forecasters to generalize. The rest of the data series was divided into four separate sets for investigating the performance of each forecaster.

The NARX model is trained with a total number of 700k samples. This was further divided in the training set, validation and testing with the ratio of 70%, 15% and 15% respectively. Selected options for the size of the hidden layer consist of 5, 7,8,10 and 15. The literature suggests different rules for selecting the size of the hidden layer for ANN model, but the hidden layer size also depends on the existence of complexity present in data patterns. Therefore, a grid search method was applied on the range of values mentioned above to set the hidden layer size. The hidden layers were activated by a sigmoid function, whereas, linear function was applied on output layer neuron. The schematic view of a generalize ANN model is illustrated in Figure.5.

The dimension of an input layer depends on a number of attributes present in the formulated feature set. The size of hidden layer corresponds to the settings applied with a grid search method. For

Journal of Theoretical and Applied Information Technology
31st December 2016. Vol.94. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the learning of a model, to understand the patterns and structure of temperature time series the model was trained to predict the corresponding temperature sample. Therefore, output layer consists of one neuron.

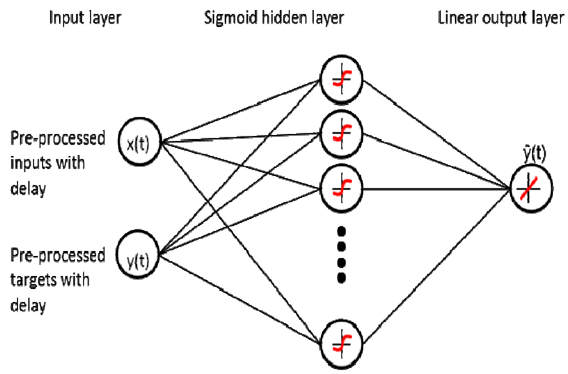

Figure 5. Schematic Diagram For ANN Training

The correct combination and selection of lag terms also places strong impact on proper forecasts. Therefore delay variable is assigned with 1:2 or 1:4 input and feedback delays. Fixed values are set for other parameters to train the network as presented in Table 1. As it is visible from the Table 1, the maximum number of iterations specified is 1000 epochs. The training procedure is terminated early, if the network performance on the validation vectors fails to improve or remains the same for 6 epochs in a row. Secondly, if the performance gradient falls below the minimum performance gradient or µ exceeds the maximum specified threshold the training will terminate. The four possibilities of µ factor are also specified in the earlier mentioned Table.

Test vectors are used as a further check to ensure that the network is generalizing well, but they do not effect on training. Once the network is trained to predict the temperature, it is converted in close loop system from open loop series-parallel to predict multi-steps ahead. In response to this the model is able to perform forecasting task for specified horizon. The generalize structure for NARX forecaster is shown in Figure 6. The rational forecaster is further tested on available historical data to investigate which scenario settings work best for 1 hour, 2 hours and 3 hours future forecasts. Table 2 demonstrates the captured results of trained models in the form of error measurements which are discussed in the next section.

Table 1: Network Parameter Settings

Parameters	Settings
Maximum number of epochs to train	1000
Minimum performance gradient	1e-7
Initial mu	0.001
µ decrease factor	0.1
µ increase factor	10
Maximum µ	1e10

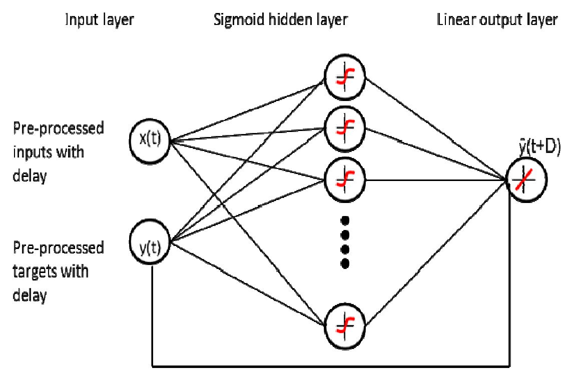

Figure 6: Schematic Diagram For ANN Forecaster

4. RESULTS AND DISCUSSION

The aim of our research was to accurately predict the outdoor temperature taken from the Meteo weather Station, Neuronica Laboratory. In order to perform this, the followed strategy is based on the consideration of several numbers of trained models. The model with the best accuracy measures was chosen as a forecaster for further deployments. The performance of each model is based on MSE which can be computed as expressed in (5).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\epsilon_i)^2 = \frac{1}{N}\sum_{i=1}^{N}[y(t) - \hat{y}(t)]^2 \quad (5)$$

Table 2 encapsulates the obtained results of eight different trained ANN models and its corresponding forecaster performance on unseen data sets is further demonstrated in Table 3. For forecaster, four different data sets were randomly taken from historical data to monitor satisfactory performance of the trained model in form of generalization. The performance of the model was evaluated on MSE values. The performance of model on training, validation and testing is far better as compared to its corresponding forecasting model. Forecaster performs meticulous job on 1 hour forecasting rather than on 2 and 3 hour forecasts. Hence the forecaster performance degrades as the duration of time for future predictions increases. This happens

Journal of Theoretical and Applied Information Technology
31st December 2016. Vol.94. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

due to increased accumulation of errors as each step estimation is computed.

The Delay parameter of the model essentially refers to the lag term in time series data. Lag terms are previous number of steps in time series data that are significantly related with any particular term. In general, the autocorrelation is higher for lag steps in the immediate past and decreases for observation in the distant past. Therefore, the input delay is kept constant having value as 2 but on the other hand feedback delay varies in between 2 and 4 delay terms for each model. It is illustrated in Table 2 that this delay setting provides much better results as compared to keeping both input and feedback delay as 2. Feedback delay steps essentially hold the values of the output variable.

4.1. Forecasting 4-steps ahead

The first model which is presented in both Tables 2 and 3 is trained with 5 hidden neurons and 2 lagged terms for both input and feedback. The rest of parameter settings were demonstrated in Table 1. These settings remain constant for each model. The training performance of first model is worthy, but as a forecaster, the network was not able to learn future patterns accurately as identified by corresponding MSE value which is summarized in Table 3.

Figure 7 plots the errors of 1 hour ahead forecasting for 4 different test sets later simulated on eight trained models specified in Table 1. The MSE value computed for each of eight models on four different unseen datasets is clearly visible in the Figure. The preferable model for 1 hour and 4 steps ahead prediction is with 15 hidden layers 2 input delays and 4 feedback delays as it is having minimum MSE on each four datasets. It is visible from the graph that this model results in considerably lower value of MSE on all randomly taken test sets as compared to other network models. The estimated temperature forecasts for 1 hour, on the first test set computed from outperforming model are shown in Figure 8.

It can be observed from the Table 3 that the second preferable model for 1 hour is NARX with hidden layer size of 10 neurons 2 input delays and 4 feedback delays. The error rate on test sets for both of the models is quite similar, albeit with a slight difference.

4.2. Forecasting 8-steps ahead

Forecasting performance for 8-steps ahead for next 2 hours onward for each test set are shown in Figure 9. It is visible from the Figure that the model that results in least possible MSE on all 4 data sets is with 7 hidden neurons, 2 input delays and 4 feedback delays. The forecasted temperature in correspondence with the original samples is further exposed in Figure 10.

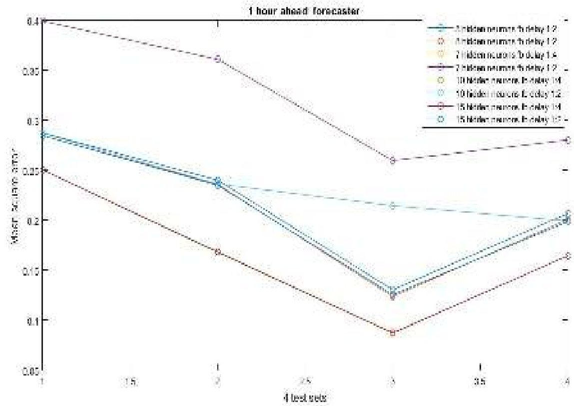

Figure 7: MSE Performance For 1 Hour Forecast On Four Test Sets

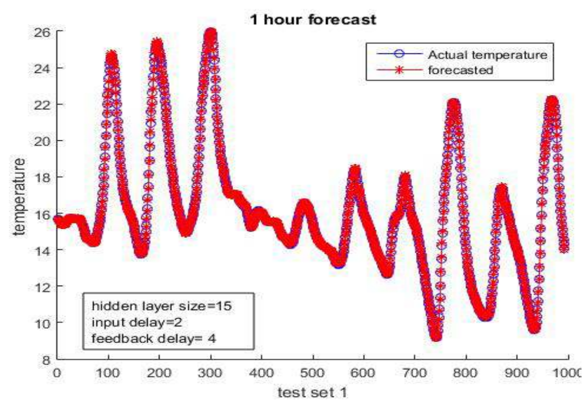

Figure 8: Temperature Forecast For 1 Hour

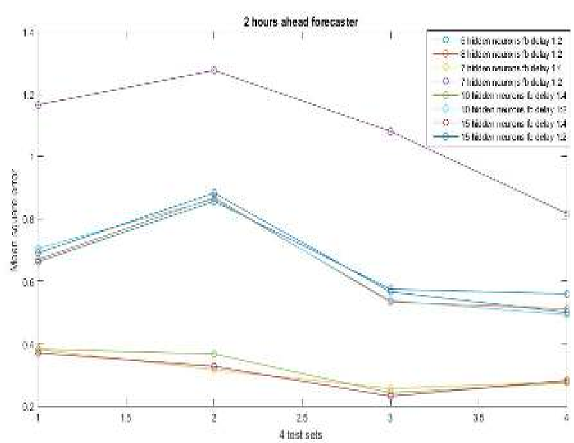

Figure 9: MSE Performance For 2 Hour Forecast On Four Test Sets

Journal of Theoretical and Applied Information Technology
31st December 2016. Vol.94. No.2

ISSN: **1992-8645** www.jatit.org E-ISSN: **1817-3195**

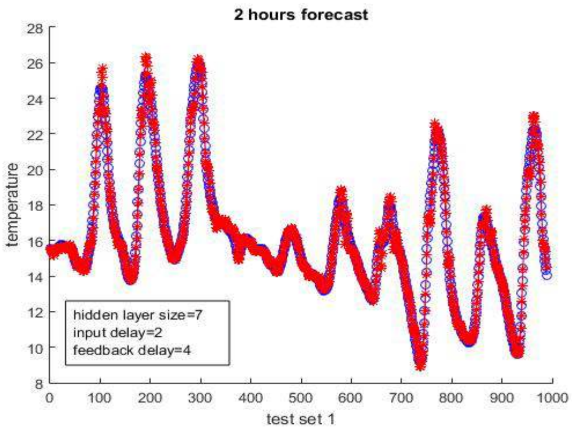

Figure 10: Temperature Forecast For 2 Hours

4.3. Forecasting 12-steps ahead

In this case, for predicting 3 hours in advance, which is equivalent to next 12 steps ahead estimations. From Table 3, it can be observed that hidden layer size 7 and 10 outperforms for 3 hours in the future as compared to the rest of the other models. Apart from this, the MSE on each of four test sets produced by all models for 3 hour forecast is shown in Figure.11. The temperature forecasts for 3 hours on both previously mentioned models on first test set are shown in Figure 12 and 13.

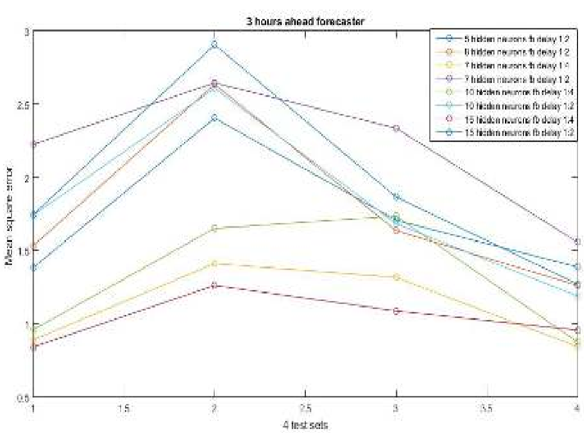

Figure 11: MSE Performance For 3 Hour Forecasting On Four Test Set.

The reported results indicate that the NARX forecasting model performs better on predicting long term future values if layer size contains more hidden neurons. On the other hand it does not predict accurate values for immediate terms. If the hidden layer size is not sufficient, the forecaster presents much higher accuracy for short term future values instead on longer horizons.

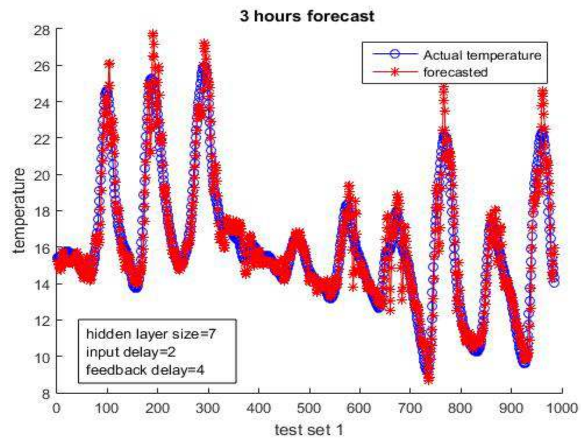

Figure 12: Temperature Forecast For 3 Hours

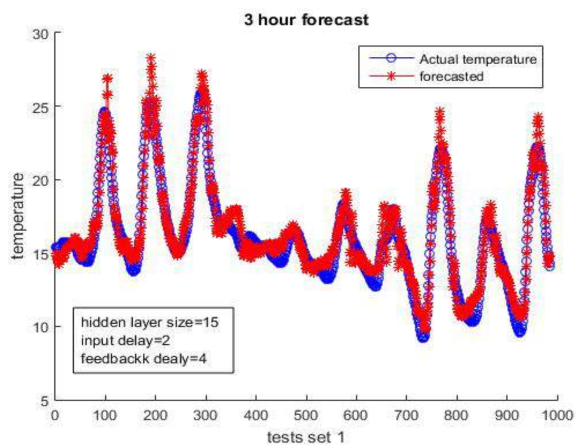

Figure 13. Temperature Forecast For 3 Hours

The satisfied configuration of neural network model highly depends on the problem. Due to this fact for each forecasting horizon, NARX model with different configuration performs accordingly. According to the authors in [6], there exist five different model follow up strategies for multi-step ahead predictions. The strategy applied in the current work lies in the category of Direct forecasting approach.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. CONCLUSION

In this work, outdoor temperature forecasting was performed using the NARX-ANN approach. Various time series dependent Recurrent NARX-ANN models were developed and trained with dynamic parameter settings to find the optimum network model according to their desired forecasting task.

The primary goal of this work was to contribute to the study and development of multi-step ahead forecasting with NARX modeling. Forecasting was computed considering the horizon of 2, 8 and 12 steps for the next 1 hour, 2 hours and 3 hours respectively in the future. The satisfied configuration of neural network model highly depends on the problem. Therefore, several NARX models were developed, trained and further extrapolated as forecasters. For each forecasting horizon, eight different configurations were considered with some optimal parameter settings.

On the whole, the general assumptions based on the current study, findings indicate that the developed forecasting model performs better in predicting long term future values if the layer size is higher in dimensions and contains more hidden neurons. Additionally, the proper lag term selection must be performed to compute input and feedback delays for specific model. This means that for multi-step ahead forecasting, when the range of horizon is longer, there should be sufficiently big hidden layer size in order to accurately forecast the long term estimations. On the other hand, if the horizon is wider in range and hidden layer size selection is not enough, the network model may not be capable to forecast long-term predictions. This indicates that in case if the hidden layer size is not sufficiently higher, but the model is well trained, the corresponding forecasting model might be better on short term multi-step forecasting rather than longer ones. Consequently, the models may present much higher accuracy for short term future values instead of longer horizons.

The time series is basically a sequential data with temporal changes. To improve the performance of forecasting model, it is necessary to understand the mechanics of time series required to be predicted. The results achieved from different forecasting models also suggest the importance of finding the appropriate correlations with input indicators or attributes. This indicates that the proper selection of lagged terms enhances the model credibility for accurate forecasts. Depending upon this, the delay terms must be selected and taken into consideration. The observations concluded from the current research demonstrated that the proposed approach is promising and can be further applied to the multi-step ahead prediction of a different set of weather parameters such as humidity, pressure and rain.

ACKNOWLEDGMENTS

Computational resources were partly provided by HPC@POLITO, (http://www.hpc.polito.it).

This project was partly funded by Italian MIUR OPLON project and supported by the Politecnico of Turin NEC laboratory.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2. Performance results of trained models

S.No	No. of neurons in hidden layer	Open loop training results	
1.	Hidden layer = 5 Delay=1:2 Feedback=1:2	Performance	6. 2124e-04
		Training	6. 1970e-04
		validation	6. 2647e-04
		Testing	6. 2322e-04
2.	Hidden layer = 8 Delay=1:2 Feedback=1:2	Performance	5. 6132e -04
		Training	5. 6287e-04
		validation	5. 875e-04
		Testing	5. 668e-04
3.	Hidden layer = 7 Delay=1:2 Feedback=1:4	Performance	4. 3382e-06
		Training	4. 3454e-06
		validation	4. 5510e-06
		Testing	4. 0916e-06
4.	Hidden layer = 7 Delay=1:2 Feedback= 1:2	Performance	5.5053e-04
		Training	5.4484e-04
		validation	5.6409e-04
		Testing	5.6355e-04
5.	Hidden layer = 10 Delay=1:2 Feedback= 1:4	Performance	4. 3506e-06
		Training	4. 3909e-06
		validation	4. 3909e-06
		Testing	4. 2819e-06
6.	Hidden layer = 10 Delay=1:2 Feedback= 1:2	Performance	5. 4976e-04
		Training	5. 5373e-04
		validation	5. 4213e-04
		Testing	5. 3888e-04
7.	Hidden layer = 15 Delay=1:2 Feedback= 1:4	Performance	4.1754e-06
		Training	4.1467e-06
		validation	4.3938e-06
		Testing	4.0908e-06
8.	Hidden layer = 15 Delay=1:2 Feedback= 1:2	Performance	5.2852e-04
		Training	5.2203e-04
		validation	5.6943e-04
		Testing	5.1792e-04

Journal of Theoretical and Applied Information Technology
31st December 2016. Vol.94. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3. Performance Result MSE Of Forecasters On Test Sets

S.No	No. of neurons in hidden layer	Close loop forecaster performance	Test results for 1 hour forecast	Test results for 2 hour forecast	Test results for 3 hour forecast
1.	Hidden layer = 5 Delay=1:2 Feedback=1:2	Test set 1	0. 2874	0. 6617	1. 3839
		Test set 2	0. 2402	0. 8560	2. 4053
		Test set 3	0. 1305	0. 5739	1. 7073
		Test set 4	0. 2070	0. 5591	1. 3878
2.	Hidden layer = 8 Delay=1:2 Feedback=1:2	Test set 1	0. 2848	0. 6685	1. 533
		Test set 2	0. 2360	0. 8678	2. 6368
		Test set 3	0. 1238	0. 5327	1. 6364
		Test set 4	0. 2009	0. 5106	1. 2616
3.	Hidden layer = 7 Delay=1:2 Feedback=1:4	Test set 1	0. 2506	0. 3804	0. 8895
		Test set 2	0. 1684	0. 3155	1. 4115
		Test set 3	0. 0876	0. 2553	1. 3168
		Test set 4	0. 1646	0. 2758	0. 8429
4.	Hidden layer = 7 Delay=1:2 Feedback= 1:2	Test set 1	0. 3993	1.1653	2.2256
		Test set 2	0.3610	1.2755	2.6469
		Test set 3	0.2598	1.0806	2.3367
		Test set 4	0.2804	0.8168	1.5574
5.	Hidden layer = 10 Delay=1:2 Feedback= 1:4	Test set 1	0. 2506	0.3822	0. 9595
		Test set 2	0.1687	0.3658	1. 652
		Test set 3	0.0874	0.2410	1. 736
		Test set 4	0.1645	0.2729	0. 8746
6.	Hidden layer = 10 Delay=1:2 Feedback= 1:2	Test set 1	0.2871	0.7042	1.7415
		Test set 2	0.2363	0.8631	2.6098
		Test set 3	0.2146	0.5372	1.6833
		Test set 4	0.1995	0.4919	1.1881
7.	Hidden layer = 15 Delay=1:2 Feedback= 1:4	Test set 1	0.2504	0.3694	0.8407
		Test set 2	0.1681	0.3264	1.2613
		Test set 3	0.0872	0.2304	1.0880
		Test set 4	0.1646	0.2811	0.9557
8.	Hidden layer = 15 Delay=1:2 Feedback= 1:2	Test set 1	0.2848	0.6896	1.7457
		Test set 2	0.2351	0.8823	2.9077
		Test set 3	0.1257	0.5651	1.8662
		Test set 4	0.1987	0.4995	1.2706

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REFRENCES:

[1] G.E. Box, G.M. Jenkins, G.C. Reinsel, and G.M. Ljung, " Time series analysis: forecasting and control. ", Holden-Day, 1976.

[2] G.P. Zhang, " Time series forecasting using a hybrid ARIMA and neural network model.", Neurocomputing, vol. 50, pp. 159-175, 2003.

[3] C. Brooks, "Univariatetime series modelling and forecasting.", Introductory Econometrics for Finance. 2nd Ed. Cambridge University Press. Cambridge, Massachusetts, 2008.

[4] J.H. Cochrane, Time series for macroeconomics and finance. Manuscript, University of Chicago, 2005.

[5] K.W. Hipel, and A.I. McLeod, Time series modelling of water resources and environmental systems, Elsevier. vol. 45, 1994.

[6] Taieb SB, Bontempi G, Atiya AF, Sorjamaa A. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. Expert systems with applications. 2012 Jun 15;39(8):7067-83.

[7] De Gooijer JG, Hyndman RJ. 25 years of time series forecasting. International journal of forecasting. 2006 Dec 31;22(3):443-73.

[8] ElMoaqet, Hisham, Dawn M. Tilbury, and Satya Krishna Ramachandran. "Multi-Step Ahead Predictions for Critical Levels in Physiological Time Series." (2016).

[9] Koskivaara, Eija. Artificial neural networks in auditing: state of the art. Turku Centre for Computer Science, 2003.

[10] Leondes, Cornelius T., ed. Intelligent Systems: Technology and Applications, Six Volume Set. Vol. 1. CRC Press, chapter 1, neural network techniques and their engineering applications, 2002.

[11] Lapedes A, Farber R. Nonlinear signal processing using neural networks: Prediction and system modelling. 1987 Jun 1.

[12] De Groot C, Würtz D. Analysis of univariate time series with connectionist nets: A case study of two classical examples. Neurocomputing. 1991 Nov 30;3(4):177-92.

[13] Chakraborty K, Mehrotra K, Mohan CK, Ranka S. Forecasting the behavior of multivariate time series using neural networks. Neural networks. 1992 Dec 31;5(6):961-70.

[14] Cai X, Zhang N, Venayagamoorthy GK, Wunsch DC. Time series prediction with recurrent neural networks trained by a hybrid PSO–EA algorithm. Neurocomputing. 2007 Aug 31;70(13):2342-53.

[15] Zhang GP. A neural network ensemble method with jittered training data for time series forecasting. Information Sciences. 2007 Dec 1;177(23):5329-46.

[16] Yu L, Wang S, Lai KK. A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates. Computers & Operations Research. 2005 Oct 31;32(10):2523-41.

[17] Mencattini A, Salmeri M, Bertazzoni S, Lojacono R, Pasero E, Moniaci W. Local meteorological forecasting by type-2 fuzzy systems time series prediction. In IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, Giardini Naxos, Italy 2005 Jul 20 (Vol. 7, pp. 75-80).

[18] Mencattini A, Salmeri M, Bertazzoni S, Lojacono R, Pasero E, Moniaci W. Short term local meteorological forecasting using type-2 fuzzy systems. In Neural Nets 2006 (pp. 95-104). Springer Berlin Heidelberg.

[19] Zhang GP, Qi M. Neural network forecasting for seasonal and trend time series. European journal of operational research. 2005 Jan 16;160(2):501-14.

[20] Maqsood I, Khan MR, Abraham A. An ensemble of neural networks for weather forecasting. Neural Computing & Applications. 2004 Jun 1;13(2):112-22.

[21] Taskaya-Temizel T, Casey MC. A comparative study of autoregressive neural network hybrids. Neural Networks. 2005 Aug 31;18(5):781-9.

[22] Ahmed NK, Atiya AF, Gayar NE, El-Shishiny H. An empirical comparison of machine learning models for time series forecasting. Econometric Reviews. 2010 Aug 30;29(5-6):594-621.

[23] Khan MR, Ondrek C. Short-term Electrical Demand Prognosis Using Artificial Neural Networks. JOURNAL OF ELECTRICAL ENGINEERING-BRATISLAVA-. 2000;51(11/12):296-300.

[24] Siegelmann HT, Horne BG, Giles CL. Computational capabilities of recurrent NARX neural networks. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics). 1997 Apr;27(2):208-15.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[25] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks. 1994 Mar;5(2):157-66.

[26] Lin T, Horne BG, Tino P, Giles CL. Learning long-term dependencies in NARX recurrent neural networks. IEEE Transactions on Neural Networks. 1996 Nov;7(6):1329-38.

[27] Salomon R. Improved convergence rate of back-propagation with dynamic adaption of the learning rate. In International Conference on Parallel Problem Solving from Nature 1990 Oct 1 (pp. 269-273). Springer Berlin Heidelberg.

[28] Vas P. Artificial-intelligence-based electrical machines and drives: application of fuzzy, neural, fuzzy-neural, and genetic-algorithm-based techniques. Oxford university press; 1999 Jan 28.