of that, reducing the feature set to the essential ones has various advantages: it reduces the computational cost of the learning process and improves the accuracy without causing overfitting. In general, it is performed either by feature extraction or by feature selection, or combinations thereof [6]. The feature extraction approach transforms the initial feature space into a lower-dimensional space. On the other hand, feature selection reduces the dimensionality problem by selecting the best possible subset of the complete input feature set.

Currently, the feature selection studies can be classified into the following directions: (a) embedded method; (b) filter-based method; (c) wrapper-based method [7]. The embedded method automatically performs feature selection, having lower computational complexity but algorithm performance depends on the choice of the predictor. The filter-based method performs feature selection based solely on the characteristic of the training set causing poor interaction with the target values and ignoring the correlation between features. The wrapper-based method is based on optimization and a forecasting model in the process of feature selection with high algorithm complexity and low computational efficiency. In contrast, in the proposed model, PEC-WNN chooses the highest correlating data to the residual error acquired by the previously trained network. In the first network, we are using highly correlated data to the target data. Input data of each network is normalized, the discrete wavelet transform is applied and wavelet coefficients are used as neural network input. The main goal of data preparation is to guarantee the quality of the data before it is used in ML algorithms [8]. The most widely used normalization methods include the min-max, the z-score, and the decimal scaling. However, these methods are unable to efficiently handle non-stationary time series data. The applicability of widely used normalization methods depends on knowing their min/max values, standard deviation, and mean. Several recent works attempt to address this issue by using carefully handcrafted stationary features or by employing more sophisticated normalization schemes. Even though proposed approaches overcome the problems of normalization methods and lead to better performance of used models, they are facing significant drawbacks since they are based on heuristically designed normalization.

In this study to overcome these limitations, we propose an average subtraction normalization that is capable to normalize the input data considering the average value of the current input. The mean value of the current input is calculated and subtracted from each distinct input, allowing the effective handling of non-stationary data. The proposed normalization is applied to each network used in this model. After the normalization, we applied discrete wavelet transform that decomposes input data into different scales, making it useful in distinguishing seasonality, revealing structural breaks and volatility clusters, and identifying local and global dynamic properties of a process at a specific timescale [9]. Further, after the training of the main network is done, we obtain the error pattern between measured and predicted values. The correlation between error patterns and supplementary data is computed. The highest correlating data to the residual error is selected as supplementary data in the second network. Generally, one or two additional feature components are sufficient [10], however, the proposed PEC-WNN model is able to select as many features as supplementary data correlates to the residual error.

In the literature, many statistical methods based on autoregression and moving average (MA) are considered state-of-the-art for time series modeling [11]. The main drawback of this method is the assumption that the data have a known distribution. More recently, several studies have demonstrated that the ML algorithms for time series prediction provide outperforming results compared to the statistical models. The artificial neural networks (ANN) are used to resolve non-linear functional dependencies between time series data in the past and its future [12]. The most used ANN model in time series predictions is the multilayer perceptron (MLP) [13]. The structure of the MLP requires a large number of parameters to solve complex non-linear problems. This results in a low learning rate and poor generalization [13]. Time series data prediction to achieve better accuracy demands the NN models to be adaptive to changes that occur over time in the data [1]. Despite the efficiency of NN in the prediction of time series, two main problems can be pointed out: network architecture and hyperparameters design, and sensitivity to estimation errors. The PEC-WNN consists of at least three fully connected and separately trained neural networks (NNs). The NN models are trained independently and therefore not prone to accumulate errors. The statistically highest correlating data are used as primary inputs. The second data in the cascaded network is selected by seeking the highest correlation to the error pattern obtained in the primary network. The second NN is trained with the past prediction errors of the primary NN, while the first NN is directly trained with the target prediction data. The third NN is trained with the past prediction errors of the second NN. Multiple NNs can be used in a similar manner through cascaded NNs that each compensates the remaining error. Cascaded networks in PEC-WNNs are ended by a NN that only compensates residual error only by using its past values. An additional, final NN merges outputs of three NNs in the cascaded part of the network. The results show the effectiveness of the proposed algorithm with respect to the competitor methods both in terms of prediction accuracy and model complexity. Additionally, without changing the network structure and hyperparameter set the proposed algorithm has the competence to find precise solutions.

In this study, we provide a description of PEC-WNN with the proposed normalization technique, present new insights on the orthogonal feature extractions using pattern similarity and correlation analysis to the residual errors. The main idea is to find the orthogonal features to the residual error in the sequence of networks and construct a network with the most correlated features to the residual error without causing overfitting or reducing the accuracy of the model.

Overall, the main contributions of this study can be summarized as:

- Improvement of the prediction accuracy using multiple neural networks without causing overfitting
- The same neural network models are applied for two different time series prediction problems
- The average subtraction normalization overcome the traditional normalization problems for non-stationary time series data
- The wavelet transformation used as a feature extraction method yields better accuracy in the proposed model
- The highly correlated mixed-frequency data are trained with the target prediction data, while additional networks use the highly correlated data to the residual errors
- PEC-WNN enable choosing the orthogonal features in data fusion applications, which appears as a promising machine learning model for spatial and temporal predictions

The rest of this paper is assembled as follows. In the next section, the proposed PEC-WNN model for time series prediction and orthogonal feature selection is explained. The time series problems, the Lorenz Attractor time series data, and wind forecasting problem are provided in section 3 along with the corresponding results and discussion. Section 4 presents the concluding remarks.

2. Predictive Error Compensating Neural Networks

The Predictive Error Compensating Wavelet Neural Network (PEC-WNN) utilized in this study comprises four separately trained neural networks, as demonstrated in Figure 1. The number of cascaded networks can be increased concerning the number of input parameters and data samples. In the first neural network, the highest correlated data are used together. The current input of each parameter is shifted to the previous values using the unit time delay operator z^{-1}. The estimated output has its own error pattern from the main network. Seeking the correlation between residual error and remaining parameters in sequences of networks gives us the network with the most orthogonal features. Each subsequent neural network uses shifted past values of residual error from the previous network. For example, the second neural network is trained with the past prediction errors of the main neural network, while the main neural network is directly trained with the target prediction data. We use additional and separately trained NNs with error data patterns of previously trained networks without applying the errors back to the same network and increasing the number of inputs or changing the network configurations. The final neural network merges outputs of neural networks in the cascaded part.

Due to the problems in the traditional normalization approaches to handle non-stationary time series data (for more details check [8]), we proposed the normalization method that uses the average values of the current input to the neural network (Figure 2.). The idea is to calculate the mean of the network inputs and subtract the obtained value from particular input data. By this, we are allowed to build a normalization method capable to represent different volatilities and preserving the original time series properties inside each input sequence.

The original time series data applied to each network are normalized and preprocessed by discrete wavelet transform (DWT). The reason for selecting the DWT is considering its ability to analyze a signal both in time and frequency domains. The wavelet transform automatically adapts itself to a suitable resolution and overcomes the limitations of the Fourier transform [9]. The input signal is presented with x[n], the low- and high-pass filters are represented by h[n] and g[n] respectively. In this way, we obtain a sequence of coefficients that characterizes and compact the original signal information. Decomposed signal y[n] consists of high and low-frequency components as shown in equation 1.

$$y[n] = y_{high}[n-1] + y_{low}[n-1] \qquad (1)$$

Low-pass outputs are recursively passed through identical filter banks in order to use different resolutions at every stage. The filtering process is expressed mathematically using equations (2) and (3). Equation (2) provides an approximation, and equation (3) provides the detailed signal.

$$y_{high}[n-1] = \sum_n g[k]\,.\,x[2n-k] \quad (2)$$

$$y_{low}[n-1] = \sum_n h[k]\,.\,x[2n-k]$$
(3)

We use Haar wavelet filters, since they beneficially diminish the rate of distortion during the signal decomposition and reconstruction, significantly reducing the processing and computational time [1].

The configuration of neural networks consists of three layers: input, hidden, and output layer for predicting the n-step-ahead time series data. Employed networks have the same network configurations. Regarding the formulas found in the literature [14], [15], [16] the number of neurons in the hidden layers are selected based on the trial and error method. The activation function used for these networks is Rectified Linear Unit (ReLU). In comparison with sigmoid and hyperbolic tangent activation functions, linear activation function ReLU (given in equation 4), notably improves the achievement of the feed-forward networks [14].

$$f(x) = x^{+} = max(0, x) \quad (4)$$

The learning rate and momentum of stochastic gradient descent (SGD) optimization algorithm are 0.05 and 0.75, respectively. In order to do the feature selection by checking the correlation of supplementary data and residual errors, we added an additional network that contains the data highly correlated to the error pattern. Otherwise, if we increase the number of input parameters in the same NN with the correlation between different inputs may rise the overfitting rate. For that purpose, we use correlation analysis as the main statistical tool to find the relationship between the residual error and remaining parameters in the additional networks. The main idea is to measure the correlation with respect to the supplementary input data to the residual error rate obtained in the previous NN. Formally, the Pearson correlation coefficient between two corresponding sets of measures $err_{res} = (err_{(res(i))})_{i=1}^{N}$ and $Y = (Y_i)_{i=1}^{N}$ is defined as:

$$r_{err_{(res)}Y} = \frac{\sum_{i=1}^{N}\left(err_{res(i)} - \overline{err_{res(\imath)}}\right)\cdot(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N}\left(err_{res(i)} - \overline{err_{res(\imath)}}\right)^2}\sqrt{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}} \quad (5)$$

Where N is number of samples, $err_{res(i)}$ and $\bar{Y}$ are the means of residual errors ($err_{res(i)}$) and additional data (Y) respectively.

Uncovering hidden patterns associated with the target variables is not trivial [10]. One of the major issues in addressing this problem is how to deal with the existence of a high correlation between supplementary and target data, which can cause a problem in overfitting. However, in this study, hidden patterns associated with the residual errors obtained in the previously trained networks and supplementary data are analyzed.

3. Experimental Setup and Results

We performed an experimental setup using two different data sets to test the proposed improvements of the PEC-WNN model and usage of supplementary data. The Lorenz Attractor, three-dimensional chaotic time series data, and meteorological data for wind speed prediction are used for this purpose.

Along with the proposed model PEC-WNN, we made a comparison between multivariate regression (MLR), neural network model (NN), wavelet neural network model (WNN), and long short-term memory (LSTM) model, and provide comparative results. Multivariable regression represents an extension of multiple regression with one dependent variable and multiple independent variables. We try to predict the output based on the number of independent variables. Long short-term memory (LSTM) represents an artificial recurrent neural network architecture used in the field of deep learning. The model presented in [17] is used to compare the performances of the PEC-WNN model. The implemented LSTM model has one hidden layer with 25 hidden units and dense output.

The root-mean-square error (RMSE) and root-mean-square percentage error (RMSPE) are used for the comparison of the experimental results. The mathematical formulations are given below, (Eq. 6 and Eq. 7) where X_{real} is real and $X_{estimated}$is estimated values in time *i*. The number of data samples is given by *n*.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{real,i} - X_{estimated,i}\right)^2}{n}} \quad (6)$$

$$RMSPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{X_{real,i} - X_{estimated,i}}{X_{real,i}}\right| \quad (7)$$

a. Lorenz Attractor

The Lorenz Attractor represents a multivariable time series prediction problem consisting of three differential equations [1] (equations are given below (8)-(10)). Close initial conditions lead to very different trajectories, making the Lorenz system a chaotic dynamical system [18]. The characteristics of the Lorenz Attractor are presented in Figure 3.

$$\frac{dx}{dt} = \sigma \cdot (y - x) \quad (8)$$

$$\frac{dy}{dt} = x \cdot (\rho - z) - \mathrm{y} \quad (9)$$

$$\frac{dz}{dt} = x \cdot y - \beta \cdot z \quad (10)$$

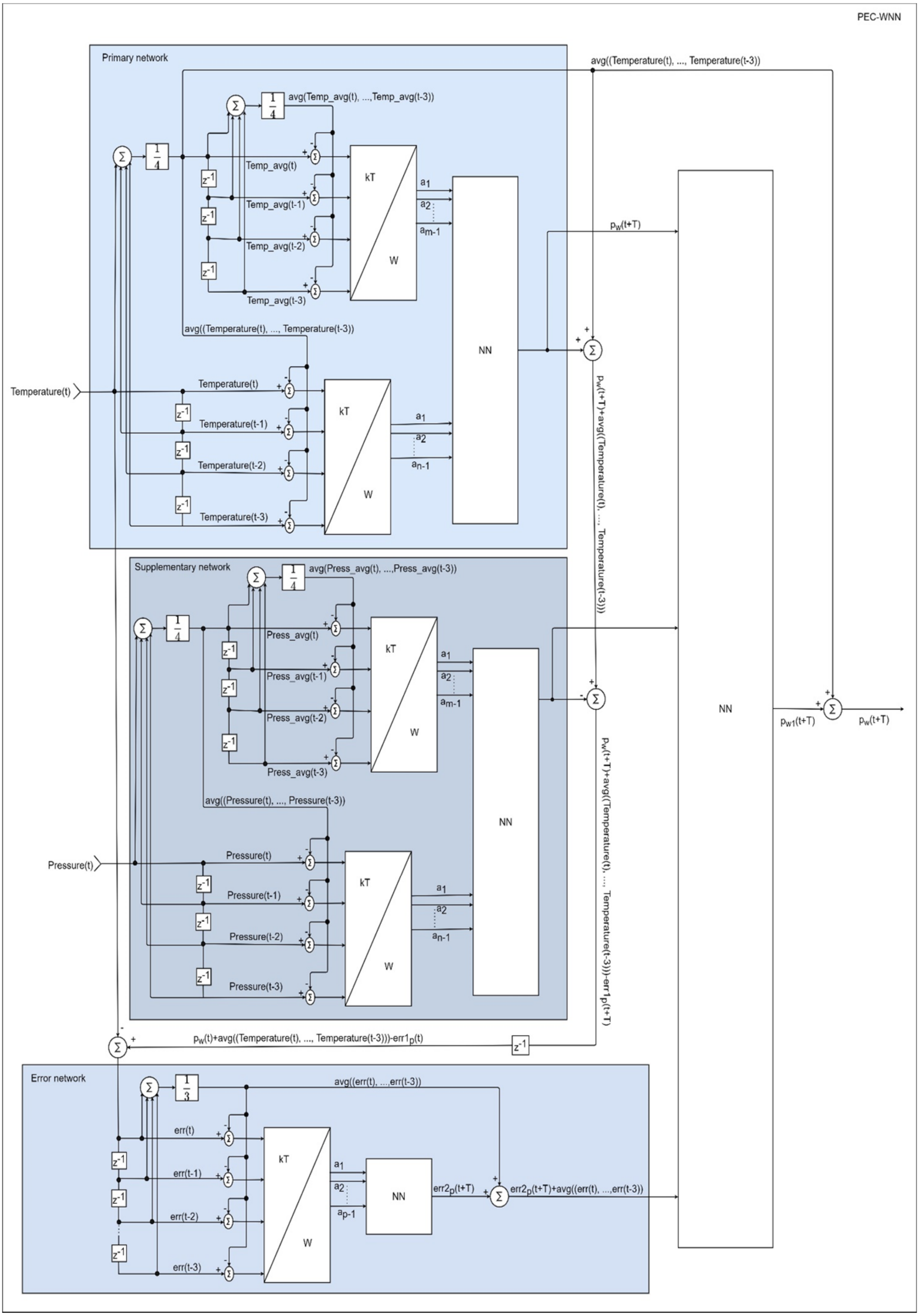

PEC-WNN
Primary network
avg((Temperature(t), ..., Temperature(t-3))
avg(Temp_avg(t), ...,Temp_avg(t-3))
Temp_avg(t)
Temp_avg(t-1)
Temp_avg(t-2)
Temp_avg(t-3)
kT
W
a_1
a_2
a_{m-1}
NN
$p_w(t+T)$
avg((Temperature(t), ..., Temperature(t-3))
Temperature(t)
Temperature(t-1)
Temperature(t-2)
Temperature(t-3)
a_{n-1}
$p_w(t+T)$+avg((Temperature(t), ..., Temperature(t-3)))
Supplementary network
avg(Press_avg(t), ...,Press_avg(t-3))
Press_avg(t)
Press_avg(t-1)
Press_avg(t-2)
Press_avg(t-3)
avg((Pressure(t), ..., Pressure(t-3))
Pressure(t)
Pressure(t-1)
Pressure(t-2)
Pressure(t-3)
$p_w(t+T)$+avg((Temperature(t), ..., Temperature(t-3)))-$err1_p(t+T)$
$p_{w1}(t+T)$
$p_w(t+T)$
$p_w(t)$+avg((Temperature(t), ..., Temperature(t-3)))-$err1_p(t)$
z^{-1}
Error network
avg((err(t), ...,err(t-3))
err(t)
err(t-1)
err(t-2)
err(t-3)
a_{p-1}
$err2_p(t+T)$
$err2_p(t+T)$+avg((err(t), ...,err(t-3))
$\frac{1}{4}$
$\frac{1}{3}$
Σ

Figure 1. Predictive error compensating neural network model for multivariable time series prediction. Supplementary data is selected by computing the correlation between residual error and remaining variables

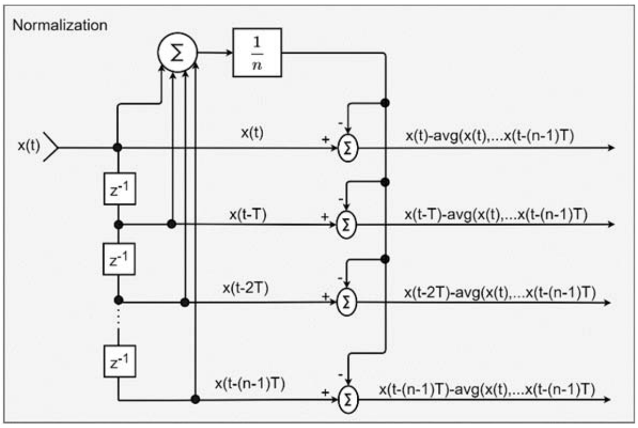

Normalization
Σ
1/n
x(t)
z⁻¹
z⁻¹
z⁻¹
x(t)
x(t-T)
x(t-2T)
x(t-(n-1)T)
x(t)-avg(x(t),...x(t-(n-1)T)
x(t-T)-avg(x(t),...x(t-(n-1)T)
x(t-2T)-avg(x(t),...x(t-(n-1)T)
x(t-(n-1)T)-avg(x(t),...x(t-(n-1)T)

Figure 2. The normalization of input values computing the average value of used inputs and subtracting it from exact input

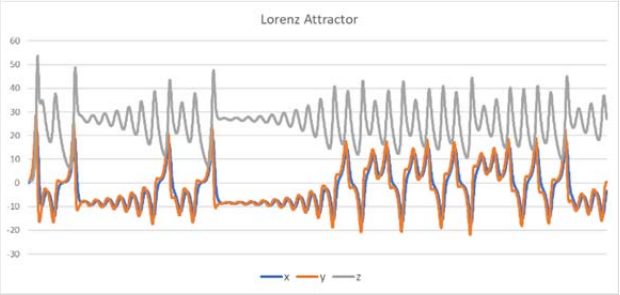


Figure 3. First 3000 values of each Lorenz Attractor dimension

The equations are derived from the Navier-Stokes equations [1]. The parameter settings are σ=10, β=8/3 and ρ=28, with initial conditions [x(0), y(0), z(0)]=[0,1,1.05] as in [18]. The dataset contains 10 000 multivariable data samples. We applied a multivariable dataset as input to predict the single variable as the output, similarly as in our previous study [1]. The main difference, comparing the previous work is applying multivariable input to the separately trained neural networks considering the orthogonality between supplementary data and residual errors. The highest correlating variable to the residual error in the primary network is selected as the additional input. The last cascaded network is trained with the remaining residual error and the previous error values are used as input. The final network merges outputs of cascaded neural networks and compensates the main predictions. We considered 80% of data for training and 20% of data for validation performances. In the first setup, mixed frequencies data of single variable x are used together in one or more neural networks. The four average values together with the last successive values are used in the primary neural network (in the table marked as 1st data configuration). After the training is done in the primary network, the error pattern is obtained between the predicted and target values. The values are shifted and used as input to the error prediction network.

In the second data setup, the supplementary data are added to the proposed model by checking the orthogonality between additional data and the residual error obtained in the previously trained network. Similarly, to the previous setup, low and high frequencies data highly correlated to the target value in the primary network are used together. The main difference is in the second network, where concerning the correlation between additional parameters and residual error we select the inputs for the second network. The correlations between residual error obtained in the primary network and additional Lorenz Attractor dimensions, y and z are 0.2249 and -0.0184, respectively. It can be seen that the highest correlated supplementary input to the residual error is y-dimension. By including the y-dimension to the PEC-WNN as supplementary data, the obtained RMSE error is reduced to 0.25. If we apply the same experiment with the lower correlated data, z-dimension the RMSE error is increased to 0.61 which is even higher than when we use only one input data.

Selecting the additional data by checking the orthogonality between supplementary data and the residual error provides us a lower error without causing overfitting in our model. The results for each setup are presented in Table 1. In Figure 4. the real and predicted values of the Lorenz Attractor test data set are presented.

Table 1. The RMSE and RMSPE results for conducted PEC-WNN experiments

PEC-WNN	RMSE	RMSPE (%)
1st input data configuration	0.4429	9.03
2nd input data configuration **(y-dimension)**	**0.2455**	**5.42**
2nd input data configuration (z-dimension)	0.6157	12.55

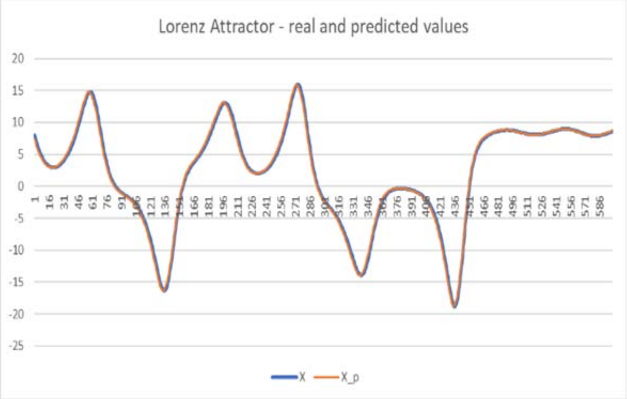


Figure 4. The real and predicted values of the first 600 values of Lorenz Attractor test data set

Table 2. Comparison of PEC-WNN RMSE to the RMSE of other implemented methods

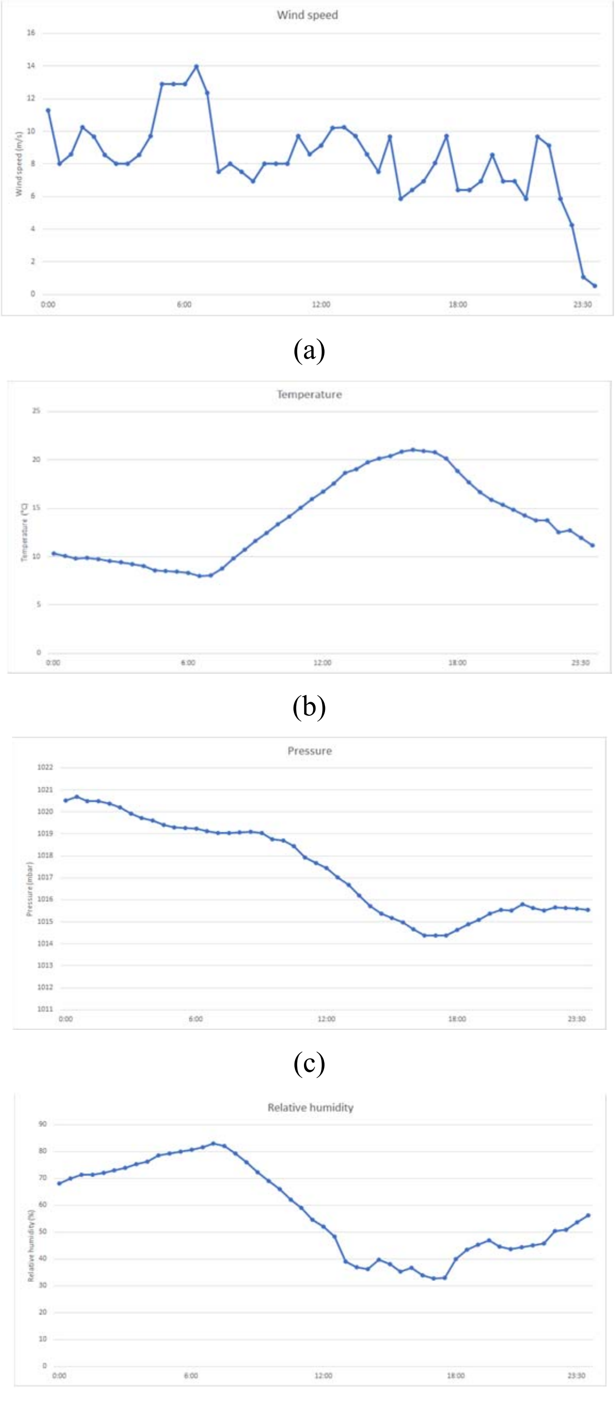
MLR	NN	WNN	LSTM	PEC-WNN
1.768	0.822	0.816	0.618	0.246

The PEC-WNN results are compared to other implemented methods, such as multivariable linear regression, neural networks, wavelet neural networks, and LSTM, the RMSE results are much higher than the PEC-WNN results. The RMSE error is higher by more than 50%, (see Table 2.). In addition to the ability to select as much as we have highly correlated supplementary parameters to the residual errors, the complexity of the proposed model and its execution time is less than the implemented models.

b. Wind speed time series prediction

In this part, to forecast the wind speed in the next 30minutes we used 30minutes of air temperature and vapor pressure data. The time sampling proposed in this work is considered as short-term wind speed prediction [19]. The data are collected by the monitoring station network of the Turkish Agricultural and Environmental Informatics Research and Application Center (TARBIL) for the period between October 2016 – July 2017. The TARBIL stations are placed next to the agricultural fields and observed 24hours/day. The relevant data are stored in the databases with 10 minutes sampling period for different parameters. The input data for our wind speed forecasting model are temperature (°C), air pressure (mbar), and relative humidity (%), similarly to [20]. The sampling frequency is 30minutes. The daily values of input parameters are given in the figures below (Figure 5. (a), (b), (c) and (d)). The correlation between input parameters is given in Figure 6. The correlation between input parameters and wind speed is less than 0.2 (Figure 6.).

In the first data setup, where we are applying mixed frequencies together, the RMSE error is 1.69 what is less more than 50% in comparison to the simple MLR, NN, WNN, or even LSTM. The supplementary data selected by checking the orthogonality between residual error obtained in the primary network and additional data reduces the error. The correlation between residual error and supplementary parameters are -0.0153 and 0.0228 for pressure and relative humidity parameters. The highest correlating parameter, relative humidity, to the residual error is used as a supplementary parameter, and the RMSE error is reduced to 1.32. In case we use pressure as a supplementary parameter, which is less correlated with residual error, the RMSE error increases to 1.58.



(a)

(b)

(c)

(d)

Figure 5. Daily meteorological data example for 30minutes average values, (a) wind speed, (b) temperature at 10m, (c) pressure, (d) relative humidity

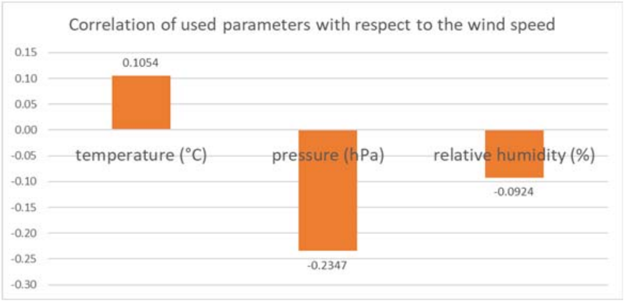


Figure 6. The correlation of input parameters with respect to the wind speed

We use 80% of data for training and 20% of data for validation performances. The same experiments were made as for the Lorentz Attractor example, which relate to the mixed frequencies, selecting additional data based on orthogonality between supplementary data and residual errors are made for the wind speed forecast model. The results and similar conclusions can be drawn by analyzing the RMSE and RMSPE errors presented in Table 3. The RMSE comparison results of the proposed PEC-WNN model to the implemented MLR, NN, WNN, and LSTM are presented in Table 4. In Figure 7. the real and predicted values of the wind speed forecasting test data set are presented.

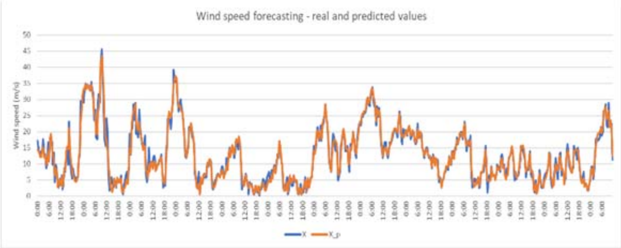


Figure 7. The wind speed forecasting measured and predicted values for test dataset

Table 3. The RMSE and RMSPE results for conducted PEC-WNN experiments

PEC-WNN	RMSE (m/s)	RMSPE (%)
1st input data configuration	1.69	22.68
2nd input data configuration (pressure)	1.58	21.64
2nd input data configuration **(humidity)**	**1.32**	**18.61**

Table 4. Comparison of PEC-WNN RMSE to the RMSE of other implemented methods

MLR	NN	WNN	LSTM	PEC-WNN
7.34	3.19	3.14	2.11	1.32

4. Conclusion

In this study, time series prediction performance improvement of PEC-WNN has been demonstrated for chaotic and stochastic dataset examples. The proposed PEC-WNN model is applied to two different time series problems and reduces the RMSE error by more than 60% compared to the multivariable regression, NN, WNN, and LSTM models. At the same time, by adding additional data the model complexity is not exponentially increased. In comparison to the implemented LSTM model, the complexity of the proposed PEC-WNN is less. On the other hand, when we use additional data if the correlation is relatively low with respect to the previous parameters, what is the case in these examples, it may increase the overfitting risk. However, choosing additional data concerning the correlation to the residual error avoids the dependency between main data, because the following network is correlated to the error pattern. PEC-WNN provides a promising prediction model for a wide range of time series prediction problems where different types of conventional networks are also possible to be imposed besides the basic schematics.

Acknowledgements

This work was supported by the research project ‘‘Development of Cyberdroid Based on Cognitive Intelligent System Applications’’ (2019-2020) that is funded by Crypttech Inc within the contract by ITUNOVA-Istanbul Technical University Technology Transfer Office. Meteorological data has been provided from Agricultural Monitoring and Information Systems Project (TARBIL 2011A020100) supported by the Turkish Ministry of Development in cooperation with Ministry of Agriculture and Forestry in Turkey.

References

[1]. Ustundag, B. B., & Kulaglic, A. (2020). High-performance time series prediction with predictive error compensated wavelet neural networks. *IEEE Access*, *8*, 210532-210541.

[2]. Kulaglic, A., & Ustundag, B. B. (2021). Stock Price Prediction Using Predictive Error Compensation Wavelet Neural Networks. *Cmc-Computers Materials & Continua*, *68*(3), 3577-3593.

[3]. Zeng, Y., Yan, E., Li, C., & Li, Y. (2008, November). Application of multivariable time series based on RBF neural network in prediction of landslide displacement. In *2008 The 9th International Conference for Young Computer Scientists* (pp. 2707-2712). IEEE.

[4]. Wang, X., & Han, M. (2014, July). Multivariate time series prediction based on multiple kernel extreme learning machine. In *2014 International joint conference on neural networks (IJCNN)* (pp. 198-201). IEEE.

[5]. Kotsiantis, S. (2011). Feature selection for machine learning classification problems: a recent overview. *Artificial Intelligence Review*, *42*(1), 157-176.

[6]. Pudil, P., & Somol, P. (2005). Current feature selection techniques in statistical pattern recognition. In *Computer Recognition Systems* (pp. 53-68). Springer, Berlin, Heidelberg.

[7]. Niu, T., Wang, J., Lu, H., Yang, W., & Du, P. (2020). Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting. *Expert Systems with Applications*, *148*, 113237.

[8]. Ogasawara, E., Martinez, L. C., De Oliveira, D., Zimbrão, G., Pappa, G. L., & Mattoso, M. (2010, July). Adaptive normalization: A novel data normalization approach for non-stationary time series. In *The 2010 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

[9]. Hwarng, H. B. (2001). Insights into neural-network forecasting of time series corresponding to ARMA (p, q) structures. *Omega*, *29*(3), 273-289.

[10]. Lee, G., & Lee, K. (2021). Feature selection using distributions of orthogonal PLS regression vectors in spectral data. *BioData Mining*, *14*(1), 1-16.

[11]. Parmezan, A. R. S., Souza, V. M., & Batista, G. E. (2019). Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information sciences*, *484*, 302-337.

[12]. Ma, Q. L., Zheng, Q. L., Peng, H., Zhong, T. W., & Xu, L. Q. (2007, August). Chaotic time series prediction based on evolving recurrent neural networks. In *2007 international conference on machine learning and cybernetics* (Vol. 6, pp. 3496-3500). IEEE.

[13]. Waheeb, W., Ghazali, R., & Herawan, T. (2016). Ridge polynomial neural network with error feedback for time series forecasting. *PloS one*, *11*(12), e0167248.

[14]. Bengio, Y., Goodfellow, I., & Courville, A. (2017). *Deep learning* (Vol. 1). Massachusetts, USA:: MIT press.

[15]. Moshiri, S., & Cameron, N. (2000). Neural network versus econometric models in forecasting inflation. *Journal of forecasting*, *19*(3), 201-217.

[16]. Patterson, D. W. (1998). *Artificial neural networks: theory and applications*. Prentice Hall PTR.

[17]. Roberts, D. (2019). Neural networks for Lorenz map prediction: A trip through time. *arXiv preprint arXiv:1903.07768*.

[18]. Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of atmospheric sciences*, *20*(2), 130-141.

[19]. Dhiman, H. S., & Deb, D. (2020). A Review of Wind Speed and Wind Power Forecasting Techniques. *arXiv preprint arXiv:2009.02279*.

[20]. Ghanbarzadeh, A., Noghrehabadi, A. R., Behrang, M. A., & Assareh, E. (2009, June). Wind speed prediction based on simple meteorological data using artificial neural network. In *2009 7th IEEE International Conference on Industrial Informatics* (pp. 664-667). IEEE.