JUNE, 2013 Master's Degree/Civil Engineering MUAZZEZ MUNİSE SOYBAKIŞ

UNIVERSITY OF GAZİANTEP
GRADUATE SCHOOL OF
NATURAL & APPLIED SCIENCES

GENERATING REGIONAL CLIMATE CHANGE SCENARIOS WITH NEAREST NEIGHBOR ALGORITHM IN WESTERN BLACK SEA BASIN, TURKEY

M. Sc. THESIS
IN
CIVIL ENGINEERING

MUAZZEZ MUNİSE SOYBAKIŞ

JUNE 2013

Generating Regional Climate Change Scenarios with Nearest Neighbor Algorithm in Western Black Sea Basin, Turkey

M.Sc. Thesis
in
Civil Engineering
University of Gaziantep

Supervisor
Assist. Prof. Dr. Nermin ŞARLAK

Muazzez Munise SOYBAKIŞ
June 2013

T.C.
UNIVERSITY OF GAZİANTEP
GRADUATESCHOOL OF
NATURAL & APPLIED SCIENCES
CIVIL ENGINEERING DEPARTMENT

Name of the thesis: GeneratingRegional Climate Change Scenarios with Nearest Neighbor Algorithm in Western Black Sea Basin Turkey

Name of the student: Muazzez Munise SOYBAKIŞ
Exam date: June 3, 2013

Approval of the GraduateSchool of Natural and Applied Sciences

Assoc. Prof. Dr. Metin BEDİR
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Mustafa GÜNAL
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Nermin ŞARLAK
Supervisor

Examining Committee Members	Signature
Prof. Dr. Mustafa GÜNAL	____________
Assoc. Prof. Dr. İsmail HALTAŞ	____________
Assist. Prof. Dr. Nermin ŞARLAK	____________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Muazzez Munise SOYBAKIŞ

ABSTRACT

GENERATING REGIONAL CLIMATE CHANGE SCENARIOS WITH NEAREST NEIGHBOR ALGORITHM IN WESTERN BLACK SEA BASIN, TURKEY

SOYBAKIŞ, Muazzez Munise
M.Sc. in Civil Engineering
Supervisor: Assist. Prof. Dr. Nermin ŞARLAK
June 2013, 40 pages

The specific objectives of this thesis are two-fold: 1) demonstrate the ability of the stochastic weather generator to simulate regional monthly maximum and minimum temperature and precipitation data based on the historic records in north western part of Turkey, 2) develop an ensemble of climate scenarios for assessment of watershed vulnerability to extreme events such as floods and droughts. K nearest neighbor weather generator is utilized for this study. A set of statistics is computed for each month and summarized with box-plot to evaluate the model results. These results are based on codes written in R programming language. Since the simulations reproduce the historic statistics reasonably well, the same method is used to produce alternative climate data sets under two different climate change scenarios. These climate scenarios are: 1) increasing precipitation, and 2) increasing temperature. While "increasing the precipitation" scenario is useful for assessing the risk of floods events, "increasing the temperature" scenario is useful for the analysis of droughts risk in the basin.

Keywords: Weather generators; climate change scenarios; nearest neighbor algorithm; hydroclimatologic data, western black sea basin.

ÖZET

BATI KARADENİZ HAVZASINDA EN YAKIN KOMŞU ALGORİTMASI İLE BÖLGESEL İKLİM DEĞİŞİMİ SENARYOLARI ÜRETMEK

SOYBAKIŞ, Muazzez Munise
Yüksek Lisans Tezi, İnşaat Mühendisliği Bölümü
Tez Yöneticisi: Yrd. Doç. Dr. Nermin ŞARLAK
Haziran 2013, 40 sayfa

Tezin amaçları iki aşamada toplanabilir: 1) Türkiye`nin kuzey batı bölgesindeki tarihi gözlem verilerine dayanarak bölgesel aylık maksimum ve minimum sıcaklık ve toplam yağış verilerinin türetilmesi için literatürdeki hava üreteçlerinin becerisini gösterebilmek, 2) kuraklık ve taşkın gibi ekstrem olaylara karşın havzanın savunmasızlığını değerlendirebilmek için iklim senaryoları topluluğu oluşturmaktır. Her bir ay için bir dizi istatistikler oluşturulmuş ve model sonuçlarını değerlendirmek için kutu grafiği ile özetlenmiştir. Model sonuçları, R programa dilinde yazılan kodlardan elde edilmiştir. Türetilen seriler tarihi serinin istatistiklerini makul ölçüde yeniden üretebildiğinden dolayı, aynı yöntem iki farklı iklim değişikliği senaryosu altında alternatif iklim veri setlerini oluşturmak için kullanılmıştır. Bu iklim senaryoları: 1) artan yağış, ve 2) artan sıcaklık durumudur. “Artan yağış” senaryosu havzadaki taşkın risklerinin değerlendirilmesinde faydalı olurken, “artan sıcaklık” senaryosu havzadaki kuraklık analizi için faydalıdır.

Anahtar kelimeler: hava üreteçleri, iklim senaryoları, en yakın komşu algoritması, hidroklimatolojik veri, batı karadeniz havzası

ACKNOWLEDGEMENT

I would like to thanks to Assist. Prof. Dr. Nermin ŞARLAK for guiding me with patience throughout this study. Without her support, inspiration, dedication of time and energy, I could never have completed this work.

I would also like to extend my appreciation to my committee member, Assoc. Prof. Dr. Ismail Haltaş, for his helpful comments and suggestions and to my committee chair, Prof. Dr. Mustafa Günal, who helped to make my defense a relaxed and valuable experience.

Finally, I would also thanks to my family (my son Ali Ege SOYBAKIŞ and my darling İbrahim SOYBAKIŞ) for their support, helping and encouragement during my study.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

LIST OF SYMBOLS/ ABBREVIATIONS

TAV	The average temperature
PCT	precipitation
Tmax	Maximum temperature
Tmin	Minimum temperature
q	Stations
N	Years
p	Variables
w	Temporal window of length
x_t	Regional means
$\overline{PCT_t}$	Regional means of precipitation
$\overline{Tmax_t}$	Regional means of maximum temperature
$\overline{Tmin_t}$	Regional means of minimum temperature
$\bar{x}_t$	The mean vectors
C_t	Covariance matrix
d_i	Maholanobis distance
p_j	Cumulative probability
$\bar{\Psi}_m$	Regional monthly means
ψ_m^i	Regional monthly means for each year
δ_m^i	Regional monthly anomalies
L	Potential neighbors block
F_t^i	Feature vector
T	Transpose of the vector
w_i	Weighted probability function
u	Uniform random number
K-NN	K nearest neighbors
$I_m^{1,N}$	The index values

CHAPTER I

INTRODUCTION

Climate is comprised by a whole of long term meteorological events in the atmosphere. In regarding the discussions of climatologists, climate is a measurement of long term weather events such as average precipitation, temperature, moisture, wind speed etc. in a region. For example, when the data of precipitation station is studied and in the event that the precipitation is detected below the average value for that region along summer months, and that situation continues for long years, this shall be regarded as a sign of climate change.

Climate and climate change has been affecting our environment and all the living creatures in the world so that issue should be worked on. Increase in global temperature is expected to raise sea level, change in precipitation and other regional climate conditions. It is clear that these changes will affect water resources negatively. Concerning the relation between temperature and saturated vapor pressure, globally a minimal temperature increase creates an overall in precipitation. In other side, increase in temperature also increase evapotranspiration. However impacts of regional climate change on regional precipitation and evaporation has not been presented clearly. In a study by Soykan (1995) assuming no change in precipitation, runoffs are assessed to decrease about 30% as a result of global warming.

Groundwater sources are also affected negatively from climate change. At dry zones, groundwater resources that are especially vital for irrigation affected dramatically by the amount and distribution of precipitation.

For the reason mentioned above, changes in climate are studied carefully and the possible reasons are presented. By the last few years, climate change has been expressed as global warming. Global warming is laid stress on manmade factors such as increase in energy usage, industrialization raise the amount of green house gases and causes global warming. Indeed, greenhouse gases that keep sun and earth radiation have a major role in warming up the atmosphere. They create greenhouse effect having a significant place in climate system. The main gases comprising the atmosphere are nitrogen (78.08%), oxygen (20.95%), argon, carbon dioxide and the others (0.97%).These gases not only prevent radiation of sun from reflecting to the space, but also cause the earth to warm up by absorbing it. The greenhouse effect can be explained as; on a cloudless and open weather, big part of short wave solar radiation passes through the atmosphere and reaches the earth, is absorbed there. However, one part of long wave earth radiation oscillated from earth’s hot surface is absorbed by many effective trace gases (greenhouse gases) by radiation standing on high levels of the atmosphere before running through the space and then is reflected again. As in general average conditions, long wave earth radiation running through the space is in balance with sun radiation, earth/atmosphere combined system will be hotter than an environment not including greenhouse gases.

However in recent studies, it is debated that greenhouse effect is not the only reason for climate change. In some of those studies, scientist showed that continental drift may be another reason for climate change. It is claimed that the movements of continents affects current systems in oceans and winds in the atmosphere and that may be a reason for climate change (Sunay 2000).

Some researchers argue that periodic extremes in volcano activities affect climate system. In these studies, it is indicated that dust in big amounts rising to the atmosphere by volcano explosions constitutes a layer and that prevents the transition of sun rays resulting in decrease in the world’s temperature. It is argues that the average temperature of the world through a year resulting from the explosion of Pinatuba in Philippines in 1991, decreased approximately 1% (Aksay et al. 2005).

Furthermore, there are researchers who look for a relation between climate change and sun spots. In fact, changes in magnetic area of sun and sun spots affect the amount of radiant energy transmitted to the earth. Even if sunspot effect on climate is debated, increase of sunspots on sun surface may reduce sun energy and this situation may reduce the amount of energy that will reach to the earth. This decrease in the energy flux may cause experiencing colder weather conditions in places especially far away from equator line.

Whatever is the reasons, meaningful changes in precipitation and temperature originated from climate changes are observed. It is an urgent and inevitable lack to take into consideration of new conditions created by current climate change effects and to plan projects for water constructions in the near future.

In that context, this study attempts to determine the statistically meaningful changes in precipitation and temperature, and the possible impacts of these changes in the future by using current data about climate. Non-parametric weather generator models, which are well described in the literature, are developed in various scenarios. Through these scenarios, how the hydroclimatic parameters such as precipitation and temperature will change in the future is estimated.

Research Objectives and Scope

This thesis describes the applicability of a K-NN based weather generator model to simulate weather data conditioned upon plausible climate scenarios in Western Black Sea basin, Turkey. The performance of weather generator model is evaluated by comparing a synthetic historical data set to the observed data. The major objectives of this study can be summarized as follows:

1) Evaluate the non-parametric weather generator model in terms of its ability to simulate historical climate characteristics at the Western Black Sea Basin.

2) Assess the potential impacts of climate change on temperature and precipitation in the Western Black Sea Basin using the same model.

CHAPTER II

LITERATURE REVIEW AND BACKROUND

Climate change is the one of the most important problems of this century as mentioned earlier. Many recent indicators such as extensive and increasing decline in glaciers, decline in stream flows at many regions in the course of time, increase in temperature, rise of global sea level, increase or decrease in precipitation and changes in local climates point out the existence of climate change. Climate scenarios generated in accordance with meaningful changes in precipitation and temperatures as a result of climate changes are clearly significant to prevent future problems regarding to excess or lack of water. For this reason, various climate change models were developed in coordination with inter-governmental climate change panel (IPCC, 2007). Furthermore, many researchers have studied the climate change and its effect on water resources. Some of these studies are discussed in the following section.

Burn and Simonovic (1996) investigated the potential impacts of climate change on performance of reservoir operations. In that context, scenarios were generated representing climate conditions and those generated data were taken as input in reservoir operation models. They used downscaling method in order to acquire climate scenarios and concluded that reservoir performances are quite sensitive to climate changes.

Westernmacott and Burn (1997) combined empirically based and process-based concepts to evaluate the effects of climatic change on hydrological variables in the Churchill–Nelson River Basin in Western-central Canada. They found out that the timing of snowmelt occurrence changed and the magnitude of hydrological events decreased over time.

Southam et al. (1999) investigated the impact of climate change in Ontario's Grand River basin under 21 different climate scenarios for future surface water supply, streamflow regulation, population and water use. They concluded that climate change might have serious impacts on the capability of the Grand River to assimilate discharged wastewater. Therefore, Grand River may not be anymore a reliable supply of water for municipal purposes while maintaining existing water quality standards.

Mortsch et al. (2000) studied the impact of changing climatic conditions on the Great Lakes region under the scenario of doubling CO_2 concentration in the atmosphere. They indicated declines in runoff and lake levels in consequence of the generated climate change scenarios by using Global Circulation Models (GCM).

There are several different approaches for generating future climate scenarios in order to develop effective climate adaptation strategies. While there is no ideal method for generating future climate scenarios, parametric, non-parametric and semi-parametric weather generator models are used for this purpose. In parametric weather generators, precipitation is used as the driving variable (Jones et al., 1972; Nicks and Harp, 1980; Richardson, 1981).In this approach researchers tried to generate the occurrence probability process such as "dry" and "wet" days of daily precipitation and the intensity process such as the sequence of precipitation amounts on wet days. Richardson (1981) fit lag-1 multivariate autoregressive model (MAR-1) to generate maximum and minimum temperature with precipitation.

The parametric models require the following assumptions: (1) normal distribution assumption, (2) searching for linear relationship between variables and (3) only conditioned variables capture the dependency in other variables generated based on a variable. However, normality assumption is difficult to justify directly from hydrological or meteorological data. A transformation function should be applied if the data does not come from normal distribution. However, it may result in losing information.

Young (1994) employed non-parametric weather generators for multiple sites in a variety of regions and found out that the model usually preserves the cross correlation between the variables, but biases were also noticed in generated series and observed series.

Rajagopalan et al.,(1997) also developed a model in order to avoid the disadvantage of the parametric models. They introduced the K- nearest neighbor (K-NN) method which is one of the promising techniques among the nonparametric weather generator models. Nearest neighbor methods are based on selecting the nearest K monitoring data randomly in order to determine the next day weather. While the algorithm developed by Rajagopalan et al., (1997) is very similar to Young's method, it has two main differences. These differences are: (1) The discriminated function is not used, and (2) K-nearest neighbors are obtained to current day's weather data and one of the neighbors is selected based on a probability metric that depends on the closeness of the neighbor.

Buishand and Brandsma et al., (2001) extended the nonparametric approach to simulate weather scenarios at multiple locations preserving the dependencies across the locations. They suggested that their nonparametric method avoids the disadvantage of parametric methods.

Yates et al., (2003) used a modification of the K-NN method developed by Lall and Sharma (1996), Rajagopalan and Lall (1999), and Buishand and Brandsma (2001) to generate synthetic climate data and generate alternative climate scenarios. They used meteorological data at several various sites differing in climate and set a successful application of synthetic series that preserved cross correlation and autocorrelation between locations. However, as historical series were used in producing synthetic series applied in the methodology of Yates et al., no new values were generated.

Hobson (2005) proposed a framework for forecasting flows and simulating historic streamflows using a stochastic K-nearest neighbor weather generator together with the Precipitation-Runoff Modeling System (PRMS). He used a deterministic, physically based watershed model to generate alternative streamflow scenarios in the Upper Truckee River Basin on the California and Nevada Border (USA). He incorporated the simulated atmospheric variables into a deterministic watershed model to generate ensembles of streamflow simulations. The generated streamflow simulations are analyzed and then used as input to PRMS. He concluded that his model was adequately capable to simulate the historical flows.

Sharif and Burn, (2006) improved K-NN weather generating model proposed by Yates et al., in order to eliminate the limitation of K-NN model. Their model is capable of generating more extreme events beyond the observed data. The model results were derived according to future climate scenarios that are consistent with projections based on GCM results. They applied their model to the Upper Thames River Basin (UTRB) in the Canadian province of Ontario. They successfully simulated extreme events that are more severe than the observed events for different climate change scenarios.

Apipattanavis et al., (2007) proposed a “hybrid” generator model. They emphasized that the nonparametric models especially the K-NN generators tend to underestimate the length of wet and dry spells. Therefore, they combined the parametric and nonparametric approaches to alleviate the underestimation problem. They defined their model as a semi parametric multivariate and multisite weather generator model. They compared their model with the conventional nonparametric models by using daily weather data from Pergamino, Argentina. They claimed that the spell statistics were captured better in their model.

King (2010) proposed a methodology for the simulation of historical and future climate data using a nonparametric K-NN block resembling weather generator with perturbation. They applied their model to the Upper Thames River basin in Ontario, Canada. They showed that their model can effectively reproduce the historical climate and produced future climate change scenarios based on Atmosphere-Ocean coupled Global Circulation (AOGC) model.

CHAPTER III
METHODOLOGY

Weather generators are stochastic models that aim to generate weather data series "statistically identical" to observed ones. The weather generator method, proposed by Yates et al. (2003), essentially reshuffles the observed data by K nearest neighbors. Therefore, application to multiple sites is achieved by selecting the corresponding day`s weather at all stations. In this way, this method is successful to preserve the spatial correlations of the climate variables. Since this method was developed according to K-NN bootstrap, the advantages and disadvantages of K-NN bootstrap are also valid for this method. The advantages of the K-NN approaches are: (1) preserving the marginal distribution; (2) reproducing linear or nonlinear dependence in the historical data; and (3) easiness in extending to higher order dependence and multi dimensions. However, K-NN approach generates only historical values since it is a resembling technique. Therefore, the synthetic values not seen in the historic record cannot be simulated.

In this weather generator method, the historic daily or monthly weather data are supposed to be available at q stations for N years and for p variables (here p is equal to 3 including precipitation (PCT), maximum temperature (Tmax), and minimum temperature (Tmin). The methodology is described as follows (Yates et al., 2003):

1. The regional means of each variable across all q stations for each month of the historical record are calculated. Regional means vector is denoted as

$$x_t = \begin{bmatrix} \overline{PCT_t} \\ \overline{Tmax_t} \\ \overline{Tmin_t} \end{bmatrix} \tag{3.1}$$

where

$$\overline{PCT_t}=\frac{1}{q}\sum_{j=1}^{q} PCT_{j,t} \quad (3.2)$$

$$\overline{Tmax_t}=\frac{1}{q}\sum_{j=1}^{q} Tmax_{j,t} \quad (3.3)$$

$$\overline{Tmin_t}=\frac{1}{q}\sum_{j=1}^{q} Tmin_{j,t} \quad (3.4)$$

2. A temporal window of length, w is chosen. It is supposed that all days within the window are potential candidates for the simulated data (Sharif and Burn, 2006). Yates et al. (2003) suggested a temporal window of 14 days, which implies that if April 11 is present day then the window covers the days between 4 April and 18 April, for all N years but excluding the value of the present day to avoid the possibility of simulating the same value. A temporal window of 3 months is selected in this study because of unavailability of daily data. The months within the given window are all potential neighbors to the feature vector. Thus, there are L=[(w+1)*N]-1months that are potential neighbors for the present month.

3. The mean vectors $\overline{x_i}$ of the L potential neighbors for each month across all q stations are computed.

4. The pxp covariance matrix, C_t for the current month t using data block of size Lxp is computed.

5. The weather on the first month t (e.g. January) comprising all p variables at q stations is randomly chosen from the set of all January values in the historic records of N years. This is the feature vector, F_t^i and represents generated weather data for month t of year i given for each station. The objective of the following algorithm is to select one of the nearest neighbors to represent the weather for month t+1 of the simulated period.

6. The Mahalanobis distances based on correlation between the current month`s weather, $\overline{x_t}$ and the mean vector of each of the potential neighbor values, $\overline{x_i}$ are calculated by for all i=1 to L:

$$d_i = \sqrt{(\overline{x_t} - \overline{x_i})C_t^{-1}(\overline{x_t} - \overline{x_i})^T} \tag{3.5}$$

Where, T and C_t^{-1} denote the transpose of the vector and the inverse of the covariance matrix, respectively.

7. The Mahalanobis distances are sorted from smallest to largest.

8. The number of K nearest neighbors is selected from Mahalanobis distances metric. Yates et al., (2003) recommend to take $K = \sqrt{L}$.

9. A weighted probability function, w_i ($0<w_i<1$) is assigned to each data within the selected K neighbors as:

$$w_j = \frac{1/j}{\sum_{i=1}^{K} 1/i} \tag{3.6}$$

The cumulative probabilities, p_j, are given by

$$p_j = \sum_{i=1}^{j} w_i \tag{3.7}$$

In this method, the closest neighbors having shortest distance gets the highest weight, whereas the longest distance gets the smallest weight.

10. A nearest neighbor from the K neighbors is selected by using uniform random numbers. The current month`s nearest neighbor is selected by uniform random number, u~U[0,1]. The random number is compared with the cumulative probability (obtained in step 9) to identify the closest data. As such, data that is more similar to current month’s value has a higher probability of being selected as the simulation one.

11. Repeat steps 6 through 10 for each month in the historical record to generate ensembles of synthetic monthly data.

The codes according to algorithm given in Figure 3.1 are written in R programming language to obtain the synthetic series from the observed ones.

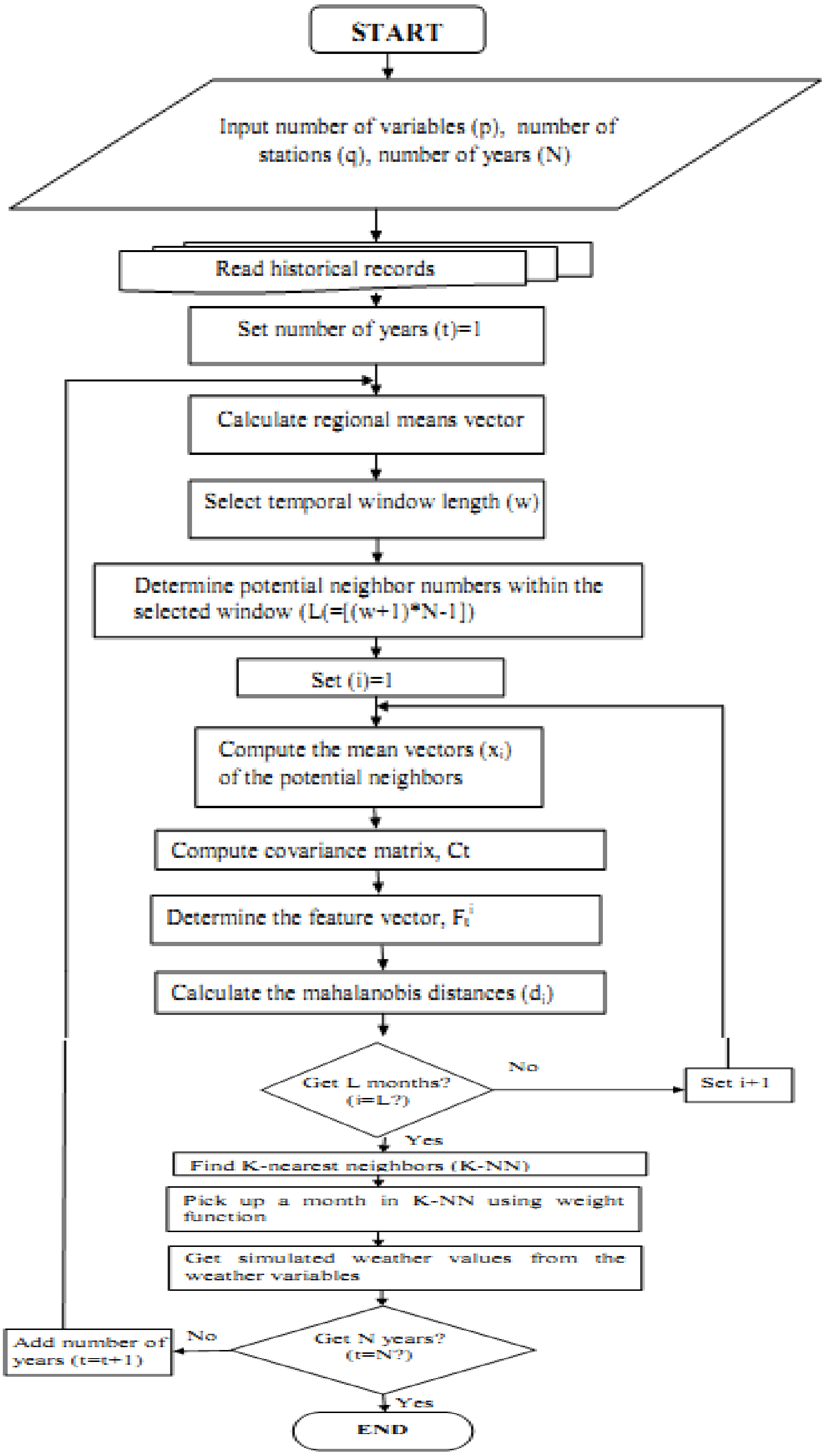

Figure 3.1 Flowchart for methodology

CHAPTER IV

STUDY AREA

The methodology mentioned in this study consists of the impact assessment of climate change in Western Black Sea Basin. Western Black Sea Basin is located on northwestern of Turkey as it can be seen in Figure 4.1. It rises from western border of Kızılırmak delta and lies to the east of Adapazarı and Bilecik.

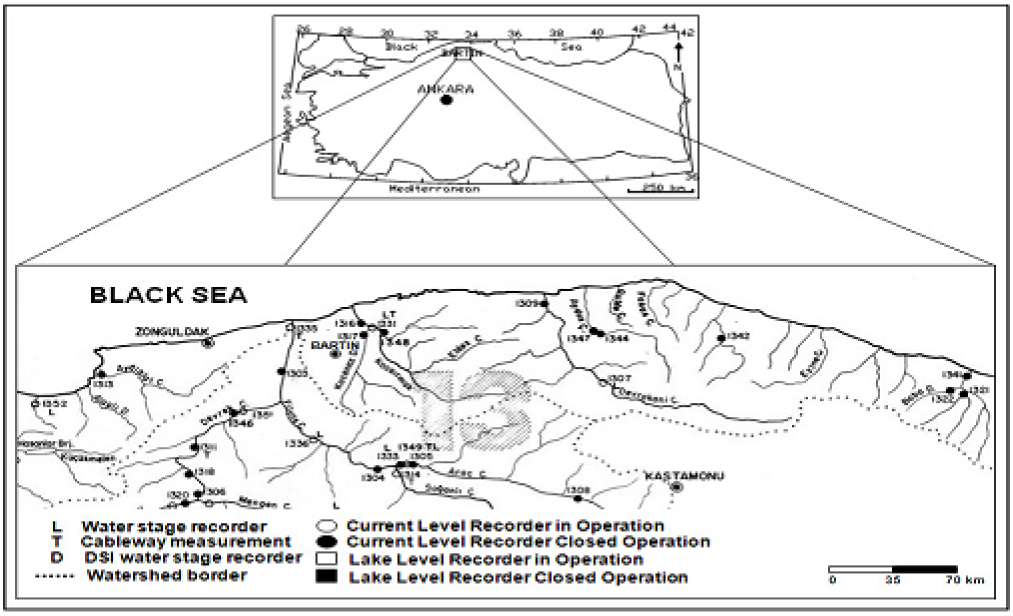

Figure 4.1 Study Area, Western Black Sea Basin

The study area is located at the western part of Black Sea Region in Turkey and has highlands. Black sea climate dominates the region. In West Black Sea region maximum rainfalls during fall months; and minimum rainfalls during spring months. While the precipitation changes month to months, October, November and December have the maximum rainfall. Annual rainfall is 1000-1500 mm. Annual mean temperature is13-15°C. Mean temperature of January is 6-7 °C. Mean temperature of July is 21-23 °C. The most frequent type of soil is podzol in the basin.

Monthly maximum and minimum temperature and precipitation data from 6 stations in and around the basin were used in this study. The selected stations are near the towns of Bartın, Zonguldak and Kastamonu. Site information, including latitude and longitude, altitude, and recording period of each station is listed in Table 4.1.

Table 4.1 List of meteorological stations used in the study

Name of Station	**Station No**	**Latitude**	**Longitude**	**Altitude**	**Period**
Bartın	17020	41.6248	32.3569	33	1975-2010
Zonguldak Bölge	17022	41.4492	31.7779	135	1975-2010
İnebolu	17024	41.9789	33.7636	64	1975-2010
Amasra	17602	41.7526	32.3827	73	1975-2010
Bozkurt	17606	41.9597	34.0037	167	1975-2010
Devrekani	17618	41.5996	33.8345	1050	1975-2010

This particular basin is selected as the study area because many natural hazard events such as flood and landslide have been seen in Western Black Sea Basin. There have been many flood events reported since 1991 by the General Directorate of Disaster Affairs (AFAD). The main reason of those natural hazard events is assumed to be the decrease of winter rainfalls together with the change of seasonal distribution of precipitation, and increase is autumn rainfalls. In addition to this, it is known that rainfalls are irregular, sudden and in high intensities. This results in high runoff rates. The high intensity rainfall over a short time than the capacity of soil absorption causes high runoff rates and flash flood. Furthermore, snowmelt in March and April raises the enhancement of floods. For example, side floods (formed by the waters accumulated in slopes of mountain) occurred in center of Kozcağız (1991), City Center of Bartın (May, 1998), Bartın Inkumu (August, 2004), Zonguldak Saltukova (May, 1998) and finally City Center of Bartın and Arıt(July, 2009). The flood disaster that occurred in the center of City Centers of Bartın in 1998 submerged 70% of the streets and caused to damages in many houses. The flood in Inkumu (2004) occurred as a result of channel overflow in mountains after a heavy rain for three or four hours and all the streets were flooded, landslides occurred, water and electricity distribution is interrupted for two days. The flood in Saltukova (1998) occurred as a consequence of accumulated rain for days and led to houses, cars and workplaces flooded. The rainfall in center of Bartın and Arıt (2009), concluded with a serious rise in stream bed and caused significant economic damage. Another reason of flood disasters in Western Black Sea Basin arises from soil materials that have low absorption rate such as silt and clay.

CHAPTER V
APLICATION AND RESULTS

5.1 Reconstruction of historical data

The objective of reconstructing the historical data was to analyze the performance of the weather generator model in producing statistical characteristics of the historical data. For this purpose, K-NN algorithm is employed by using n data which is less than or equal to (n<N) the available data set, N. The chosen data sub group of size n is considered as the driving data for the model. In order to choose the sub groups, regional monthly anomalies from available data are considered for each variable. To compute the regional monthly deviations (or anomalies) observed in the historic record, firstly regional monthly means are computed as follows:

$$\bar{\psi}_m = \frac{1}{Nqm} \sum_{i=1}^{N} \sum_{j=1}^{q} \sum_{t=1}^{m} \psi_{j,t}^{i} \qquad (5.1)$$

where i, j and t represent the year, the station, and t the month in the year, respectively. Regional monthly means for each year are calculated as:

$$\psi_m^i = \frac{1}{qm} \sum_{j=1}^{q} \sum_{t=1}^{m} \psi_{j,t}^{i} \qquad (5.2)$$

Therefore, regional monthly anomalies are computed as $\delta_m^i = \psi_m^i - \bar{\psi}_m$. Then those anomalies are sorted from minimum to maximum according to their relevant value. As an example, regional precipitation anomalies can be sorted from minimum to maximum for all months and N years such as A={$\delta_1^{1,N}$: (2001 8.9, 1989 50.2,…., 1987 204.8, 1983 227.2); $\delta_2^{1,N}$: (2007 10.7, 1994 18.3,…, 2004 166.6); ….$\delta_m^{1,N}$: (,)…..;$\delta_{12}^{1,N}$: (2006 51.5,…….,1987 202.1, 1980 242)}. The minimum precipitation value for January is under the mean of January data observed as 8.9 mm in 2001 and the maximum precipitation value for January is above the average of 36 years data observed as 227.2 mm in 1983. These values for each month is then assigned a rank index where the first index I_1^1, corresponding to δ_1^1, gets ranking number 1; and the last index I_1^N gets ranking number N (Yates et all., 2003).

A random number, u is generated between 0 and 1 from uniform distribution so as to form sub group years for each month. The index values are computed randomly from 1 to N (=36 in this study) as follows:

$$I_m^{1,N}=INT[N(1-u)]-1 \tag{5.3}$$

These numbers correspond directly to the years in the ranked list. In this way, since data are selected randomly, any year in the data set has the same probability to be chosen. There is the probability that some years could be repeatedly selected and appear multiple times in the sub group.

After obtaining a new sub group as a driving data for the model, K-NN methodology is applied to obtain the synthetic series. Then a set of statistics such as means, standard deviations and skewness coefficients for each month are calculated. These statistics are used to compare the simulated data with the historic data. Box plot of statistics of the reconstructed data for precipitation and average temperature along with the statistics of the observed data are used for the comparison.

In these box plot figures, the box indicates the inter quartile range, the whiskers indicates the 5th and 95th percentile of the simulation, dots show the values outside this range and the horizontal line within each box indicates the median value. The values of the statistics from the observed data are represented by a solid line and marks. The box plots are used to represent the uncertainty range of the mean estimates. One can conclude that the model can adequately simulate the statistics of the historical value when the statistics of the observed data lie within the box of the simulated values. This comparison is conducted for both precipitation and temperature data, which are collected within the study area.

5.1.1 Precipitation

Once the driving data for the model has been determined, the synthetic samples of precipitation are reconstructed. 900 synthetic series, each having the same length as the historical records (i.e., 34), have been generated. For the validation of the model, some statistics of the reconstructed series are compared to those computed from the historical record. Box plot of statistics of reconstructed data for precipitation along with the statistics of observed data are presented in Figure 5.1 through 5.4 for comparison.

It can be seen in Figure 5.1 that the mean of the observed series are generally within the box except for January and October. This indicates that the model preserve the mean of the historical series reasonably well. The largest difference between means is 10 mm that is observed in October. Although there is some bias in the mean values, it can be concluded that the model has reasonably good performance for capturing mean value of observed precipitation data.

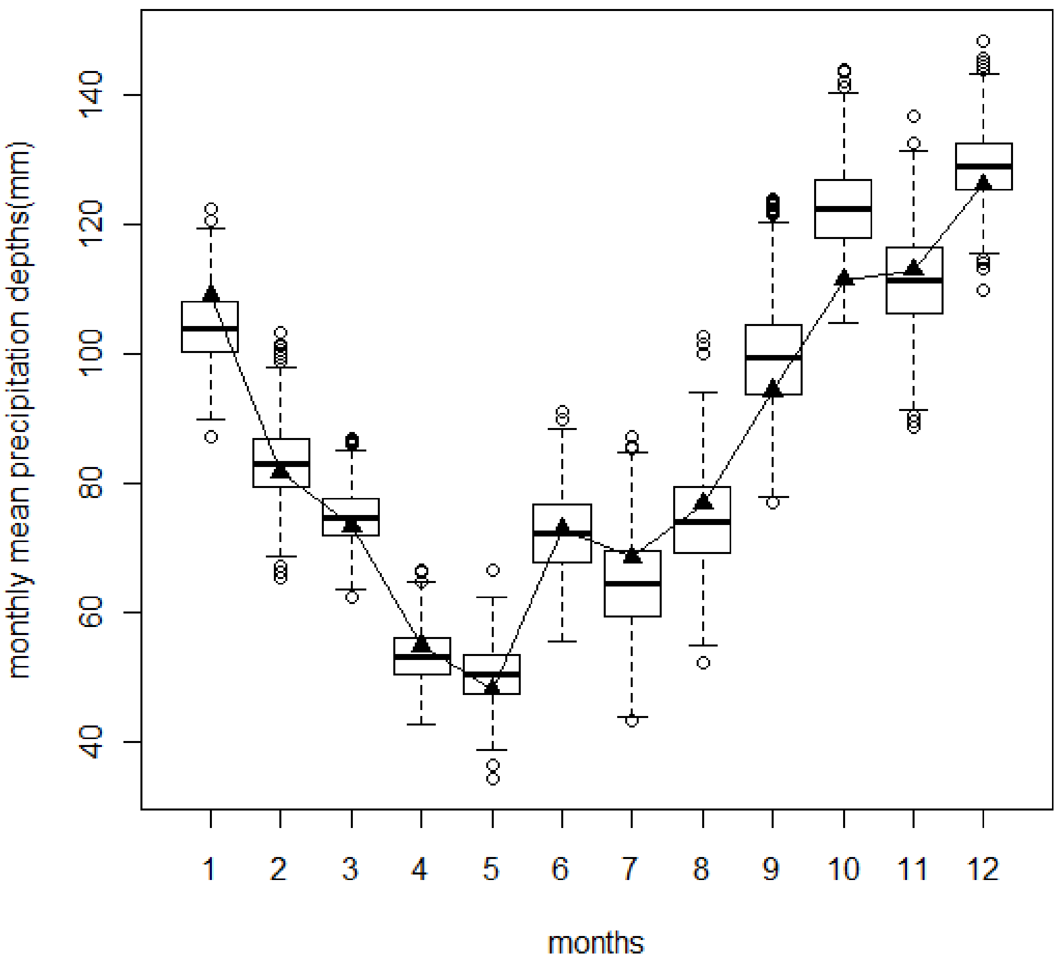

Figure 5.1 Reconstructed monthly mean precipitation depths for Bartın.

In Figure 5.1 the triangles represent the historical mean monthly precipitation not the sub group data series obtained by index values.

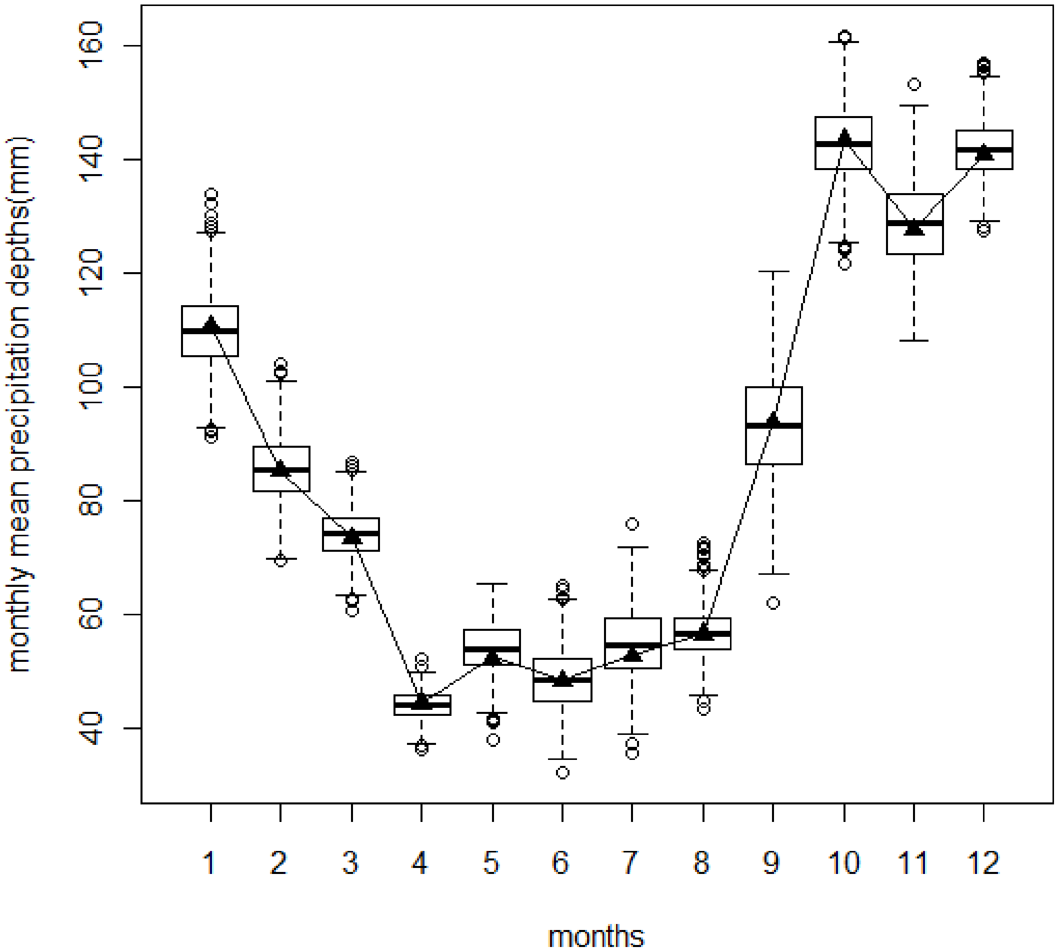

Figure 5.2 Reconstructed monthly mean precipitation depths for İnebolu

Figure 5.2 shows that the median of the simulated data matches with the median of the historical data for each month.

The plots of Figure 5.3 and 5.5 show the standard deviations and skewness coefficients of monthly data for Bartın, respectively. Although the model does a fairly good job at capturing mean of historical data, this is not the case for the other statistics. The higher order moments i.e. variance and skewness show rather high variability. This fact can be explained by the use of monthly data instead of daily data. Porter and Pink (1991) pointed out this issue and suggested the disaggregation method to obtain daily data from observed monthly data (Porter and Pink, 1991).

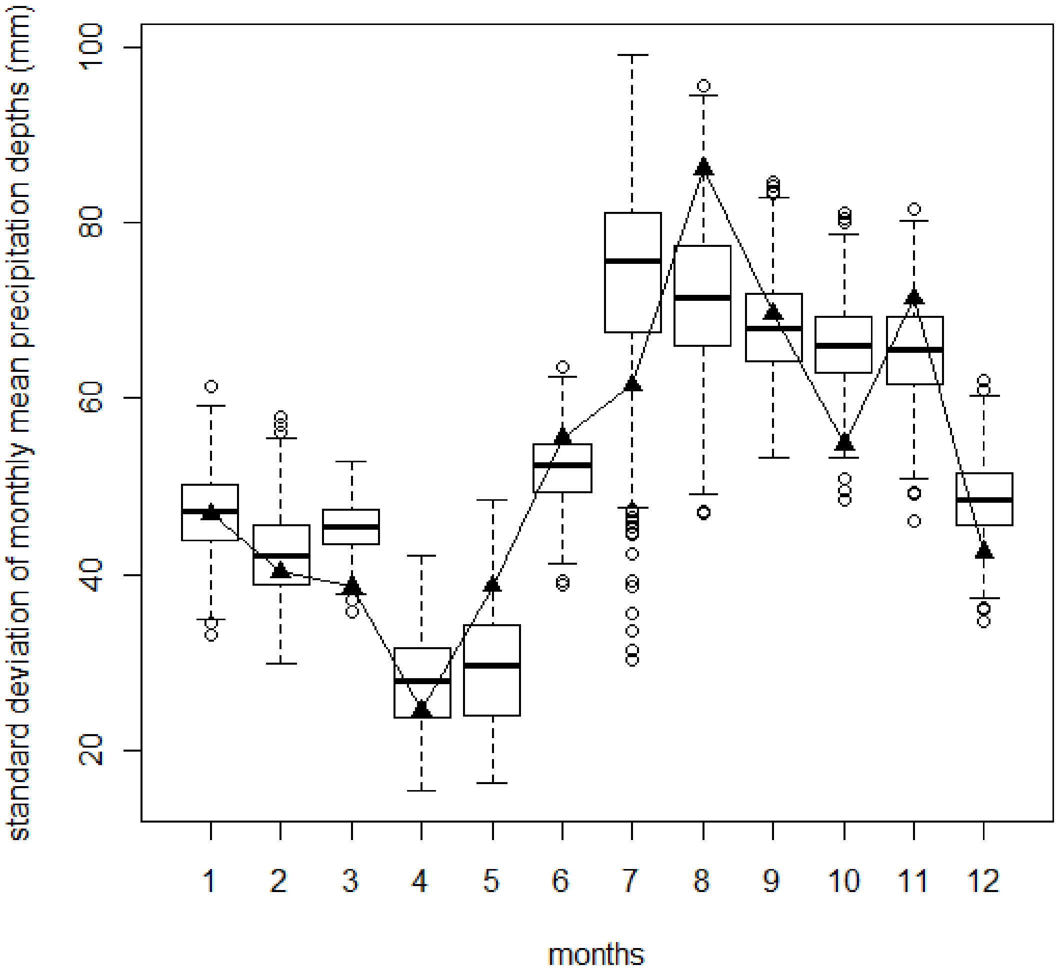

Figure 5. 3 Reconstructed standard deviation of monthly mean precipitation depths for Bartın

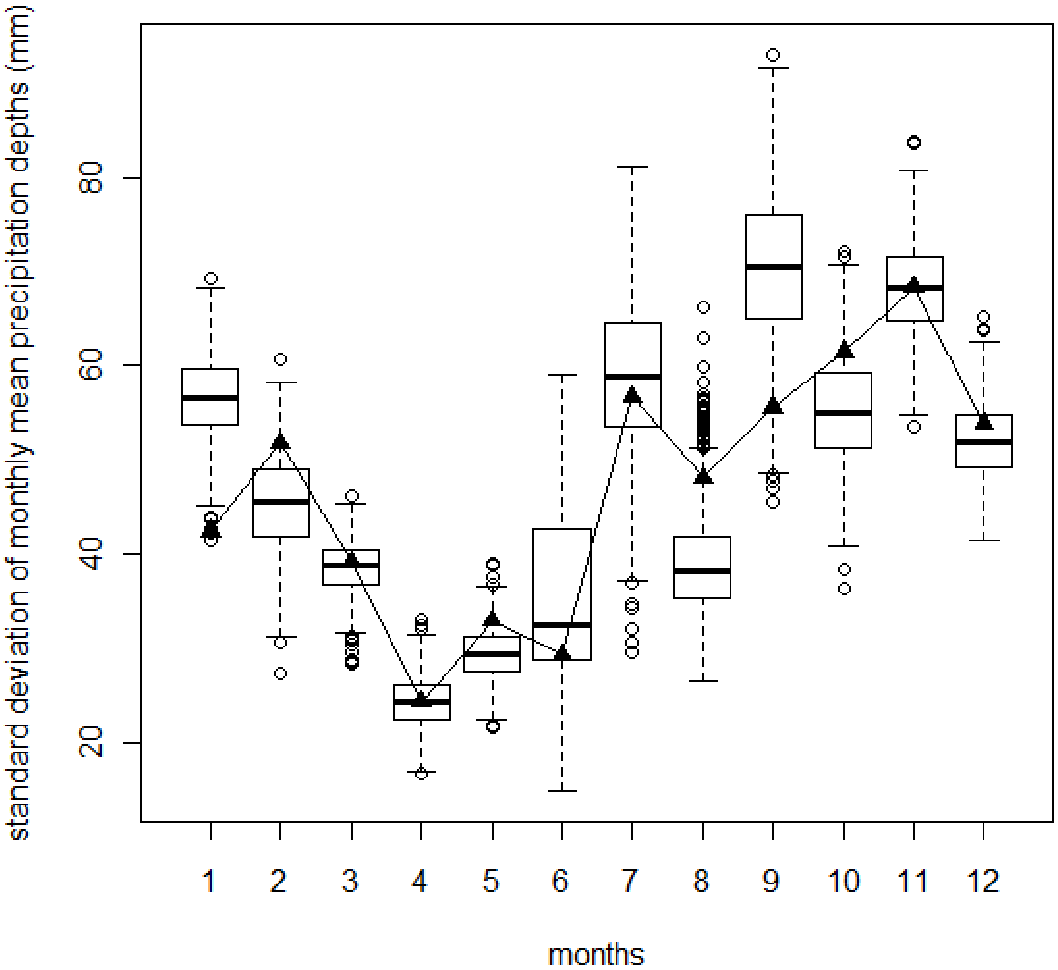

Figure 5.4 Reconstructed standard deviation of monthly mean precipitation depths for İnebolu

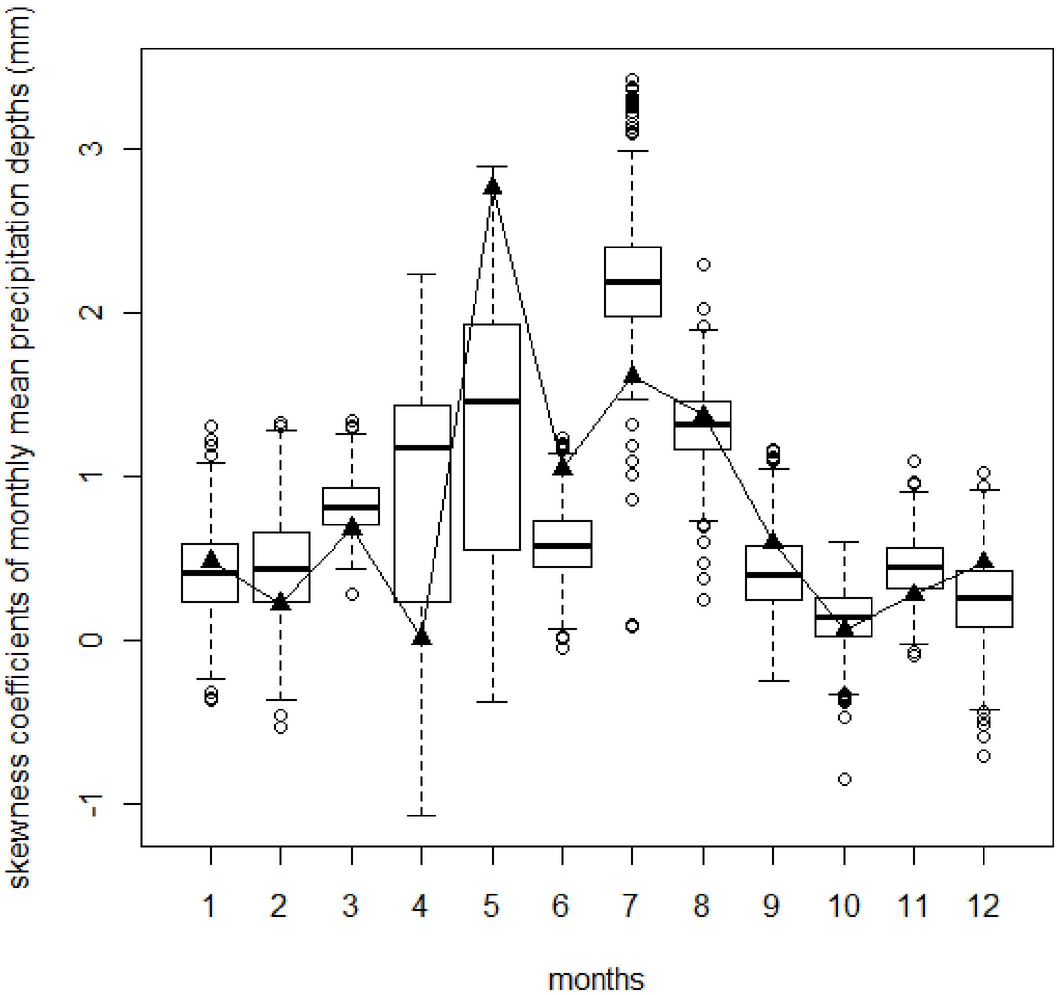

Figure 5.5 Reconstructed skewness coefficients of monthly mean precipitation depths for Bartın

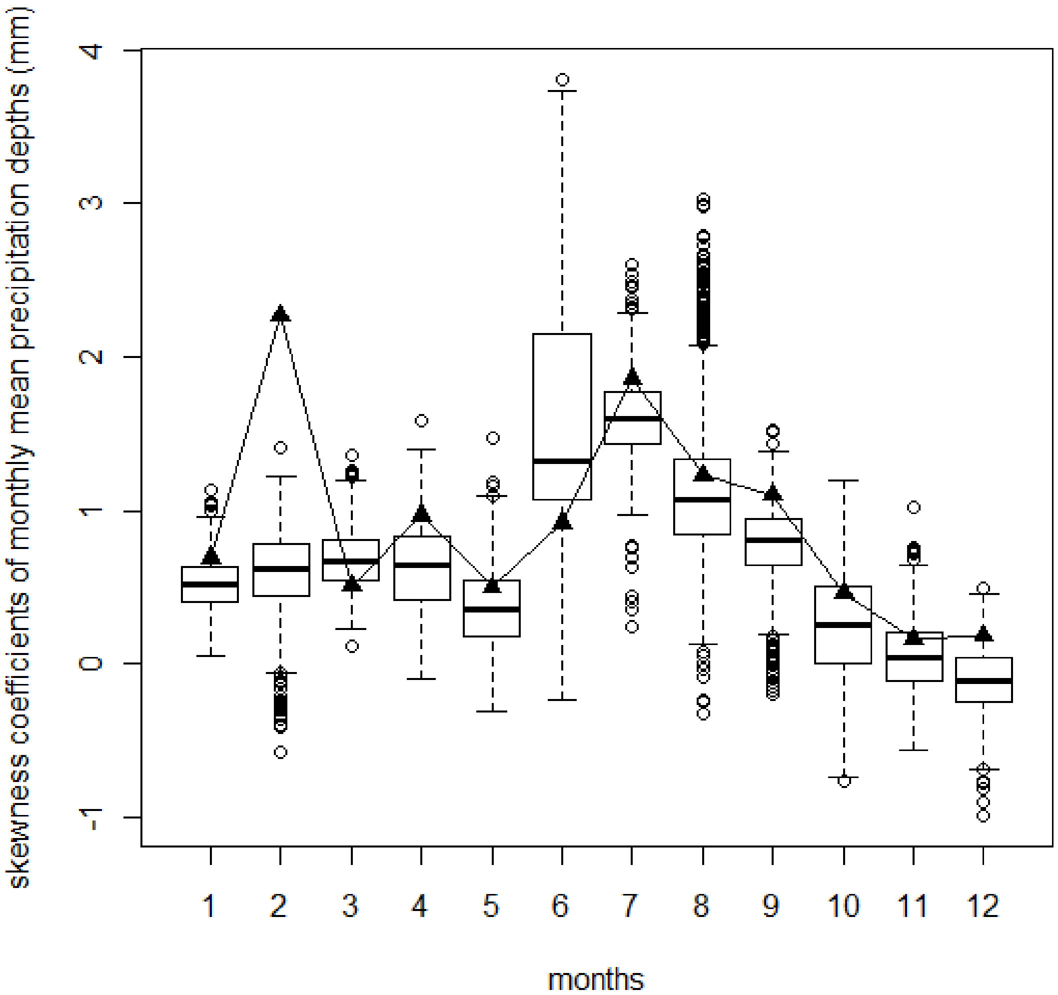

Figure 5.6 Reconstructed skewness coefficients of monthly mean precipitation depths for İnebolu

In fact, underestimation or overestimation problem for standard deviation and skewness commonly arises with weather generator methods. Clark et al., (2004) reported that the method suggested by Yates et al., (2003) have problems with under-prediction of precipitation when they are extended to multiple sites even in daily data (Clark et al., 2004). Yates et al. (2003) noticed these issues and they argued that the shortcoming of the model is acceptable considering the overall performance.

5.1.2 Temperature

For evaluation of the model performance, the monthly mean value of the temperature are computed from both synthetic and observed series and then compared by box-plot method, which is shown in Figure 8. Based on visual inspection, it can be concluded that the model performs well except in June and July.

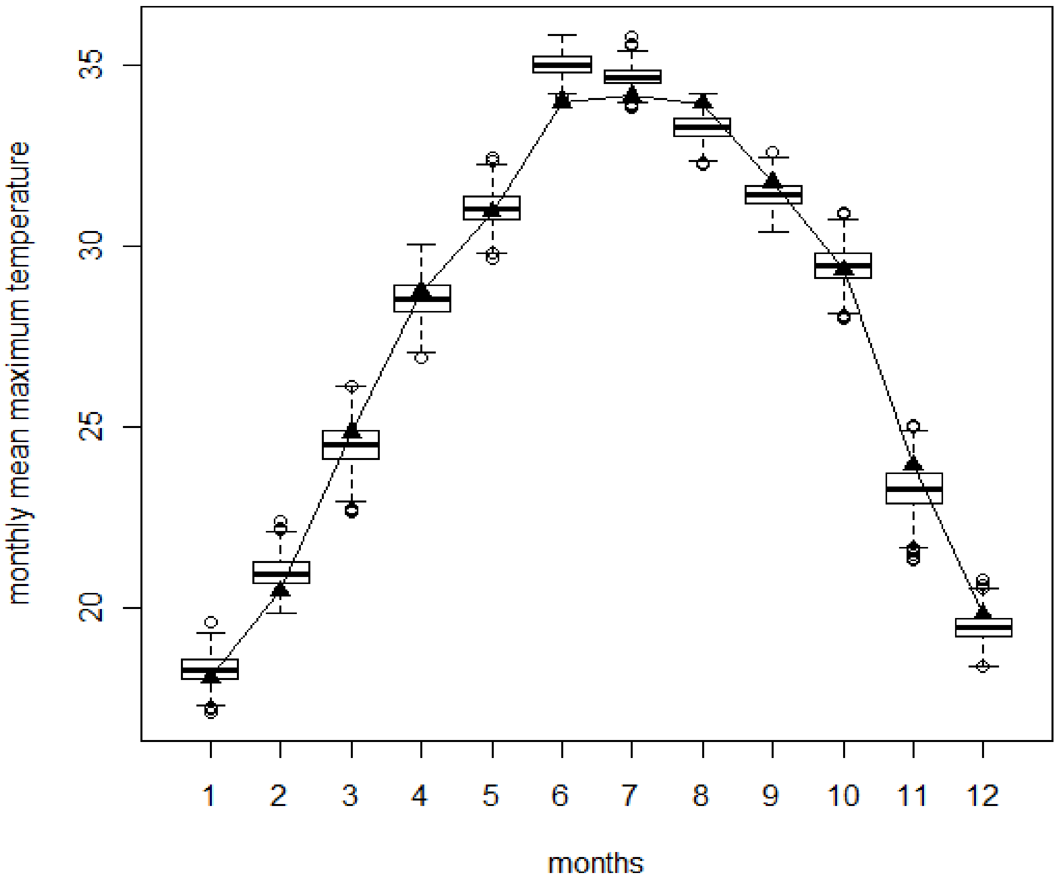

Figure 5.7 Reconstructed monthly mean maximum temperatures for Bartın

The spread of reconstructed values is centered on the historical observation in the case of temperature unlike precipitation. The results are promising for further work.

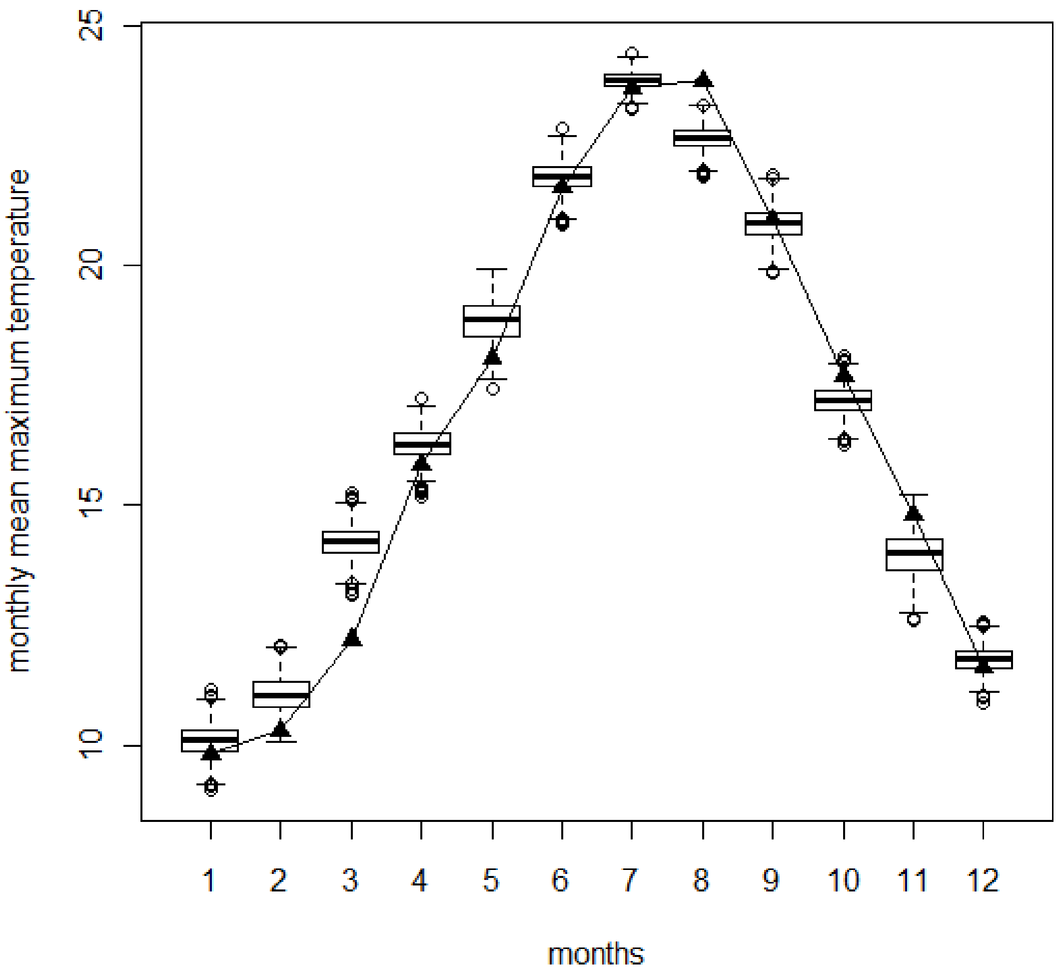

Figure 5.8 Reconstructed monthly mean maximum temperatures for İnebolu

5.1.3 Correlation between Variables at a Station

Lag 1 correlation is used to show the ability of the model to simulate the temporal variability at particular station. Figure 5.9 shows the Lag 1 correlation of precipitation at Bartın station. There are overestimation and underestimation for some months.

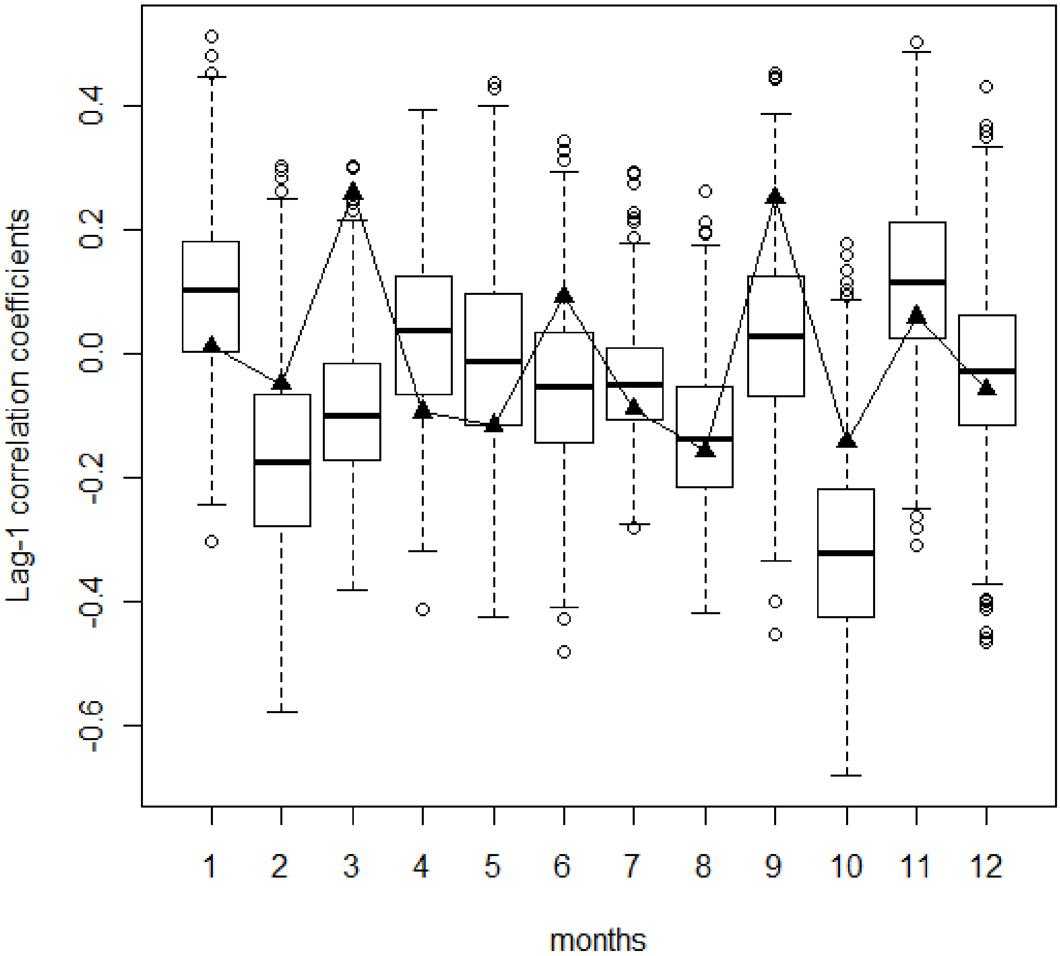

Figure 5.9 Reconstructed Lag-1 correlation coefficients of precipitation for Bartın

Lag 0 correlations between precipitation and maximum temperature at a Bartın station is calculated. The correlation coefficient is used to indicate the relationship between variables at a station on a given month and they are shown in Figure 5. 10.

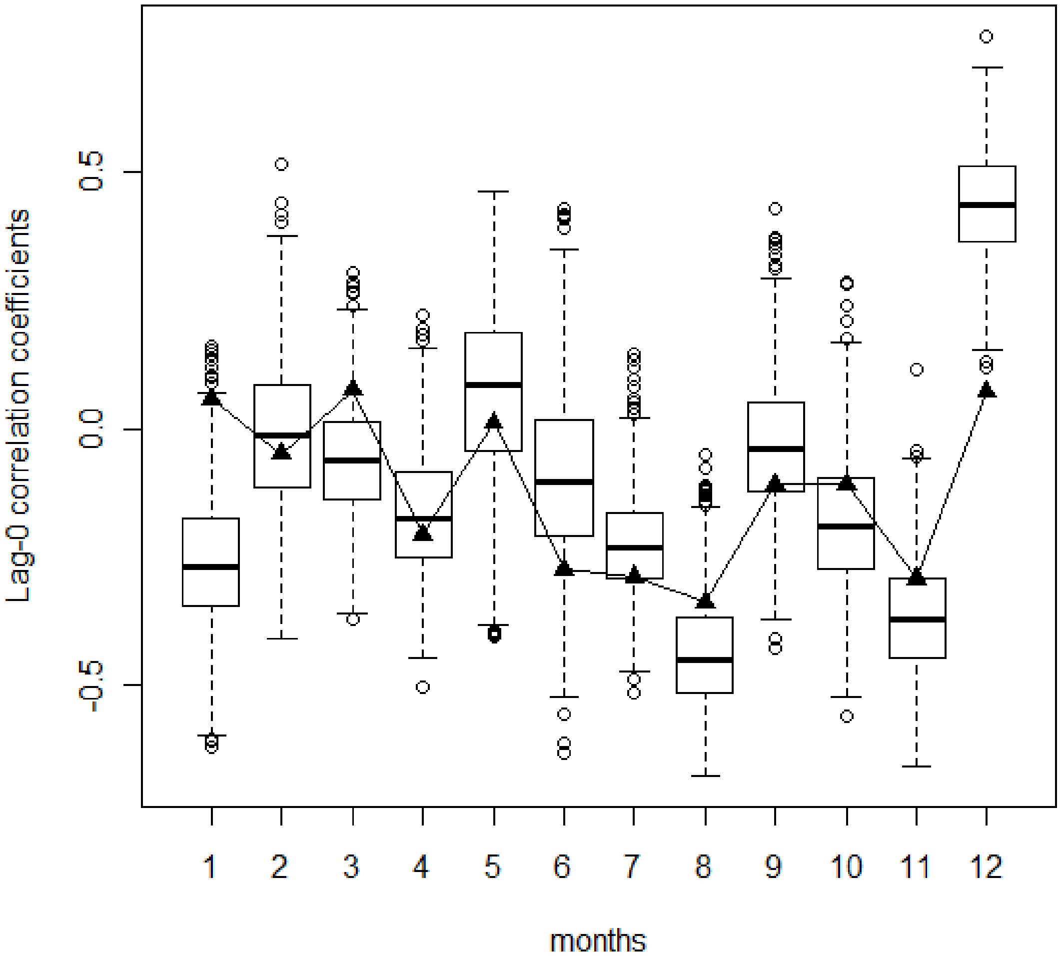

Figure 5.10 Reconstructed Lag-0 correlation between precipitation and maximum temperature for Bartın

The model could not produce the historical statistics well. There are overestimation for the December and underestimation for January and August. This can be explained by having small spatial correlation among observed data in the region. Since the aim is obtaining reconstructed data from regional means values, the observed spatial correlation values play an important role in improving the model performance. Furthermore, the elevation differences among the considered stations and measurement errors may have led to the biased estimates.

5.2 Increasing Average Temperature Scenario

In Section 5.1, sub group selected from the historical data is used to test the model performance. Each data has the same probability to constitute the sub group. We tested the model performance by comparing the model statistics with the statistics of the historical data. Since we were able to reproduce the historical statistics reasonably well, we decided that our model could be considered successful. Actually this is valid under stationary conditions. With stationary condition, researches assume that statistical characteristics of time series do not change with time. However the stationarity assumption is under discussion in recent years because of the evidences of climate change. The generated data is questionable because one cannot say that current climate conditions will be same in the future. Actually it is expected that temperature, precipitation and evapotranspiration will change in time. Some of the researchers claim that overall temperature will increase approximately 2 ^{o}C by the year 2050 and 3-5 ^{o}C by the year 2100. Until now, a lot of studies related with regional-scale simulations under greenhouse scenarios have been conducted as discussed in Chapter 2.

In this study, a new data set comprising years with increased average temperature is constituted using strategic resampling that involves a ranked list of years on the basis of the deviations of mean monthly average temperature from the long term historical mean (Sharif and Burn, 2006) to produce plausible climate change scenarios. The average temperature (TAV) is computed by TAV=T_{max}*0.6+T_{min}*0.4 as in the study by Sharif and Burn (2004).

The integer index given in Eq. 5.3 is rearranged as:

$$I_m^{1,N} = INT\left[N\left(1 - u^{\lambda_m^i}\right)\right] + 1 \qquad (5.4)$$

where λ_m^{i} is a shape parameter of the function. The desired amount of biasing can be achieved by adjusting the shape parameter. When shape parameter is equal to 1.0, we could not notice the bias. if we want to consider the increasing scenario, we should select the shape parameter values greater than 1.0, otherwise we should select shape parameter values less than 1.0 since it tends to bias selection of cold years in ranked series. Yates et al., (2003) suggested that shape parameter value is 2.0 for selection of warm years. Likewise, a value of 0.66 was suggested for selection of bias cold years.

Once the new sub group with increasing temperature is determined, the proposed model is utilized to produce synthetic data condition upon plausible climate change scenarios. The reconstructions of both maximum temperature and mean precipitation under increasing average temperature are shown in Figure 5.11 and Figure 5.12, respectively.

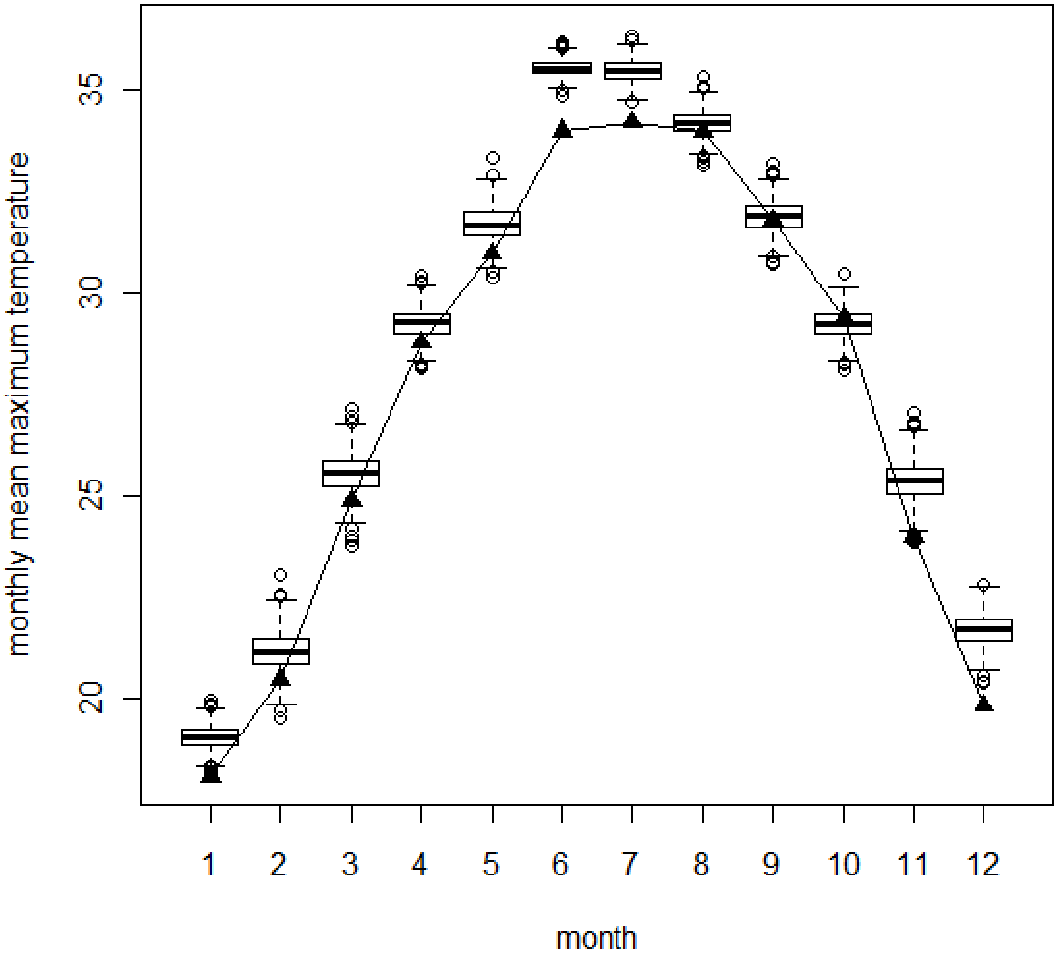

Figure 5.11 Reconstructed monthly maximum temperatures under increasing TAV scenario for Bartın

Figure 5.11 shows that model produce increased monthly maximum temperature over the historical values for all the months. Therefore, we can conclude that our model is capable of producing desired climate scenario.

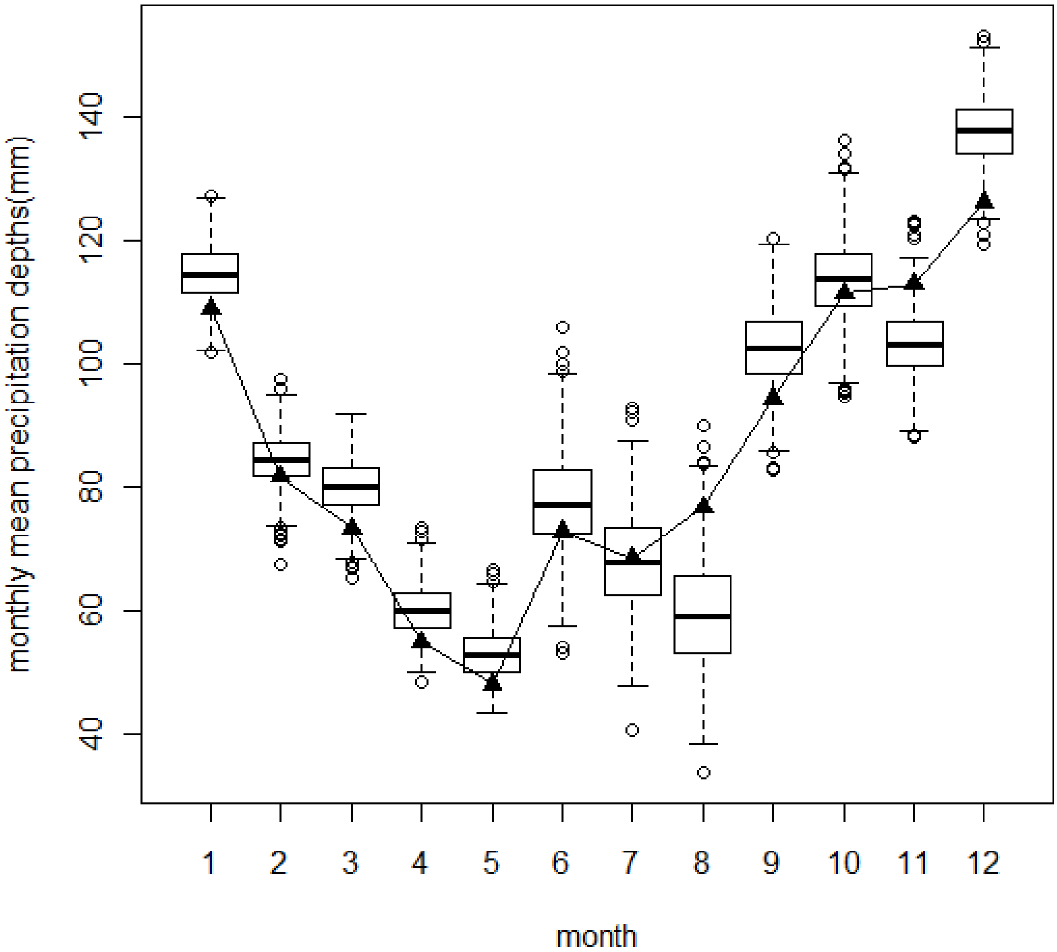

Figure 5.12 Reconstructed monthly mean precipitation depths under increasing TAV scenario for Bartın

Figure 5.12 shows an increase in precipitation for all the months except for August and November. The reason of this observation will be discussed in the following section.

5.3 Increasing Precipitation Scenario

A resampling procedure is carried out to generate scenario with increasing precipitation over months. This procedure is similar to the increasing temperature scenario except that the deviations are computed from the precipitation rather than temperature data. A shape parameter is also selected as 2.0 since it tends to bias the selection of wetter years. After obtaining this new data set, the data sets are used as the driving data set for the K-NN model (Sharif and Burn, 2006). The reconstructions of both mean precipitation and maximum temperature are presented in Figure 5.13 and Figure 5.14, respectively.

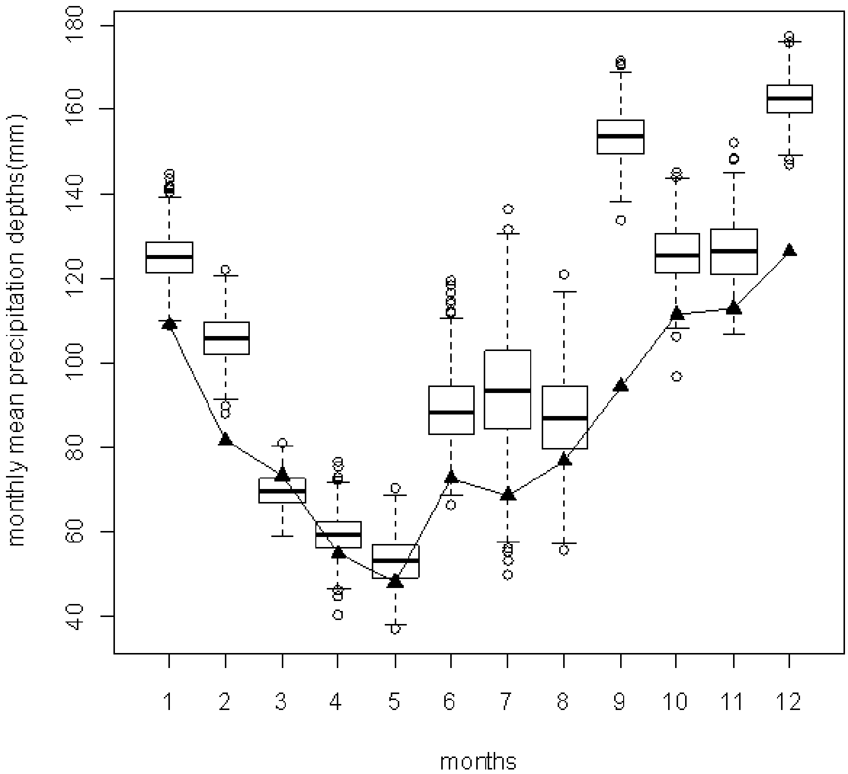

Figure 5.13 Reconstructed monthly mean precipitation depths under increasing precipitation scenario for Bartın

Figure 5.13 shows the model performance under increasing precipitation scenario. The model performs well in generating the climate change scenario except for March. In this month, model underestimates the precipitation depth.

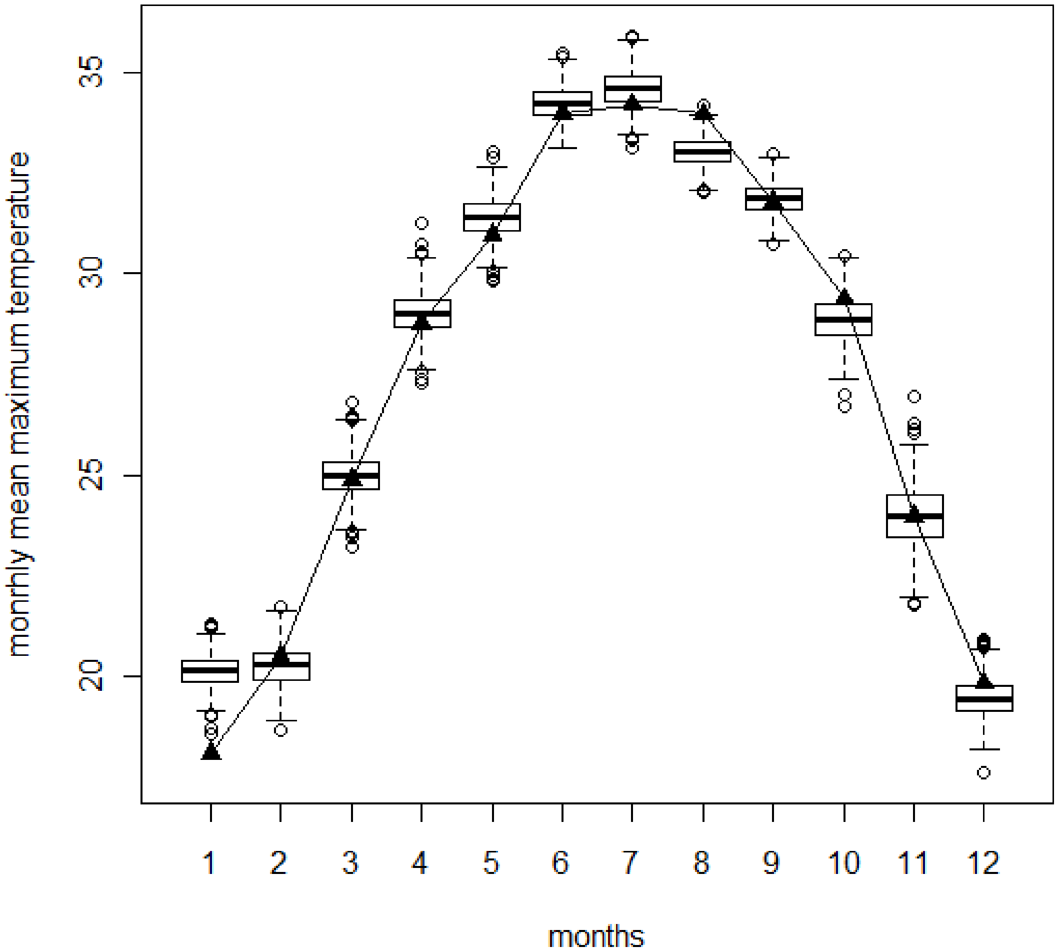

Figure 5.14 Reconstructed monthly maximum temperature under increasing precipitation scenario for Bartın

Model produces some decrease in maximum temperature under increasing precipitation scenario or vice versa.

To evaluate the model performance under different climate change scenario, we should consider the correlation coefficient between precipitation and maximum temperature. Because the magnitude of the correlation coefficient shows the degree of association and sign indicates whether the relationship is direct or inverse. As seen from Figure 5.10 historically there is a positive correlation between maximum temperature and precipitation during December. Because of the positive correlation one can expect an increase in precipitation accompanied by an increase in temperature. However this correlation slightly affects on the temperature under increasing precipitation scenario. For the rest of the months, there is no significant correlation between precipitation and temperature except for August and November. In August and November, a negative correlation between precipitation and temperature is observed. Hence, model produces increase in maximum temperature corresponding to the decrease in precipitation or vice versa. From the model results, it may be concluded that the correlation structure is adequately preserved for the climate change scenarios.

CHAPTER VI
CONCLUSIONS

In this study, the possible impacts of climate change have been examined for Western Black Sea Basin in the event of increasing precipitation and temperature. For this reason, average temperature and average precipitation data were produced for each station in accordance with regional stations selected under the scenarios of various climate changes such as increasing temperature and precipitation with integer index.

With the purpose of evaluating the model performance for each station, sub groups of regional average values were obtained without introducing any shape parameter into the model. The values obtained for each station and the methodology based on K-NN algorithm were utilized and average monthly precipitation and maximum temperature values were simulated. Simulated and historically observed values were shown with the box-plot. Since moment values obtained from historically observed mean monthly precipitation depths and maximum temperature values were within box plots obtained from 900 produced synthetic data, it is concluded that the model is successful in producing historical data. These results motivated us to produce alternative climate change scenarios.

The correlation between monthly mean precipitation depths and maximum temperature values observed in stations were calculated. It was found out that there is no any statistically meaningful correlation between them. However, while a positive correlation value obtained between them in December, negative correlation values were obtained in August and November. It was seen that the model is able to preserve these correlation while generating synthetic series under possible increasing precipitation and temperature scenarios.

The sub groups of each station under possible increasing average temperature scenario are obtained taking shape parameter as two and then these sub group values were used in generating monthly mean temperature and precipitation depths. The aim of choosing shape parameter as two was to be able to select high observed average temperature values in sub groups. When selected sub group values for each station were reproduced with K-NN methodology, the model was capable of producing higher values than observed average temperature values. The same procedure was repeated under increasing precipitation scenario.

It should be emphasized that the model produced mean monthly precipitation depths corresponding to the produced synthetic average temperature values under increasing average temperature scenario successfully. From this point of view, the correlation between monthly total precipitation and temperature values seems to be significant. Also, it is observed that the model showed increase in monthly average precipitation values corresponding to increasing average temperature values in December. Since the correlation coefficient is positive in that month, this result is not surprising. On the other hand, the model produced lower values than observed data under increasing average temperature scenario in August and November. This situation could not be observed clearly in synthetic average temperatures under increasing total precipitation scenario.

As a result, producing synthetic series gains important role in assessment of watershed vulnerability to extreme events such as floods and droughts under changing climate conditions of the basin. Since reconstructed series can be utilized as input parameters in the hydrologic models they will be generated in accordance with the concerned climate change scenarios. It is certain that produced or reconstructed series will play an important role in the planning and management of water resources under changing climate conditions.

REFERENCES

Aksay, C. S., Ketenoğlu, O., and Kurt, L. (2005). Global Warming and Climate Change. Selcuk University, *Faculty of Arts and Science Journal*, in Turkey **25**, 29-41.

Apipattanavis, S., Podesta, G., Rajagopalan, B., and Katz, R. W. (2007). A semiparametric multivariate and multisite weather generator, *Water Resour. Res*., **43**, W11401, doi:10.1029 /2006WR005714.

Başaran, M., Karakaş, S., Demirel, B., Yalçınkaya, R., Uzun, A.S., and Güneş, A. (1998). Bartın city inundated technical report written by Trade chamber and Democratic Mass organization, in Turkey.

Buishand, T. A., and Brandsma, T. (2001). Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling, *Water Resour. Res.*, **37** (11), 2761– 2776.

Burn, D.H., Simonovic, S.P. (1996). Sensitivity of reservoir operation performance to climatic change. Water Resources Management **10**, 463–478.

Clark ,M. P., Gangopadhyay, S., Brandon , D., Werner, K., Hay, L., Rajagopalan, B., and Yates, D. (2004). A resampling procedure for generating conditioned daily weather sequences,*Water Resour.Res*., 40, W04304, doi:10, 1029/2003wr002747.

Hobson, N. A. (2005). Use of a Stochastic Weather Generator in a Watershed Model for Stream flow. Simulation. M.Sc. Thesis, Faculty of the Graduate School, University of Colorado. IPCC (2007). Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the

Intergovernmental Panel on Climate Change”. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp.

Jones, W., Rex, R. C., and Threadgill, D. E. (1972). A simulated environmental model of temperature, evaporation, rainfall, and soil moisture, Trans. ASAE, **15**, 366– 372.

King, L.M. (2010). Application of a K-nearest neighbor weather generator for simulation of historical and future climate variables in the upper Thames river basin. M.Sc. Thesis, Graduate Program in Civil and Environmental Engineering, The University of Western Ontario.

Lall, U., and Sharma, A. (1996). A nearest neighbor beat strap for time series resembling. *Water Research* **32** (3), 679-693.

Mortsch, L., Hengeveld, H., Lister, M., Lofgren, B., Quinn, F., Slivitzky, M., Wenger, L. (2000). Climate change impacts on the hydrology of the Great Lakes—St. Lawrence System. Canadian Water Resources Journal **25** (2), 153–179.

Nicks, A.D., Harp, J.F. (1980). Stochastic generation of temperature and solar radiation data. *Journal of Hydrology* **48**, 1–7.

Porter, J.W., and Pink, B.J. (1991). A method of synthetic fragments for disaggregation in stochastic data generation. International Hydrology and Water Resources Symposium, Perth, Australia, 781-786.

Rajagopalan, B., Lall, U., Tarboton, D. G., and Bowles, D. S. (1997). Multi-variate nonparametric resampling scheme for generation of daily weather variables, Stochastic Hydrol. Hydraul., **11** (1), 523–547.

Rajagopalan, B., and Lall U. (1999). A k-nearest-neighbor simulator for daily precipitation and other variables, Water Resour. Res., **35** (10), 3089–3101.

Richardson, C.W. (1981). Stochastic simulation of daily precipitation, temperature and solar radiation. Water Resources Research **17** (1), 182–190.

Sharif, M., and Burn, D. H. (2004). Development and Application of K-nearest neighbor weather generating model. 57th Canadian Water Resources Association Annual Congress, June 16-18, Montreal, Qc, Canada.

Sharif, M., and Burn, D. H. (2006). Simulating climate change scenarios using an improved K-nearest neighbor model. *Journal of Hydrology* **325**: 179-196.

Southam, C.F., Mills, B.N., Moulton, R.J., Brown, D.W. (1999). The potential impact of climate change in Ontario's Grand River basin: water supply and demand issues. Canadian Water Resources Journal **24** (4), 307–330.

Soykan, İ. (1995). Climatic Changes in The World. General Directorate of Rural Services, Water Management Research Group Assembly, Conference, Menemen, Turkey.

Sunay, Ç. (2000). Climate is changing. *TUBITAK Scientific and Technical Journal* , **392**.(in Turkey)

Westernmacott, J.R., and Burn, D. H. (1997). Climate change effects on the hydrological regime within the Churchill–Nelson river Basin. *Journal of Hydrology* **202**, 263–279

Yates, D., Gangopadhyay, S., Rajagopalan, B., and Strzepek, K. (2003). A technique for generating regional climate scenarios using a nearest-neighbor algorithm. *Water Resources Research*, **39** (7), 1199-1213.

Young, K.C. (1994). A Multivariate chain model for simulating climatic parameters from daily data. *Journal of Applied Meteorology* **33**:661-671.