

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**DEEP WAVELET NEURAL NETWORK FOR
SPATIO-TEMPORAL DATA FUSION**



Ph.D. THESIS

Ajla KULAGLIC

Department of Computer Engineering

Computer Engineering Programme

JULY 2022

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**DEEP WAVELET NEURAL NETWORK FOR
SPATIO-TEMPORAL DATA FUSION**

Ph.D. THESIS

**Ajla KULAGLIC
(504112516)**

Department of Computer Engineering

Computer Engineering Programme

Thesis Advisor: Prof. Dr. Burak Berk ÜSTÜNDAĞ

JULY 2022

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**UZAMSAL-ZAMANSAL VERİ FÜZYONU İÇİN
DERİN DALGACIK SİNİR AĞLARI**

DOKTORA TEZİ

**Ajla KULAGLIC
(504112516)**

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Burak Berk ÜSTÜNDAĞ

TEMMUZ 2022

Ajla KULAGLIC, a Ph.D. student of ITU Graduate School student ID 504112516 successfully defended the thesis entitled “DEEP WAVELET NEURAL NETWORK FOR SPATIO-TEMPORAL DATA FUSION”, which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Prof. Dr. Burak Berk ÜSTÜNDAĞ**
Istanbul Technical University

Jury Members : **Assoc. Prof. Dr. Sanem SARIEL UZER**
Istanbul Technical University

Asst. Prof. Dr. Tolga ENSARİ
Istanbul University-Cerrahpaşa

Prof. Dr. Deniz Turgay ALTILAR
Istanbul Technical University

Assoc. Prof. Dr. Taner ARSAN
Kadir Has University

Date of Submission : **16 June 2022**

Date of Defense : **19 July 2022**





To my father and mother,



FOREWORD

Once upon a time, a wise man told me that we never change our destination after launching a rocket. Although the direction of the string can be changed, sometimes only the route can be adjusted for the same goal. My paths changed many times during my Ph.D. journey, but today we are at the destination. Sometimes the doing is often more important than the outcome. It seemed impossible until it was over. I'm sure everyone's path is unique. For others, it may just be a doctoral degree. For me, this is proof that "impossible" is misspelled and that, in my case, it means "I'm possible." As Rumi once said: "It is your path and yours alone. Others can walk with you, but no one can walk for you."

Thanks to all who were walking with me on this road... part of it, the whole of it, who gave a contribution to it.

Still, some of them deserve to be extra mentioned. First and foremost, I am incredibly grateful to my supervisor Prof. Dr. Burak Berk ÜSTÜNDAĞ, for giving me the opportunity to work with him and for his unlimited patience, support, belief, encouragement, and selfless exchange of knowledge. The wide knowledge, strong research enthusiasm, and hard-working attitude have inspired me and will profoundly affect my career forever. His insightful support and feedback pushed me to sharpen my thinking and brought my work to a higher level. Besides my advisor, I would like to thank my committee members: Assoc. Prof. Dr. Sanem SARIEL UZER and Asst. Prof. Dr. Tolga ENSARI for their valuable guidance, comments, and encouragement to widen my research from various perspectives and successfully complete my dissertation. Prof. Dr. Deniz Turgay ALTILAR and Assoc. Prof. Dr. Taner ARSAN for their comments and support during the final stage of this dissertation.

I am sincerely grateful to my many friends and colleagues, both past and present, who have collaborated to any part of this Ph.D. Their encouragement and support helped me to pass through this journey. Specially thanks to my extraordinary friends from Sarajevo, scattered all over the world, for always being by my side, helping me to grow, challenge myself and discover my passions.

I would end up with my family, where the most fundamental source of my energy resides. This dissertation I dedicated to my amazing family, unique and perfect in many ways. Their support has been unconditional all these years, they have celebrated with me every great moment and supported me whenever I needed it, and they have given up many things for me to be here where I am now. My deepest gratitude goes to my Mother, my Father, and my Sister. Thank you for being my safety net and, ever since, my most significant support. Thank you for your moral support and unconditional love. You always have a special place in my heart. This is for you...

July 2022

Ajla KULAGLIC
(Computer Engineer)

TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	ix
TABLE OF CONTENTS	xii
ABBREVIATIONS	xiv
SYMBOLS	xv
LIST OF TABLES	xviii
LIST OF FIGURES	xxi
SUMMARY	xxiii
ÖZET	xxv
1. INTRODUCTION	1
1.1 Motivation	3
1.2 Purpose Of Thesis	4
1.3 Contributions Of This Research	5
1.4 Thesis Organization	6
2. HIGH-PERFORMANCE TIME SERIES PREDICTION WITH PREDICTIVE ERROR COMPENSATED WAVELET NEURAL NETWORKS	9
2.1 Introduction	9
2.2 Predictive Error Compensated Wavelet Neural Network Model	11
2.3 Materials and Results	16
2.3.1 The Mackey-Glass chaotic time series data	17
2.3.2 The Box-Jenkins time series data	21
2.3.3 The Lorenz Attractor	23
2.3.4 Drought forecasting	26
2.4 Discussion	29
2.5 Conclusion	30
3. STOCK PRICE PREDICTION USING PREDICTIVE ERROR COMPENSATION WAVELET NEURAL NETWORKS	33
3.1 Introduction	33
3.2 Predictive Error Compensated Wavelet Neural Network Model	38
3.3 Data	42
3.4 Results and Discussion	44
3.5 Conclusion	48
4. PREDICTIVE ERROR COMPENSATING WAVELET NEURAL NETWORK MODEL FOR MULTIVARIABLE TIME SERIES PREDICTION	51
4.1 Introduction	51
4.2 Predictive Error Compensating Wavelet Neural Networks	54
4.3 Experimental Setup and Results	57
4.3.1 Lorenz Attractor	58
4.3.2 Wind speed time series prediction	61
4.4 Conclusion	62
5. PREDICTIVE ERROR COMPENSATING NEURAL NETWORKS FOR SOIL MOISTURE ESTIMATION	65
5.1 Introduction	65
5.2 Study Area and Data Set	66
5.2.1 Study area	66
5.2.2 Data set	68
5.3 Predictive Error Compensating Neural Network Model for Soil Moisture	72
5.4 Results and Discussion	73
5.5 Conclusion	77
6. FREQUENCY DECOMPOSITION IN PREDICTIVE ERROR COMPENSATING WAVELET NEURAL NETWORK	81
6.1 Introduction	81

6.2	Predictive Error Compensated Neural Network Model with Butterworth Filter	83
6.3	Model Evaluations	86
6.4	Conclusion	88
7.	CONCLUSIONS	91
7.1	Practical Application of This Study	91
7.2	Future Research	92
	REFERENCES	95
	CURRICULUM VITAE	110





ABBREVIATIONS

AI	: Artificial Intelligence
ANN	: Artificial Neural Network
AR	: Auto-Regressive
ARIMA	: Auto-regressive Integrated Moving Average
ARMA	: Auto-regressive Moving Average
BF	: Butterworth Filter
BP	: Back-propagation Algorithm
DNN	: Deep Neural Networks
DA	: Directional Accuracy
DWT	: Discrete Wavelet Transform
EMH	: Efficient Market Hypothesis
EP	: Estimation point
FANN	: Feed-forward Artificial Neural Network
FT	: Fourier Transform
GRU	: Gated Recurrent Unit
HF	: High-Frequency
LSTM	: Long Short-Term Memory
ML	: Machine Learning
MAE	: Mean Absolute Error
MAPE	: Mean Absolute Percentage Error
MF	: Mixed-Frequency
MLR	: Multivariate Linear Regression
MSAVI	: Modified Soil Adjusted Vegetation Index
NN	: Neural Network
NDMI	: Normalized Difference Moisture Index
NDVI	: Normalized difference vegetation index
PECNET	: Predictive Error Compensating Network
PEC-NN	: Predictive Error Compensating Neural Network
PEC-WNN	: Predictive Error Compensating Wavelet Neural Network
RNN	: Recurrent Neural Networks
ReLU	: Rectified Linear Unit
RS	: Reference station
RMSE	: Root Mean Square Error
RMSPE	: Root Mean Square Percentage Error
SARIMA	: Seasonal Auto-regressive Integrated Moving Average
SGD	: Stochastic Gradient Descent
SM	: Soil Moisture
WNN	: Wavelet Neural Network
WT	: Wavelet Transform

SYMBOLS

t	: Time
ψ	: Mother wavelet function
ϕ	: Scaling function
$c_{i,j,k}$	: Scaling coefficients
$d_{i,j,k}$	: Wavelet coefficients
$x[n], y[n]$	: Signal
$h[n]$	: High frequencies
$g[n]$	: Low frequencies
$w_{j,i}$	: Weight vector
b_j	: Bias
$f(x)$	: Activation function
q	: Number of hidden neurons
m	: Number of inputs
p	: Number of output



LIST OF TABLES

	<u>Page</u>
Table 2.1 : The mathematical equations of used evaluation metrics.	17
Table 2.2 : The RMSE results of The Mackey-Glass time series data considering applied equations and forecasting time interval.....	19
Table 2.3 : The MAPE (%) results of The Mackey-Glass time series data considering applied equations and forecasting time interval.....	20
Table 2.4 : The DA (%) results of The Mackey-Glass time series data considering applied equations and forecasting time interval.....	20
Table 2.5 : The RMSE error comparison for the Mackey-Glass time series data found in the literature for prediction time interval $t = \{+1, +6, +84\}$	21
Table 2.6 : The RMSE results of the Box-Jenkins time series data considering applied models and equations.	23
Table 2.7 : The MAPE (%) results of the Box-Jenkins time series data considering applied models and equations.	23
Table 2.8 : The DA(%) results of the Box-Jenkins time series data considering applied models and equations.	23
Table 2.9 : The RMSE error comparison for Box-Jenkins time series data found in the literature for prediction time $t = \{1\}$	24
Table 2.10 : The RMSE results for Lorenz Attractor data considering applied models and equations.	26
Table 2.11 : The MAPE(%) results for Lorenz Attractor data considering applied models and equations.	26
Table 2.12 : The DA(%) results for Lorenz Attractor data considering applied models and equations.	27
Table 2.13 : The RMSE error comparison for Lorenz Attractor found in the literature for prediction time $t = \{1\}$	27
Table 2.14 : The RMSE results for applied models and equations for SPEI data set.	29
Table 2.15 : The MAPE (%) results for applied models and equations for SPEI data set.	29
Table 2.16 : The DA (%) results for applied models and equations for SPEI data set.	29
Table 2.17 : The SPEI RMSE error comparison for PEC-WNN, LSTM and LR...	30
Table 2.18 : The number of parameters of compared models.....	30
Table 3.1 : The parameter settings of Neural Network models.....	39
Table 3.2 : Attributes of used data set.	42
Table 3.3 : The Company information and period of the data used in this study. .	43
Table 3.4 : The mathematical equations of used evaluation metrics.	44
Table 3.5 : The RMSE(\$) training results for the Ford company.....	45
Table 3.6 : The MAPE(%) training results for the Ford company.....	46

Table 3.7 : The MAE(\$) training results for the Ford company.	46
Table 3.8 : The RMSE(\$) test results for the Ford company.....	46
Table 3.9 : The MAPE(%) test results for the Ford company.	46
Table 3.10 : The MAE(\$) test results for the Ford company.	46
Table 3.11 : The RMSE(\$) test results for the Apple company.	46
Table 3.12 : The MAPE(%) test results for the Apple company.	47
Table 3.13 : The MAE(\$) test results for the Apple company.	47
Table 3.14 : The RMSE(\$) test results for the Micron company.....	47
Table 3.15 : The MAPE(%) test results for the Micron company.	47
Table 3.16 : The MAE(\$) test results for the Micron company.	47
Table 3.17 : The RMSE (\$) test results for the Bank of America company.	47
Table 3.18 : The MAPE (%) test results for the Bank of America company.	48
Table 3.19 : The MAE (\$) test results for the Bank of America company.	48
Table 4.1 : The RMSE and RMSPE results for conducted PECNET experiments.	59
Table 4.2 : Comparison of RMSE of PECNET to other implemented methods. ..	60
Table 4.3 : The RMSE and RMSPE results for conducted PECNET experiments.	62
Table 4.4 : Comparison of RMSE of PECNET to other implemented methods. ..	62
Table 5.1 : Interpretation of tensiometric soil moisture readings.....	68
Table 5.2 : Correlation matrix of all input parameters.	74
Table 5.3 : The distances between selected stations to the central point.	74
Table 5.4 : The sample data used as input for the selected 63.02 station.	75
Table 5.5 : The RMSE error (cb) of soil moisture at 45cm with respect to the conducted experiments.....	76
Table 5.6 : Test RMSE error results for soil moisture estimation using Inverse Distance Weighted Interpolation (IDWI), Multi-variable Linear Regression (MLR), Neural Networks (NN), and the proposed PECNET model.	77
Table 6.1 : The RMSPE (%) average filtering results for PECNET.....	87
Table 6.2 : The RMSE (TL) average filtering results for PECNET.	87
Table 6.3 : The RMSPE results for increasing frequency applied to the PECNET.	87
Table 6.4 : The RMSE results for increasing frequency applied to the PECNET.	88
Table 6.5 : The RMSPE results for decreasing frequency applied to the PECNET.	88
Table 6.6 : The RMSE results for decreasing frequency applied to the PECNET.	89
Table 6.7 : The RMSPE results for when the Butterworth filter is applied to the input signal.	89
Table 6.8 : The RMSE results for when the Butterworth filter is applied to the input signal.	89
Table 6.9 : The correlation between Far-eastern stock indices to the BIST30 index.	89
Table 6.10 : The RMSPE results when the Far-eastern index is used.	89
Table 6.11 : The RMSE results when the Far-eastern index is used.....	90
Table 6.12 : The RMSPE results when the Far-eastern index is used with the Butterworth filter.	90
Table 6.13 : The RMSE results when the Far-eastern index is used with the Butterworth filter.	90



LIST OF FIGURES

	<u>Page</u>
Figure 1.1 : Spatio-temporal soil moisture mapping.....	3
Figure 1.2 : Spatio-temporal stock closing price forecasting using temporal features of different stock markets that are mutually influenced.	4
Figure 2.1 : 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.....	13
Figure 2.2 : The DWT decomposition of the input signal using Mallat's algorithm.	15
Figure 2.3 : The chaotic Mackey-Glass time series data.	18
Figure 2.4 : The CO_2 concentration in the gas in the furnace data set, the output of the Box-Jenkins time series data.	22
Figure 2.5 : The Lorenz Attractor.....	25
Figure 2.6 : The Lorenz Attractor trajectories projection of each time series for the first 3000 values.....	25
Figure 2.7 : 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.....	28
Figure 3.1 : Predictive error compensation wavelet neural network (PEC-WNN) model for stock price forecasting.	40
Figure 3.2 : The schematic representation of training the networks.	41
Figure 3.3 : The schematic representation of update training.	42
Figure 4.1 : 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.	55
Figure 4.2 : Graphical representation of average normalization. The average value computed based on the number of inputs is subtracted from the current value.	56
Figure 4.3 : The Lorenz Attractor trajectories projection of each time series for the first 3000 values.....	60
Figure 4.4 : The real and predicted values of first 600 values of Lorenz Attractor test data set.	60
Figure 4.5 : Daily meteorological data example for 30minutes average values, (a)wind speed, (b)temperature at 10m, (c)pressure, (d)relative humidity.	63
Figure 4.6 : The correlation of input parameters with respect to the wind speed.	64

Figure 4.7 : The wind speed forecasting measured and predicted values for test data set.....	64
Figure 5.1 : Location of selected TARBIL monitoring stations. The red area represents the Sanliurfa/Hilvan province, the area where the stations are positioned. The blue square represents the footprint of Landsat scene 173/034. On the right side, the selected stations with the mapping points are presented.....	67
Figure 5.2 : The acquisition of spatial and temporal data.	69
Figure 5.3 : Remotely sensed indices for the 63.02 station: (a) NDVI, (b) NDMI, and (c) MSAVI.	70
Figure 5.4 : Remotely sensed indices for the selected station after harmonization and smoothing, 63.02 station: (a) NDVI, (b) NDMI, and (c) MSAVI.	71
Figure 5.5 : Predictive Error Compensating Neural Network for soil moisture estimation, where soil moisture data together with the remote sensing indices are used as input.	73
Figure 5.6 : Predictive Error Compensating Neural Network for soil moisture estimation, where agro-meteorological data (ET_0 -RF) with the remote sensing indices are used as input.	74
Figure 5.7 : Graphical representation of conducted experiments. (a) the soil moisture measurements are known in the reference points, the distance from the estimation point is known (x_{EP}, y_{EP}) and the model is trained with the soil moisture at estimation point; (b) soil moisture measurements are known at some reference points and remote sensing data at estimation points; and (c) agro-meteorological data from reference stations and the remote sensing data estimation points are used together with the distances for soil moisture estimation at the estimation point.	76
Figure 5.8 : PECNET model for spatio-temporal data fusion. (a_training) and (a_test) - The PECNET results for training and test data set when only soil moisture data is used; (b_training) and (b_test) - The PECNET results for training and test data set when soil moisture data together with remote sensing indices are used; and (c_training) and (c_test) - The PECNET results for training and test data set when ET_0 -RF data together with remote sensing indices are used.	78
Figure 5.9 : The spatio-temporal soil moisture changes map for time period from 28.03.2017 until 04.04.2017.	79
Figure 6.1 : The PECNET model with Butterworth filtering.....	84



DEEP WAVELET NEURAL NETWORK FOR SPATIO-TEMPORAL DATA FUSION

SUMMARY

Machine Learning (ML) algorithms have recently gained prominence in prediction problems. The construction of an accurate machine learning model becomes a real challenge concerning the nature of the data, the number of data samples as well as the accuracy and complexity of the model. This study introduced a new machine learning structure for temporal and spatio-temporal, univariate, and multivariate prediction problems. The predictive error compensation neural network model (PECNET), which combines spatio-temporal data, has been developed. Temporal data contains information within the observation time window, and its bandwidth is limited by the sampling rate.

On the other hand, spatial data provide information regarding spatial location, while spatio-temporal data combine temporal and spatial resolution together. The PECNET model can capture both time dependencies and the spatial relationships between different data resources by fusing multivariate input patterns at multiple lengths and the sampling resolution. The PECNET achieves reliable prediction performance with relatively low model complexity and minimizes the overfitting and underfitting problems. In the proposed model, additional networks are used to predict the error of previously trained networks to compensate the overall prediction. The main network uses high correlation data with the target through moving frames in multiple scales. The PECNET improves time series prediction accuracy by enhancing orthogonal features within a data fusion scheme. The same structure and hyperparameter sets are applied to quite different problems to verify the proposed model's robustness and accuracy. Root-zone soil moisture, wind speed, financial time series data, and stationary and non-stationary time series benchmark problems are selected to evaluate the PECNET model. The results have shown improvement in the prediction accuracy and overfitting prevention using multiple neural networks for distinctive types of problems. The first part of this dissertation focuses on designing and implementing the proposed PECNET model. The algorithm is implemented in the Python programming language, and the performance of the proposed algorithm is evaluated on stochastic and chaotic time series benchmark problems found in the literature. Results have highlighted some aspects of PECNET implementations.

The major contributions of the proposed method can be seen in improving the prediction accuracy for distinct types of time series data (chaotic and stochastic) using multiple neural networks where the secondary network is trained by shifted time series prediction error of the primary network. The overfitting is avoided due to an increase in recurrence-related feedback. The same structure and hyperparameter sets are applied for a wide range of time series prediction problems with moving frames

in multiple scales. The discrete wavelet transform (DWT) used for preprocessing the input data yields better accuracy improvement than directly applying the time series data to the neural network in predictive error compensation. The PECNET for the stock price prediction problem is introduced in the second part of the dissertation. The selected data represent the non-stationary time series data. Due to the difficulties in the traditional normalization techniques that deal with non-stationary time series data, the average normalization method is proposed. The average value of the current input to the neural networks is computed and subtracted from particular input data. The proposed normalization method is able to represent the different volatilities and preservation of the original properties within each input sequence. The different frequencies of stock price time series data are used together in one neural network, while an additional network uses the previous residual errors as inputs. The updated learning method is applied in this part, enhancing the overall prediction performance. In the third part, the improved PECNET model enables choosing orthogonal features in data fusion applications. Different data types can be fused into one single model by extracting valuable knowledge from multivariate time series data. The extraction of valuable knowledge is done by checking the correlation between the remaining features and residual error. The PECNET chooses the highest correlating data to the residual error acquired by the previously trained network. It is well known that irrelevant features cause overfitting in forecasting models, representing a critical issue considering the number of samples and the number of available features. Because of that, selecting the proper feature set to the essential ones will reduce the learning process's computational cost and improve the accuracy by minimizing the overfitting.

In the fourth chapter, the root-zone soil moisture problem is introduced. For this purpose, in-situ agrometeorological measurements and satellite remote sensing indices are used. The distance between the central point and known stations is calculated. The root-zone soil moisture estimation is done using only accumulated ground-based measurements as input data, using only remotely sensing indices, and combining both. Applying the PECNET to the spatio-temporal root-zone soil moisture estimation problem shows promising results. The results can be used to obtain the soil moisture map of neighboring points where sensor information is unavailable. The fifth part examines the decomposition into frequency bands of input time series data and the applicability of different filtering methods. For this purpose, the Butterworth filter is implemented and used as an additional filtering method. Besides the closing stock price as input data, the Far-Eastern stock market indices to obtain the spatial dimensional for the financial time series forecasting example has been included. The overall results showed that fusing spatial and temporal data together into a separately trained cascaded PECNET model can achieve promising results without causing overfitting or reducing the model performances. The proposed wavelet preprocessed PECNET also leaves room for improvement using various preprocessing techniques as well as different types of neural networks.

UZAMSAL-ZAMANSAL VERİ FÜZYONU İÇİN DERİN DALGACIK SINIR AĞLARI

ÖZET

Makine öğrenimi algoritmaları günümüzde tahmin konularında oldukça önem kazanmıştır. Güçlü bir makine öğrenmesi oluşturabilmek; modelin karmaşıklığı ve isabetliğinin yanında, veri sayısının çokluğu ve verinin işlenme doğası gereği, oldukça zorlayıcı bir süreç olabilmektedir.

Bu çalışmada zamansal ve uzamsal-zamansal tek değişkenli ve çok değişkenli tahmin problemleri için yeni bir makine öğrenimi yapısı tanıtılmakta olup bu amaçla, uzamsal-zamansal veriyi harmanlayan, uzamsal zamansal hata kompanzasyonlu yapay sinir ağı modeli (PECNET), geliştirilmiştir.

Zamansal veriler, gözlenen süre penceresinde ki verileri içerir ve örnekleme oranı bantgenişliğini sınırlar. Öte yandan, uzamsal veriler uzamsal konumla ilgili veriler sağlarken, bu iki veri uzamsal-zamansal veri kombinasyonu çözünürlüğü sağlar. PECNET modeli, çoklu boyutlarda ve farklı örnekleme çözünürlüklerini birleştirerek zamana bağlı ve uzamsal ilişkili farklı veri tiplerini yakalar. PECNET, görece düşük model karmaşıklığı ile güvenilir tahmin performansı elde eder ve bunun yanında fazla ve eksik uydurma sorunlarını en aza indirmeyi amaçlamaktadır. Önerilen modelde, önceden eğitilen ağların hata oranını tahmin eden ek ağlar kullanılarak, nihai tahminin optimizasyonu sağlanır. Ana ağ, çoklu ölçeklerde hareketli çerçeveler aracılığıyla yüksek korelasyon verisi kullanır.

PECNET, ortogonal özellikleri genişleterek zaman serisi tahmin isabet oranını geliştirmektedir. Bu çalışmada, sistemin gücünün ve isabetinin doğrulanması için, benzer yapı ve hiperparametre örnekleri birçok değişik alanda uygulanmaktadır. PECNET modelini değerlendirmek için literatürden kök bölgesi toprak nemi tahmini, rüzgar hızı tahmini, finansal zaman serisi verileri, durağan ve durağan olmayan zaman serileri kıyaslama problemleri gibi konular seçilmiştir. Sonuçlar; farklı türdeki problemler için çoklu sinir ağları kullanımının, tahmin doğruluğunun artırılmasında ve fazla uydurmanın önlenmesinde büyük katkı sağladığını göstermiştir. Bu tezin ilk bölümünde, önerilen PECNET modelinin dizaynı ve uygulamasına odaklanılmıştır. Algoritma Python programlama dilinde uygulanmış ve önerilen algoritmanın performansı literatürde bulunan stokastik ve kaotik zaman serisi kıyaslama problemleri üzerinde değerlendirilmiştir. Sonuçlar, PECNET uygulamalarının bazı avantajlarını öne çıkarmaktadır. Önerilen modelin farklı tip zaman serilerinde (stokastik ve kaotik) çoklu yapay sinir ağları kullanımının, isabet oranını artırması en önemli kazanım olarak görülmektedir. Bu ağ önceki ağın kaydırılmış zaman serisi tahmin hatalarının girdi olarak eğitilmesi ile oluşturulur. Tekrarlama kaynaklı geri bildirimin kontrolsüz artışından dolayı fazla uydurmadan kaçınılmaktadır. Aynı yapı

ve hiperparametre örnekleri geniş bir alanda zaman serisi tahmin problemlerinde hareketli zamanlarda çoklu ölçülerde uygulanmıştır. Ayrık dalgacık dönüşümü (DWT) tahminsel hata kompanzasyonu yaparken, yapay sinir ağına zaman serisi verisini direk uygulamaktansa veriyi önce işleyerek isabet oranını artırır.

İkinci bölümde, PECNET modelinin hisse fiyatı tahmininde kullanımı ortaya konmuştur. Seçilen veri durağan olmayan zaman serisi verisini temsil etmektedir. Durağan olmayan zaman serisi verisini ele alan geleneksel normalleştirme tekniklerinin zorluğundan dolayı, ortalama normalleştirme metodu kullanılmıştır. Yapay sinir ağına gelen mevcut girdinin ortalama değeri hesaplanarak belirli girdi verilerinden çıkarılır. Önerilen normalleştirme yöntemi, her bir girdi dizisi içindeki orijinal özelliklerin farklı oynaklıklarını ve orijinal özelliklerin korunmasını temsil etmektedir.

Farklı frekanslardaki hisse fiyatı zaman serisi verileri tek bir ağda kullanılırken, ek bir ağ önceki artık hataları girdi olarak kullanılır. Güncellenen öğrenme metodu bu bölümde uygulanır ve böylece nihai tahmin performansının artırılması amaçlanmaktadır.

Üçüncü kısımda, geliştirilen PECNET modeli veri füzyon uygulamalarında ortogonal öz niteliklerin seçilebilmesine olanak verilmektedir. Çok değişkenli zaman serisi verilerinden faydalı bilgiler elde edilerek farklı veri türleri tek bir modelde birleştirilebilmektedir. Değerli bilginin elde edinişi, kalan özellikler ve ortaya çıkan hatanın korelasyonunun kontrol edilmesiyle sağlanır. PECNET modeli önceki eğitilen ağdan ortaya çıkan hata ile en çok korele veriyi seçmektedir. Uyumsuz özellikler tahmin modellerinde fazla uydurmaya neden olduğu iyi bilinmektedir. Örneklerin sayısı ve mevcut özelliklerin sayısı göz önüne alındığında bu kritik bir mesele olarak karşımıza çıkmaktadır. Bu nedenle uygun ve gerekli özellik örneklerinin seçimi, öğrenme işlemlerinin hesaplama maliyetini düşürerek fazla uydurmanın önüne geçilmesi ve isabet oranını artırmada kilit rol oynamaktadır.

Dördüncü bölümleri birlikte kullanılmıştır. İlk olarak merkez nokta ve bilinen istatistiksel kısımda PECNET modeli kök bölgesi toprak nem oranı ölçümlerine uygulanmıştır. Bu amaçla, yerinde ölçümler ve uydusal uzaktan ölçüm indikatörleri arasında mesafe ölçülmüştür. Kök bölgesi toprak nem tahmini, sadece gelen arazi verisini girdi olarak alarak, sadece uydu indikatörleri verisi olarak ve son olarak ikisini kombine ederek yapılmaktadır. PECNET modelinin uzamsal-zamansal kök bölgesi toprak nem tahmin problemine uygulanması gelecek vadede sonuçlar göstermektedir. Elde edilen veriler sensör bilgisinin olmadığı komşu noktaların haritasının çıkarılmasında kullanılabilir.

Beşinci kısımda önceki kısmı genişletip giriş verisinin frekans analizine odaklanmaktadır. Model performansının iyileşip iyileşmediğinin kontrolü için Butterworth filtresi eklenmiştir. Ayrıca bu kısımda Uzakdoğu borsa endeksi bilgisi uzamsal boyuta ek parametre olarak eklenmiştir. Genel sonuçlar, uzamsal ve zamansal verilerin ayrı ayrı eğitilmiş kademeli bir PECNET modelinde birleştirilmesi ile, aşırı uydurma problemlerine yol açmadan ve model performansını düşürmeden umut verici sonuçlar elde edilebileceğini göstermiştir. Önerilen ön işleme tabi tutulmuş dalgacık dönüşümlü PECNET, farklı sinir ağlarının yanı sıra çeşitli ön işleme tekniklerinin iyileştirilmesine de ışık tutmaktadır.

Tez katkısı. Önceki bölümlerde bahsedildiği gibi, bir makine öğrenmesi algoritması geliştirilmiştir. Bu model aşırı uydurma problemini minimize ederek ve farklı uzamsal-zamansal veri tiplerini birleştirerek isabetli sonuçlar vermektedir. İlk araştırma tahmin isabetini geliştirmeye yönelik kaotik ve stokastik zaman serilerini kullanan PECNET modelini içerir. Ana NN en çok korele veriyi hedef alır. Öğrenme sonrası hatalar hesaplanır ve ilave ağ hata paternleri eğitimi için oluşturulur. Amaçlanan algoritma, geniş bir alanda data serisi tahmin problemlerinde hareketli çerçevede ve çoklu ölçeklerde ana yapıyı ve hiperparametre örneklerini değiştirmeden uygulanır. İkinci çalışma ek özelliklerin artık hataların ortogonalitesi kontrol edilerek oluşturulur. Ek olarak, giriş verisiyle düşük korelasyonlu olsa bile, önceki eğitilen ağların en çok korelasyonlu ek özellikler seçilerek oluşturulan artık hataları, model kompleksliğini artırmadan ve isabet oranını düşürmeden performansı yükseltebilmektedir. Üçüncü araştırma zamansal ve uzamsal veri kullanımıyla aynı modeli sınamaktadır. Seçilen model, farklı tipte uzamsal ve zamansal veriyi zamana ve uzaya bağlı veriyi tek bir modelde birleştirmektedir. Seçilen modelin avantajı, artık hata ile korele gerekli miktarda özelliğin seçilebilmesidir. Uzamsal ve zamansal verinin birlikte uygulanmasıyla, zamana bağlı uzamsal ve zamansal varyasyonlar anlaşılabilir ve tüm anlık durumlar için bir harita oluşturulabilir. Farklı örnekleme frekanslarında ve filtrelerdeki etkiler izlenmektedir. İsbet oranına artırmaya yönelik, ayrık dalgacık girişimi dönüşümü ve Butterworth filtresi çalışılmıştır. Farklı filtrelerin uygulanması ve karışık frekanslar yeni bir konu olmamakla birlikte, bu çalışma farklı frekanslar kullanan birleşik modellerin kompleksliği artırmadan ve aşırı uydurmadan kaçınarak isabet oranını artırılmasını amaçlamaktadır.

Tez organizasyonu. Bu tezin bölümleri, hakemli dergilerde makale olarak yayınlanmıştır. Her çalışmanın içeriği bağımsız bir bölümdür.

İkinci kısım verimli bir makine öğrenme yapısının detaylı açıklamasıdır. Yapay sinir ağında tahminsel hata kompanzasyonu komponentleriyle açığa kavuşturulmuştur ve farklı zaman serilerinde karmaşık dalgacık sinir ağı algoritması dizayn ve uygulamasına odaklanılmıştır. Bu bölüm geleneksel algoritmaların yüksek isabet ve düşük zaman kompleksliğine vurgu yapmaktadır. Stokastik ve kaotik zaman serilerinin isabet oranı tahmininde, seçilen model ağın yapısını ve hiperparametre örneklerini bozmadan başarı elde edilmektedir.

Üçüncü kısım önceki kısım çalışmalarını genişleterek seçilen algoritmanın durağan olmayan zaman serileri için tahminini açıklamaktadır. Seçilen algoritma borsa fiyat tahmini problemine güncellenen eğitimle uygulanır. Geleneksel eğitim ve güncellenen eğitim şeması performans sonuçlarıyla birlikte karşılaştırmalı gösterilmektedir. Dördüncü kısımda çoklu değişken verinin kullanımı açıklanır. Amaçlanan model ortogonal özelliklerin veri füzyonu uygulamalarında kullanımına olanak sağlar. Bu bölüm ortalama eksiltme normalizasyonu gelişmelerini göstermektedir. Bundan sonra, çoklu değişkenli zaman serisinin tahmini için PECNET hesaplaması sonuçları gösterilmektedir.

Beşinci kısım, PECNET modelinin güncel toprak nemi ölçülmesi görevinde model verileri ve arazi ölçümlerini birleştirerek tahminini kapsamaktadır. Temel bir tarımsal komponent olarak toprak nemi, gelen veri ve uzak indikatör verisinin birleşimidir.

Korele girdi verileri PECNET modeli merkez noktaya uygulanarak çevresel toprak nemi haritası çıkarılmaktadır.

Altıncı kısımda bir önceki çalışma genişletilerek giriş verisinin frekans çözümüne odaklanmaktadır. Butterworth filtresi seçilen modele uygulanarak model performansının artıp artmadığı kontrol edilmektedir. Ayrıca bu bölümde uzamsal boyut ek parametre olarak Uzakdoğu hisse endeksi bilgisi olarak girilmiştir. Uzakdoğu borsası zaman farkından ve analiz tamamlanmadan önce ilk olarak kapandığından seçilmiştir.

Sonuç olarak yedinci kısımda gerekli konseptlerin altı çizilip, ana bulgular ve gelecek araştırmalara yer verilmiştir.



1. INTRODUCTION

Data fusion represents a multilevel and multifaceted process dealing with the automatic detection, association, correlation, estimation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates and complete and timely assessments of situations and threats and their significance [1]. It combines data from several sources to produce more consistent, reliable, and relevant information. In order to achieve better and more quality inferences between multiple data sources, fusing different data types can provide more useful information than information provided by any individual data source [2]. Data fusion methods are widely used to support decision-making in dynamic and complex environments. Decisions are made based on the analysis of multiple data sources, either from similar or different types of sensors. The collection of data over time and space represents spatio-temporal data. The spatial attributes provide the information with respect to spatial locations, while temporal features explore the data over time. Temporal data represents the state of time, such as regular time series (e.g., stock prices, EEG), event sequences (e.g., sensor readings, media records, twitter data), and temporal databases (e.g., timestamp relationships) sampled at regular or irregular time interval [3]. These time values can also indicate the length of time when a particular event occurred.

On the other hand, data on terrestrial entities with global properties, parts, and related geographical entities are described by spatial data [4]. They represent one of the content-richest forms of data because they can be collected on a much larger scale [5] and represent spatial information from high-level entities such as agricultural fields to low-level entities such as points or pixels [4], also known as ground sampling distance (GSD) [6]. However, spatio-temporal data understanding and availability remain significant limitations. Spatio-temporal studies aim to understand how spatial variation differs through time and how and why the temporal variation may differ from point to point on a map [7]. The merging of spatial-temporal data gives significant

complexity to the data analysis process for two reasons: continuous and discrete adjustments in spatial and non-spatial properties of spatial-temporal data and their impact on mutually arranged neighboring spatial-temporal data. In the dissertation, an integrated modeling approach is proposed, the Predictive Error Compensating Neural Network (PECNET) model, where spatial and temporal data are used together in one model. The dissertation aims to develop a conceptual machine learning (ML) model for spatio-temporal data fusion. The thesis addresses directionality - a key issue that arises while trying to merge together spatial and temporal data. The proposed PECNET model reduces dimensionality by checking the orthogonality between the input features and residual errors, enabling the selection of the orthogonal features. Appropriate selection and data pre-processing of patterns significantly affect the performance. The PECNET minimizes overfitting without increasing the dimensionality or complexity of the model. Spatio-temporal soil moisture mapping, wind forecasting, closing stock price prediction, and chaotic and stochastic benchmark time series forecasting problems are selected and analyzed. As an important indicator, soil specifies the land surface-atmosphere inter- actions by contributing precisely to the surface energy and water balance. It represents dynamic land surface characteristics demonstrating spatial variation as a function of vegetation, land use, and topographic position. Soil mapping can be defined as a computer-assisted production of digital representations of soil type or soil properties, in our case, soil moisture, which involves the creation and population of spatially-explicit information by the use of field and laboratory methods, coupled with spatial and non-spatial soil inference systems [8]. Soil moisture mapping aims to create and populate spatial soil information systems that could help users to make decisions over territories to deal with agricultural and agro-environmental problems [9]. Fusing multiple data types is more effective than using only one data type, Figure 1.1. Further, the PECNET is applied to estimate the soil patterns, identify the potential lack and construct the soil map for a given area. The main goal is to estimate the changes in the soil so that early detection and intervention can be done. Soil observations are scarce and costly, which means that a careful selection of the observed sites is necessary to reduce the cost of soil mapping. Analyzing the soil represents an essential input to the profitable application of fertilizer,

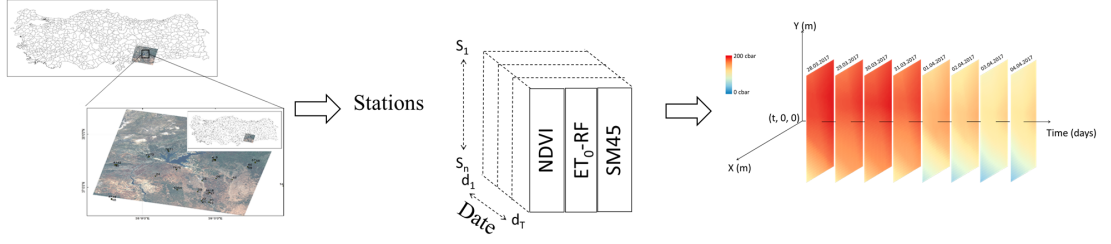


Figure 1.1 : Spatio-temporal soil moisture mapping.

lime, and other soil amendments and precision irrigating. Stochastic and chaotic time series data such as The Mackey-Glass chaotic time series, the gas furnace data (series J) of the Box-Jenkins benchmark problem, the Lorenz Attractor time series data, drought forecasting problem (SPEI index), wind forecasting and closing stock prices are selected as additional data to analyze the performance of the PECNET model. The historical data from the National Association of Securities Dealers Automated Quotations (NASDAQ) and Istanbul Stock Exchange (BIST) are used for the stock price time series problems, together with different Far-eastern indices of the markets that are closing before the stock markets mentioned above. Time-series data represent a sequence of values recorded at some time intervals. The time-series data contain a range of information having a sequence of repetition, changes, and correlation amongst cycles. The similarities and patterns in the data have been captured and investigated how the chaotic and stochastic time-series data and different sampling intervals affect the results of the proposed PECNET model. Sampling frequency or periodicity of time-series data is a fundamental characteristic. For this purpose, the data sets are arranged in different frequency frames from daily to weekly, up to monthly, and seasonally. As our target is the convolution decomposition of spatio-temporal data, to achieve spatial resolution, the impact of one on the other time series has been considered, Figure 1.2.

1.1 Motivation

The advancement in deep learning and various problems to be solved have moved the development of machine learning models to a higher level. In particular, the huge innovation in machine learning algorithms and the increase in computation power have led to an even greater interest in developing machine learning algorithms. The usage

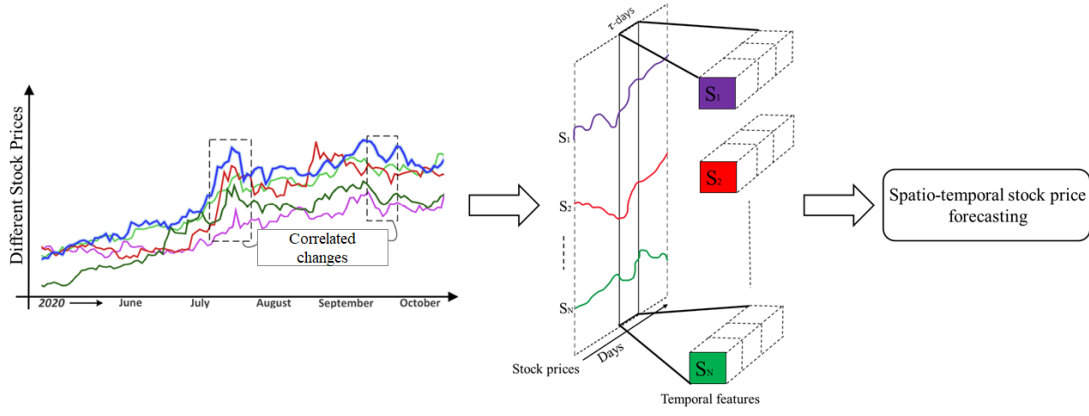


Figure 1.2 : Spatio-temporal stock closing price forecasting using temporal features of different stock markets that are mutually influenced.

of one type, taking into account the available samples as well as the complexities of the ML models, prompted us to combine spatio-temporal data into one model, considering the problems that may arise when complex data and models are used together. Given the number of input features and available data samples, the developed algorithm can solve various problems with acceptable precision. Evaluating the input data in multiple dimensions and frequencies, the proposed PECNET model, by selecting the most orthogonal features to the residual errors, can achieve decent accuracy without causing overfitting or increasing the complexity compared to different ML algorithms.

1.2 Purpose Of Thesis

The objective of this dissertation is to use different types of spatio-temporal data to assemble a machine learning algorithm that is able to achieve appropriate accuracy without causing overfitting or increasing complexity. In addition, this dissertation aims to examine the effect of different frequencies of input data, as well as additional filtering methods for spatio-temporal data. In particular, our ultimate goal is to apply the proposed model the diverse types of problems without changing the hyper-parameters and structure of the algorithm. The new heights have been outlined of innovation in using preprocessed signal patterns in the cascaded PECNET model and checking the orthogonality between input features and residual errors.

1.3 Contributions Of This Research

As mentioned in the previous sections, a machine-learning forecasting algorithm has been developed. It can achieve accurate predictions by minimizing the overfitting problem and combining different spatio-temporal data types. The first research contribution involves the improvement of the prediction accuracy for chaotic and stochastic time series data using the proposed PECNET model. The primary NN is trained with the most correlated data to the targets. After training, errors are computed, and an additional network is constructed for training error patterns. The proposed algorithm is applied for a broad range of time series prediction problems with moving frames in multiple scales without changing structure and hyperparameter sets. Our second research contribution is a selection of additional features by checking the orthogonality to the residual errors. Herewith, even with the low correlation between input data, by selecting the most correlated additional features to the residual errors of previously trained networks, the model performances can be improved without increasing the model complexity or reducing the accuracy results. The third research contribution examines spatial and temporal data usage in the same model. The proposed model fuses different types of spatial and temporal data capturing both the time dependencies of the data and the spatial dimension in one single model. The advantage of the proposed model is the selection of as many features as needed that correlate to the residual errors. By applying spatial and temporal data together, spatial and temporal variation over time can be understood, and a map can be generated for all current situations. The influence of different sampling frequencies and filters on the proposed model have been investigated. The effect of discrete wavelet transformation and Butterworth filter, which yielded accuracy improvement, have been studied. Though implementing different filters and mixed frequencies is not a new topic, our work provides cascaded models that use different frequencies that significantly improve the model's accuracy without increasing the complexity or causing overfitting.

1.4 Thesis Organization

Excerpts of this dissertation have been published or submitted in peer-reviewed journals as articles. The content of each study is a standalone chapter.

Chapter 2 presents a detailed explanation of the proposed efficient machine learning structure. The predictive error compensating neural network model is clarified with its components in detail, with the main focus on designing and implementing deep wavelet neural network algorithms for different time series prediction problems. The chapter highlights the considerably high accuracy and low time complexity concerning conventional algorithms. Along with improving the prediction accuracy for chaotic and stochastic time series data, the proposed model can find accurate solutions without changing the network structure or the hyperparameter set.

Chapter 3 extends the previous chapter and describes implementing the proposed algorithm for forecasting non-stationary time series data. The proposed algorithm is applied to the stock price forecasting problem by employing the updated training. The difference between conventional (fixed) training and the updated training scheme has been shown, as well as the performance results.

Chapter 4 introduced the usage of multi-variable data. It has been shown that the proposed model enables choosing the orthogonal features in a data fusion application. This chapter proposes and presents the development of average subtraction normalization. After that, the evaluation of PECNET for multi-variable time series prediction is shown at the end of the chapter.

Chapter 5 applies the PECNET model for the current soil moisture assessment in the root zone, where in-situ measurements are combined with remote sensing data. As an essential agricultural component, soil moisture is used in conjunction with accumulated agrometeorological data and remote indices related to soil moisture. Correlated input parameters are applied to the PECNET model with respect to the distance from the center point, generating a soil moisture map for adjacent points.

Chapter 6 extends the previous chapter and focuses on the frequency decomposition of input data. The Butterworth filter is added to the proposed model in order to check if the model performances are improved. Also, in this chapter, spatial dimension in the

form of an additional parameter is included where the additional parameter represents the far-eastern stock market information. The far-eastern stock market indices are selected due to the time difference and closing these markets before once that are analyzed as prime in the proposed model.

Finally, in concluding remarks, chapter 7 outlines the essential concepts, main findings, and future research directions.





2. HIGH-PERFORMANCE TIME SERIES PREDICTION WITH PREDICTIVE ERROR COMPENSATED WAVELET NEURAL NETWORKS ¹

2.1 Introduction

Time series forecasting is an important area that has attracted the attention of researchers from various fields, such as business, economics, finance, science, and engineering [10], [11]. In this study, an efficient ML structure for time series prediction problems has been proposed. The proposed method provides considerably higher accuracy and low time complexity compared to conventional algorithms such as long short-term memory (LSTM) and convolutional neural networks (CNN). In addition, the proposed algorithm has the ability to find accurate solutions to different types of problems without changing the hyperparameter set or network structure. The main aim of time series prediction is to collect and analyze past observations of time series data in order to develop a model that describes the behavior of the relevant system [10]. Time series problems found in the literature, such as sunspots, runoff, electrical loads, temperature, gas furnaces, drought, and rainfall, can be interpreted through chaos theory [10] - [13]. Various methods for predicting time series have been developed using linear models. Conventional statistical methods such as Auto-Regression (AR) and Auto-regressive Integrated Moving Average (ARIMA) assume linear relationships between past values. ARIMA models are relatively robust and more efficient than complex models. The development and implementation of linear methods are relatively simple; however, they are not able to capture non-linear relationships in the data [14]. In recent decades, ML methods, including artificial neural network (ANN) models, have attracted more attention in the domain of time series forecasting. These models are widely used compared to various traditional time series models. ANN models are intended to resolve nonlinear functional dependencies

¹This chapter is based on the paper B. B. Ustundag and A. Kulaglic, "High-Performance Time Series Prediction With Predictive Error Compensated Wavelet Neural Networks," in IEEE Access, vol. 8, pp. 210532-210541, 2020, doi: 10.1109/ACCESS.2020.3038724.

between time series data in the past and their future [15]. ANN models can be classified based on network structure into feed-forward and recurrent neural networks [16]. The most commonly used feed-forward neural networks in time series prediction are multilayer perceptrons (MLP) [14]. The MLP structure demands a large number of parameters to solve complex nonlinear problems. This results in a low learning rate and poor generalization [14]. To achieve better accuracy, the prediction of time series data requires NN models to be adaptable to changes that occur over time in the data. Several neural networks (NNs) and their modified models have been applied to predict nonlinear time series to overcome these drawbacks [17] - [24]. Convolutional neural networks (CNN) are widely used to learn nonlinear mapping functions from complex data [25]. CNNs can learn and extract the most important features due to special convolutional operations. Recurrent neural networks (RNNs) possess internal memory and make them capable of incorporating changes due to internal recurrence [26]. RNNs are computationally more powerful than feed-forward networks. Despite the effectiveness of NN, CNN, and RNN models in predicting time series, two main problems can be addressed. The performance of networks depends highly on the architecture and hyper-parameters of networks. 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. This study proposes a predictive error compensated wavelet neural network (PEC-WNN) model consisting of two NNs. The motivation to use two separate NNs comes from the following perspectives. First, forecasting models face expanding uncertainties such as a lack of information for more accurate predictions and the accumulation of errors. A well-known drawback of recursive methods is the susceptibility to estimation errors since their predicted values are used in the model instead of the target values [27]. On the contrary, in the proposed approach, models are trained independently and are therefore not prone to accumulation of errors. The compensation of the predicted error through the second NN improves the overall prediction performance. The PEC-WNN uses time series input data in multiple time windows. The sampled time series data in the moving time window are first transformed into a set of wavelet coefficients using discrete wavelet transform (DWT) and then fed into the NNs. DWT is applied

separately to each window by analyzing the signal over time as well as in the frequency domain. The results show that using a multidimensional time window improves prediction performance without increasing the algorithm's complexity. PEC-WNN improves accuracy while simultaneously preventing overfitting by taking advantage of multi-resolution DWT and NN. The main contributions of the proposed method can be summarized as follows:

- 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 an increase of recurrence-related feedback input.
- The same structure and hyperparameter sets can be applied for a broad range of time series prediction problems with moving frames in multiple scales.
- 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 (2.2), the proposed PEC-WNN model for time series prediction has been explained. 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 2.3 along with 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 2.4. Section 2.5 presents the concluding remarks.

2.2 Predictive Error Compensated Wavelet Neural Network Model

The predictive error compensated wavelet neural network model (PEC-WNN) utilized in this study comprises two separate wavelet preprocessed neural networks, as demonstrated in figure 2.1. The current input is shifted to the previous value using the unit delay operator z^{-1} (see figure 2.1.a)). Along with the four consecutive values, the average values of the different time intervals have been obtained by applying the unit delay operator (Figure 2.1.b)) in the same manner as in figure 2.1.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, which 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 [28] - [29]. 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 the time and frequency domain is represented by $\psi(t)$, a scaling function is defined by $\phi(t)$, 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 2.1 and 2.2.

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}} \psi(2^{-j}t - k) \quad (2.1)$$

$$\phi_{j,k}(t) = 2^{-\frac{j}{2}} \phi(2^{-j}t - k) \quad (2.2)$$

Formerly, the data are sampled into a window W_i as in (2.21):

$$W_i = \begin{cases} \text{None}, & i < l_x \\ x(i - l_x + 1), x(i - l_x + 2), \dots, x(i), & i > l_x \end{cases} \quad (2.3)$$

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, the double-scaling equations are obtained as given in 2.4 and 2.5:

$$\psi(2^{-j}t) = \sqrt{2} \sum_k h_k \psi(2^{-(j-1)}t - k) \quad (2.4)$$

$$\phi(2^{-j}t) = \sqrt{2} \sum_k g_k \phi(2^{-(j-1)}t - k) \quad (2.5)$$

The DWT applied to the current window, the scaling coefficients $c_{i,j,k}$, and the wavelet coefficients $d_{i,j,k}$ are computed using the following equations:

$$c_{i,j,k} = \sum_n h_{n-2k} c_{i,j-1,n} \quad (2.6)$$

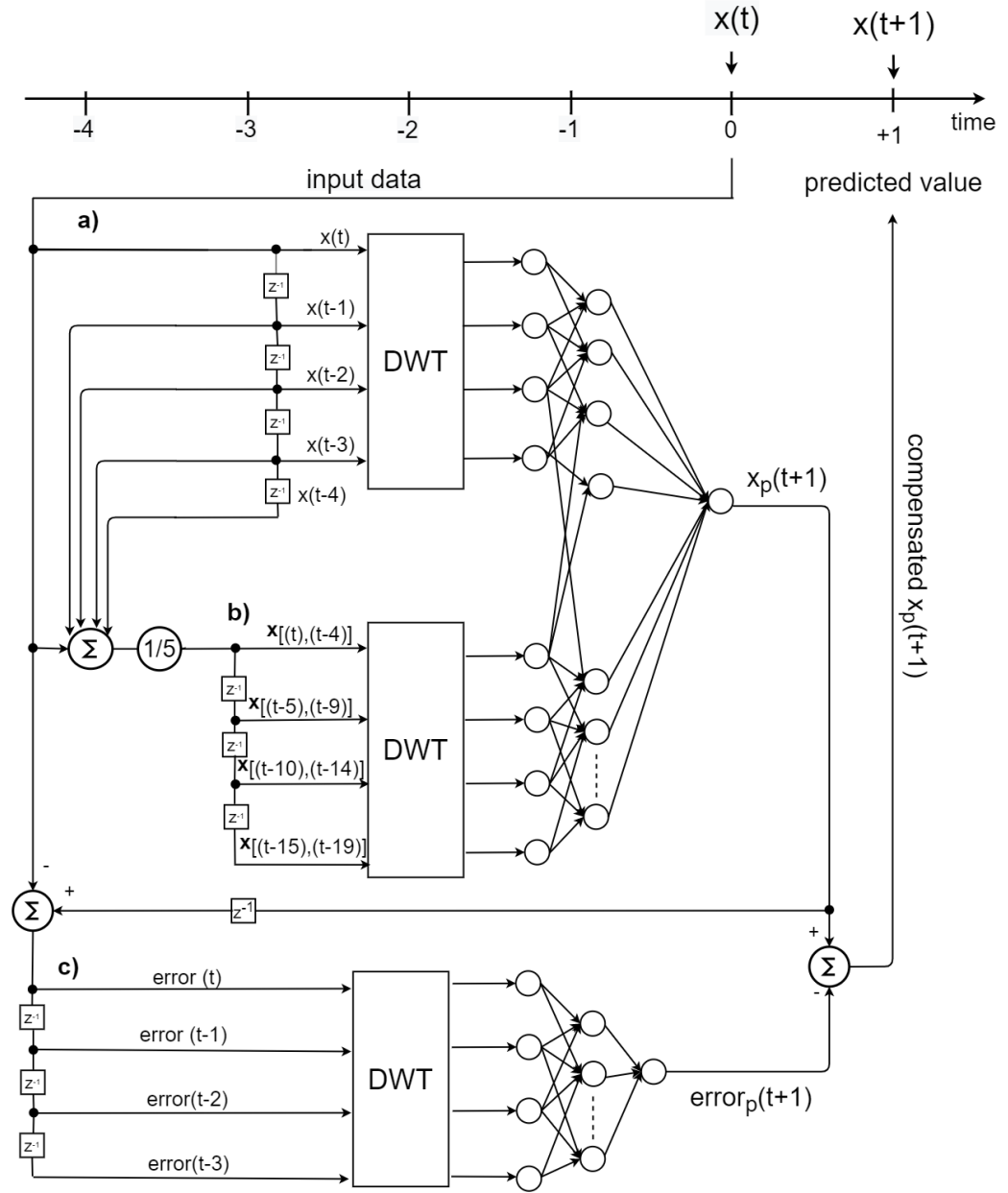


Figure 2.1 : The Predictive Error Compensated Wavelet Neural Network model (PEC-WNN): a) Consecutive input values shifted using the unit delay operator z^{-1} , b) Average values of size five computed applying the unit delay operator three times, c) Forecasting error calculation and input preparation for the compensation network.

$$d_{i,j,k} = \sum_n g_{n-2k} c_{i,j-1,n} \quad (2.7)$$

where the coefficients h_{n-2k} and g_{n-2k} can be performed by using (2.4) and (2.5):

$$h_{n-2k} = \langle \phi_{j,k}, \phi_{j-1,k} \rangle \quad (2.8)$$

$$g_{n-2k} = \langle \psi_{j,k}, \psi_{j-1,k} \rangle \quad (2.9)$$

Equations 2.6 and 2.7 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:

$$\hat{c}_{i,j-1,n} = \sum_k h_{n-2k} c_{i,j,k} + \sum_k g_{n-2k} d_{i,j,k} \quad (2.10)$$

This work uses the symmetric Haar wavelet function as a common wavelet basis function. The Haar wavelet function beneficially diminishes the distortion rate during the signal decomposition and reconstruction, significantly reducing the processing and computational time [30]. Mallat's pyramidal algorithm provides high- (h_n) and low- (g_n) frequencies from a given signal used to decompose 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 multilevel wavelet decomposition block diagram is presented in figure 2.2, with the coefficients used as input to the first NN. The prediction of n -step-ahead time series data is obtained using the main network characterized by three layers: input, hidden, and output. Mathematically, a hidden layer with activation function $g(\cdot)$ and k hidden neurons can be represented as given in 2.11:

$$g\left(\sum_{i=1}^n w_{ji} \cdot x_i + b_j\right) = y_j, \quad j = 1, 2, 3, \dots, k, \quad (2.11)$$

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 2.12:

$$g\left(\sum_{i=1}^k \beta_{ji} \cdot y_i + b_j\right) = O_j, \quad j = 1, 2, 3, \dots, N, \quad (2.12)$$

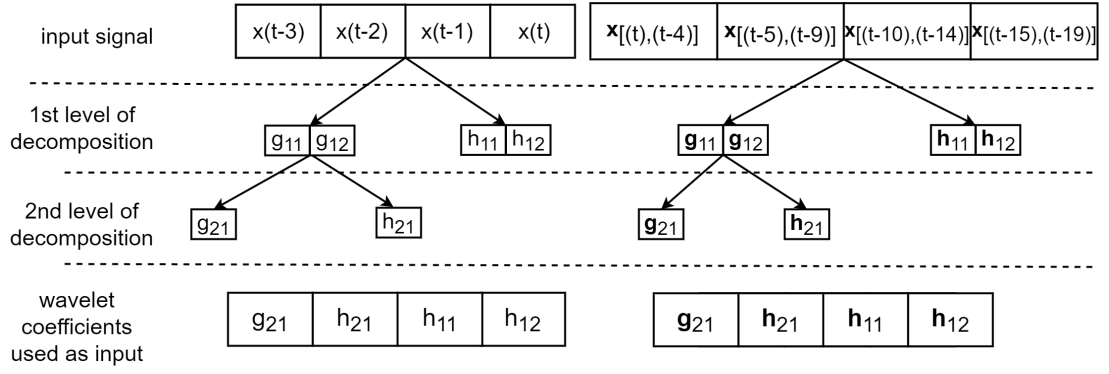


Figure 2.2 : The DWT decomposition of the input signal using Mallat's algorithm.

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 β_i , w_i , and b_i such that:

$$\sum_{i=1}^n \beta_{ji} g(w_i \cdot x_j + b_j) = t_j, \quad j = 1, 2, 3, \dots, n \quad (2.13)$$

The employed network uses Rectified Linear Unit (ReLU) activation function. The ReLU activation function, compared to the widely used activation functions (sigmoid and hyperbolic tanged), significantly improves the performance of the feed-forward networks [31]. The ReLU is a linear function that returns the value provided as input if the value is higher than zero as given in (2.14).

$$f(x) = x^+ = \max(0, x) \quad (2.14)$$

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 by applying the unit delay operator z^{-1} (Figure 2.1.c)). The output of the second NN is the prediction of the error at a time $(t + 1)$ using the prediction errors at the 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:

$$x_p(t + 1) = f(x(t), x(t - 1), x(t - 2), x(t - 3), \bar{x}_{[t, t-4]}, \bar{x}_{[t-5, t-9]}, \bar{x}_{[t-10, t-14]}, \bar{x}_{[t-15, t-19]}) \quad (2.15)$$

$$\overline{x[i, j]} = \frac{\sum_i^j x_i}{(j - i) + 1} \quad (2.16)$$

$$error_p(t + 1) = f(err(t), err(t - 1), err(t - 2), err(t - 3)) \quad (2.17)$$

$$error(t) = real\ value(t) - forecasted\ value(t) \quad (2.18)$$

$$compensated\ value(t + 1) = x_p(t + 1) - error_p(t + 1) \quad (2.19)$$

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 $(x[t, t - 4], x[t - 5, t - 9], x[t - 10, t - 14], x[t - 15, t - 19])$ computed using 2.15, represent the input of the first NN. The average values of an interval $[i, j]$ are computed using 2.16. The input data of the second NN contains four errors obtained using the predicted values from the first NN using 2.17. 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 2.19.

2.3 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 the 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

Table 2.1 : The mathematical equations of used evaluation metrics.

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

are scaled by using the minimum/maximum normalization method given in equation 2.20:

$$x(t)_{norm} = (max_2 - min_2) \cdot \left(\frac{x(t) - min_1}{max_1 - min_1} \right) + min_2 \quad (2.20)$$

where $x(t)$ represents the actual 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 table 2.1, 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:

$$W_i = \begin{cases} 1 & (X_{model,i} - X_{model,i-1})(X_{obs,i} - X_{obs,i-1}) \geq 0 \\ 0 & otherwise \end{cases} \quad (2.21)$$

2.3.1 The Mackey-Glass chaotic time series data

The chaotic Mackey-Glass time series data (Figure 2.3) has been typically used as a benchmark problem before considering the suitability of a specific approach to real-world forecasting problems [32]. The time series data have been generated from the following differential equation 2.22:

$$\frac{dx(t)}{dt} = \frac{\alpha x(t - \tau)}{1 + x^{10}(t - \tau)} - bx(t) \quad (2.22)$$

where x (unitless) is the series in time t , and τ is the time delay. The parameters α , β and τ are set as $\alpha = 0.2$, $\beta = 0.1$, $\tau = 17$. Note that, for $\tau > 17$, the time series show chaotic behavior [33]. 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 [33] uses non-consecutive values with the constant time interval, $T = 6$, to

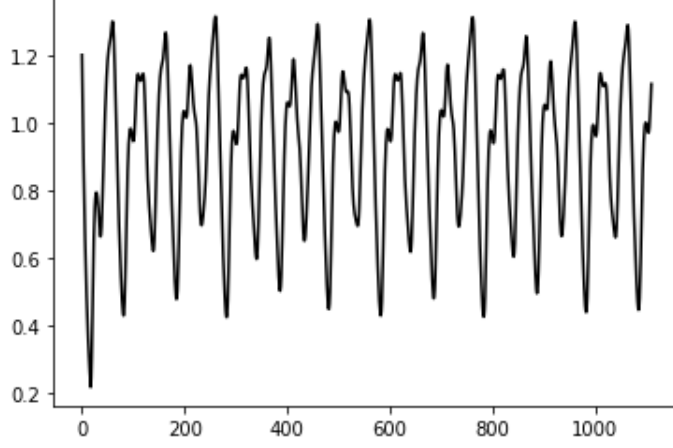


Figure 2.3 : The chaotic Mackey-Glass time series data.

predict 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, [32] 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 were served for training and the remaining half for testing. Unlike the previous studies, in this work, constructed data sets contain the averages of different window sizes together with sequential values. The first data set contains only four successive values obtained by 2.23. In order to observe the effects of the average values on the forecasting performance in the following data sets, the average values of window size 5 and size 10 are included, obtained using equations 2.24 and 2.25, respectively, which can be considered as daily data. Hence by using the average values of size 5, the business week resolution of data has been gained. Moreover, the monthly resolution is obtained by extending it to the four shifted average values. 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 [33].

$$MG_1 : y_1(t+T) = f(x(t-3), x(t-2), x(t-1), x(t)) \quad (2.23)$$

$$MG_2 : y_2(t+T) = f(x(t), x(t-1), x(t-2), x(t-3), x_{[0,4]}, x_{[5,9]}, x_{[10,14]}, x_{[15,19]}) \quad (2.24)$$

Table 2.2 : 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_1 (2.23)	0.013	0.006	0.013	0.005
	MG_2 (2.24)	0.003	0.001	0.002	0.001
	MG_3 (2.25)	0.010	0.007	0.011	0.006
x+5	MG_1 (2.23)	0.072	0.029	0.071	0.029
	MG_2 (2.24)	0.005	0.004	0.005	0.004
	MG_3 (2.25)	0.0031	0.019	0.029	0.018
x+6	MG_1 (2.23)	0.087	0.036	0.086	0.035
	MG_2 (2.24)	0.004	0.003	0.003	0.002
	MG_3 (2.25)	0.006	0.004	0.005	0.003
x+42	MG_1 (2.23)	0.136	0.042	0.131	0.034
	MG_2 (2.24)	0.033	0.026	0.024	0.014
	MG_3 (2.25)	0.018	0.012	0.014	0.010
x+84	MG_1 (2.23)	0.159	0.040	0.160	0.038
	MG_2 (2.24)	0.124	0.035	0.082	0.030
	MG_3 (2.25)	0.067	0.028	0.41	0.025

$$MG_3 : y_3(t+T) = f(x(t), x(t-1), x(t-2), x(t-3), x_{[0,9]}, x_{[10,19]}, x_{[20,29]}, x_{[30,39]}) \quad (2.25)$$

The RMSE results are presented concerning the constructed data sets (explained in the above equations) and its forecasting time interval in table 2.2. The results show that using 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 window size selection 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, compared to the MAPE (Table 2.3), appending the averages to the consecutive values improves the results and reduces forecasting errors. The results for directional accuracy, DA in table 2.4, show descent results for applied models where the PEC-WNN model precedes. The Mackey-Glass time series problem results showed that the PEC-WNN achieves the lowest RMSE error for different forecasting intervals. The second-best result found in the literature [34], for short-term forecasting ($t+1$) is 0.0327. The same time-interval forecasting done with PEC-WNN obtained the RMSE

Table 2.3 : 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_1 (2.23)	1.253	0.547	1.266	0.486
	MG_2 (2.24)	0.264	0.152	0.200	0.118
	MG_3 (2.25)	0.925	0.552	1.001	0.643
x+5	MG_1 (2.23)	7.200	2.519	6.991	2.49
	MG_2 (2.24)	0.486	0.366	0.542	0.407
	MG_3 (2.25)	2.912	1.747	2.722	1.534
x+6	MG_1 (2.23)	8.767	3.475	8.584	3.100
	MG_2 (2.24)	0.374	0.260	0.311	0.241
	MG_3 (2.25)	0.568	0.381	0.513	0.369
x+42	MG_1 (2.23)	13.446	3.809	12.605	2.938
	MG_2 (2.24)	3.120	1.840	2.161	1.203
	MG_3 (2.25)	1.722	0.970	1.252	0.970
x+84	MG_1 (2.23)	16.307	3.660	16.351	3.524
	MG_2 (2.24)	12.113	2.517	7.935	2.782
	MG_3 (2.25)	6.610	2.692	3.535	2.414

Table 2.4 : 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_1 (2.23)	96.267	96.673	96.493	96.719
	MG_2 (2.24)	98.484	98.592	98.672	9.709
	MG_3 (2.25)	95.682	95.482	95.227	95.341
x+5	MG_1 (2.23)	69.432	70.682	72.612	74.659
	MG_2 (2.24)	96.816	96.432	96.226	96.755
	MG_3 (2.25)	80.909	87.343	87.913	89.369
x+6	MG_1 (2.23)	62.841	63.841	62.955	64.091
	MG_2 (2.24)	97.642	98.231	97.642	98.426
	MG_3 (2.25)	95.873	95.991	94.108	96.187
x+42	MG_1 (2.23)	61.457	66.510	64.160	67.568
	MG_2 (2.24)	87.912	89.499	87.424	90.005
	MG_3 (2.25)	88.889	91.209	91.941	93.592
x+84	MG_1 (2.23)	49.572	58.752	58.384	59.241
	MG_2 (2.24)	74.268	75.924	76.815	77.834
	MG_3 (2.25)	74.268	82.293	75.921	86.242

Table 2.5 : 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	[34]
6th-order Polynom		0.04	0.85	[34]
BPNN		0.02	0.06	[36]
DSC-LMM		0.0055	0.03	[35]
Cascade correlation NN		0.06		[37]
PEC-WNN	0.0013	0.0027	0.028	
LSTM	0.0576	0.0644	0.1642	

of 0.0013, which reduces the RMSE by 95%. Similar results are noticed for the $(x + 6)$ forecasting term, where the best result found in the literature is 0.0055 [35]. This result is obtained using the dynamic cell structures and local linear models (DSC-LMM) proposed in [35]. The PEC-WNN RMSE for a time interval $(t + 6)$ was 0.0027, which is 49% better than 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. [35] reached the RMSE of 0.03 using the DSC-LMM model.

2.3.2 The Box-Jenkins time series data

The Box-Jenkins time series data set is another frequently used benchmark example in the prediction algorithms [33]. 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 [38]. The first stage determines the stationarity of the data. The dependent time series data plots are used to decide which autoregressive or moving average components should be applied. In the second stage, the parameters of the selected model are estimated using the maximum likelihood or non-linear least square estimation. In the last step, the statistical model checking examines 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 problems. The output of the Box-Jenkins gas furnace time series data set is given in figure 2.4. The inputs proposed in the literature, as given in equation 2.26, are applied first. Subsequently, successive values of the methane gas flow (Equation 2.27) are used to forecast the

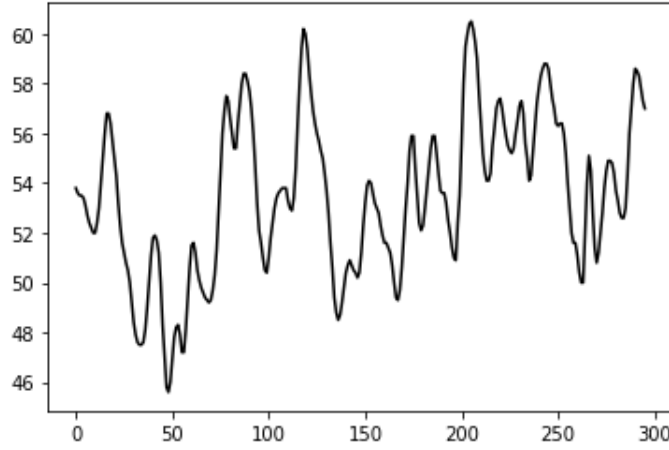


Figure 2.4 : The CO_2 concentration in the gas in the furnace data set, the output of the Box-Jenkins time series data.

successive value. The investigation is expended by checking how the increased amount of input data with its average values affect the forecasting performances. For that purpose, the averages of five and ten window sizes with four successive values of methane gas flow with CO_2 concentration in the gas are applied. Note that the forecasting value is always the next $(t + 1)$ value of CO_2 concentration in the gas, while the input data set differs. The input equations for applied experiments are given below (Equations 2.26 -2.29):

$$BJ_1 : y_1(t + 1) = f(v(t - 3), y(t)) \quad (2.26)$$

$$BJ_2 : y_2(t + 1) = f(v(t), v(t - 1), v(t - 2), v(t - 3), y(t)) \quad (2.27)$$

$$BJ_3 : y_3(t + 1) = f(v(t), v(t - 1), v(t - 2), v(t - 3), v_{[0,9]}, v_{[10,19]}, v_{[20,29]}, v_{[30,39]}, y(t)) \quad (2.28)$$

$$BJ_4 : y_4(t + 1) = f(v(t), v(t - 1), v(t - 2), v(t - 3), v_{[0,4]}, v_{[5,9]}, v_{[10,19]}, v_{[20,29]}, y(t)) \quad (2.29)$$

The results indicate that the proposed PEC-WNN shows the lowest RMSE (Table 2.6) when the data set with four successive and four average values with window size ten is used. Compared to the MAPE (Table 2.7) and DA results (Table 2.8), the PEC-WNN shows the lowest MAPE and the highest DA results. The applied average values of a

Table 2.6 : The RMSE results of the Box-Jenkins time series data considering applied models and equations.

Equation	NN	PEC-NN	WNN	PEC-WNN
BJ_1 (2.26)	0.082	0.081		
BJ_2 (2.27)	0.077	0.076	0.074	0.071
BJ_3 (2.28)	0.074	0.068	0.066	0.059
BJ_4 (2.29)	0.070	0.069	0.064	0.061

Table 2.7 : The MAPE (%) results of the Box-Jenkins time series data considering applied models and equations.

Equation	NN	PEC-NN	WNN	PEC-WNN
BJ_1 (2.26)	25.268	28.022		
BJ_2 (2.27)	20.126	18.707	27.680	37.744
BJ_3 (2.28)	17.927	21.232	16.577	23.513
BJ_4 (2.29)	19.401	17.104	19.163	21.150

different time window to the consecutive values improve the results. In the literature, the RMSE results for the gas furnace (series J), or the Box-Jenkins problem, change 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.

2.3.3 The Lorenz Attractor

The Lorenz Attractor represents a classical time series multivariate prediction problem consisting of three ordinary differential equations given in 2.30-2.32.

$$\frac{dx}{dt} = \sigma(y - x) \quad (2.30)$$

$$\frac{dy}{dt} = x(\rho - z) - y \quad (2.31)$$

Table 2.8 : The DA(%) results of the Box-Jenkins time series data considering applied models and equations.

Equation	NN	PEC-NN	WNN	PEC-WNN
BJ_1 (2.26)	90.5	90.5		
BJ_2 (2.27)	91.0	91.0	90.5	91.0
BJ_3 (2.28)	91.0	91.5	91.5	91.5
BJ_4 (2.29)	90.0	90.5	90.0	91.0

Table 2.9 : 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	[38]
Tong's model	0.685	[39]
Xu's model	0.573	[40]
Sugeno's model	0.596	[41]
Surmans' model	0.4	[42]
Adaptive network-based fuzzy inference system (ANFIS)	0.405	[43]
Generalized fuzzy NN	0.273	[43]
PEC-WNN	0.0589	
LSTM	0.9730	

$$\frac{dz}{dt} = xy - \beta z \quad (2.32)$$

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 [44]. Different Lorenz maps with the same general dynamics can be obtained using distinctive initial conditions and parameter values. The Lorenz map is given in figure 2.5. The data set contains 10,000 multivariate data samples. The plots of each trajectory interdependencies between the time series can be seen (Figure 2.6). Xiu et al. [45] applied a multivariate data set as inputs to predict the single variable as the output. Similar single and multivariate inputs have been used for the proposed model. The input equations are given below in 2.33-2.35. The output represents the next $(t + 1)$ value of a single variable.

$$LA_1 : x(t + 1) = f(x(t), y(t), z(t)) \quad (2.33)$$

$$LA_2 : x(t + 1) = f(x(t), y(t), y(t - 1), y(t - 2), \\ y(t - 3), z(t), z(t - 1), z(t - 2), z(t - 3)) \quad (2.34)$$

$$LA_3 : x(t + 1) = f(x(t), y(t), y(t - 1), y(t - 2), y(t - 3), z(t), \\ z(t - 1), z(t - 2), z(t - 3), \bar{y}_{[0,9]}, \bar{y}_{[10,19]}, \bar{y}_{[20,29]}, \\ \bar{y}_{[30,39]}, \bar{z}_{[0,9]}, \bar{z}_{[10,19]}, \bar{z}_{[20,29]}, \bar{z}_{[30,39]}) \quad (2.35)$$

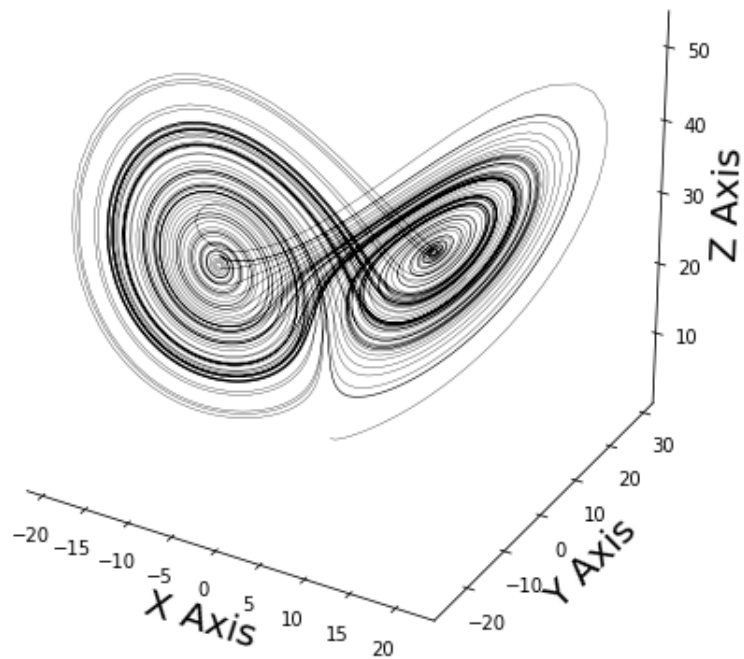


Figure 2.5 : The Lorenz Attractor.

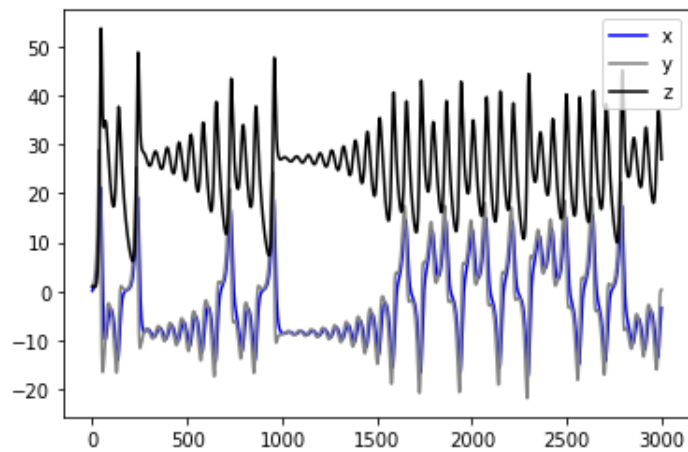


Figure 2.6 : The Lorenz Attractor trajectories projection of each time series for the first 3000 values.

Table 2.10 : The RMSE results for Lorenz Attractor data considering applied models and equations.

Equation	NN	PEC-NN	WNN	PEC-WNN
LA_1 (2.33)	0.0015	0.0014		
LA_2 (2.34)	0.0359	0.0340	0.0356	0.0308
LA_3 (2.35)	0.0609	0.0476	0.0378	0.0113

Table 2.11 : The MAPE(%) results for Lorenz Attractor data considering applied models and equations.

Equation	NN	PEC-NN	WNN	PEC-WNN
LA_1 (2.33)	0.024	0.014		
LA_2 (2.34)	1.179	0.526	1.5570	0.4287
LA_3 (2.35)	2.909	1.500	1.6632	0.3002

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 (Table 2.10) is achieved using only one previous trajectory value for forecasting the next value. Similar to previous experiments, the impact of adding average values of different time-window to the consecutive values has been observed herein. The growth in the number of variables, from 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 compared to the successive multivariate input. The lowest value of MAPE (Table 2.11) is found in the PEC-WNN model, with the utilization of consecutive and average values together. In the end, very high DA (Table 2.12) 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 than a similar experiment found in Xiu et al. literature [45]. 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.

2.3.4 Drought forecasting

In this section, the performances of the proposed model have been demonstrated when it is applied to the stochastic time series data. The benchmark problems

Table 2.12 : The DA(%) results for Lorenz Attractor data considering applied models and equations.

Equation	NN	PEC-NN	WNN	PEC-WNN
LA_1 (2.33)	99.975	99.988		
LA_2 (2.34)	99.875	99.900	99.9250	99.9750
LA_3 (2.35)	99.738	99.213	99.6250	99.7650

Table 2.13 : The RMSE error comparison for Lorenz Attractor found in the literature for prediction time $t=\{1\}$.

Models:	x+1	Source:
Neural Network	0.072	[46]
NARX(multivariate)	0.000295	[45]
NARX(single variable)	0.068	[45]
MKELM(multivariate)	0.0312	[45]
MKELM(single variable)	0.0973	[45]
PEC-WNN	0.0113	
LSTM	0.3148	

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. [47] is selected for that purpose. Drought identification and forecasting are essential in limiting their effects. However, accurate drought prediction remains a scientific issue due to the nature of the 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 [46]. 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 [48]. 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;

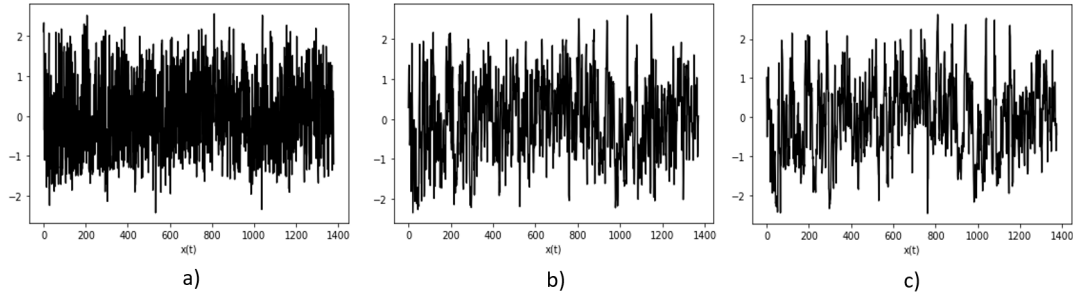


Figure 2.7 : The Standardized Precipitation Evapotranspiration Index (SPEI) for the period from January 1901 until December 2015 within a) month, b) four months and c) six months.

and, (c) it enables the identification of various drought types, given the multi-scalar character of the SPEI [48]. The analyzed period is from January 1901 until December 2015. The 1-month, 4-months, and 6-months data were used (Figure 2.7). The input-output functions for the prediction model are given in equations 2.36 and 2.37. The input data sets consist of eight inputs with different window sizes. The first data set contains one-month and four-month data (Equation 2.36); the second data set contains one-month and six-month data (Equation 2.37). In both cases, the next six months' drought period has been forecasted. 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 PEC-WNN model contains the same hyperparameters and the number of inputs applied to the previous chaotic time series problems.

$$SPEI_1 : y_1(6t + 1) = f(x(t), x(t-1), x(t-2), x(t-3), x(4t), x(4t-1), x(4t-2), x(4t-3)) \quad (2.36)$$

$$SPEI_2 : y_2(6t + 1) = f(x(t), x(t-1), x(t-2), x(t-3), x(6t), x(6t-1), x(6t-2), x(6t-3)) \quad (2.37)$$

As an alternative to the previously mentioned models, for drought forecasting problems, the LSTM model proposed in [22] and multivariate linear regression (LR) have been used for performance comparison. The proposed PEC-WNN accomplished significantly low RMSE (Table 2.14) with monthly and four-month data used as inputs

Table 2.14 : 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	$SPEI_1$ (2.36)	0.157	0.151	0.150	0.139	0.353	0.409
	$SPEI_2$ (2.37)	0.216	0.147	0.221	0.146	0.386	0.485

Table 2.15 : 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	$SPEI_1$ (2.36)	0.301	0.287	0.296	0.278	0.356	0.429
	$SPEI_2$ (2.37)	0.351	0.289	0.348	0.286	0.378	0.493

given in equation 2.36. The PEC-WNN model also provides the lowest MAPE, as seen in table 2.15. 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 simultaneously improves the proposed method's performance at different time scales.

2.4 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 shows 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 used and found in the literature. The PEC-WNN has been applied to different problems without changing the network structure and hyperparameters. The PEC-WNN, although in its structure,

Table 2.16 : 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	$SPEI_1$ (2.36)	61.26	63.18	64.21	64.95	33.94	45.33
	$SPEI_2$ (2.37)	60.0	57.46	55.86	56.61	32.17	42.76

Table 2.17 : 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

Table 2.18 : The number of parameters of compared models.

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

uses two NNs are less computationally expensive and time-consuming than other ML methods found in the literature. The PEC-WNN complexity concerning the number of parameters is relatively low (Table 2.18). The results are consistent with one of the conclusions found in [49], which states that simple models 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.

2.5 Conclusion

This work proposes a predictive error compensated wavelet preprocessed NN model for time series prediction problems. The model consists 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 in all benchmark problems. The Mackey-Glass, Box-Jenkins, and Lorenz Attractor problems are used to evaluate the prediction performance for chaotic time series cases and the 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. The proposed model's time and space complexity were less than 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.





3. STOCK PRICE PREDICTION USING PREDICTIVE ERROR COMPENSATION WAVELET NEURAL NETWORKS¹

3.1 Introduction

The stock markets are 80% controlled by machines, and over the next ten years, robots will replace 200,000 banking jobs, according to an article published in Forbes [50]. The high-frequency trading technologies represent a type of algorithmic trading that uses machine learning algorithms to implement investment strategies in extremely small-time intervals. Algorithmic trading has a significant effect on the way financial assets change hands [51]. Chague et al. [52] found that 97% of day traders end up losing money as a year approach. Similar studies found that only 1% of traders make money and have some profitability after costs [53] - [54]. Despite the evidence of negative consequences discussed in [50], the full effects of algorithm trading are yet to be seen. The stock price prediction mechanisms in the stock market are fundamental to forming investment strategies and developing risk management models [55]. As the stock market influences individual and national economies, predicting the stock market is essential when making the proper decision [56]. However, due to the uncertainty in financial time series data, the accurate prediction of stock market changes represents a challenging task. For this reason, the next day's closing stock price has been forecasted in the proposed study by evaluating input data in multiple networks. The networks are used in an additive manner. The first network is the primary predictor; an additional network is used to predict the main network error to compensate the overall daily stock prediction error. The prediction performance is significantly improved using the proposed model by reducing the overfitting and without increasing the complexity of the proposed algorithm. Early stock market prediction studies were based on random walk theory and the Efficient Market Hypothesis (EMH) [57]. The EMH states that current stock prices reflect all the available information, showing that it is

¹This chapter is based on the paper A. Kulaglic and B. B. Ustundag, "Stock Price Prediction using Predictive Error Compensation Wavelet Neural Networks," in *Computers, Materials & Continua* DOI : 10.32604/cmc.2020.014768.

impossible to predict future stock prices using past information. Furthermore, Malkiel and Fama [57] argued that any new information is immediately reflected in price changes without delay, and therefore future asset price movements are independent of past and current information. The suggestions made in [57] are that stock prices cannot be predicted since they are driven by external and new information rather than just historical or current prices. Moreover, the stock price data are disposed to frequent changes that cannot be derived from a historical trend. On the other hand, numerous studies have attempted to experimentally disprove the study of Malkiel and Fama [57] by showing that stock markets are predictable. Bachelier [58] first proposed the efficient market theory and described the stock price movement in the random walk manner. Later, the random walk characteristics of price changes were empirically tested by Cootner [59] and Fama [60]. The changes are influenced by real-world factors, such as political, social, and environmental factors [24]. In addition, the noise-to-signal ratio is very high in such conditions, and it is difficult to analyze and forecast future data. The use of econometric models is convenient for describing and evaluating the relationships between variables using statistical inference, with some limitations. These limitations can be seen in not capturing the nonlinear nature of stock prices. In addition, [61] in their study assumed to have constant variance while the financial time series is boisterous and has time-varying volatility. Thus, by the work done in the [62], it is concluded that the stock market price will follow a random walk, and prediction accuracy cannot exceed 50%. The statistical methods, such as Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), and vector autoregression have generally achieved reasonable predicting results based on the literature results [63]- [68]. These statistical models map linear relationships, but they are not very useful in stock market prediction due to the nature of stock market data. ARIMA is one of the most popular and widely used statistical techniques for making predictions using past observations [65]. In the work of Nochai and Nochai [66], they intend to find an appropriate ARIMA model. The empirical analysis of the study showed that ARIMA (2,1,0), (1,0,1), and (3,0,0) are the best models for predicting the price of palm oil. Reddy [67], in his work, tries to check the stationarity in time series data and predict the direction of change

in the stock market index using the ARIMA model. The best results are obtained for ARIMA (0,1,0). In his study, the author confirmed the ARIMA model's perspective to forecast future time series in a short time. Adebisi et al. [68], in their research, predict share prices in the short run. The results of this study showed that ARIMA models have the power to predict stock prices quickly. Despite the above-mentioned statistical models, Artificial Neural Networks (ANNs) are one of the most accurate prediction models [69]. According to [70], the ANNs, unlike statistical models with sufficient amounts of data, can approximate any finite and continuous function based on the universal approximation theorem. A forecasting system based on the radial basis function Neural Network (NN) proposed by Lendasse et al. [71] showed that the system could capture the nonlinear relationship in financial time series data. The first significant study of neural network models for stock price return prediction was accomplished by White [51], who introduced the predictive model based on IBM's daily common stock and achieved promising results. In order to increase prediction performance, hybrid models have shown significant achievements. Different hybrid systems were proposed by using ANNs, the Hidden Markov Model (HMM) [72], exponential smoothing, and ARIMA [73], and ANN with exponential generalized autoregressive conditional heteroscedasticity model [74]. Yao et al. [75] compared the back-propagation NN model with the ARIMA model. They found that the NN results in better prediction accuracy compared to the ARIMA models. Adebisi et al. [76] compared the performances of the ARIMA and ANN models for stock price prediction and found the superiority of the NN model over the ARIMA model. On the other hand, Nitin et al. in [77] did a comparative study of a three-layer feed-forward NN model and the ARIMA model to predict future stock price data and revealed that the ARIMA models perform better than NN models. Similar to [77], Lee et al. [78] developed the NN model and the Seasonal Autoregressive Integrated Moving Average (SARIMA) model for stock index prediction. Their results also found that the ARIMA model outperforms ANN models for stock prediction. The results show that the model results depend on the data. Furthermore, using ARIMA and ANN as a new hybrid model improves the prediction performances. In this work, the empirical results reveal that hybrid systems outperform all individual systems providing more

accurate prediction performances. The two most popular deep-learning architectures for stock market forecasting in recent years are the Long Short-Term Memory (LSTM) model and the Gated Recurrent Unit (GRU) model with its hybridization [79]. The LSTM models are appropriately structured to learn temporal patterns and over-perform the conventional recurrent neural networks (RNNs) as it overcomes the problem of vanishing gradients. Shahi et al. [79] proposed the LSTM and GRU deep-learning architectures and compared the performances of these two models for stock market prediction. In their study, they made a comparison of the performances of the LSTM and GRU models under the same conditions and also showed that by including the financial news sentiments together with the stock market features, the predicting model can be significantly improved. Bao et al. [30] used the LSTM for stock price forecasting using different types of sequential data. Using the sentiment features, Li et al. [80] showed that LSTMs outperform benchmark models of SVM and improve the prediction accuracy of the next day's open price. The problem of overfitting and getting stuck in local optima are additional issues that must be considered in prediction models. The problem occurs due to data amount limitations and appropriate model configuration. Even though the NN models achieve better generalization, they are prone to overfitting due to their high capacity. In financial time series forecasting using Deep Neural Networks (DNNs), overfitting occurs due to a lack of data [31]. The financial time series data in a year daily obtain approximately 252 data points. However, it is insufficient for the DNN models compared to the number of model parameters. Sufficient data is needed as the number of model parameters increases as the number of used features is enlarged. Considering that overfitting impairs prediction accuracy, regularization techniques, such as dropout, early stopping, data augmentation, or reducing the network size and learning rate are needed to avoid this problem [81]. The regularization techniques can prevent overfitting but cannot improve generalization performance. Hence, data augmentation is a method used in order to avoid overfitting while improving generalization accuracy. However, when it is about the financial time series, the data augmentation distorts the original data, which is not a simple task. Instead, in recent times, signal processing techniques have been used to transform the data into a format that reveals certain characteristics. The results

showed that the extracted features could achieve more accurate predictions than the data without feature extraction. The Fourier Transformation (FT) enables a signal to be expressed with a specific characteristic; however, a severe disadvantage is that the time resolution is missing. In addition, Wavelet Transform (WT) is proposed to overcome this disadvantage, and the time series mostly uses different variations of the WT. Good local representation of a signal in both the time and frequency domain simultaneously is one of the main advantages of WT. This study proposes the Wavelet Neural Network Model (WNN) for stock price time series data using a Predictive Error Compensation WNN (PEC-WNN) model. The research is conducted based on two separately trained NNs where input data are preprocessed using a Discrete Wavelet Transform (DWT). The motivation for using two separate NNs comes from the following perspectives. Firstly, the forecasting models are facing expanding uncertainties such as a lack of input data for making more accurate predictions. Secondly, a well-known drawback in the recursive methods is the accumulation of errors since the predicted values are used in the model instead of the target values. The proposed model is independently trained and does not incline the accumulated errors. The compensation of the predicted error through the second NN enhances the overall prediction performance. This chapter does not propose trading strategies despite the evidence of asset price predictability. The prediction performances of the proposed model are presented and evaluated based on different companies in which data are publicly available using the Yahoo Finance website [82]. The remainder of this chapter is organized as follows. Section 3.2 describes the proposed model. The data set specification and construction are explained in subsection 3.3. The experimental results with discussion are given in section 3.4. Conclusion and suggested future work remarks are given in section 3.5.

3.2 Predictive Error Compensated Wavelet Neural Network Model

When machine learning methods are used for time series prediction of stochastic data sets, their time series error patterns usually include residual information. These error patterns may depend on the inconsistency of sampling time interval, scale, inappropriate machine learning structure, overfitting, underfitting, or time-varying characteristics of the data resource. A conventional way to reduce the prediction error is using past prediction errors through additional inputs or recurrences in NNs [30]. This also increases the requirement for larger training data sets and more weights to avoid overfitting and better characterization of the input data patterns. Instead of applying the error data patterns back into the same network through the increased amount of inputs and nodes, using an additional NN that is trained by the error of the first NN is proposed. Specifically, overall accuracy significantly rises when the WT of the input data patterns and the error data patterns are used in this method (PEC-WNN). At the same time, the time complexity remains less than the unified network equivalent of the solution [83]. This proposed data efficiency raising strategy can then be extended by using more NNs trained by the error pattern of the superposed prediction. Additional data can also be fused. The PEC-WNN model is shown in figure 3.1. This section explains the proposed PEC-WNN structure and demonstrates its performance in predicting the next day's closing price on the stock exchange. A critical factor in this improvement is the schematic of training the networks, as shown in figure 3.2. The main network, in the figure, represented as WNN 1, uses the closing stock price data through moving frames in single and multiple-time scales as inputs for closing price predictions. Figure 3.1 illustrates the eight consecutive values in a single time scale. An additional network, in the figure presented as WNN 2, uses the error patterns computed with the predictions of the main network (WNN 1). Finally, the predictive closing price from the WNN 1 and the predictive error from WNN 2 are used together to acquire the compensated predictive closing stock price. The proposed networks are characterized by three-layer NN architecture. The optimal number of neurons in the hidden layers is obtained based on the trial and error method considering the formulas proposed in the literature [84] - [86]. The formula proposed by Patterson

Table 3.1 : The parameter settings of Neural Network models.

Parameters:	1st Neural Network	2nd Neural Network
Number of Inputs	8	4
Number of Outputs	1	1
Number of Hidden Layers	1	1
Number of neurons in Hidden Layers	Computed using 3.1	Computed using 3.1
Learning Algorithm	Back-propagation	Back-propagation
Activation Function	ReLU in the hidden layer, Linear in the output layer	ReLU in the hidden layer, Linear in the output layer
Learning Rate	0.05	0.05
Momentum	0.75	0.75
Optimizing Algorithm	Stochastic gradient descent	Stochastic gradient descent
Training Epochs	600	300

(Equation 3.1) in [86] is used since it generates the lowest prediction error.

$$q = \frac{N}{10 \cdot (m + p)} \quad (3.1)$$

where q is the number of hidden neurons, m and p are the numbers of inputs and output, respectively, and N is the number of observations in the training data set. The employed networks use the Rectified Linear Unit (ReLU) activation function (Equation 3.2) that significantly improves the performance of the network compared to the widely used activation functions (sigmoid and hyperbolic tanged) [31].

$$f(x) = x^+ = \max(0, x) \quad (3.2)$$

The Stochastic Gradient Descent (SGD) algorithm is used as an optimization algorithm for both networks. 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. The learning rate and momentum are 0.05 and 0.75, respectively. In the Feed Forward Back-Propagation (FF-BP) model, the “generalized delta rule” is used to update the weight for each unit as follows (Equation 3.3):

$$w(t+1) = w(t) + \eta \cdot \nabla \quad (3.3)$$

where the $w(t)$ is the weight at time t , η is the learning rate, and ∇ is the gradient vector associated with the weights. The properties of the NNs used in this study are shown in table 3.1. This work investigates how the learning of networks affects the results besides different data set construction strategies for the proposed method. In traditional learning, fixed training separates the input data set into fixed training and test data sets [55]. Arnerić et al. [87] examine different ratios of separating input data

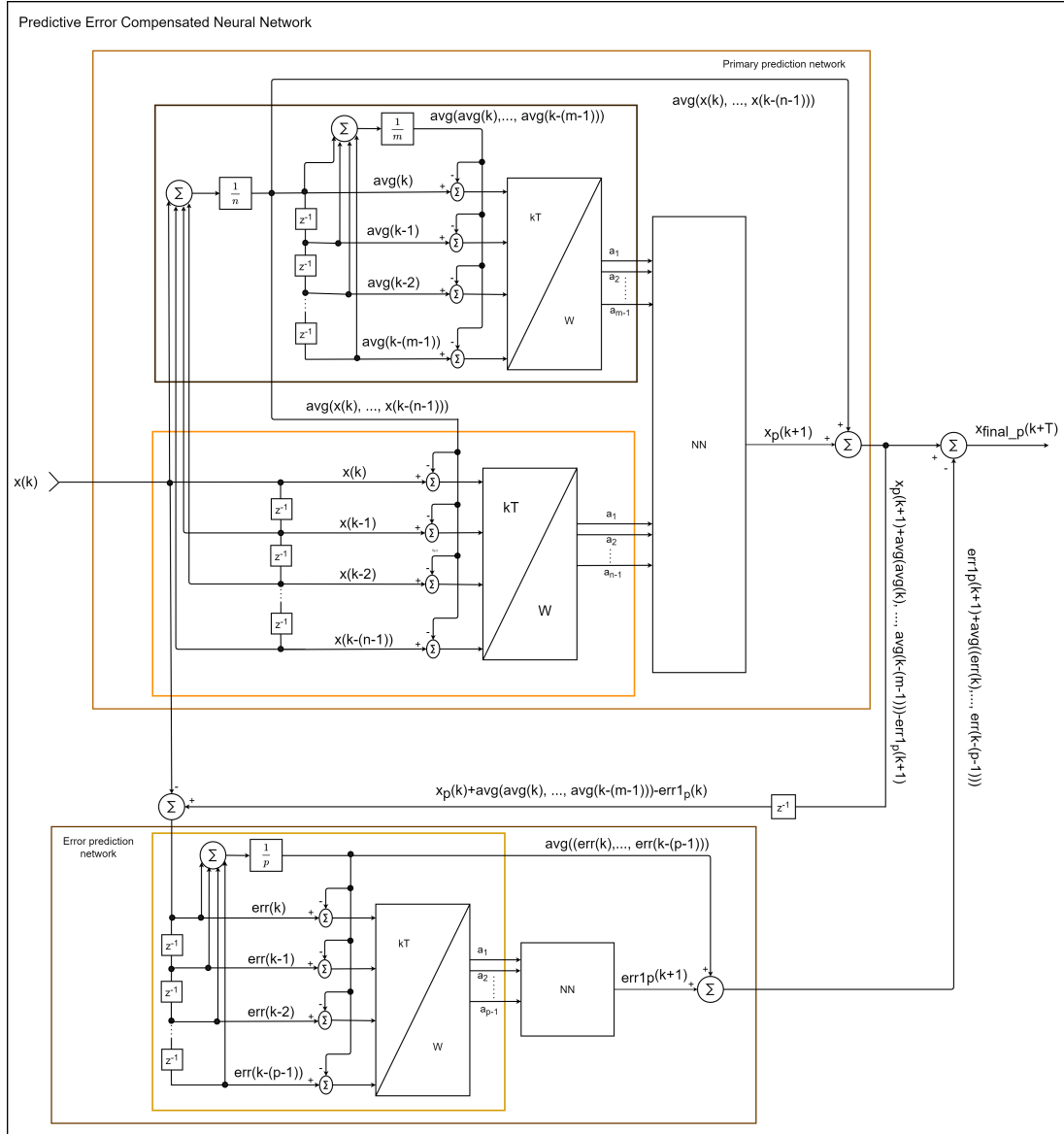


Figure 3.1 : Predictive error compensation wavelet neural network (PEC-WNN) model for stock price forecasting.

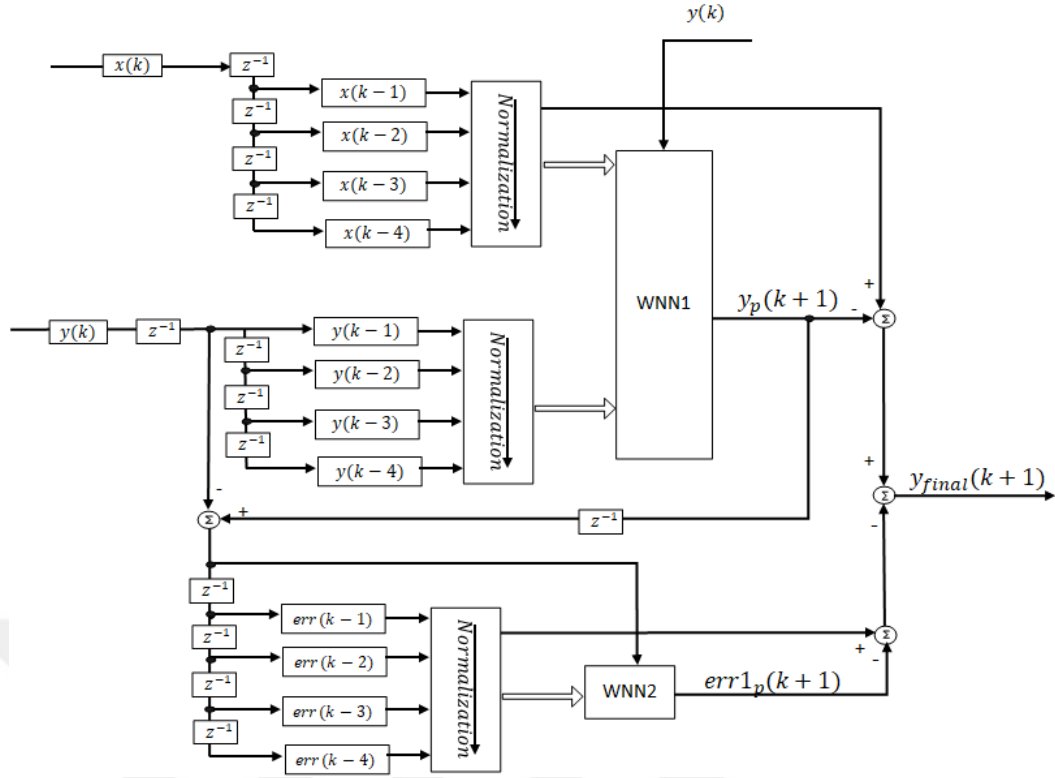


Figure 3.2 : The schematic representation of training the networks.

sets to training and test samples (90/10, 80/20, 70/30, 60/40, and 50/50) and conclude that the lowest error is achieved using a 70/30 ratio. In this study, for fixed training, the data set is divided using a 70/30 ratio (the first 70% of data is used for training and the next 30% to test the model). Continual training, on the other side, updates the training data set after a certain amount of data has been tested and retrained the network again. This experiment's initial training uses the first 70% of the data. The initial test set is equally divided where the first 15% is used for initial testing. The training data set is increased by 15% of the initial test data in the second stage. The previously constructed model is retrained with 85% of the data, and the test is done on the last 15% of the data. The scheme of the used learning algorithm is given in figure 3.3.

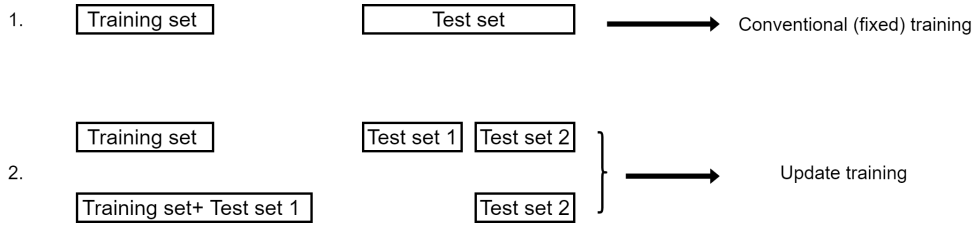


Figure 3.3 : The schematic representation of update training.

Table 3.2 : Attributes of used data set.

Feature:	Description:
Date	Corresponding Data for Stock values
Open	The opening price of a stock on a particular day
High	The highest selling stock value for a day
Low	The lowest value of the selling price of a stock on a given day
Close	Contains the closing value of a stock on a given day
Volume	The number of shares traded or bought on a given day
Adjusting Close	The closing price of a stock after paying dividends to the investors

3.3 Data

Historical stock price data for different companies are downloaded from Yahoo finance [82]. The page contains multiple stock markets of multiple companies with financial news, reports, and the facility to download historical data. The attributes of the downloaded data set are given in table 3.2. In this work, the daily closing prices are examined for the prediction of the next day's closing price and used for the construction of data sets. The selected stocks are shown in table 3.3. The daily prices of each of the stocks are collected from March 22, 2007, to April 3, 2020, which is a total of 13 years. The construction of data sets is done based on two strategies. The first is a single time window of consecutive values with different input sizes. The unit delay operator z^{-1} is used to construct the consecutive input data set. The input size varies from four to eight business days. The second strategy involves the different time windows beside the successive values. At this stage, two different time frames are included where the average values of different time intervals are applied together with four consecutive values. In order to generate sub-sampled data, separate averaging is applied in time series data, similar to the average pooling used in the Convolutional Neural Networks

Table 3.3 : The Company information and period of the data used in this study.

Name:	Symbol:	Minimum closing price:	Maximum closing price:	Historical Data Period:	
				From:	To:
Apple Inc.	AAPL	2.79\$	81.8\$	22 March 2007	3 April 2020
Micron Technology, Inc.	MU	1.69\$	62.62\$		
Bank of America Corporation	BAC	3.14\$	52.71\$		
Ford Motor Company	F	1.26\$	18.79\$		

(CNNs). The time interval for calculating the average values is five. The calculations are organized in the same manner using the unit delay operator z^{-1} . The reason for choosing five is to obtain the weekly resolution as there are five closing prices in one week. The four-weekly average values acquire the monthly resolution of input data. The constructed data sets from both strategies are normalized and preprocessed. Data normalization is a fundamental preprocessing step for mining and learning from the data [88]. Most traditional normalization methods assume that the time series data are stationary and their volatility is uniform. However, these assumptions do not hold for most time series, especially not for the financial and economic time series. In the proposed algorithm, to avoid the problems caused by traditional normalization, the average value of all current inputs is subtracted from the single value at the input. For the proposed strategy, for example, for four days, the mean value of those four days has been calculated and subtracted from each day's average value separately. The subtracted average value is added at the end of the forecasting process. The schematic representation of the proposed normalization is presented in figure 3.1. The preprocessing is done with a discrete wavelet transform (DWT) regarding the window frames. The aim of preprocessing is to extract features found within the time windows similar to the convolutional layers in CNNs. The gathered wavelet coefficients are used as inputs to the prediction model. The Haar wavelet basis function is utilized since it beneficially diminishes the distortion rate during the signal decomposition and reconstruction. Also, it significantly reduces processing and computational time. Mallat's pyramidal algorithm [89], of the second level that provides high (h_n) and low (g_n) frequencies from a given signal, is applied for decomposition. Both components are used together as inputs to the proposed model to capture valuable information during the training process.

Table 3.4 : The mathematical equations of used evaluation metrics.

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

3.4 Results and Discussion

The forecasting performances of the proposed model are evaluated using the mean absolute percentage error (MAPE), the mean absolute error (MAE), and the root mean squared error (RMSE). The MAE considers the absolute deviation as the loss. It is more sensitive to minor variations and much less sensitive to large ones than the squared error. The MAE is also scale-dependent and thus unsuitable for comparing prediction performance across different variables or ranges. An average measure of errors in predicting stock market indices represents the MAPE [90]. The average error is calculated without considering the directions of the set of predictions, and each set of differences has equal weight. The RMSE is a quadratic score principle used to determine the average magnitude of estimation error in stock market trends [90]. In addition, the RMSE depends on the scales and is sensitive to outliers. The formulas are given in table 3.4 where X_{obs} is observed and X_{model} is modeled values in time i . The number of data samples is given by n . In the following parts, the obtained results are explained and discussed. The section considers closing stock price prediction results for six different companies. The results are analyzed concerning the proposed input data strategies and applied models and by considering the training mechanism. The distinct input data sets are constructed concerning previously mentioned strategies and applied to five different models, including the simple NN model (used below as NN), PEC-NN, WNN, LSTM, and PEC-WNN. The first data set contains four, whereas the second data set consists of eight consecutive values of closing stock prices. The third data set consists of two different window frames, the four successive daily closing prices, and the four-weekly average values of the current and previous three weeks. The four weekly average values of closing prices are used since it gives the monthly resolution of price changes. The output of our model is the next day's closing

Table 3.5 : The RMSE(\$) training results for the Ford company.

Input size\Model:	NN	WNN	PEC-NN	LSTM	PEC-WNN
4 days	0.1597	0.1641	0.1541	0.1729	0.1349
8 days	0.1879	0.1731	0.16437	0.1743	0.1417
4days+4weeks	0.1632	0.1532	0.1509	0.1837	0.1201

price. The network configuration of NN models is the same for each case. The main difference is in the preprocessing where the DWT is used. The LSTM model proposed by Roberts in [22] compares the proposed model results with the model used in deep learning. The LSTM model consists of one LSTM layer with 25 hidden units and a dense output. Moreover, the dropout regularization technique is applied to the hidden layer. The prediction results for Ford (F) company are presented in tables 3.5, 3.6 and 3.7 for the training data set and tables 3.8, 3.9, and 3.10 for the test data set. The remaining results are presented as an average error with respect to the applied models in figure 3.5. The results have shown that the RMSE error is reduced when simultaneously different time frames are included with the proposed model. Increasing the number of inputs in a single window from four to eight business days raises the RMSE error and does not show any improvements. On the other hand, with multiple time windows, the forecasting error is decreased. The proposed model PEC-WNN used together with the simultaneously different window frames achieves the lowest prediction error. The RMSE error for Ford stock price reduces by 42% compared to the LSTM model when the predictive error compensation model was applied. The examination for update training is obtained regarding the lowest error results with the best data set. The second strategy is used for that purpose, simultaneously different window frames with four consecutive values. In this part, the first 70% is used for initial training, and the next 15% is applied as an initial test part. Later on, the updated training is performed by 85% of the data set (70% from initial training and +15% from the initial test), and the last 15% is used as a continual test part. When the updated training mechanism is used, the RMSE for the Ford closing prices is improved by 31.25%.

Table 3.6 : The MAPE(%) training results for the Ford company.

Input size\Model:	NN	WNN	PEC-NN	LSTM	PEC-WNN
4 days	1.4521	1.3253	1.2684	2.9742	1.2606
8 days	1.5435	1.3428	1.2635	3.8974	1.2604
4days+4weeks	1.3654	1.3245	1.2135	2.8476	1.1149

Table 3.7 : The MAE(\$) training results for the Ford company.

Input size\Model:	NN	WNN	PEC-NN	LSTM	PEC-WNN
4 days	0.2513	0.25686	0.2143	0.2043	0.1998
8 days	0.2948	0.29360	0.2278	0.2876	0.1999
4days+4weeks	0.1158	0.0985	0.10960	0.26640	0.0759

Table 3.8 : The RMSE(\$) test results for the Ford company.

Input size\Model:	NN	WNN	PEC-NN	LSTM	PEC-WNN
4 days	0.1694	0.1681	0.1691	0.2431	0.1677
8 days	0.1718	0.1677	0.1716	0.2398	0.1669
4days+4weeks	0.1612	0.1669	0.1611	0.2287	0.1609

Table 3.9 : The MAPE(%) test results for the Ford company.

Input size\Model:	NN	WNN	PEC-NN	LSTM	PEC-WNN
4 days	1.2621	1.2329	1.1243	2.3156	1.1181
8 days	1.2673	1.2352	1.1881	3.9647	1.1599
4days+4weeks	1.1283	1.1245	1.1261	2.4435	1.1149

Table 3.10 : The MAE(\$) test results for the Ford company.

Input size\Model:	NN	WNN	PEC-NN	LSTM	PEC-WNN
4 days	0.2756	0.2863	0.2135	0.2734	0.2103
8 days	0.3092	0.2997	0.2217	0.2891	0.2113
4days+4weeks	0.1218	0.0963	0.1093	0.2713	0.0867

Table 3.11 : The RMSE(\$) test results for the Apple company.

Input size\Model:	NN	WNN	PEC-NN	LSTM	PEC-WNN
4 days	1.4371	1.4658	1.3239	1.515	1.2851
8 days	1.7326	1.4941	1.2299	1.651	1.4313
4days+4weeks	0.9829	0.8635	0.8293	1.226	0.7997

Table 3.12 : The MAPE(%) test results for the Apple company.

Input size\Model:	NN	WNN	PEC-NN	LSTM	PEC-WNN
4 days	1.9376	1.843	1.7743	1.755	1.7264
8 days	2.1605	2.039	1.7667	2.028	1.6993
4days+4weeks	0.8804	0.805	0.7785	1.138	0.7501

Table 3.13 : The MAE(\$) test results for the Apple company.

Input size\Model:	NN	WNN	PEC-NN	LSTM	PEC-WNN
4 days	1.0012	1.0326	0.8944	0.978	0.8646
8 days	1.2248	1.0486	0.8318	1.045	0.8509
4days+4weeks	0.8519	0.8112	0.7396	0.906	0.7362

Table 3.14 : The RMSE(\$) test results for the Micron company.

Input size\Model:	NN	WNN	PEC-NN	LSTM	PEC-WNN
4 days	1.6242	1.6032	0.7407	1.963	0.7361
8 days	1.6463	1.5713	0.6834	2.036	0.6682
4days+4weeks	0.4315	0.453	0.3989	0.5387	0.3477

Table 3.15 : The MAPE(%) test results for the Micron company.

Input size\Model:	NN	WNN	PEC-NN	LSTM	PEC-WNN
4 days	2.1045	2.0928	2.1036	4.005	2.0921
8 days	2.2084	2.1787	2.1983	4.831	2.1781
4days+4weeks	2.1455	2.0518	2.1406	2.9641	2.0462

Table 3.16 : The MAE(\$) test results for the Micron company.

Input size\Model:	NN	WNN	PEC-NN	LSTM	PEC-WNN
4 days	1.4801	1.4783	0.6298	0.6391	0.6248
8 days	1.5043	1.4415	0.5819	0.6391	0.5571
4days+4weeks	0.4486	0.4201	0.2813	0.4198	0.2219

Table 3.17 : The RMSE (\$) test results for the Bank of America company.

Input size\Model:	NN	WNN	PEC-NN	LSTM	PEC-WNN
4 days	1.1012	0.9154	0.7266	1.012	0.7189
8 days	0.7692	0.8666	0.6701	0.934	0.6269
4days+4weeks	0.4645	0.4221	0.4595	0.533	0.3763

Table 3.18 : The MAPE (%) test results for the Bank of America company.

Input size\Model:	NN	WNN	PEC-NN	LSTM	PEC-WNN
4 days	1.192	1.1632	1.1778	3.4563	1.1578
8 days	1.310	1.2696	1.3019	2.5102	1.2685
4days+4weeks	1.192	1.1626	1.1599	1.1789	1.1147

Table 3.19 : The MAE (\$) test results for the Bank of America company.

Input size\Model:	NN	WNN	PEC-NN	LSTM	PEC-WNN
4 days	0.7323	0.7409	0.4863	1.024	0.5143
8 days	0.5657	0.5345	0.4545	1.234	0.4999
4days+4weeks	0.2648	0.2822	0.2556	0.315	0.1957

3.5 Conclusion

The stock market prediction represents a challenging but important task in analyzing the behavior of the financial market. Accurate predictions are essential for building a profitable financial market transaction strategy. Computationally less complex NN models, using simultaneously different time windows past prices, are developed to predict the next day's closing price using a predictive error compensated wavelet preprocessed model. The proposed method uses in an additive manner two separately trained NN models. The first network performs as the main predictor for the primary estimation of the next day's closing price. The second network predicts the next day's closing price error. The second network compensates for the error of the first one by subtracting the error prediction. The overall prediction performances are improved using the compensation of predicted error through the second NN. In addition to the proposed model PEC-WNN, four other models were implemented for comparison: simple NN, WNN, PEC-NN, and LSTM. The RMSE error compared to the implemented LSTM model reduces by 53.3% for Apple, 54.9% for Micron, 41.6% for Bank of America, and 42.1% for Ford stocks. The accumulation problem in error feedback neural networks is avoided through independent networks. An essential contribution of this study is the usage of the second network in an additive manner. In this respect, future research will consider adding additional networks that

process additional data to improve prediction performances. The second significant contribution of this work is the implementation of updated training. The acceptable prediction accuracy can be achieved by applying fixed training; however, the prediction performances can be improved with updated training. The limitation of this article can be seen in considering predictions by closing prices. Future studies may concern data variation patterns inside the session and a larger number of stocks and markets.





4. PREDICTIVE ERROR COMPENSATING WAVELET NEURAL NETWORK MODEL FOR MULTIVARIABLE TIME SERIES PREDICTION ¹

4.1 Introduction

Multivariate machine learning (ML) models are increasingly used for time series prediction. However, avoiding over- and underfitting in the ML-based time series prediction model requires special attention depending on the size and characteristics of the available training data set. Depending on the amount and nature of available data, the parameters used can make it challenging to build an accurate ML model. In this study, prediction performance improvement of predictive error-compensated wavelet neural networks (PECNET), an effective ML structure for predicting time series [83] - [91]. The PECNET model allows the selection of orthogonal features in data fusion applications. Different data types can be combined into the same model by extracting valuable knowledge from time series data. Methods for predicting time series depend on past observations of data that contain helpful information describing the behavior of the relevant system [83]. A single-variable time series can reconstruct a system if the embedding dimension is large enough for an automated system. Still, it cannot guarantee that any single time series variable is effective [92]. However, multivariable time series models have more dynamic information and can more effectively reflect essential aspects of the system by taking into account multivariate factors [93]. From the ML point of view, irrelevant features in prediction models impair their accuracy and ability to generalize. The presence of redundant features, which can cause overfitting, is a critical issue, especially when considering the number of samples and the number of features. It is often the case that the number of available features is extensive. However, most of the features available are redundant or irrelevant, which can negatively affect the learning process and

¹This chapter is based on the paper A. Kulagic and B. B. Ustundag, "Predictive Error Compensating Wavelet Neural Network Model for Multivariable Time Series Prediction ," in TEM JOURNAL - Technology, Education, Management, Informatics DOI : 10.18421/TEM104 – 61

model accuracy, as they can add more noise than useful information [94]. Therefore, reducing the set of features to the essential ones has various advantages: it reduces the computational costs of the learning process and improves accuracy without causing overfitting. It is generally performed either by feature extraction, selection, or combinations thereof [95]. The feature extraction approach turns the initial feature space into a lower-dimension space. On the other hand, feature selection reduces the dimensionality problem by selecting the best possible subset of the complete input feature set. Feature selection studies can be classified into the following directions: embedded method, filter-based method, and wrapper-based method [96]. The embedded, built-in method automatically selects a feature with less computational complexity; however, algorithm performance depends on the choice of the predictor. The filter-based method determines a feature based solely on the characteristics of the training set, causing poor interaction with target values and neglecting the correlation between the components. The wrapper-based method is based on an optimization and prediction model with high algorithm complexity and low computational efficiency in the feature selection process. In general, one or two additional feature components are sufficient [97]; however, the proposed PECNET model can select as many features as additional data correlate with residual error. In contrast, the selection of the best possible subset of input data in the proposed PECNET model is chosen by selecting the highest correlation data to the residual error. A previously trained network obtains a sample of the error between the measured and predicted values. Another important step on which the model's success depends, besides the input data selection and preparation, is normalization. The most commonly used normalization methods include min-max, z-score, and decimal scaling. However, these methods are unable to handle non-stationary time series data efficiently. The applicability of the previously mentioned normalization methods depends on the knowledge of their min/max values, standard deviation, and mean values. Several recent papers attempt to solve this problem by using carefully hand-crafted stationary features or employing more sophisticated normalization schemes. Although the proposed approaches overcome the problems of normalization methods and lead to better performance of used models, they face significant drawbacks as they are based on heuristically designed

normalization. In this study, normalization of average subtraction to overcome these limitations has been used, which can normalize input data concerning the average value of current input. The mean value of the current input is calculated and subtracted from each distinct input, allowing efficient handling of non-stationary data. After normalization, a discrete wavelet transform that decomposes the input data into different scales has been applied, making it useful in differentiating seasonality, detecting structural breaks and volatility clusters, and identifying the process's local and global dynamic properties over time [84]. In this study, a description of PECNET with a proposed normalization technique has been demonstrated with new insights into orthogonal feature extractions using pattern similarities and correlation analysis to the residual errors. The main idea is to find orthogonal features according to the residuals in the grid sequence and construct the network with the most correlated features to the residual error without causing overfitting or reduction of model accuracy. Overall, the main contributions of this study can be summarized as follows:

- Improvement of the prediction accuracy using multiple neural networks without causing overfitting
- The highly correlated mixed-frequency data are trained with the target prediction data, while additional networks use the highly correlated data for the residual errors
- PECNET enables 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 chapter is assembled as follows. In the next section, the proposed PECNET model for time series prediction and orthogonal feature selection is explained 4.2. The time series problems, the Lorenz Attractor time series data, and wind forecasting problem are provided in section 4.3 along with the corresponding results and discussion. Section 4.4 presents the concluding remarks.

4.2 Predictive Error Compensating Wavelet Neural Networks

The Predictive Error Compensating Wavelet Neural Network (PECNET) utilized in this study comprises four separately trained neural networks, as demonstrated in figure 4.1. The number of cascaded networks can be increased concerning the number of input parameters and data samples. Uncovering hidden patterns associated with the target variables is not trivial [97]. One of the significant issues in addressing this problem is how to deal with the existence of a high correlation between supplementary and target data. However, this study analyzes hidden patterns associated with residual errors obtained in the previously trained networks and supplementary data. 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} . Each subsequent neural network uses shifted past residual error values from the previous network. Seeking the correlation between residual error and remaining parameters in sequences of networks gives us the network with the most orthogonal features. Otherwise, increasing the number of input parameters in the same NN with the correlation between different inputs may increase the overfitting rate. The main idea is to measure the correlation between the supplementary input data and the residual error rate obtained in the previous NN. For that purpose, the Pearson correlation coefficient is used. 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:

$$r_{err_{res}Y} = \frac{\sum_{i=1}^N (err_{res(i)} - \overline{err_{res(i)}})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (err_{res(i)} - \overline{err_{res(i)}})^2 (Y_i - \bar{Y})^2}} \quad (4.1)$$

Where N is number of samples, $err_{res(i)}$ and $\bar{Y}$ are the means of residual errors (err_{res}) and additional data (Y) respectively. Additional and separately trained NNs with error data patterns of previously trained networks are used without applying errors back to the same network and increasing the number of inputs or changing network configurations. The final neural network merges outputs of neural networks in the cascaded part. Due to the problems in the traditional normalization approaches for handling non-stationary time series data, the normalization method that uses the

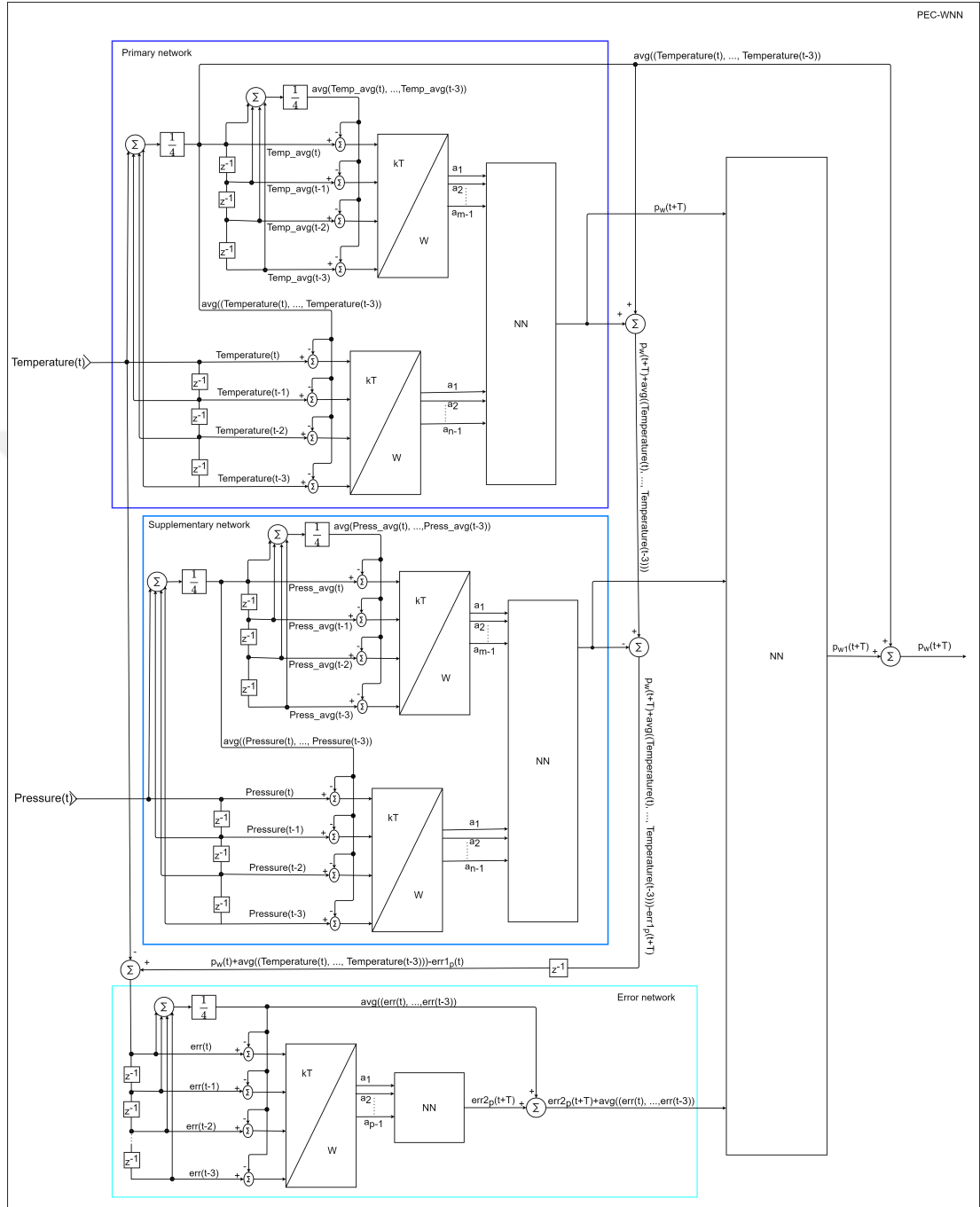


Figure 4.1 : Predictive error compensating neural network model for multivariable time series prediction. Supplementary data is selected by computing the correlation between residual error and remaining variables.

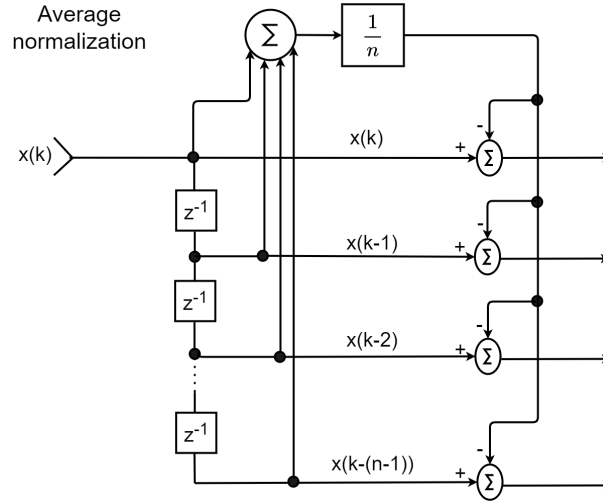


Figure 4.2 : Graphical representation of average normalization. The average value computed based on the number of inputs is subtracted from the current value.

average values of the current input to the neural network has been proposed (Figure 4.2). The idea is to calculate the mean of the network inputs and subtract the obtained value from particular input data. By this, a normalization method capable of representing different volatilities and preserving the original time series properties inside each input sequence has been introduced. 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 due to 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 [84]. The input signal is presented with $x[n]$, and the low- and high-pass filters are represented by $h[n]$ and $g[n]$, respectively. This way, a sequence of coefficients that characterizes and compacts the original signal information have been obtained. Decomposed signal $y[n]$ consists of high and low-frequency components, as shown in equation 4.2.

$$y[n] = y_{high}[n-1] + y_{low}[n-1] \quad (4.2)$$

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 4.3 and 4.4. Equation 4.3 provides an approximation, and equation 4.4 provides the detailed signal.

$$y_{high}[n-1] = \sum_n g[k] \cdot x[2n-k] \quad (4.3)$$

$$y_{low}[n-1] = \sum_n h[k] \cdot x[2n-k] \quad (4.4)$$

The Haar wavelet filters has been used since they beneficially diminish the distortion rate during the signal decomposition and reconstruction, significantly reducing the processing and computational time [83]. 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 [31], [85], [86], 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, the linear activation function ReLU (given in equation 4.5) notably improves the achievement of the feed-forward networks [31].

$$f(x) = x^+ = \max(0, x) \quad (4.5)$$

The learning rate and momentum of stochastic gradient descent (SGD) optimization algorithm are 0.05 and 0.75, respectively.

4.3 Experimental Setup and Results

An experimental setup using two different data sets to test the proposed PECNET model improvements and the use of additional data is performed. Lorenz Attractor, three-dimensional chaotic time series data, meteorological data for wind speed prediction, and different financial time series data frequencies are used for this purpose. In addition to the proposed PECNET model, a comparison between multivariate regression (MLR), neural network model (NN), wavelet preprocessed neural network model (WNN), and long-term memory model (LSTM) has been made and provided comparative results. Multivariable regression is an extension of multiple regression with one dependent variable and multiple independent variables. Long short-term memory (LSTM) is an artificial recurrent neural network architecture used in the field of deep learning. The model presented in [22] is used to compare the performance of the PECNET model. The implemented LSTM model has one hidden layer with 25 hidden units and a 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 (Equation 4.6 and 4.7), where X_{real} is accurate, and $X_{estimated}$ is estimated values in time i . The number of data samples is given by n .

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_{real,i} - X_{estimated,i})^2}{n}} \quad (4.6)$$

$$RMSPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{X_{real,i} - X_{estimated,i}}{X_{real,i}} \right| \quad (4.7)$$

4.3.1 Lorenz Attractor

The Lorenz Attractor represents a multivariable time series prediction problem consisting of three differential equations [83] (equations are given below 4.8-4.10).

$$\frac{dx}{dt} = \sigma(y - x) \quad (4.8)$$

$$\frac{dy}{dt} = x(\rho - z) - y \quad (4.9)$$

$$\frac{dz}{dt} = xy - \beta z \quad (4.10)$$

The equations are derived from the Navier-Stokes equations [83]. The parameter settings are $\sigma = 10$, $\beta = 8/3$ and $\rho = 28$, with initial conditions $[x(0), y(0), z(0)] = [0, 1, 1.05]$ as in [44]. The data set contains 10 000 multivariable data samples. A multivariable data set has been applied as input to predict the single variable as the output, as in our previous study [83]. Compared to the previous work, the main difference is applying multivariable input to the separately trained neural networks considering the orthogonality between supplementary data and residual errors. The primary network's highest correlating variable to the residual error 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. The 80% of data for training and 20% of data for validation performances have been considered. In the first setup, mixed frequencies data of single variable x are used together in one or more

Table 4.1 : The RMSE and RMSPE results for conducted PECNET experiments.

PECNET	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

neural networks. The four average values and the last successive values are used in the primary neural network (in the table marked as 1st data configuration). After the training 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. The supplementary data are added to the proposed model in the second data setup by checking the orthogonality between additional data and the residual error obtained in the previously trained network. Similar 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 the inputs for the second network are selected concerning the correlation between additional parameters and residual error. 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 PECNET as supplementary data, the obtained RMSE error is reduced to 0.25. If the same experiment with the lower correlated data, z-dimension is applied, the RMSE error is increased to 0.61, which is even higher than when only one input data is used. Selecting the additional data by checking the orthogonality between supplementary data and the residual error provides us with a lower error without causing overfitting in our model. The results for each setup are presented in table 4.1. In figure 4.4 the real and predicted values of the Lorenz Attractor test data set are presented. The PECNET 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 PECNET results. The RMSE error is higher by more than 50% (see Table 4.2.). The proposed model and its execution time are less than the implemented models.

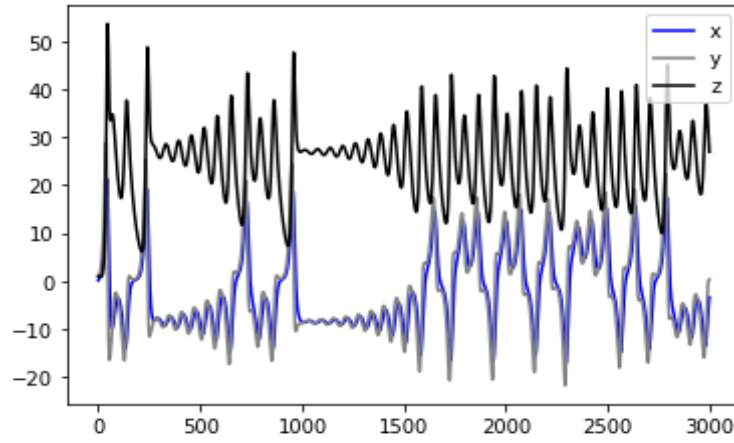


Figure 4.3 : The Lorenz Attractor trajectories projection of each time series for the first 3000 values.

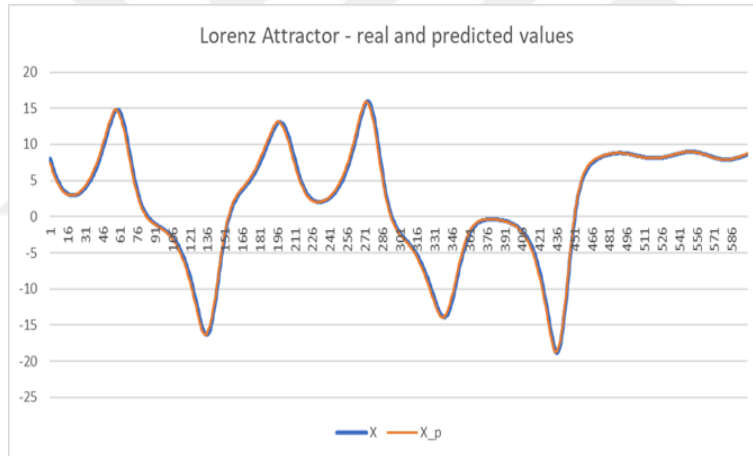


Figure 4.4 : The real and predicted values of first 600 values of Lorenz Attractor test data set.

Table 4.2 : Comparison of RMSE of PECNET to other implemented methods.

MLR	NN	WNN	LSTM	PECNET
1.768	0.822	0.816	0.618	0.246

4.3.2 Wind speed time series prediction

In this part, to forecast the wind speed in the next 30 minutes, the air temperature and vapor pressure 30 minutes data are used. The time sampling proposed in this work is considered a short-term wind speed prediction [98]. The data are collected by the Turkish Agricultural and Environmental Informatics Research and Application Center (TARBIL) monitoring station network between October 2016 and July 2017. The TARBIL stations are placed next to the agricultural fields and observed 24 hours/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 ($^{\circ}\text{C}$), air pressure (mbar), and relative humidity (%), similarly to the [99]. Ghanbarzadeh et al. [99] showed that usage of these features greatly improves the performance of wind speed forecasting models. The sampling frequency is 30 minutes. The daily values of input parameters are given in the figures below (Figure 4.5 (a), (b), (c), and (d)). The correlation between input parameters is given in figure 4.6, where it can be seen that the correlation between input parameters and wind speed is less than 0.2. Previously, wind speed forecasting was done by using physical models like numerical weather prediction [100], combining different meteorological data. Afterward, statistical models such as ARIMA were utilized to forecast wind speed. With the development of machine learning and deep learning methods, the interest in wind speed forecasting increased [101] - [102]. Zhang et al. in [103] used decomposition forecasting aggregation (DFA) and decomposition selection forecasting (DSF) with linear regression as a feature evaluator for short-term wind forecasting using a support vector machine (SVM) and NN. Feature evaluator methods and SVM and NN outperformed the single NN and SVM models. Another novel prediction model was introduced by Liu et al. [104], where de-noising auto-encoder and LSTM with mutual information feature selection are used. The results showed that feature selection reduces complexity and improves forecasting accuracy. In their study, Chen et al. in [105] predicted the n-step ahead wind speed using a denoising auto-encoder, variance, and extreme learning machine-based ensemble predictors. In the first data setup, where the mixed frequencies are applied together, the RMSE error is 1.69,

Table 4.3 : The RMSE and RMSPE results for conducted PECNET experiments.

PECNET	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 4.4 : Comparison of RMSE of PECNET to other implemented methods.

MLR	NN	WNN	LSTM	PECNET
7.34	3.19	3.14	2.11	1.32

which is less than 50% compared 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 additional parameters is -0.0153 and 0.0228 for pressure and relative humidity parameters. The highest correlating parameter to the residual error relative humidity and residual error is used as an additional parameter, and the RMSE error is reduced to 1.32. When pressure is used as an additional parameter, which is less correlated with residual error, the RMSE error increases to 1.58. The same experiments for the wind speed forecast model were conducted for the Lorentz Attractor example, which relates to the mixed frequencies, selecting additional data based on orthogonality between supplementary data and residual errors. The results and similar conclusions can be drawn by analyzing the RMSE and RMSPE errors presented in table 4.3. The RMSE comparison results of the proposed PECNET model to the implemented MLR, NN, WNN, and LSTM are shown in table 4.4. Figure 4.7 show the actual and predicted values of the wind speed forecasting test data set.

4.4 Conclusion

This study presents time series prediction performance improvements using PECNET for chaotic and stochastic data sets examples. The proposed model reduces the RMSE error by more than 60% compared to the multivariate regression models, NN, WNN, and LSTM. At the same time, adding additional data does not exponentially increase the model's complexity. Compared to the implemented LSTM model, the

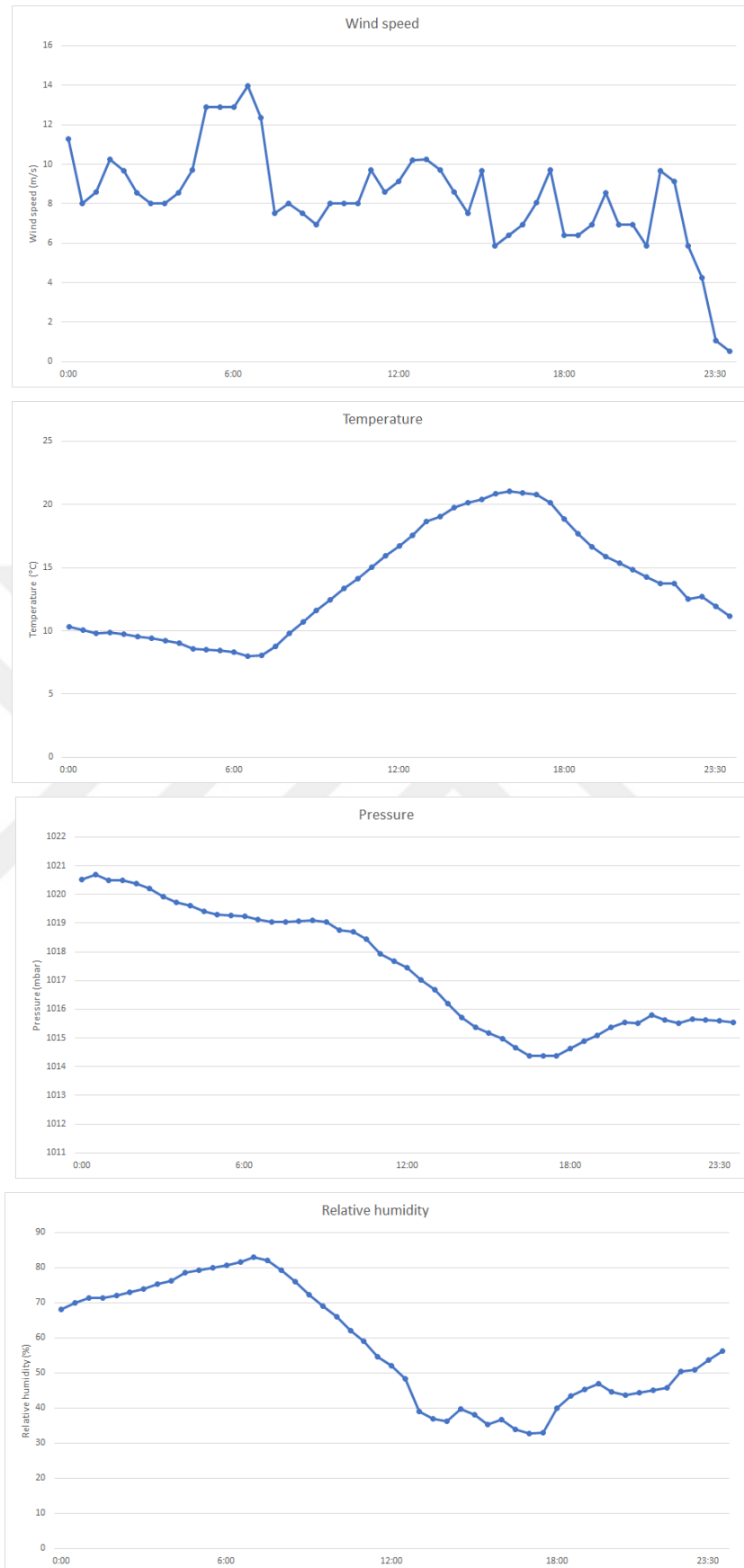


Figure 4.5 : Daily meteorological data example for 30minutes average values, (a)wind speed, (b)temperature at 10m, (c)pressure, (d)relative humidity.

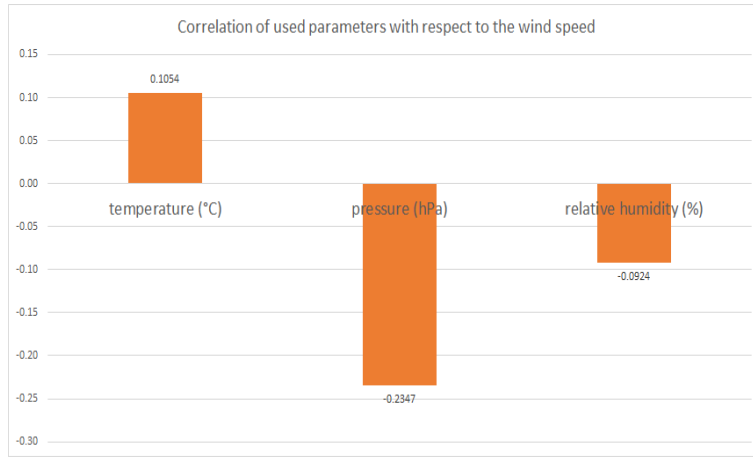


Figure 4.6 : The correlation of input parameters with respect to the wind speed.

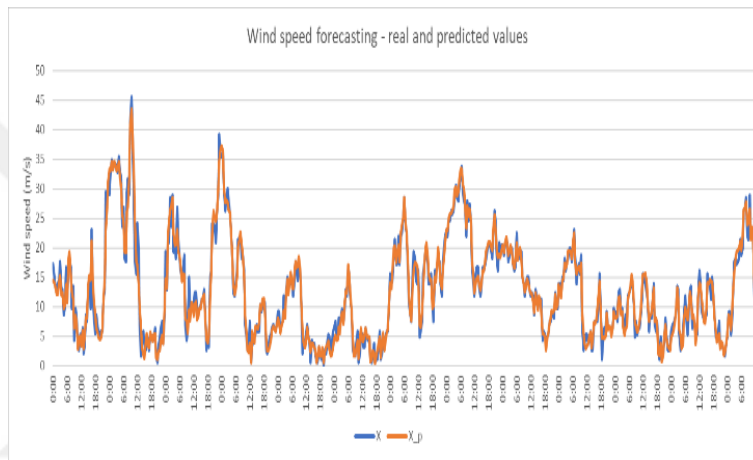


Figure 4.7 : The wind speed forecasting measured and predicted values for test data set.

complexity of the proposed PECNET is less. On the other hand, when additional data is used, if the correlation is relatively low compared to the previous parameters, which is the case in these examples, it may increase the risk of overfitting. However, selecting additional data related to the correlation of the residual error avoids the dependency between the primary data because the following network is correlated with the error pattern. PECNET provides a promising prediction model for a wide range of time series prediction problems where it is possible to impose different types of conventional networks in addition to the proposed scheme. The proposed PECNET model addresses uncertainty, dimensionality, overfitting, and non-linearity issues and improves forecasting performance.

5. PREDICTIVE ERROR COMPENSATING NEURAL NETWORKS FOR SOIL MOISTURE ESTIMATION ¹

5.1 Introduction

Soil moisture is a soil surface variable demonstrating spatial variations as a function of vegetation, land use, and topographic position [105]. Soil observations are scarce and costly and represent an important input to the profitable application of fertilizer, lime, and other soil amendments and precision irrigating. Reducing water in the soil below the minimum balance leads to a loss of yield. On the other hand, water consumption for irrigation leads to water losses. Soil surface moisture generally refers to water content in the soil layer's upper 5-15cm. Although the water in this thin layer forms a small part of global water reserves, it embodies fundamental hydrological, biochemical, physiological, agricultural, and other Earth-process [106]. This study uses the predictive error compensation neural network (PECNET) model for estimating root-zone soil moisture by fusing the satellite remote sensing and in-situ agro-meteorological monitoring data. Conventional in-situ measurements are based on single-point observations of a specific location and are not feasible for mapping the soil moisture in large-scale agricultural areas [107]. On the other hand, remote sensing data can even be used to estimate soil moisture at large and regional scales negatively affected by vegetation coverage [108]. The main idea is to fuse remote sensing and in-situ agro-meteorological data to overcome these limitations when the data are applied separately. In recent years, the potential of remote sensing for the estimation of soil moisture has been demonstrated by many researchers [106] - [113]. Iodice et al. [109], proposed a Polarimetric Two-Scale Model (PTSM), which can effectively describe depolarization and cross-polarization phenomena observed in actual data.

¹This chapter is based on the work A. Kulaglic, B.B. Ustundag, "Predictive Error Compensating Neural Networks for Soil Moisture Estimation" presented at "The Ninth International Conference on Agro-Geoinformatics, Virtual Conference (July 26-29 2021)", the presented paper is improved, submitted, and under revision in a Special Issue "Progresses in Agro-Geoinformatics", special edition of Remote Sensing (ISSN 2072-4292). This special issue belongs to the section "Remote Sensing in Agriculture and Vegetation"

MODIS and in-situ data are used by Ma et al. [113], to retrieve soil moisture using real thermal inertia remote sensing information model. Compared with the traditional soil moisture models, remote sensing data can well acknowledge comprehensive information about the earth's surface [114]. In monitoring and managing agricultural dynamics both spatially and temporally, remote sensing techniques combined with emergent machine learning (ML) algorithms play a key role. Similarly, machine learning algorithms establish to be useful tools for estimating different hydrological variables [115] - [119]. Due to their simple methodology and accurate results, regression models remain popular. However, some statistical assumptions need to be made that may lead to limited use due to outlier, non-linearity, and multicollinearity in the data [106]. Neural network models are nonlinear models commonly used in the field of prediction, resulting in more reliable predictions [116]. The proposed PECNET model consists of at least three fully connected and separately trained neural networks (NNs). Input data are normalized, and discrete wavelet transformation is applied to the input data as preliminary feature extraction. The PECNET error rate gradually decreases compared to the different machine learning models implemented in this study. Also, one of the major problems in machine learning applications, overfitting, is avoided by applying networks in a cascading manner and distributing the input data to separately trained networks. The rest of this chapter is organized as follows. Section 5.2 describes the study area and the data set. Proposed model details can be found in 5.3. The results and discussion are in Section 5.4. The conclusion and proposal of future remarks on the work are given in section 5.5.

5.2 Study Area and Data Set

5.2.1 Study area

Agricultural and Environmental Informatics Research and Application Center (TARBIL-ITU) provided the soil moisture data located in Şanlıurfa/Hilvan province, in the Southeastern Anatolia region (Figure 5.1). The area of research is the arid and-semi arid continental climate. There are two seasons within a year, whereas crop rotation and fallow season are standard practices in this area. Winter wheat is the

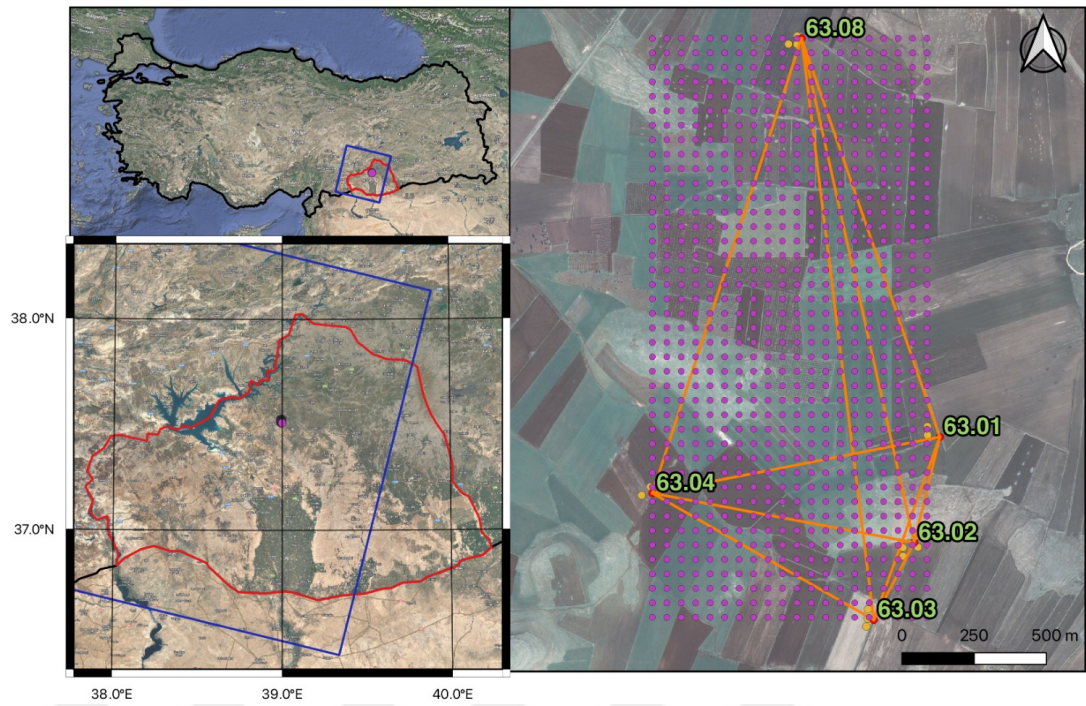


Figure 5.1 : Location of selected TARBIL monitoring stations. The red area represents the Sanliurfa/Hilvan province, the area where the stations are positioned. The blue square represents the footprint of Landsat scene 173/034. On the right side, the selected stations with the mapping points are presented.

primary cultivated crop within these parcels, mostly under rainy conditions. Barley, lentil, and corn are the secondary products. The region has its own characteristics of growing small farming and a fragmented crop pattern. Agro-meteorological stations are set up next to the agricultural fields and monitored 24 hours/day. Relevant data are stored in databases with a sampling period ranging from 10 to 30 minutes for different parameters. In-situ measurements are acquired at heights of 10m, 2m, 5cm, and depths of 15cm and 45cm. In this study, the five TARBIL stations are selected (63.01, 63.02, 63.03, 63.05, and 63.08), as shown in Figure 5.1. The center point, shown in Figure 5.1 as a red point, represents the central point relative to which the distance of the other stations is measured. Figure 5.1 (a) shows distribution of the stations, (b) NDVI image selected area, (c) distribution of spatial points in relation to which the soil map is constructed, and (d) spatial point distribution when the NDVI image is removed.

Table 5.1 : Interpretation of tensiometric soil moisture readings.

Soil Moisture (cbars)	Interpretation
0-10	Saturated soil
10-30	Soil is adequately wet
30-60	Usual range for irrigation
60-100	Usual range for irrigation in heavy clay
100-200	Soil is becoming dangerously dry for maximum production

5.2.2 Data set

In-situ soil moisture is measured every 10 minutes at 15cm and 45cm depths, corresponding to the root zone of crops. It was sampled concurrently with the satellite data, corresponding to approximately 08:00 AM with an 8-day interval. The accumulated rainfall, evapotranspiration, and soil moisture data are collected concerning the dates of remotely sensed images. Soil moisture tensiometers, measured in centibars (cbars), were used instead of volumetric soil moisture. Tensiometers measure water potential by replicating the root's behavior while extracting water from the soil. Its values vary between 0 and 200 cbar (Table 5.1). Rainfall (RF) and evapotranspiration (ET_0) data are collected. The parcel-based evapotranspiration (ET_0) is computed using the Penman-Monteith method [120], which calculates evapotranspiration based on meteorological data using daily mean temperature, relative humidity, wind speed, relative humidity, and solar radiation (Equation 5.1).

$$\lambda ET = \frac{\Delta(R_n - G) + \rho_a \cdot c_p \cdot \frac{e_s - e_a}{r_a}}{\Delta + \gamma \cdot (1 + r_s r_a)} \quad (5.1)$$

Where R_n represents net radiation, G soil heat flux, ρ_a is the mean air density at constant pressure, c_p is the specific heat in the air, $(e_s - e_a)$ vapor pressure deficit, Δ the slope of the saturation vapor pressure-temperature relationship, γ psychrometric constant, and r_s and r_a the (bulk) surface and aerodynamic resistances, respectively. Vegetation indices Normalized Vegetation Index (NDVI), Normalized Difference Moisture Index (NDMI), and Modified Soil Adjusted Vegetation Index (MSAVI) are calculated from Landsat-7 and Landsat-8 satellite imagery, with scan area 173/034 (Path/Row) downloaded as Surface Reflectance (SR) products provided by the U.S.

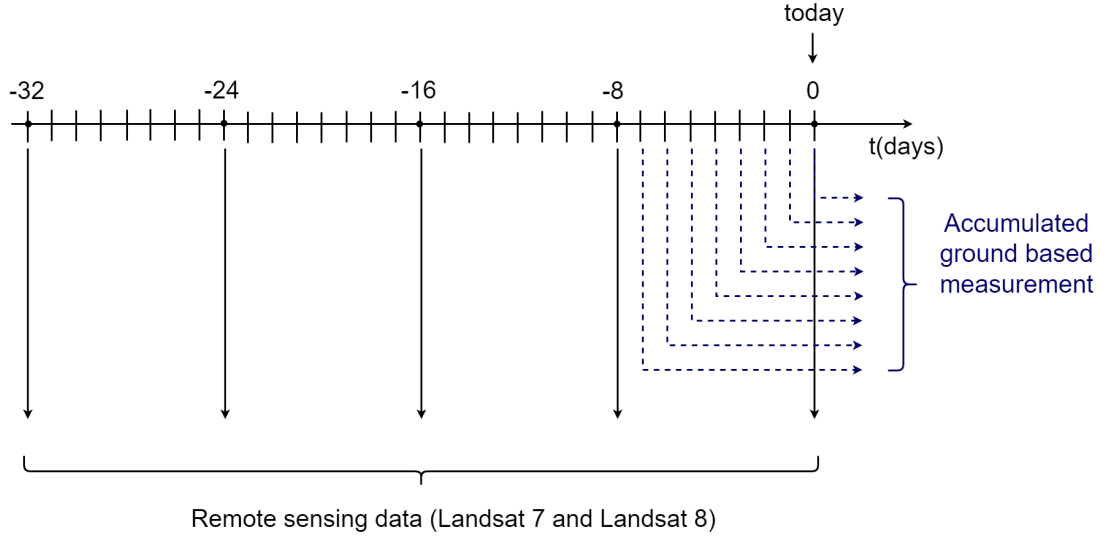


Figure 5.2 : The acquisition of spatial and temporal data.

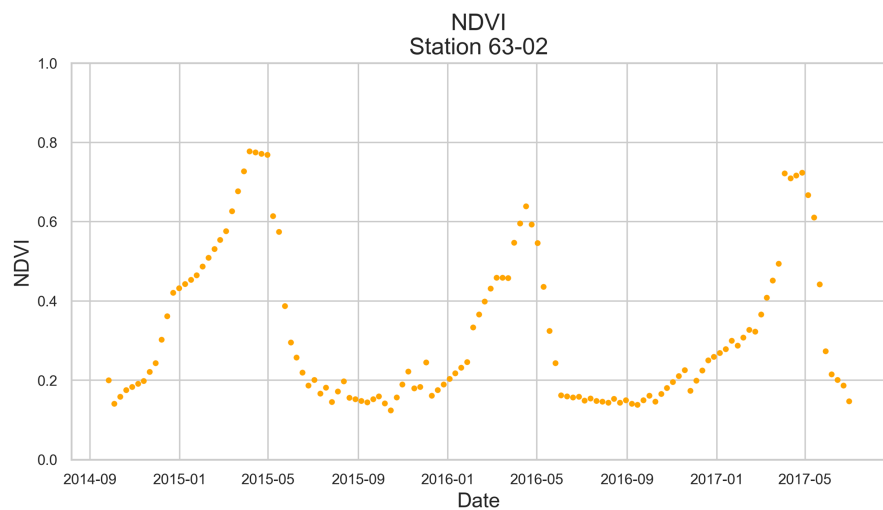
Geological Survey (USGS) [121]. Using both Landsat satellites enables an 8-day temporal resolution that detects agricultural changes occurring in relatively short time intervals [122]- [123]. The ground-based measurements are segmented with respect to the obtained remotely sensed images. Figure 5.3 represents computed indices from the Landsat imagery. In order to overcome the band interval differences, the remotely sensed images are harmonized, and Landsat 7 is harmonized with Landsat 8. The adaptive Savitzky-Golay filter is applied to smoothen the remotely sensed indices when the harmonization is done, using equations 5.2 and 5.3:

$$y' = \sum_{j=-n}^n c_j y_{i+j} \quad (5.2)$$

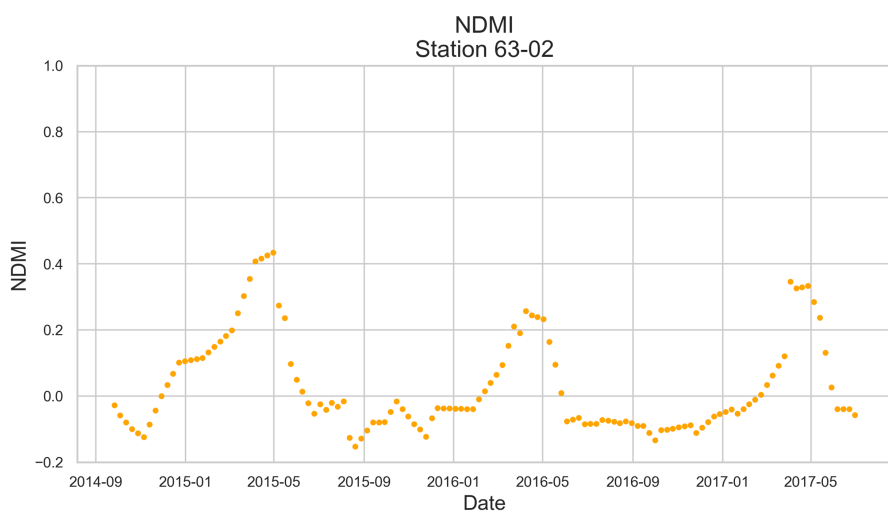
$$c_j = \frac{1}{2n+1} \quad (5.3)$$

Where c_j corresponds to the Savitzky-Golay coefficient and n is the half-width of the smoothing window. The smoothen remotely sensed indices are presented in the Figure 5.4. NDVI is the most well-known and widely used vegetation index (Equation 5.4). It quantifies vegetation greenness by measuring photosynthetically active biomass in plants. The drawbacks of NDVI are its sensitiveness to soil brightness, the impact of atmospheric effects, and saturated values in densely vegetated areas.

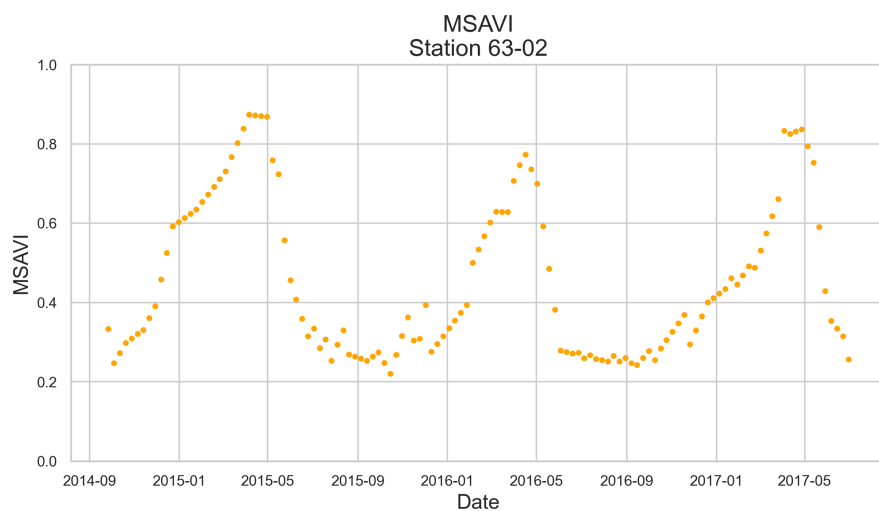
$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} \quad (5.4)$$



(a)

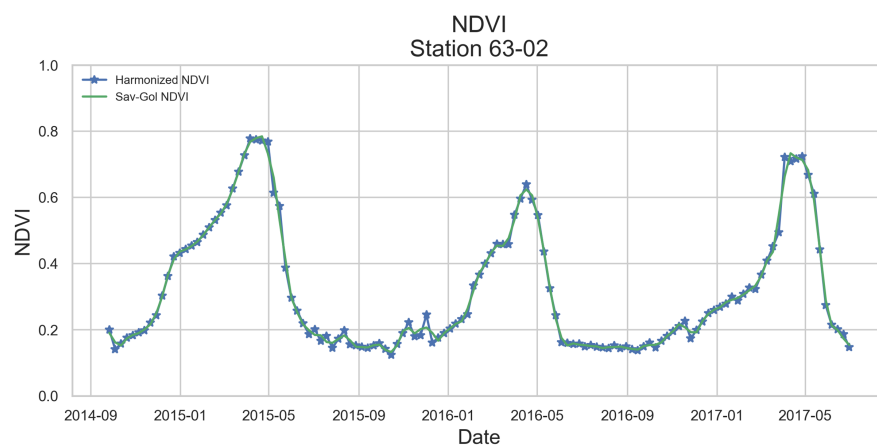


(b)

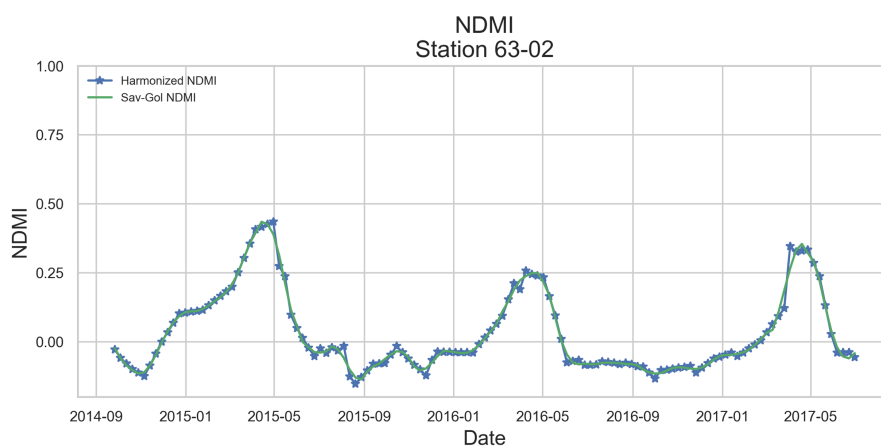


(c)

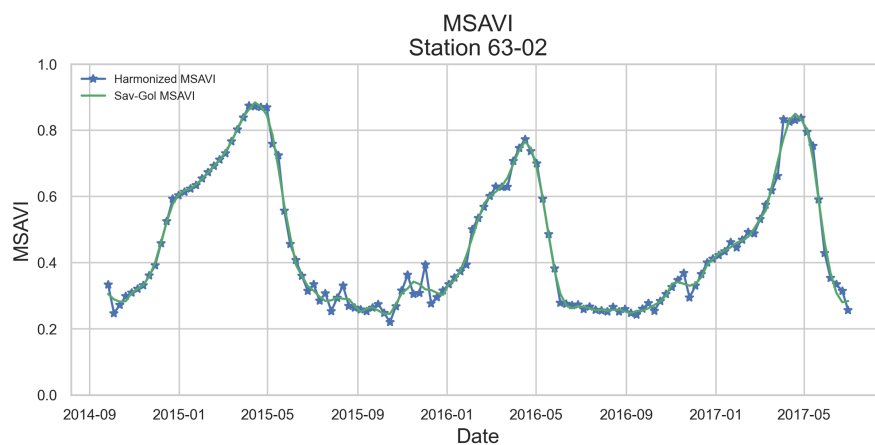
Figure 5.3 : Remotely sensed indices for the 63.02 station: (a) NDVI, (b) NDMI, and (c) MSAVI.



(a)



(b)



Window Length 99

(c)

Figure 5.4 : Remotely sensed indices for the selected station after harmonization and smoothing, 63.02 station: (a) NDVI, (b) NDMI, and (c) MSAVI.

ρ_{NIR} represents the reflectance in the near-infrared band and ρ_{RED} reflectance in the red band. The NDVI values range from -1 to 1. Extreme negative values represent water and values around zero represent bare soil. Besides the NDVI, and in order to mitigate above mentioned drawbacks, other vegetation indices such as MSAVI, and NDMI are computed. MSAVI provides higher accuracy when the soil percentage is higher than the vegetation percentage corresponding to the early phenological stages of crop development (Equation 5.5). MSAVI is computed using the near-infrared and red bands.

$$MSAVI = \frac{2 \cdot \rho_{NIR} + 1 - \sqrt{(2 \cdot \rho_{NIR} + 1)^2 - 8 \cdot (\rho_{NIR} - \rho_{RED})}}{2} \quad (5.5)$$

NDMI expresses the crop's water stress, independent of crop type, and is used for the determination of vegetation water content. It is calculated as a ratio between the near-infrared (NIR) and short-wave infrared spectral channel (SWIR) values.

$$NDMI = \frac{(NIR - SWIR)}{(NIR + SWIR)} \quad (5.6)$$

The remotely sensed indices are parcel based computed. The average of 3 points for each station is used where all points are chosen based on the homogeneity and closeness to the station in order to maintain the representation of station location. In the presented results only NDVI remotely sensed index had been used in the proposed model. However, in future studies above mentioned vegetation indices will be used.

5.3 Predictive Error Compensating Neural Network Model for Soil Moisture

The predictive error compensating neural network model (PECNET) proposed in [83] is used for nowcasting soil moisture at the root-zone using soil moisture and agro-meteorological data of reference stations together with the remotely sensed data of estimation points. The proposed model contains of at least three separately trained NNs. The highest correlated data to the target soil moisture is used as input to the primary NN. In contrast, the highest correlating parameter concerning Table 5.2 is the ET0-RF when the agro-meteorological parameter is selected. In case the reference points' soil moisture is used, the correlation to the target point is computed, and the highly correlated soil moisture data are used as input. The distances of the stations

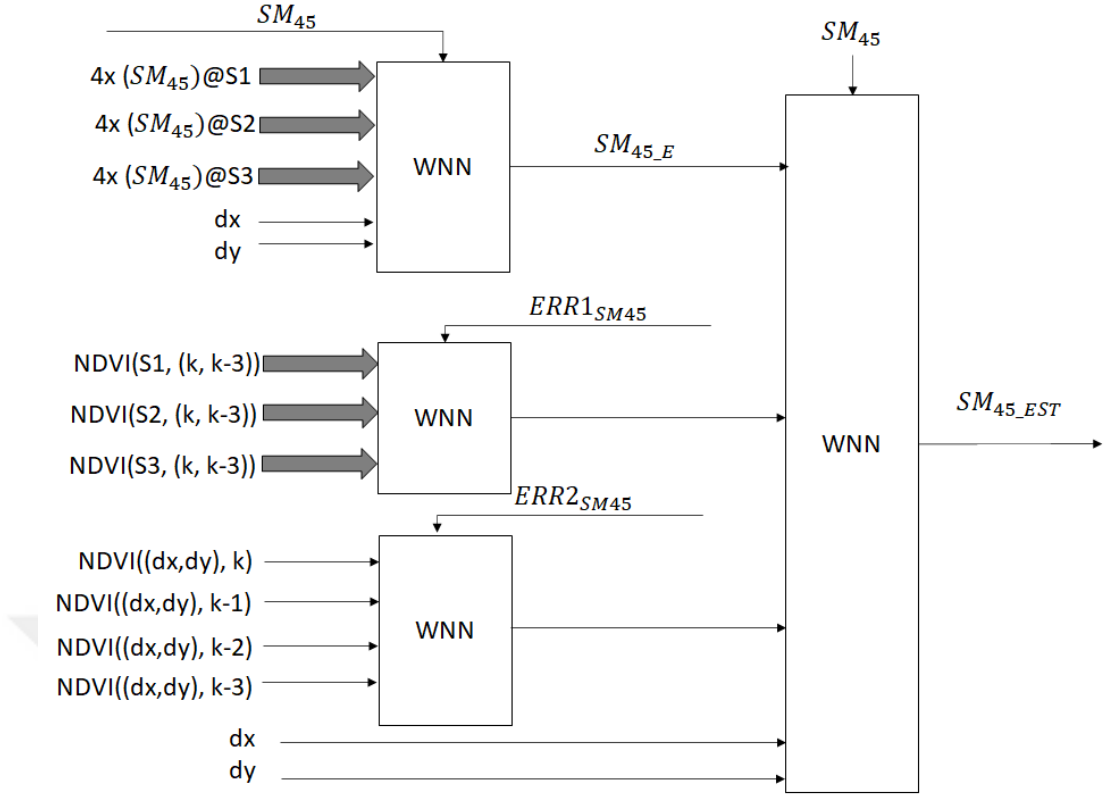


Figure 5.5 : Predictive Error Compensating Neural Network for soil moisture estimation, where soil moisture data together with the remote sensing indices are used as input.

considering the central point and the sample input data for station 63.02 are given in Tables 5.3, and 5.4. Error patterns from the main network are computed, and the correlation between residual errors and remaining input parameters is computed to obtain the network model with the most orthogonal features. Similar to the previous works [91] and [124], the input data are normalized and preprocessed using the discrete wavelet transform (DWT).

5.4 Results and Discussion

In this study, the proposed PECNET model is applied to three different cases (Figure 5.7). In the first experiment, reference soil moisture data are used to train the model with the spatial information of the estimation point. In the second experiment, besides the in-situ soil moisture data of reference stations, the remotely sensed indices of the estimation point are used as input. Firstly the soil moisture data of reference stations are applied to the PECNET. Afterward, the remote sensing indices are added

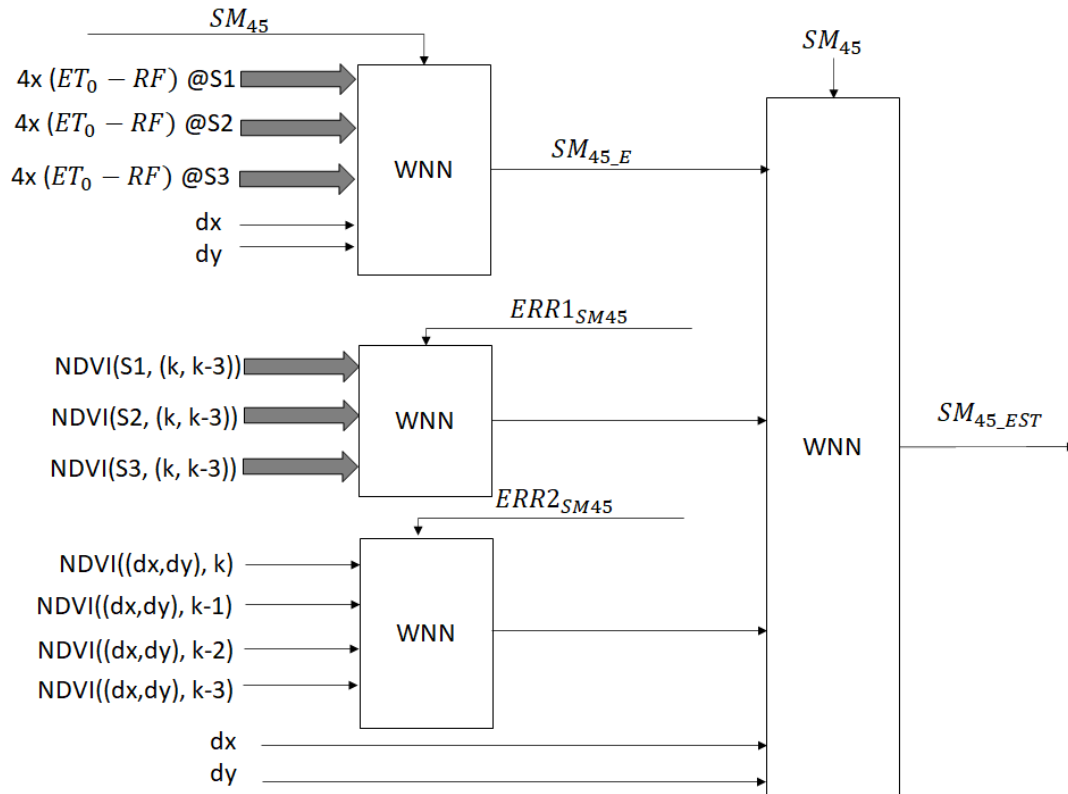


Figure 5.6 : Predictive Error Compensating Neural Network for soil moisture estimation, where agro-meteorological data (ET_0 -RF) with the remote sensing indices are used as input.

Table 5.2 : Correlation matrix of all input parameters.

	NDVI	NDMI	MSAVI	RF	ET0	RF-ET0	SM45 %	SM45
NDVI	1.000	0.974	0.992	0.203	-0.091	0.146	0.409	-0.409
NDMI	0.974	1.000	0.960	0.196	-0.027	0.092	0.346	-0.346
MSAVI	0.992	0.960	1.000	0.197	-0.095	0.147	0.399	-0.399
RF	0.203	0.196	0.197	1.000	-0.397	0.679	0.433	-0.433
ET0	-0.091	-0.027	-0.095	-0.397	1.000	-0.943	-0.583	0.583
RF-ET0	0.146	0.092	0.147	0.679	-0.943	1.000	0.623	-0.623
SM45 %	0.409	0.346	0.399	0.433	-0.583	0.623	1.000	-1.000
SM45	-0.409	-0.346	-0.399	-0.433	0.583	-0.623	-1.000	1.000

Table 5.3 : The distances between selected stations to the central point.

Station	dx(m)	dy(m)
63.01	519.21	-376.94
63.02	430.84	-741.82
63.03	290.93	-1008.76
63.04	-475	-571.03
63.08	41.09	1000.00

Table 5.4 : The sample data used as input for the selected 63.02 station.

Station No	Date	NDVI	NDMI	MSAVI	RF	ET0	ET0-RF	SM45
63-02	3/25/2017	0.5326	0.1761	0.6957	0	4.826	4.826	161
63-02	3/26/2017	0.5465	0.1862	0.7052	0	4.293	4.293	161
63-02	3/27/2017	0.5611	0.1961	0.7139	0	4.445	4.445	162.5
63-02	3/28/2017	0.5756	0.2059	0.7226	0	4.318	4.318	164.5
63-02	3/29/2017	0.5902	0.2158	0.7314	0	3.556	3.556	167.5
63-02	3/30/2017	0.6048	0.2257	0.7401	0	3.480	3.480	172
63-02	3/31/2017	0.6194	0.2356	0.7488	2.2	2.083	-0.117	107
63-02	4/1/2017	0.6340	0.2455	0.7575	17.4	2.337	-15.063	107
63-02	4/2/2017	0.6486	0.2554	0.7662	2.2	2.159	-0.041	108
63-02	4/3/2017	0.6631	0.2653	0.7750	1.4	2.769	1.369	115
63-02	4/4/2017	0.6719	0.2735	0.7818	1.2	3.683	2.483	122

as additional information with the distances of the estimation point (Figure 5.5). In-situ measurements exist as accumulated values at specific points, while remote sensing data exist at all points but only for a certain period of time. In the third case, agro-meteorological data from the reference stations and remotely sensed data are used as inputs to the PECNET for the root-zone soil moisture estimation (Figure 5.6). The input data for each of the experiments are selected, finding the highest correlating parameter to the target soil moisture and is used as input, in one of the cases, it is ET_0 -RF. The distances from the reference stations and estimation points to the central point are used in each experiment beside the agro-meteorological and remotely sensed data. It would be time-consuming to train the network for each point separately, so based on the distance of points from the estimation point and known stations is possible to predict soil moisture. The root-mean-square error (RMSE) is used to compare the experimental results. Mathematical formulations are given below (Equation 5.7), where X_{real} is real and $X_{estimated}$ is estimated values in time i . The number of data samples is given by n .

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_{real,i} - X_{estimated,i})^2}{n}} \quad (5.7)$$

In Table 5.5 RMSE (cb) errors of soil moisture at 45cm depth are presented. The graphical representation of the results are shown in Figure 5.9. The Figures (a_training) and (a_test) represent the results when only soil moisture of reference stations is used; the first experiment. (b_training) and (b_test) show results when

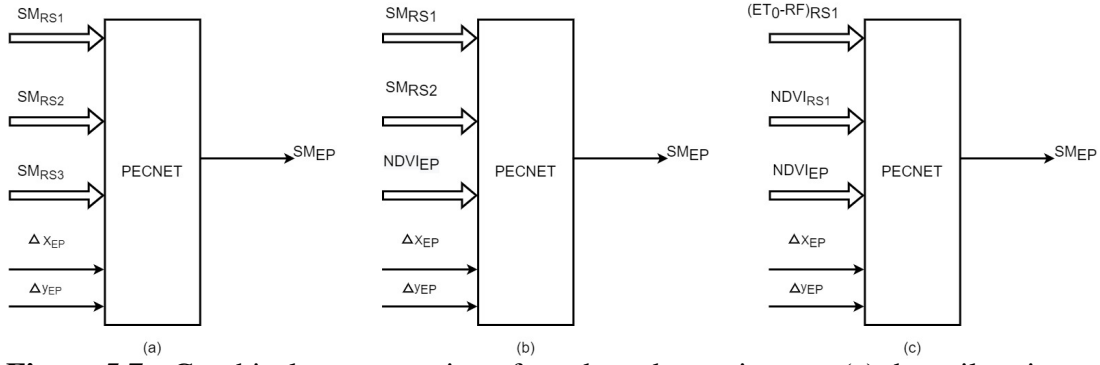


Figure 5.7 : Graphical representation of conducted experiments. (a) the soil moisture measurements are known in the reference points, the distance from the estimation point is known (x_{EP}, y_{EP}) and the model is trained with the soil moisture at estimation point; (b) soil moisture measurements are known at some reference points and remote sensing data at estimation points; and (c) agro-meteorological data from reference stations and the remote sensing data estimation points are used together with the distances for soil moisture estimation at the estimation point.

Table 5.5 : The RMSE error (cb) of soil moisture at 45cm with respect to the conducted experiments.

Input data	PECNET (RMSE(cb))	
	Training	Test
Soil moisture	18.50	20.84
Soil moisture + remote sensing	9.53	14.23
Agro-meteorological data + remote sensing	10.56	14.49

remote sensing indices are used together with the soil moisture data, and (c_training) and (c_test) represent the graphical results when agro-meteorological data are used together with remote sensing indices. Multiple NNs can be used similarly through cascaded NNs that each compensates not for the absolute prediction error but the remaining error. Therefore, orthogonal features in the additional data become more effective. Different models are implemented for the comparison performances, and the lowest RMSE error is obtained using the proposed PECNET model (Table 5.6). The implemented models with respect to the proposed PECNET model, using only agro-meteorological and remote sensing data as inputs, obtain a higher RMSE error of more than 50%. Soil moisture used as additional data together with the agro-meteorological and remote sensing data, applying to the PECNET model still gives the lowest RMSE error compared to the implemented models. The introduced model overall reduces the RMSE and surpasses different machine learning models

Table 5.6 : Test RMSE error results for soil moisture estimation using Inverse Distance Weighted Interpolation (IDWI), Multi-variable Linear Regression (MLR), Neural Networks (NN), and the proposed PECNET model.

MODEL	IDWI	MLR	NN	PECNET
RMSE(cb) Test				
Soil moisture	25	57.65	22.51	20.84
Soil moisture+remote sensing		47.76	24.07	14.23
Agro-meteorological + remote sensing		55.06	26.64	14.49

such as multi-variable linear regression and neural network models. Fusing spatial and temporal data, it is possible to obtain the map of spatio-temporal data fusion by constructing day-by-day soil moisture maps. However, their sampling rate is lower than remotely sensing data. The daily spatio-temporal soil moisture map is constructed for the period from 28.03.2017 until 04.04.2017.

5.5 Conclusion

The PECNET machine learning algorithm for the soil moisture estimation problem using fused spatial and temporal data has been applied in this work. The model has at least three separately trained NNs with the most correlated data as the primary input data. The soil moisture estimation performances at 45cm depths were analyzed. The results demonstrate that applying the proposed PECNET model yields reliable and improved prediction performance. The PECNET model provides 50% less RMSE compared to the various implemented machine learning models. The proposed model can be applied in space where soil moisture sensors are not installed, combining spatial and temporal data together. The PECNET is a promising machine learning model for temporal and spatial predictions in agricultural applications. The usage of agricultural forecasting models becomes an essential tool for predicting future events early enough, allowing appropriate decisions to be made to minimize damage and maximize productivity.

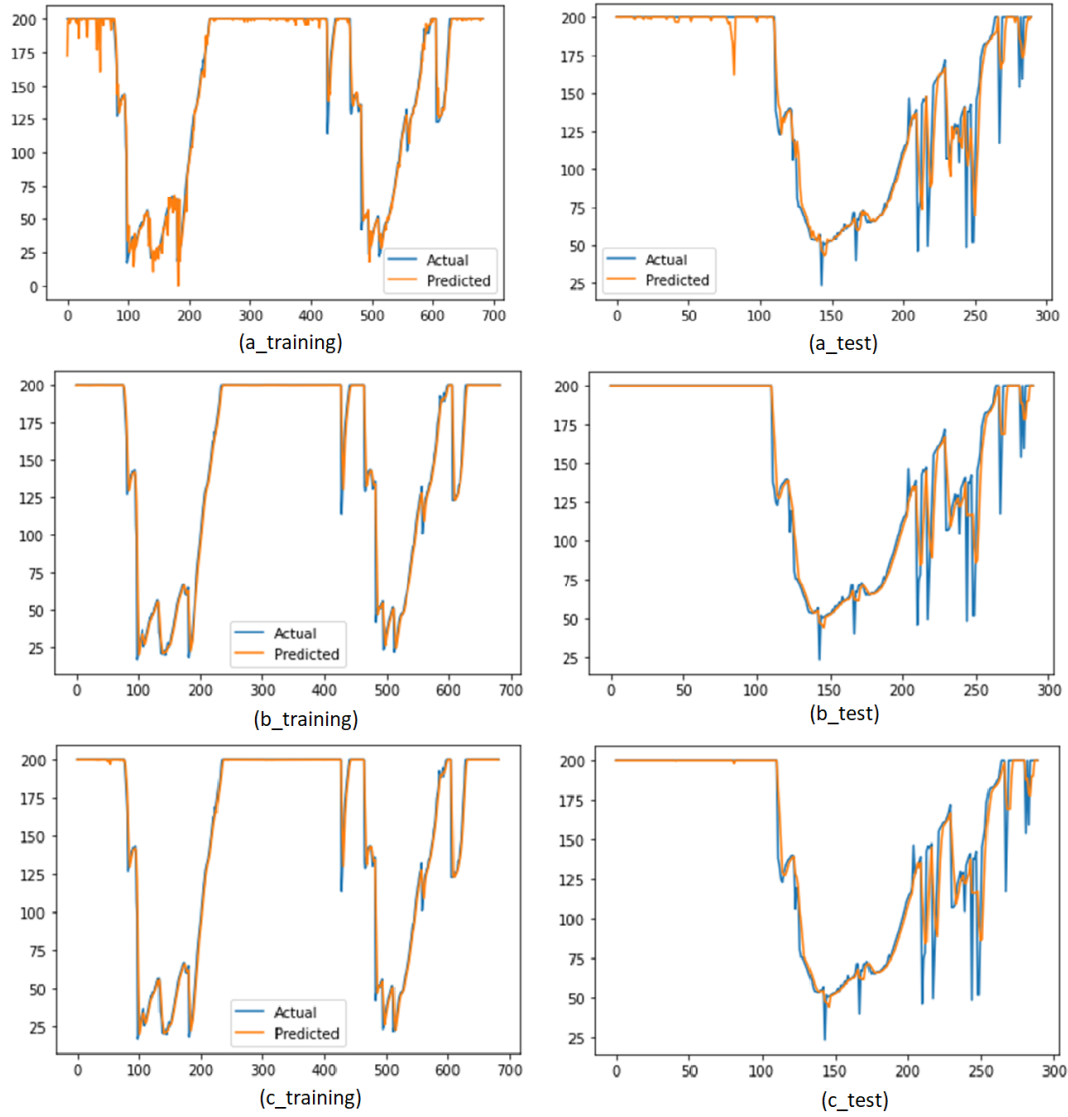


Figure 5.8 : PECNET model for spatio-temporal data fusion. (a_training) and (a_test) - The PECNET results for training and test data set when only soil moisture data is used; (b_training) and (b_test) - The PECNET results for training and test data set when soil moisture data together with remote sensing indices are used; and (c_training) and (c_test) - The PECNET results for training and test data set when ET_0 -RF data together with remote sensing indices are used.

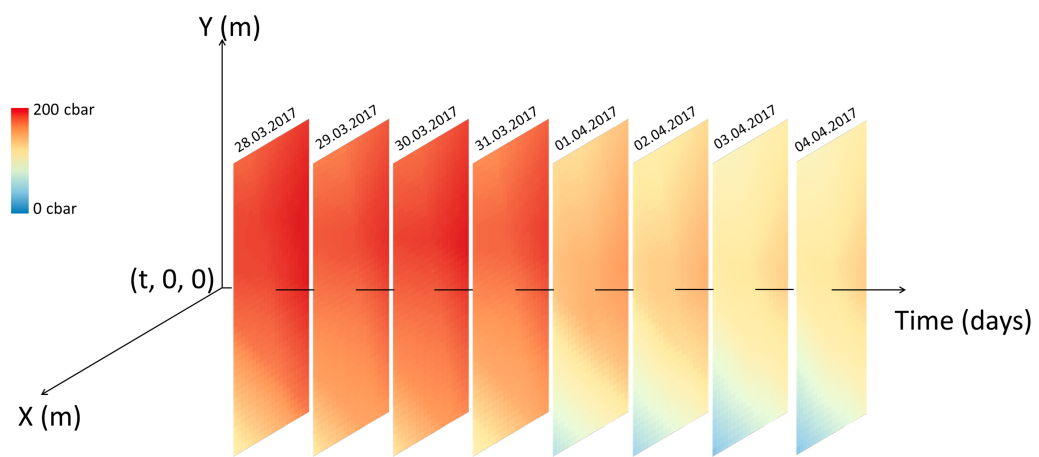


Figure 5.9 : The spatio-temporal soil moisture changes map for time period from 28.03.2017 until 04.04.2017.



6. FREQUENCY DECOMPOSITION IN PREDICTIVE ERROR COMPENSATING WAVELET NEURAL NETWORK ¹

6.1 Introduction

In the time of global financial changes, forecasting financial time series data is a significant challenge that even trading robots can hardly predict. 80% of the stock markets nowadays are controlled by machines, and according to Forbes [50], robots will replace 200,00 banking businesses in the next ten years. High-Frequency trading technologies are a type of algorithmic trading that uses machine learning algorithms to implement investment strategies in brief time intervals. Stock market prediction is essential when making the proper decision [56]. However, evaluating input data and their appropriate frequencies is critical regarding machine learning problems. For this purpose, more relevant frequencies have been determined that can be used to improve the forecasting performance without causing overfitting problems or increasing the complexity of the proposed algorithm. Stock price data are disposed to frequent changes that cannot be derived from the historical trend. Changes are influenced by real-world factors, such as political, social, and environmental factors [24]. In addition, the noise-to-signal ratio is very high in such conditions, and it is difficult to analyze and forecast future data. The use of econometric models is convenient for describing and evaluating the relationships between variables using statistical inference, with some limitations. These limitations can be seen in the inability to capture the nonlinear nature of stock prices. In addition, [61] in their study assumed constant variance while the financial time series are boisterous and have time-varying volatility. Statistical methods such as Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), and vector autoregression have generally achieved reasonable predictive results based on the results found in the literature [63] - [68]. However, according to the statistical models, Artificial Neural Networks (ANNs) are one of the most accurate prediction models [69]. According

¹This chapter represent a draft version for the journal submission

to [70], ANNs with a given sufficient amount of data can approximate any finite and continuous function based on the universal approximation theorem. The first significant study of a neural network model to predict stock price returns were made by White [51], where he introduced a prediction model based on IBM's daily common stock and achieved promising results. Various hybrid systems using ANN have been proposed to increase prediction performance, the Hidden Markov Model (HMM) [72], exponential smoothing, and ARIMA [73], and ANN with exponentially generalized autoregressive conditional heteroscedasticity model [74]. The two most popular deep-learning architectures for stock market forecasting in recent years are the Long Short-Term Memory (LSTM) model and the Gated Recurrent Unit (GRU) model with its hybridization [79]. LSTM models are appropriately structured to learn temporal patterns and overperform the conventional recurrent neural networks (RNNs). Shahi et al. [79] proposed LSTM and GRU deep-learning architectures and compared the performances of these two models for stock market prediction. In their study, they compared the performance of the LSTM and GRU models under the same conditions. Also, they showed that the predicting model could be significantly improved by including contextual information such as financial news sentiments and stock market features. Bao et al. [30] used the LSTM for stock price forecasting using different types of sequential data. Using the sentiment features, Li et al. [80] have shown that LSTMs surpass benchmark models of SVM and improve the accuracy of next-day opening price predictions. In prediction models, the problem of overfitting and getting stuck in local optima are additional issues that must be considered. The problem is due to the limited amount of data and appropriate model configuration. The financial time series data yearly obtain approximately 252 data points. However, this is insufficient for the Deep Neural Networks (DNN) models compared to the number of model parameters [31]. Sufficient data is needed as the number of model parameters increases as the number of features used is enlarged. Although NN models achieve better generalization, they are prone to overfitting due to their high capacity. Regularization techniques can prevent overfitting, but they cannot improve generalization performance. Hence, data augmentation is a method used in order to avoid overfitting while improving generalization accuracy. However, regarding

financial time series, the data augmentation distorts the original data. Consequently, in recent times, signal processing techniques have recently been used to transform data into a format that reveals certain characteristics. The results showed that the extracted features could achieve more accurate predictions than the data without feature extraction. In this study, the proposed studies in [83], [91] and [124] are extended by using different frequencies as input to the Predictive Error Compensation Neural Network (PECNET) model. In addition, the use of the Butterworth filter on the input data is tested, and the model performances are compared. The rest of this chapter is organized as follows. Chapter 6.2 describes the differences between the proposed model. The Butterworth filter is a signal processing technique with a frequency response as flat as mathematically possible in the bandwidth [125]. Experimental results with discussion are given in section 6.3. Conclusions and proposed remarks on future work are given in section 6.4.

6.2 Predictive Error Compensated Neural Network Model with Butterworth Filter

The Predictive Error Compensating Neural Network (PECNET) model presented in this section is an extension of the previously proposed model. The model utilized in this study also consists of four separately trained neural networks, as demonstrated in Figure 6.1. The main discrepancy is manifested in the first network, wherein place of the average filtering that was previously applied, the Butterworth filter (BF) is used. Butterworth filters have the sharpest roll-off possible without inducing a peak in the Bode plot, and because of that, they are called maximally flat filters [126]. The general formula for BF depends on the order of the applied filter. For continuous-time Butterworth filters, the poles associated with the squares of the frequency response magnitude are equally distributed in the angle on the circle in the s-plane, concentric with the origin and radius equal to the cut-off frequency. The poles that characterize the system function are readily obtained when the cutoff frequency and filter order are specified. Once the poles are specified, getting the differential equation characterizes the filter is straightforward. The response of the Butterworth filter is given in Equation 6.1:

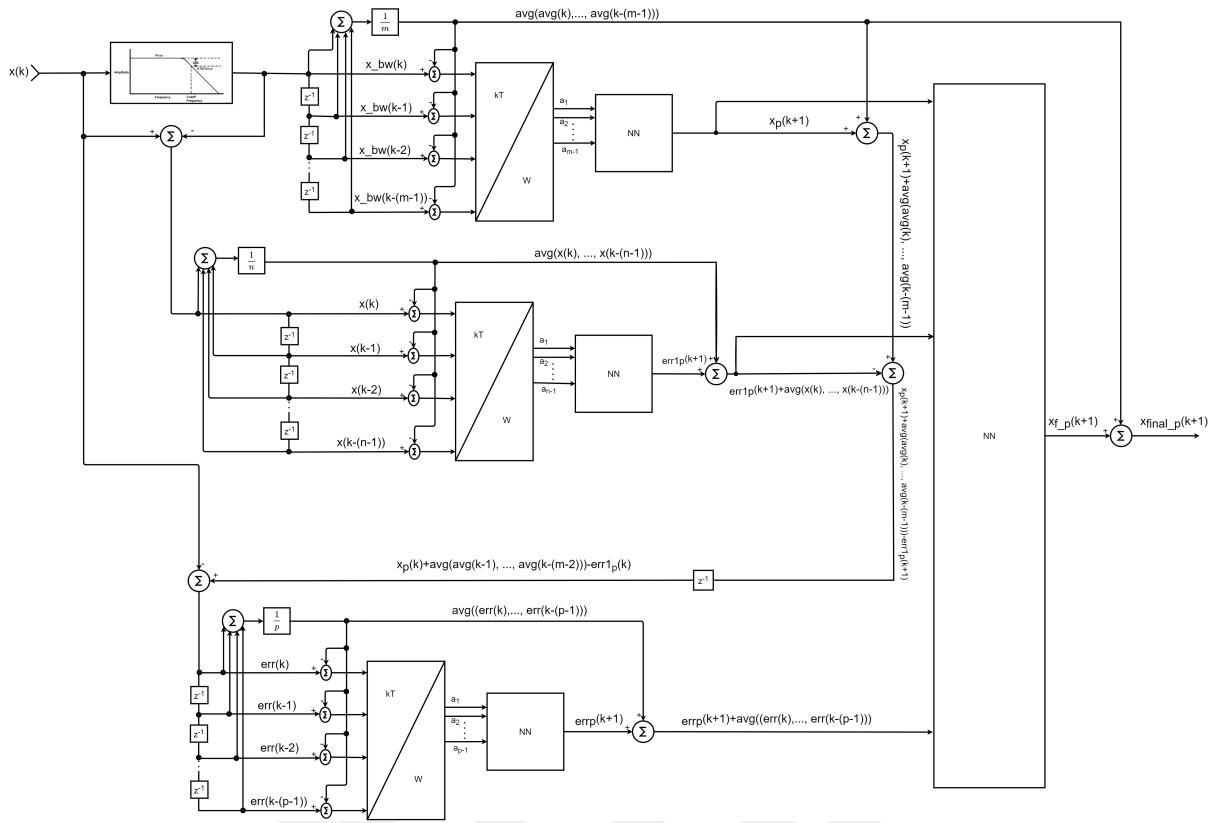


Figure 6.1 : The PECNET model with Butterworth filtering.

$$|B(j \cdot \omega)|^2 = \frac{1}{1 + \left(\frac{j \cdot \omega}{j \cdot \omega_c}\right)^{2n}} \quad (6.1)$$

Where the constant is the 68 400 sample frequency (one day is 68 400 seconds). It is easy to show that the first $2n - 1$ derivative of $B(j \cdot \omega)^2$ at ω is equal to zero ($\omega=0$). For this reason, Butterworth's response is maximally flat at $\omega=0$. Later, the current input of each parameter is shifted to the previous values using the unit time delay operator z^{-1} , as in previous studies. The error pattern obtained in the main network is applied as the output of the second network, where the difference between the primary data and the BF filtered data is used as input. Additional and separately trained NN with error data patterns of previously trained network have been used. The final neural network merges the outputs of the neural networks in a cascaded part. The average subtraction normalization used as a normalization technique due to the problems in traditional normalization approaches has been proposed and discussed in [124]. Normalization of the average subtraction allows us to build a normalization method representing differed volatilities and preserving the original time series properties inside the input

sequence. The normalized time series data are preprocessed by discrete wavelet transform (DWT), where the obtained coefficients are used as input to the NNs. The decomposed signal $y[n]$ consists of high- and low-frequency components, as shown in Equation 6.2. The input signal is presented with $x[n]$. The low- and high-pass filters are represented by $h[n]$ and $g[n]$, respectively. The Haar wavelet filters have been used as they significantly reduce distortion rate during signal decomposition and reconstruction, and also considerably reducing processing and computational time [83].

$$y[n] = y_{high}[n-1] + y_{low}[n-1] \quad (6.2)$$

The low-pass outputs recursively pass through an identical group of filter banks in order to use different resolutions in each phase. The filtering process is mathematically expressed using Equations 6.3 and 6.4. Equations 6.3 and 6.4 provide an approximation and detailed signal, respectively.

$$y_{high}[n-1] = \sum_n g[k] \cdot x[2n-k] \quad (6.3)$$

$$y_{low}[n-1] = \sum_n h[k] \cdot x[2n-k] \quad (6.4)$$

The neural network configuration consists of input, hidden, and output layers for predicting n-step-ahead time series data. Employed networks have the same network configurations. Regarding different formulas found in the literature [31], [85] and [86], the number of neurons in the hidden layers are selected using [86]. The activation function used for these networks is Rectified Linear Unit (ReLU). In comparison with sigmoid and hyperbolic tangent activation functions, the linear activation function ReLU notably improves the achievement of feed-forward networks [31]. The learning rate and momentum of the Stochastic Gradient Descent (SGD) optimization algorithm are 0.05 and 0.75, respectively. In addition to the effect of different filtering methods, the impact of frequency decomposition on the presented model has been investigated. In this regard, high, medium, and low frequencies have been applied to the financial data and ways to improve model performance without compromising model accuracy or causing overfitting have been explored. In order to obtain the spatial resolution, we also included the additional parameters to see the performance of the proposed

model. An additional network with supplementary data is added to the proposed model. In order not to increase complexity or cause the overfitting rate by applying the supplementary data to the same NN, we added new parameters into the additional network.

6.3 Model Evaluations

The experimental setup uses the stock price data obtained from the Istanbul Stock exchange, Borsa Istanbul. For this purpose, data from the banking sector and the stock exchange index are used, Is Bank, Garanti Bank, AK Bank, and BIST30 (index of 30 companies). Data were collected using web services Investing [127] for daily closing prices and Matriks [128] for hourly data. The experimental results are measured with the root-mean-square error (RMSE) and root-mean-square percentage error (RMSPE) (Equations 6.5 and 6.6). Mathematical formulations are given below, where X_{real} is accurate, and $X_{estimated}$ is estimated values in time i . The number of data samples is given by n .

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_{real,i} - X_{estimated,i})^2}{n}} \quad (6.5)$$

$$RMSPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{X_{real,i} - X_{estimated,i}}{X_{real,i}} \right| \quad (6.6)$$

The RMSPE (Table 6.1) and RMSE (Table 6.2) errors for disseminated frequencies are presented. First, the results for the average filtering used in the primary network are shown. Secondly, the results of the improvements done by applying different frequencies to the PECNET model are presented. The frequency decomposition is done using weekly, daily, and hourly data. The results are presented in Table 6.3, (RMSPE) and in Table 6.4 (RMSE). The improvements in forecasting performances can be seen by increasing the frequencies in the proposed model comparing the results where only weekly and daily data are used. If the frequencies are reduced, and the amount of information and data used increases, the model performances are significantly reduced (Tables 6.6 and 6.5). Improvements in performance results are also noticed when the Butterworth filter is applied. First, the appropriate frequencies were selected. The sampling frequency, since daily data are used was $\frac{1}{days(inseconds)}$.

Table 6.1 : The RMSPE (%) average filtering results for PECNET.

W/D	AK	IS	GARANTI	BIST30
1NN	2.420932	1.684884	1.929106	1.581213
2NN	1.884708	1.735879	1.673334	1.247089
3NN	1.831507	1.621118	1.658614	1.168715
4NN	1.615228	1.117738	1.634853	0.842569

Table 6.2 : The RMSE (TL) average filtering results for PECNET.

W/D	AK	IS	GARANTI	BIST30
1NN	0.131046	0.090225	0.176524	0.247961
2NN	0.09956	0.07617	0.240385	0.191119
3NN	0.098091	0.066546	0.242836	0.170376
4NN	0.089519	0.059838	0.150628	0.130173

One day has 86 400 seconds, so the sampling frequency is $\frac{1}{86400}$. The cut-off frequency $\frac{1}{n \cdot \text{days}(\text{inseconds})}$ where n is selected as 4. The lowest error is obtained when the ninth order of the BF filter is applied. The results indicated that applying BF filter to the input data set can further improve the model performances (6.7, 6.8). Last but not least, the spatial dimension is included in the proposed model. The spatial dimension is obtained using the different indices from stock exchanges that close before the stock market in Turkey. For this purpose, a parameter that contains information from Far Eastern stock market indices has been constructed. This idea is based on the correlation between Far East market indices and the Istanbul stock market index (BIST30). Only a few Far Eastern indices are used at this stage, such as the Nikkei index (N225), the Tokyo Stock Exchange index, Hang Seng Index (HSI), the stock market index in Hong Kong, and the Australian Securities Exchange index (AXS), and its correlation with the BIST30 index is given in Table 6.9.

Table 6.3 : The RMSPE results for increasing frequency applied to the PECNET.

W/D/H	AK	IS	GARANTI	BIST30
1NN	2.422607	1.581548	1.925032	1.587106
2NN	1.750624	1.229699	1.679927	1.266596
3NN	1.710756	1.220028	1.662537	0.960022
4NN	1.311037	1.07663	1.556721	0.860772
5NN	1.263475	1.070055	1.219931	0.797137

Table 6.4 : The RMSE results for increasing frequency applied to the PECNET.

W/D/H	AK	IS	GARANTI	BIST30
1NN	0.131534	0.08455	0.176067	0.24241
2NN	0.09478	0.065632	0.191322	0.194238
3NN	0.092534	0.065115	0.170537	0.147811
4NN	0.069717	0.057155	0.143131	0.131245
5NN	0.062936	0.057059	0.112433	0.128643

Table 6.5 : The RMSPE results for decreasing frequency applied to the PECNET.

H/D/W	AK	IS	GARANTI	BIST30
1NN	2.962316	2.228355	2.779509	2.085103
2NN	1.496557	1.489793	1.578333	0.918139
3NN	1.309792	1.438456	1.502039	0.747594
4NN	1.276866	1.368016	1.313222	0.811673
5NN	1.205566	1.238978	1.30025	0.737473

The new parameter is constructed from the average normalized values of Far Eastern indices. The constructed parameter is used as an input to the proposed model. Performance improvement is noticed by applying new parameters (Table 6.10 and Table 6.11) and together with BF filtering (Table 6.12 and Table 6.13).

6.4 Conclusion

This work introduces an improved PECNET machine learning algorithm that yields reliable and improved prediction performance for the closing stock price prediction model. The model has been enhanced by including the Butterworth filter in the proposed model. In addition, the spatial dimension has been included in the proposed model by constructing additional parameters. The constructed parameter contains the average normalized values of selected Far Eastern indices. Performance improvement of the proposed PECNET model is noticed by applying different filtering methods as well as including the additional parameters to the model.

Table 6.6 : The RMSE results for decreasing frequency applied to the PECNET.

H/D/W	AK	IS	GARANTI	BIST30
1NN	0.163362	0.121518	0.25242	0.333386
2NN	0.082705	0.080917	0.245359	0.246327
3NN	0.07862	0.077063	0.239065	0.216419
4NN	0.083928	0.074037	0.221789	0.19813
5NN	0.080243	0.067509	0.210701	0.181385

Table 6.7 : The RMSPE results for when the Butterworth filter is applied to the input signal.

D	AK	IS	GARANTI	BIST30
1NN	3.716282	2.55976	3.189103	2.760657
2NN	2.607084	2.077328	2.519306	2.009513
3NN	1.541605	1.02821	1.576386	0.91528
4NN	1.520222	0.998691	1.57281	0.774124

Table 6.8 : The RMSE results for when the Butterworth filter is applied to the input signal.

D	AK	IS	GARANTI	BIST30
1NN	0.202757	0.137869	0.286422	0.436493
2NN	0.144097	0.113562	0.229782	0.318334
3NN	0.078491	0.05527	0.146178	0.200641
4NN	0.072492	0.053673	0.143009	0.140798

Table 6.9 : The correlation between Far-eastern stock indices to the BIST30 index.

	N225	HIS	ASX	BIST30
N225	1			
HIS	0.525851	1		
ASX	0.779896	0.404533	1	
BIST30	0.861042	0.652458	0.718566	1

Table 6.10 : The RMSPE results when the Far-eastern index is used.

Far-eastern index	AK	IS	GARANTI	BIST30
1NN	2.523778	1.563303	1.743394	1.535997
2NN	2.102215	1.272339	1.935211	1.089865
3NN	2.024797	1.271858	1.88175	1.265789
4NN	1.577906	1.101763	1.699738	0.993911
5NN	1.369125	1.088723	1.361959	0.802536

Table 6.11 : The RMSE results when the Far-eastern index is used.

<i>Far-eastern index</i>	<i>AK</i>	<i>IS</i>	<i>GARANTI</i>	<i>BIST30</i>
1NN	0.139754	0.084788	0.161343	0.252468
2NN	0.097526	0.067915	0.181596	0.155434
3NN	0.102322	0.072752	0.195592	0.16116
4NN	0.070979	0.058526	0.163982	0.158479
5NN	0.069592	0.054038	0.155418	0.147533

Table 6.12 : The RMSPE results when the Far-eastern index is used with the Butterworth filter.

<i>Far-eastern index+BW</i>	<i>AK</i>	<i>IS</i>	<i>GARANTI</i>	<i>BIST30</i>
1NN	3.556565	2.436843	3.092161	2.686676
2NN	2.68649	2.299159	2.397482	2.276949
3NN	2.599607	2.501083	2.322024	2.262921
4NN	1.362777	0.83993	2.16978	0.900213
5NN	1.281084	0.71907	1.359017	0.762704

Table 6.13 : The RMSE results when the Far-eastern index is used with the Butterworth filter.

<i>Far-eastern index+BW</i>	<i>AK</i>	<i>IS</i>	<i>GARANTI</i>	<i>BIST30</i>
1NN	0.195383	0.131386	0.277857	0.426849
2NN	0.14941	0.125319	0.216807	0.363872
3NN	0.144235	0.135966	0.237084	0.361765
4NN	0.068886	0.044471	0.197342	0.139435
5NN	0.060181	0.038127	0.168986	0.117781

7. CONCLUSIONS

This dissertation proposes and describes the spatio-temporal data fusion model PECNET that applies wavelet preprocessed neural networks. An algorithm that merges spatial and temporal time series data together has been proposed, as well as experimental evaluations with its improvements. The robustness of the proposed algorithm has been evaluated concerning different methods implemented and proposed in the literature. The essentials of the PECNET implementation, properties, advantages, and disadvantages were surveyed and studied. The core development of PECNET was done through the Python programming language. The results have highlighted some aspects of the PECNET model. The major key summaries among this research work are: The proposed model achieves more accurate predictions and shows less sensitivity to hyperparameters and structural settings for a broad range of time series spatio-temporal forecasting and nowcasting problems. The PECNET model minimizes one of the main problems in machine learning, overfitting. In addition, the proposed model architecture is less complex. The average subtraction normalization overcomes the traditional normalization problems characteristic of nonstationary time series data. Selecting additional input features by checking the correlation to the residual error of previously trained networks and reducing the frequency of input data as more networks are added reduces the computational cost of the learning process and improves the accuracy without causing overfitting. Finally, PECNET has better error performance over traditional machine learning models found in the literature.

7.1 Practical Application of This Study

Bearing in mind that the proposed PECNET structure can be applied to various spatio-temporal time series problems, this dissertation may have several key applications. Accordingly, the work done in this dissertation can be seen that different stochastic and chaotic time series data can be applied to the proposed model and give promising results. As with temporal data, spatio-temporal time series forecasting and

nowcasting applications can benefit from the PECNET model. The PECNET checks the correlation between input parameters to obtain the orthogonal features used in the proposed model. The provided cascaded network architecture prevails sensitivity to estimation errors by separate training networks. Different types of temporal and spatio-temporal data and sampling rates have been applied to the proposed PECNET structure. According to the results of our work, using different sampling frequencies of the same data type, or applying additional spatial data in the proposed model by selecting the highest correlation to the residual error, helps us to achieve better results without increasing the exponential complexity of the model.

7.2 Future Research

The possibilities of future research concerning the proposed PECNET model for spatio-temporal time series forecasting are both wide. Hereby, the possible improvements for any future work in this field are pointed out.

- In the dissertation, an efficient machine learning model for spatio-temporal data fusion has been developed. In order to meet the timeline and dissertation scope has been on designing the PECNET model. Although the proposed model can be improved by adding additional hyperparameters and changing the current structure of the PECNET model, the main issue is that the hyperparameter selection and learning process will consume more time. The problem of building an accurate ML model with the proper structure and hyper-parameters can also be achieved by selecting different types of deep learning models for the proposed PECNET structure, such as convolutional neural networks and long short-term memory networks, etc.
- Furthermore, supplementary networks can be added concerning the orthogonality between the data and residuals of previously trained networks. The residual errors' orthogonality always increases with respect to the chosen data selection method. On the other hand, the highest correlating orthogonal feature can be selected through the additional layer. In this work, a wavelet transform as feature selection has been used, and in the future, alternatively, the convolutional filters can be applied to

choose the correct features. The main target here was using a sample filter through residual error and how to merge them for spatial and temporal data fusion.

- Motivated by recent advances and the remarkable success of different normalization techniques for non-stationary time series prediction models can lead to performance improvements of PECNET.
- Different spatial and temporal data types can be used together by increasing the proposed model with supplementary networks. Also, different filtering methods, such as the Butterworth filter, can be applied on the input signal or in the place of the proposed wavelet transform, and the PECNET model can be improved.





REFERENCES

- [1] **Klein, L.A.** (2004). *Sensor and data fusion: a tool for information assessment and decision making*, SPIE press.
- [2] **Castenedo, F.** (2013). A review of data fusion technique, *The Scientific World Journal*.
- [3] **Mamoulis, N.** (2009). *Temporal Data Mining*, Springer US, Boston, MA, pp.2948–2952.
- [4] **Shapiro, L.G.** (1980). Design of a spatial information system, 101–117.
- [5] **Xiong, G., Zhu, F., Dong, X., Fan, H., Hu, B., Kong, Q., Kang, W. and Teng, T.** (2016). A kind of novel ITS based on space-air-ground big-data, *IEEE intelligent transportation systems magazine*, 8(1), 10–22.
- [6] **Hessah, A. and Rongjun, Q.** (2020). Spatiotemporal fusion in remote sensing.
- [7] **Legendre, P. and Gauthier, O.** (2014). Statistical methods for temporal and space-time analysis of community composition data., *Proceedings of the Royal Society B: Biological Sciences*, 281(1778), 20132728, <https://hal.univ-brest.fr/hal-00948693>.
- [8] **Grunwald, S.** (2009). Multi-criteria characterization of recent digital soil mapping and modeling approaches, *Geoderma*, 152(3-4), 195–207.
- [9] **Lagacherie, P. and McBratney, A.** (2006). Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping, *Developments in soil science*, 31, 3–22.
- [10] **Raicharoen, T., Lursinsap, C. and Sanguanbhokai, P.** (2003). Application of critical support vector machine to time series prediction, 5, V–V.
- [11] **Deb, C., Zhang, F., Yang, J., Lee, S.E. and Shah, K.W.** (2017). A review on time series forecasting techniques for building energy consumption, *Renewable and Sustainable Energy Reviews*, 74, 902–924.
- [12] **Zhang, G.P.** (2003). Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing*, 50, 159–175.
- [13] **Sorjamaa, A., Hao, J., Reyhani, N., Ji, Y. and Lendasse, A.** (2007). Methodology for long-term prediction of time series, *Neurocomputing*, 70(16-18), 2861–2869.

- [14] **Waheeb, W., Ghazali, R. and Herawan, T.** (2016). Ridge polynomial neural network with error feedback for time series forecasting, *PloS one*, 11(12), e0167248.
- [15] **Ma, Q.L., Zheng, Q.L., Peng, H., Zhong, T.W. and Xu, L.Q.** (2007). Chaotic time series prediction based on evolving recurrent neural networks, *2007 international conference on machine learning and cybernetics*, volume 6, IEEE, pp.3496–3500.
- [16] **Haykin, S.** (2010). *Neural Networks and Learning Machines*, 3/E, Pearson Education India.
- [17] **Chen, Y., Yang, B., Dong, J. and Abraham, A.** (2005). Time-series forecasting using flexible neural tree model, *Information sciences*, 174(3-4), 219–235.
- [18] **Waheeb, W. and Ghazali, R.** (2019). 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.
- [19] **López-Caraballo, C.H., Salfate, I., Lazzús, J.A., Rojas, P., Rivera, M. and Palma-Chilla, L.** (2016). Mackey-Glass noisy chaotic time series prediction by a swarm-optimized neural network, *Journal of Physics: Conference Series*, 720, 012002, <https://doi.org/10.1088/1742-6596/720/1/012002>.
- [20] **Pal, S.S. and Kar, S.** (2019). 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.
- [21] **Loia, V., Tomasiello, S., Vaccaro, A. and Gao, J.** (2020). Using local learning with fuzzy transform: application to short term forecasting problems, *Fuzzy Optimization and Decision Making*, 19(1), 13–32.
- [22] **Roberts, D.** (2019). Lorenz Trajectories Prediction: Travel Through Time, *arXiv preprint arXiv:1903.07768*.
- [23] **Oord, A.v.d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A. and Kavukcuoglu, K.** (2016). Wavenet: A generative model for raw audio, *arXiv preprint arXiv:1609.03499*.
- [24] **Gorenc Novak, M. and Velušček, D.** (2016). Prediction of stock price movement based on daily high prices, *Quantitative Finance*, 16(5), 793–826.
- [25] **Lecun, Y. and Bengio, Y.** (1995). *Convolutional networks for images, speech, and time-series*, MIT Press.
- [26] **Cai, X., Zhang, N., Venayagamoorthy, G.K. and Wunsch II, D.C.** (2007). Time series prediction with recurrent neural networks trained by a hybrid PSO–EA algorithm, *Neurocomputing*, 70(13-15), 2342–2353.

- [27] **Bontempi, G., Ben Taieb, S. and Le borgne, Y.A.** (2013). Machine learning strategies for time series forecasting, *Lecture Notes in Business Information Processing*, Springer, pp.62–77, 2nd European Business Intelligence Summer School, eBISS 2012 ; Conference date: 15-07-2012 Through 21-07-2012.
- [28] **Reis, A.R. and Da Silva, A.A.** (2005). Feature extraction via multiresolution analysis for short-term load forecasting, *IEEE Transactions on power systems*, 20(1), 189–198.
- [29] **Alzaq, H.Y. and Ustundag, B.B.** (2018). An optimized two-level discrete wavelet implementation using residue number system, *EURASIP Journal on Advances in Signal Processing*, 2018(1), 41.
- [30] **Bao, W., Yue, J. and Rao, Y.** (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory, *PloS one*, 12(7), e0180944.
- [31] **Goodfellow, I., Bengio, Y. and Courville, A.** (2016). *Deep learning*, MIT press.
- [32] **Meng, K., Dong, Z. and Wong, K.** (2009). Self-adaptive radial basis function neural network for short-term electricity price forecasting, *IET generation, transmission & distribution*, 3(4), 325–335.
- [33] **Chen, Y., Yang, B. and Dong, J.** (2006). Time-series prediction using a local linear wavelet neural network, *Neurocomputing*, 69(4-6), 449–465.
- [34] **Gers, F.A., Eck, D. and Schmidhuber, J.** (2002). Applying LSTM to Time Series Predictable Through Time-Window Approaches, **R. Tagliaferri and M. Marinaro, editors**, *Neural Nets WIRN Vietri-01*, Springer London, London, pp.193–200.
- [35] **Chudy, L. and Farkas, I.** (1998). Prediction of chaotic time-series using dynamic cell structures and local linear models, *Neural Network World*, 8(5), 481–489.
- [36] **Day, S.P. and Davenport, M.R.** (1993). Continuous-time temporal back-propagation with adaptable time delays, *IEEE Transactions on neural networks*, 4(2), 348–354.
- [37] **Crowder III, R.S.** (1991). Predicting the Mackey-Glass timeseries with cascade-correlation learning, 117–123.
- [38] **Box, G.E.P. and Jenkins, G.** (1990). *Time Series Analysis, Forecasting and Control*, Holden-Day, Inc., USA.
- [39] **Tong, R.** (1980). The evaluation of fuzzy models derived from experimental data, *Fuzzy sets and systems*, 4(1), 1–12.
- [40] **Xu, C.W. and Lu, Y.Z.** (1987). Fuzzy model identification and self-learning for dynamic systems, *IEEE Transactions on Systems, Man, and Cybernetics*, 17(4), 683–689.

- [41] **Sugeno, M. and Yasukawa, T.** (1991). Linguistic modelling based on numerical data, *Proceedings of the IFSA'91*, 264–267.
- [42] **Surmann, H., Kanstein, A. and Goser, K.** (1993). Self-Organizing and Genetic Algorithms for an Automatic Design of Fuzzy Control and Decision Systems, *In Proc. EUFIT'93*, pp.1097–1104.
- [43] **Gao, Y. and Er, M.J.** (2005). NARMAX time series model prediction: feedforward and recurrent fuzzy neural network approaches, *Fuzzy sets and systems*, 150(2), 331–350.
- [44] **Lorenz, E.N.** (1963). Deterministic nonperiodic flow, *Journal of the atmospheric sciences*, 20(2), 130–141.
- [45] **Xiu, Y. and Zhang, W.** (2017). Multivariate Chaotic Time Series Prediction Based on NARX Neural Networks, *2017 2nd International Conference on Electrical, Automation and Mechanical Engineering (EAME 2017)*, Atlantis Press.
- [46] **Elsner, J.B. and Tsonis, A.A.** (1992). Nonlinear Prediction, Chaos, and Noise, *Bulletin of the American Meteorological Society*, 73(1), 49–60, [https://doi.org/10.1175/1520-0477\(1992\)073<0049:NPCAN>2.0.CO;2](https://doi.org/10.1175/1520-0477(1992)073<0049:NPCAN>2.0.CO;2), [https://doi.org/10.1175/1520-0477\(1992\)073<0049:NPCAN>2.0.CO;2](https://doi.org/10.1175/1520-0477(1992)073<0049:NPCAN>2.0.CO;2).
- [47] **Beguiría, S., Vicente-Serrano, S.M. and Angulo-Martínez, M.** (2010). 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, <https://doi.org/10.1175/2010BAMS2988.1>, https://journals.ametsoc.org/bams/article-pdf/91/10/1351/3737355/2010bams2988_1.pdf.
- [48] **Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., Angulo, M. and El Kenawy, A.** (2010). 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, <https://doi.org/10.1175/2010JHM1224.1>, https://journals.ametsoc.org/jhm/article-pdf/11/4/1033/4107385/2010jhm1224_1.pdf.
- [49] **Ahmed, N.K., Atiya, A.F., Gayar, N.E. and El-Shishiny, H.** (2010). An empirical comparison of machine learning models for time series forecasting, *Econometric Reviews*, 29(5-6), 594–621.
- [50] **URL-1.** <<https://www.forbes.com/sites/bethkindig/2020/04/10/new-age-of-stock-market-volatility-driven-by-machines>; "New age of stock market volatility driven by machines"; Retrieved date 09.10.2020.

- [51] **White, H.** (1988). Economic prediction using neural networks: The case of IBM daily stock returns, 2, 451–458.
- [52] **Fernando, De-Losso, R. and Giovannetti, B.** Day Trading for a Living?, <http://dx.doi.org/10.2139/ssrn.3423101>, <https://ssrn.com/abstract=3423101>.
- [53] **Barber, B.M., Lee, Y.T., Liu, Y.J. and Odean, T.** (2014). The cross-section of speculator skill: Evidence from day trading, *Journal of Financial Markets*, 18, 1–24, <https://www.sciencedirect.com/science/article/pii/S1386418113000190>.
- [54] **Jordan, D.J. and Diltz, J.D.** (2003). The Profitability of Day Traders, *Financial Analysts Journal*, 59(6), 85–94, <https://doi.org/10.2469/faj.v59.n6.2578>, <https://doi.org/10.2469/faj.v59.n6.2578>.
- [55] **Henrique, B.M., Sobreiro, V.A. and Kimura, H.** (2018). Stock price prediction using support vector regression on daily and up to the minute prices, *The Journal of Finance and Data Science*, 4(3), 183–201, <https://www.sciencedirect.com/science/article/pii/S2405918818300060>.
- [56] **Fama, E.F. and French, K.R.** (1993). Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics*, 33(1), 3–56, <https://www.sciencedirect.com/science/article/pii/03044405X93900235>.
- [57] **Fama, E.F.** (2021). Efficient capital markets a review of theory and empirical work, *The Fama Portfolio*, 76–121.
- [58] **Bachelier, L.** (1900). Théorie de la spéculation, *Annales scientifiques de l'École Normale Supérieure*, 3e série, 17, 21–86, <http://www.numdam.org/articles/10.24033/asens.476/>.
- [59] **Cootner, P.H.** (1964). *The random character of stock market prices.*, Cambridge, Mass.: M.I.T. Press, [1964].
- [60] **Fama, E.F.** (1965). The behavior of stock-market prices, *The journal of Business*, 38(1), 34–105.
- [61] **Abu-Mostafa, Yaser S. and Atiya, A.F.** (1996). Introduction to financial forecasting, *Applied Intelligence*, 6, 205–213.
- [62] **Bollen, J., Mao, H. and Zeng, X.** (2011). Twitter mood predicts the stock market, *Journal of Computational Science*, 2(1), 1–8, <https://www.sciencedirect.com/science/article/pii/S187775031100007X>.
- [63] **Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M.** (2015). *Time series analysis: forecasting and control*, John Wiley & Sons.

- [64] **Wang, J.H. and Leu, J.Y.** (1996). Stock market trend prediction using ARIMA-based neural networks, *4*, 2160–2165.
- [65] **Meyler, A., Kenny, G. and Quinn, T.** (1998). Forecasting Irish inflation using ARIMA models.
- [66] **Nochai, R. and Nochai, T.** (2006). ARIMA Model Forecasting Oil Palm Price.
- [67] **Reddy, C.** (2019). Predicting the Stock Market Index using Stochastic Time Series ARIMA Modelling: The Sample of BSE and NSE, *Indian Journal of Finance*, *13*, 7–25.
- [68] **Ariyo, A.A., Adewumi, A.O. and Ayo, C.K.** (2014). Stock Price Prediction Using the ARIMA Model, *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, pp.106–112.
- [69] **Khashei, M. and Bijari, M.** (2010). An artificial neural network (p, d, q) model for timeseries forecasting, *Expert Syst. Appl.*, *37*, 479–489.
- [70] **Hornik, K., Stinchcombe, M. and White, H.** (1989). Multilayer feed-forward networks are universal approximators, *Neural Networks*, *2*(5), 359–366, <https://www.sciencedirect.com/science/article/pii/0893608089900208>.
- [71] **Lendasse, A., Bodt, E., Wertz, V. and Verleysen, M.** (2002). Non-linear financial time series forecasting - Application to the Bel 20 stock market index, <http://dx.doi.org/10.1051/ejess:2000110>, *14*.
- [72] **Hassan, M.R., Nath, B. and Kirley, M.** (2007). A fusion model of HMM, ANN and GA for stock market forecasting, *Expert Systems with Applications*, *33*(1), 171–180, <https://www.sciencedirect.com/science/article/pii/S0957417406001291>.
- [73] **Wang, J.J., Wang, J.Z., Zhang, Z.G. and Guo, S.P.** (2012). Stock index forecasting based on a hybrid model, *Omega*, *40*(6), 758–766, <https://www.sciencedirect.com/science/article/pii/S0305048311001435>, special Issue on Forecasting in Management Science.
- [74] **Hajizadeh, E., Seifi, A., Zarandi, M.F. and Turksen, I.** (2012). A hybrid modeling approach for forecasting the volatility of S&P 500 index return, *Expert Systems with Applications*, *39*(1), 431–436.
- [75] **Yao, J., Tan, C.L. and Poh, H.I.** (1999). Neural Networks For Technical Analysis: A study on KLCI, *International Journal of Theoretical and Applied Finance*, *2*.
- [76] **Adebiyi, A.A., Adewumi, A.O. and Ayo, C.K.** (2014). Comparison of ARIMA and artificial neural networks models for stock price prediction, *Journal of Applied Mathematics*, *2014*.

- [77] **Merh, N., Saxena, V. and Pardasani, K.** (2010). A comparison between Hybrid Approaches of ANN and ARIMA for Indian Stock Trend Forecasting, *Business Intelligence Journal*, 3, 23–44.
- [78] **Yoo, S.** (2007). Neural Network Model vs. SARIMA Model In Forecasting Korean Stock Price Index (KOSPI), *Issues in Information Systems*, (3).
- [79] **Shahi, T.B., Shrestha, A., Neupane, A. and Guo, W.** (2020). Stock Price Forecasting with Deep Learning: A Comparative Study, *Mathematics*, 8(9), 1441.
- [80] **Li, J., Bu, H. and Wu, J.** (2017). Sentiment-aware stock market prediction: A deep learning method, *2017 international conference on service systems and service management*, IEEE, pp.1–6.
- [81] **Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.** (2014). Dropout: a simple way to prevent neural networks from overfitting, *The journal of machine learning research*, 15(1), 1929–1958.
- [82] **URL-2.** <<https://finance.yahoo.com>>; Yahoo Finance; Retrieved date 13.10.2020.
- [83] **Ustundag, B.B. and Kulaglic, A.** (2020). High-Performance Time Series Prediction With Predictive Error Compensated Wavelet Neural Networks, *IEEE Access*, 8, 210532–210541.
- [84] **Hwarng, H.B.** (2001). Insights into neural-network forecasting of time series corresponding to ARMA (p, q) structures, *Omega*, 29(3), 273–289.
- [85] **Moshiri, S. and Cameron, N.** (2000). Neural network versus econometric models in forecasting inflation, *Journal of forecasting*, 19(3), 201–217.
- [86] **Patterson, D.W.** (1996). *Artificial Neural Networks: Theory and Applications*, Prentice Hall.
- [87] **Arnerić, J., Poklepović, T. and Aljinović, Z.** (2014). GARCH based artificial neural networks in forecasting conditional variance of stock returns, *Croatian Operational Research Review*, 329–343.
- [88] **Ogasawara, E., Martinez, L.C., De Oliveira, D., Zimbrão, G., Pappa, G.L. and Mattoso, M.** (2010). Adaptive normalization: A novel data normalization approach for non-stationary time series, *The 2010 International Joint Conference on Neural Networks (IJCNN)*, IEEE, pp.1–8.
- [89] **Mallat, S.G.** (2009). *A theory for multiresolution signal decomposition: the wavelet representation*, Princeton University Press.
- [90] **Willmott, C.J. and Matsuura, K.** (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, *Climate research*, 30(1), 79–82.

- [91] **Kulaglic, A. and Ustundag, B.B.** (2021). Stock Price Prediction Using Predictive Error Compensation Wavelet Neural Networks, *Computers, Materials & Continua*, 68(3), 3577–3593, <http://www.techscience.com/cmc/v68n3/42451>.
- [92] **Zeng, Y., Yan, E., Li, C. and Li, Y.** (2008). Application of multivariable time series based on RBF neural network in prediction of landslide displacement, *2008 The 9th International Conference for Young Computer Scientists*, IEEE, pp.2707–2712.
- [93] **Wang, X. and Han, M.** (2014). Multivariate time series prediction based on multiple kernel extreme learning machine, *2014 International joint conference on neural networks (IJCNN)*, IEEE, pp.198–201.
- [94] **Kotsiantis, S.** (2011). Feature selection for machine learning classification problems: a recent overview, *Artificial Intelligence Review*, 42(1), 157–176.
- [95] **Pudil, P. and Somol, P.** (2005). Current feature selection techniques in statistical pattern recognition, *Computer Recognition Systems*, Springer, pp.53–68.
- [96] **Niu, T., Wang, J., Lu, H., Yang, W. and 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.
- [97] **Lee, G. and Lee, K.** (2021). Feature selection using distributions of orthogonal PLS regression vectors in spectral data, *BioData Mining*, 14(1), 1–16.
- [98] **Dhiman, H.S. and Deb, D.** (2020). A Review of Wind Speed and Wind Power Forecasting Techniques, *arXiv preprint arXiv:2009.02279*.
- [99] **Ghanbarzadeh, A., Noghrehabadi, A., Behrang, M. and Assareh, E.** (2009). Wind speed prediction based on simple meteorological data using artificial neural network, *2009 7th IEEE International Conference on Industrial Informatics*, IEEE, pp.664–667.
- [100] **Senthil Kumar, P. and Lopez, D.** (2015). Feature selection used for wind speed forecasting with data driven approaches, *Journal of Engineering Science and Technology Review*, 8(5), 124–127.
- [101] **Swaroop, G., Senthil Kumar, P. and Muthamil Selvan, T.** (2014). An efficient model for share market prediction using data mining techniques, *International Journal of Applied Engineering Research*, 9(17), 3807–3812.
- [102] **Liu, H., Mi, X. and Li, Y.** (2018). Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network, *Energy Conversion and Management*, 166, 120–131.

- [103] **Zhang, C., Wei, H., Zhao, J., Liu, T., Zhu, T. and Zhang, K.** (2016). Short-term wind speed forecasting using empirical mode decomposition and feature selection, *Renewable Energy*, 96, 727–737.
- [104] **Liu, X., Zhang, H., Kong, X. and Lee, K.Y.** (2020). Wind speed forecasting using deep neural network with feature selection, *Neurocomputing*, 397, 393–403.
- [105] **Flores, A., Istanbuluoglu, E., Bras, R. and Entekhabi, D.** (2006). *A framework for the prediction of soil moisture*, World Scientific.
- [106] **Adab, H., Morbidelli, R., Saltalippi, C., Moradian, M. and Ghalhari, G.A.F.** (2020). Machine learning to estimate surface soil moisture from remote sensing data, *Water*, 12(11), 3223.
- [107] **Bogena, H.R., Huisman, J.A., Oberdörster, C. and Vereecken, H.** (2007). Evaluation of a low-cost soil water content sensor for wireless network applications, *Journal of Hydrology*, 344(1-2), 32–42.
- [108] **Aktas, A. and Üstündağ, B.B.** (2020). Soil Moisture Monitoring of the Plant Root Zone by Using Phenology as Context in Remote Sensing, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 6051–6063.
- [109] **Iodice, A., Natale, A. and Riccio, D.** (2013). Polarimetric two-scale model for soil moisture retrieval via dual-pol HH-VV SAR data, *IEEE Journal of selected topics in applied earth observations and remote sensing*, 6(3), 1163–1171.
- [110] **Baghdadi, N., Zribi, M., Loumagne, C., Ansart, P. and Anguela, T.P.** (2008). Analysis of TerraSAR-X data and their sensitivity to soil surface parameters over bare agricultural fields, *Remote sensing of environment*, 112(12), 4370–4379.
- [111] **Guo, P., Shi, J., Liu, Q. and Du, J.** (2013). A new algorithm for soil moisture retrieval with L-band radiometer, *IEEE Journal of selected topics in applied earth observations and remote sensing*, 6(3), 1147–1155.
- [112] **Brown, M.E. and Escobar, V.M.** (2013). Assessment of soil moisture data requirements by the potential SMAP data user community: Review of SMAP mission user community, *IEEE Journal of selected topics in applied earth observations and remote sensing*, 7(1), 277–283.
- [113] **Ma, C., Han, X. and Wang, W.** (2012). Soilmoisture retrieval using thermal inertia method in Heihe River Basin, China, *2012 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, pp.694–697.
- [114] **Zhang, F., Zhang, L.W., Jing-Jing, S. and Huang, J.F.** (2014). Soil moisture monitoring based on land surface temperature-vegetation index space derived from MODIS data, *Pedosphere*, 24(4), 450–460.

- [115] **Partal, T. and Cigizoglu, H.K.** (2009). Prediction of daily precipitation using wavelet—neural networks, *Hydrological sciences journal*, 54(2), 234–246.
- [116] **Sarmadian, F. and Taghizadeh Mehrjardi, R.** (2008). Modeling of some soil properties using artificial neural network and multivariate regression in Gorgan province, north of Iran, *Global Journal of Environmental Research*, 2(1), 30–35.
- [117] **Tokar, A.S. and Markus, M.** (2000). Precipitation-runoff modeling using artificial neural networks and conceptual models, *Journal of Hydrologic Engineering*, 5(2), 156–161.
- [118] **Al-saffar, R., Khattab, S. et al.** (2013). Prediction of soil's compaction parameter using artificial neural network, *Al-Rafidain Engineering Journal (AREJ)*, 21(3), 15–27.
- [119] **Kulaglic, A., Berk Üstündağ, B. and Bağış, S.** (2013). Spatiotemporal soil moisture estimation for agricultural drought risk management, *2013 Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, pp.76–81.
- [120] **Allen, R.G., Pereira, L.S., Raes, D., Smith, M. et al.** (1998). Crop evapotranspiration-Guidelines for computing crop water requirements FAO Irrigation and drainage paper 56, *Fao, Rome*, 300(9), D05109.
- [121] **URL-3.** <<https://earthexplorer.usgs.gov>>; Earth Exploerer; Retrieved date 15.08.2022.
- [122] **Roy, D., Kovalskyy, V., Zhang, H., Vermote, E., Yan, L., Kumar, S. and Egorov, A.** (2016). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity, *Remote Sensing of Environment*, 185, 57–70, <https://www.sciencedirect.com/science/article/pii/S0034425715302455>, landsat 8 Science Results.
- [123] **Flood, N.** (2014). Continuity of Reflectance Data between Landsat-7 ETM+ and Landsat-8 OLI, for Both Top-of-Atmosphere and Surface Reflectance: A Study in the Australian Landscape, *Remote Sensing*, 6(9), 7952–7970, <https://www.mdpi.com/2072-4292/6/9/7952>.
- [124] **Kulaglic, A. and Ustundag, B.B.** (2021). Predictive Error Compensating Wavelet Neural Network Model for Multivariable Time Series Prediction.
- [125] **Bansal, M., Sharma, R. and Grover, P.** (2010). Performance evaluation of Butterworth Filter for Signal Denoising, *International Journal of Electronics & Communication Technology*, December.
- [126] **Ellis, G.** (2012). *Chapter 9 - Filters in Control Systems*, Butterworth-Heinemann, Boston, fourth edition edition, <https://www.sciencedirect.com/science/article/pii/B9780123859204000096>.

[127] **URL-4.** <<https://www.investing.com>>; Investing; Retrieved date 15.08.2022.

[128] **URL-5.** <<https://www.matriksdata.com/website>>; Information Distribution services Matrix; Retrieved date 15.08.2022.





CURRICULUM VITAE

Name Surname: Ajla KULAGLIC

EDUCATION:

- **B.Sc.:** 2008, University of Sarajevo, Faculty of Electrical Engineering, Computer Science and Informatics Department, Sarajevo, Bosnia and Herzegovina.
- **M.Sc.:** 2010, University of Sarajevo, Faculty of Electrical Engineering, Computer Science and Informatics Department, Sarajevo, Bosnia and Herzegovina.

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2006 Award for the most hardworking students in academic year 2005/2006 issued by the University of Sarajevo, Faculty of Electrical Engineering
- 2007 Award for the most hardworking students in academic year 2006/2007 issued by the University of Sarajevo, Faculty of Electrical Engineering
- 2010-2015 Turkish Government Scholarship
- 2011-2015 Measurement and system expert, Agricultural Informatics Research and Application Center (TARBIL), Istanbul Technical University, Istanbul, Turkey
- 2011-2017 Research Assistant, Agricultural Informatics Research and Application Center (TARBIL), Istanbul Technical University, Istanbul, Turkey
- 2017-2019 Senior Teaching Assistant, Lecturer, University Sarajevo School of Science and Technology, Sarajevo, Bosnia and Herzegovina
- 2022 Completed Doctorate at Istanbul Technical University

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- B. B. USTUNDAG and A. KULAGLIC, 2020. High-Performance Time Series Prediction With Predictive Error Compensated Wavelet Neural Networks. IEEE Access, vol. 8, pp. 210532-210541, 2020, doi:10.1109/ACCESS.2020.3038724.
- A. KULAGLIC, B.B. USTUNDAG, 2021. Improvement of Stock Price prediction Performance by Predictive Error Compensation in Wavelet Neural Networks, Computers, Materials, and Continua Journal, DOI:10.32604/cmc.2021.014768

- **A. KULAGLIC**, B.B. USTUNDAG, 2021. Predictive Error Compensating Wavelet Neural Network Model for Multivariable Time Series Prediction, TEM JOURNAL - Technology, Education, Management, Informatics DOI: 10.18421/TEM104-61
- **A. KULAGLIC**, A. AKTAS, B. Berk USTUNDAG, 2021. Predictive Error Compensating Neural Networks for Improvement of Soil Moisture Estimation Accuracy, the Ninth International Conference on Agro-Geoinformatics : July 26-29, virtual conference
- **A. KULAGLIC**, B. Berk USTUNDAG, 2018. Stock Price Forecast using Wavelet Transformations in Multiple Time Windows and Neural Networks, 3rd International Conference on Computer Science and Engineering UBMK'18, pp. 518-521, doi: 10.1109/UBMK.2018.8566614.
- **A. KULAGLIC** , B. Berk USTUNDAG, 2018, Stock Index Forecasting using Wavelet Decomposed Neural Networks. Computational Methods and Telecommunication in Electrical Engineering and Finance, Sarajevo, BiH, pp. 18-20
- **A. KULAGLIC**, B. Berk USTUNDAG, Alper YILMAZ, 2021., PECNET. Patent application.

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- Begic E. Hadzidedic S., **KULAGLIC A.**, Ramic-Brkic B., Begic Z., Causevic M., "SOMA SOMAScan-based proteomic measurements of plasma brain natriuretic peptide are decreased in mild cognitive impairment and in Alzheimer's dementia patients." PLoS ONE 14(2):e0212261. <https://doi.org/10.1371/journal.pone.0212261>
- E. Sehovic, L. Spahic, **A. KULAGLIC**, L. Smajlovic-Skenderagic, A. Hajdarpasic-Saracevic, "Analysis of miRNA Targets in Correlation to Neurodevelopment and Diagnosis of Autism Spectrum Disorder (ASD)", part of the IFMBE Proceedings book series (IFMBE, volume 73)
- **A. KULAGLIC**, B. Berk Ustundag, "Effect of Soil Property Changes on Soil Moisture Profile Estimation using Cascaded Wavelet Pre-processed Neural Networks ", Agro-Geoinformatics 2019, The Eighth International Conference on AgroGeoinformatics, Istanbul, Turkey, 16-19 July 2019
- **A. KULAGLIC**, B. Berk Ustundag, " Stock Price Forecast using Wavelet Transformations in Multiple Time Windows and Neural Networks ", 3rd International Conference on Computer Science and Engineering UBMK'18
- **A. KULAGLIC**, B. Berk Ustundag, " Stock Index Forecasting using Wavelet Decomposed Neural Networks", Computational Methods and Telecommunication in Electrical Engineering and Finance, May 2018, Sarajevo, BiH
- **A. KULAGLIC**, B. Berk Ustundag, "Spatio-temporal Soil Moisture Estimation Using Neural Network with Wavelet Preprocessing ", Agro-Geoinformatics 2017, The Sixth International Conference on AgroGeoinformatics, Fairfax, VA, USA, 7-10 August 2017

- **A. KULAGLIC**, B. Berk Ustundag, "Neural Network with Wavelet Preprocessing for Wheat Growth Stage Estimation", Agro-Geoinformatics 2016, The Fifth International Conference on AgroGeoinformatics, Tjianin, China, 18-20 July 2016
- **A. KULAGLIC**, B. Berk Ustundag, " Effect of Soil Property Change on Soil Moisture Profile Estimation Through Data Fusion", Agro-Geoinformatics 2015, The Fourth International Conference on AgroGeoinformatics, Beijing China, 20-24 July 2015; Pages: 178 - 182, DOI: 10.1109/Agro- Geoinformatics.2015.7248131
- **A. KULAGLIC**, Adnan Ozdemir, B. Berk Ustundag, "Reconstruction of Missing Sensor Data Using Different Interpolation Techniques", Agro-Geoinformatics 2014, The Third International Conference on AgroGeoinformatics, Beijing China, 11-14 August 2014 ;
- **A. KULAGLIC**, B. Berk Ustundag, "Estimation of Soil Moisture Profile Using Wavelet Neural Networks", Agro-Geoinformatics 2014, The Third International Conference on AgroGeoinformatics, Beijing China, 11-14 August 2014 ; Pages: 1 - 6, DOI: 10.1109/Agro- Geoinformatics.2014.6910632
- **A. KULAGLIC**, B. Berk Ustundag, Serdar Bagis, "Spatiotemporal Soil Moisture Estimation for Agricultural Drought Risk Management", Agro-Geoinformatics 2013, The Second International Conference on AgroGeoinformatics, Fairfax, VA USA; Pages: 76 - 81, DOI: 10.1109/Argo- Geoinformatics.2013.6621883
- Bekir Sitki Ertugrul, Hasan Kivrak, Evren Daglarli, **A. KULAGLIC**, Ahmet Tekelioglu, Rabia Yorgancı, Hatice Kose, "iSign: An Interactive Sign Imitation Game with Humanoid Robots." HBU 2012, 3rd International Workshop on Human Behavior Understanding

A. KULAGLIC

DEEP WAVELET NEURAL NETWORK FOR SPATIO-TEMPORAL DATA FUSION

2022