DOKUZ EYLÜL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

FORECASTING NONLINEAR TIME SERIES USING PARTIAL LEAST SQUARES METHOD

by

Mefharet ERTAÇ

February, 2014

İZMİR

FORECASTING NONLINEAR TIME SERIES USING PARTIAL LEAST SQUARES METHOD

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Statistics Program

by

Mefharet ERTAÇ

February, 2014

İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"FORECASTING NONLINEAR TIME SERIES USING PARTIAL LEAST SQUARES METHOD"** completed by **MEFHARET ERTAÇ** under supervision of **ASSOC. PROF. DR. ESİN FİRUZAN** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Assoc. Prof. Dr. Eşin FİRUZAN

Supervisor

Prof. Dr. Efendi NASİBOĞLU

Thesis Committee Member

Assist. Prof. Dr. Nalan ÖZKURT

Thesis Committee Member

Prof. Dr. Saban Eren

Examining Committee Member

Prof. Dr. [illegible] KURT

Examining Committee Member

Prof.Dr. Ayşe OKUR
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

First of all, I would like to express my appreciation to my supervisor Assoc. Prof. Dr. Esin FİRUZAN for her great help and constructive criticism. She encouraged me to develop own ideas and to stand on one's leg in academic life. Moreover, she did not only supervise my thesis but also she was always interested in my life.

I acknowledge Kandilli Observatory and Earthquake Research Institute especially Şenol Solum for supplying long term meteorological data of İstanbul.

I want to thank to my dear friend Soner SEZGİN for helping me greatly to prepare MATLAB codes.

I am thankful to my family. I can not thank them enough for their love, encouragement, and patience. Especially my sister Yağmur ÖZKILÇIK standed always by me at every stage of my PhD process.

Last, but not least, I would like to thank my husband Özgür ERTAÇ who always encouraged me to prepare this work. I am grateful for his great love, patience, and understanding. Thanks to my husband for being my great source of motivation.

Mefharet ERTAÇ

FORECASTING NONLINEAR TIME SERIES USING PARTIAL LEAST SQUARES METHOD

ABSTRACT

The goals of this thesis are (a) to develop a new nonlinear forecasting adaptation by combining partial least square (PLS) method and multivariate embedding; (b) to improve the forecast ability by evaluating the forecasting errors depending upon the parameter and forecasting method (local or global) changes; (c) to forecast İstanbul monthly mean temperature by benefitting from this new method and meteorological variables like temperature, humidity and rainfall.

A new PLS forecasting technique based on chaotic analysis is developed by using nonlinear time series and variable selection methods. Approximately 100 years of meteorological data obtained from 54 automatic meteorology observation stations of İstanbul that is the mega city of Turkey are analyzed to infer about the meteorological behaviour of the city. The proposed model is also compared with artificial neural networks (ANN) which model nonlinearly the relation between inputs and outputs by working neurons like human brain. Ordinary least square (OLS), PLS and ANN methods are used for nonlinear time series forecasting in this thesis. Major findings are best performance values of the proposed PLS model and the chaotic nature of the meteorological data of İstanbul.

Keywords: Meteorological data, İstanbul, nonlinear time series, forecasting errors, ordinary least square (OLS), partial least square (PLS), artificial neural network (ANN)

DOĞRUSAL OLMAYAN ZAMAN SERİLERİNİN KISMİ EN KÜÇÜK KARELER YÖNTEMİ İLE TAHMİNLENMESİ

ÖZ

Bu tezin amacı (a) kısmi en küçük kareler (KEKK) ve çok değişkenli gömme yöntemlerini kullanarak yeni bir doğrusal olmayan tahminleme adaptasyonu geliştirmek; (b) parametre veya tahminleme yöntemi değişikliklerine (yerel veya global) bağlı olarak tahminleme hatalarını değerlendirerek tahmin yeteğini iyileştirmek; (c) bu yeni yöntem ve sıcaklık, nem ve yağış gibi meteorolojik değişkenlerden yararlanarak İstanbul'un aylık ortalama sıcaklığını tahminlemektir.

Doğrusal olmayan zaman serisi ve değişken seçimi yöntemleri kullanılarak kaotik analizi temel alan yeni bir KEKK tahminleme yöntemi geliştirilir. Türkiye'nin mega şehri olan İstanbul'un 54 otomatik meteoroloji gözlem istasyonundan elde edilen yaklaşık 100 yıllık meteorolojik veri şehrin meteorolojik davranışı hakkında yorum yapmak için analiz edilmiştir. Ayrıca, önerilen yöntem insan beyni gibi çalışan nöronlar sayesinde girdilerle çıktılar arasındaki ilişkiyi doğrusal olmayan şekilde modelleyen yapay sinir ağları (YSA) ile karşılaştırılır. Bu tezde, doğrusal olmayan zaman serisinin tahminlenmesi için sıradan en küçük kareler (EKK), KEKK ve YSA yöntemleri kullanılır. Temel bulgular önerilen KEKK yönteminin en iyi performans değerlerine sahip olması ve İstanbul'un meteorolojik verisinin kaotik yapısıdır.

Anahtar Kelimeler : Meteorolojik veri, İstanbul, doğrusal olmayan zaman serisi, tahminleme hataları, sıradan en küçük kareler (EKK), kısmi en küçük kareler (KEKK), yapay sinir ağı (YSA)

CONTENTS

Page

THESIS EXAMINATION RESULT FORM ii
ACKNOWLEDGEMENTS iii
ABSTRACT iv
ÖZ v
LIST OF FIGURES viii
LIST OF TABLES ix

CHAPTER ONE – INTRODUCTION 1

1.1 Introduction 1
1.2 Objective of the Dissertation 4

CHAPTER TWO –NONLINEAR TIME SERIES TOOLS 5

2.1 Introduction 5
2.1.1 Linear versus Nonlinear Dynamical Systems 6
2.1.2 Phase Space Methods 7
2.2 Delay Reconstruction 7
2.2.1 False Neighbours: Choice of Dimension 8
2.2.2 Mutual Information: Choice of Time Delay 9
2.2.3 Phase Space Plot 9
2.3 Correlation Dimension 10
2.4 Recurrence Plots 11
2.5 Lyapunov Exponent 12
2.5.1 Calculation of Lyapunov Exponents from Time Series 14
2.5.2 Calculation of Lyapunov Exponents by Jacobian Approach 15
2.6 Symbolic Time Series and Entropy 17
2.6.1 Other Symbol Sequence Statistics 18
2.6.2 Relation between Lyapunov Exponent and Entropy 19

CHAPTER THREE –FORECASTING METHODS 21

3.1 Introduction 21

3.2 Ordinary Least Square 24
3.3 Partial Least Square 25
3.3.1 Linear PLS 26
3.3.1.1 NIPALS Algorithm 27
3.3.1.2 Quadratic Partial Least Square Method (QPLS) 28
3.3.2 Nonlinear PLS 29
3.3.2.1 Neural Network Partial Least Square Method (NNPLS) 30
3.3.2.2 Kernel Partial Least Square Method (KPLS) 31
3.3.3 PLS Variable Selection Method 32
3.4 Artificial Neural Networks (ANN) 33
3.4.1 Focused Time-Delay Neural Network 34
3.4.2 Spatio-Temporal Time-Lagged Networks 36
3.5 Proposed Methods of the Thesis 37
3.5.1 Adapted Partial Least Square Method 37
3.5.2 Proposed Method 37
3.6 Performance Measures of Forecasting Methods 38

CHAPTER FOUR –AN APPLICATION 40

4.1 Introduction 40
4.2 Analysis 41
4.2.1 Data 42
4.2.1 Nonlinear Time Series Analysis 48
4.3 Forecasting Results 51
4.3.1 OLS Results 52
4.3.2 PLS Results 53
4.3.3 Proposed Method Results 55
4.3.4 Comparing Proposed Method with Artificial Neural Networks 58
4.3.5 Forecasting Results of Proposed Method 65

CHAPTER FIVE –CONCLUSION 67

REFERENCES 70

LIST OF FIGURES

Page

Figure 2.1 3-dimensional Lorenz attractor.....6
Figure 2.2 Time series and phase space plot of Lorenz map 10
Figure 2.3 Recurrence plot of Sine function 12
Figure 2.4 Process of symbolizing a real sequence..... 17
Figure 3.1 PLS analysis of the relationship between X and Y matrices 26
Figure 3.2 Illustration of PLS variable selection methods.....33
Figure 3.3 General architecture of FTDNN 35
Figure 4.1 Meteorological stations of Istanbul 42
Figure 4.2 Boxplots of meteorological variables (temperature: orange, humidity: green and rainfall: blue) for 5-year periods between 1921 and 2010 44
Figure 4.3 Time series plots of monthly temperature, humidity and rainfall data for Istanbul in February 1914-December 2010..... 48
Figure 4.4 Recurrence plots for temperature, humidity and rainfall data.....49
Figure 4.5 Mutual Information graphs 50
Figure 4.6 False Nearest Neighbours graphs 50
Figure 4.7 Phase Space Plot for Temperature (τ=3, m=8) Phase Space Plot for Humidity (τ=3, m=6) Rainfall (τ=2, m=1).....51
Figure 4.8 Forecasting error values versus k values 52
Figure 4.9 Forecasting error values versus k values and number of components..... 54
Figure 4.10 Observed vs. Forecasted values.....57
Figure 4.11 Comparing forecasting methods errors..... ..58
Figure 4.12 Architecture of NARX neural network..... 59
Figure 4.13 NARX neural network results 61
Figure 4.14 Performance plot..... ..62
Figure 4.15 Coefficients of determination 63
Figure 4.16 Time series response..... ..64
Figure 4.17 Analyzing the Temperature by Five Years Mean Values between 1961 and 2015..... 66

LIST OF TABLES

Page

Table 2.1 Possible types of motion and the corresponding maximal Lyapunov exponents ..... .14
Table 4.1 Forecasting Errors of Temperature (OLS Method)..... ..52
Table 4.2 Forecasting Temperature by PLS Method (k=20) 53
Table 4.3 Forecasting Temperature by PLS Method (k=50) 53
Table 4.4 Forecasting Temperature by PLS Method (k=100) 54
Table 4.5 Multicollinearity statistics of selected variables 56
Table 4.6 Forecasting Errors of Temperature by PLS and Variable Selection Method.56
Table 4.7 NARX neural network performance values (RMSE) 65
Table 4.8 Comparison between the Forecast Values and the Past Temperature Data ..65

CHAPTER ONE
INTRODUCTION

1.1 Introduction

In the real life, innumerable time series has behaviours that are irregular, seen as random and unpredictable, in other words, chaotic. Since chaos means irregular behaviour of dynamical systems arising from a deterministic time evolution with sensitivity to initial conditions, analyzing chaotic time series by linear methods and interpreting are almost impossible. Linear time series methods can't consider irregular behaviours of the systems and these behaviours has to be taken into account, too. Nonlinear time series analysis uses chaos theory and nonlinear dynamics to understand such unpredictable behaviour. So, nonlinear time series methods has become one of the most exciting and rapidly expanding research topics of recent decades.

Zhang et al. (2001) presented an experimental evaluation of neural networks for nonlinear time series forecasting. Bertels et al. (2001) studied neural networks from a dynamical systems point of view and examined its properties. Konstantinou and Lin (2004) applied nonlinear methods to volcanic tremor events recorded at Sangay volcano, Ecuador in order to obtain more information about this series. Shang et al. (2007) applied nonlinear time series techniques to analyze suspended sediment data. Behnia et al. (2008) investigated the application of nonlinear analysis in differentiating between patients and healthy persons using dynamics of the heart. Sakai et al. (2008) proposed a way to reconstruct local dynamics from a very short ecological time series whose data point is smaller than ten.

Nonlinear methods that have developed in parallel with developments in computer technology provide to exhibit chaotic attributes. Recent studies about predicting chaotic time series have shown that local linear prediction models mostly perform better than global forecasting models like Artificial Neural Network (ANN). Especially Casdagli's paper is one of the most important literatures in this context.

Casdagli (1991) introduced a technique for forecasting and investigating the amount of the irregularities in low dimensional chaotic, stochastic or high dimensional chaotic dynamics and applied the technique to various time series data. In the study, Casdagli's method is focused in this thesis, since the amount of the studies related with local linear prediction in the literature has increased after Casdagli's method.

Alexander and Giblin (1997) studied on a method based on the nearest neighbour algorithm. Jaditz and Sayers (1998) also used this type forecasting methodology as a prediction test for evidence of a noisy, chaotic data generating process underlying the Divisia monetary aggregate series and compared the forecasting accuracy of this method to the other methods. Aparicio et al. (2002) applied the nearest neighbour forecasting method to five series taken from the US Stock Market to test whether these series display chaotic dynamics and some of the fluctuations of the series could be explained by chaos in their study. Diaz (2008) used different generalizations of standard nearest neighbours to predict the dynamic of exchange rates and compared their results with the global methods results as neural networks and genetic algorithms.

Hougland (2010) studied on local linear prediction and local noise reduction methods in the TISEAN package program. Dhanya and Kumar in the same year used local method for forecasting daily rainfall data of three regions Malaprabha, Mahanadi and All India for the period 1955-2000. Yamamura and Shoji also presented a nonparametric model that combined the local linear model and the Gaussian kernel for multi-step ahead forecasting in diffusion processes and had better forecasting performance than the global linear model. Kugiumtzis (2000) described the the general features of local linear prediction. He dealed with the statistical properties of the OLS and regularization methods (Principal Component Regression (PCR), Partial Least Square (PLS) and Ridge Regression (RR)) and demonstrated these using simulated and financial data. Currenti et al. (2004) used local and the global linear methods to understand and describe the dynamics of volcanomagnetic signals of Etna volcano. Laio et al. (2004) also studied on some nature data and tried to infer about the dynamics of the rainfall-runoff transformation

considering riverflow time series and using deterministic versus stochastic (DVS) plots.

Diaz and Sbert (2011) used different nearest neighbour predictors (a simple baricentric predictor, a weighted baricentric predictor, local regression, tricube weighted local regression and exponential weighted local regression) to predict the tourist arrivals in Mallorca Island. Recently, Zhang and Yue (2012) studied on local prediction and used engine exhaust temperature (EGT) time series for their study. They adapted the partial least square (PLS) based on the cubic spline function or the kernel function transformation to the data and showed that their PLS chaotic prediction algorithm based on bi-weight kernel function transformation had less Normalized Mean Square Error (NMSE) values than the ordinary least square method.

In this thesis, two methods that approach from different views to analyze meteorological observations are combined. The first method is nonlinear time series analysis that indicates the predictability of meteorological data which is irregular and has chaotic structure. The second method is adaptation of partial least square model (PLS) for forecasting. Since the lagged time series values might be highly correlated, PLS is adapted to Casdagli's forecasting method. Right along with being a good local prediction method, Partial Least Square (PLS) has the advantages of managing multicollinearity and avoiding cumulative error with local single-step linear model. The studies in nonlinear time series literature have generally studied on univariate prediction by reconstructing the phase space. In addition, multivariate prediction is also dealt in this thesis. The proposed method is applied to a dataset which include 100 years of observations that was obtained from Kandilli Observatory and Earthquake Research Institute that 53 local stations in Istanbul that continuously collected meteorological data.

1.2 Objective of the Dissertation

In this dissertation, nonlinear time series forecasting by different statistical as OLS, PLS, PLS with variable selection and ANN methods, comparing these methods by interpreting the results and drawing conclusions about the methods and the future trends of the series are focused on. The goals of this thesis are to develop a new nonlinear forecasting adaptation by combining PLS method and multivariate embedding, to improve the forecast ability by evaluating the forecasting errors depending upon the parameter changes and local or global forecasting methods and to forecast Istanbul monthly mean temperature by benefitting from this new method and meteorological variables like temperature, humidity and rainfall.

The organization of this dissertation is as follows: In the following chapter, the most widely used nonlinear time series analysis tools are described. In the third chapter, OLS, PLS and ANN that are the most used nonlinear time series forecasting methods are described, examined and compared by their performances. In chapter four, the proposed forecasting methods are applied to meteorological data of Istanbul. The methods are illustrated and the main results of this chapter are given with the concluding remarks. In the final chapter, the conclusions that include results and future research subjects are provided. In this dissertation, MATLAB software is used for statistical computations. Visual Reccurrence Analysis that is nonlinear time series package software is used for nonlinear time series parameters' computations.

CHAPTER TWO
NONLINEAR TIME SERIES TOOLS

2.1 Introduction

Nonlinear time series methods have currently become one of the most interesting and rapidly growing research topics. Nonlinear time series tools benefit from chaos theory and dynamical systems to understand such a complex behaviour. In deterministic systems context, all future states of the system can be determined if the present state is exact. And when we are analyzing nonlinear time series, we need to benefit from these dynamical systems and their concepts to define these complex series behaviors. If we can define the current situation of the series in a space, we can infer about the future. Hence, the dynamics of the series can be worked on by analyzing the dynamics of the related phase space points. From this point of view, we firstly interested in some dynamical systems tools that are used in nonlinear time series analysis to be able to specify the related time series in a new space and determine the governing rule of this series.

Dynamical systems include a space and some rule describing the evolution of a point in that space. This evolution rule of the dynamical system is a fixed rule that determines the future states of the current state. "A dynamical system is a manifold *M* called the phase space and smooth evolution function ϕ^t that for any element of $t \in T$, the time, maps a point of the phase space back into the phase space. There are several options for the set T. When T is real numbers, the dynamical system is called a flow, and if T is restricted to the nonnegative real numbers, then the dynamical system is a semiflow. When T represents with integer numbers, it is cascade or a map; and the restriction to the nonnegative integer is semi-cascade" (Izhikevich, 2011). The solution x(t) of an ordinary differential equation, $dx/dt=f(x)$, where $x \in R^n$, from a given initial condition $z_0=x(t=0)$, plotted in phase space called a *trajectory* or *orbit*. A trajectory of a flow is also referred to as an orbit (Kapitaniak & Bishop, 1998). Attractors become by a dynamical system evolving

after very long time. So, related points get close enough to the attractor and form a pattern.

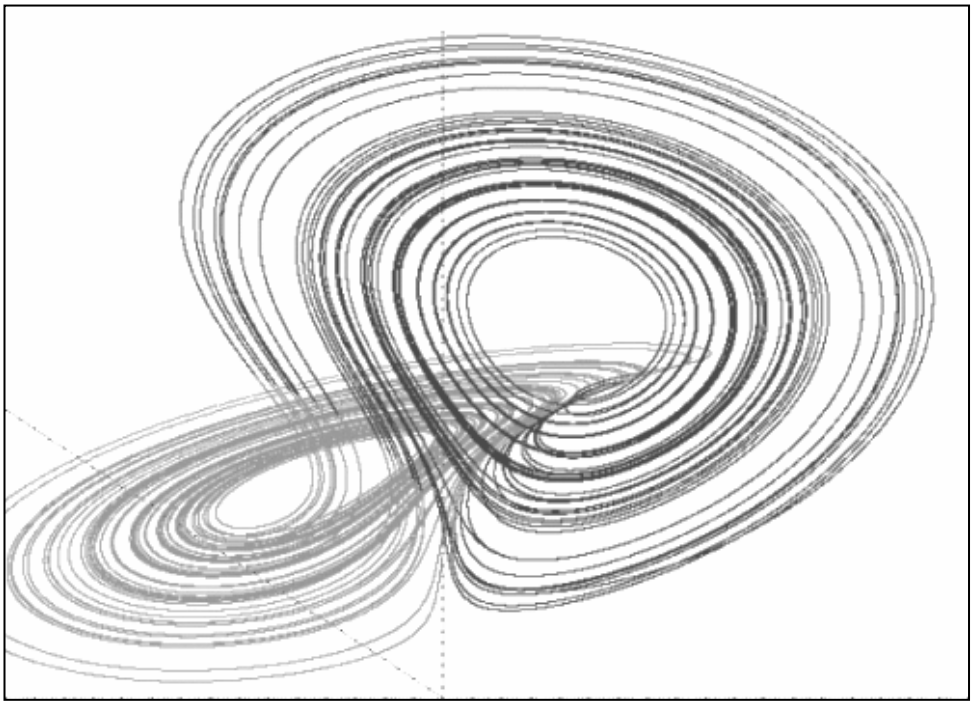


Figure 2.1 3-dimensional Lorenz attractor (Bradley, 2010)

In Figure 2.1, 3-dimensional Lorenz attractor is presented as an example attractor.

2.1.1 Linear versus Nonlinear Dynamical Systems

A nonlinear system is a system like a linear system except that one or more terms are nonlinear. The difference between linear and nonlinear is based on the details of the governed equations. Linear systems generally include one of following four predictable behaviours:

1. Grows exponentially and heads toward infinity.
2. Decays exponentially and heads toward zero.
3. Cycles periodically.
4. Exhibits any combination of the above behaviors.

For nonlinear systems, finding a solution is much more complex and need more detailed especially visual analysis.

2.1.2 Phase Space Methods

Deterministic dynamical systems describe the time evolution of a system in some phase space. The introduction of phase space enables one to study the dynamics of a complicated system geometrically. These systems can be expressed by two forms:

- For discrete time variable

$$x_{n+1} = f(x_n) \quad n \in Z \tag{2.1}$$

- For continous time variable by ordinary differential equations:

$$x(t) = F(x(t)) \tag{2.2}$$

$$\frac{d}{dt}x(t) = f(x(t)) \quad t \in R \tag{2.3}$$

A sequence of points x_n or $x(t)$, solving the above equations is called a trajectory of the system with x_0 or $x(0)$ initial conditions. Discrete time situation is found more convenient on many occasions. For simplicity these situations are restricted to the case where the phase space is a finite dimensional vector space $\Re^m$ and m is the embedding dimension. Phase space reconstruction is the first stage of nonlinear time series analysis and it is a dynamical system tool. Phase space provides us to see simply even the governing rule of a complex system. Phase space plots will be detailed examined in this chapter.

2.2 Delay Reconstruction

In nonlinear time series analysis, corresponding time series is not a phase space object and the observations must firstly be converted into state vectors. This problem

is solved by the method of delays. Takens (1981) proved that the attractor from a time series of a single component can be reconstructed by embedding vectors.

By phase space reconstruction, s_n vectors that are sequences of scalar measurements depend on the current state of the system and taken at fixed sampling times as in Equation 2.4 reformed in m dimensions as in Equation 2.5 (Kantz & Schreiber, 2004):

$$s_n = s(x(n\Delta t)) + \eta_n. \tag{2.4}$$

$$s_n = (s_{n-(m-1)\tau}, s_{n-(m-2)\tau}, \ldots, s_{n-\tau}, s_n). \tag{2.5}$$

where lag or delay time *(τ)* is time difference between the components of the delay vectors.

The embedding theorems state that there exists a sufficiently large dimension *m* which the vectors s_n are equivalent to phase space vectors for every time series data without noise. In the following parts, brief information about *m* and *τ* is mentioned.

2.2.1 False Neighbours: Choice of Dimension

The false neighbours concept was first introduced by Kennel, Brown & Abarbanel (1992). The basic idea behind this concept is: neighbouring states should have almost the same time evolution. For each point of time series its closest neighbour in *m* dimensions is found and the ratio of the distance between these two points is calculated in *m+1* dimensions and in *m* dimensions to obtain false neighbours:

$$X_{fnn}(r) = \frac{\sum_{n=1}^{N-m-1} \Theta\left(\frac{\left| s_n^{(m+1)} - s_{k(n)}^{(m+1)} \right|}{\left| s_n^{(m)} - s_{k(n)}^{(m)} \right|} - r \right) \Theta\left(\frac{\sigma}{r} - \left| s_n^{(m)} - s_{k(n)}^{(m)} \right| \right)}{\sum_{n=1}^{N-m-1} \Theta\left(\frac{\sigma}{r} - \left| s_n^{(m)} - s_{k(n)}^{(m)} \right| \right)} \tag{2.6}$$

σ: standard deviation of the data

$s_{k(n)}^{(m)}$: the closest neighbour to s_n in m dimensions

$k(n)$: the time series element k different from n for which $|s_n - s_k| = \min$

$X_{fnn}(r)$: ratio of false nearest neighbours based upon threshold r

If obtained ratio is larger than a determined threshold *r*, then the neighbour is false (Kantz & Schreiber, 2004).

2.2.2 Mutual Information: Choice of Time Delay

Estimation of lag time is also a difficult step in nonlinear time series analysis and Mutual Information Method relies on the information theoretic concept of mutual information helps us. The information *I(T)* can be defined by:

$$I(T) = \sum_{n=1}^{N} P(y_n, y_{n+T}) \log_2 \frac{P(y_n, y_{n+T})}{P(y_n)P(y_{n+T})} \tag{2.7}$$

$P(y_n, y_{n+T})$: the probability of observing y_n and y_{n+T},

$P(y_n)$: the probability of observing y_n.

I(T) is the amount of information we have about y_n by observing y_{n+T} and optimum τ is the first local minima of *I(T)*. In Mutual Information calculation case, deciding the histogram binning is very important since wrong decision can easily lead to poor results and the Mutual Information will be a good guide when this condition is provided (Small, 2005).

2.2.3 Phase Space Plot

Visual inspection of a systems dynamics is an important stage in understanding that system. There are two popular types of graph show a system's dynamics: time series plot and phase space plot. While first plot is a two-dimensional plot of some variable versus time, time is not directly plotted and shown by the sequence of

plotted points in the second one. Phase space plots involve more than one variable; X_t values and lagged values of X_t and provide more obvious view of the time evolution. Phase space plot of Lorenz map that is one of the most known dynamical systems is shown as in Figure 2.2.

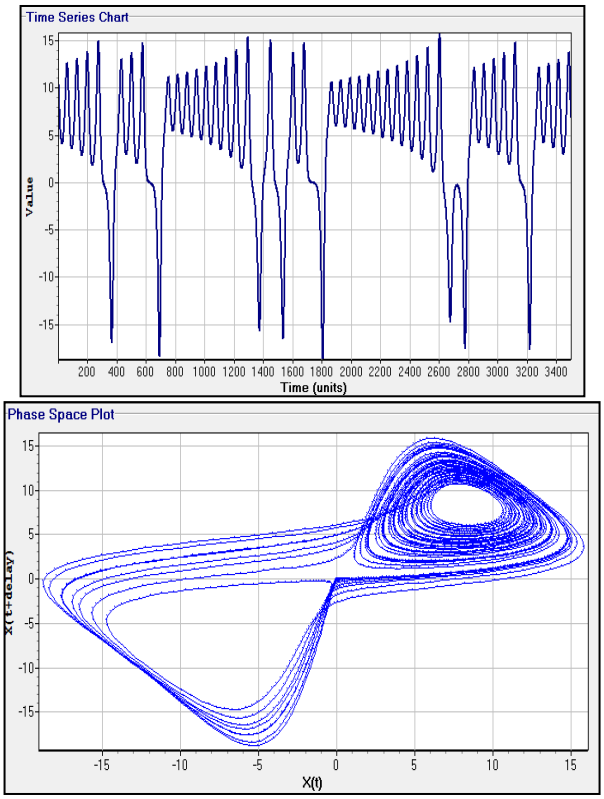


Figure 2.2 Time series and phase space plot of Lorenz map

As seen in the figure, time evolution of the map can be noticed more explicitly by phase space plots.

2.3 Correlation Dimension

The correlation dimension that was introduced by Grassberger & Procaccia (1983) and (1983a) is denoted by *v* is usually used in nonlinear time series. It is a dimension measure of the space consists of a set of random points. The correlation dimension is often used in the case of reconstructing the geometrical object from a finite sample of data points.

Let x_n be a collection of points in some vector space. Then the formula for correlation sum is:

$$C(\varepsilon) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta(\varepsilon - \left\| x_i - x_j \right\| \tag{2.8}$$

Θ is the Heaviside step function; Θ (x)=0 if x≤0 and Θ (x)=0 for x>0.
Then, the correlation dimension D is defined by

$$d(N,\varepsilon) = \frac{\partial \ln C(\varepsilon, N)}{\partial \ln \varepsilon}$$
$$D = \lim_{\varepsilon \to 0} \lim_{N \to \infty} d(N,\varepsilon) \tag{2.9}$$

(Kantz, H.& Schreiber, T., 2004).

The last step is to calculate $d(N,\varepsilon)$ and plot it against ε. For a low-dimensional chaotic sytem, when the dimension become sufficient the curves saturate at a particular value of $d(N,\varepsilon)$. This $d(N,\varepsilon)$ is called as the correlation dimension ν of the system. The advantage of the correlation dimension is computation easiness and time.

2.4 Recurrence Plots

Recurrence is that coming back of a point to a neighbourhood of where it has been before in the trajectory. Visualizing these reccurences is very important for identifying intermittency and nonstationarity structures in the data. Recurrence plots were introduced by Eckmann et al. (1987). While for a stationary process points spread uniformly, the points cluster near the diagonal for a nonstationary one.

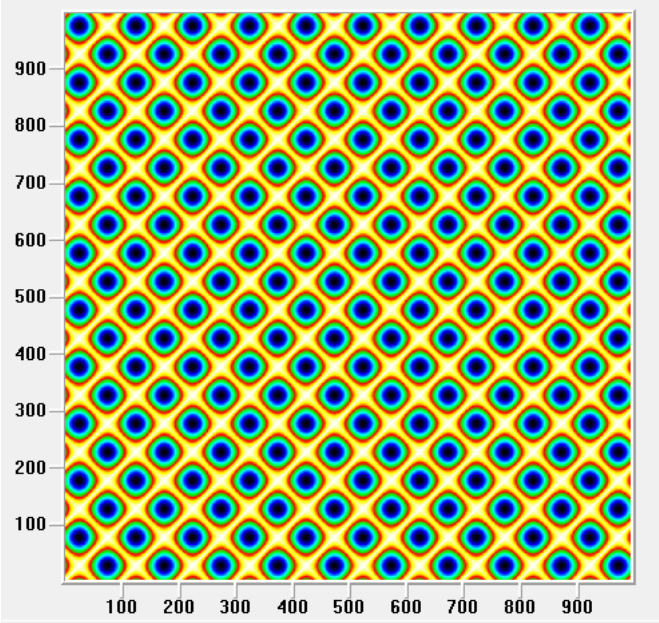


Figure 2.3 Recurrence plot of Sine function

Recurrence plot of Sine function which is a stationary process is presented as in Figure 2.3. Uniformly spreaded patterns are obviously seen in the figure.

2.5 Lyapunov Exponent

Being quickly unpredictable with increasing prediction horizon and even in this case exhibiting statistical regularities are two important properties of chaos. Calculating the Lyapunov exponent, average divergence rate (λ), is the simplest method of quantifying chaos. The difference in the initial condition grows exponentially with a rate of $\exp(\lambda m)$ after m iterates, if λ is positive. In other words, $\lambda m = \log 2$, so the doubling time equals $\log 2/\lambda$ and this means when λ increases, the doubling period gets shorter and the system becomes quickly unpredictable. Lyapunov exponent is very important in chaos theory, since a system is called to be chaotic only if it has a positive Lyapunov exponent (Chan & Tong, 2001).

There are several methods to estimate Lyapunov exponents. A simple procedure that estimates the spectrum of Lyapunov exponents were developed by Benettin et al. in 1976. But this method was not robust to some conditions like noise in data. Since a positive maximal Lyapunov exponent is a strong sign of chaos, computation of its

value is very practical way for deciding existence of chaos. The first algorithm was suggested by Wolf et al. (1985) and then Rosenstein (1993), Kantz and Schreiber (2004) presented the algorithm for calculating the largest Lyapunov exponent.

Let X_0 and $X_0 + \Delta x_0$ be two orbits that are the subsets of a space. If one of the orbits is used as a reference orbit, separation is also a function and has the form $\Delta x(X_0, t)$ and this function $\Delta x(X_0, t)$ for a chaotic system behaves irregularly. The mean exponential rate of divergence of these initially close orbits is calculated as below formula:

$$\lambda = \lim_{\substack{t\to\infty \\ |\Delta x_0|\to 0}} \frac{1}{t} \ln \frac{|\Delta x(x_0, t)|}{|\Delta x_0|} \tag{2.10}$$

(Elert, 2007).

In the case of calculating the maximal Lyapunov exponent λ, let s_{n1} and s_{n2} be two points in state space with distance $\|s_{n1} - s_{n2}\| = \delta_0 << 1$ and $\delta_{\Delta n}$ is the distance time Δn ahead between the two trajectories.

$$\delta_{\Delta n} = \|s_{n1+\Delta n} - s_{n2+\Delta n}\| \tag{2.11}$$

Then λ is computed by:

$$\delta_{\Delta n} \cong \delta_0 e^{\lambda \Delta n} \ , \ \delta_{\Delta n} << 1, \ \Delta n >> 1 \tag{2.12}$$

The types of motion for corresponding maximal Lyapunov exponent values are given in Table 1 (Kantz, H. & Schreiber, T., 2004).

Table 2.1 Possible types of motion and the corresponding maximal Lyapunov exponents

Type of motion	Maximal Lyapunov exponent
Stable fixed point	$\lambda<0$
Stable limit cycle	$\lambda=0$
Chaos	$0<\lambda<\infty$
Noise	$\lambda=\infty$

There are several numerical methods to estimate Lyapunov exponents. Two general approaches for computing the Lyapunov exponents from time series are geometrical and Jacobian approaches. Calculating Lyapunov exponents directly from time series and Jacobian approach are dealed in this study.

2.5.1 Calculation of Lyapunov exponents from time series

In this method, with defining the relation between the steps along a trajectory is started,

$$\delta x(t) = e^{\lambda t}\delta x(0) \tag{2.13}$$

where $\delta x(t) = x_1(t) - x_2(t)$ and $\delta x(0) = x_1(0) - x_2(0)$.

So the Lyapunov exponent is found as:

$$\lambda = \frac{1}{t}\ln\frac{\delta x(t)}{\delta x(0)}. \tag{2.14}$$

By using logarithm, the derivative of the function f(x):

$$\lambda = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\ln\left|\frac{df(x)}{dx}\right|_{x=x_i} \tag{2.15}$$

If the logarithm is dropped, the Lyapunov number (e^{λ}) is obtained:

$$L = \lim_{n\to\infty}\left(\prod_{i=1}^{n}\left|\frac{df(x)}{dx}\right|_{x=x_i}\right)^{\frac{1}{n}} \tag{2.16}$$

This Lyapunov number is simplified by a finite N:

$$\lambda = \frac{1}{N}\sum_{i=1}^{N}\ln\left|\frac{df(x)}{dx}\right|_{x=x_i} \tag{2.17}$$

Let two trajectories with initial points separated by ε_0 that result in a separation of ε_n after n iterations x_0, $x_0+\varepsilon_0 \rightarrow x_n$, $x_n+\varepsilon_n$ and λ for one dimensional map is calculated as:

$$\lambda = \lim_{n\to\infty}\frac{1}{n}\ln\frac{\varepsilon_n}{\varepsilon_0} \tag{2.18}$$

For an analytic approach, $\varepsilon_n = f^n(x+\varepsilon_0) - f^n(x)$ is taken and the equation becomes:

$$\lambda = \lim_{n\to\infty}\lim_{\varepsilon_0\to\infty}\frac{1}{n}\ln\frac{f^n(x+\varepsilon_0)-f^n(x)}{\varepsilon_0} = \lim_{n\to\infty}\frac{1}{n}\ln\left|\frac{df^n}{dx}\right| \tag{2.19}$$

This method is successful for one dimesional systems. Since the related systems are generally higher dimensional, Jacobian approach is often used (Penner, 2004).

2.5.2 Calculation of Lyapunov exponents by Jacobian approach

In this approach, local Jacobian matrices are estimated and then the long term product of matrices is computed. Let the discrete dynamical system be as below:

$$\underline{x}_{k+1} = F(\underline{x}_k) \text{ for k=0,1,…} \tag{2.20}$$

$\underline{x}_k$: the state vector

F(.) : continuously differentiable nonlinear function. Then the linearized system around the trajectory in the phase space can be rewritten as:

$$\underline{\delta x}_{k+1} = J_k . \underline{\delta x}_k \tag{2.21}$$

$$J_k = \left. \frac{\partial F}{\partial \underline{x}} \right|_{\underline{x}_k} \in R^{m \times m} \text{ for k=0,1,…} \tag{2.22}$$

J_k: Jacobian matrix in point k. In the multiplication stage, let $Y^k = J_{k-1} . J_{k-2} ... J_0$, then the symmetric positive definite matrix is as below:

$$\Lambda = \lim_{k \to \infty} ((Y^k)^T . Y^k)^{\frac{1}{2k}} \tag{2.23}$$

and Lyapunov exponents are the logarithms of eigenvalues. But this method has some disadvantages; Y^k may go to very large values for large k and the linear independency of the columns should maintain for the computation of Y^k. QR factorisation algorithm is used to manage these problems.

The steps of QR factorisation algorithm are as below:

1. Orthogonal Q_0 such that $Q_0{}^T . Q_0 = I$ is given.
2. $Z_{k+1} = J_k . Q_k$, k=0,1,… is solved and the decomposition $Z_{k+1} = Q_{k+1} . R_{k+1}$ is obtained.

   Q_{k+1} : an orthogonal matrix

   R_{k+1} : upper triangular matrix with positive diagonal elements.
3. At last, Lyapunov exponent can be calculated:

$$\lambda_j = \lim_{k\to\infty} \frac{1}{k} \log((R_k)_{jj} \ldots (R_1)_{jj}) \quad \text{for j=1,…,m} \qquad (2.24)$$
$$= \lim_{k\to\infty} \sum_{i=1}^{k} \log((R_i)_{jj})$$

(Ataei, Khaki-Sedigh, Lohmann, & Lucas, 2003).

2.6 Symbolic Time Series and Entropy

When we are analyzing a time series data, our main objective is to extract the features of the system and determine the governing rule. One common method that is used for this purpose is symbolic time series analysis that is also related to the mathematical discipline symbolic dynamics first introduced by Hadmard in 1896. This method converts the measured time series into a corresponding sequence of symbols and gives a symbolic representation of the system. In 1951 C. Shannon discussed analyzing these symbol sequences in his seminal paper "Predictions and Entropy of Printed English". And then, Shannon's method was applied in many areas (Steuer, Molgedey, Ebeling, & Jimenénez-Montano, 2001).

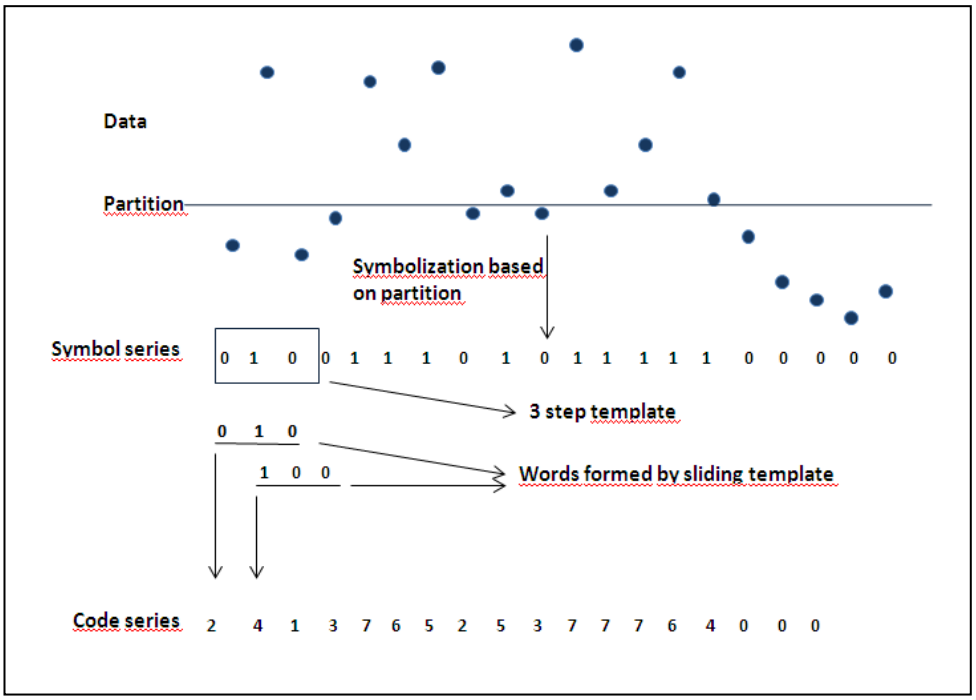


Figure 2.4 Process of symbolizing a real sequence

In Figure 2.4, a symbolic time series analysis process is represented. In 1959, Kolmogorov applied Shannon's theory to symbolic encodings of the phase space of nonlinear dynamical systems.

Let S be a sequence of length N and combined of symbols from a finite alphabet of λ letters. If n letters are formed n-words, $p_i^{(n)}$ is the probability of any n-word(i) to occur. In these conditions, entropies of words of length n are calculated as:

$$H_n = -\sum p_i^{(n)} \log_2 p_i^{(n)}. \tag{2.25}$$

H_n entropies define information necessary to predict a subsequence of length n. Then the conditional entropies h_n , so the average information necessary to predict the next symbol, is computed by the formula:

$$h_n = H_{n+1} - H_n. \quad h_0 := H_1 \tag{2.26}$$

$$h_{n+1} \le h_n.$$

And then the limit of the conditional entropies h_n for large n is defined as below:

$$h_{ks} := \lim_{n\to\infty} h_n = \lim_{n\to\infty} \frac{H_n}{n}. \tag{2.27}$$

This limit entropy h_{ks} called Kolmogorov Sinai entropy is the average amount of information necessary to predict the next symbol when having information about the history of the system (Steuer, Molgedey, Ebeling, & Jimenénez-Montano, 2001).

2.6.1 Other Symbol Sequence Statistics

Although the most important symbol sequence statistics are information theoretical measures such as Shannon entropy and Kolmogorov Sinai (KS) entropy,

estimator Effective Measure of Complexity (EMC) defined by Grassberger also represents the dynamical entropy of the source which generated the time series. Firstly, conditional entropies are calculated by using Equation 2.26. Then the EMC is obtained as below:

$$EMC = \sum_{n=1}^{\infty} n(h_{n-1} - h_n) \tag{2.28}$$

The EMC is a lower bound for the measure of complexity (Daw, Finney, & Tracy, 2002).

Average predictability is also a useful symbol sequence statistics. By using conditional entropies, average predictability of a word (r_n) is obtained as below:

$$r_n = 1 - h_n \tag{2.29}$$

Local predictability and uncertainty for different prediction horizons provide information about the significance of predictions (Velickov, Price, & Solomatine, 2003).

2.6.2 Relation between Lyapunov Exponent and Entropy

The entropy and the Lyapunov exponent are very important and interrelated characteristics of complex dynamical systems as nonlinear or chaotic time series. As the Lyapunov exponents measure the divergence of initial conditions and the entropy quantifies the loss of information about the initial conditions, the relation between entropy and Lyapunov exponent is formulized by Pesin's identity:

$$h_{KS} \leq \sum_{i=1}^{k} \lambda_i \tag{2.30}$$

where k is number of positive Lyapunov exponents.

Pesin's identity identifies upper bound to the KS entropy (Shashidhar, 2006). Pesin's identity is also useful to find the KS entropy since the direct computation of Kolmogorov Sinai entropy fails in lack of data case (Kantz & Schreiber, 2004).

CHAPTER THREE
FORECASTING METHODS

3.1 Introduction

Forecasting future is an important subject in many areas like meteorology, resources or economics in the real life. In time series forecasting, a model describing the future is constructed by analyzing the past time values and this model is used for forecasting. For linear systems, forecasting stage is relatively simple. But, most of systems in the real life has behaviours that are irregular, unpredictable and seemed as random, in other words, chaotic. So nowadays, the studies of forecasting chaotic systems are very important. Nonlinear time series analysis methods make these series to be predictable by benefitting from chaos theory and dynamical systems and so using system state observations.

Local linear models have been popular to predict nonlinear time series, since these series can be predicted in short term because of the governed attractor. Especially Casdagli's paper is one of the most significant studies in this context. Casdagli (1992) introduced a technique for forecasting and investigating the amount of the irregularities in low dimensional chaotic, stochastic or high dimensional chaotic dynamics. He also applied the technique to a variety of time series data (experimental or natural).

According to Casdagli's method, the dynamics of the series is first reconstructed in the phase space with appropriate embedding dimension m and delay time τ. Local linear models are then fitted using the k nearest neighbours of the delay vector x_i. The goal of this study is to investigate the effects of parameter changes and the forecasting method of the dynamical system to the error functions. Forecasting steps of Casdagli's method are briefly presented below:

1. Firstly, the related series (N) is divided to two parts as a fitting set $(\boldsymbol{x_1}, \dots, \boldsymbol{x_{N_f}})$ and a testing set $(\boldsymbol{x_{N_{f+1}}}, \dots, \boldsymbol{x_{N_f}} + \boldsymbol{x_{N_t}})$.

2. Delay parameters (*m*: embedding dimension, τ: delay time) and forecasting time *T* are decided.

3. A delay vector $\boldsymbol{x_i = x_i, x_{i-\tau}, \dots, x_{i-(m-1)\tau}}$ is chosen for $\boldsymbol{i \geq N_f}$ to make *T*-step ahead forecasts.

4. Then using one of the distance calculation methods, the distances $\boldsymbol{d_{ij}}$ between the test vector $\boldsymbol{x_i}$ and the delay vectors $\boldsymbol{x_j}$ in the fitting set are calculated for $\boldsymbol{1 + (m-1)\tau \ll j \ll N_f - T}$.

5. The distances $\boldsymbol{d_{ij}}$ are ordered from minimum to maximum and *k* nearest neighbours $\boldsymbol{x_{j(1)}, x_{j(2)}, \dots, x_{j(k)}}$ **of** $\boldsymbol{x_i}$ are found for selected *k* value. $\boldsymbol{2(m+1) < k < N_f - T - (m-1)\tau}$.

6. A model is fitted as Equation 1 and the parameters ($\boldsymbol{\alpha_0, \dots, \alpha_n}$) are found by using Ordinary Least Square Method:

$$x_{j(l)+T} \approx \alpha_0 + \sum_{n=1}^{m} \alpha_n \, x_{j(l)-(n-1)\tau} \quad l=1,\dots,k \tag{3.1}$$

7. T step-ahead forecasts $\hat{x}_{i+T}(k)$ for $\boldsymbol{x_i}$ vectors are estimated by Equation 3.1 and errors of these $e_i(k) = |x_{i+T} - \hat{x}_{i+T}(k)|$ are obtained.

8. By using steps 3-7, the errors are computed for all *i* and so normalized root mean square error values ($E_m(k)$) are found by Equation 3.2 (Casdagli, 1991).

$$E_m(k) = \{\sum_i e_i^2(k)\}^{1/2} / \sigma \tag{3.2}$$

If *k* equals to the total size of the fitting set, the model becomes a global linear model. In this study, *T* (lead time) is determined as one. Maximum norm distance is used to compute the distances due to computation effectiveness.

In statistics, regression analysis is a common method that uses the normal equations for forecasting. As for Ordinary Least Square (OLS) is a method that estimates the unknown parameters in a linear regression model. OLS method minimizes the sum of squared errors between the observed and the predicted values assuming:

1. Linearity of parameters
2. No linear relationship between independent variables
3. Errors having zero mean, constant variance
4. Independency and identically distribution of error terms

Partial Least Squares (PLS) is also a regression method that models relations between sets of observed variables through latent variables. PLS assumes that the observed data belongs to a system that has a few latent variables and creates orthogonal latent vectors (components) by maximizing the covariance between different sets of variables (Rosipal & Kramer, 2006). Partial least square methods are divided to two basic groups; Linear Partial Least Square and Nonlinear Partial Least Square. Linear PLS technique has been successfully employed for prediction.

In this study, Ordinary Least Square and Linear Partial Least Square Methods are discussed as linear forecasting methods. Quadratic Partial Least Square (QPLS), Neural Network Partial Least Square (NNPLS) and Kernel Partial Least Square (KPLS) that are the nonlinear partial least square methods are also examined. On the other hand, artificial neural network models that are commonly used in nonlinear forecasting are dealt. Since presence of high multicollinearity among the lagged variables is decided, PLS is adapted to Casdagli's method in this study. ANN that is a common nonlinear time series forecasting method is also used to compare the models error results.

3.2 Ordinary Least Square

Ordinary Least Square is a simple regression method that minimizes the residuals between the predicted and observed values while estimating the dependent variable linearly. The linearity means that both parameters and variables are dependent linearly. The linear model for a dependent variable and *p* independent variables is formulized as:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \epsilon_i. \quad (3.3)$$

i: observational unit
n: sample size
X: independent variable
Y: dependent variable
p: number of independent variables
β_j (j=0,...,p): model parameters

The linear model can also be briefly written as:

$$Y = X\beta + \epsilon, \quad (3.4)$$

In matrix notation, the normal equations are indicated as below:

$$X'X\hat{\beta} = X'Y. \quad (3.5)$$

Though the normal equations are always consistent, they always have a solution:

$$\hat{\beta} = (X'X)^{-1}X'Y. \quad (3.6)$$

Then, the vector of dependent variable Y for the matrix of the independent variables is calculated as:

$$\hat{Y} = X\hat{\beta}. \tag{3.7}$$

At last, the residuals vector that is the difference between the observed Y and the estimated $\hat{Y}$

$$e = Y - \hat{Y} \tag{3.8}$$

(Rawlings, Pantula, & Dickey, 1989).

The residuals and the other performance measures related with the residuals are used to decide how the model fits the data and represent the performance of the method.

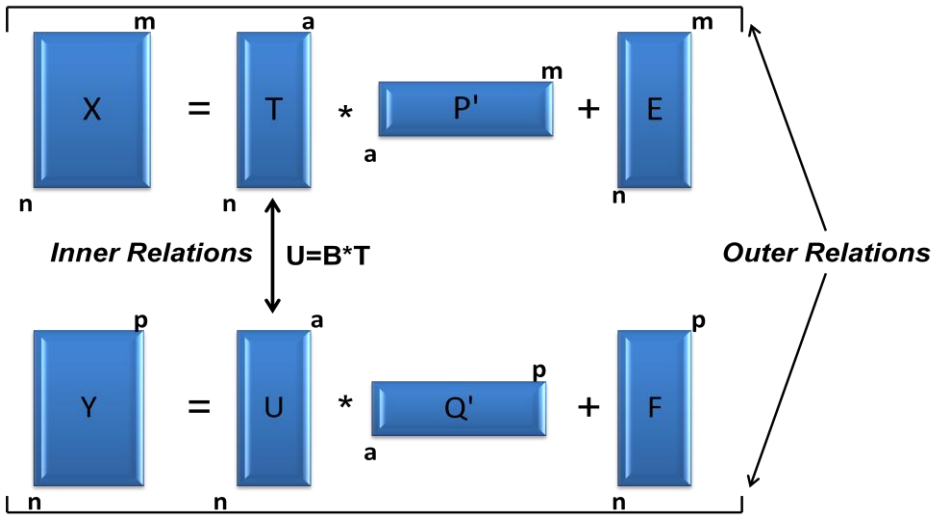
3.3 Partial Least Square

PLS is a robust and reliable method to model the relationship between one or multiple response (Y) and large number of X variables in the collinear data case that any two of columns are linearly dependent and the inverse of the matrix is non-existent. PLS was first introduced by H. Wold in econometrics and has become an important technique in many areas like social sciences, chemometrics, biology, neuroimaging and geosciences. PLS eliminates collinearity by transforming the predictor variables to a set of orthogonal variables. So, this method avoids larger differences between the forecasts and real values that arise from oversize of variances and has been successfully employed for prediction (Fantoni et al., 2002; Bhattacharya, 2005).

PLS is a very powerful tool applied in many areas of research and industry. Computation and implementation of PLS is simple and this is the first strong aspect of this approach. Second important aspect of PLS is representing high-dimensional data through the set of extracted latent variables (Rosipal & Kramer, 2006). Partial least square methods are examined by two basic groups; Linear Partial Least Square and Nonlinear Partial Least Square.

3.3.1 Linear PLS

Linear PLS is a good predictive method that relates two data matrices, X and Y, by a linear multivariate model and also models the structure of X and Y while considering outer relation, inner relation, exchange of scores and mixed relation. PLS decomposes the matrix X and the matrix Y as outer relations as in Figure 3.1 and minimizes the norm of F while keeping the correlation between X and Y by the inner relation U = BT.



X: matrix of predictors
Y: matrix of responses
U: projections of X
T: projections of Y
P and Q: orthogonal loading matrices
E and F: error terms

Figure 3.1 PLS analysis of the relationship between X and Y matrices (Geladi & Kowalski, 1986)

The outer relation step is similar to Principal Component Analysis (PCA) in terms of decomposition of deflation of X and Y data matrices into their principal components. But, only the X matrix is deflated in PLS:

$$X = TP' + E = \sum t_h \, p_h{}' + E$$
$$Y = UQ' + F = \sum u_h \, q_h{}' + F \qquad (3.9)$$

where E and F are the residual matrices which are to be minimised. The principal components of PLS are called Latent Variables (LVs) (Bhattacharya, 2005). The optimum number of LVs is determined by examining prediction errors and the number of LVs with minimum prediction error is chosen.

3.3.1.1 NIPALS Algorithm

Although there are several different algorithms of performing a partial least squares regression, a common used iterative algorithm is the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm. So, NIPALS algorithm is used as PLS calculation steps in this study. NIPALS algorithm finds the two project vectors of *w* and *c*, meeting that

$$[cov(t,u)]^2 = [cov(Xw,Yc)]^2 = max_{|r|=|s|=1}[cov(Xr,Ys)]^2 \qquad (3.10)$$

where $[cov(t,u)] = t'u/n$ is the covariance between the latent component vectors *t* and *u*.

NIPALS algorithm steps:

1. Firstly, the Y-latent-component vector *u* is initialized and the following process is repeated until convergence:

$$w = \frac{X'u}{u'u} \quad \|w\| \to 1 \qquad (3.11)$$
$$t = Xw$$
$$c = \frac{Y't}{t't} \quad \|c\| \to 1$$
$$u = Yc$$

The project vector *w* corresponds to the first eigenvector of the following eigenvalue equation:

$$X'YY'Xw = \lambda w \tag{3.12}$$

The X-latent-component vector t can be obtained as $t=Xw$, which is also the first eigenvector about the equation $XX'YY't = \lambda t$. Latent component vector u can be estimated as $u = YY't$. The above algorithm can be repeated p times with the deflation of X,Y, defined as

$$X \leftarrow \mathrm{X} - \mathrm{tt}'\mathrm{X} \qquad Y \leftarrow \mathrm{Y} - \mathrm{tt}'\mathrm{Y} \tag{3.13}$$

2. Then T, U, the $(N \times p)$ matrix W and the $(M \times p)$ matrix C can be obtained, consisting of the columns from the vectors $\{t_i\}_{i=1}^{p}, \{u_i\}_{i=1}^{p}, \{w_i\}_{i=1}^{p}$ and the PLS regression model can be written in matrix form

$$\hat{Y} = XB_{pls} \tag{3.14}$$

where coefficient matrix B_{pls} can be calculated as

$$B_{pls} = W(P'W)^{-1}C' \tag{3.15}$$

and P is the (N × p) matrix consisting of loadings vectors $\{p_i = X't_i/(t_i't_i\}_{i=1}^{p}$ (Liu et al., 2006). In Equation 3.1, α values will be changed with corresponding B_{pls} values and $x_{(j(l)+T)}$ values will be changed with corresponding Y values.

3.3.1.2 Quadratic Partial Least Square Method (QPLS)

In 1989, quadratic polynomial modeling method was developed by S. Wold. Quadratic Partial Least Square (QPLS) calculation equations are the same as the linear PLS except linking inner relation by nonlinear polynomial functions. Assume the matrix X and Y are the independent data matrix and the dependent data matrix respectively. QPLS model equations are given below:

$$X = t_1 p_1^{'} + t_2 p_2^{'} + \cdots + t_A p_A^{'} + E \tag{3.16}$$
$$Y = u_1 q_1^{'} + u_2 q_2^{'} + \cdots + u_A q_A^{'} + F$$

where t_A is the latent variable for the independent data matrix X, u_A is the latent variable for dependent data matrix Y. p_A and q_A are also defined as the loading vector corresponding to t_A and u_A and E and F are the model residuals of X and Y respectively.

k^{th} components latent variables, t_k and u_k, are calculated using formulas as follows:

$$t_k = \{X - (t_1 p_1^{'} + t_2 p_2^{'} + \cdots + t_{k-1} p_{k-1}^{'})\} w_k \tag{3.17}$$
$$u_k = \{Y - (u_1 q_1^{'} + u_2 q_2^{'} + \cdots + u_{k-1} q_{k-1}^{'})\} c_k$$

The nonlinear inner relation is the important and determinative part of the QPLS modeling and it is as follows:

$$u_{ja} = b_{0a} + b_{1a} t_{ja} + b_{2a} t_{ja}^2 + h \tag{3.18}$$
$$u_{ja} = r_{ja} = b_{0a} + b_{1a} t_{ja} + b_{2a} t_{ja}^2$$

All parameters are determined to minimize the sum of squares for residuals of inner relation (Kimura, Miyashita, Funatsu & Sasaki, 1996).

3.3.2 Nonlinear PLS

The linear PLS approach is always not effective due to the inability of linear regression techniques to account for nonlinearity and dynamics. There are several different nonlinear PLS methods in the literature. Nonlinear PLS methods can be classified into two groups: Type I and Type II.

Type I: Nonlinear projection of X:

The observed X matrix of independent variables is projected onto a nonlinear surface but the inner relation between score vectors t and u is kept linear in Type I group models. It is performed by mapping the data by a nonlinear function to a new representation like the expansion of the X matrix by considering component square terms x_i^2 and cross terms $x_i x_j$ of the input vector x (Lodhi &Yamanashi, 2010).

Type II: Nonlinear inner relation:

The nonlinear relation between X and Y data sets is modeled by a nonlinear form in Type II ones. S.Wold et al. (1989) expansed the linear PLS model to nonlinear by replacing the linear inner relation between the score vectors *t* and *u*:

$$u = g(t) + h = g(X, w) + h \tag{3.19}$$

where g and h respectively represent a continuous nonlinear function and h denotes residuals vector. g can be modelled by related methods like polynomial functions and artificial neural networks. Keeping the score vectors *t* and *u* are linear projections of the original variables is assumed in this method. Since Type I is easily operable, computationally easy and fast and able to model even complex nonlinear relations, it looks advantageous. But, it generally causes to lose of interpretations compared to the original data (Rosipal, 2011).

3.3.2.1 Neural Network Partial Least Square Method (NNPLS)

Qin and McAvoy (1992) and Qin (1993) suggested Neural Network PLS (NNPLS) algorithm that have both neural networks approximation abilities and robustness of PLS regression. NNPLS combines the neural networks with PLS to model nonlinear systems contain collinearity. The input and output variables are projected onto the latent space as classical PLS method, but then each latent variable pair is mapped with a neural network as follows:

$$u_i = N(t_i) + v_i \tag{3.20}$$

where N(.) represents the inner relation performed by a neural network and *v* represents the residuals (Lee et al., 2006).

3.3.2.2 Kernel Partial Least Square Method (KPLS)

Kernel Partial Least Square (KPLS) nonlinearly maps original *X*-space data into a feature space *F* corresponding to a reproducing kernel Hilbert space as Equation 3.21 and then develops linear PLS models in this space.

$$x \in X \rightarrow \Phi(x) \in F \tag{3.21}$$

This mapping into a feature space *F* provides less linear algebra operations and simplicity, and it is called as kernel trick. Because, the kernel function can compute a dot product between two vectors in *F* as Equation 3.22 and Equation 3.23 and *i*, *j* values are equal to the values of the kernel function:

$$k(x, y) = \Phi(x)'\Phi(y), \ \forall\, x, y \ \in X \tag{3.22}$$

$$K(x_i, x_j) = \Phi(x_i)'\Phi(x_j) \tag{3.23}$$

Gram matrix K, $K = \Phi\Phi'$, is stated as the cross dot products between all mapped input data points, where Φ is the matrix of mapped *X*-space data (Rosipal & Kramer, 2006).

After extraction of *u* components, the deflation of K matrix is obtained as:

$$K = (I - uu')K(I - uu') \tag{3.24}$$

I: an m-dimensional identity matrix.

Then, the training data prediction of KPLS model considering normalized scores u is formulized as:

$$\hat{Y} = KW(U'KW)^{-1}U'Y = UU'Y \tag{3.25}$$

So, new observations can be predicted by the regression model below:

$$\hat{Y}_t = K_t W(U'KW)^{-1}U'Y \tag{3.26}$$

where K_t is the test matrix and its elements are $K_{ij} = K(x_i, x_j)$, x_i: the testing data, $x_{j:}$ the training data (Jo et al., 2010).

KPLS is a strong method owing to its simplicity stated as a nonlinear data in the original space is most likely to be linear after high-dimensional nonlinear mapping as stated in Cover's theorem (Jo et al., 2010). And KPLS has advantage of not involving nonlinear optimization as all kernel-based algorithms and requiring only linear algebra as linear PLS (Noorizadeh et al., 2011).

3.3.3 PLS Variable Selection Method

Variable selection is important for determining the best prediction model in forecasting. Original PLS does not have variable selection step, but there are several variable selection methods that are efficient in finding the appropriate variables in PLS.

Variable selection methods that are used in PLS can be basically classified to three categories as below and Figure 3.2:

• Filter methods: These methods use the PLS output to decide a subset of important variables for variable identification.
• Wrapper methods: These methods refit the variables in PLS that are selected by filter methods for obtaining reduced models.

• Embedded methods: These methods do the variable selection at component level (Mehmood *et al.*, 2012).

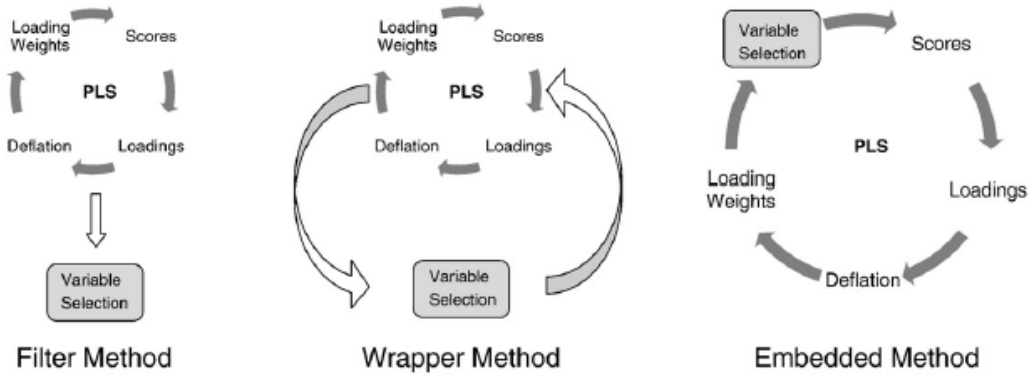


Figure 3.2. Illustration of PLS variable selection methods (Mehmood *et al.*, 2012)

Since its computation simplicity and time performance, filter method was used for variable selection in this thesis.

3.4 Artificial Neural Networks (ANN)

Artificial Neural Network (ANN) is a collection of interconnected neurons like human brain that learns from their data and it is relatively successful to model especially nonlinear patterns. Neural networks that are universal approximators keep information in neurons that operate in parallel and are connected into a network by way of weights and then in learning process networks obtain knowledge, at last network can make predictions by this obtained information (Samarasinghe, 2007).

In this study, ANN is used for nonlinear time series forecasting and nonlinear relationship between the current and next observations by ANN can be formulized as:

$$y_t = f\left(y_{t-1}, y_{t-2}, \dots, y_{t-p}\right) + \varepsilon_t \tag{3.27}$$

where $f(y_{t-1}, y_{t-2}, \dots, y_{t-p})$ is the nonlinear function and ε_t is the error that is a random variable has a mean of zero and variance σ^2 (Samarasinghe, 2006). The parameters of a neural network are identified in training step and the most popular algorithm is back-propagation. But suggesting the best learning algorithm is difficult and hence the successful training often needs substantial amount of experience (Kantz & Schreiber, 2004).

ANNs can be classified into two categories as dynamic and static. While static networks don't have feedback mechanism, dynamic networks don't only consider the current input, but also the current or previous inputs, outputs, or states. Since dynamic networks have memory advantage and they can be trained to learn time varying patterns or series like our condition, they are more powerful than static networks. Dynamic networks can also be divided with regard to connection directions into two categories: those that have only feedforward connections and those that have feedback (recurrent) connections (Demuth et al., 2008). In this study, Focused Time Delay Network and NARX Network that are popular dynamic networks and are powerful in nonlinear time series forecasting are examined.

3.4.1 Focused Time-Delay Neural Network

Focused Time Delay Neural Network (FTDNN) is a direct dynamic network and consist of a feed forward network with a tapped delay line at the input layer. In these networks, the dynamic appear only at the input layer of a static multilayer feed forward network (Abed, El-Shafie, & Osman, 2010). In Figure 3.3, the architecture of FTDNN and single neuron calculations are represented.

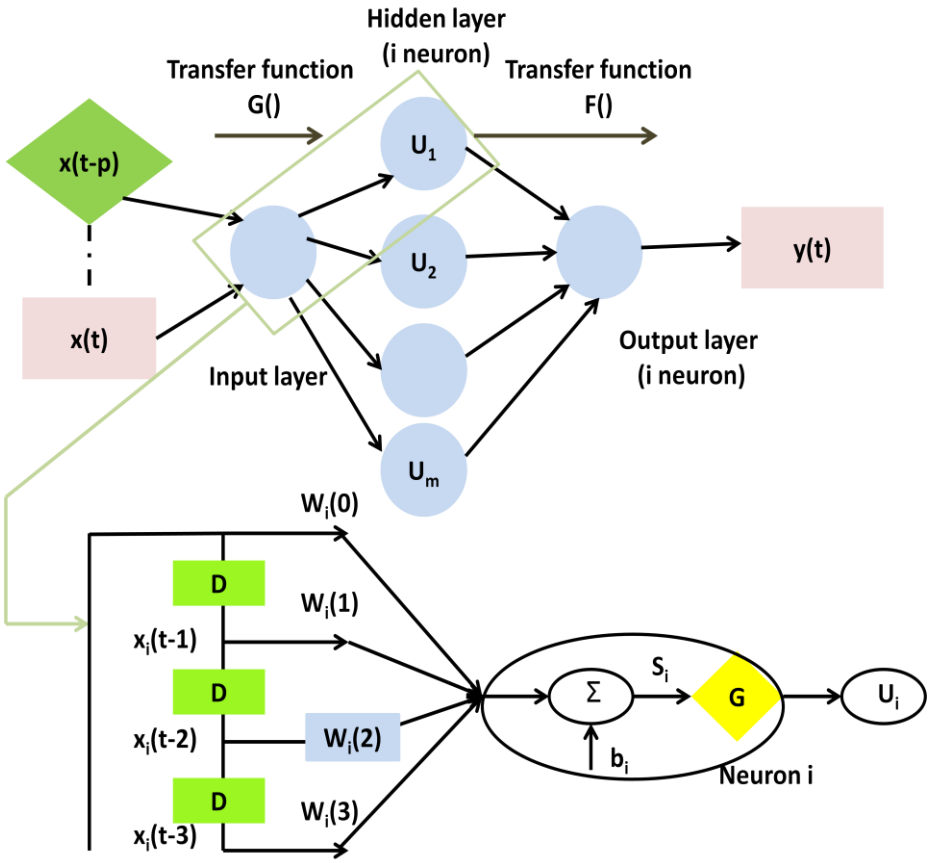


Figure 3.3 General architecture of FTDNN (Abed et al., 2010)

As seen in Figure 3.3, FTDNN has two components: a memory structure and a nonlinear associator. While the memory structure is a time delay line includes the *p* most recent inputs generated by the delay element (D) and so keeps the related past information, nonlinear associator used this memory to make the future predictable by conventional feed-forward network (Demuth et al., 2009). In these networks, input variables enter to a feedforward network, at time *t* are fed again with unit time delay at the next time step until the required lag length. Focused time delay means that the entire memory structure is located at the input end of a neuron that operates as a filter of short term response, takes lagged inputs and embeds that memory structure into its own weights (Samarasinghe, 2006).

FTDNN is a powerful and explicit method in time series prediction, since the output of the network depends not only on the connection weights and the current input signal, but also on the previous states of the network. FTDNN has also an

advantage that it does not require dynamic backpropagation to compute the network gradient. Because, the tapped delay line appears only at the input of the network, and contains no feedback or adjustable parameters. As a result, this network has faster training performance than other dynamic networks (Demuth et al., 2009).

3.4.2 Spatio-Temporal Time-Lagged Networks

Spatio-Temporal Time-Lagged Networks are commonly used in time series forecasting case is affected by one or more time series and needs association of memory dynamics of each for time series forecasting of such interaction. These networks incorporate such interacting time series effects by filtering short term memory response for each spatial dimension. Filtered short term memories are received by the neuron via the weights a_{mi} (Samarasinghe, 2006).

The short term memory is embedded in the weighted sum of the time lags ($s_{mi}(t)$). For example, the weighted sum for input one is computed as:

$$s_{m1}(t) = \sum_{l=1}^{p} a_{m1}(l) z_{t-l} \tag{3.28}$$

where $a_{m1}(l)$ is the weight from input one with lag one to neuron m and z_{t-l} is the input corresponding to lag one. The total weight vector a_{m1} and the input vector of input one are respectively defined as below:

$$a_{m1} = [a_{m1}(1), a_{m1}(2), \dots, a_{m1}(p)] \qquad z = [z_{t-1}, z_{t-2}, \dots, z_{t-p}] \tag{3.29}$$

By summing Equation 3.28 for i=1,...,n, the total weighted sum of short-term responses for time t, ($S_m(t)$) for that neuron are calculated as:

$$S_m(t) = \sum_{i=1}^{n} s_{mi}(t) \tag{3.30}$$

Nonlinear transfer function is used for the total weighted sum uses and the nonlinear response is obtained from the nonlinear processing in the hidden neurons.

After training, the temporal dynamics of nonlinearly interacting time series are embedded in its weights. The spatio-temporal time lagged network is a nonlinear extension of a NARX model that is a class of nonlinear AR model with endogenous and exogenous input variables (Samarasinghe, 2006).

The nonlinear autoregressive network with exogenous inputs (NARX) is a recurrent dynamic network with feedback connections. The simply equation of the NARX model is as below:

$$y_{(t)} = f(y_{(t-1)}, y_{(t-2)}, ..., y_{(t-n_y)}, u_{(t-1)}, u_{(t-2)}, ...; u_{(t-n_u)}) \quad (3.31)$$

where the next value of the dependent output signal ($y(t)$) is regressed on previous values of the output signal and previous values of an independent input signal ($u(t)$) (Demuth et al., 2008).

3.5 Proposed Methods of the Thesis

3.5.1 Adapted Partial Least Square Method

In the adapted PLS method, values of α coefficients in Casdagli's method that is discussed in Forecasting Methods issue with the aforementioned B_{pls} values and $x_{(j(l)+T)}$ values correspond the Y values were changed and PLS was applied to these data. Because recent studies show that regularisation methods as Principal Component Analysis, PLS and Ridge Regression give better forecasting results than OLS. In nonlinear time series context, the advantage of PLS may be avoiding multi collinearity problem even with high embedding dimension values.

3.5.2 Proposed Method

Forecasting steps of the proposed method are briefly as below:

1. Firstly, the delay parameters; embedding dimension (m) and delay time (τ) of

the explatory and predicted variables are obtained. Then, for these variables delay variables as $y_t, y_{t-\tau}, \dots, y_{t-m\tau}$ and $x_{i,t}, x_{i,t-\tau}, \dots, x_{i,t-m\tau}$ are constructed.

2. And then, the most significant variables among these variables are found by variable selection methods related to PLS method. Since its computation simpleness and time performance, filter method was used for variable selection in this thesis.

3. The regression model that used these selected variables and obtains the one time ahead forecasts by using Partial Least Square Method is constructed:

$$\hat{y}_{t+1} = b_0 + b_{1,y} y_t + \sum_{n=1}^{m} b_{n+1,y} y_{t-n\tau} + b_{1,x_{i,t}} x_{i,t} + \sum_{n=1}^{m} b_{n+1,x_{i,t}} x_{i,t-n\tau} \quad (3.32)$$

i:indices of selected input variables ($x_{i,t}$)
$\boldsymbol{b_{1,y}}$:b_1 coefficient of y
$\boldsymbol{b_{1,x_{i,t}}}$: b_1 coefficient of x_i

4. After the beta coefficients are obtained by PLS, one step-ahead forecasts $\hat{y}_{ti+1}$ and errors of these $e_i = |y_{t+1} - \hat{y}_{t+1}|$ are obtained.

5. After the all errors are computed by using steps 1-4, $E_m(k)$ ($k=N_{fitting}$) error performance and the other forecasting performance measures are obtained. So, the forecasting methods can be compared.

3.6 Performance Measures of Forecasting Methods

Determining the model performance is one of the most significant stages in forecasting since forecasts of a model with a low performance will be unmeaningful. There are a lot of performance measures that evaluate forecast model as $E_m(k)$ in Casdagli's method. But Mean Absolute Error (MAE), Mean Square Error (MSE) and Root Mean Square Error (RMSE) are the most common ones and these are

respectively computed as below:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{3.33}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{3.34}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{3.35}$$

where N, y and $\hat{y}$ are respectively number of observations, observed values and forecasted values. RMSE, MAE, and MSE values that are equal to zero is the most ideal condition and mean a perfect fit. As these values increase, the performance of the model decreases.

CHAPTER FOUR
AN APPLICATION

4.1 Introduction

Weather, especially temperature forecasting has always been a popular subject since its effects are valid for all of us and meteorological data include uncertainty mostly chaos. Verification of the meteorological forecasts that have nonlinear and non-normal distributions is possible by following and developing the forecasting system and the errors produced by this system. Testing the forecast and creating new hypotheses have increased developing new statistical methods to promote the meteorology knowledge or the tendencies of application of new developed statistical methods to meteorological data. In this sense, Ramage (1993) stated the importance of statistics in his work named "Statistics has generally been accepted as an essential component of weather forecasting". Ebert et al. (2011) investigated the historical development process of the statistical methods used the extant meteorological studies and emphasized the significance of diagnostic verification.

Gluhovsky and Agee (2007) reported that the time series is crucial to describe and forecast the atmospheric and climate changes, but it is insufficient to understand the nature of the physical processes that do not satisfy the assumptions of conventional time series analysis. They dealt with invalidity of the statistical inferences based on the linear models and studied on analysis of meteorological incidences by nonparametric techniques.

After the studies that presented invalidity of the statistical methods based on linearity and normality (Bountis et al., 1993; Katz, 2002, Percival et al., 2004), meteorological data that have chaotic and irregular behaviours have been tried to analyze by nonlinear time series methods. Sivakumar et al. (1999) worked on forecasting rainfall and analyzed the chaotic behaviour in the series by using data from Singapore. Munuzuri and Gelpi (2000) nonlinearly forecasted cloud coverage time series from four years of cloud absorption data of half hourly Meteosat infrared images of Northwestern Spain, used the forecasted values for daily weather

prediction and compared the results with the real measurements. Tang et al. (2010) analyzed temperature change in China for the past 100 years by combining several surface air temperature time series of the country.

Recent studies about predicting chaotic nature time series have also shown that local linear prediction models perform well too. Currenti et al. (2004) used local and the global linear methods to understand and describe the dynamics of volcano magnetic signals of Etna volcano. Laio et al. (2004) studied on some nature data and tried to infer about the dynamics of the rainfall-runoff transformation considering river flow time series. Dhanya and Kumar (2010) used local method for forecasting daily rainfall data of three regions Malaprabha, Mahanadi and All India for the period 1955-2000.

As distinct from literature, two methods that approach from the different angles to analysis of meteorological observations are combined in this thesis. Nonlinear time series analysis that indicates the predictability dimensions of the meteorological data known with its irregular and chaotic structure is first of these and partial least square model (PLS) for forecasting is the second. In this stage, meteorological data of Istanbul is forecasted by using PLS as both local and global method. In global forecasting context, from an effective variable selection method is benefitted to select other meteorological independent data as rain and humidity and forecast ability is improved. ANN models are also constructed to evaluate the success of proposed method by comparing with benchmark values.

4.2 Analysis

In analysis step, approximately 100 years long meteorological data from 54 automatic meteorology observation stations of Istanbul that is the metropolis city of Turkey is analyzed to infer about the meteorological behavior of Istanbul. Monthly temperature values are firstly tried to be forecasted by own lagged values, then by both own and other meteorological data (humidity and rainfall) lagged values.

4.2.1 Data

In this thesis, monthly mean temperature, monthly mean humidity and monthly total rainfall data between February 1914 and December 2010 (N=1,163) for Istanbul, Turkey is used for forecasting temperature and analyzing the parameter changes and comparing forecasting methods. The meteorological stations of Istanbul which related data is obtained from is seen in Figure 4.1.

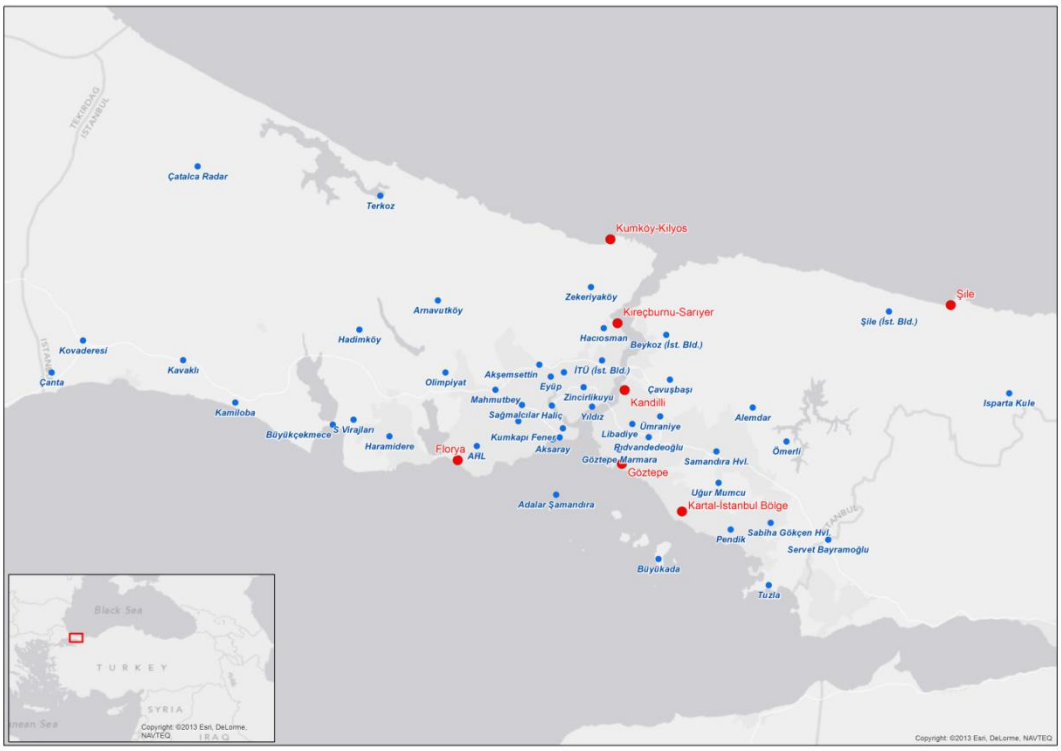


Figure 4.1 Meteorological stations of Istanbul

Altough Istanbul has 54 Automatic Meteorology Observation Stations (blue and red points), continuous data stream is provided from only seven stations (Şile, Göztepe, İstanbul Bölge, Florya, Kumköy, Sarıyer, Kandilli) seen as red in the map. In four of these stations (Florya, İstanbul Bölge, Kumköy, Sarıyer), daily climate that includes rainfall, meteorolojical incidents, evaporation and insolation durations every morning is also observed.

Firstly, boxplots of the variables are drawn by MINITAB 15 Software as in Figure 4.2 to infer about the variability of the variables according to 5-year and recent periods. The information obtained from these plots is useful for determining the reliability of the forecasts.

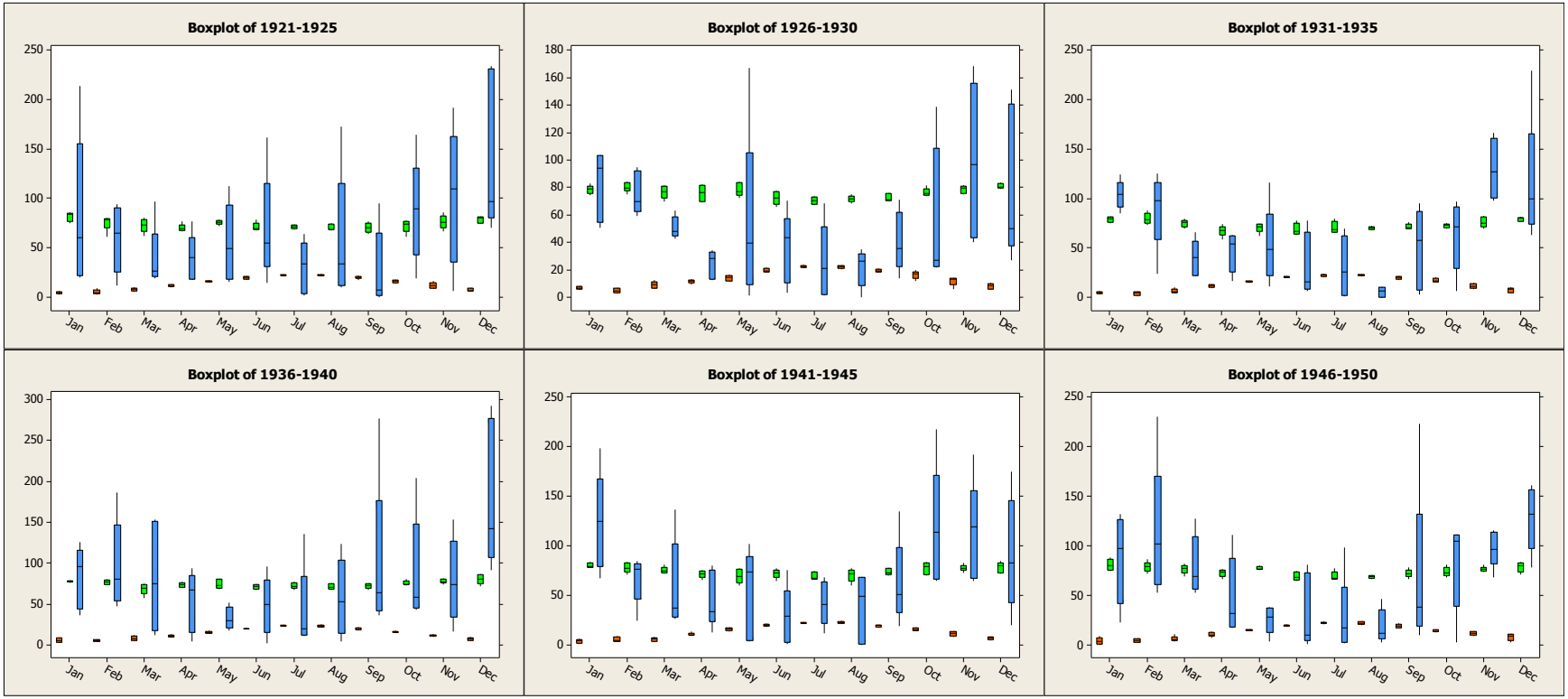


Figure 4.2 Boxplots of meteorological variables (temperature: orange, humidity: green and rainfall: blue) for 5-year periods between 1921 and 2010

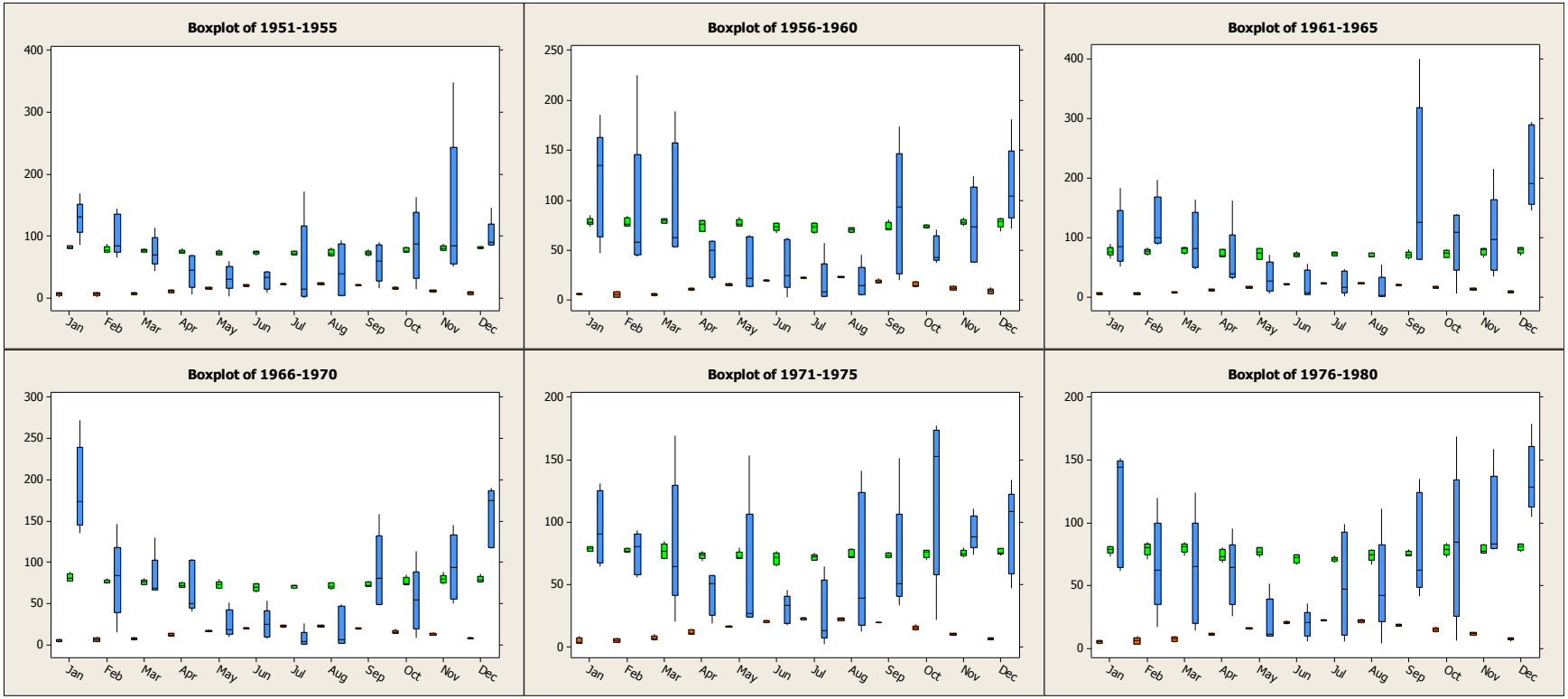


Figure 4.2 Boxplots of meteorological variables (temperature: orange, humidity: green and rainfall: blue) for 5-year periods between 1921 and 2010

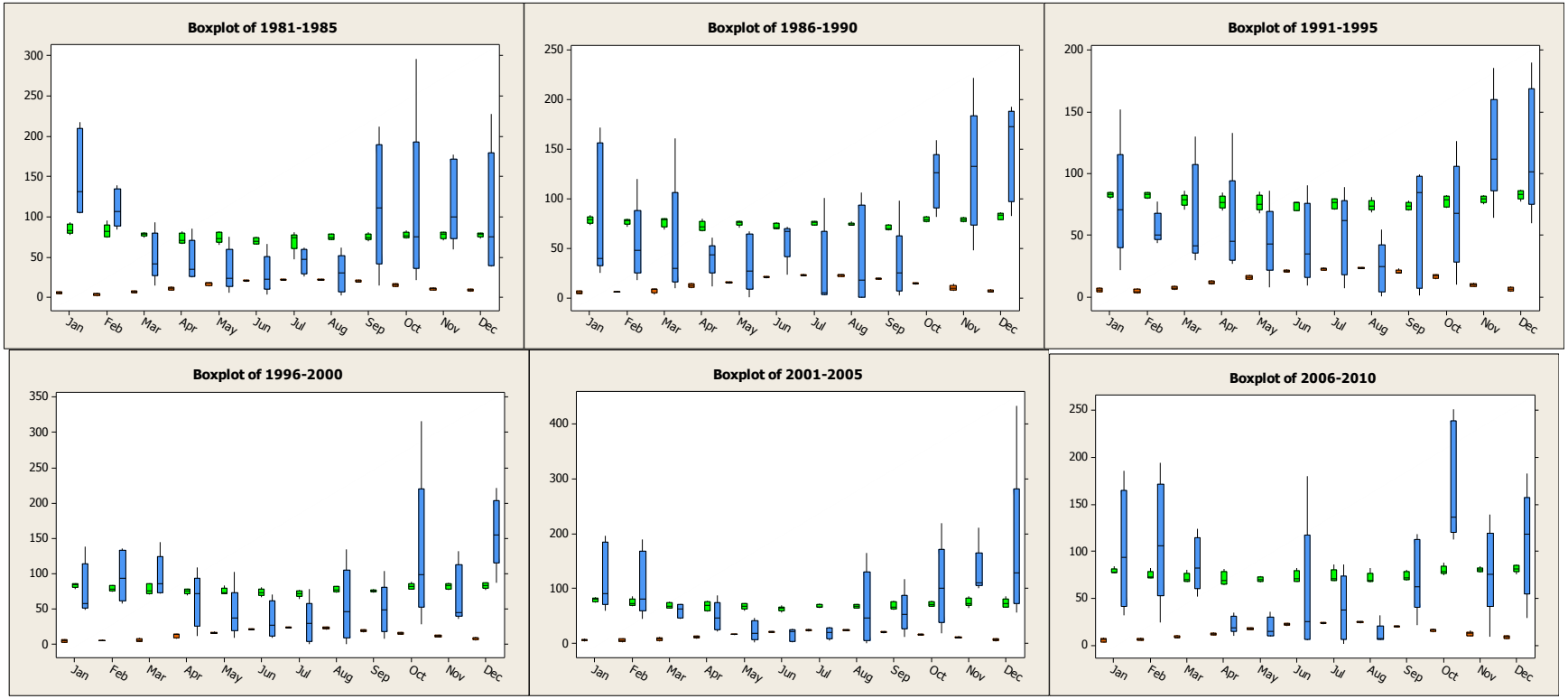


Figure 4.2 Boxplots of meteorological variables (temperature: orange, humidity: green and rainfall: blue) for 5-year periods between 1921 and 2010

Temperature and humidity variables have generally small variations for 5-year periods between 1921 and 2010. In the past periods, temperature has high variations in winter months (especially 1921-1925 November, 1926-1930 October and November, 1946-1950 January and December, 1956-1960 February, 1971-1975 January, 1976-1980 February). In recent periods (2001-2010), low variations are observed for temperature in summer (especially July and August and 2006-2010 June). Humidity has generally small variations in spring and summer for past periods. Humidity reachs minimum variations in 1926-1930 April, May and June, 1981-1985 July and 1991-1995 May. But it has more variation than previous periods in 2006-2010 April, June and July. Rainfall variable has extremely high variations in contrast to the other variables. In 1926-1930 November and December, 1961-1965 September and 1976-1985 October, it has highest variations and it also has higher values than previous periods for 2006-2010 January, February and June. Rainfall has the lowest variability in 1931-1935 August and 1966-1970 July and also has lower values for 2001-2005 June and July and 2006-2010 April, May and August than previous periods. In conclusion, rainfall variable has much high variations in comparison with temperature and humidity. Humidity has recently (2006-2010) higher variations for only April, June and July. Since the recent values are more significant in forecasting and the higher variations indicate that the data values are wide spread, these inferences will be useful in forecasting stage.

Time series plots which visualize simply behaviours of data sets and are first diagnosis for series are respectively obtained in Figure 4.3 for temperature, humidity and rainfall.

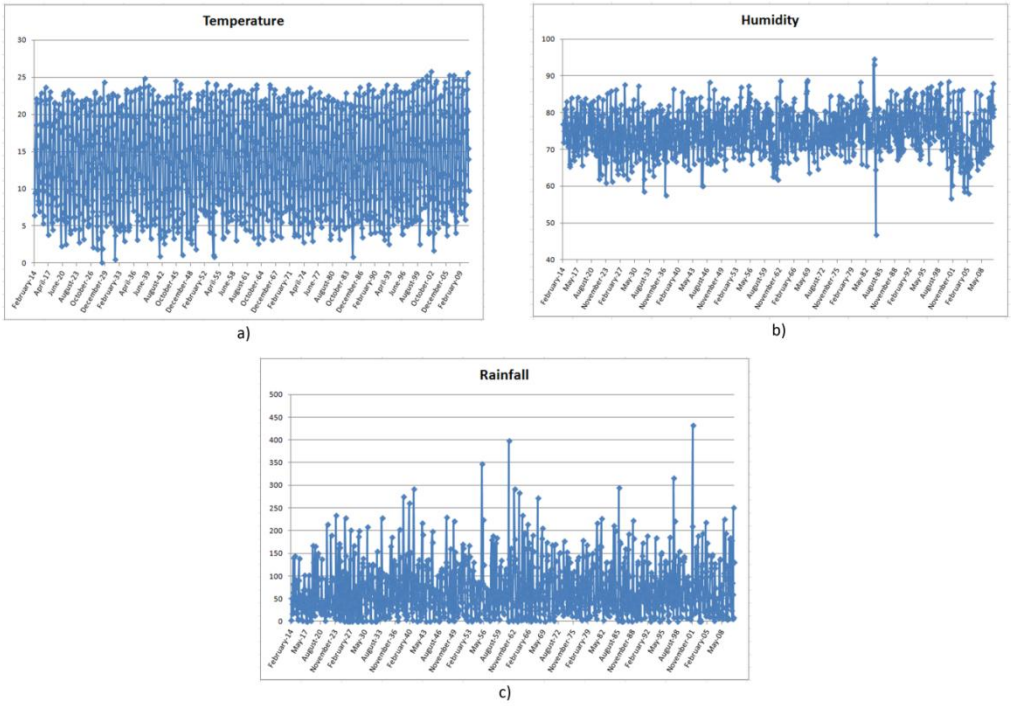


Figure 4.3 Time series plots of monthly temperature, humidity and rainfall data for Istanbul in February 1914-December 2010.

When the plots are investigated, it is realized that temperature and humidity data have periodic behaviours as distinct from rainfall that has also different data structure as monthly total amount.

4.2.2 Nonlinear Time Series Analysis

Lyapunov exponent values for the meteorological data are computed and maximum Lyapunov values are respectively found 0.2782, 0.9827, 1.9707 for temperature, humidity and rainfall. Since this value was greater than zero, the chaos presence of related data was noticed. Figure 4.4 that include reccurence plots of these variables also demonstrates these Lyapunov values and presence of chaos.

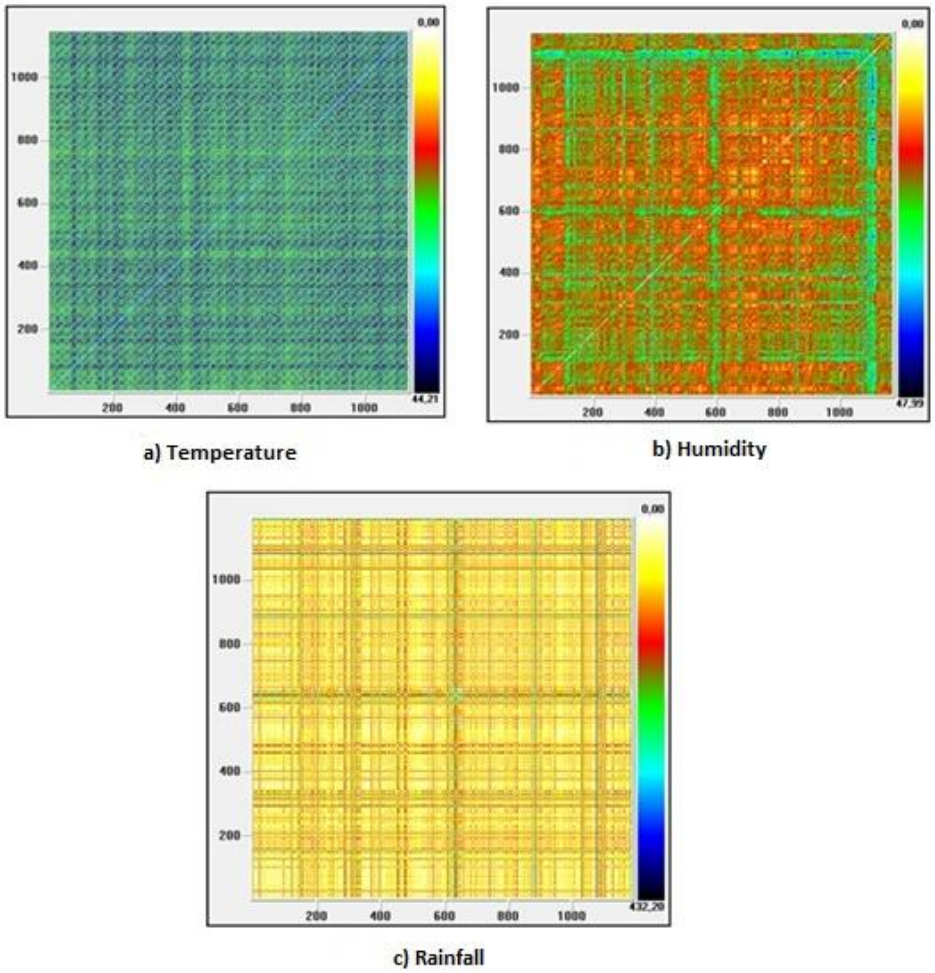


Figure 4.4 Recurrence plots for temperature, humidity and rainfall data

As Figure 4.4, the views of temperature, humidity and rainfall recurrence plots prove that they belong to chaotic systems. As it is mentioned before, chaotic systems are complex systems that must be analyzed by reconstructing the system in a phase space. For this purpose the parameters of the new space were determined by Mutual Information and False Nearest Neighbours methods.

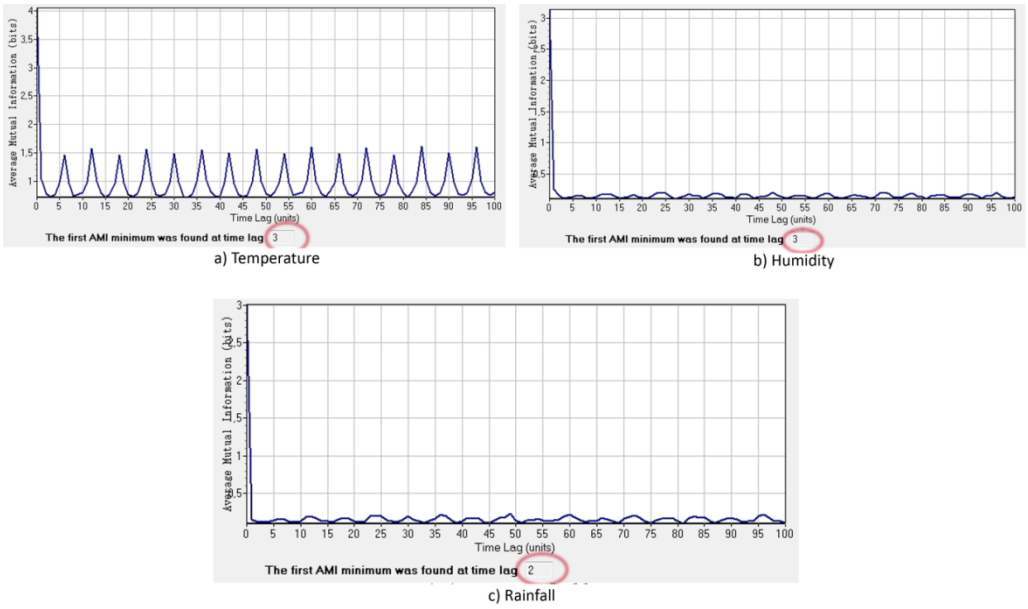


Figure 4.5 Mutual Information graphs

Optimum delays for Temperature, Humidity and Rainfall are respectively obtained 3, 3 and 2 by examining the first minimum values of the graphs as seen in Figure 4.5. Mutual Information method (Equation 2.7) is used for calculating optimum delays.

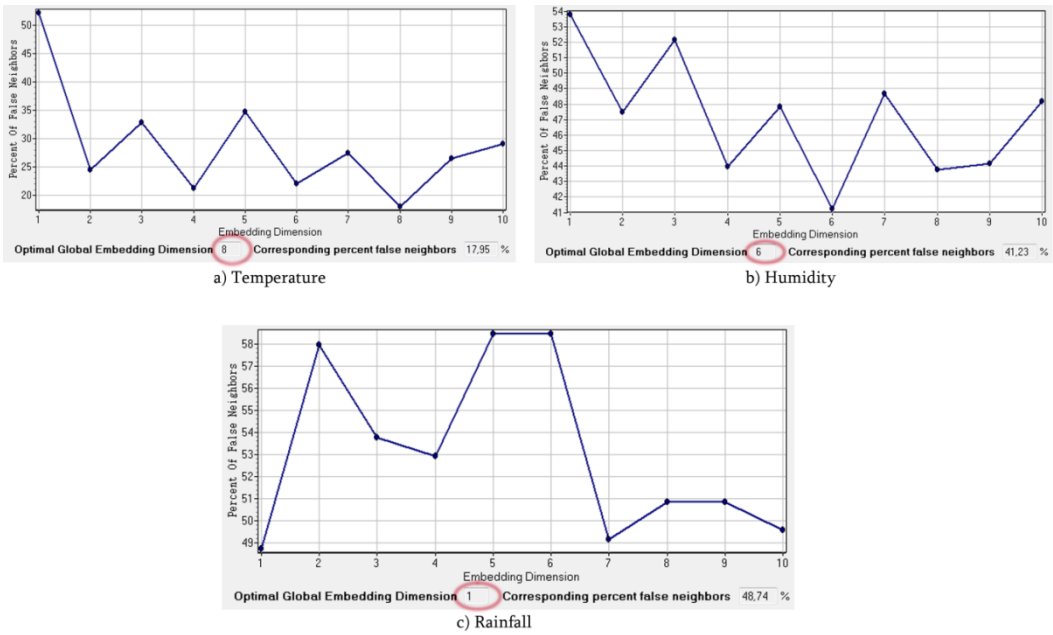


Figure 4.6 False Nearest Neighbours graphs

Optimum embedding dimensions that define the series for temperature, humidity and rainfall are also respectively 8, 6 and 1 by examining the first minimum of the graphs as seen in Figure 4.6. Optimum embedding dimensions are computed by using False Nearest Neighbour method (Equation 2.6). Then these variables were also visualized by drawing phase space plots as in Figure 4.7.

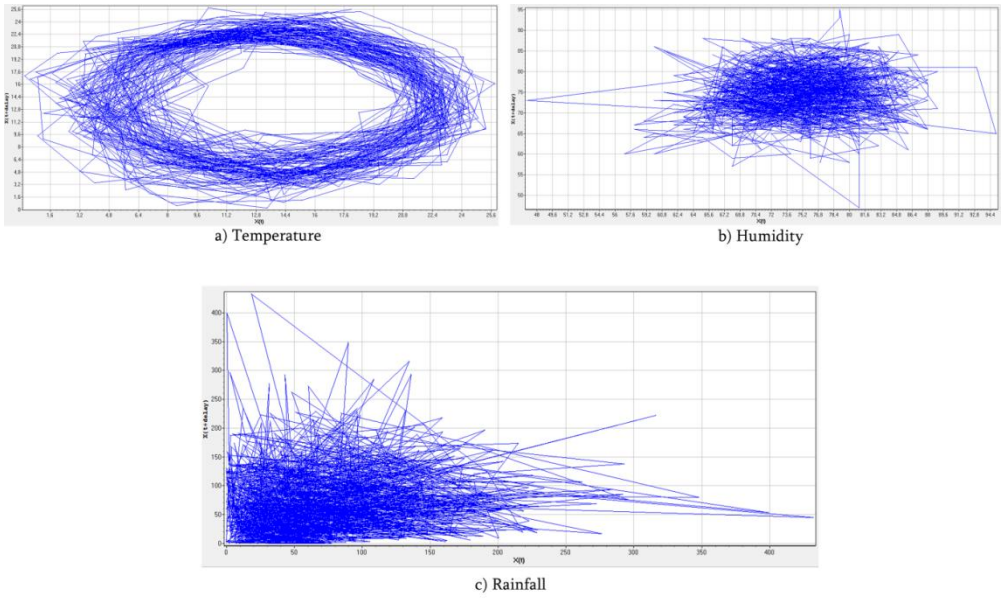


Figure 4.7 Phase Space Plot for Temperature (τ=3, m=8) Phase Space Plot for Humidity (τ=3, m=6) Rainfall (τ=2, m=1)

When Figure 4.7 is examined, chaotic attractor structure of Temperature, noisy complex structures of Humidity and Rainfall are seen.

4.3 Forecasting Results

In the forecasting step, Casdagli's OLS method is firstly applied to temperature data and 10 test values were forecasted by using 1,153 fitting values for different *k* neighbour numbers. And then, adapted PLS version of this method is developed. After the performance of this adaptation, a new improved PLS method that benefit from both multivariate embedding and variable selection methods is defined.

4.3.1 OLS Results

Temperature values are forecasted by Casdagli's OLS method and the model in Equation 3.3 for different values of *k* in this step. The forecasting results by means of $E_m(k)$, MAE, MSE, RMSE for various *k* values are shown in Table 4.1 and Figure 4.8.

Table 4.1 Forecasting Errors of Temperature (OLS Method)

		E_mk	**MAE**	**MSE**	**RMSE**
k	**20**	3.101356	5.276792	5.310913	2.304542
	50	3.020937	5.317384	4.924759	2.219180
	100	**3.020815**	**5.425057**	**4.757670**	**2.181208**

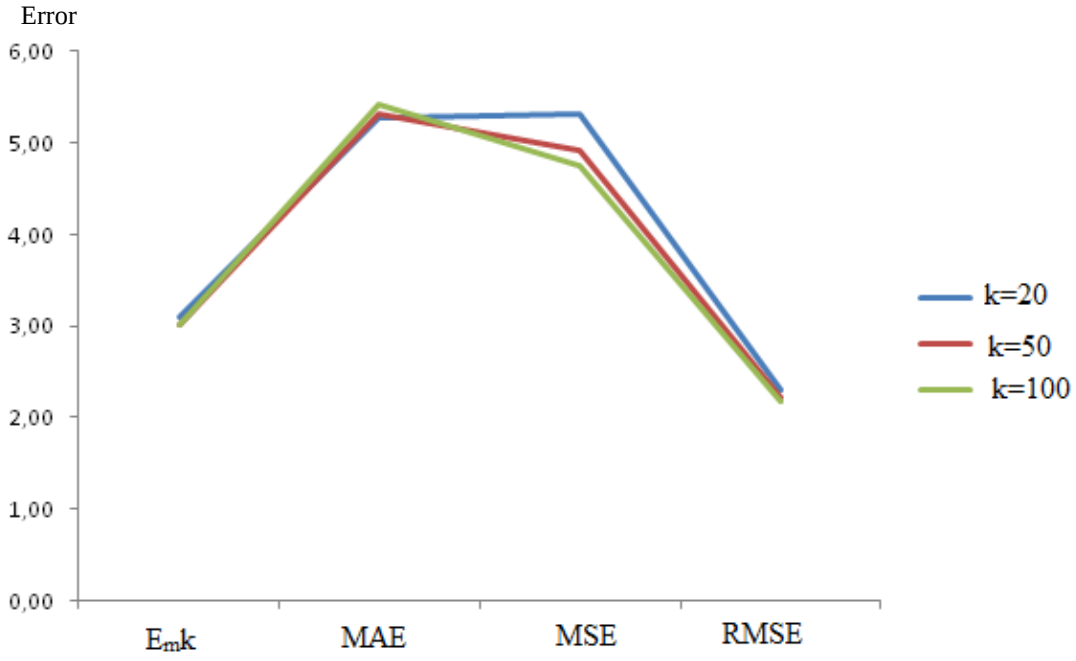


Figure 4.8 Forecasting error values versus k values

As seen above, E_mk values are obtained about 3. While *k* is increasing, forecast error values of the model generally decrease except MAE. The lowest RMSE value which is a common performance criteria in forecasting studies is observed at k=100 (RMSE=2.18).

4.3.2 PLS Results

Since forecasting errors of OLS method evaluated insufficient, it is believed PLS will give lower $E_m(k)$ and forecast performance values. Because, phase space is reconstructed from the delay vectors and these delays are the interrelated variables, so multicollinearity in the data can be high. PLS method version of Casdagli's method is applied to data for same *k* values (k=20, 50 and 100) and the results are computed as in Table 4.2, Table 4.3 and Table 4.4.

Table 4.2 Forecasting Temperature by PLS Method (k=20)

		E_mk	MAE	MSE	RMSE
Number of Components	**1**	**0.877996**	**1.452247**	**3.276206**	**1.810029**
	2	0.996797	1.692361	4.222797	2.054945
	3	1.101794	1.816301	5.159258	2.271400
	4	1.104170	1.804225	5.181532	2.276298
	5	1.107562	1.772608	5.213415	2.283290
	6	1.106125	1.769712	5.199902	2.280329
	7	1.107185	1.770527	5.209871	2.282514
	8	1.10757	1.771871	5.213496	2.283308

Table 4.3 Forecasting Temperature by PLS Method (k=50)

		E_mk	MAE	MSE	RMSE
Number of Components	**1**	0.955258	1.432934	3.878178	1.969309
	2	**0.947118**	**1.493051**	**3.812371**	**1.952529**
	3	0.966672	1.546267	3.971409	1.992839
	4	0.959005	1.557042	3.908666	1.977035
	5	0.961137	1.583426	3.926061	1.981429
	6	0.961677	1.588773	3.930476	1.982543
	7	0.962193	1.590061	3.934695	1.983607
	8	0.962157	1.590056	3.934402	1.983533

Table 4.4 Forecasting Temperature by PLS Method (k=100)

		E_mk	MAE	MSE	RMSE
Number of Components	**1**	0.926154	1.436658	3.645466	1.909310
	2	**0.921535**	**1.374643**	**3.609190**	**1.899787**
	3	0.937454	1.371047	3.734961	1.932605
	4	0.942460	1.382669	3.774959	1.942925
	5	0.938601	1.378742	3.744112	1.934971
	6	0.938963	1.381020	3.746997	1.935716
	7	0.938999	1.381357	3.747282	1.935790
	8	0.939088	1.381452	3.747996	1.935974

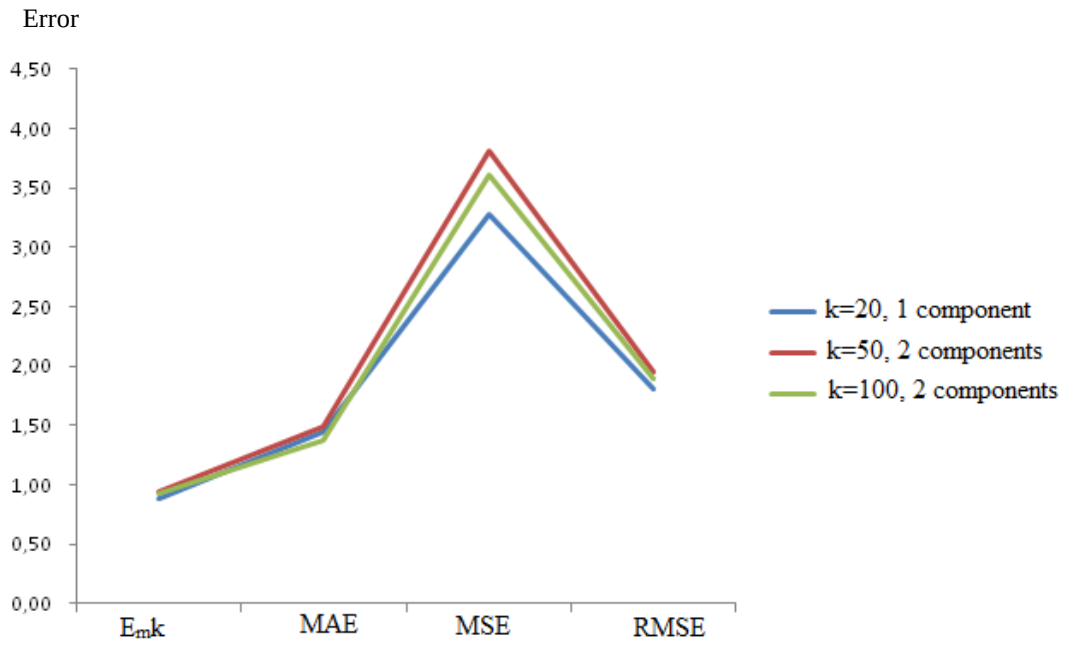


Figure 4.9 Forecasting error values versus k values and number of components

Number of component values for PLS are equal to one to optimum embedding dimension value. Number of components that have the lowest error values are marked red. As distinct from OLS method results, the lowest error values were obtained from low nearest neighbour value (k=20) and largest error values are observed for nearest neighbour value (k=50) as seen in Figure 4.9. Much more successful values are generally reached by adapted PLS method. A decrease at the

rate of approximately 70% in the error values leaps out, if Table 4.1 and the last three tables (Table 4.2, Table 4.3 and Table 4.4) are compared.

4.3.3 Proposed Method Results

So far, Casdagli's method is used as a local linear method. After the performance of PLS model, testing PLS as a global model by adding the other variables Humidity and Rainfall variables is also necessary. Since Casdagli's method is restricted about applying multivariate embedding, a global forecasting methodology is developed by ignoring *k* nearest neighbours. In this part, identifying the minimum number of variables enough for forecasting the temperature data is needed and filter variable selection method is used for PLS. If X_t: Temperature, M_t: Humidity and N_t: Rainfall define the current values, the minimum number of fundamental variables is determined as seven among 18 variables (X_t, $X_{t-\tau}$, $X_{t-2\tau}$, $X_{t-3\tau}$, $X_{t-4\tau}$, $X_{t-5\tau}$, $X_{t-6\tau}$, $X_{t-7\tau}$, $X_{t-8\tau}$, M_t, $M_{t-\tau}$, $M_{t-2\tau}$, $M_{t-3\tau}$, $M_{t-4\tau}$, $M_{t-5\tau}$, $M_{t-6\tau}$, N_t, $N_{t-\tau}$) based upon this selection. Since statistically the insignificant variables are removed from the model, rainfall or its delays are not included in the model. Variable selection results are also consistent with the boxplot results in Figure 4.2. Most of the selected variables are related with temperature that has the smallest variability and one of the variables is humidity that has the second smallest variations in the boxplots.

The selected variables obtained by applying filter method are:

X_1: X_t
X_2: $X_{t-2\tau}$
X_3: $X_{t-3\tau}$
X_4: $X_{t-4\tau}$
X_5: $X_{t-7\tau}$
X_6: $X_{t-8\tau}$
X_7: M_t

The presence of multicollinearity among selected variables is also checked by computing Tolerance and Variance Inflation Factor (VIF) values that are the most common multicollinearity statistics as in Table 4.5.

Table 4.5 Multicollinearity statistics of selected variables

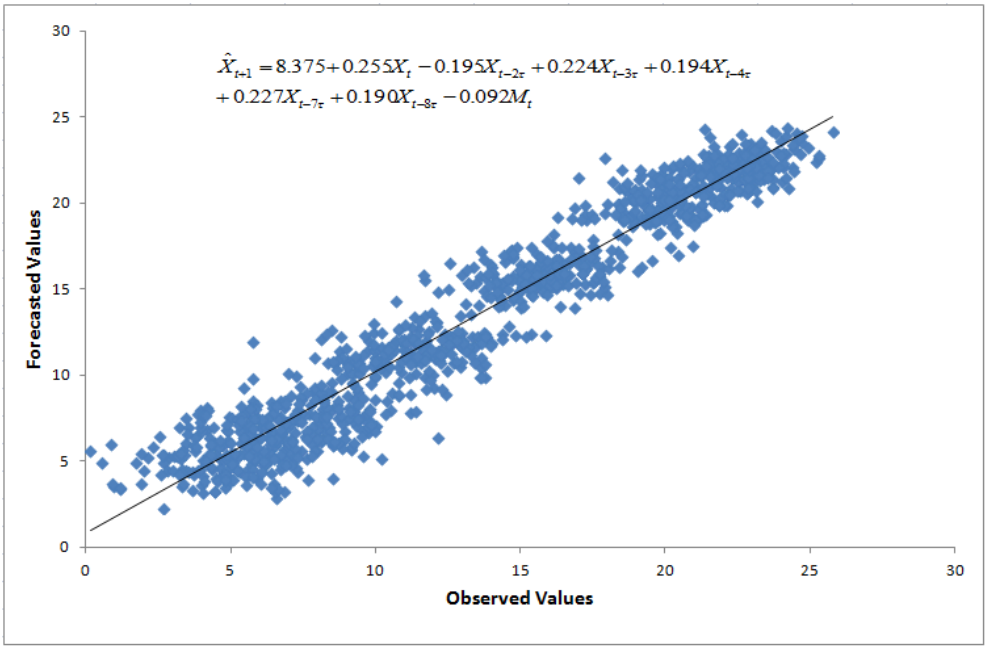
	Beta	**Tolerance**	**VIF**
(Constant)	7.883		
X_t	0.427	0.077	12.913
$X_{t-2\tau}$	-0.146	0.088	11.322
$X_{t-3\tau}$	0.209	0.111	8.977
$X_{t-4\tau}$	0.139	0.078	12.846
$X_{t-7\tau}$	0.236	0.111	9.031
$X_{t-8\tau}$	0.119	0.077	12.935
M_t	-0.102	0.715	1.399

Four of VIF values and the most of the VIF values are greater than 10 and 5 as seen in Table 4.5. This situation is a significant evidence of multicollinearity problem and supports applying PLS decision.

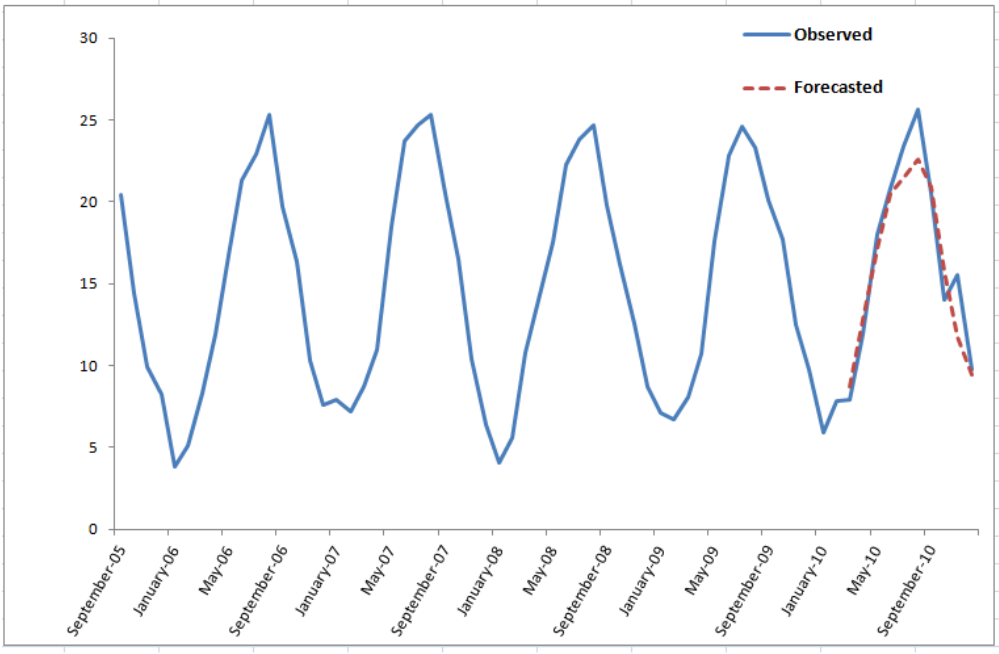
Table 4.6 Forecasting Errors of Temperature by PLS and Variable Selection Method

	$\mathbf{E_mk}$	**MAE**	**MSE**	**RMSE**	$\mathbf{R^2}$	$\mathbf{R^2_{adj}}$
3 Components	0.873382	1.404654	3.241868	1.800519	0.938276	0.937902

Much lower error values and better fitted forecasts were observed by selecting variables among a set of lagged independent variables and applying PLS as a global method (k=$N_{fitting}$) to these data as in Table 4.6 and Figure 4.10.



a) Regression line of the model



b) Time series plot of the forecasted values

Figure 4.10 Observed vs. Forecasted values

Since PLS model with three components provided the minimum error values from all models, it was identified as the forecasting model. This model explains approximately 94% of the variability in temperature as R^2 and R^2_{adj} coefficients of determination are seen in Table 4.5. So, temperature for Istanbul can be forecasted by following model:

$$X_{t+1} = 8.375 + 0.255X_t - 0.195X_{t-2\tau} + 0.224X_{t-3\tau} + 0.194X_{t-4\tau} + 0.227X_{t-7\tau} + 0.190X_{t-8\tau} - 0.092M_t \quad (4.1)$$

X_t: Temperature and M_t: Humidity, τ: Delay of temperature

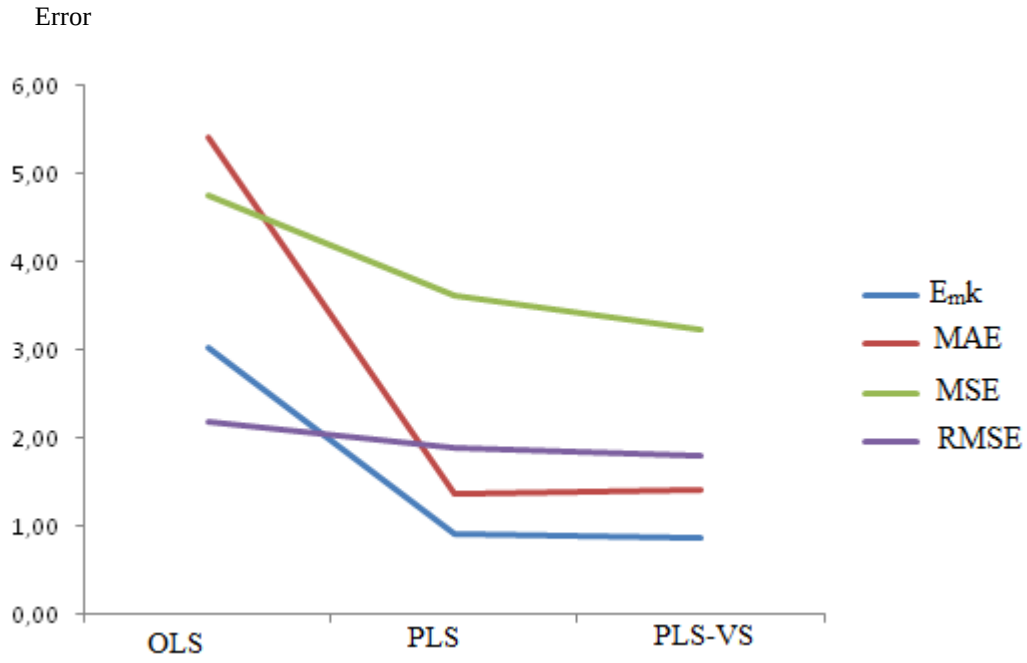


Figure 4.11 Comparing forecasting methods errors

When the comparison of forecasting methods according the errors was summarized in Figure 4.11 and examined, PLS methods had lower error values than OLS method and proposed PLS method with variable selection provided the best performance with lowest errors.

4.3.4 Comparing Proposed Method with Artificial Neural Networks

Different artificial neural network models are constructed to be able to compare the forecasting results with a common method and to validate the efficiency of the methods. The most appropriate artificial neural networks in time series forecasting are dynamic networks. NARX networks among dynamic neural networks are chosen

in this study, since they combine the delays of inputs and the past values of the output. These type neural networks are often used for nonlinear filtering and modeling of nonlinear dynamic systems.

NARX network model are defined by the Equation 4.2:

$$y_{(t)} = f(y_{(t-1)}, y_{(t-2)}, \dots, y_{(t-n_y)}, x_{(t-1)}, x_{(t-2)}, \dots, x_{(t-n_x)}) \quad (4.2)$$

$y_{(t)}$ is the dependent variable (output) and regressed on the previous values of its own values and the previous values of an independent variable ($x(t)$). In this thesis, temperature is the output while the humidity is the input variable. Because, these variables and their delays were found significant in PLS variable selection step.

Firstly, a NARX neural network with 10 hidden neurons and d=2 (delay) is constructed. The architecture for training related NARX network is represented below in Figure 4.12.

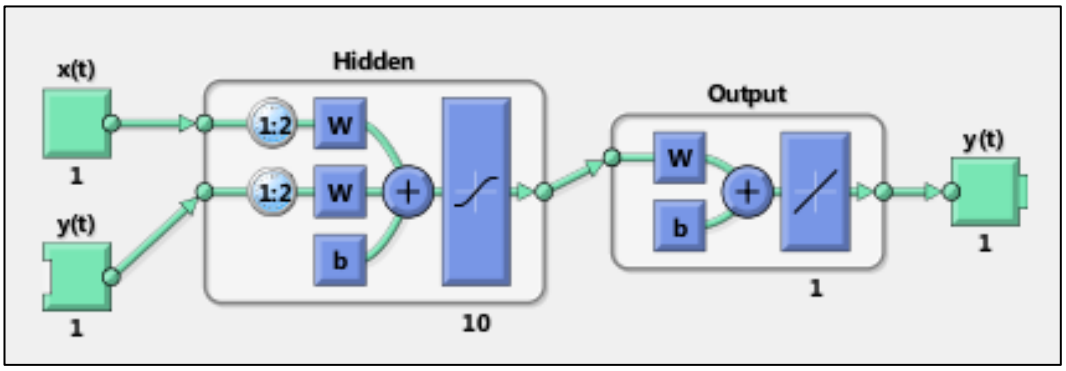


Figure 4.12 Architecture of NARX neural network

In NARX neural networks, the next value of the output signal $y_{(t)}$ is regressed on previous values of the output signal and previous values of an input (exogenous) $x_{(t)}$ signal as shown in Figure 4.12. For percentage of fitting, validation of network and test of network generalization, 70%, 15% and 15% of the data are respectively used.

Then the network is trained by using Levenberg-Marquart (trainlm) training function that is based on maximum neighbourhood and least squares.

MATLAB software is used for neural network computations and NARX neural network results are shown in Figure 4.13, Figure 4.14, Figure 4.15 and Figure 4.16. The numbers of neurons affect the weights and flexibility of the network and there is no standard rule for selecting number of neurons. So different neural networks for different number of hidden neurons (5, 10, 15) and different delays (d=2,...,8) were tried. These NARX neural networks with different parameters are used for comparison.

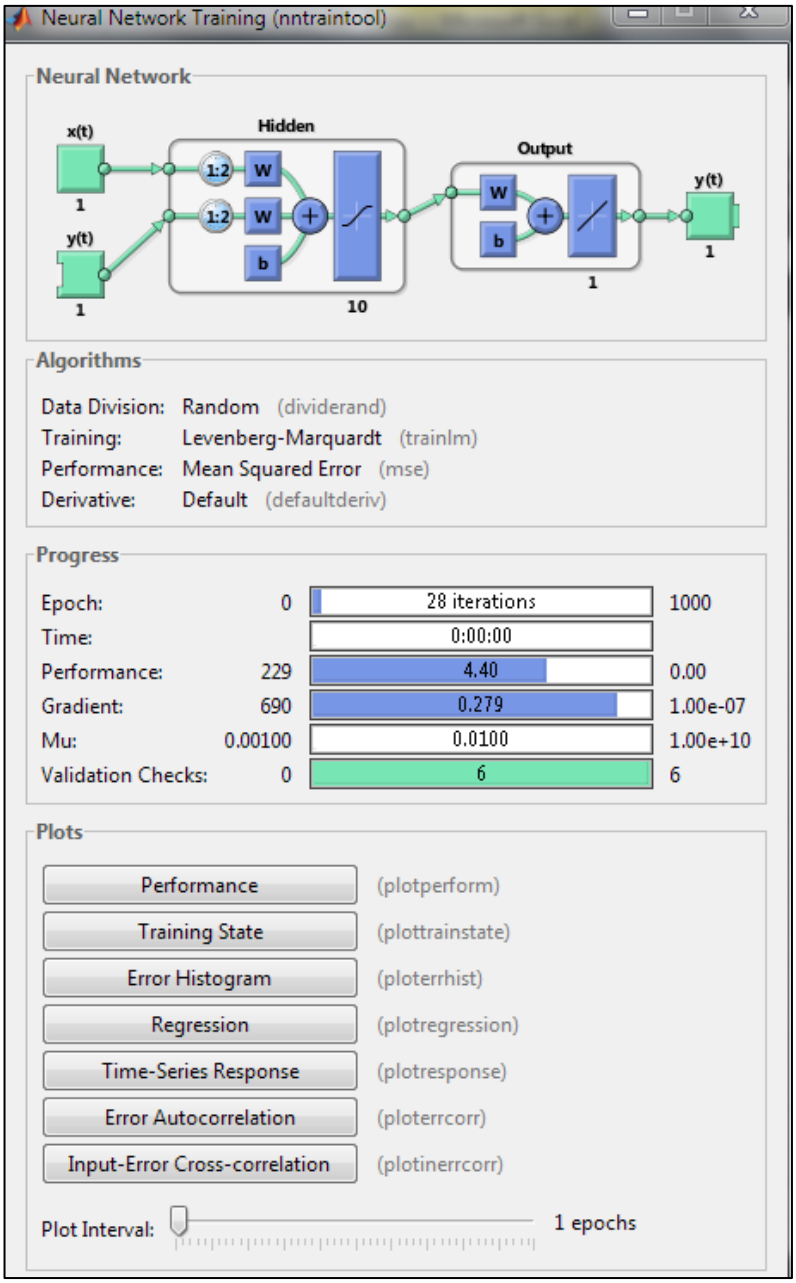


Figure 4.13 NARX neural network results

The training process was completed in 28 iterations as shown in Figure 4.13.

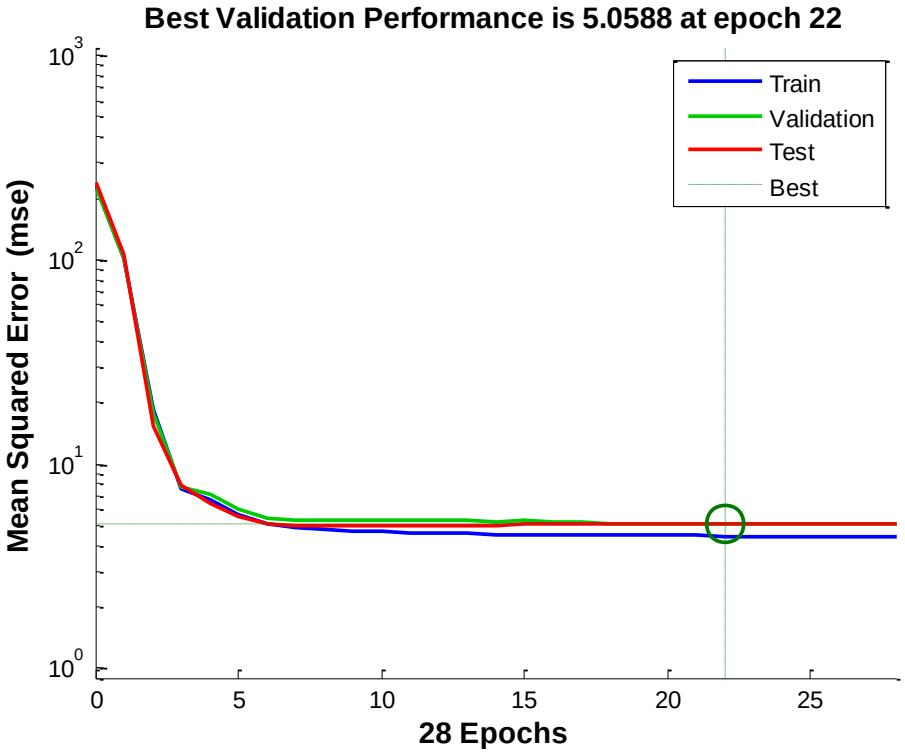


Figure 4.14 Performance plot

Training, validation and testing errors decrease as iteration increases and best performance is obtained at iteration 22 with a Mean Squared Error (MSE)=5.0588 as seen in Figure 4.14. To show the model explanation percent, regression plot of the network is drawn in Figure 4.15.

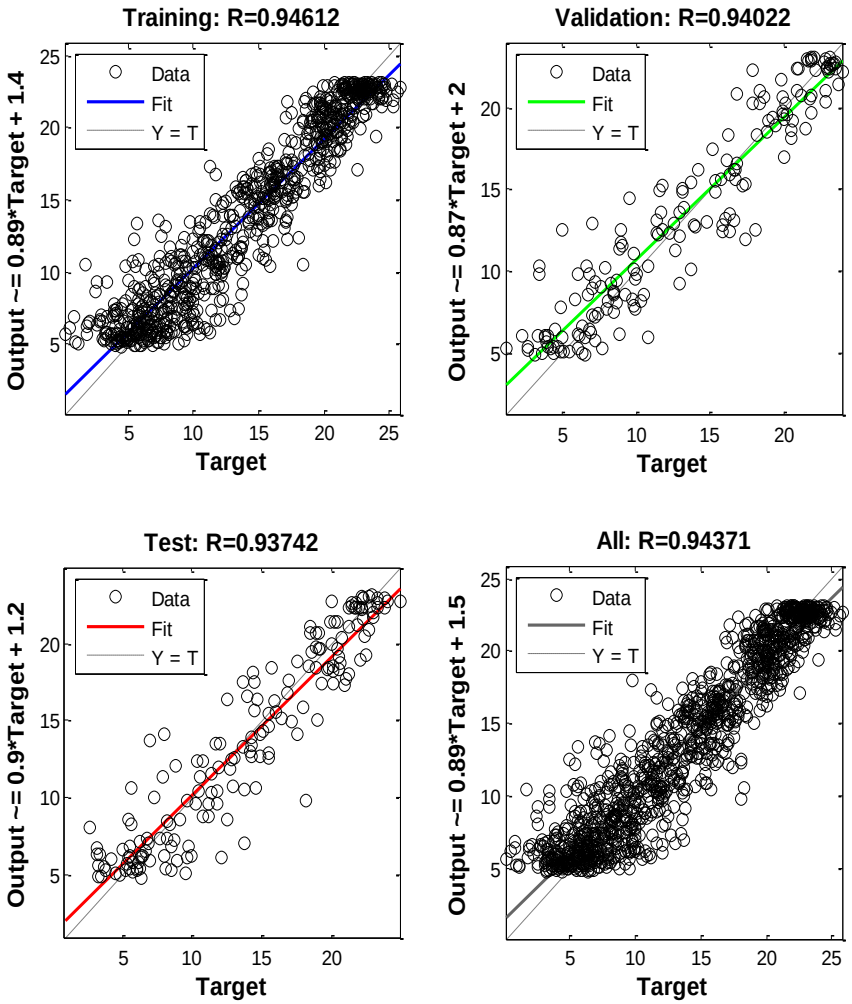


Figure 4.15 Coefficients of determination

Training, validation, test and overall coefficients of determination are respectively 0.94612, 0.94022, 0.93742 and 0.94371. Hence, the NARX neural network with 10 hidden neurons and 2 delays explains 94% of the variability in temperature. This coefficient of determination value (0.937902) is almost the same as proposed PLS model with variable selection.

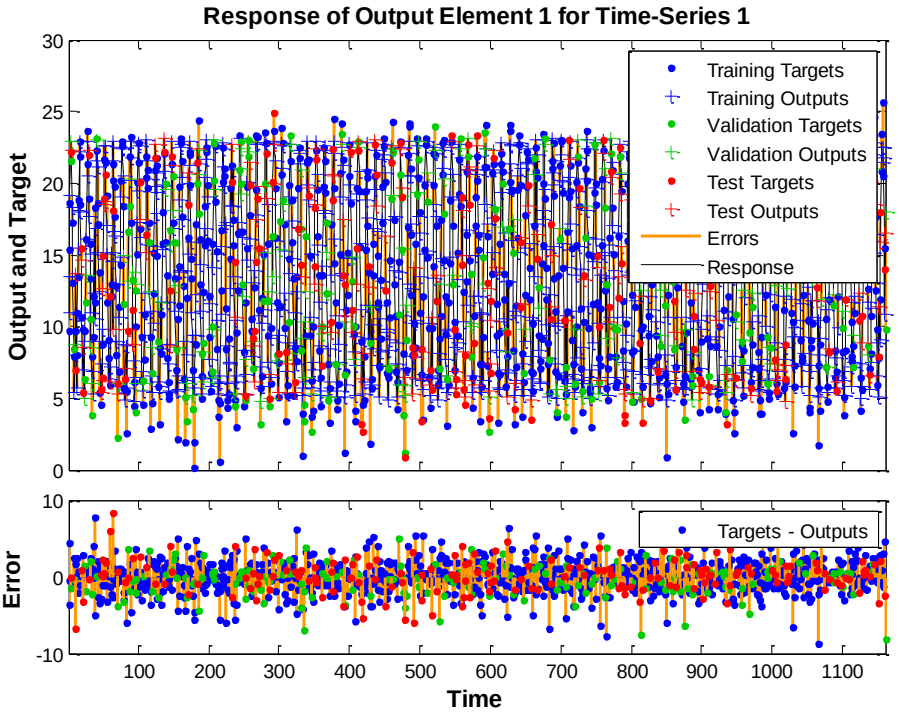


Figure 4.16 Time series response

In Figure 4.16, inputs, targets and errors for training, validation, test and overall data according to time are displayed. When the figure is examined, it is seen that the outputs generally distribute in a more restricted interval. While the differences between the targets and outputs (errors) do not variate too much, there are some outlier values.

In Table 4.7, computed RMSE values for NARX neural networks which have different parameters; number of hidden neurons (5, 10, 15), delays (d=2,...,8), are shown.

Table 4.7 NARX neural network performance values (RMSE)

Delays/ Number of hidden neurons	5	10	15
d=2	3.3906	3.6226	3.4123
d=3	2.2889	2.0954	2.0045
d=4	1.9187	2.4117	1.9601
d=5	1.9660	2.0348	2.0718
d=6	2.0551	2.1809	1.8969
d=7	2.3155	2.1376	2.0906
d=8	2.3199	1.8758	2.0886

RMSE values for 21 different NARX network architectures varied between 1.8969 and 3.6226 and even the lowest RMSE was greater than the value of proposed PLS method (1.8005) and the proposed method provided better performance with a 5% lower error value than NARX neural network architectures.

4.3.5 Forecasting Results of Proposed Method

Monthly mean temperatures for 2013, 2014 and 2015 years are also forecasted by the model in Equation 4.1. The forecasted values are obtained as Table 4.8.

Table 4.8 Comparison between the Forecast Values and the Past Temperature Data

	Mean Temperature Between 2008-2012	2013	2014	2015
January	6.0	**7.3**	**7.4**	**7.5**
February	6.2	**6.4**	**7.2**	**7.2**
March	8.5	7.5	7.9	8.5
April	12.2	11.8	11.5	11.2
May	16.7	15.4	15.7	15.2
June	21.3	18.8	18.6	18.8
July	23.5	22.5	21.5	20.7
August	23.7	22.7	22.4	21.7
September	20.9	20.4	20.5	20.6
October	16.7	**17.8**	**17.4**	**17.3**
November	13.1	**14.1**	**14.0**	**13.9**
December	9.5	**9.6**	**10.2**	**10.7**

Some temperature values especially mean temperature of autumn and winter (between October and February (marked red)) seasons are much higher than past 4 year mean temperature values (2008-2012). From this point of view, temperature values between 1961 and 2015 are analyzed in detail by means of 5 year mean values according to the months and are summarized as Figure 4.17 to infer about the future temperature and climate change of Istanbul.

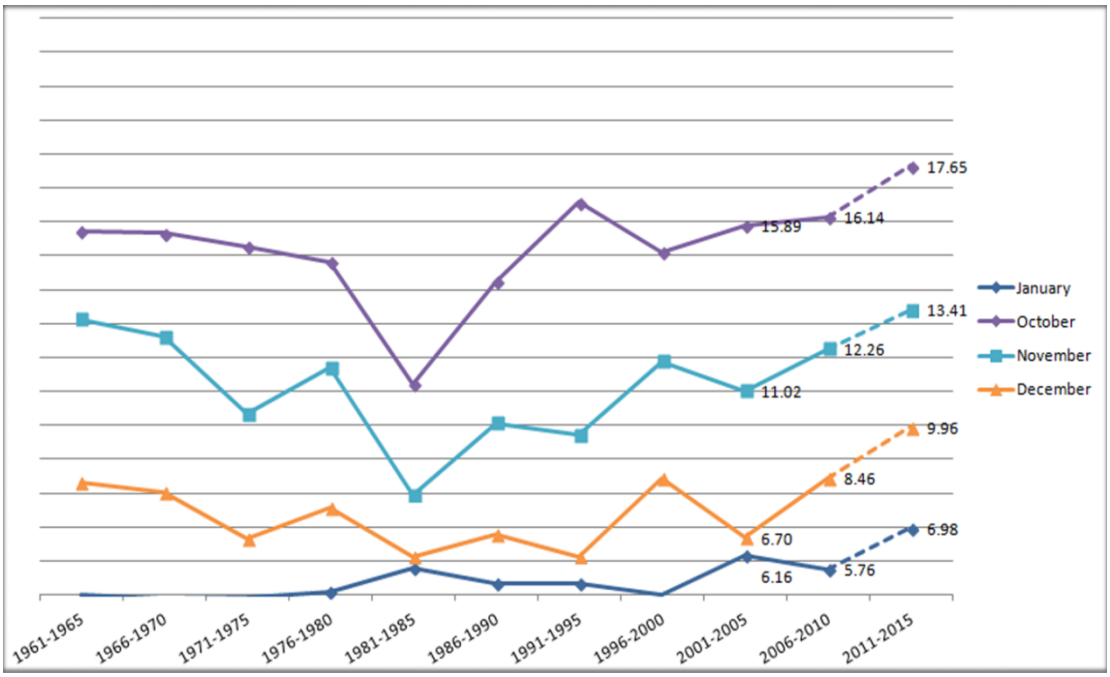


Figure 4.17 Analyzing the Temperature by Five Years Mean Values between 1961 and 2015

As seen in Figure 4.17, autumn and winter temperature values for Istanbul continuingly increase after 2001-2005 period. In 2011-2015 period, the temperature values reach the values that can not rise in any of past 5 year period. If this climate change continues for long years, global warming problem will be present for Istanbul in the near future.

CHAPTER FIVE
CONCLUSION

The goals of this study were (a) to develop a new nonlinear forecasting adaptation by combining partial least square (PLS) method and nonlinear time series analysis especially multivariate embedding; (b) to improve the forecast ability by evaluating the forecasting errors depending upon the parameter and forecasting method (local (OLS, PLS) or global (ANN)) changes; (c) to forecast İstanbul monthly mean temperature by benefitting from this new method and meteorological variables like temperature, humidity and rainfall.

This thesis contained two reviews of literatures, one on nonlinear time series analysis and the other one on meteorological studies. The study concentrated on statistical models in nonlinear time series forecasting, artificial neural network was also applied. Monthly mean temperature of Istanbul was also forecasted by presented prediction methods (OLS, PLS and ANN) using meteorological data between February 1914 and December 2010.

In addition, a multivariate embedding based new PLS forecasting method was developed by using nonlinear time series and variable selection methods. The effectiveness of this proposed method for forecasting in nonlinear patterns is performed in meteorological data.

Then, parameter estimations (m, τ) were obtained for the data by using Casdagli's method. Then Partial Least Square method was used to construct the regression model for obtaining the forecasts and the parameter estimations and error values were obtained. Many combinations depending on parameter (m, τ, k) changes were constructed, these combinations were computed by using MATLAB software and the changes in $E_m(k)$ values were examined.

OLS and PLS methods were adapted Casdagli's method and PLS method provided better forecasting performance in both local and global context. It was

found that PLS for nonlinear time series especially for high embedding dimension case - was a very efficient method. PLS gave not only low error value (RMSE=1.80), but it also explained temperature well (R^2=94%) by aforementioned independent variables.

NARX models among ANN models that have different parameters (number of hidden neurons=5,10,15, delays:2,...,8) were constructed to compare the forecasting results of the proposed method with the results of ANN as benchmark values and evaluate the performance of the proposed method. RMSE values for 21 different NARX network architectures were obtained and even the lowest RMSE (1.8969) was greater than the value of proposed PLS method (1.8005) and the proposed method provided better performance with a 5% lower error value.

Future temperature values were also forecasted by the proposed method. Forecast values of 2013, 2014 and 2015 were also compared the past temperature data of Istanbul and it was seen that some temperature values especially autumn and winter values, between October and January, were much higher than past temperature values. It was concluded that if the present temperature increase continued for long years, global warming might be a problem for Istanbul.

Briefly, in this thesis:

- The temperature variable was forecasted in both univariate and multivariate case (using other meteorological variables) by combining nonlinear time series and partial least square methods.
- Forecasting performance was improved using embedding method and variable selection methods.
- Approximately 100 years long data from 54 automatic meteorology observation stations in the megacity Istanbul was analyzed
- Monthly temperature values of 2014 and 2015 years were forecasted.

For further research, this thesis will also contribute to nonlinear time series analysis area since it deals with both multivariate case and different forecasting methods benefitting from long term time series data.

REFERENCES

Abed, M.M., El-Shafie, A., & Osman, S.A. (2010). Creep predicting model in masonry structure utilizing dynamic neural network. *Journal of Computer Science*, 597-605.

Alexander, C., & Giblin, I. (1997). Multivariate embedding methods: forecasting high-frequency financial data in the first INFFC. *Journal of Computational Intelligence in Finance*, *5* (6), 17-24.

Aparicio, T., Pozo, E., & Saura, D. (2002). The nearest neighbour method as a test for detecting complex dynamics in financial series. *Applied Financial Economics, 12* (7), 517-525.

Ataei, M., Khaki-Sedigh, A., Lohmann, B., & Lucas, C. (2003). Estimating the Lyapunov exponents of chaotic time series: A model based method. European Control Conference.

Baffi, G., Martin, E.B., & Morris, A.J. (1999). Non-linear projection to latent structures revisited: the quadratic PLS algorithm. *Computers and Chemical Engineering, 23*, 395-411.

Baffi, G., Martin, E.B., & Morris, A.J. (1999). Non-linear projection to latent structures revisited (the neural network PLS algorithm). *Computers and Chemical Engineering, 23*, 1293-1307.

Behnia, S., Akhshani, A., Mahmodi, H., & Hobbenagi, H. (2008). On the calculation of chaotic features for nonlinear time series. *Chinese Journal Of Physics, 46* (4), 394-404.

Benettin, G., Galgani, L., & Strelcyn, J.M. (1976). Kolmogorov entropy and numerical experiments. *Physical Review A, 14*, 2338–2345.

Bertels, K., Neuberg, L., Vassiliadis, S., & Pechanek, D.G. (2001). On chaos and neural networks: the backpropagation paradigm. *Artificial Intelligence Review, 15*(3), 165-187.

Bhattacharya, T. (2005). Prediction of silicon content in blast furnace hot metal using partial least squares (PLS). *ISIJ International, 45* (12).

Bilgili, M., & Sahin, B. (2009). Prediction of long-term monthly temperature and rainfall in Turkey. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 32* (1), 60-71.

Bountis, T.C., Drossos, L.B., Lakhsmanan, M., & Parthasarathy, S. (1992). On the non-integrability of a family of Duffing-van der Pol oscillators. *Journal of Physics A: Mathematical and General, 26,* 6927-6942.

Brock, W.A., Dechert, D., Lebaron, B., & Scheinkman, J. (1996) A test for independence based on a correlation dimension. *Econometric Review, 15,* 197-235.

Broomhead, D., & King, G. (1986). Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena, 20* (2-3), 217-236.

Caneco, A., Gracio, C., & Rocha, J. L. (2008). Symbolic dynamics and chaotic synchronization in coupled duffing oscillators. *Journal of Nonlinear Mathematical Physics, 15,* 102-111.

Casdagli, M. (1989). Non-linear prediction of chaotic time series. *Physica D: Nonlinear Phenomena, 35*, 335-356.

Casdagli, M. (1991). Nonlinear forecasting, chaos and statistics. *Santa Fe Institute Working Papers, 91*-05-022.

Casdagli, M. (1991). Chaos and deterministic versus stochastic non-linear modelling. *Journal of the Statistical Society, 54* (2), 303-328.

Chan, K.S, & Tong, H. (2001). *Chaos: a statistical perspective*. New York: Springer-Verlag,.

Chi-Hyuck, J., Lee, S., Park, H., & Lee, J. (2009). Use of partial least squares regression for variable selection and quality prediction. *Computers & Industrial Engineering*, 1302-1307.

Currenti, G., Negro, C.D, Fortuna, L., Napoli, R., & Vicari, A. (2004). Non-linear analysis of geomagnetic time series from Etna volcano. *Nonlinear Processes in Geophysics, 11*, 119-125.

Daw, C.S., Finney, C.E., & Tracy, E.R. (2002). A review of symbolic analysis of experimental data. *Review of Scientific Instruments*, 915-930.

Deffeyes, J. E. (2009). Nonlinear Dynamics of Infant Sitting Postural Control. Nebraska: University of Nebraska.

Demuth, H., Beale, M., & Hagan, M. (2008). *Matlab neural network toolbox user's guide*. Natick MA, USA: MathWorks.

Dhanya, C.T., & Kumar, D.N. (2010). Nonlinear ensemble prediction of chaotic daily rainfall. *Advances in Water Resources, 33*, 327–347.

Dharmasena, L. S. (2008). Sequential Procedures for Nonparametric Kernel Regression. School of Mathematical and Geospatial Sciences : RMIT University Science.

Diaz, M.A. (2008). Exchange rates forecasting: local or global methods? *Applied Economics, 40* (15), 1969-1984.

Diaz, M.A, & Sbert, J.M. (2011). Forecasting daily air arrivals in Mallorca Island using nearest neighbour methods. *Tourism Economics, 17* (1), 191-208.

Ding, M. Z., Grebogi, C., Ott, E., Sauer, T., & York, J. A. (1993). Estimating correlation dimension from a chaotic time series: When does plateau onset occur? *Physica D: Nonlinear Phenomena, 69*, 404-424.

Ebert, E.E., Turk, M., Kusselson, S.J., Yang, J., Seybold, M., Keehn, P.R., & Kuligowski, R.J. (2011). Ensemble tropical rainfall potential (eTRaP) forecasts. *Weather and Forecasting, 26*, 213-224.

Eckmann, J. P., Kamphorst, S. O., & Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhysics Letters, 5*, 973-977.

Elert, G. (1995). Measuring Chaos. Retrieved November 2009, from The Chaos Hypertextbook: http://hypertextbook.com/chaos/

Fantoni, P., Hoffmann, M. Rasmussen, B., Hines, J.W., & Kirschner, A. (2002). The use of non linear partial least square methods for on-line process monitoring as an alternative to artificial neural Networks. 5th International Conference on Fuzzy Logic and Intelligent Technologies in Nuclear Science.

Frank, G. W. (1990). Recovering the Lyapunov Exponent From Chaotic Time Series. Ontario: The University of Western Ontario.

Geladi, P., & Kowalski, B. (1986). Partial least squares regression: A tutorial. *Analytica Chimica Acta, 185*, 1-17.

Gluhovsky, A., & Agee, E. (2007). On the analysis of atmospheric and climatic time series. *Journal of Applied Meteorology and Climatology, 46*, 1125-1129.

Guo, H., Wang, L., Ping-liu, H. (2005). Kernel Partial Least Squares Based On Least Squares Support Vector Machine Primal Dual Optimization Problems. *International Conference on Intelligent Computing*, China.

Hardle, W. (1994). *Applied Nonparametric Regression*. Berlin: Humboldt University.

Hasegawa, K., Kimura, T., Miyashita, Y., & Funatsu, K. (1996). Nonlinear partial least squares modeling of phenyl alkylamines with the monoamine oxidase inhibitory activities. *Journal of Chemical Information and Computer Sciences, 36*, 1025-1029.

Haykin, S. (2005). *Neural Networks*. London: Pearson.

Hegger, R., Kantz, H., & Schreiber, T. 1999. Practical implementation of nonlinear time series methods: the TISEAN package. *Chaos, 9* (2), 413-435.

Hegger, R., Kantz, H., & Schreiber, T. (2007). *TISEAN sources.* Retrieved February 01, 2009, from http://www.mpipks-dresden.mpg.de/~tisean/

Hougland, J. (2010). Nonlinear forecasting and noise removal for chaotic time-series.

Izhikevich, E.M. (2013). *Encyclopedia of dynamical systems*. Retrieved May, 2012, from http://www.scholarpedia.org/article/Encyclopedia_of_dynamical_systems

Jaditz, T., & Sayers, C.L. (1998). Out-of-sample forecast performance as a test for nonlinearity in time series. *Journal of Business & Economic Statistics*, *16* (1), 110-117.

Jianping, J. (2012). Application of non-linear partial least-squares regression in the prediction of carbonization depth of concrete. *Advanced Materials Research*, 53-57.

Jo, H.J., Jo, E.M., Park, D., Lee, D.S., & Woo, S.H. (2010). Application of kernel partial least square to predict biological hydrogen production by enterobacter aerogenes. *Journal of Nanoelectronics and Optoelectronics*, *5*, 203-207.

Kantz, H., & Schreiber, T. (2004). *Nonlinear time series analysis* (2nd ed.). United Kingdom: Cambridge University Press.

Kapitaniak, T. & Bishop, S. R., (1998) *The Illustrated Dictionary of Nonlinear Dynamics and Chaos*, New York: Wiley.

Kasabov, N.K. (1998). *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*. Cambridge, Massachusetts: The MIT Press.

Katz, R.W. (2002). Techniques for estimating uncertainty in climate change scenarios and impact studies. *Climate Research, 20,* 167-185.

Kimura, T., Miyashita, Y., Funatsu, K. & Sasaki, S. (1996). Quantitative structure-activity relationships of the synthetic substrates for elastase enzyme using nonlinear partial least squares regression. *Journal of Chemical Information and Computer Sciences, 36,* 185-189.

Kolmogorov, A. N. (1959). Entropy per unit time as a metric invariant of automorphism. *Doklady of Russian Academy of Sciences, 124,* 754-755.

Konstantinou, K.I., & Lin, C.H. (2004). Nonlinear time series analysis of volcanic tremor events recorded at Sangay Volcano, Ecuador. *Pure and Applied Geophysics, 161,* 145-163.

Kugiumtzis, D., Lingjaerde, O.C., & Christophersen, N. (1998). Regularized local linear prediction of chaotic time series. *Physica D: Nonlinear Phenomena, 112,* 344-360.

Kugiumtzis, D. (2000). *State Space Local Linear Prediction*. Contribution for the book "Nonlinear Deterministic Modelling and Forecasting of Economic and Financial Time Series" by A. Soofi and L. Cao, Glasgow, UK.

Laio, F., Porporato, A., Ridolfi, L., & Tamea, S. (2004). Detecting nonlinearity in time series driven by non-Gaussian noise: the case of river flows. *Nonlinear Processes in Geophysics, 11,* 463-470.

Lee, D.S., Lee, M.W., Woo, S.H., Kim, Y., & Park, J.M. (2006). Nonlinear dynamic partial least squares modeling of a full-scale biological wastewater treatment plant. *Process Biochemistry, 41,* 2050-2057.

Liu, Z., Xie, X., Zhang, D., & Liu, H. (2006). Local partial least squares multi-step model for short-term load forecasting. *IEICE Transactions., 89-A* (10), 2740-2744.

Lodhi, H.M., & Yamanashi, Y. (2010). *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques*. USA: IGI Global.

Loun, W. A. (2006). *Variability Analysis of Physiological Signals Using Nonlinear Time Series Analysis Techniques*. Islamabad: Pakistan Institute of Engineering and Applied Sciences.

Mehmood, T., Liland, K.H., Snipen, L., & Sæbø, S. (2012). A review of variable selection methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems, 118*, 62-69.

Mendes, D., Mendes, V.M., Ferreira, N., & Menezes, R. (2009). Symbolic shadowing and the computation of entropy for observed time series. *Econophysics Colloquium 2006*. Tokyo: International Christian University (ICU).

Menezes, J. M., & Barreto, G. A. (2008). Long-term time series prediction with the NARX network: An empirical evaluation. *Source Neurocomputing*, 3335-3343.

Meng, M. & Niu, D. (2011). Annual electricity consumption analysis and forecasting of China based on few observations methods. *Energy Conversion and Management, 52*, 953-957.

Munuzuri, V.P., & Gelpi, I.R. (2000). Application of nonlinear forecasting techniques for meteorological modelling. *Annales Geophysicae, 18*, 1349-1359.

Nayfeh, A.H., & Balachandran, B. (1995). *Applied Nonlinear Dynamics*. New York: John Wiley and Sons.

Nerenberg, M.A., & Essex, C. (1990). Correlation dimension and systematic geometric effects. *Physical Review A, 42*, 7065-7074.

Packard, N., Crutchfield, J., Farmer, J., & Shaw, R. (1980). Geometry from a time series. *Physical Review Letters, 45*, 712.

Parker, T., & Chua, L. (1987). Chaos: A tutorial for engineers. *Proceedings of the Institute of Electrical and Electronics Engineer, 75*, 982.

Penner, A.J. (2004). *Nonlinear Analysis of Complicated Physical Systems*. Manitoba: University of Manitoba.

Percival, D.B., Overland, J.E., & Mofjeld, H.O. (2004). Modeling North Pacific climate time series. *The IMA Volumes in Mathematics and Its Applications Series, 139*, 151-167.

Qin, S. J., & McAvoy, T. J. (1992). Non-linear PLS modelling using neural networks. *Computers and Chemical Engineering*, *16*, 379–391.

Qin, S. J. (1993). A statistical perspective of neural networks for process modelling and control. *Proceedings of the 1993 International Symposium on Intelligent Control*, 559–604.

Ramage, C.S. (1993). Forecasting in meteorology. *Bulletin of the American Meteorological Society*, *74*, 1863-1871.

Ramzan S., Zahid, F.M., & Ramzan S. (2010). Prediction methods for time series regression models with multicollinearity. *World Applied Sciences Journal*, *11* (4), 443-450.

Redaelli, S., & Macek, W. M. (2001). Lyapunov exponent and entropy of the solar wind flow. *Planetary and Space Science, 49*, 1211-1218.

Rosenstein, M.T., Collins, J.J., & Luca, C.J.D. (1993). A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena, 65*, 117-134.

Rosipal, R. (2011). Nonlinear partial least squares: an overview. *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques* (169-189). ACCM, IGI Global.

Sahin, M., Yildiz, B.K., Senkal, O., & Pestemalci, V. (2012). Forecasting of air temperature based on remote sensing. *Gazi University Journal of Science, 25* (2), 327-335.

Sakai, K., Noguchi, Y., & Asada, S. (2008). Detecting chaos in a citrus orchard: reconstruction of nonlinear dynamics from very short ecological time series. *Chaos, Solitons & Fractals, 38* (5), 1274-1282.

Samarasinghe, S. (2007). *Neural Networks for Applied Sciences and Engineering-from Fundamentals to Complex Pattern Recognition.* New York: Taylor&Francis Group.

Sevil, H. E. (2006). On the Predictability of Time Series by Metric Entropy. Izmir: Izmir Institute of Technology.

Shang, P., Na, X., & Kamae, S. (2007). Chaotic analysis of time series in the sediment transport phenomenon. *Chaos, Solitons & Fractals, 41* (1), 368-379.

Shashidhar, A. (2006). *Generalized Volterra-Wiener and Surrogate Data Methods for Complex Time Series Analysis*. Massachusetts: Massachusetts Institute of Technology.

Sinai, Y. G. (1959). On the notion of entropy of a dynamical system. *Doklady of Russian Academy of Sciences, 124*, 768-771.

Sivakumar, B., Liong, S.Y., Liaw, C.Y., Phoon, K.K. (1999). Singapore rainfall behaviour: chaotic? *Journal of Hydrologic Engineering, 4*(1), 38-48.

Small, M. (2005). *Applied Nonlinear Time Series Analysis- Applications in Physics, Physiology and Finance*. Hong Kong Polytechnic University.

Sprott, J.C. (2008). *Chaos and Time Series Analysis*. Oxford: Oxford University Press.

Steuer, R., Molgedey, L., Ebeling, W., & Jimenénez-Montano, M. A. (2001). Entropy and optimal partition for data analysis. *The European Physical Journal, B* (19), 265-269.

Strogatz, S.H. (1994). *Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry and Engineering*. Addison Wesley Publishing.

Tang, G., Ding, Y., Wang, S., Ren, G., Liu, H., & Zhang, L. (2010). Comparative analysis of China surface air temperature series for the past 100 years. *Advances in Climate Change Research, 1* (1), 11-19.

Tektas, M. (2010). Weather forecasting using ANFIS and ARIMA models: a case study for Istanbul. *Environmental Research, Engineering and Management, 1* (51), 5-10.

Titchener, M. R., & Ebeling, W. B. (2001). Deterministic chaos and information theory. *Proceedings of the Data Compression Conference*.

Turkish State Meteorological Service-TSMS. (2013). Retrieved May 09, 2013, from http://dmi.gov.tr/kurumsal/istasyonlarimiz.aspx?sSirala=AL&m=ISTANBUL#sfB.

Ustaoglu, B., Cigizoglu, H.K., & Karaca, M. (2008). Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. *Meteorological Applications, 15*, 431-445.

Velickov, S., Price, R. K., & Solomatine, D. P. (2003). Prediction of Nonlinear Dynamical Systems Based on Time Series Analysis: Issues of Entropy, Complexity and Predictability. *Proceedings of the XXX IAHR Congress. IAHR*.

Wilks, D.S. (2011). *Statistical Methods in the Atmospheric Sciences*. USA: Elsevier.

Williams, J. M. (2004). *Lyapunov Exponents, Entropy and Dimension*. Texas: University of North Texas.

Wold, S., Kettaneh-Wold, N., & Skagerberg, B. (1989). Nonlinear PLS modeling. *Chemometrics and Intelligent Laboratory Systems, 7*, 53-65.

Wolf, A., Swift, J.B., Swinney, H.L., & Vastano, J.A. (1985). Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena, 16*, 285-317.

Xue-feng, X., Bao-chuan, F., Wei-zhong, L., & An-yong, L. (2010). The Application and Study of A Neural Network Model Based on Multivariate Phase Space Reconstruction. *Software Engineering and Data Mining (SEDM), 2010 2nd International Conference*, 361-365.

Xue-feng, X., Bao-chuan, F., & An-yong, L.(2011). Applying one fuzzy neural network model based on multivariate phase space reconstruction to predict deformation of deep foundation pit. *International Journal of Advancements in Computing Technology, 3* (2).

Yamamura, M., & Shoji, I. (2010). A nonparametric method of multi-step ahead forecasting in diffusion processes. *Physica A: Statistical Mechanics and Its Applications, 389*, 2408-2415.

Zhang, C., & Yue, J. (2012). The chaotic prediction for aero-engine performance parameters based on nonlinear PLS regression. *Journal of Applied Mathematics*.

Zhang, G.P., Patuwo, B.E., & Hu, M.Y. (2001). A simulation study of artifcial neural networks for nonlinear time-series forecasting. *Computers & Operations Research, 28* (4), 381-396.

Zhang, R., Cai, W., Ni, L., & Lebby, G L. (2008). Power System Load Forecasting Using Partial Least Square Method. *40th Southeastern Symposium on System Theory*, 169-173.