

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

# High-Performance Time Series Prediction with Predictive Error Compensated Wavelet Neural Networks

**BERK USTUNDAG<sup>1</sup> (Member, IEEE), AJLA KULAGLIC<sup>2</sup> (Member, IEEE)**

<sup>1</sup>Computer Engineering Department, Istanbul Technical University, Istanbul, Turkey

<sup>2</sup>Graduate School of Science Engineering and Technology, Istanbul Technical University, Istanbul, Turkey

Corresponding author: Ajla Kulaglic (e-mail: kulaglic@itu.edu.tr).

**ABSTRACT** Machine learning (ML) algorithms have gained prominence in time series prediction problems. Depending on the nature of the time series data, it can be difficult to build an accurate ML model with the proper structure and hyperparameters. In this study, we propose a predictive error compensation wavelet neural network model (PEC-WNN) for improving the prediction accuracy of chaotic and stochastic time series data. In the proposed model, an additional network is used for the prediction of the main network error to compensate the overall prediction error. The main network takes as inputs the time series data through moving frames in multiple-scales. The same structure and hyperparameter sets are applied for quite distinct four types of problems for verification of the robustness and accuracy of the proposed model. Specifically, the Mackey-Glass, Box-Jenkins, and Lorenz Attractor benchmark problems, as well as drought forecasting are used to characterize the performance of the model for chaotic and stochastic data cases. The results show that the PEC-WNN provides significantly more accurate predictions for all compared benchmark problems with respect to conventional machine learning and time series prediction methods without changing any hyperparameter or the structure. In addition, the time and space complexity of the PEC-WNN model is less than all other compared ML methods, including long short-term memory (LSTM) and convolutional neural networks (CNNs).

**INDEX TERMS** Box-Jenkins, Discrete Wavelet Transform, Drought Forecasting, Lorenz Attractor, Mackey-Glass, Neural Networks, Predictive Error Compensated Wavelet Neural Network, Standardized Precipitation Evapotranspiration Index (SPEI), Time Series

## I. INTRODUCTION

Time series prediction is an important area that has attracted the attention of researchers from different fields, such as business, economics, finance, science, and engineering [1], [2].

In this study, we propose an efficient ML structure for time series prediction problems that provide considerably higher accuracy and low time complexity with respect to conventional algorithms such as long short-term memory (LSTM) networks, and convolutional neural networks (CNNs). Besides, the proposed algorithm has capability to find accurate solutions for different types of problems without changing the hyperparameter set or the network structure.

The main aim of time series prediction is to collect and analyze the past observations of the time series data to develop a model that describes the behavior of the relevant

system [1]. Time series problems found in the literature, such as sunspots, runoff, electric loads, temperature, gas furnace, drought, and rainfall, can be interpreted through chaos theory [1] - [4]. The various methods for time series prediction have been developed by using linear models. A conventional statistic methods such as Auto-Regressive (AR) and Autoregressive Integrated Moving Average (ARIMA) assume linear relationships between past values. The ARIMA models are relatively robust and more efficient than complex models. Development and implementation of linear methods are relatively simple, however, they are not able to capture non-linear relationships in the data [5].

In recent decades, ML methods, including artificial neural network (ANN) models have attracted more attention in the domain of time series forecasting. These models have been widely used compared to various traditional time series

models. The ANNs models are intended to resolve non-linear functional dependencies between the time series data in the past and its future [6]. ANN models can be classified based on network structure into feed-forward and recurrent neural networks [7]. The most used feedforward neural networks in time series predictions are the multilayer perceptrons (MLPs) [5]. The MLP structure demands a large number of parameters to solve complex non-linear problems. This results in a low learning rate and poor generalization [5]. The prediction of time series data to achieve better accuracy requires the NN models to be adaptive to changes that occur over time in the data. Several neural networks (NNs) and their modified models have been applied for non-linear time series prediction to overcome these drawbacks [8], [9], [10], [11], [12], [13], [14], [15].

Convolutional neural networks (CNNs) [16] are widely used for learning nonlinear mapping functions from complex data. They can be applied to a variety of problems, from image data, time series to the outputs. The CNNs can learn and extract the most important features due to special convolutional operations.

Recurrent neural networks (RNNs) possess an internal memory and makes them capable of incorporating changes due to internal recurrence [17]. RNNs are computationally more powerful than feedforward networks. Despite the efficiency of NN, CNN, and RNN models in the prediction of time series, two main problems can be addressed. The performance of networks highly depends on the architecture and hyperparameters of networks. The appropriate design of CNN, RNN, and NN models becomes more difficult regarding the nature of time series data. Therefore, the prediction performance is affected by appropriate network parameters.

In this study, we propose a predictive error compensated wavelet neural network (PEC-WNN) model consisting of two NNs. The motivation for using two separated NNs comes from the following perspectives. Firstly, the forecasting models are facing with expanding uncertainties such as the lack of information for making more accurate predictions and the accumulation of errors. A well-known drawback in the recursive methods is sensitivity to the estimation errors since their predicted values are used in the model instead of the target values [18]. In contrary, in the proposed approach the models are trained independently and hence not prone to accumulate errors. We show that compensating the predicted error through a second NN enhances the overall prediction performance. The PEC-WNN uses time series input data in multiple-time windows. Sampled time series data in the moving time window are first transformed into a set of wavelet coefficients using a discrete wavelet transform (DWT) and then fed into the NNs. DWT is applied separately to each window by analyzing signal in time as well as in frequency domain. The results show that using a multi-dimensional time window improves the prediction performance without increasing the algorithm complexity. The PEC-WNN improves accuracy while at the same time prevents overfitting by taking the advantage of multi-resolution DWT and NN.

The main contributions of the proposed method can be summarized as:

1. Improvement of the prediction accuracy for chaotic and stochastic time series data using multiple neural networks where the secondary network is trained by shifted time series prediction error of the primary network so that overfitting can be avoided due to increase of recurrence related feedback input.
2. The same structure and hyperparameter sets can be applied for broad range of time series prediction problems with moving frames in multiple-scales.
3. The DWT yields better accuracy improvement than directly applying the time series data to the neural network in predictive error compensation.

In the next section, we explained the proposed PEC-WNN model for time series prediction. The time series problems: the Mackey-Glass chaotic time series, the gas furnace data (series J) of the Box-Jenkins benchmark problem, the Lorenz Attractor time series data, and drought forecasting problems are provided in Section III along the corresponding results. The time and space complexity of the proposed model with respect to the models found in the literature have been discussed in section IV Section V presents the concluding remarks.

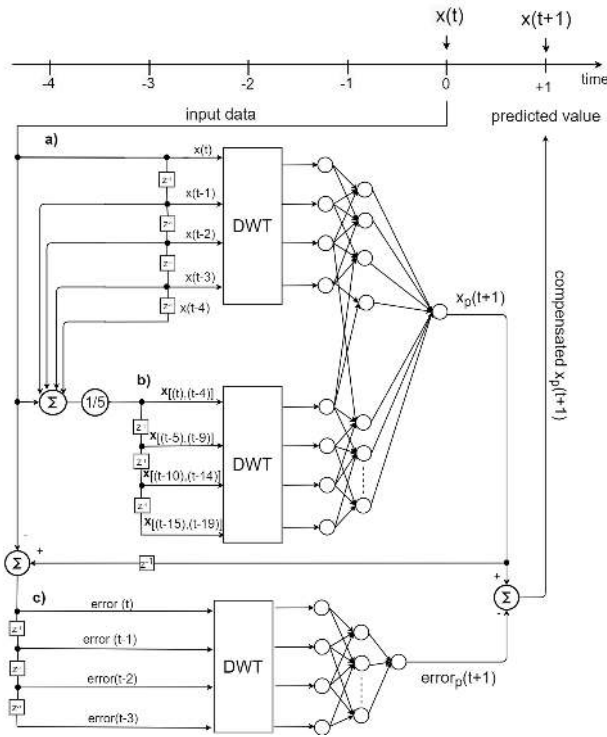
## II. PREDICTIVE ERROR COMPENSATED WAVELET NEURAL NETWORK MODEL

The predictive error compensated wavelet neural network model (PEC-WNN) utilized in this study comprises of two separate wavelet preprocessed neural networks, as demonstrated in Fig. 1. The current input is shifted to the previous value using the unit delay operator  $z^{-1}$  (see Fig. 1.a)). Along with the four consecutive values, we compute the average values of the different time intervals obtained by applying the unit delay operator (Fig. 1.b)), in the same manner as in (Fig. 1.a)). The input data consist of two different time windows that are preprocessed in accordance with the time frame using the discrete wavelet transform (DWT). The rationale for applying DWT is due to its ability to analyze a signal both in time and frequency domains. Unlike FT that provides insight just into frequency content, the wavelet analysis can automatically adapt itself to a suitable resolution and overcome the limitations found in the FT [19], [20].

The DWT is a linear signal processing technique that transforms a signal from the time domain to the "wavelet" domain. The wavelets characterize a family of functions generated from one single function  $\psi(t)$  by the operation of dilation and translation. The mother wavelet function localized both in time and frequency domain is represented by  $\psi(t)$ , a scaling function is defined by  $\varphi(t)$  and the parameters  $j_i$  and  $k_i$  can be defined as the scale and translation parameters to generate the families of functions, respectively, as given in (1) and (2).

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}} \psi(2^{-j}t - k) \quad (1)$$

$$\varphi_{j,k}(t) = 2^{-\frac{j}{2}} \varphi(2^{-j}t - k) \quad (2)$$



**FIGURE 1.** The Predictive Error Compensated Wavelet Neural Network model (PEC-WNN): a) Consecutive input values shifted using the unit delay operator  $z^{-1}$ , b) Average values of size five computed applying the unit delay operator three times, c) Forecasting error calculation and input preparation for the compensation network.

Formerly, we sample the data into a window  $W_i$  as in (3):

$$W_i = \begin{cases} \text{None}, & i < l_x \\ x(i - l_x + 1), x(i - l_x + 2), \dots, x(i), & i > l_x \end{cases} \quad (3)$$

where  $x(i)$  is the  $i$ -th value of the input,  $l_x$  is the length of the signal part using the current window. Based on the selected scaling functions and multi-resolution analysis, we obtain the double-scaling equations as given in (4) and (5):

$$\psi(2^{-j}t) = \sqrt{2} \sum_k h_k \psi(2^{-(j-1)}t - k) \quad (4)$$

$$\varphi(2^{-j}t) = \sqrt{2} \sum_k g_k \varphi(2^{-(j-1)}t - k) \quad (5)$$

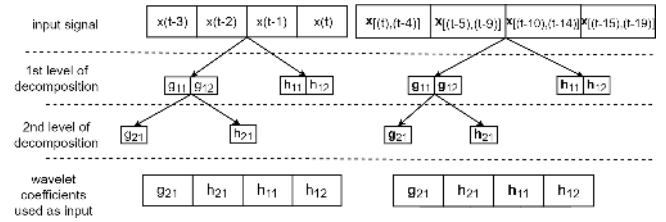
Applying the DWT to the current window, we compute the scaling coefficients  $c_{i,j,k}$ , and the wavelet coefficients  $d_{i,j,k}$  using the following equations:

$$c_{i,j,k} = \sum_n h_{n-2k} c_{i,j-1,n} \quad (6)$$

$$d_{i,j,k} = \sum_n g_{n-2k} c_{i,j-1,n} \quad (7)$$

where the coefficients  $h_{n-2k}$  and  $g_{n-2k}$  can be performed by using (4) and (5):

$$h_{n-2k} = \langle \varphi_{j,k}, \varphi_{j-1,k} \rangle \quad (8)$$



**FIGURE 2.** The DWT decomposition of the input signal using Mallat's algorithm.

$$g_{n-2k} = \langle \psi_{j,k}, \psi_{j-1,k} \rangle \quad (9)$$

Equations (6) and (7) are mathematical expressions of filtering a signal through a high-pass ( $h[n]$ ) and low-pass ( $g[n]$ ) filters, which corresponds to convolution with an impulse response of  $k$ -tap filters. Subsequently, the signal reconstruction can be computed by:

$$\hat{c}_{i,j-1,n} = \sum_k h_{n-2k} c_{i,j,k} + \sum_k g_{n-2k} d_{i,j,k} \quad (10)$$

In this work, as a common wavelet basis function, the symmetric Haar wavelet function is used. It beneficially diminishes the distortion rate during the signal decomposition and the signal reconstruction. The Haar wavelet function also reduces the processing and computational time significantly [21]. The Mallat's pyramidal algorithm that provides high- ( $h_n$ ) and low- ( $g_n$ ) frequencies from a given signal that are used for the decomposition of the input signal. The low- and high- frequency components are used together as input to the forecasting model to capture valuable information during the training process. A block diagram of multilevel wavelet decomposition is presented in Fig. 2, together with the coefficients used as input to the first NN.

The prediction of  $n$ -step-ahead time series data is obtained by using the main network characterized by three layers: input, hidden, and output layer. Mathematically, a hidden layer with activation function  $g(\cdot)$  and  $k$  hidden neurons can be represented as given in (11):

$$g\left(\sum_{i=1}^n w_{ji} \cdot x_i + b_j\right) = y_j, \quad j = 1, 2, 3, \dots, k, \quad (11)$$

where  $w_{ji} = [w_{j1}, w_{j2}, \dots, w_{jk}]^T$  characterizes the weight vector that connects the  $j$ th hidden neurons with the inputs, and  $b_j$  is the bias value of the  $j$ th hidden neuron. The result of the  $j$ th output neuron mathematically can be computed as represented in (12):

$$g\left(\sum_{i=1}^k \beta_{ji} \cdot y_i + b_j\right) = O_j, \quad j = 1, 2, 3, \dots, N, \quad (12)$$

where  $\beta_{ji} = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jm}]^T$  denotes the weight vector connecting the  $j$ th hidden and output neurons, and  $b_j$  is the bias value of the  $j$ th output neuron.

The total number of output neurons is given by  $N$ . The activation function  $g(\cdot)$  approximate the relationship between

the input  $x_i$  and the output, target  $t_i$ . Consequently, there are  $\beta_i$ ,  $w_i$ , and  $b_i$  such that:

$$\sum_{i=1}^n \beta_{ji} g(w_i \cdot x_j + b_j) = t_j, \quad j = 1, 2, 3, \dots, n \quad (13)$$

The employed network uses Rectified Linear Unit (ReLU) activation function. The ReLU activation function, compared to the widely used activation functions (sigmoid and hyperbolic tangent), significantly improves the performance of the feed-forward networks [22]. The ReLU is a linear function that returns the value provided as input if the value is higher than zero as given in (14).

$$f(x) = x^+ = \max(0, x) \quad (14)$$

The stochastic gradient descent (SGD) is used for optimizing, where the learning rate and momentum are 0.05 and 0.75, respectively. The SGD maintains a single learning rate for all weight updates without varying during the training. The learning rate is maintained for each network weight, whereas it is distinctly adopted as learning folds. Secondly, the PEC-WNN is used to improve forecasting performances obtained in the first NN. The input data of the second NN is constructed using the DWT preprocessed prediction errors. The prediction error at a time  $(t + 1)$  is shifted applying the unit delay operator  $z^{-1}$  (Fig. 1.c)). The output of the second NN is the prediction of the error at a time  $(t + 1)$  using the prediction errors at time  $(t)$ ,  $(t - 1)$ ,  $(t - 2)$  and  $(t - 3)$ .

Finally, the predictive value from the first NN at a time  $(t + 1)$  and predictive error  $(t + 1)$  from the second NN, are used together to acquire the compensated predictive value at a time  $(t + 1)$ . The main equations of the proposed model can be expressed as given below:

$$x_p(t + 1) = f(x(t), x(t - 1), x(t - 2), x(t - 3), \bar{x}_{[t, t-4]}, \bar{x}_{[t-5, t-9]}, \bar{x}_{[t-10, t-14]}, \bar{x}_{[t-15, t-19]}) \quad (15)$$

$$\overline{x[i, j]} = \frac{\sum_i^j x_i}{(j - i) + 1} \quad (16)$$

$$error_p(t + 1) = f(err(t), err(t - 1), err(t - 2), err(t - 3)) \quad (17)$$

$$error(t) = real\ value(t) - forecasted\ value(t) \quad (18)$$

$$compensated\ value(t + 1) = x_p(t + 1) - error_p(t + 1) \quad (19)$$

where  $x_p(t + 1)$  characterizes predicted value at time  $(t + 1)$ . Four consecutive values,  $(x(t), x(t - 1), x(t - 2), x(t - 3))$  and its average values of size five  $(\overline{x[t, t - 4]}, \overline{x[t - 5, t - 9]}, \overline{x[t - 10, t - 14]}, \overline{x[t - 15, t - 19]})$  computed using (15), represent the input of the first NN. The average values of an interval  $[i, j]$  are computed using (16). The input data of the second NN contains four errors obtained using the predicted values from the first NN using (17). The compensated predicted value is computed by subtracting the predictive value at a time  $(t + 1)$  from the predictive error at  $(t + 1)$ , which is given in the (19).

TABLE 1. The mathematical equations of used evaluation metrics.

Evaluation metrics:	Formulation
MAPE(%)	$\frac{100}{n} \sum_{i=1}^n \frac{ X_{model,i} - X_{obs,i} }{X_{obs,i}}$
RMSE	$\sqrt{\frac{\sum_{i=1}^n (X_{obs,i} - X_{model,i})^2}{n}}$
DA(%)	$\frac{100}{n} \sum_{i=0}^n d_i$

### III. MATERIALS AND RESULTS

The performances of the proposed model are verified using the Mackey-Glass, the Box-Jenkins gas furnace (series J), the Lorenz Attractor time series data, and for the drought forecasting problem the global SPEI index. The data sets are applied to the different models such as simple neural network model (hereafter NN), predictive error compensated neural network model (PEC-NN), wavelet neural network (WNN), and predictive error compensated wavelet neural network (PEC-WNN).

The data sets are scaled by using the minimum/maximum normalization method given in (20):

$$x(t)_{norm} = (max_2 - min_2) \cdot \left( \frac{x(t) - min_1}{max_1 - min_1} \right) + min_2 \quad (20)$$

where  $x(t)$  represents the real value,  $min_1$  and  $max_1$  are the minimum and maximum values of observations. The  $max_2$  and the  $min_2$  refer to the desired maximum and minimum of the new scaled values. The mean absolute percentage error (MAPE), root-mean-square error (RMSE), and directional accuracy (DA) are used for the comparison of the experimental results. In Tab. 1, the mathematical formulations of evaluation metrics are shown.

where  $X_{obs}$  is observed value and  $X_{model}$  is modeled value in time  $i$ . The number of data samples is given by  $n$ . The  $d_i$  is given by:

$$W_i = \begin{cases} 1 & (X_{model,i} - X_{model,i-1})(X_{obs,i} - X_{obs,i-1}) \geq 0 \\ 0 & otherwise \end{cases} \quad (21)$$

#### A. THE MACKEY-GLASS CHAOTIC TIME SERIES DATA

The chaotic Mackey-Glass time series data (Fig. 3) has been typically used as a benchmark problem before considering the suitability of a specific approach to real-world forecasting problems [23]. The time series data have been generated from the following differential equation (22):

$$\frac{dx(t)}{dt} = \frac{\alpha x(t - \tau)}{1 + x^{10}(t - \tau)} - bx(t) \quad (22)$$

where  $x$  (unitless) is the series in time  $t$ , and  $\tau$  is the time delay. The parameters  $\alpha$ ,  $\beta$  and  $\tau$  are set as  $\alpha = 0.2$ ,  $\beta = 0.1$ ,  $\tau = 17$ . Note that, for  $\tau > 17$ , the time series show chaotic behavior [24]. The initial condition  $x(0) = 1.2$  is used to generate the data points by using the fourth-order Runge-Kutta method with time step 0.1.

The work from [24], uses non-consecutive values with the constant time interval,  $T = 6$  for prediction of the short-term outputs. They performed experiments by considering inputs as  $x(t - 18)$ ,  $x(t - 12)$ ,  $x(t - 6)$ ,  $x(t)$  to predict  $x(t + 6)$ . Out of 1000 samples, the authors used 500 for training the model and 500 for testing performance. Similarly, [23] considered sequential four input variables,  $x(t - 3)$ ,  $x(t - 2)$ ,  $x(t - 1)$ ,  $x(t)$  to estimate single output variable at time  $x(t + 5)$ . Out of 300 samples, half of the samples served for training and the remaining half for testing.

Different than the previous studies, in this work, we construct the data sets which contains the averages of different window sizes together with sequential values. The first data set contains only four successive values obtained by (23). In order to observe the

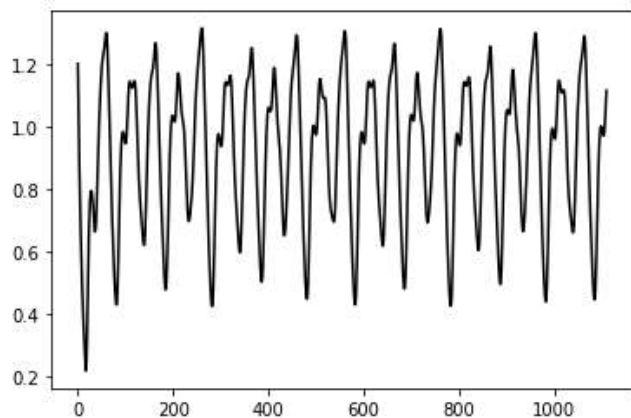


FIGURE 3. The chaotic Mackey-Glass time series data.

effects of the average values on the forecasting performance in the following data sets, we include the average values of window size 5 and size 10, obtained using equations (24) and (25), respectively, which can be considered as daily data. Hence by using the average values of size 5, we gain the business week resolution of data. Moreover, extending it to the four shifted average values, we obtain the monthly resolution.

In our study the forecasting intervals differ from the next value ( $t + 1$ ), the fifth ( $t + 5$ ), the sixth ( $t + 6$ ), the forty-second ( $t + 42$ ) until the eighty-fourth ( $t + 84$ ) value. Equally divided into training and test sets, 1000 samples are used as in [24].

$$MG_1 : y_1(t + T) = f(x(t - 3), x(t - 2), x(t - 1), x(t)) \quad (23)$$

$$MG_2 : y_2(t + T) = f(x(t), x(t - 1), x(t - 2), x(t - 3), x_{[0,4]}, x_{[5,9]}, x_{[10,14]}, x_{[15,19]}) \quad (24)$$

$$MG_3 : y_3(t + T) = f(x(t), x(t - 1), x(t - 2), x(t - 3), x_{[0,9]}, x_{[10,19]}, x_{[20,29]}, x_{[30,39]}) \quad (25)$$

The RMSE results are presented concerning the constructed data sets (presented in the above equations) and its forecasting time interval in Tab. 2. The results show that the use of averages along with consecutive values significantly reduces the error. In addition, the average values used in conjunction with the successive values improve the forecasting performances of the proposed model (PEC-WNN). The selection of window size has a huge impact, considering the predicting time interval. Small window size averages show better results for short-term predictions, while window size expansion shows better forecast performance for long-term forecasts. Similarly, in comparison to the MAPE (Tab. 3) we confirmed that appending the averages to the consecutive values improve the results and reduce the forecasting errors. The results for directional accuracy, DA in Tab. 4, shows descent results for applied models where the PEC-WNN model precedes.

The Mackey-Glass time series problem results showed that at different forecasting intervals, the PEC-WNN achieves the lowest RMSE error. The second-best result found in the literature [25], for short-term forecasting ( $t + 1$ ) is 0.0327. The same time-interval forecasting done with PEC-WNN obtain the RMSE of 0.0013, which reduces the RMSE by 95%. Similar results are noticed for ( $x + 6$ ) forecasting term, where the best result found in the literature is 0.0055 [26]. This result is obtained by using the dynamic cell

TABLE 2. The RMSE results of The Mackey-Glass time series data considering applied equations and forecasting time interval.

Forecasting interval	Equation	NN	PEC-NN	WNN	PEC-WNN
x+1	MG <sub>1</sub> (23)	0.013	0.006	0.013	0.005
	MG <sub>2</sub> (24)	0.003	0.001	0.002	0.001
	MG <sub>3</sub> (25)	0.010	0.007	0.011	0.006
x+5	MG <sub>1</sub> (23)	0.072	0.029	0.071	0.029
	MG <sub>2</sub> (24)	0.005	0.004	0.005	0.004
	MG <sub>3</sub> (25)	0.0031	0.019	0.029	0.018
x+6	MG <sub>1</sub> (23)	0.087	0.036	0.086	0.035
	MG <sub>2</sub> (24)	0.004	0.003	0.003	0.002
	MG <sub>3</sub> (25)	0.006	0.004	0.005	0.003
x+42	MG <sub>1</sub> (23)	0.136	0.042	0.131	0.034
	MG <sub>2</sub> (24)	0.033	0.026	0.024	0.014
	MG <sub>3</sub> (25)	0.018	0.012	0.014	0.010
x+84	MG <sub>1</sub> (23)	0.159	0.040	0.160	0.038
	MG <sub>2</sub> (24)	0.124	0.035	0.082	0.030
	MG <sub>3</sub> (25)	0.067	0.028	0.41	0.025

TABLE 3. The MAPE (%) results of The Mackey-Glass time series data considering applied equations and forecasting time interval.

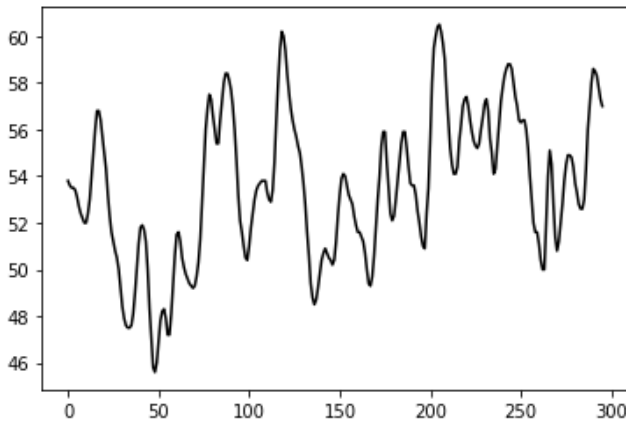
Forecasting interval	Equation	NN	PEC-NN	WNN	PEC-WNN
x+1	MG <sub>1</sub> (23)	1.253	0.547	1.266	0.486
	MG <sub>2</sub> (24)	0.264	0.152	0.200	0.118
	MG <sub>3</sub> (25)	0.925	0.552	1.001	0.643
x+5	MG <sub>1</sub> (23)	7.200	2.519	6.991	2.49
	MG <sub>2</sub> (24)	0.486	0.366	0.542	0.407
	MG <sub>3</sub> (25)	2.912	1.747	2.722	1.534
x+6	MG <sub>1</sub> (23)	8.767	3.475	8.584	3.100
	MG <sub>2</sub> (24)	0.374	0.260	0.311	0.241
	MG <sub>3</sub> (25)	0.568	0.381	0.513	0.369
x+42	MG <sub>1</sub> (23)	13.446	3.809	12.605	2.938
	MG <sub>2</sub> (24)	3.120	1.840	2.161	1.203
	MG <sub>3</sub> (25)	1.722	0.970	1.252	0.970
x+84	MG <sub>1</sub> (23)	16.307	3.660	16.351	3.524
	MG <sub>2</sub> (24)	12.113	2.517	7.935	2.782
	MG <sub>3</sub> (25)	6.610	2.692	3.535	2.414

TABLE 4. The DA (%) results of The Mackey-Glass time series data considering applied equations and forecasting time interval.

Forecasting interval	Equation	NN	PEC-NN	WNN	PEC-WNN
x+1	MG <sub>1</sub> (23)	96.267	96.673	96.493	96.719
	MG <sub>2</sub> (24)	98.484	98.592	98.672	97.09
	MG <sub>3</sub> (25)	95.682	95.482	95.227	95.341
x+5	MG <sub>1</sub> (23)	69.432	70.682	72.612	74.659
	MG <sub>2</sub> (24)	96.816	96.432	96.226	96.755
	MG <sub>3</sub> (25)	80.909	87.343	87.913	89.369
x+6	MG <sub>1</sub> (23)	62.841	63.841	62.955	64.091
	MG <sub>2</sub> (24)	97.642	98.231	97.642	98.426
	MG <sub>3</sub> (25)	95.873	95.991	94.108	96.187
x+42	MG <sub>1</sub> (23)	61.457	66.510	64.160	67.568
	MG <sub>2</sub> (24)	87.912	89.499	87.424	90.005
	MG <sub>3</sub> (25)	88.889	91.209	91.941	93.592
x+84	MG <sub>1</sub> (23)	49.572	58.752	58.384	59.241
	MG <sub>2</sub> (24)	74.268	75.924	76.815	77.834
	MG <sub>3</sub> (25)	74.268	82.293	75.921	86.242

**TABLE 5.** The RMSE error comparison for the Mackey-Glass time series data found in the literature for prediction time interval  $t = \{+1,+6,+84\}$

Models:	x+1	x+6	x+84	Source:
Linear Predictor	0.0327	0.7173	1.50351	[ [25]]
6th-order Polynom		0.04	0.85	[ [25]]
BPNN		0.02	0.06	[ [27]]
DSC-LMM		0.0055	0.03	[ [26]]
Cascade correlation NN		0.06		[ [28]]
PEC-WNN	0.0013	0.0027	0.028	
LSTM	0.0576	0.0644	0.1642	



**FIGURE 4.** The  $CO_2$  concentration in the gas in the furnace data set, the output of the Box-Jenkins time series data.

structures, and local linear models (DSC-LMM) proposed in [26]. The PEC-WNN RMSE for time interval  $(t + 6)$  was 0.0027, which is 49% better in comparison to the results found in the literature. The long-term forecasting interval  $(t + 84)$  using the proposed model achieves the RMSE of 0.028. On the other hand, for the same forecasting interval, Cudy et al. [26], reached the RMSE of 0.03 using the DSC-LMM model.

### B. THE BOX-JENKINS TIME SERIES DATA

The Box-Jenkins time series data set is another frequently used benchmark example in the prediction algorithms [24]. The method refers to the iterative application of a three-stage modeling approach: 1) model-identification and selection, 2) estimation, and 3) statistical model checking [29]. The first stage determines the stationarity of the data. The plots of the dependent time series data are used to decide which autoregressive or moving average components should be applied. In the second stage, the estimation of parameters of the selected model is obtained by using the maximum likelihood or non-linear least square estimation. In the last stage, the statistical model checking, we examine whether the model follows the conditions of a stationary univariate process. The data used in this study are well known as gas furnace data (series J) prediction problem. The output of the Box-Jenkins gas furnace time series data set is given in Fig. 4.

The inputs proposed in the literature, given as in 26 are applied first. Subsequently, we used successive values of the methane gas flow (27) to forecast the successive value. We expanded our investigation by checking how the increased amount of input data with its average values, affect the forecasting performances. For that purpose, we applied the averages of five and ten window sizes with four successive values of methane gas flow with  $CO_2$  concentration in the gas. Note that, the forecasting value is always the next  $(t + 1)$  value of  $CO_2$  concentration in the gas, while the input data set

**TABLE 6.** The RMSE results of the Box-Jenkins time series data considering applied models and equations.

Equation	NN	PEC-NN	WNN	PEC-WNN
$BJ_1$ (26)	0.082	0.081		
$BJ_2$ (27)	0.077	0.076	0.074	0.071
$BJ_3$ (28)	0.074	0.068	0.066	0.059
$BJ_4$ (29)	0.070	0.069	0.064	0.061

**TABLE 7.** The MAPE (%) results of the Box-Jenkins time series data considering applied models and equations.

Equation	NN	PEC-NN	WNN	PEC-WNN
$BJ_1$ (26)	25.268	28.022		
$BJ_2$ (27)	20.126	18.707	27.680	37.744
$BJ_3$ (28)	17.927	21.232	16.577	23.513
$BJ_4$ (29)	19.401	17.104	19.163	21.150

**TABLE 8.** The DA(%) results of the Box-Jenkins time series data considering applied models and equations.

Equation	NN	PEC-NN	WNN	PEC-WNN
$BJ_1$ (26)	90.5	90.5		
$BJ_2$ (27)	91.0	91.0	90.5	91.0
$BJ_3$ (28)	91.0	91.5	91.5	91.5
$BJ_4$ (29)	90.0	90.5	90.0	91.0

differs.

The input equations for applied experiments are given below (26-29):

$$BJ_1 : y_1(t + 1) = f(v(t - 3), y(t)) \quad (26)$$

$$BJ_2 : y_2(t + 1) = f(v(t), v(t - 1), v(t - 2), v(t - 3), y(t)) \quad (27)$$

$$BJ_3 : y_3(t + 1) = f(v(t), v(t - 1), v(t - 2), v(t - 3), v_{[0,9]}, v_{[10,19]}, v_{[20,29]}, v_{[30,39]}, y(t)) \quad (28)$$

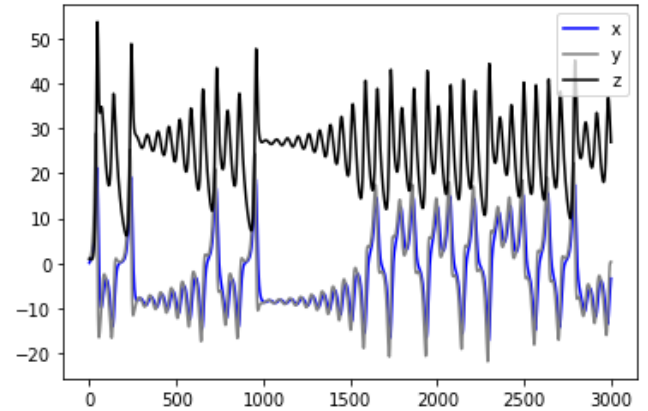
$$BJ_4 : y_4(t + 1) = f(v(t), v(t - 1), v(t - 2), v(t - 3), v_{[0,4]}, v_{[5,9]}, v_{[10,19]}, v_{[20,29]}, y(t)) \quad (29)$$

The results indicate that the proposed PEC-WNN shows the lowest RMSE (Tab. 6) when the data set with four successive and four average values with window size ten is used. Compared to the MAPE (Tab. 7) and DA results (Tab. 8), the PEC-WNN shows the lowest MAPE and the highest DA results. The applied average values of a different time window to the consecutive values improves the results.

In the literature, the RMSE results for the gas furnace (series J), or the Box-Jenkins problem changes from 0.273 to 0.843, while the lowest RMSE using the PEC-WNN is 0.059. The PEC-WNN achieves 78% less RMSE compared to the second-best results found in the literature.

**TABLE 9.** The RMSE error comparison for Box-Jenkins time series data found in the literature for prediction time  $t = \{1\}$

Models:	y+1	Source:
Arima	0.843	[ [29]]
Tong's model	0.685	[ [30]]
Xu's model	0.573	[ [31]]
Sugeno's model	0.596	[ [32]]
Surmans' model	0.4	[ [33]]
Adaptive network-based fuzzy inference system (ANFIS)	0.405	[ [34]]
Generalized fuzzy NN	0.273	[ [34]]
PEC-WNN	0.0589	
LSTM	0.9730	



**FIGURE 6.** The Lorenz Attractor trajectories projection of each time series for the first 3000 values.

**TABLE 10.** The RMSE results for Lorenz Attractor data considering applied models and equations.

Equation	NN	PEC-NN	WNN	PEC-WNN
$LA_1$ (33)	0.0015	0.0014		
$LA_2$ (34)	0.0359	0.0340	0.0356	0.0308
$LA_3$ (35)	0.0609	0.0476	0.0378	0.0113

given below in (33)-(35). The output represents the next  $(t + 1)$  value of a single variable.

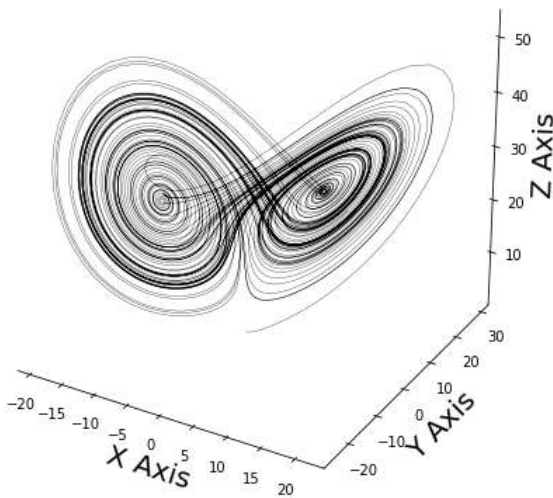
$$LA_1 : x(t + 1) = f(x(t), y(t), z(t)) \quad (33)$$

$$LA_2 : x(t + 1) = f(x(t), y(t), y(t - 1), y(t - 2), y(t - 3), z(t), z(t - 1), z(t - 2), z(t - 3)) \quad (34)$$

$$LA_3 : x(t + 1) = f(x(t), y(t), y(t - 1), y(t - 2), y(t - 3), z(t), z(t - 1), z(t - 2), z(t - 3), \bar{y}_{[0,9]}, \bar{y}_{[10,19]}, \bar{y}_{[20,29]}, \bar{y}_{[30,39]}, \bar{z}_{[0,9]}, \bar{z}_{[10,19]}, \bar{z}_{[20,29]}, \bar{z}_{[30,39]}) \quad (35)$$

The constructed data sets are divided into training and test sets, where 80% of the data is used for training and 20% for testing performances. The significantly low RMSE (Tab. 10) is achieved using only the one previous value of each trajectory for forecasting the next value. Similar to previous experiments, herein we try to observe the impact of adding average values of different time-window to the consecutive values. The growth in the number of variables, from a single to multivariate, increases the RMSE. On the other hand, the usage of multivariate average values together with consecutive multivariate values reduces the RMSE error in comparison to the successive multivariate input. The lowest value of MAPE (Tab. 11) is found in the PEC-WNN model, with the utilization of consecutive and average values together. In the end, very high DA (Tab. 12) results are noticed for the applied input data set.

The best result for the Lorenz Attractor data set using the multivariate input data is 0.0013 for the PEC-WNN model. The PEC-WNN result obtained 64% less RMSE compared to a similar experiment found in the literature, Xiu et. al. [36]. The Lorenz multivariate time series data with the natural structure used as the input overperform the predicting results where the single variable sequence is used.



**FIGURE 5.** The Lorenz Attractor.

### C. THE LORENZ ATTRACTOR

The Lorenz Attractor represents a classical time series multivariate prediction problem consisting of three ordinary differential equations given in (30)-(32).

$$\frac{dx}{dt} = \sigma(y - x) \quad (30)$$

$$\frac{dy}{dt} = x(\rho - z) - y \quad (31)$$

$$\frac{dz}{dt} = xy - \beta z \quad (32)$$

The equations are obtained from the Navier-Stokes equations and used in fluid mechanics. The parameter settings to exhibit the chaotic behavior are  $\sigma = 10$ ,  $\beta = 8/3$  and  $\rho = 28$ , with initial conditions  $[x(0), y(0), z(0)] = [0, 1, 1.05]$  as studied in Lorenz [35]. Different Lorenz maps with the same general dynamics can be obtained by using distinctive initial conditions and parameter values. The Lorenz map is given in the Fig. 5.

The data set contains 10,000 multivariate data samples. From the plots of each trajectory interdependencies between the time series can be seen (Fig. 6). Xiu et al. [36] applied a multivariate data set as inputs to predict the single variable as the output. We applied similar single and multivariate inputs to our model. The input equations are

**TABLE 11.** The MAPE(%) results for Lorenz Attractor data considering applied models and equations.

Equation	NN	PEC-NN	WNN	PEC-WNN
$LA_1$ (33)	0.024	0.014		
$LA_2$ (34)	1.179	0.526	1.5570	0.4287
$LA_3$ (35)	2.909	1.500	1.6632	0.3002

**TABLE 12.** The DA(%) results for Lorenz Attractor data considering applied models and equations.

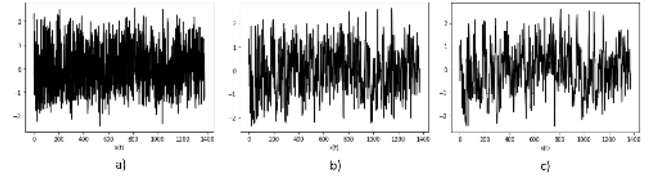
Equation	NN	PEC-NN	WNN	PEC-WNN
$LA_1$ (33)	99.975	99.988		
$LA_2$ (34)	99.875	99.900	99.9250	99.9750
$LA_3$ (35)	99.738	99.213	99.6250	99.7650

**TABLE 13.** The RMSE error comparison for Lorenz Attractor found in the literature for prediction time  $t = \{1\}$

Models:	$x+1$	Source:
Neural Network	0.072	[ [37]]
NARX(multivariate)	0.000295	[ [36]]
NARX(single variable)	0.068	[ [36]]
MKELM(multivariate)	0.0312	[ [36]]
MKELM(single variable)	0.0973	[ [36]]
PEC-WNN	0.0113	
LSTM	0.3148	

#### D. THE DROUGHT FORECASTING

In this section, we also demonstrate the performance of the proposed model when it is applied to the stochastic time series data. The benchmark problems previously explained, the Mackey-Glass, the Box-Jenkins gas furnace (series J), and the Lorenz Attractor represent chaotic problems with deterministic models. Their output can be determined based on representing mathematical models when the initial conditions and the model parameters are known. The proposed PEC-WNN model for performance comparison is also applied to the stochastic problem. The standardized precipitation-evapotranspiration index (SPEI) drought index developed by Begueria et. al [38] is selected for that purpose. Drought identification and forecasting are very important in limiting their effects. However, accurate drought prediction remains a scientific issue due to the nature of data. The SPEI represents an index that quantifies the drought condition over a given area. The index can be calculated in several time scales to adopt the characteristic drought response time of the target natural and economic systems, by determining their drought resistance [38]. The data set evaluates accumulated precipitation minus potential evapotranspiration (PET) over multiple time scales between 1 and 48 months, with global coverage at a 0.5-degree resolution [39]. The advantages of a used data set are that (a) it improves the spatial resolution of the unique global drought data set at a global scale; (b) it is spatially and temporally comparable to other data sets, given the probabilistic nature of the SPEI; and, (c) it enables the identification of various drought types, given the multiscalar character of the SPEI [39]. The analyzed period is from January 1901 until December 2015. The 1-month, 4-months, and 6-months data were used (Fig. 7). The input-output functions for the prediction model are given in the (36) and (37). The input data sets consist of eight inputs with different window sizes. The first data set contains one-month and four-month data (36); the second data set one-month and six-month data (37). In both cases, we tried to forecast the next six months' drought period. In equations,  $t$  represents the SPEI values of one-month data,  $4t$  the SPEI values of four-month data, and  $6t$  the SPEI values of six-month data. The used PEC-WNN model contains the same hyperparameters and the



**FIGURE 7.** The Standardized Precipitation Evapotranspiration Index (SPEI) for the period from January 1901 until December 2015 within a) month, b) four months and c) six months.

**TABLE 14.** The RMSE results for applied models and equations for SPEI data set.

Forecasting period	Equation	NN	PEC-NN	WNN	PEC-WNN	LSTM	LR
6-months SPEI forecasting	$SPEI_1$ (36)	0.157	0.151	0.150	0.139	0.353	0.409
	$SPEI_2$ (37)	0.216	0.147	0.221	0.146	0.386	0.485

number of inputs that are applied to the previous chaotic time series problems.

$$SPEI_1 : y_1(6t + 1) = f(x(t), x(t - 1), x(t - 2), x(t - 3), x(4t), x(4t - 1), x(4t - 2), x(4t - 3)) \quad (36)$$

$$SPEI_2 : y_2(6t + 1) = f(x(t), x(t - 1), x(t - 2), x(t - 3), x(6t), x(6t - 1), x(6t - 2), x(6t - 3)) \quad (37)$$

As an alternative to the previously mentioned models, for drought forecasting problems we additionally used the LSTM model proposed in [13] and multivariate linear regression (LR) for performance comparison. The proposed PEC-WNN accomplished significantly low RMSE (Tab. 14) with monthly and four-month data used as inputs given in (36). The PEC-WNN model also provides the lowest MAPE as seen in table 15.

The results of forecasting the SPEI index have shown reasonable prediction accuracy for a six-month time scale considering the uncertainty level of stochasticity. The accuracy of the proposed method with increasing the scale of SPEI input data from four-months to six-months average data does not show better performances. Evaluated SPEI prediction at different time scale simultaneously used increases the performance of the proposed method.

#### IV. DISCUSSION

The time series prediction model where a separate NN model is used for predictive error correction of the main NN, PEC-WNN has been applied to different kinds of deterministic, chaotic, and stochastic time series problems. The introduced method, PEC-WNN, has been compared to twenty time series prediction methods found in the literature, to demonstrate the prediction performance where the PEC-WNN model demonstrates the lowest RMSE error. The predictive error compensation model overall reduces the RMSE but when applied together with wavelet transform as a preprocessing mechanism surpasses the other methods applied and found in the literature. The PEC-WNN has been applied to different problems without changing the network structure and hyperparameters. The

**TABLE 15.** The MAPE (%) results for applied models and equations for SPEI data set.

Forecasting period	Equation	NN	PEC-NN	WNN	PEC-WNN	LSTM	LR
6-months SPEI forecasting	$SPEI_1$ (36)	0.301	0.287	0.296	0.278	0.356	0.429
	$SPEI_2$ (37)	0.351	0.289	0.348	0.286	0.378	0.493



**TABLE 16.** The DA (%) results for applied models and equations for SPEI data set.

Forecasting period	Equation	NN	PEC-NN	WNN	PEC-WNN	LSTM	LR
6-months SPEI forecasting	$SPEI_1$	61.26	63.18	64.21	64.95	33.94	45.33
	$SPEI_2$	60.0	57.46	55.86	56.61	32.17	42.76

**TABLE 17.** The SPEI RMSE error comparison for PEC-WNN, LSTM and LR.

Input data set	6-months
PEC-WNN	0.139
LSTM	0.353
LR	0.409

PEC-WNN, although in its structure, uses two NNs is less computationally expensive and time-consuming with respect to other ML methods found in the literature. The PEC-WNN complexity concerning the number of parameters is relatively low (Tab. 18). The results are consistent with one of the conclusions found in [40], which states that simple models tend to outperform complex models. The proper arrangement of input data sets can significantly improve the forecasting performance of the proposed model. The results show that different sizes of input data frames used together with consecutive values improve the forecasting performances.

## V. CONCLUSION

In this work, a predictive error compensated wavelet preprocessed NN model for time series prediction problems is proposed. The model is consisting of at least two separate NNs, where the input data are preprocessed using DWT in both of them. It has been demonstrated that the second predictive error compensating network significantly improved the overall accuracy of the proposed model at all benchmark problems. The Mackey-Glass, Box-Jenkins, and Lorenz Attractor problems are used to evaluate the prediction performance for chaotic time series case and global drought forecasting problem for a stochastic case. The results show that the PEC-WNN model provides 64% less RMSE for the Lorenz Attractor, 78% less RMSE for the Box-Jenkins, and 95% less RMSE for the Mackey-Glass benchmark problems. The proposed method achieves reasonable results also in forecasting the global drought SPEI index. An additional advantage of the proposed model is less sensitivity to its hyperparameters and structural settings for a broad range of time series prediction problems. The same network structure of PEC-WNN has been used in all given benchmark evaluations. Both the time and space complexity of the proposed model was less than the compared other machine learning methods in all cases. Though the proposed PEC-WNN method demonstrated promising results, more improvements can also be achieved through fusion with additional cascaded predictive error compensating networks for multidimensional data sets.

## ACKNOWLEDGMENT

This study is based on the research project "Development of Cyberdroid based on Cognitive Intelligent system applications" (2019-2020) funded by Crypttech company

**TABLE 18.** The number of parameters of compared models.

Models	Number of parameters
NN	51
PEC-NN	112
WNN	51
PEC-WNN	112
LSTM	1036

(<https://www.crypttech.com/en/>) within the contract by ITUNOVA-Istanbul Technical University Technology Transfer Office.

## REFERENCES

- [1] Thanapant Raicharoen, Chidchanok Lursinsap, and Paron Sanguanbhokai. Application of critical support vector machine to time series prediction. In Proceedings of the 2003 International Symposium on Circuits and Systems, 2003. ISCAS'03., volume 5, pages V–V. IEEE, 2003.
- [2] Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, and Kwok Wei Shah. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74:902–924, 2017.
- [3] G Peter Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003.
- [4] Antti Sorjamaa, Jin Hao, Nima Reyhani, Yongnan Ji, and Amaury Lendasse. Methodology for long-term prediction of time series. *Neurocomputing*, 70(16-18):2861–2869, 2007.
- [5] Waddah Waheeb, Rozaida Ghazali, and Tutut Herawan. Ridge polynomial neural network with error feedback for time series forecasting. *PloS one*, 11(12), 2016.
- [6] Qian-Li Ma, Qi-Lun Zheng, Hong Peng, Tan-Wei Zhong, and Li-Qiang Xu. Chaotic time series prediction based on evolving recurrent neural networks. In 2007 international conference on machine learning and cybernetics, volume 6, pages 3496–3500. IEEE, 2007.
- [7] Simon Haykin. *Neural Networks and Learning Machines*, 3/E. Pearson Education India, 2010.
- [8] Yuehui Chen, Bo Yang, Jiwen Dong, and Ajith Abraham. Time-series forecasting using flexible neural tree model. *Information sciences*, 174(3-4):219–235, 2005.
- [9] Waddah Waheeb and Rozaida Ghazali. Forecasting the behavior of gas furnace multivariate time series using ridge polynomial based neural network models. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(5):126–133, 2019.
- [10] C H López-Caraballo, I Salfate, J A Lazzús, P Rojas, M Rivera, and L Palma-Chilla. Mackey-glass noisy chaotic time series prediction by a swarm-optimized neural network. *Journal of Physics: Conference Series*, 720:012002, may 2016.
- [11] Shanoli Samui Pal and Samarjit Kar. A hybridized forecasting method based on weight adjustment of neural network using generalized type-2 fuzzy set. *International Journal of Fuzzy Systems*, 21(1):308–320, 2019.
- [12] Vincenzo Loia, Stefania Tomasiello, Alfredo Vaccaro, and Jinwu Gao. Using local learning with fuzzy transform: application to short term forecasting problems. *Fuzzy Optimization and Decision Making*, 19(1):13–32, 2020.
- [13] Denisa Roberts. Lorenz trajectories prediction: Travel through time. arXiv preprint arXiv:1903.07768, 2019.
- [14] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. arXiv preprint arXiv:1609.03499, 2016.
- [15] Marija Gorenc Novak and Dejan Velušček. Prediction of stock price movement based on daily high prices. *Quantitative Finance*, 16(5):793–826, 2016.
- [16] Yann Lecun and Yoshua Bengio. *Convolutional networks for images, speech, and time-series*. MIT Press, 1995.
- [17] Xindi Cai, Nian Zhang, Ganesh K Venayagamoorthy, and Donald C Wunsch II. Time series prediction with recurrent neural networks trained by a hybrid pso-ea algorithm. *Neurocomputing*, 70(13-15):2342–2353, 2007.
- [18] Gianluca Bontempi, Souhaib Ben Taieb, and Yann-Aël Le Borgne. Machine learning strategies for time series forecasting. In European business intelligence summer school, pages 62–77. Springer, 2012.
- [19] AJ Rocha Reis and AP Alves Da Silva. Feature extraction via multiresolution analysis for short-term load forecasting. *IEEE Transactions on power systems*, 20(1):189–198, 2005.
- [20] Husam Y Alzaq and Berk Ustundag. An optimized two-level discrete wavelet implementation using residue number system. *EURASIP Journal on Advances in Signal Processing*, 2018(1):41, 2018.
- [21] Wei Bao, Jun Yue, and Yulei Rao. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PloS one*, 12(7):e0180944, 2017.
- [22] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

- [23] K Meng, ZY Dong, and KP Wong. Self-adaptive radial basis function neural network for short-term electricity price forecasting. *IET generation, transmission & distribution*, 3(4):325–335, 2009.
- [24] Yuehui Chen, Bo Yang, and Jiwen Dong. Time-series prediction using a local linear wavelet neural network. *Neurocomputing*, 69(4-6):449–465, 2006.
- [25] Felix A Gers, Douglas Eck, and Jürgen Schmidhuber. Applying lstm to time series predictable through time-window approaches. In *Neural Nets WIRN Vietri-01*, pages 193–200. Springer, 2002.
- [26] Lucius Chudy and Igor Farkas. Prediction of chaotic time-series using dynamic cell structures and local linear models. *Neural Network World*, 8(5):481–489, 1998.
- [27] Shawn P Day and Michael R Davenport. Continuous-time temporal back-propagation with adaptable time delays. *IEEE Transactions on neural networks*, 4(2):348–354, 1993.
- [28] R Scott Crowder III. Predicting the mackey-glass timeseries with cascade-correlation learning. In *Connectionist Models*, pages 117–123. Elsevier, 1991.
- [29] George EP Box and Gwilym M Jenkins. *Time series analysis: Forecasting and control san francisco*. Calif: Holden-Day, 1976.
- [30] RM Tong. The evaluation of fuzzy models derived from experimental data. *Fuzzy sets and systems*, 4(1):1–12, 1980.
- [31] Chen-Wei Xu and Yong-Zai Lu. Fuzzy model identification and self-learning for dynamic systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 17(4):683–689, 1987.
- [32] Michio Sugeno. Linguistic modelling based on numerical data. *Proceedings of the IFSA'91*, 1991.
- [33] Hartmut Surmann, Andreas Kanstein, and Karl Goser. Self-organizing and genetic algorithms for an automatic design of fuzzy control and decision systems. In *In Proc. EUFIT'93*, 1993.
- [34] Yang Gao and Meng Joo Er. Narmax time series model prediction: feedforward and recurrent fuzzy neural network approaches. *Fuzzy sets and systems*, 150(2):331–350, 2005.
- [35] Edward N Lorenz. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141, 1963.
- [36] Yan Xiu and Wei Zhang. Multivariate chaotic time series prediction based on narx neural networks. In *2017 2nd International Conference on Electrical, Automation and Mechanical Engineering (EAME 2017)*. Atlantis Press, 2017.
- [37] J. B. Elsner and A. A. Tsonis. Nonlinear prediction, chaos, and noise. *Bulletin of the American Meteorological Society*, 73(1):49–60, 1992.
- [38] Santiago Beguería, Sergio M. Vicente-Serrano, and Marta Angulo-Martínez. A Multiscalar Global Drought Dataset: The SPEIbase: A New Gridded Product for the Analysis of Drought Variability and Impacts. *Bulletin of the American Meteorological Society*, 91(10):1351–1356, 10 2010.
- [39] S. M. Vicente-Serrano, S. Beguería, J. I. López-Moreno, M. Angulo, and A. El Kenawy. A New Global 0.5° Gridded Dataset (1901–2006) of a Multiscalar Drought Index: Comparison with Current Drought Index Datasets Based on the Palmer Drought Severity Index. *Journal of Hydrometeorology*, 11(4):1033–1043, 08 2010.
- [40] Nesreen K Ahmed, Amir F Atiya, Neamat El Gayar, and Hisham El-Shishiny. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5-6):594–621, 2010.



**BERK USTUNDAG** is a Professor of Computer Engineering Department at Faculty of Computer and Informatics, Istanbul Technical University (ITU). Dr.Ustundag has received Ph.D. and M.Sc. in control systems and computer engineering at ITU and B.Sc. in electrical engineering. He has served as Science and Technology Advisor to Ministers and Governmental Institutions for more than 15 years. He has more than 100 scientific publications and patents. Dr.Ustundag is a member of the IEEE Communication Society and Computational Intelligence. His research interests involve Data Fusion, Artificial Intelligence, Signal Processing, Global Optimization, Cognitive Communication, and Agricultural Information Systems.



**AJLA KULAGLIC** She received a B.Sc. degree and M. Sc. Degree in Computing and Informatics, both from the University of Sarajevo (Faculty of Electrical Engineering), Sarajevo, Bosnia, and Herzegovina, in 2008 and 2010 respectively. She is currently a Ph.D. candidate and research assistant at the Computer Engineering Department, Graduate School of Science Engineering and Technology, Istanbul Technical University, Istanbul, Turkey, under the supervision of Prof. Dr. Burak

Berk Ustundag. She has published papers at international conferences and journals interested in data fusion. The published works are mainly about data fusion models for time series, remotely sensed, and ground-based measurement data of agricultural sensor networks using preprocessing techniques, such as Wavelet transformation together with Neural Networks for estimation performances. She is interested in doing further research in the development of deep wavelet networks for estimation performances. Her current primary fields of investigation are forecast of time series data, natural events, data fusion, and machine learning techniques.

...