Taylor-based Optimized Recursive Extended Exponential Smoothed Neural Networks Forecasting Method

Emna Krichene [1], Wael Ouarda [2], Habib Chabchoub [2], Ajith Abraham [2], Abdulrahman M. Qahtani [2], Omar Almutiry [2], habib dhahri [2], and Adel Alimi [2]

[1]National School of Engineering of Sfax
[2]Affiliation not available

October 30, 2023

Abstract

A newly introduced method called Taylor-based Optimized Recursive Extended Exponential Smoothed Neural Networks Forecasting method is applied and extended in this study to forecast numerical values. Unlike traditional forecasting techniques which forecast only future values, our proposed method provides a new extension to correct the predicted values which is done by forecasting the estimated error. Experimental results demonstrated that the proposed method has a high accuracy both in training and testing data and outperform the state-of-the-art RNN models on Mackey-Glass, NARMA, Lorenz and Henon map datasets.

Taylor-based Optimized Recursive Extended Exponential Smoothed Neural Networks Forecasting Method

Emna Krichene[1*], Wael Ouarda[2], Habib Chabchoub[3], Ajith Abraham[4], Abdulrahman M. Qahtani[5], Omar Almutiry[6], Habib Dhahri[6] and Adel M. Alimi[1,7]

[1*]REsearch Groups in Intelligent Machines (REGIM Lab), University of Sfax, National Engineering School of Sfax (ENIS), BP 1173, Sfax, 3038, Tunisia.
[2]Digital Research Center of Sfax, B.P. 275, Sakiet Ezzit, 3021 Sfax, Tunisia.
[3]College of Business, Al Ain University of Science and Technology.
[4]Machine Intelligence Research Labs (MIR Labs), Scientific Network for Innovation and Research Excellence, Washington 98071-2259, USA.
[5]Department of Computer Science, College of Computers and Information Technology, Taif University, P.O.Box. 11099, Taif 21944, Saudi Arabia..
[6]College of Applied Computer Science, King Saud University, Riyadh, Saudi Arabia..
[7]Department of Electrical and Electronic Engineering Science, Faculty of Engineering and the Built Environment, University of Johannesburg, South Africa..

*Corresponding author(s). E-mail(s): emna.krichen@enis.tn;

Abstract

A newly introduced method called Taylor-based Optimized Recursive Extended Exponential Smoothed Neural Networks Forecasting method is applied and extended in this study to forecast numerical values. Unlike traditional forecasting techniques which forecast only future values, our

proposed method provides a new extension to correct the predicted values which is done by forecasting the estimated error. Experimental results demonstrated that the proposed method has a high accuracy both in training and testing data and outperform the state-of-the-art RNN models on Mackey-Glass, NARMA, Lorenz and Henon map datasets.

1 Introduction

Nowadays, data are increasingly growing. Hence, data computation becomes a major challenge. Companies dealing with Big Data need sophisticated decision assistance in order to optimize the planning of these tasks and tools which have got several problems to achieve their goals. First, the heterogeneity of the data induces the complexity of forecasting future values. Second, the diversity of the methods involved in the computation (i.e., in addition to the several inherent constraints) increases this complexity.

Forecasting is an important data analysis field that aims to examine historical data in order to extend and predict its future values. Thus, forecasting activities play an important role in our daily life and involve various fields so that much research aim to develop tools for forecasting and decision making. We often forecast weather ([11] [18]([19][37]), wind speed ([2] [12]), stock market [8], electricity ([20] [21][22][23][36]), etc.

Meanwhile, forecasting literature contains a wide diversity of techniques that can be organized into two main families (see citeShao10[36]):

1. Qualitative methods ([42][43][44]) where judgmental forecasts are principally based on experts' opinions. These methods provide in forecasting any information that cannot be explained in the past. Furthermore, these techniques are, especially, used when sufficient information and data are not available; so that quantitative methods cannot be used i.e., when case study is vague and only few data exist. For example, bearing on apparition of a new product or a new event, this type of forecasting techniques seems very adequate. The most known qualitative methods are: Consumer Survey [47], Consumer Survey-Sample Survey Method [47], Delphi Method ([45] [46]) and Past Analogies [48]. However, in spite of their robustness, these methods present some disadvantages. As they are totally based on the intuitions of experts, these techniques are not only expensive but also not adequate for most cases especially in big data case; i.e., human capability is not able to predict future values from a wide range and observations. For these reasons and others, another type of forecasting techniques appears which is totally based on mathematical fundamentals later called as quantitative methods.

2. Quantitative methods (also called statistical methods) ([38][39][40][41]) do not relate to experts' intuitions but they mostly rely on quantitative historical data that can be extrapolated to make our forecasts. This type of techniques is used when the case study is stable and historical data exist. The most known statistical techniques are: Naive method [52], Moving Average (MA) [50], Weighted Moving Average (WMA) [50], Exponential Smoothing [49] and linear regression method [51]. These techniques are commonly used when they handle products or phenomena that already exist and when the historical data are of course available in order to study it.

   Comparing to qualitative methods, these techniques are less expensive and faster. Even though, their effectiveness is limited to some applications as they still be not appropriate when observations are very different and chaotic i.e non linear and unorganized observations. This point motivated researchers to go even further and to move toward artificial intelligence as prediction tools.

Recently, different methods have been presented to the task of forecasting as intelligent techniques; AI based systems are conceived to take advantage of the two traditional forecasting approaches: quantitative and qualitative. In fact, Neural Networks (NN) are the most-used techniques as they are nonparametric and nonlinear methods [70]. NNs are able to model and to map complex problems by training them in order to determine parameters and explore relationships between data ([1][10][27]). Accordingly, NNs have been considered as a promising and useful method applied to a variety of forecasting problems.

From an architectural point of view, there are various types of neural networks that share the same basic concept. Each problem needs to be molded by admitting its own neural network architecture and network configuration since there is no single learning model that can be suitably used for all fields. However, in NN, we distinguish two main categories: (i) Multi-Layer perceptron (MLP) which is the most popular method applied in time series forecasting task and (ii) Recurrent NN (RNN) ([1] [4] [9]) where its context-layer serves to store previous states within a dynamic memory. Authors in [27] believe that the interest of this supplementary layer is its ability to better underline the relationships between past observations and future values and so that a better generalization is enhanced.

A brief comparison of these techniques is summarized and presented in table 1.

The purpose of this paper is to forecast time series data using a variant of Recurrent Neural networks (RNN). Yet, to build the recurrent architecture, we drew inspiration from the Exponential Smoothed method (ES) [49]. Thereby, Recursive Extended Exponential Smoothed Neural Networks (REESNN) represents the first part conception of our proposed method. The primary objective of this work is an assessment of this method as a tool of forecasting. Furthermore, a secondary part of this paper consists in optimizing

Table 1 Forecasting Techniques Comparison

Family of the technique	Description (main concept and advantages)
Qualitative methods ([42][43][44][45][46][47] [48])	- Depend upon experience of individual experts. - Suitable for new phenomena, product, event and when few data exists. - Not able to model complex data especially in big data case.
Quantitative methods ([38] [39][40][41] [49] [50][51] [52])	- Rely on mathematical formulas - Suitable when historical data exist. - Not able to model complex data.
Intelligent methods [70]	- Rely on intelligent theories. - Suitable for complex and chaotic data. - Capability to map the relationship between data.

the results already obtained in the first step. The task of optimization is performed by predicting the estimated error to be then tailored with forecasting results by applying Taylor expansion principle.

This paper is organized as follows: In Section 2, we start by a review in which we will discuss recent researches using Recurrent Neural network architectures. Then, Section 3 gives an extensive description of the propounded approach based on Taylor theorem to optimize forecasting future values. Then, we will illustrate the robustness of our approach by giving an exhaustive experimental result on four different datasets: Mackey-Glass times series, NARMA, Lorenz and Henon-Attractor in Section 4. Finally, Section 5 draws the conclusion of the paper.

2 Related works

Nowadays, significant feature related to RNN is the subject of numerous forecast researches. The literature is vast and growing and it is difficult for a researcher to sweep all works contributing to forecasting literature done until now. In this section, we will present some works done over the last ten years that apply RNN as forecasting techniques. Hence, different architectures of RNN have been studied. However it seems that the most influential models are the Elman RNN (ERNN) [4] and Jordan RNN (JRNN) [6] as they are the most used models. This section is divided into three parts: In the first one, we explain the fundamental principle of ERNN and we present some works contributing with ERNN architecture. In the second part, we briefly describe the architecture of Jordan RNN and we introduce a brief survey of research activities contributing to forecasting literature using Jordan RNN over the last 10 years. Lastly, we present different works that were presented to the literature with other variants of RNN such as LSTM, BLSTM, GRU, etc..

2.1 Elman Recurrent Neural Networks

The simplest and the most known Recurrent Neural Network is ERNN architecture [4]. ERNN is designed as three layers expanded by a context-layer

which receives inputs from the hidden units (see Fig. 1). The role of this supplementary layer is to store previous states of the hidden neurons within a dynamic memory.
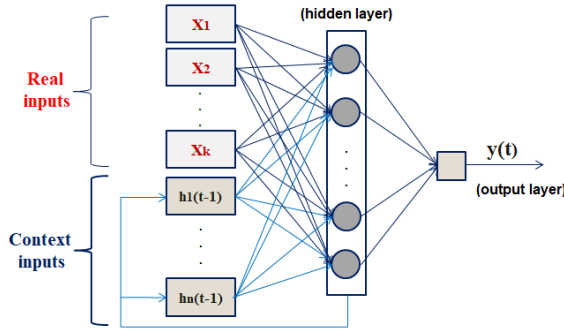
Fig. 1 The architecture of Elman RNN model

ERNN is the most popular and widely-used technique on Artificial Intelligence applications. Nevertheless, many researchers use ERNN as forecasting tool. Among them, authors in [8] applied four different types of RNN for sales forecasting (i.e., Elman RNN, Jordan RNN, Hierarchical Elman recurrent neural network (HERNN) and Multi-recurrent network). Then, they propose to analyze dissimilarities between the obtaining results and other quantitative forecasting methods (i.e., Exponential smoothing method [49] and SARIMA method [60]. Empirical results provide that HERNN outperforms other types of recurrent neural networks, whereas it is noticed that the Exponential method accurate better on longer horizon than other methods.

In the same context, researchers in [71] applied a hybrid approach based on ERNN with stochastic time effective function (STNN) as forecasting technique. Compared with different methods, experiment results show that the proposed technique gives better performance.

In [72], two different architectures of NN was applied as predictor methods to forecast the flood water level earlier so that precaution steps can be taken. To do so, ERNN and NARX were compared to specify which technique gives more accurate results. Both architectures were fed with the same recorded water level. Based on Experiment results, ERNN accurate better than NARX not only during the training step but also during the testing one.

Yet, the work of [73] consists in predicting electricity consumption. The main purpose is to allocate suitable power resources and to help electric power companies to present reasonable sales plans. Elman recurrent neural network is applied in this work serving by extra inputs which affect the electricity consumption such as gross domestic product, temperature, and Spring Festival. By comparing the results with and without considering effect factors, it can be shown that the proposed architecture with the extra inputs accurate better than the simple ERNN.

Another application of Elman recurrent neural network contributing to the literature of forecasting was applied by [74]. The proposed technique aims to forecast load. Authors employ Elman architecture taking consideration of some related factors which enhance the performance of the model.

In the same sense, authors in [75] applied also the Elman recurrent neural network architecture based on Particle Swarm Optimization to predict the wind speed and compare the obtained results to those obtained by applying simple ERNN. Experiments prove that the optimized topology outperforms the simple one.

In the same context, researchers in [76] design an architecture that merges two Elman NN structures to forecast wind power. The architecture is designed as follows: The first Elman NN part is realized to predict wind power component. Then, a second part is designed to forecast the guidance of wind direction. As a final step, the fusion of the two parts gives accurate results about wind power.

Since Elman RNNs have a high performance, another work adopts its architecture to predict the electric load. Thereby, authors in [77] used two types of Elman RNN: the first one with single hidden layer and the second with dual hidden layers. Experimental results show that the first typology is more accurate as the average of the smallest error was given by it.

In [78], Elman RNN was also applied to forecast the Photovoltaic power. The gradient descent back propagation algorithm was used during the learning phase in order to settle the parameters and building the optimized architecture. To prove the efficiency of the system, three tests were made and evaluated through Mean Absolute Error and Root Mean Square Error.

However, based on the Jordan recurrent neural network architecture, many other papers contribute to the literature. In this section and the rest of the paper, we focus our interest on the Jordan RNN model and its variations.

2.2 Jordan Recurrent Neural Networks

The major difference between Jordan RNN's architecture and those of other RNNs is that the context-layer is designed as a copy of the output-layer as shown in Fig. 2.
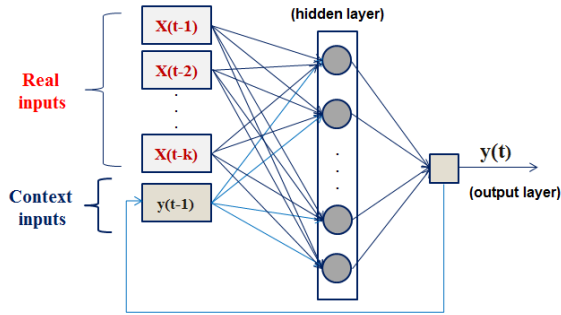
Fig. 2 The architecture of Jordan RNN model

The Jordan RNN [6] models are used in diverse prediction tasks especially in time series forecasting problems because of their inherent competence (effectiveness) to map input-output problems. In [13], authors build a variant architecture of Jordan RNN using exogenous input to forecast wave data; experimental results show that the proposed architecture presents more definite results when the input data has been smoothed; applying the smoothing is two-fold: first to eliminate the noise from the data, and second to give more importance to recent observations.

In the same sense, another forecasting problem has been solved by Jordan RNN. Authors in [2] applied two different architectures of Jordan recurrent neural networks in order to analyze the wind speed prediction. Their study compares the proposed scheme with ARIMA model and the testing results confirms the assumption that recurrent neural network outperforms in most cases others statistical models. Another important application of recurrent neural network was proposed by [23], where the authors proposed a variation of Jordan recurrent neural network called Recurrent Cartesian Genetic Programming evolved Artificial Neural Network (RCGPANN). Comparing to other previous studies, the evaluation results show that the implemented method accurate better.

In [24], researchers present a new Jordan RNN model based on the Hybrid Complex neural network; the proposed architecture is built as a recurrent multilayer perceptron as it has two hidden layers in addition to the recurrent connections. Results were then compared to those of feed-forward architecture and fully RNN model and it was noticed that the proposed scheme outperforms their results even in the worst case. Moreover, authors in [25] applied the Jordan RNN and have compared it to the MLP architecture. Thereby, although testing results were very similar, it can be noticed that the best forecasts were given by the recurrent architecture.

Likewise, [26] proposed a novel technique called wavelet RNN in which its principle is inspired from the Jordan RNN. Briefly describing, the proposed architecture is a multilayer recurrent network with output feedback connection to the first hidden layer.

Another application of Jordan recurrent neural network is presented to the literature was applied by [79]. The proposed technique aims to forecast Stock Market Price. To enhance the performance of the model, Unsupervised method is used in the first step to reduce the dimension of inputs. As a second step, JRNN is used to predict the closing price of each day serving by the information of the previous day. Empirical results illustrate the efficiency of the proposed technique; Evaluation was made serving by two metrics: mean square error and mean absolute percentage error.

In [80], researchers propose to combine two methods in one architecture to predict the plant output. Their system is composed of Balanced Truncation (BT) cascaded with Jordan network. A comparative study was then performed to different MOR models and shows that the proposed system accurates better in most cases.

2.3 Other Variants of Recurrent Neural Networks

Although our focus is on Elman and Jordan RNN as they represent an important part Of our methodology, it should be mentioned that LSTM as well as its derivatives also perform good results in prediction tasks.

However, predicting Stock market is considered as one of the most challenging issues to resolve. In this context, authors in [67] present to the literature an evaluation of Bi-LSTM for stock market forecasting. They compare different variants of LSTM to each others: Unidirectional, bidirectional, shallow neural network and stacked LSTM. Experiment results showed that BLSTM and SLSTM perform better than other compared techniques.

Another work was presented to the literature using LSTM ([64]), where authors propose using Deep LSTM to forecast petroleum production. The proposed architecture is optimally configured serving by the genetic algorithm.

Yet, [66] suggest using LSTM methodology in order to forecast real time data. The proposed architecture is a LSTM-based regression structures extended by an additional gate to predict real time series data.

To the best of our knowledge, researchers in ([65]) were the first to apply Bidirectional LSTM to build blocks for deep architecture in order to exploit backward dependencies of traffic data for forecasting.

Lately, another variant of RNN named Gated Recurrent Units (GRU) has increasing its popularity on timeseries applications. Researchers in [68] empower the usefulness of these different variants of RNN by applying them on Turkish electricity load prediction. The comparison between obtained results with those obtained by existing researches based on Auto Regressive integrated moving average (ARIMA) and other ANN architectures prove the efficiency of RNN as well as its derivatives.

Another work recently added to the literature [69], is applied to the Agriculture sector as it is considered as one of the highest energy consuming. RNN, LSTM and GRU are used to predict hourly short term Agriculture load. Empirical results show that these techniques are more feasible than Moving Average model, Auto Regressive model and ARIMA.

To conclude this section, a brief recapitulation of these contributing papers is summarized and presented in table 2.

Nevertheless, thanks to their specific design, Jordan RNN is supposed to be a perfect alternative model to forecast time series even if their elements are hard to predict. Its structure is supposed to have a more powerful competence than other typical RNN architectures to train and predict the given data [14]. In this paper, to address the aforementioned problem, we propose a newly Recurrent NN architecture inspired from Taylor theory as a forecasting technique. Specifically, we have aligned our contribution by focusing on two points: (1) Function approximation using the Recursive Extended Exponential Smoothed method and (2) error approximation in order to optimize the forecasted values using the Taylor expansion [53].

Table 2: Summary of Different Contributing Papers Over Last Ten Years.

	Reference	Data	Brief Description
	[8]	Real-world sales data	Partial RNN was applied to prove the effectiveness of PRNN in business decision support and sales forecasting.
ELMAN RNN	[71]	Financial time series	Hybrid model based on ERNN with STNN.
	[72]	Flood water level	Comparison between the performance of ERNN and NARX.
	[73]	Monthly electricity data	ERNN is applied in this work serving by extra inputs affecting electricity consuming.
	[74]	Guangdong electric power grids	Authors employ Elman architecture taking consideration of some related factors which enhance the performance of the model.
	[75]	Wind speed	Authors suggest to optimize the Elman typology by the practical swarm algorithm
	[76]	Wind power	The fusion of two architectures of Elman NN is applied to get accurate wind speed prediction.
			Continued on next page

Table 2 – continued from previous page

	Reference	Data	Brief Description
	[77]	Electricity price	Comparison between Elman and other existing forecasting methods to predict electricity prices. Empirical results prove that the forecasted error obtained by the proposed method was lower than the reported errors of other compared techniques.
	[78]	Photovoltaic power	Elman architecture is used to forecast the future power of the photovoltaic station. The system is fed by four inputs: Temperature, Humidity, Wind-speed and Radiation which are all normalized. Three tests were made to prove the efficiency of the system.
	[2]	Wind speed	Comparison between two architectures of Jordan RNN and SARIMA method.
JORDAN RNN	[13]	Wave data	A variant of Jordan RNN with exogenous data to forecast wave data.
	[23]	Electricity	A variation of Jordan RNN called: Recurrent Cartesian Genetic Programming evolved ANN was applied to forecast electrical load.
	[24]	11 Economic time series of the NN5 Forecasting Competition	A new Jordan RNN based on hybrid complex NN was applied and compared to MLP and multi-recurrent NN.
	[25]	Flash floods without rainfalls	Comparison between Jordan RNN and MLP to forecast Flash floods without rainfalls.
	[26]	Solar Radiation	A variation of Jordan RNN was proposed called Wavelet RNN and applied to forecast Solar Radiation.

Table 2 – continued from previous page

|	Reference	Data	Brief Description
	[79]	Stock Market Price	Unsupervised method is used in the first step to reduce the dimension of inputs, then serving by the information of the previous day, JRNN is used to predict the closing price of each day.
	[80]	Plant output	BT method and Jordan network are combined in one architecture to predict plant production. The system's response is then compared to the original signal as well as the output of different other techniques such as IV, JRNN and [81]. The proposed method illustrates better results over competing methods named above.
Other variants of RNN	[64]	Petroleum production	A deep LSTM optimally configured by the use of genetic algorithm is adopted to predict the petroleum production.
	[65]	Traffic data	Building blocks for deep Bidirectional LSTM architecture in order to exploit backward dependencies of traffic data for prediction.
	[66]	Real time data	Forecasting real time-series data serving by a LSTM-based regression structures extended by an extra gate.
	[67]	Stock market	A literature comparison and evaluation of different variants of LSTM: Unidirectional, bidirectional, shallow neural network and stacked LSTM. Experiments prove that BLSTM and SLSTM perform better than other variants of LSTM.
			Continued on next page

Table 2 – continued from previous page

	Reference	Data	Brief Description
	[68]	Turkish electricity load	A Comparison between the efficiency of GRU and others forecasting methods in the case of Turkish electriciy load prediction.
	[69]	Hourly short term Agriculture load	Comparison between RNN, LSTM and GRU to other forecasting techniques such as MA, AR and ARIMA. Experiments prove the effectiveness of the different variants of RNN.

3 Proposed Approach

The proposed methodology is divided into three main steps given as follow:

- **Step 1:** The proposed method Recursive Extended Exponential Smoothed Neural Networks (REESNN) inspired from the Jordan RNN & ES is applied in order to approximate the function and forecast values.
- **Step 2:** Elman RNN Classifier is trained to approximate the forecasted error.
- **Step 3:** The final forecasting results are obtained by tailoring the two previous results according to the principle of Taylor expansion.

These aforementioned steps are more detailed in next sections. The forecasting flowchart of the proposed methodology is shown in Fig. 3.
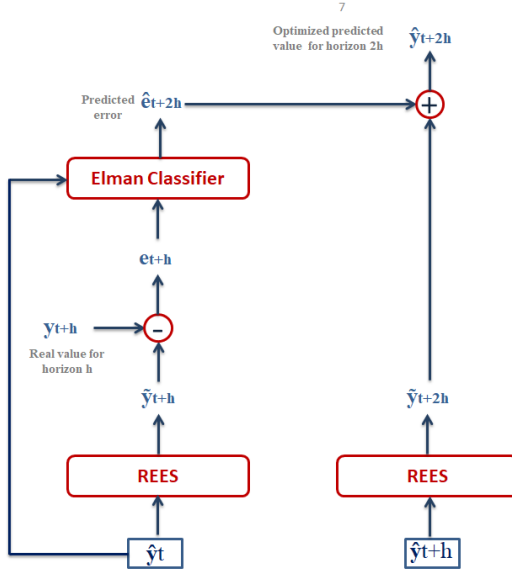
Fig. 3 The flowchart of a general TOREESNN architecture

3.1 Function approximation and forecasting values based on REESNN

A Jordan RNN (JRNN) is a neural network that uses the received previous network output as a new input to be later processed [16]. Hence, a JRNN is a simple recurrent network so that a specific group of neurons (later called context neurons) receives feedback signals from the previous time step. Thereby, input layer is composed of two parts: true input neurons and context neurons which are a duplicate of the outputs provided from the previous time step. Hence, the structure of the JRNN model can be pictorial as illustrated in Fig. 2.

Basically, each unit in a particular layer is connected with all neurons in the next layer and then summed and multiplied by the appropriate importance degree called w_{ij} which characterizes the connection and the degree of importance between the ith and the jth neuron. Hence, the nonlinear mapping function F masters the amplitude of the resulting output $\hat{y}$ which is given by Equation 1.

$$\hat{y}_i = F(\sum_{j=1}^{i} w_{ij} h_j) \tag{1}$$

where F is called the activation function. In general, the transfer function accurately reflects the nonlinearity degree of most preprocessed data by NN. We note that in theory, the activation function can be any differentiable function but in practice only four transfer functions are defined and can be computed according to Equations 2- 5.

14

- The linear function:
$$F(x) = x \tag{2}$$

- The sigmoid bipolar function (hyperbolic)

$$F(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3}$$

-The cos-sin function:

$$F(x) = \sin(x) or F(x) = \cos(x) \tag{4}$$

- The sigmoid function:
$$F(x) = \frac{1}{1 + e^{-cx}} \tag{5}$$

where c is the adaptive gain parameter of sigmoid function.

Meanwhile, in JRNN structure, the activation of the jth hidden neuron h_j is computed by summing the activation of context units in addition to those of true inputs. It holds that:

$$h_j(t) = F(\sum_{i=1}^{k} w_{ij} x_i(t) + \sum_{m=1}^{n} v_{mj} \hat{y}_i(t-1)) \tag{6}$$

where w_{ij} is the connection strength between the jth hidden unit h_j and the ith real input x_i , and $v_m j$ is the connection's weight between the jth hidden unit h_j and the $i_t h$ output past value at time $t-1$. Based on the modeled architecture in 2, the JRNN structure with activation function (5) in both hidden layer and output layer can be expressed by Equation 7.

$$\hat{y} = \sum_{h=1}^{H} u_h [\sum_{i=1}^{k} w_{ij} x_i(t) + \sum_{m=1}^{n} v_{mj} \hat{y}_i(t-1)] \tag{7}$$

where $\hat{y}$ is the predicted output and u_h are the connections weights on output layer.

Subsequently, all connection weights in the network are are merely calculated and updated by applying the back-propagation learning algorithm, as expressed in Equation 8,

$$\Delta w_{ij} = -\alpha \frac{\partial E}{\partial w_{ij}} \tag{8}$$

where α is the learning rate and E is the associated error with the ith neuron in layer j.

Hence, our goal is to properly adjust the network's connection weights and update them according to (9) in order to understate the value of error.

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij} \tag{9}$$

In the same way, all connection strengths in JRNN architecture are so updated. In our study, to approximate the value of a function, we considered recursive architecture inspired from the Exponential Smoothed (ES) method [49] which is a statistical forecasting technique. Its main principle assumes that each observation at time t relies on the previous observation and the variation between that forecast and the actual value of the series at that point. The principle of the ES can be formulated by the equation:

$$\hat{y}(t+1) = \hat{y}(t) + \alpha(x(t) - \hat{y}(t)) \tag{10}$$

where $\hat{y}(t)$ is the previous forecasted value , x is the actual one and $\alpha \in [0, 1]$ is the smoothing factor.

Our experiments aim to convert the ES method into an Extended ES designed as a Jordan Recurrent neural architecture. This task is accomplished within two steps: First, the ES method is modified and extended so that each observation does not only depend on solely one past value as ES assumes, but on many past values. Hence, we consider Extended ES (EES) as a generalization case of ES. Mathematically, the equation of EES can be formulated as:

$$\hat{y}(t+1) = \sum_{k=0}^{n} (\hat{y}(t-k) + \alpha(x(t-k) - \hat{y}(t-k))) \tag{11}$$

The second step consists on conceiving the EES as a recurrent architecture since it adopts properly the same principle; i.e the prediction of the future forecasted value depends on the previous estimated output as the JRNN supposes. Moreover, the smoothing factor is supposed been replaced by the weight connection. Hence, the simplest method for this purpose is to project the EES dynamic equation into a mapping function that can be depicted as:

$$\hat{y}(t+1) = F[\hat{y}(t), ..., \hat{y}(t-k); \ (x(t) - \hat{y}(t)), ..., (x(t-k) - \hat{y}(t-k))] \tag{12}$$

Graphically, the Recursive Extended Exponential Smoothed Neural Networks (REESNN) is considered as a set of Long-Short Exponential Smoothed units (LSES) where the input layer is a set of two types of inputs: The previous forecasted values $\hat{y}$ at time $t-1$, $t-2$,.., $t-k$ and, the estimated error e outlined as the variance between the actual real value x and the forecasted one $\hat{y}$ at time $t-1$, $t-2$,.., $t-k$ as illustrated in Fig. 4.
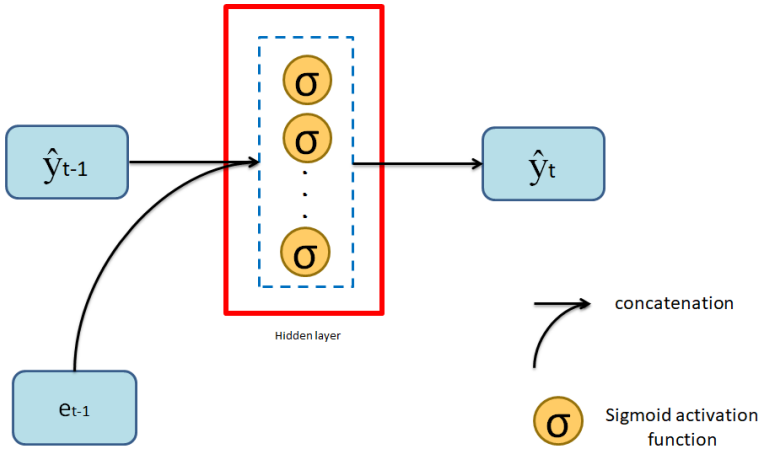
**Fig. 4** Architecture of long-short exponential smoothed unit (LSES).

After being processed by sigmoid activation functions, the information passes through many such LSES units as Fig. 5 depicts.

3.2 Forecasting error estimation

A powerful part of this approach is the prediction of error estimation in order to improve the forecasting precision. Forecasting error estimation is defined as the prediction of the difference between the actual value x and the predicted one $\hat{y}$.

However, the uncertainty associated with forecasting is one of the most powerful factors influencing the resulting values. This point inspires us to examine in advance the task of forecasting by predicting not only the future values but also the uncertainty associated with each one.

In our study, Elman RNN Classifier (ERNNC) [4] model is used to model and approximate the error estimation associated to each forecasted value. This choice is justified by two points: first, when examining the error distribution, we notice that the behavior of the resulting errors is chaotic and oscillates around close values which makes the task of prediction hard to model. This points motivates us to solve the problem as a classification task, i.e a set of error classes is defined and associated to each residual value to be then trained. Second, comparing to other existing functions in classification forecasting problems, Elman has one of the best learning rates.
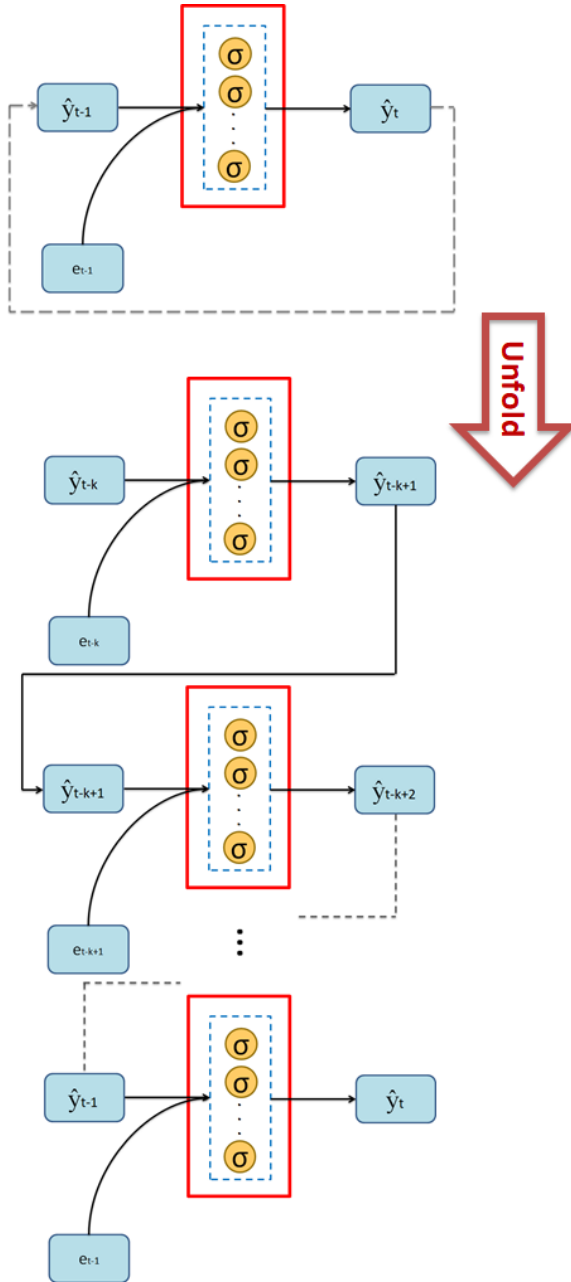
Fig. 5 Architecture of Recursive Extended Exponential Smoothed Neural Networks (REESNN).

In our study, Elman RNN Classifier (ERNNC) [4] model is used to model and approximate the error estimation associated to each forecasted value. This choice is justified by two points: first, when examining the error distribution, we notice that the behavior of the resulting errors is chaotic and oscillates around close values which makes the task of prediction hard to model. This points motivates us to solve the problem as a classification task, i.e a set of error classes is defined and associated to each residual value to be then trained. Second, comparing to other existing functions in classification forecasting problems, Elman has one of the best learning rates.

The suggested model is a three layered NN in addition to a hidden context layer and a classification function as shown in Fig. 6.
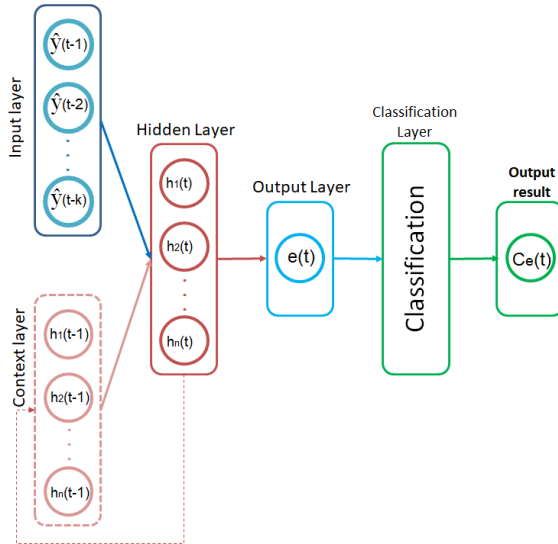
Fig. 6 Elman Recurrent Neural Network Classifier architecture.

Firstly, to train the Elman RNN as a Classifier method, it is necessary to preprocess the data and transform it on the adequate form. Hence, the transformation of the training Dataset S_c for the Elman RNN Classifier is introduced as follows:

$$S_c = \{(Y_t, C(e))\} \in [0, 1] \times \{-1, 0, 1\} \tag{13}$$

where Y_t is an input vector and $C(e)$ is a classification function of the estimated error e defined by Equations (14) and (15).

$$Y_t = [\hat{y}(t - \tau), \hat{y}(t - 2\tau), ..., \hat{y}(t - d\tau)] \tag{14}$$

$$C(e) = \begin{cases} -1 & e < -\epsilon \\ 0 & -\epsilon \leq e \leq -\epsilon \\ +1 & e > \epsilon \end{cases} \tag{15}$$

where τ is the time-delay and ϵ is an appropriate threshold that divides estimated errors into three scales: positive values, negative values and values near to zero. We note that ϵ is chosen, graphically, from the estimated error histogram as Fig. 7 depicts.
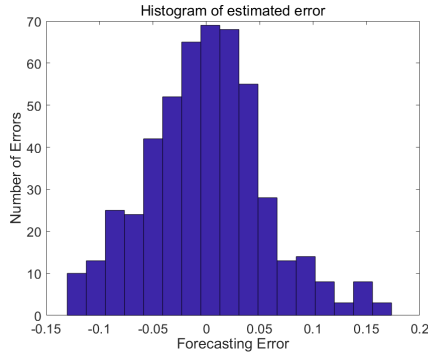
Fig. 7 Histogram of estimated error.

After the histogram of the estimated error was established, the Elman RNN Classifier is trained using S_c to predict the error class associated to each sample in the vector X_t .Then, to estimate the forecasting error, IF-THEN rules are adopted.

$$\begin{cases} \text{IF } C(e) = 1 \text{ THEN } f(e) = +\omega \\ \text{IF } C(e) = 0 \text{ THEN } f(e) = 0 \\ \text{IF } C(e) = -1 \text{ THEN } f(e) = -\omega \end{cases} \tag{16}$$

where $f(e)$ is the estimated forecasting error and ω is an experimental value measured during the learning phase where different values near to ϵ were tested; the one which gives minimal error is adopted for the rest of the algorithm . During the optimization phase, different values of ω were trained; the value giving the best result was memorized to be then applied during the final step.

3.3 Taylor-based forecasting results optimization

After predicting the estimated error, a step of optimization results was done. The optimized forecasted future value $\hat{y}_{t+1}$ is calculated according to an approximation function expressed between the first forecasted value $\hat{y}_t$ resulting from the first part of our architecture denoted as REESNN and the estimated forecasted error $f(e)$ resulting from the second part denoted as Elman Classifier (see Fig. 3). Though, a Taylor series approximation represents a number as a polynomial that has a very similar value to the number in a neighborhood around a specified value. Hence, the key idea in our method is to optimize the forecasted results serving by the Taylor expansion [53].
Motivation: The Taylor expansion

The main principle of the Taylor expansion [53] aims to approximate a function that is many times differentiable in the neighborhood of a point x. Its equation is expressed as:

$$f(x) = f(x_0) + \frac{x-x_0}{1!}f'(x_0) + \frac{(x-x_0)^2}{2!}f^2(x_0) + .. + \frac{(x-x_0)^n}{n!}f^n(x_0) + (x-x_0)^n\delta(x-x_0) \quad (17)$$

where $\delta(x - x_0)$ is a function that gets closer and closer to 0 as $(x - x_0)$ tends to 0.

From the above definition, this formula looks pretty similar to the optimization formulation; i.e the key idea behind optimization is to approximate a number by representing it by its nearest accurate neighborhood. Hence, a new problem formularization is proposed which translates the Taylor expansion formula into a combination of optimization task.

In fact, to resolve y_{t+1} and y_t by Equation 17 a step of integration of the error estimation function $f(e)$ is needed before time $t + 1$. This integration expression shows that the proposed optimization model can perform the dynamic backward optimization to better exploit the historical knowledge in the studied time series.

Inspired from Taylor principle, and provided the ϵ is small enough, the previous equation denotes that this combinational optimization problem can be approximately formulated as:

$$\hat{y}_{t+1} = \hat{y}_t + \frac{(t+1)-t}{1!}y'_t + 0(t) \quad (18)$$

where $\hat{y}_{t+1}$ denotes the optimized forecasted value and $\hat{y}_t$ represents the predicted value before optimization. We note that in our study we will limit to the first stage of Taylor principle and we hypothesize that y'_t is equal to the error approximation $f(e)$ as given in the following equation:

$$y'_t = f(e) = \frac{\hat{y}_{t+1} - \hat{y}_t}{(t+1) - t} \quad (19)$$

4 Results and Discussion

To validate the efficiency of our proposed method, we have based on three data sets. The first training data is composed of 500 samples which represent 35% from the whole data. The second training data is composed of 500 another patterns to train the estimated error which represents the second part of our architecture. Finally 500 new samples are used as testing data in order to improve the performance of our proposed scheme with new cases.

4.1 Mackey-Glass time series

The performance of the proposed methodology is demonstrated via the well-known benchmark time series called the Mackey-Glass time series (MG) [7].

Mathematically, the MG formula is established by a simple dynamic equation that can be noted as:

$$\frac{dx(t)}{dt} = \frac{ax(t-\tau)}{1+x^c(t-\tau)}bx(t) \tag{20}$$

where a, b and c are constants, and τ represents the time delay.

In this paper, the used parameters a, b and c are set to $0.2, 0.1$ and 10 respectively, whereas the most used τ in the literature are 17 and 30. The aim of this paper is to forecast the actual value $x(t)$ based on some specific historical past values. We note that a step of normalization was done to optimize the training phase.

In order to prove the efficiency of our proposed architecture, different forecasting RNN models were compared and testing results are outlined in tables 3 and 4.

Table 3 Comparison of Different Forecasting RNN Model to Our Proposed Method (MG Benchmark $\tau = 17$)

| Methods	MSE test
Elman [27]	$3.00e-03$
Elman & Elman Classifier	$2.30e-03$
ESN-LARS[31]	$7.80e-04$
ESN-FSR[31]	inf^*
REESNN	$2.07e-04$
TOREESNN	$1.91e-04$

*According to authors, the results are unstable and the error is very high.

**Table 4** Comparison of Different Forecasting RNN Model to Our Proposed Method (MG Benchmark $\tau = 30$)

Methods	MSE test
ESN-RR[31]	$8.25e-03$
ESN-LARS[31]	$3.20e-02$
ESN-FSR[31]	$7.72e-03$
ESN-LASSO[31]	$4.31e-03$
ESN-EN [31]	$4.31e-03$
REESNN	$3.50e-03$
TOREESNN	$2.90e-03$

According to tables 3 and 4, it is revealed that the proposed (TOREESNN) performs better than other RNN methods. Moreover, it can be noticed that optimizing forecasting task by predicting estimated errors influences greatly the testing results.

Fig. 8 illustrates the superposition between our results and the corresponding desired outputs.
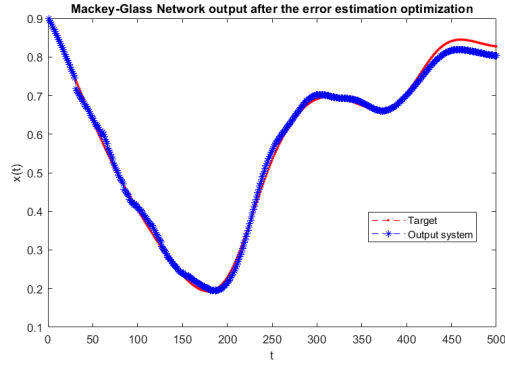
Fig. 8 Mackey-Glass Network output after the error estimation optimization $tau = 17$
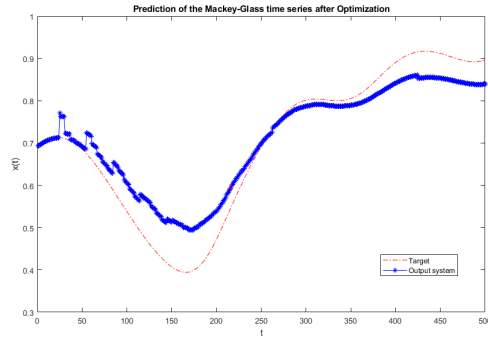
Fig. 9 Mackey-Glass Network output after the error estimation optimization $tau = 30$

4.2 The Nonlinear Auto-Regressive Moving Average time series

The Nonlinear Auto-Regressive Moving Average later called NARMA [59] is a renowned benchmark which is differentiated by its chaotic behavior. Further, the non-correlation between its values makes the task of learning more difficult to achieve. Moreover, the dynamic equation of NARMA is –mathematically-dependent on many parameters which makes it hard to model. In light of these facts, NARMA is considered among the most complex studied benchmarks; this point motivates us to further improve the efficiency of our method by testing NARMA patterns. Mathematically, the NARMA formula is expressed as:

$$y(t+1) = c_1 y(t) + c_2 y(t) \sum_{i=1}^{k} y(t-i) + c_3 x(t-k-1)x(t) + c_4 \qquad (21)$$

where $x(t)$ represents the input, k characterizes the order of the time series and the parameters c_i are set to 0.3, 0.05, 1.5 and 0.1 respectively.

The calculated errors between the output values and the desired one are visualized in figure 10. Seen that the error signal is displayed around ± 0.05 it clearly reflects that the network response is miming the required one. By the same token as for the foregoing benchmark, a MSE based comparison study with other existent forecasting methods is applied for NARMA series prediction with order $k = 10$.
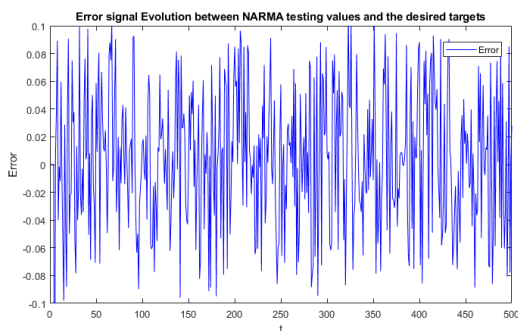
Fig. 10 Evolution of Error signal between NARMA testing values and the desired targets.

As illustrated by the preceding test, the output system and the target values are plotted in the same figure. The proportionality between the two curves further justifies the obtained results.

Table 5 presents the comparison of testing results given by the already existent approaches in the literature for NARMA Benchmark forecasting.

Table 5   MSE Based Comparison Between Our Proposed Approach and Other Literature Methods (NARMA Benchmark)

Methods	MSE test
ESN-FSR[31]	7.29e-03
ESN-LARS[31]	3.79e-03
ESN-RR[31]	3.03e-03
ESN-LASSO[31]	2.89e-03
LSTM[63]	0.01
REESNN	**1.20e-02**
TOREESNN	**2.00e-03**

4.3 Lorenz Attractor

Lorenz can be defined as a system of three nonlinear differential equations characterizing a fluid motion between a hot and a cool surface.(22-24) ([28]; [29]).

$$\frac{dx}{dt} = \alpha(y - x) \tag{22}$$

$$\frac{dy}{dt} = -y - xz - rx \tag{23}$$

$$\frac{dz}{dt} = xy - bz \tag{24}$$

where $\alpha = 10$, $r = 28$ and $b = 8/3$.

As in the previous time series, our task is to forecast $x(t)$ according to a set of its own historical past values.

In the same way as the MG test, the resulting tests are displayed in the same figure with the target values to be then compared. The superposition of the two curves is depicted in figure 11. It can be noticeable the network outputs are mimicking the output line of the target values.

As revealed by the previous benchmarks, a MSE based comparison with some existent literature RNN methods is applied for Lorenz attractor. Table 6 recapitulates testing results and proves that our proposed method represents a strong competitive forecaster in terms of accuracy.
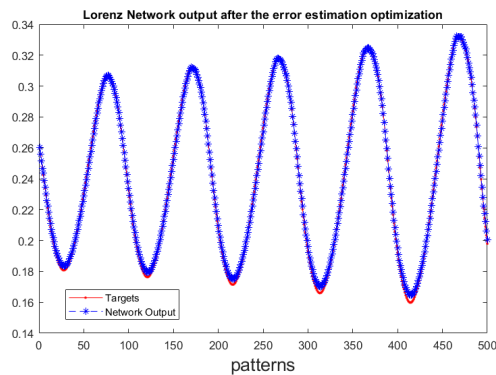
Fig. 11 Lorenz Network output after the error estimation optimization

4.4 Henon Attractor

The Henon map is a discrete-time dynamic attractor. Litteraturaly, this model is considered among the most common studied dynamical systems. It is typified by its chaotic behavior defined concretely by the following equations in 25 and 26 [28].

Table 6 MSE Based Comparison Between our Proposed Approach and Other Literature Methods (LORENZ Benchmark)

Methods	MSE test
SVR	2.80e-01
Canonical ESN [30]	7.23e-04
PSO-ESN [30]	4.04 e-06
LSTM[63]	6.4 e-03
REESNN	**3.93e-06**
TOREESNN	**3.90e-06**

$$x(t+1) = y(t) - ax^2(t) + 1 \qquad (25)$$
$$y(t+1) = bx(t) \qquad (26)$$

where $a = 1.4$ and $b = 0.3$. The dataset was normalized in the interval $[0, 1]$ to get more accurate results during the implementation.

Fig. 12 presents a depiction of the MSE signal calculated between desired values and system outputs. According to the figure, the signal of calculated errors is displayed around ± 0.04 which extremely reflects that the network response is following the desired one.
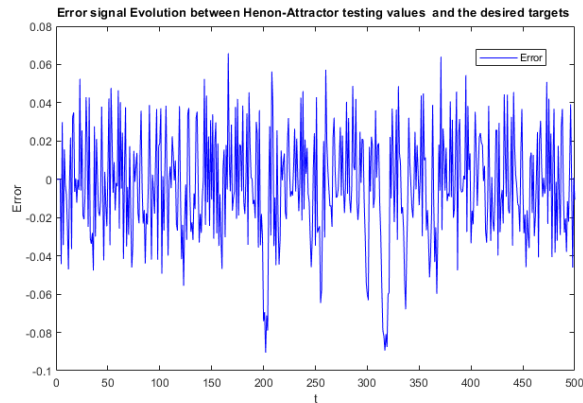

Fig. 12 Evolution of Error signal between Henon-Attractor testing values and the desired targets.

Table 7 presents the comparison of testing results given by the already existent approaches in the literature for Henon attractor forecasting. According to Table 7, we remark that TOREESNN outperforms the other works in terms of accuracy.

Table 7 MSE Based Comparison Between our Proposed Approach and Other Literature Methods (HENON Benchmark)

Methods	MSE test
Canonical ESN [30]	3.40e-3
SGD-ESN[32]	3.67e-2
RLDDE[33]	4.70e-3
REESNN	**9.70e-04**
TOREESNN	**8.23e-04**

5 Conclusion

In literature, the task of Forecasting was investigated with different Artificial Neural Network methods inter alia Recurrent Neural Networks architectures. In this paper, we applied a novel JRNN architecture (named TOREESNN) inspired from the ES method extended by error approximation to predict and optimize one of the most well-known time series such as Mackey-Glass time series, Henon Attractor, Lorenz and NARMA. We have compared the obtained results to other RNN architectures and experimental results proved the efficiency of our proposed method.

In summary, this paper warrants the assumption above-mentioned at the beginning of this paper which assumes that optimizing the task of forecasting by prediction the error estimation can be a good alternative model to deal hard time series. In fact, the step of optimization offers the possibility to make full use of historical knowledge in past data as it suits ideally for memorizing not only past forecasted values but also the error made during the training phase.

References

[1] Ammar, B., Chérif, F. & Alimi, A. Existence and Uniqueness of Pseudo Almost-Periodic Solutions of Recurrent Neural Networks With Time-Varying Coefficients and Mixed Delays. *IEEE Trans. Neural Netw. Learning Syst.*. **23**, 109-118 (2012)

[2] Cao, Q., Ewing, B. & Thompson, M. Forecasting wind speed with recurrent neural networks. *European Journal Of Operational Research*. **221**, 148-154 (2012,8)

[3] Brezak, D., Bacek, T., Majetic, D., Kasac, J. & Novakovic, B. A comparison of feed-forward and recurrent neural networks in time series forecasting.

CIFEr. pp. 1-6 (2012)

[4] Elman, J. Finding structure in time. *COGNITIVE SCIENCE.* **14**, 179-211 (1990)

[5] Hopfield, J. Neurocomputing: Foundations of Research. (MIT Press,1988)

[6] Jordan, M. Artificial Neural Networks. (IEEE Press,1990)

[7] Lapedes, A. & Farber, R. Nonlinear Signal Processing Using Neural Networks: Prediction and System Modeling. (Los Alamos National Laboratory,1987)

[8] Muller-Navarra, M., Lessmann, S. & Voss, S. Sales Forecasting with Partial Recurrent Neural Networks: Empirical Insights and Benchmarking Results. *2015 48th Hawaii International Conference On System Sciences (HICSS).* pp. 1108-1116 (2015)

[9] Ulbricht, C. Multi-Recurrent Networks for Traffic Forecasting. *Proc. Of AAAI-94.* pp. 883-888 (1994)

[10] Wallace, M. Neural Networks and Their applications to Finance. *Business Intelligence Journal.* **1** pp. 67-77 (2008)

[11] Mustafaraj, G., Lowry, G. & Chen, J. Prediction of room temperature and relative humidity by autoregressive linear and nonlinear neural network models for an open office. *Energy & Buildings.* **43**, 1452-1460 (2011)

[12] Ji, G., Han, P. & Zhai, Y. Wind Speed Forecasting Based on Support Vector Machine with Forecasting Error Estimation. *Proceedings Of The Sixth International Conference On Machine Learning And Cybernetics.* pp. 2735 - 2739 (2007,9)

[13] Park, D. A Time Series Data Prediction Scheme Using Bilinear Recurrent Neural Network. *2010 International Conference On Information Science And Applications.* pp. 1-7 (2010,4)

[14] Wutsqa, D., Kusumawati, R. & Subekti, R. The application of Elman recurrent neural network model for forecasting consumer price index of education, recreation and sports in Yogyakarta. *2014 10th International Conference On Natural Computation (ICNC).* pp. 192-196 (2014,8)

[15] Forney, E. & Anderson, C. Classification of EEG during imagined mental tasks by forecasting with Elman Recurrent Neural Networks. *The 2011 International Joint Conference On Neural Networks.* pp. 2749-2755 (2011,7)

[16] Lin, C. & Lee, C. Neural Fuzzy Systems: A Neuro-fuzzy Synergism to Intelligent Systems. (Prentice-Hall, Inc.,1996)

[17] Lecun, Y., Bottou, L., Orr, G. & Müller, K. Efficient BackProp. (1998)

[18] Samin, R., Kasmani, R., Khamis, A. & Isa, S. Forecasting Sunspot Numbers with Recurrent Neural Networks (RNN) Using 'Sunspot Neural Forecaster' System. *2010 Second International Conference On Advances In Computing, Control, And Telecommunication Technologies.* pp. 10-14 (2010,12)

[19] Fang, C., Wang, X., Murphey, Y., Weber, D. & MacNeille, P. Specific humidity forecasting using recurrent Neural Network. *2014 International Joint Conference On Neural Networks (IJCNN).* pp. 955-960 (2014,7)

[20] Siddaramdhwara, N., Yelamali, A. & Byahatti, K. Electricity Short Term Load Forecasting Using Elman Recurrent Neural Network. *2010 International Conference On Advances In Recent Technologies In Communication And Computing.* pp. 351-354 (2010,10)

[21] Xia, C., Yang, Z. & Li, H. Electric load forecasting using virtual instrument based on dynamic recurrent Elman neural network. *2012 Power Engineering And Automation Conference.* pp. 1-4 (2012)

[22] Anbazhagan, S. & Kumarappan, N. Day-Ahead Deregulated Electricity Market Price Forecasting Using Recurrent Neural Network. *IEEE Systems Journal.* **7**, 866-872 (2013,12)

[23] Khan, G., Khattak, A., Zafari, F. & Mahmud, S. Electrical load forecasting using fast learning recurrent neural networks. *The 2013 International Joint Conference On Neural Networks (IJCNN).* pp. 1-6 (2013,8)

[24] Garcia-Pedrero, A. & Gomez-Gil, P. Time series forecasting using recurrent neural networks and wavelet reconstructed signals. *2010 20th International Conference On Electronics Communications And Computers (CONIELECOMP).* pp. 169-173 (2010,2)

[25] Artigue, G., Johannet, A., Borrell, V. & Pistre, S. Flash floods forecasting without rainfalls forecasts by recurrent neural networks. Case study on the Mialet basin (Southern France). *2011 Third World Congress On Nature And Biologically Inspired Computing.* pp. 303-310 (2011,10)

[26] Capizzi, G., Napoli, C. & Bonanno, F. Innovative Second-Generation Wavelets Construction With Recurrent Neural Networks for Solar Radiation Forecasting. *IEEE Transactions On Neural Networks And Learning Systems.* **23**, 1805-1815 (2012,11)

[27] Krichene, E., Masmoudi, Y., Alimi, A., Abraham, A. & Chabchoub, H. Forecasting Using Elman Recurrent Neural Network. *Intelligent Systems Design And Applications.* pp. 488-497 (2017)

[28] Dhahri, H., Alimi, A. & Abraham, A. Designing of Beta Basis Function Neural Network for optimization using cuckoo search (CS). *2014 14th International Conference On Hybrid Intelligent Systems.* pp. 110-116 (2014,12)

[29] Connor, J., Martin, R. & Atlas, L. Recurrent neural networks and robust time series prediction. *IEEE Transactions On Neural Networks.* **5**, 240-254 (1994,3)

[30] Chouikhi, N., Ammar, B., Rokbani, N. & Alimi, A. PSO-based analysis of Echo State Network parameters for time series forecasting. *Appl. Soft Comput..* **55** pp. 211-225 (2017)

[31] Ceperic, V. & Baric, A. Reducing Complexity of Echo State Networks with Sparse Linear Regression Algorithms. *2014 UKSim-AMSS 16th International Conference On Computer Modelling And Simulation.* pp. 26-31 (2014,3)

[32] Soh, H. & Demiris, Y. Spatio-Temporal Learning With the Online Finite and Infinite Echo-State Gaussian Processes. *IEEE Transactions On Neural Networks And Learning Systems.* **26**, 522-536 (2015,3)

[33] Liu, F., Ng, G. & Quek, C. RLDDE: A Novel Reinforcement Learning-based Dimension and Delay Estimator for Neural Networks in Time Series Prediction. *Neurocomput..* **70**, 1331-1341 (2007,3)

[34] Wang, J., Z, L., Guo, Q. & Yi, Z. Recurrent Neural Networks With Auxiliary Memory Units. *IEEE Transactions On Neural Networks And Learning Systems.* **PP** pp. 1-10 (2017,3)

[35] Shao, C. & Wang, L. Quantitative and Qualitative Forecasting Applied to Supply-Chain Inventory. *MultiMedia And Information Technology, International Conference On.* **2** pp. 184-187 (2010,4)

[36] Cerjan, M., Krzelj, I., Vidak, M. & Delimar, M. A literature review with statistical analysis of electricity price forecasting methods. *IEEE EuroCon 2013.* pp. 756-763 (2013,7)

[37] Maqsood, I., Khan, M. & Abraham, A. Weather Forecasting Models Using Ensembles of Neural Networks. *Intelligent Systems Design And Applications.* pp. 33-42 (2003)

[38] Turner, L. Practical tourism forecasting: D.C. Frechtling, 1996, (Butterworth Heineman, Oxford), + 240 pp., Softcover, ISBN 0 7506 0877 3. *International Journal Of Forecasting.* **13**, 296-297 (1997)

[39] Song, H., Witt, S. & Li, G. Modelling and Forecasting the Demand for Thai Tourism. *Tourism Economics.* **9**, 363-387 (2003)

[40] Montgomery, A., Zarnowitz, V., Tsay, R. & Tiao, G. Forecasting the U.S. Unemployment Rate. *Forecasting Journal Of The American Statistical Association.* **93**, 478-493 (1998)

[41] Burger, C., Dohnal, M., Kathrada, M. & Law, R. A practitioners guide to time-series methods for tourism demand forecasting - a case study of Durban, South Africa. *Tourism Management.* **22**, 403 - 409 (2001)

[42] Dalkey, N. & Helmer, O. An experimental application of the Delphi method to the use of experts. *Santa Monica, CA: RAND Corporation.* (1963)

[43] Archer, B. Forecasting demand: Quantitative and intuitive techniques. *International Journal Of Tourism Management.* **1**, 5 - 12 (1980)

[44] Liu, J. Hawaii tourism to the year 2000: A Delphi forecast. *Tourism Management.* **9**, 279 - 290 (1988)

[45] Brown, B. Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts. *Santa Monica, CA: RAND Corporation.* (1968)

[46] Sackman, H. Delphi Critique; Expert Opinion, Forecasting, and Group Process. (Lexington Books,1974)

[47] Brown, B. Survey Methods and Practices. (Canada,2003)

[48] Green, K. & Armstrong, J. Structured analogies for forecasting. *International Journal Of Forecasting.* **23**, 365 - 376 (2007)

[49] Brown, R. Exponential Smoothing for Predicting Demand. (Little,1956)

[50] Chou, Y. Statistical Analysis. (Holt International,1975)

[51] Abraham, B. & Ledolter, J. Statistical forecasting methods. (New Jersey,1983)

[52] Hyndman, R. & Athanasopoulos, G. Forecasting : principles and practice. (OTexts.com [Heathmont?, Victoria],2014)

[53] Brook, T. Direct and Reverse Methods of Incrementation . (London,1715)

[54] Chaabane, C., Mellouli, D., Hamdani, T., Alimi, A. & Abraham, A. Wavelet Convolutional Neural Networks for Handwritten Digits Recognition. *Hybrid Intelligent Systems - 17th International Conference On Hybrid Intelligent Systems (HIS 2017) Held In Delhi, India, December 14-16, 2017.* pp. 305-310 (2017)

[55] Mellouli, D., Hamdani, T., Ayed, M. & Alimi, A. Morph-CNN: A Morphological Convolutional Neural Network for Image Classification. *Neural Information Processing - 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part II.* pp. 110-117 (2017)

[56] Baccour, L., Alimi, A. & John, R. Intuitionistic Fuzzy Similarity Measures and Their Role in Classification. *J. Intelligent Systems.* **25**, 221-237 (2016)

[57] Dhahri, H. & Alimi, A. The Modified Differential Evolution and the RBF (MDE-RBF) Neural Network for Time Series Prediction. *Proceedings Of The International Joint Conference On Neural Networks, IJCNN 2006, Part Of The IEEE World Congress On Computational Intelligence, WCCI 2006, Vancouver, BC, Canada, 16-21 July 2006.* pp. 2938-2943 (2006)

[58] Bouaziz, S., Dhahri, H., Alimi, A. & Abraham, A. A hybrid learning algorithm for evolving Flexible Beta Basis Function Neural Tree Model. *Neurocomputing.* **117** pp. 107-117 (2013)

[59] Basterrech, S., Alba, E. & Snásel, V. An Experimental Analysis of the Echo State Network Initialization Using the Particle Swarm Optimization. *CoRR.* **abs/1501.00436** (2015)

[60] Box, G. & Jenkins, G. Time Series Analysis, Forecasting and Control. (Holden-Day, Inc.,1990)

[61] Dhieb, T., Ouarda, W., Boubaker, H., Ben Halima, M. & Alimi, A. Online Arabic Writer Identification based on Beta-Elliptic Model. (2015,12)

[62] Ouarda, W., Trichili, H., Alimi, A. & Solaiman, B. Bag of face recognition systems based on holistic approaches. *2015 15th International Conference On Intelligent Systems Design And Applications (ISDA).* pp. 201-206 (2015)

[63] Yao, W., Huang, P. & Jia, Z. Multidimensional LSTM Networks to Predict Wind Speed. (2018,7)

[64] Sagheer, A. & Kotb, M. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing.* **323** pp. 203 - 213 (2019), http://www.sciencedirect.com/science/article/pii/S0925231218311639

32

[65] Cui, Z., Ke, R. & Wang, Y. Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction. *ArXiv.* **abs/1801.02143** (2018)

[66] Ergen, T. & Kozat, S. Efficient Online Learning Algorithms Based on LSTM Neural Networks. *IEEE Transactions On Neural Networks And Learning Systems.* **29**, 3772-3783 (2018)

[67] Althelaya, K., El-Alfy, E. & Mohammed, S. Evaluation of bidirectional LSTM for short-and long-term stock market prediction. *2018 9th International Conference On Information And Communication Systems (ICICS).* pp. 151-156 (2018)

[68] Tokgöz, A. & Ünal, G. A RNN based time series approach for forecasting turkish electricity load. *2018 26th Signal Processing And Communications Applications Conference (SIU).* pp. 1-4 (2018)

[69] Saini, U., Kumar, R., Jain, V. & Krishnajith, M. Univariant Time Series forecasting of Agriculture load by using LSTM and GRU RNNs. *2020 IEEE Students Conference On Engineering Systems (SCES).* pp. 1-6 (2020)

[70] Tealab, A. Time series forecasting using artificial neural networks methodologies: A systematic review. *Future Computing And Informatics Journal.* **3**, 334 - 340 (2018), http://www.sciencedirect.com/science/article/pii/S2314728817300715

[71] Wang, J., Wang, J., Fang, W. & Niu, H. Financial Time Series Prediction Using Elman Recurrent Random Neural Networks. *Computational Intelligence And Neuroscience.* **2016** pp. 1-14 (2016,1)

[72] Zainorzuli, S., Afzal Che Abdullah, S., Adnan, R. & Ruslan, F. Comparative Study of Elman Neural Network (ENN) and Neural Network Autoregressive With Exogenous Input (NARX) For Flood Forecasting. *2019 IEEE 9th Symposium On Computer Applications Industrial Electronics (ISCAIE).* pp. 11-15 (2019)

[73] Mengying, H., Jiandong, D., Zequan, H., Peng, W., Shuai, F., Peijia, H. & Chaoyuan, F. Monthly Electricity Forecast Based on Electricity Consumption Characteristics Analysis and Multiple Effect Factors. *2019 IEEE 8th International Conference On Advanced Power System Automation And Protection (APAP).* pp. 1814-1818 (2019)

[74] Di, P., Dong, K., Du, J., Dong, C., He, X., Guan, Y., Gao, H., Li, J. & Liang, Y. Ultra-Short Term Load Forecasting Based on Elman Neural Network. *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia).* pp. 911-915 (2019)

[75] Xu, L. & Mao, J. Short-term wind power forecasting based on Elman neural network with particle swarm optimization. *2016 Chinese Control And Decision Conference (CCDC)*. pp. 2678-2681 (2016)

[76] Su, Y., Wang, S., Xiao, Z., Tan, M. & Wang, M. An Ultra-Short-Term Wind Power Forecasting Approach Based on Wind Speed Decomposition, Wind Direction and Elman Neural Networks. *2018 2nd IEEE Conference On Energy Internet And Energy System Integration (EI2)*. pp. 1-9 (2018)

[77] Vardhan, N. & Chintham, V. Electricity price forecasting of deregulated market using Elman Neural Network. *2015 Annual IEEE India Conference (INDICON)*. pp. 1-5 (2015)

[78] Khan, I., Zhu, H., Yao, J. & Khan, D. Photovoltaic power forecasting based on Elman Neural Network software engineering method. *2017 8th IEEE International Conference On Software Engineering And Service Science (ICSESS)*. pp. 747-750 (2017)

[79] Gozalpour, N. & Teshnehlab, M. Forecasting Stock Market Price Using Deep Neural Networks. *2019 7th Iranian Joint Congress On Fuzzy And Intelligent Systems (CFIS)*. pp. 1-4 (2019)

[80] Baziyad, M., Jarndal, A. & Bettayeb, M. A Model Order Reduction Technique Based on Balanced Truncation Method and Artificial Neural Networks. *2019 8th International Conference On Modeling Simulation And Applied Optimization (ICMSAO)*. pp. 1-5 (2019)

[81] Tsuji, T. Model reduction with time delay combining the least-squares method with the genetic algorithm. *IEE Proceedings - Control Theory And Applications*. **143**, 247-254(7) (1996,5), https://digital-library.theiet.org/content/journals/10.1049/ip-cta19960361

[82] Figueredo, A. & Wolf, P. Assortative pairing and life history strategy - a cross-cultural study.. *Human Nature*. **20** pp. 317-330 (2009)

[83] Science, A. Designs for Science Literacy. (Oxford University Press,2001,3)

[84] Albert, L. Curriculum design: Finding a balance. *Journal Of Rheumatology*. **34**, 458-459 (2007)

[85] Archambault, I., Janosz, M., Fallu, J. & Pagani, L. Student engagement and its relationship with early high school dropout. *Journal Of Adolescence*. **32**, 651-670 (2009)

[86] Busato, V., Prins, F., Elshout, J. & Hamaker, C. Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education . *Personality And Individual Differences* . **29**, 1057 - 1068 (2000), http://www.sciencedirect.com/science/article/pii/S0191886999002536

[87] Campbell, J., DeBlois, P. & Oblinger, D. Academic analytics: A new tool for a new era. *Educause Review.* **42**, 40 (2007)

[88] Caulkins, J., Larkey, P. & Wei, J. Adjusting GPA to reflect course difficulty. (1996)

[89] Denton, J., Franke, V. & Surendra, K. Curriculum and course design: a new approach using quality function deployment. *Journal Of Education For Business.* **81**, 111-117 (2005)

[90] Ester, M., Kriegel, H. & Schubert, M. Web site mining: a new way to spot competitors, customers and suppliers in the world wide web. *Proceedings Of The Eighth ACM SIGKDD International Conference.* pp. 249-258 (2002)

[91] Ferguson, R. Learning analytics: drivers, developments and challenges. *International Journal Of Technology Enhanced Learning.* **4**, 304-317 (2012)

[92] Jordan, W., Lara, J. & McPartland, J. Exploring the causes of early dropout among race-ethnic and gender groups. *Youth & Society.* **28**, 62-94 (1996)

[93] Koning, B., Loyens, S., Rikers, R., Smeets, G. & Molen, H. Generation Psy: Student characteristics and academic achievement in a three-year problem-based learning bachelor program . *Learning And Individual Differences* . **22**, 313 - 323 (2012), http://www.sciencedirect.com/science/article/pii/S104160801200012X

[94] McNeal Jr, R. Extracurricular activities and high school dropouts. *Sociology Of Education.* pp. 62-80 (1995)

[95] Parthasarathy, S., Zaki, M., Ogihara, M. & Dwarkadas, S. Incremental and interactive sequence mining. *Proceedings Of The Eighth International Conference On Information And Knowledge Management.* pp. 251-258 (1999)

[96] Pei, J., Han, J., Mortazavi-Asl, B. & Zhu, H. Mining access patterns efficiently from web logs. *Knowledge Discovery And Data Mining. Current Issues And New Applications.* pp. 396-407 (2000)

[97] Pukkila, P., DeCosmo, J., Swick, D. & Arnold, M. How to engage in collaborative curriculum design to foster undergraduate inquiry and research

in all disciplines. *Developing And Sustaining A Research-Supportive Curriculum: A Compendium Of Successful Practices*. pp. 341-357 (2007)

[98] R Core Team R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing,2013), http://www.R-project.org/

[99] Revelle, W. psych: Procedures for Psychological, Psychometric, and Personality Research. (Northwestern University,2013), http://CRAN.R-project.org/package=psych, R package version 1.3.10

[100] Rumberger, R. Dropping out of high school: The influence of race, sex, and family background. *American Educational Research Journal*. **20**, 199-220 (1983)

[101] Siemens, G. & Long, P. Penetrating the fog: Analytics in learning and education. *Educause Review*. **46**, 30-32 (2011)

[102] Singh, K., Granville, M. & Dika, S. Mathematics and Science Achievement: Effects of Motivation, Interest, and Academic Engagement. *The Journal Of Educational Research*. **95**, 323-332 (2002), http://www.tandfonline.com/doi/abs/10.1080/00220670209596607

[103] Sunderman, J. Curriculum Incubation : Data-driven Innovative Instructional Design. *ASEE Annual Conference*. (2012)

[104] Tabachnick, B. & Fidell, L. Using Multivariate Statistics: International Edition. (Pearson,2012)

[105] Turner, F., Clutterbuck, D., Semple, C. & Others POCUS: mining genomic sequence annotation to predict disease genes. *Genome Biology*. **4**, R75-R75 (2003)

[106] Wolf, P. A model for facilitating curriculum development in higher education: A faculty-driven, data-informed, and educational developer–supported approach. *New Directions For Teaching And Learning*. **2007**, 15-20 (2007)

[107] Wolff, A., Zdrahal, Z., Nikolov, A. & Pantucek, M. Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. *Proceedings Of The Third International Conference On Learning Analytics And Knowledge*. pp. 145-149 (2013)

[108] Zaki, M. SPADE: An efficient algorithm for mining frequent sequences. *Machine Learning*. **42**, 31-60 (2001)