JULY 2019

M.Sc. Thesis in Civil Engineering

ABDULHADI PALA

REPUBLIC OF TURKEY
UNIVERSITY OF GAZİANTEP
GRADUATE SCHOOL OF
NATURAL & APPLIED SCIENCES

MODELING THE CLIMATE CHANGE IMPACT ON A RIVER BASINS BY USING GLOBAL CLIMATE CHANGE, HEURISTIC METHODS, AND WATERSHED MODELING SYSTEM

M. Sc. THESIS
IN
CIVIL ENGINEERING

BY
ABDULHADİ PALA
JULY 2019

MODELING THE CLIMATE CHANGE IMPACT ON A RIVER BASINS BY USING GLOBAL CLIMATE CHANGE, HEURISTIC METHODS, AND WATERSHED MODELING SYSTEM

M.Sc. Thesis

in

Civil Engineering

University of Gaziantep

Supervisor

Prof. Dr. Aytaç GÜVEN

By

Abdulhadi PALA

July 2019

REPUBLIC OF TURKEY
UNIVERSITY OF GAZİANTEP
GRADUATE SCHOOL OF NATURAL & APPLIED SCIENCES
CIVIL ENGINEERING DEPARTMENT

Name of the Thesis : Modeling the climate change impact on a river basins by using Global Climate Change, Heuristic Methods, and Watershed Modeling System

Name of the Student : Abdulhadi PALA

Exam Date : 17.07.2019

Approval of the graduate school of natural and applied science.

Prof. Dr. Ahmet Necmeddin YAZICI
Director

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

Prof. Dr. Hanifi ÇANAKÇI
Head of Department

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

Prof. Dr. Aytaç GÜVEN
Supervisor

Examining Committee Members:	Signature
Prof. Dr. Aytaç GÜVEN	…..........………
Prof. Dr. Mustafa GÜNAL	…..........………
Assist. Prof. Dr. Adem YURTSEVER	…..........………

I hereby declare that all information in this document has 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.

Abdulhadi PALA

ABSTRACT

MODELING THE CLIMATE CHANGE IMPACT ON A RIVER BASINS BY USING GLOBAL CLIMATE CHANGE, HEURISTIC METHODS, AND WATERSHED MODELING SYSTEM

ABDULHADİ PALA
M.Sc. in Civil Engineering
Supervisor: Prof. Dr. Aytaç GÜVEN
July 2019
113 pages

Investigation of the hydrological impacts of climate change at the local scale hydropower potential requires the use of a statistical downscaling technique. Global Circulation Models (GCMs) are essential tools used for the future change of climate impact including precipitation and temperature using different scenarios. The GCMs outputs cannot be used directly due to the uncertainty in the spatial resolution between the GCMs and hydrological models. Statistical downscaling techniques involves the development of relationships between the large-scale climatic parameters and the local variables such as temperature and precipitation. In this study, statistical downscaling of monthly areal mean precipitation for Goksun River basin, which is delineated by using Watershed Modeling System (WMS) program, in Turkey was constructed using the Group Method of Data Handling (GMDH), Gene Expression Programming (GEP), and Support Vector Machine (SVM) techniques and compared results to Linear Regression model. The R-value for precipitation SVM model is 0.863, 0.788 for training and testing periods, respectively. The results showed that SVM has the best model performance than the other proposed downscaling models, however, AIC values showed the GEP model has the lowest AIC value. The SVM downscaling result of the GCM scale simulated under both GCM A1B and A2 emission scenarios for areal mean precipitation for study area has been explored and analyzed. The simulated results for both scenarios show a similarity in their average precipitation. Generally, both scenarios anticipate a decrease in the average monthly precipitation during the simulated periods. Therefore, the results of future projections show that the precipitation will decrease during the period of 2021-2100.

Keywords: GCM, Statistical downscaling, GEP, GMDH, SVM, Precipitation, WMS

ÖZET

KÜRESEL İKLİM DEĞİŞİKLİĞİ, SEZGİSEL YÖNTEMLER VE HAVZA MODELLEME SİSTEMİ KULLANARAK İKLİM DEĞİŞİKLİĞİNİN NEHİRLER ÜZERİNDEKİ ETKİSİNİN MODELLENMESİ

ABDULHADİ PALA
Yüksek Lisans Tezi, İnşaat Mühendisliği Bölümü
Danışman: Prof. Dr. Aytaç GÜVEN
Temmuz 2019
113 sayfa

İklim değişikliğinin hidrolojik etkilerinin yerel ölçekte hidroelektrik potansiyelinde araştırılması, istatistiksel bir ölçek küçültme tekniğinin kullanılmasını gerektirir. Küresel Dolaşım Modelleri (GCM'ler), farklı senaryolar kullanarak yağış ve sıcaklık dahil olmak üzere gelecekteki iklim etkisinin değişimi için kullanılan temel araçlardır. GCM verileri, GCM'ler ve hidrolojik modeller arasındaki uzaysal çözünürlükteki belirsizlik nedeniyle doğrudan kullanılamaz. İstatistiksel ölçek küçültme teknikleri, büyük ölçekli iklim parametreleri ile sıcaklık ve yağış gibi yerel değişkenler arasındaki ilişkilerin geliştirilmesini içerir. Bu çalışmada, Türkiye'de Havza Modelleme Sistemi (WMS) programı kullanılarak tanımlanan Göksun Nehri Havzası için aylık ortalama yağış istatistiksel olarak ölçek küçültülmesi, GMDH, GEP ve SVM teknikleri kullanılarak yapıldı ve sonuçlar doğrusal regresyon modeli ile karşılaştırıldı. Yağış SVM modeli için R-değeri sırasıyla kalibrasyon ve validasyon modelleri için 0.863, 0.788'dir. Sonuçlar, SVM'nin önerilen diğer ölçek küçültme modellerinden daha iyi model performansına sahip olduğunu gösterdi, ancak AIC değerleri GEP modelinin en düşük AIC değerine sahip olduğunu gösterdi. SVM ölçek küçültme teknik sonuçları GCM ölçeğinde simulasyonu hem A1B hem de A2 emisyon senaryoları altında çalışma bölgesindeki alansal ortalama yağış için araştırılmış ve analiz edilmiştir. Her iki senaryo için simüle edilen sonuçlar, ortalama yağışlarında benzerlik göstermektedir. Genel olarak, her iki senaryo da simüle edilen dönemlerde ortalama aylık yağışta bir düşüş öngörmektedir. Bu nedenle, gelecekteki projeksiyonların sonuçları, 2021-2100 döneminde yağışların azalacağını göstermektedir

Keywords: GCM, İstatistiksel ölçek küçültme, GEP, GMDH, SVM, Yağış, WMS

ACKNOWLEDGEMENT

I would like to express my sincere gratitude and honest respect to my supervisor, Prof. Dr. Aytaç Güven for his endless supports and contributions during this thesis. There is no doubt that I would not have been completed this thesis without his advice and criticism. The door to Prof. Aytaç GÜVEN office was always open whenever I ran into a trouble spot or had a question about my research or writing.

TABLE OF CONTENTS

Page

LISTS OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

IPCC	Intergovernmental Panel on Climate Change
NASA	National Aeronautics and Space Administration
EPA	Environmental Protection Agency
GCM	General Circulation Model
RCM	Regional Climate Model
ANN	Artificial Neural Networks
GMDH	Group Method of Data Handling
GEP	Gene Expression Programming
SVM	Support Vector Machine
ERM	Empirical Risk Minimization
LR	Linear Regression
P_M	Mean Precipitation
NCEP	National Centers for Environmental Prediction
NCAR	National Center for Atmospheric Research
SRES	Special Report on Emissions Scenarios
CGCM3	3rd form of the Coupled Global Climate Model
HadCM3	Hadley Centre 3rd Generation
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
R	Correlation Coefficient
AIC	Akaike Information Criterion
MSE	Mean Square Error
WMS	Watershed Modelling System
DEM	Digital Elevation Model

CHAPTER 1

INTRODUCTION

1.1 General

The world's climates are principally determined by the effects of solar radiation. It can be claimed that the balance between incoming and outgoing solar radiation is the reason for the same average temperature for centuries. Although the incident solar radiation remained the same, the rate of outgoing radiation (longwave radiation) reduced due to the augmentation concentrations of aerosols, clouds, and greenhouse gases in the atmosphere (IPCC, 2013, NASA, 2014, EPA, 2014). Comparing the latest observation of the climate standard with initial observations and research shows that global warming continues. Whatever the cause of this phenomenon, climate change is one of the biggest environmental challenges we face this century.

The lack of information has the combined effects that climate change can have on a watershed. With rising temperatures, the dynamics of winter precipitation may change, snowfall tends to melt (although more winter precipitation fills this gap), evaporation can increase and higher water vapor pressures can produce more rainfall. Due to the complex interactions between these different processes, the separation of the various complex bonds in the development of hydrological regime and flood risk requires a hydrological model based on physical processes. The effects of warming on flood risk are not yet fully understood but are largely related to society.

The interest in generating hydropower has increased in recent years as it is a renewable, efficient and reliable source of energy and has an advantage in reducing gas concentrations in the atmosphere. Greenhouse effect occurs as a result of human activities. At the same time, hydropower is one of the most vulnerable global warming industries because water resources are closely linked to climate change. In fact, the impact of climate change on water availability is expected to affect hydropower generation.

Climate change is likely to increase the incidence of drought and water scarcity in some regions suffering from water stress, such as in South and Southeastern Europe (Bernstein et al. 2007)

The primary task of environmental researchers is to collect all information about the climate and its driving forces and to anticipate future changes. Temperature and precipitation are essentially key features of climate change. Changes in individual parameters can have a direct effect on evaporation magnitude and on quantity /natural of runoff amount (Konig, 2008).

Climate change is expected to have both global and local effects on the surface of the world. One of these effects would be to reduce the amount of fresh water and thus increase the demand for water (Kusangaya et al., 2014, Lespinas et al., 2014). The management of water resources have become a very significant topic. Human activities have changed the land cover of the earth and, as a result, the emission of some important gases, aerosols, and atmospheric gas concentration is changing (EPA, 2014). These weather events are changing the global energy budget. Many studies have shown the multiple impacts of climate change on human activity. The main impacts of climate change occur in the atmosphere, the land surface, the glaciers, and oceans.

The General Circulation Model (GCM) is regarded as the most widely used model to simulate the response of the global climate system to increase the greenhouse gas concentrations and to supply an overall estimate of climatic variables (Ghosh and Mujumdar, 2007). GCMs have a horizontal resolution of 300-500 km worldwide. This value limits the capability to estimate the impact of climate change at the regional or local level. GCM outputs cannot generate local climate details in the finer spatial resolution due to the uncertainty in the spatial resolution between the GCM and hydrological models. For this reason, several downscaling methods have been employed to transform the GCM outputs from coarse spatial resolution to a finer spatial resolution that can be used directly for forecasting the climate change at the local scale output.

Downscaling can be described as the technique of getting information from large-scale climate known as large-scale atmospheric variables. By using downscaling models with large-scale climate variables, to make estimation at local scales can be done. They can be classified as dynamical and statistical downscaling.

Statistical downscaling methods developed an empirical relationship between large-scale weather factors (GCM variables) and local-scale factors such as precipitation, temperature (Anandhi et al., 2013). Statistical downscaling models are comparatively low-cost and need less computations. Substantial development has been already done in improving statistical downscaling models. (Huth 1999; Hashmi et al., 2011) The statistic downscaling use Artificial Neural Networks (ANNs) and Genetic programming. Dynamic downscaling which relates to regional climate model (RCM) and limited zone, nesting by GCM outputs, which also has a resolution in horizontal axis about 10-100 km.

The power of the computer can be used for prediction of climate. Different complex programs are run on the supercomputer for providing different information about climate, for example, the Artificial Neural Network that was used to predict the rainfall (Choudhary, 2016). The artificial intelligence application gave also the possibility to have a new forecasting model on the earth temperature increase due to climate change (Bevilacqua, 2008).

In this study, suggested statistical downscaling method can be classified as Group Method of Data Handling (GMDH), Gene Expression Programming (GEP) which is widely used on hydrology to simulate downscaling techniques, and Support Vector Machine (SVM) to examine the climate change effect on hydropower potential. All proposed downscaling models are based on regression statistical downscaling because of its ease to use and require less computation. And lastly, a simple linear regression is done for comparison.

1.2 Objectives

This thesis examines and explores the impact of climate change on Goksun River basin using the proposed downscaling model between large-scale GCM factors and monthly areal mean precipitation of the basin.

The objectives of this study can be outlined as follows;

- To propose the most appropriate GCM predictors among the 26 large-scale weather variables, which can more describe areal mean precipitation for past, current and future period.

- To develop four different statistical downscaling models, that correlates the local variable P_M and downloaded GCM variables at the centroid of the basin.

- To train and validate the models, so as to select the best downscaling model for the future projections.

- To simulate the future projection of monthly P_M based on different emission scenarios.
- To assess the projected monthly P_M and compared with the observed baseline periods.

1.3 Importance of study

The following subjects describe the primary benefits to be made out of this study:

- To evolve the administration of water resources and more exploratory of hydropower potential.
- To conduce to the improvement of irrigation projects and agricultural production.
- To present appropriate lands for human habitation in advance, and warning people to alter the lifestyle in the event of disasters.

1.4 Scope of the study

This thesis examines and explores the effect of climate change on Goksun River basin under different future GCM emission scenarios. The statistical downscaling and the General Circulation Model (GCM) are widely utilized to estimate the climate change effects on the basin. Four different statistical downscaling models will be proposed to predict monthly areal mean precipitation as outputs using downloaded large-scale GCM factors as inputs. Model evaluation criteria will be displayed for the proposed model's performance to obtain the best-fit model. Consequently, the best-fit model will be simulated to estimate areal mean precipitation (P_M) for the future projection under future emission scenarios to investigate the climate change effect on the basin.

1.4 Layout of thesis

This thesis is formed of six chapters. The scope of the chapters are outlined below:

Chapter One: This chapter is a general introduction to the study.

Chapter Two: This chapter covers the literature review regarding the thesis.

Chapter Three: This chapter includes the methodology of the study.

Chapter Four: This chapter comprises information about the case study.

Chapter Five: This chapter includes results and discussion.

Chapter Six: This chapter contains conclusions and recommendations.

CHAPTER 2

LITERATURE REVIEW

2.1 Climate change

Climate change describes the weather conditions and the fluctuations of its properties. In recent years, Scientists are concerned with global warming, which is related to the impact of human activities on the climate system, although natural greenhouse effects are increasing. In human life, the importance of climate is associated with the positive or negative effects of climate on social and economic life and its consequences (Demir et al., 2013). To sustain people's lives in better condition and a healthier way, many institutions, and organizations, both national and international, central and local governments and non- governmental organizations, have made different efforts to identify climate change and impact of these changes.

Climate change forecasting provides the key input data to be made in planning for adaptation, mitigation, and prevention efforts in all sectors; meaning that the future planning of stakeholders must be based on climate and climate model outputs (Demircan et al., 2014).

In the context of climate change, various scenarios of global climate models for the future need to be reduced to provide a high-resolution data set for Turkey and its surrounding. To better understand the potential impacts of climate change, researchers need future socio-economic information's. Studying the implications of different socio-economic scenarios on land use at regional and local levels, climate and demography will identify key adaptation issues and priorities. Its purpose is to support politically by providing medium- and long-term planning for climate change adaptation. Scenario exploration has become an essential tool in support of regional policy discussions (Nakicenovic and Swart, 2000a,b; Carpenter et al., 2006). Climate models are the most important tools for studying the response of the climate system to various influences, for making climate forecasts on seasonal to decadal timescales, and predicting future climate in the coming century and beyond (IPCC, 2013). Otherwise, the climate system may be in a stable state when the amount of sunlight absorbed by the atmosphere is equal to that reflected by the surface of the earth. Most of them are infrared rays. When released into space, it is infrared radiation, which encounters many barriers, such as clouds and absorbing gases, and has resulted in the formation of a protective suppression providing a temperature appropriate to the continuity of life on Earth. This phenomenon is called the greenhouse effect and the absorbing gases called greenhouse gases (Kiehl and Trenberth, 1997).

2.2 Climate change impacts

Considering the strong effect of climate change on the environment, the economy and society is growing. Owing to the increasing gap between greenhouse gases, especially carbon dioxide, climate change has been found in continents, regions, and seas. These include climate change as the arctic region, the wind, and strong weather such as heavy rainfall, drought, and waves (IPCC, 2007).

On the report of Arnell, (1999) 62.5% of the world's population will be living in water-stressed countries by 2025. In addition, the rise in temperature caused by climate change has led to a general decline in precipitation rates, including in waters such as oceans and seas, as well as in large parts of the country where rainfall is decreasing despite the increase in global average rainfall due to climate change.

In Turkey, there have been investigated many studies on the impact of climate change. For example, Okkan and Inan, (2015) utilized multi-model ensemble, statistical methods, a number of factors and state-of-the-art speakers to study the GCM statistics at the monthly inputs in Kemer Dam, in the Aegean zone in Turkey under A2, B1, and A1B. GCM has been used to investigate the interaction of NCEP / NCAR selected in GCMs for 20C3M. The GCM is used to examine the relationship between the informations selected in the data analysis NCEP / NCAR and those in the GCM platform for 20C3M between 1979 and 1999. The change tangible is known by test statistics based on the assessment thoughtful. At long last, allowing their results to say that reducing the flows trends in the autumn, spring and winter seasons have been planned over the study area for the period between 2010 and 2039. Yano et al., (2007) described the impact of climate change on crop growth and irrigation due to the need for wheat-maize cutting sequences in the Mediterranean environment in Turkey. Climate scenarios were analyzed using data from three general circulation models-GCMs (CGCM2, ECHAM4, and MR) for the period of 1990 up to 2100 and for regional climate model 1990 - 2100. RCM) and data from 2070 until 2079. The scenario A2 was estimated in the Special Report on Emissions Scenarios (SRES) for potential climate change impacts based on GCMs. The results showed that both CGCM2 and RCM data predict an increase in wheat crop yield given the increase in CO2 concentration and temperature. Wheat is probably tolerant of climate change. Unlike the CGCM2 data predicted a slight increase in the future climate in corn production and RCM data.

Fujihara et al., (2008) looked systematically the potential influences of climate change on the hydrology and Seyhan River basin in Turkey. They put to a practice purpose a dynamic downscaling a systematic way, known as the global pseudo-warming method (PGWM), to link the results of GCMs and hydrological models of watersheds. The GCMs utilized in this study were MRI-CGCM2 and CCSR / NIES / FRCGC-MIROC under the SRES A2 scenario and the reduced-scale data covered two 10-year time periods matching to the present (1990) and the future (2070). Consequently, precipitation and temperature, which were dynamically reduced by a biased correction, were in good accord with the observed data. The results of the hydrological simulation were consistent with this observed dam discharge, flow, and reservoir volume. Hence, it has been deduced that PGWM is extremely useful for generating input data for hydrological simulations in combination with bias correction.

Okkan, (2015) examined a strategy to estimate the amount of monthly rainfall over Tahtali watershed in Turkey for climate change scenarios using the neural network. From this study, The author considered that the results of this study would contribute to authorities to develop new strategies for water resources in Turkey. The study is also considered as a source of inspiration for other climate research that can be conducted in this area.

2.3 Climate general circulation models (GCMs)

In the recent past, the topic of climate change has become an important issue of environmental scientists. Therefore, adequate information about climate change should be gathered to find the future and appropriate solution for its risk. Weather scientists cannot implement climate experiments within small laboratories. The Earths atmosphere is made up of several high complexity systems that are not applicable to apply in reactors.

For this reason, they proposed basis physical laws for a mathematical model for the climate which can be called as a General Circulation Model (GCM) or Global Climate Model. As stated by Dibike and Coulibaly (2005) can model numerically/physically operations of the ocean, atmosphere, and land cover system on a global scale. İt has been modeled to recreate present, past and future climate variables. İn addition, it distributes ocean and atmosphere in a horizontal mesh with a resolution of about 2.5-degree x 2.5 - degrees (longitude and latitude) alternative a vertical resolution of 10 – 20 layers of the atmosphere, and up to 30 layers of down deepness into the ocean (Carter et al., 2007).

Climate models were commenced by Phillips (1956) in the last century, who calculated GCM models to simulate an average, a synoptic scale (that is a spatial scale of 104 to 106 km2). And atmospheric circulation types for certain external forcing cases. İn this model, other models have been created which is more sophisticated and complex compared to the GCM model called AOGCM or Atmospheric – Ocean general circulation model (McAvaney et al., 2001).

Thus Wilby et al., (1999) made a comparison between three sets of current and future rainfall-runoff scenarios, implicated for climate change scenarios in the San Juan River basin, in Colorado. The scenarios were made by using: (1) statistically downscaled GCM output; (2) raw GCM output; and (3) raw GCM output corrected for elevational biases. Their results were discussed by the relative benefits of the three technologies for regional scenario refinement and climate change impact assessment. Wilby and Wigley, (1997), Dawson and Wilby, (2007) and Fowler et al., (2007) proposed the use of a strategy called Downscaling, which reduces the overall resolution of the GCM in fine resolution or produces the correct station information for a given region. Araji et al., (2018) predicted future climate change by collecting five and seven general circulation patterns (GCMs) in the Fourth and Fifth Assessment Reports of the IPCC. Emission scenarios B1, A1B, and A2 for AR4 and RCP2.6 and RCP8.5 for AR5 have been utilized. They finally discovered that soybeans could reach an optimal temperature threshold that could lead to increased yields in the 2030s.

Table 2.1 presents the types of GCM models which are commonly used to imitate global-scale and aerial circulation.

Table 2.1 Types of GCM

GCM pattern	Definition
CGCM1	Canadian Centre for Climate Modelling and Analysis Global Model 1
CGCM2	Canadian Centre for Climate Modelling and Analysis Global Model 2
CGCM3	Canadian Centre for Climate Modelling and Analysis Global Model 3
HadCM2	Hadley Centre Coupled Model 2
HadCM3	Hadley Centre Coupled Model 3
GFDLR15	Geophysical Fluid Dynamics Laboratory R15
GFDLR30	Geophysical Fluid Dynamics Laboratory R30
ECHAM4	European Centre/ Hamburg Model 4
NCARPCM	National Centre for Atmospheric Research

2.4 Climate emission scenarios

According to the Special Report on Emission Scenarios (SRES) (IPCC, 2000), climate change scenarios are potential future scenarios, not future predictions or forecast, but each scenario provides a way to see how the future can look. Scenarios provide important information about environmental quality and development, decisions and are certainly a useful tool for decision makers, details and experts.

The four scenarios describe demographic conditions, environmental conditions, social conditions, economic conditions, technologies, and future policies. The four scenarios are defined by IPCC (2007) as follows:

- A1 storyline and scenario family: "This scenario describes the future of the world with very rapid economic growth, a world population that peaks in the middle of the century and then declined, and the rapid introduction of new and more efficient technologies. Regional differences in per capita income. The three A1 groups are characterized by a technological focus: the balance between fossil energy sources (A1F1), non-fossil energy sources (A1T) or all energy sources (A1B). "

- A2 storyline and scenario family: They describe a very diverse world, with a growing global population and regional economic growth. This scenario is more disconnected and slower than in the other scenarios. It also represents a high emissions scenario and illustrates the worst case scenario (Andersen et al., 2006).

- B1 story and scenario family: It describes a world living with the people around the world as in the A1 only, but with a very quickly change in the aspect economic toward service and economic information. Reducing the intensity of the material and introducing clean and resource-efficient technologies. The focus is on global solutions to social, economic, and environmental sustainability, improving equity, but without additional climate first.

  This storyline mostly shows represents the best state of the SRES scenarios through the 21st century (IPCC, 2001).

- The B2 storyline and scenario family explains the world in the family. Regional, economic and environmental solutions stand out. A2 is less global population, modern economic development and B1 and A1 less technological change. The department also focuses on local and regional issues, and on environmental and social security.

Future emission scenarios play an important role in the extent to which climate change has an impact on water resources, both globally and regionally (Wilby et al., 2002). Certains climate change mitigation scenarios also point to a possible future in which global warming can be reduced through planned actions such as the use of green energies. For example, one of the climate change mitigation scenarios could be to select actions to achieve the goal of defining a long-term goal of the desired carbon dioxide concentration and to limit greenhouse gas emissions. at the international and national levels. They can also be compared with other climate change scenarios, such as climate change mitigation scenarios, A2 and B2 about the effects on reducing global warming. The concentration of carbon dioxide has reached about 375 ppm (Pacalal S. and R. Socolow, 2004). The limitations of carbon dioxide emissions bring many problems, such as cost, which would make it difficult to reach an agreement between governments and institutions. Therefore, the 550 ppm rule is one of the applicable rules that politicians and scientists can take into account to project temperature. The SRES has introduced four storylines, that are tagged as B1, B2, A1, and A2 as shown in Figure 2.1.

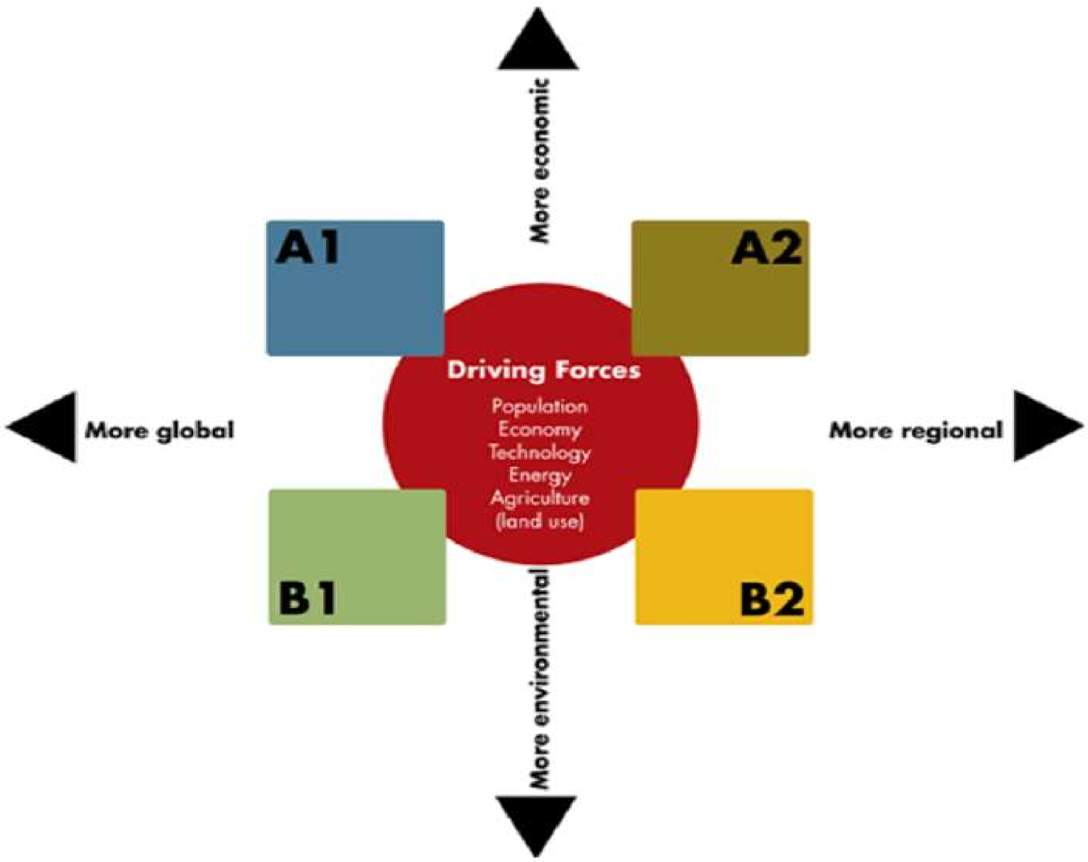

Figure 2.1 SRES storylines representation

2.5 Downscaling

The goal of downscaling is to create locally accurate climate information from global data. In other words, we are increasing data globally by placing it in the context of the observed local climatological conditions and improving spatial and temporal resolutions along the way. In the past few decades, many scale-modeling models have been improved, all having benefits and disadvantages. There are many research tools on how to work differently, most importantly, the Intergovernmental Panel on Climate Change (IPCC) Reporting Quotes (Solomon et al., 2007, Wigley 2004).

Also, many different methods will be used to collect the divisions that the GCMs can offer and that the wealthy society for the decision-making process. The derivation of the climatical information on the scales shows that the local climate is conditioned by interaction on atmospheric temperatures (circuits, temperatures, humidity, etc.) and local variable (mountain ranges, water bodies, etc.). They can model the interaction and stability of the local climate and the atmospheric conditions through the rhythmic process. It's a great way to handle the rhythm of information on the CGM producers, in terms of the most realistic information on the most scalable, the contrast and the domains of the scales of the scooters. A series of downscaling methods could be utilized to elaborate and perfect the CGM producers to find the highest quality adhesives.

Downscaling techniques can be divided into two categories: dynamical and statistical downscaling.

2.6 Dynamical downscaling

Climate models are not only important for understanding how the climate system works, but also for making decisions about issues that have a huge economic impact. Global Climate Models (GCMs) are made at a coarse resolution that cannot resolve atmospheric mesoscale characteristics and heterogeneity of the soil surface. As a result, GCMs may not accurately represent extremes of temperature and precipitation**.** The dynamic downscaling method is a technique used to liberate fine-scale data based on the use of (RCM) with the large GCM output used as a border condition (Nguyen, 2002; Prudhomme et al., 2002). It is usually base according to the regional climate models(RCMs), which produce more potential products based on the state of using GCM gaming facilities (e.g. Giorgi and Mearns, 1991 and 1999).

The choice of the individual will identify the differences between the RCM and the GCMs rule (Jones et al., 1997). Due to the decision at a high level, the RCM provides a good description more about these important events like the orographic product

(Christensen and Christensen, 2007). Furthermore, the most robust approach in the RCM is to create more modern circulation patterns (Buonomo et al., 2007).

On the other hand, RCMs have no plans to capture the edge of space flight observed at thin cell scales. In the literature, it was found that the development of many studies Rauscher et al. (2009) RCM skills depend on the resolution of RCM, on the region and the season. While RCMs can provide feedback to operating GCMs, many dynamic reduction approaches are based on a one-way nesting approach (Maraun et al.,2010).

The principal problem of RCMs is that considerable bias in large-scale mean rainfall simulation is inherited from GCM (Durman et al., 2001). Frei and Digg (2006) investigated that model differences are closed to model bias. Also, Christensen et al. (2008) cancel the bias by taking the difference between the GCM bias not linear and the control adopted in many studies and the future scenario.

2.7 Statistical downscaling

To transfer information from larger to smaller scales may be executed by statistics. Moreover, the relationships between local-scale predictands and large-scale predictors are developed and applied to the simulations of climate. The climate simulation can either be global or regional. According to the particular statistical relationships between the coarse GCMs and fine observed data, statistical downscaling is a truthful means of getting high-resolution climate scenarios (Wilby et al., 2004).

The statistical downscaling consist of two-step process such as i) the development of statistical relationships between local climate variable and large scale predictors and ii) the application of such relationships to the output of global climate model(GCM) experiments to simulate local climate characteristics in the future.

This method can be used in regional and local climate impact assessment and when appropriate data are available to derive statistical correlations.

Finally, a comparison between dynamical and statistical downscaling is displayed in Table 2.2 by their advantages and disadvantages.

Table 2.2 Advantages and disadvantages of statistical and dynamical downscaling (Wilby and Wigley, 1997).

	Advantages	Disadvantages
Dynamical downscaling	• Consistent downscaled variables • Physically based; • Data readily available; • Allows changes in probability density functions • Most of the climatic variables considered	• Spatial information poorly resolved • High uncertainties; • Computationally intensive
Statistical downscaling	• Generates information on high resolution; • Assess uncertainties; • Computationally cheap • Study of extremes can be given separate focus	• Assumes consistency of statistical relationship in future; • Only limited number of variables can be considered

2.8 Application of GEP on hydrology

GEP is an evolutionary algorithm designed to improve computer programs based on the research and analysis of similarities using natural selection and natural evolution. GEP seizes the best elements of the genetic algorithm (GA) and the genetic program (GP) but throws the limitations of established affiliates. GA and GP law enforcement know this, but GEP only uses growth (Steeb et al., 2005).

In the use of GEP, the individual expressions are encoded in linear chromosomes which in turn constitute expression trees, although in GA the procedure is the reverse, ie the individuals are parsed analysis trees, which can be somewhat cumbersome, and then be expressed as a linear string.

Many researchers have used GEP program, such the studies of (Alam and Danish, 2015) whom made an assess to the performance of the gene expression programming model, ANN model, and the Gumbel's method for flood frequency analysis for a typical Indian river gauging site. They found that the GEP model was the best among the various prediction models under consideration. Furthermore, the GEP model also provided a mathematical equation which makes it superior over those soft computing techniques. The results are very promising and show that GEP technology is a diverse tool and characterized by a more general approach to flood analysis. (Zhou et al., 2003) proposed a new approach to discovering classification rules using gene expression programming (GEP), a new linear programming genetic programming technique (GP). The history of the discovered rules may include many different combinations of attributes. They proposed a fitness function that takes into account both the consistency of the rules and their completeness. Their experimental results were carried out with several reference datasets and showed that the approach allows for an improvement in validation accuracy of up to 20% and a much smaller size of rule sets compared with C4.5 algorithms. The proposed GEP approach was more efficient and tended to lead to shorter solutions compared to canonical tree-based GP classifiers.

(Guven and Talu, 2010) investigated the analysis of the suspended sediment yield in Euphrates River. The proposed models were developed to extrapolate the flow data from five stations in the Central Euphrates basin. A detailed sensitivity analysis was performed to select time-lagged flow and sediment yield variables. They concluded that the GEP model is clear and simple: it can be used by anyone who does not focus on GEP technology. The GEP proposal provided a means to organize the right product Incineration and production and to promote the use of GEP in other technical research areas.

Guven and Azamathulla, (2012) studied on modelization techniques in the prediction and prediction of scour depth downstream of flip bucket spillways. The research was based on dimensional and normalized scour parameters, two GEP models were simulated and compared with earlier predictions of other GP. The results of their studies led to conclude that the study is truly promising and supported the use of GEP technology in the design of spillway and plunge pool structures.

Seckin and Guven, (2012) investigated the forecast peak flood discharges for 543 ungauged sites across Turkey using linear genetic programming (LGP), GEP, and logistic regression (LR). Four input variables were selected such as elevation, drainage area, longitude, latitude, and return period as independent variables, and the peak flood discharge as the target variable. The results indicated that GEP and LGP models were better than the LR model because of the correlation coefficient between the predicted and observed peak flow.

Traore and Guven, (2013) used GEP to modeled reference evapotranspiration in West Africa in Burkina Faso in simulating weather data from the tropical seasonal dry zones.

Azamathulla et al., (2011) used GEP for developing a stage-discharge link for the Pahang River. GEP model seems to be a helpful tool could be applied for determining the daily discharge data in flood situations.

Guven and Aytek, (2009) investigated two stations of the Schuylkill River in the US. GEP was compared with more traditional methods, multiple linear regression (MLR) and stage rating curve (SRV) methods. the results were shown that GEP was more accurate than the two others.

Zakaria et al., (2010) used GEP to analyze the transportation total load in three rivers which are Kurau, Langat, and Muda. Fernando, et al. (2012) studied the combination of four rainfall-runoff model outputs with the GEP model to predict the stream flow. four models were selected, linearly varying gain factor model, linear perturbation model, soil moisture and routing model. Moreover, the data from four different watersheds with different hydrological, geographical and climate conditions were utilized. The results proved the efficiency of GEP for multimodal combination. Furthermore, GEP was broadly utilized in different topics in planning water resources engineering (Zorn and Shameldin 2015; Wang et al. 2013; Kisi et al., 2012; Guven, 2009).

2.9 Application of GMDH on hydrology

Gholami et al., (2017) predicted three geometric variables of stable channels by using the Group Method of Data Handling (GMDH) models. the results exhibited good accuracy during the analyze.

Shaghaghi et al., (2017) developed a genetic algorithm (GA) to boost the multi-objective Pareto optimal design of the GMDH neural network outcomes. Ebtehaj et al., (2015) investigated the purpose of estimating the discharge coefficient in a side weir by using the GMDH. The results showed that the model was more precise in estimating than the feed-forward neural network model and existing nonlinear regression equations.

Najafzadeh and Barani, (2011) presented a novel application of GMDH in the estimation of scour depth around a vertical pier. According to the result, the proposed method generated a better estimation of scour depth than GMDH-BP in calibration and validation stages.

Garg, (2013) made an attempt to develop the sediment yield estimation using GMDH. He found that the inductive GMDH-NN could adequately seize the sediment yield. Moosavi et al., (2017) estimated daily runoff quantities by using the GMDH. Results showed that the GMDH model had a moderate performance.

Najafzadeh et al., (2013) utilized the GMDH network to estimate abutments scour depth for both clear-water and live-bed conditions. Their result utilized that the GMDH network estimated with a lower error and more accurate.

Najafzadeh and Zahiri, (2014) predicted the flow discharge in straight compound channels by using neuro-fuzzy-based group method of data handling (NF-GMDH). The evaluation of the proposed model provided a more certain estimation of flow discharge.

Najafzadeh and Azamathulla, (2013) predicted the scour depth around bridge piers by using the GMDH with quadratic polynomial. the results showed that with GMDH an accurate prediction could be obtained than those using conventional equations. Parsaie et al., (2017) utilized GMDH to estimate the discharge coefficient of cylindrical weir-gate. The results showed that all developed models have an appropriate performance.

2.10 Application of SVM on hydrology

Sivapragasamet al., (2001) proposed a simple and influential estimation model based on singular spectrum analysis (SSA) coupled with support vector machine (SVM) for rainfall and runoff forecasting. The results were compared to the non-linear prediction(NLP) method. They found that the proposed technique generated no doubt higher accurate result in the prediction than of NLP.

Raghavendra and Deka, (2014) investigated many SVM-based models related to the field of hydrology. The SVM has also been widely used during the last decade for rainfall-runoff modeling. They concluded that the regression models were ineffective compared to SVM models in many cases as analyzed in the literature.

Dibike et al., (2001) reviewed the statistical learning theory and SVM, and the potential of the SVM for feature classification and multiple regression problems was proved by applying the method to two different cases of model induction from empirical data. It was shown that SVMs generalize the more precise estimation of runoff on test data.

Bray and Han, (2004) researched the use of SVM models, in flood forecasting, focused on the identification of a suitable model structure and its relevant parameters for rainfall-runoff modeling. Lin et al., (2010) studied the SVM model to estimate long-term flow discharges in Manwan, based on the historical record. their investigation informed that the SVM model could generate a good prediction.

Shabri, (2012) used least-squares support vector machine (LS SVM) model to ameliorate the accuracy of streamflow forecasting. they found that the LSSVM model was a useful device to predict the flow. Zakaria, (2012) studied the ability of support vector machines (SVM) model for streamflow forecasting at ungauged sites. Behzad et al., (2010) compared the approach of SVMs to ANNs for estimating transient groundwater levels. It was concluded that SVM has a beneficial and convenient tool where less measured data are existing for future estimation.

Liu et al, ,(2009) investigated the use of SVM on an estimation of groundwater level on the basis of groundwater physical process study. Their study showed that the model has good learning ability and has fast training and testing period.

Kumar and Ojha, (2011) studied the performance of support vector machine to predict the scour downstream of a ski-jump bucket. Cimen, (2010) worked to predict the concentration/load of suspended sediments in rivers. Their results indicated that this approach may perform better than those described in the literature, based on different methodologies. Safavi and Esmikhani, (2013) presented conjunctive use of surface water and groundwater at a basin-wide scale by using an SVM and Genetic Algorithms.

Tripathi et al., (2006) applied an SVM technique for statistical downscaling of precipitation at a monthly time scale. The results were indicated that SVMs are convenient for conducting climate impact researches.

Chen et al., (2010) investigated a statistical downscaling technique based on the Smooth Support Vector Machine (SVM) technique to estimate daily precipitation of the changing climate in Hanjiang Basin. The result showed that SSVM was suitable for conducting climate impact studies on the point of a statistical downscaling model in this region.

Aggarwal et al., (2012) employed SVM and ANN technique to the forecasting of stage and discharge. their results showed that SVM model was more accurate over a longer time frame than by the other two models. Ghorbani et al. (2016) studied the estimation of river flow time series on the applicability of multilayer perceptron (MLP), radial basis function (RBF) and SVM models.

Kisi, (2015) investigated the capacity of least squares support vector regression (LSSVR) and adaptive neuro-fuzzy embedded fuzzy c-means clustering (ANFIS-FCM) in forecasting and estimation of monthly streamflows. Based on the results, the LSSVR was explored to be better than the ANFIS-FCM and successfully used.

Seyam et al., (2017) explored the capacity of SVM to estimate hourly stream flow (SF) in the downstream area from the upstream water level and rainfall records in a humid tropical area. The results showed that SVM did well in real-time SF prediction. Guo et al., (2011) investigated the forecasting monthly streamflow values by using an SVM and concluded that SVM generated more accurate results than ANNs.

Wang et al., (2009) employed Genetic Programming (GP), ARMA, ANN, Adaptive Neural-Based Fuzzy Inference System (ANFIS), and SVM. Based on the results it was shown that SVM has the highest coefficient of determination of Manwan Hydropower. Samsudin et al., (2011) indicated how the monthly river flow could be displayed by a hybrid model combining the GMDH and LSSVM models.

CHAPTER 3

METHODOLOGY

3.1 General

In this chapter, the methodology employed to downscale monthly precipitation from CGCM3 data with the A2 and A1B future scenarios by using downscaling models.

Firstly, downscaling is explained and downscaling techniques are derived. Statistical downscaling has selected as a downscaling technique because it is easy to use and it needs low computation requirements. Secondly, large-scale weather factor (predictors) is defined and obtained from GCM and future emission scenarios. The predictors can be downloaded directly from CGCM box grids. Thirdly, statistical downscaling models Support Vector Machines (SVMs), Group Method of Data Handling (GMDH), and Gene Expression Programming (GEP) have been investigated to form a strategy for calibration (1971-1990) and validation (1991-2000) process. Finally, models evaluation criteria are specified to assess these models such as graphical comparison and statistical calculations between the predicted and observed monthly P_M. This comparison will be used to select the best model performance for forecasting future monthly mean precipitation to evaluate climate change effects on the basin. Figure 3.1 shows the flow diagram of the methodology adopted for this thesis for the downscaling process.

3.2 Downscaling

Downscaling can be described as a tool for making a relationship between the variables representing large-scale weather variables (predictors) and local-scale variables (predictands). Basically, there are two major forms of downscaling techniques using in scientific studies: dynamic downscaling and statistical downscaling as we have mentioned in Chapter 2. The statistical downscaling technique is cheaper, faster, and can present an immediate solution with significant lower computing requirements.

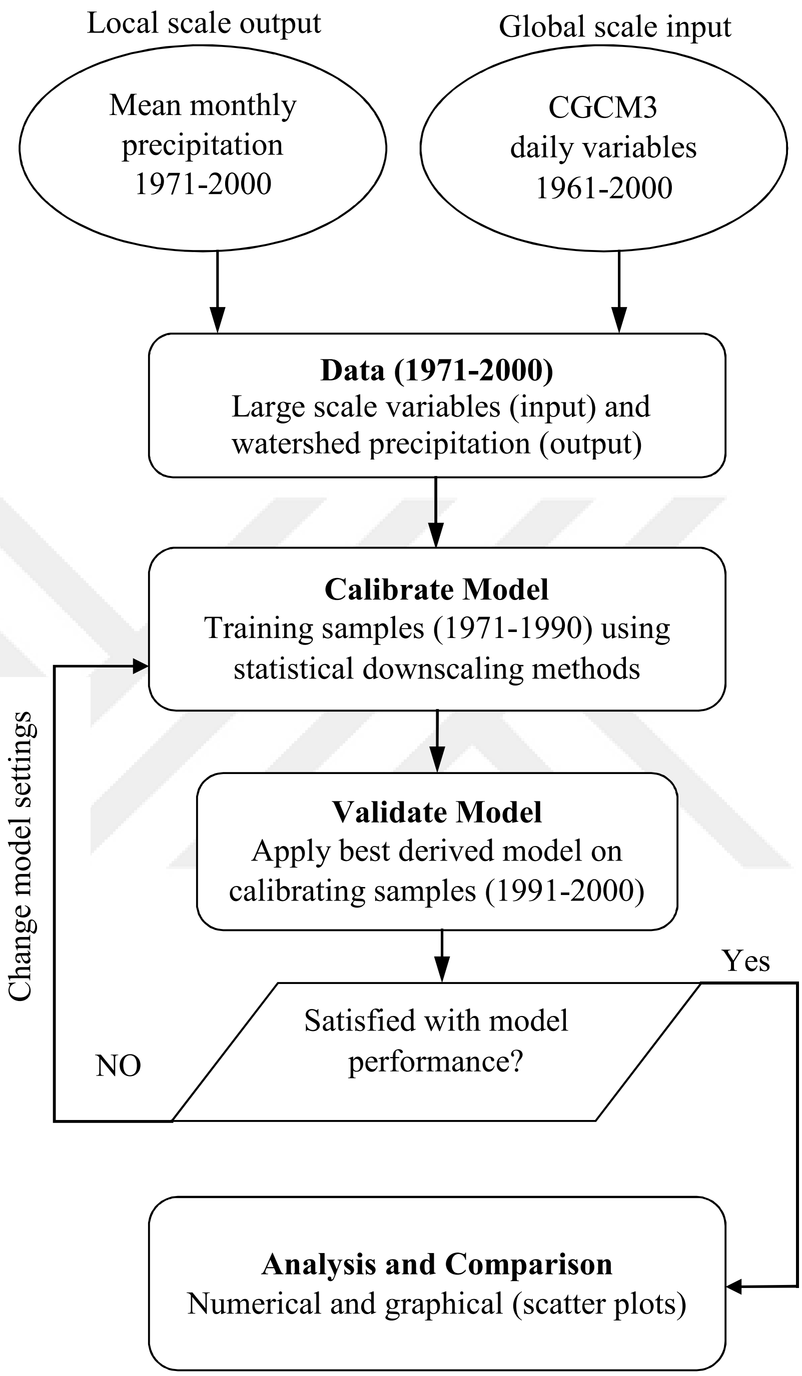

Figure 3.1 Flow diagram of downscaling process for this study

For these reasons, we used statistical downscaling technique in this thesis. Statistical downscaling builds up a quantitative link between the predictors and predictands. Predictors are the input data of statistical downscaling techniques while the predictand is the output data, such as precipitation and temperature. The input-output data relationship can be given as an equation form where f is a function which relates input and output:

$$\text{Output} = f\,(\text{input}) \qquad (3.1)$$

Figure 3.2 displays the schematic representation of the downscaling process.

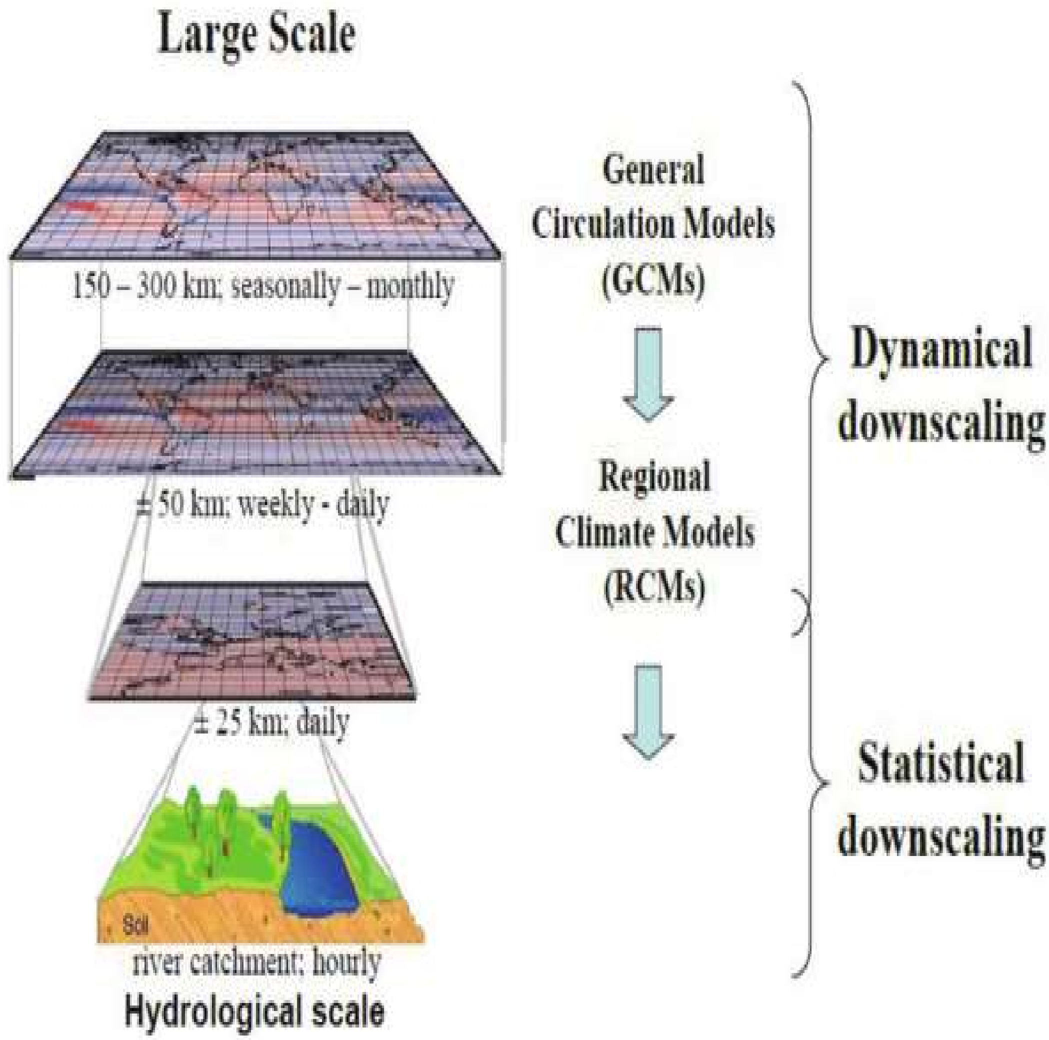

Figure 3.2 A schematic representation of the downscaling process (Adapted from Willems,2011)

Statistical downscaling can be shared out into almost four categories: stochastic air generators, limited area modeling, stochastic air generators, and regression methods (Wilby et al., 1997). A regression-based statistical downscaling technique establishes a linear or non-linear function between input and output. A regression method is preferred among these approaches because it is easy to implement and it has low computation requirements.

Predictor variables can be obtained from General Circulation Models. There are different types of GCM and future emission scenarios that have been progressed recently like UK Hadley Centre Coupled Model Version 2 (HadCM2), (HadCM3), and National Center for Atmospheric Research (NCAR). Each model can be used for the prediction process. In this thesis, Coupled Global Climate Model (CGCM3) with A2 and A1B future scenarios have been preferred. CGCM3 variables (input) can be taken from the Canadian Centre for Climate Modeling and Analysis (CCCma). Although CGCM3 utilizes the same ocean item as CGCM2, CGCM3 performs use of the essentially updated atmospheric component ACGCM3 (Atmospheric Canadian GCM, version 3) (Flato and Boer, 2001)

3.3 Support Vector Machine (SVM)

The support vector machine (SVM) was introduced by Vapnik in 1995, which has gained popularity in data-driven study fields because of its many alluring features and promising empirical performance. Support vector machines (SVMs) essentially are classification and regression methods that have been obtained from statistical learning theories. The regression version of the SVM Framework follows the principle of Structural Risk Reduction (SRM), which is proven to be superior to the traditional empirical risk reduction (ERM) principle used by traditional modeling techniques. The traditional ERM concentrates on minimizing training data error, whereas SRM minimizes the upper limit of expected risk, hence providing SVM with the ability to generalize and this is the essential aim of statistical learning (Vapnik 1999). There are four fundamental advantages of SVM. Firstly, it has a regularization parameter, which makes the user thinking about escaping over-fitting. Secondly, an SVM is decided by a convex optimization problem for which there are accurate methods (e.g., SMO). Thirdly, it is estimated to bound on the test error rate, and there is a significant body of theory behind it which suggests it should be a good idea. Final and main advantages of SVM is that it utilizes the kernel trick (function), to construct expert knowledge about the investigated phenomenon so that the model complexity together with estimation error is concurrently minimized. Support vector regression (SVR) is summarily outlined below.

Consider data examples with N number of samples (X, Y) be a given training set, where $X = \{x_1, x_2, \ldots., x_n\}$ is the input vector and $Y = \{y_1, y_2, \ldots., y_n\}$ is the its corresponding output value. The objective is to find the SVM estimator (f) function on regression (SVR) can be expressed as:

$$f(x) = \boldsymbol{w} \cdot x + b \tag{3.2}$$

Equation 3.2 is the best describes the observed output y with an error tolerance, where **w** is the weight vector and b is the bias term, and **x** is a non-linear function. The term **w**.x is the dot product of **w** and x. This function can later be used to find y_i depend on some new inputs (x_i).

In SVR, a penalized loss function,$L_\varepsilon(y, f(x))$, is used when data are outside of the tube of error tolerance. There are different types of loss functions in the literature. The loss function is used here, defined by the Vapnik's ε -insentive loss function.

$$L_\varepsilon(\boldsymbol{y}, f(\boldsymbol{x})) = \begin{cases} 0, & |\boldsymbol{y} - f(\boldsymbol{x})| \leq \varepsilon \\ |\boldsymbol{y} - f(\boldsymbol{x})| - \varepsilon, & |\boldsymbol{y} - f(\boldsymbol{x})| > \varepsilon \end{cases} \tag{3.3}$$

This SVR problem can be written in terms of the following convex quadratic optimization problem:

$$Minimize \qquad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{3.4}$$

$$\begin{aligned} Subject\ to \qquad & y_i - (\boldsymbol{w} \cdot x_i + b\,) \leq \varepsilon + \xi_i \\ & (\boldsymbol{w} \cdot x_i + b\,) - y_i \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, i = 1,2,..,n \end{aligned}$$

Where ξ_i and ξ_i^* represent slack variables stating the upper and lower training errors subject to an error tolerance ε. The slack variables are zero for points inside the ε-insensitive zone and increase progressively for points outside the zone. C denotes a positive constant that detects the degree of penalized loss when a training error occurs. Figure 3.3 shows the concept of SVR stand on equation 3.2 and 3.3.

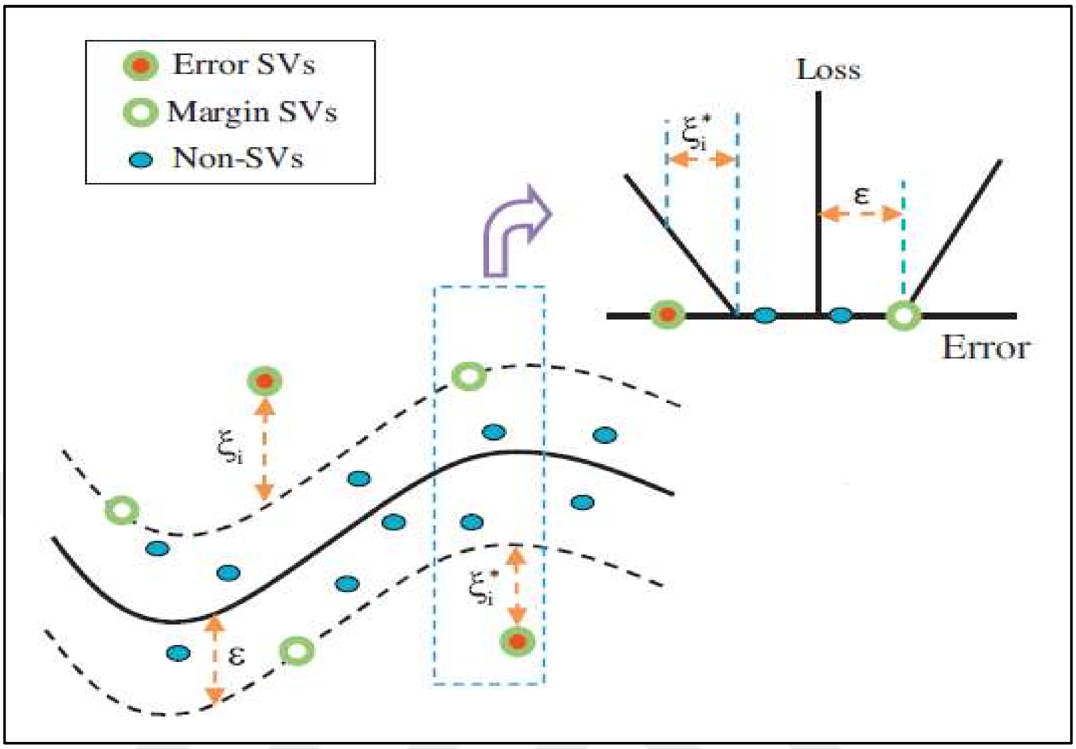

Figure 3.3 Nonlinear SVR with Vapnik's ε-insensitive loss function.

It seems that the optimization problem can easily be solved in the dual form. The Lagrange function is established from the primal objective function and the corresponding constraints, by suggesting a dual set of variables, α_i and $\bar{\alpha}_i$. It can be stated that this function has a saddle point with respect to the primal and dual variables of the optimal solution.

After getting the Lagrange multipliers, α_i and $\bar{\alpha}_i$, the Karush-Kuhn-Tucker (KKT) method is used to get the parameters **w** and b. Prediction in a linear regression function can be formulated as:

$$f(x) = \sum_{i=1}^{n} \alpha_i - \bar{\alpha}_i \langle x_i . x_j \rangle + b \tag{3.5}$$

In order to utilize SVR in nonlinear cases, the decision space variable (X) can be transferred to a higher-dimensional space, using function ϕ. Applying this transformation, the dual problem can be expressed as:

$$Maximize \qquad -\frac{1}{2}\sum_{i,j=1}^{n} \alpha_i - \bar{\alpha}_i(\alpha_i - \bar{\alpha}_i)\langle\phi(x_i).\phi(x_j)\rangle \tag{3.6}$$

$$-\varepsilon\sum_{i=1}^{n}(\alpha_i - \bar{\alpha}_i) + \sum_{i=1}^{n} y_i(\alpha_i - \bar{\alpha}_i)$$

$$\text{Subject to } \sum_{i=1}^{n}(\alpha_i - \bar{\alpha}_i) = 0$$

$$0 \le \alpha_i \le C, i = 1,2,..,n$$

$$0 \le \bar{\alpha}_i \le C, i = 1,2,..,n$$

Where $k(x_i, x_j) = \langle\phi(x_i).\phi(x_j)\rangle$ is the kernel function. SVR utilizes the kernel function to output the inner products in feature space, circumventing the problems intrinsic in evaluating the feature space. Input space needs a kernel function to display the computation. Functions that satisfy the Mercer's condition can be demonstared to correspond to dot products in a feature space. Thus, any functions that corresponds to Mercer's theorem rules can be used as a kernel function. (Vapnik, 1999) The most common kernels in SVM are:

1. Linear kernel

$$k(x_i, x_j) = x_i \cdot x_j \tag{3.7}$$

2. Polynomial kernel

$$k(x_i, x_j) = \lfloor\gamma(x_i \cdot x_j) + c\rfloor^d \tag{3.8}$$

3. Sigmoid kernel

$$k(x_i, x_j) = tanh\lfloor\gamma(x_i \cdot x_j) + c\rfloor \tag{3.9}$$

4. Radial basis function kernel

$$k(x_i, x_j) = exp\left(-\gamma|x_i - x_j|^2\right) \tag{3.10}$$

where γ, c and d are the kernel parameters. These parameters ought to be attentively selected as they implicitly describe the structure of the high dimensional feature space $\phi(x)$ and would be in charge of the complexity of the final solution. RBF can effectively be used when the link between and predictand predictors is nonlinear Tripathi et al., (2006). Furthermore, the RBF needs less computation while the polynomial kernel has more parameters Chen et al.,(2010). Therefore, the radial basis function is utilized as a kernel function in this study, given by γ in Equation 3.10.

Finally, the kernel function lets the decision function of nonlinear SVR to be presented as:

$$f(x_i) = \sum_{i=1}^{n} (\alpha_k + \bar{\alpha}_k) k(x_i, x_k) + b \qquad (3.11)$$

Where x_k represents the support vector, and n donates the number of support vectors. The SVR has three parameters C, ε and γ. These parameters that need to be trained. These parameters, the Lagrange coefficients $\bar{\alpha}_k$ and the bias term b can be solved analytically in Equation 3.11, and thus the optimal network architecture is obtained.

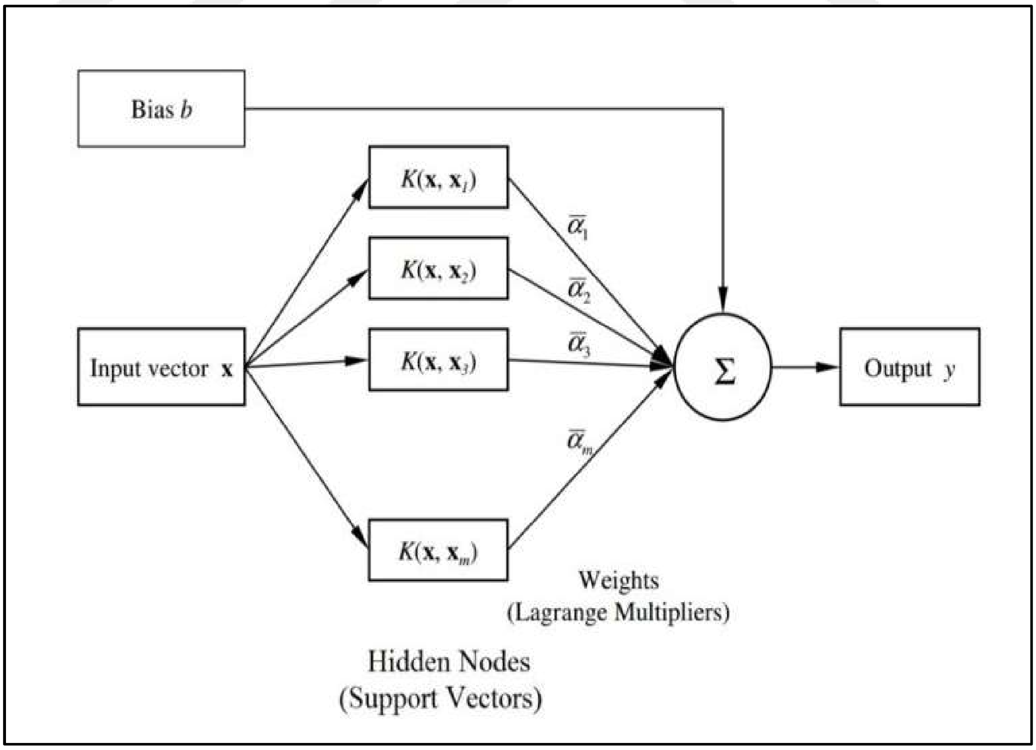

Figure 3.4 Network architecture of SVM

The SVR that employes the radial basis function as a kernel is same as the radial basis function neural network shown in Figure 3.4.

The hyper-parameters such as the SVR constant ε, regularization constant C and γ affect the success of the SVR. These parameters are mutually dependent when one parameter changes, it will affect other parameters. The parameter C regulates the smoothness or flatness of the approximation function. A greater C value points that the objective is only to minimize the empirical risk, that makes the learning machine more complex. A smaller C value may increase the number of training errors, yielding a learning machine with poor approximation. If the data are noisy, then a small value of C, which punish the errors less, can be preferred.

The parameter ε controls the width of the ε -insensitive zone and affects the number of support vectors. Hence, its value influences both complexity and generalization capability of the approximation function. Low values of ε caused to a larger number of support vectors and raise the complexity, whereas high values of ε caused to a smaller number of support vectors and yield in more flat estimates of the regression function. The performance of SVR is depend on these parameters, so it is crucial to find appropriate values for C and ε (Kim, 2003). Determining the adequate value of these parameters is frequently a heuristic trial-and-error process (Raghavendra and Deka, 2014). An interested reader who is looking for more information about the discussion of SVM can consult Chen et al., (2006) and Goyal et al., (2017) articles.

3.4 Group Method of Data Handling (GMDH)

The algorithm of Group Method of Data Handling (GMDH) was introduced in 1966 by Ivakhnenko as an inductive learning algorithm for modeling and describing complex systems containing a data set with multiple inputs and one output with the nonlinear relationship between simple mathematical relationship parameters. GMDH is the advancement of self-regulating inductive computer technology to solve more complex practical problems. The basic principle of GMDH is to maintain a function using a learning algorithm based on the quadratic node transfer function (Farlow, 1984). A number of different neuron pairs that interact with second-order polynomials to form the next layer of neurons help to map the input and output parameters. This method has been specially developed to solve high-level regression polynomials to solve modeling and classification problems. The general link between the input and output variables can be expressed by complex polynomial sequences in the form of the Volterra series known as the Kolmogorov-Gabor polynomial (Ivakhnenko, 1971):

$$y = f(x_i) = a_0 + \sum_{i=1}^{K} a_i\, x_i + \sum_{i=1}^{K}\sum_{j=1}^{K} a_{ij}\, x_i x_j + \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{k=1}^{K} a_{ijk} x_i x_j x_k + \ldots.. \quad (3.12)$$

where x is the input, K is the number of inputs and a_i are coefficients. But, many of the applications of the quadratic form are called partial descriptions (PD) where only two of the variables are used in the following form:

$$y = f(x_i) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 {x_i}^2 + a_5 {x_j}^2 \quad (3.13)$$

to estimate the output. To determine the value of the coefficients a_i for each k models, a system of Gauss normal equations is solved. The coefficient a_i of nodes in each layer are presented:

$$\boldsymbol{A} = (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T Y} \quad (3.14)$$

Where $\boldsymbol{Y} = [y_1\ y_2 \ldots.\ y_K]^T$ $\quad \boldsymbol{A} = a_0,\ a_1,\ a_2,\ a_3,\ a_4,\ a_5$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{1p} & x_{1q} & x_{1p}x_{1q} & {x_{1p}}^2 & {x_{1q}}^2 \\ 1 & x_{2p} & x_{2q} & x_{2p}x_{2q} & {x_{2p}}^2 & {x_{2q}}^2 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & x_{Kp} & x_{Kq} & x_{Kp}x_{Kq} & {x_{Kp}}^2 & {x_{Kq}}^2 \end{bmatrix}$$

and K is the number of observations in the training set.

The purpose of GMDH is to obtain the coefficient a_i in Eq. (5) using regression in order to minimize the difference between the actual (y) and the predicted output ($\hat{y}$) for each pair ($x_i x_j$) of input variables (Ivakhnenko, 1971). Accordingly, to optimize the coefficients of each quadratic function, the principle of least-squares error is displayed as:

$$E = \frac{\sum_{i=1}^{K} (\hat{y}_i - y_i)^2}{K} \rightarrow Min \quad (3.15)$$

The primary function of GMDH is to transmit the signal along with the nodes of the network as in conventional neural networks. Each layer comprises of simple nodes, each of which generates its own polynomial transfer function and then transfers its output to the nodes in the next layer. The network structure of GMDH structure is shown in Figure 3.5 and 3.6. An interested reader who is looking for more information about the discussion of GMDH can consult Samsudin et al., (2011).

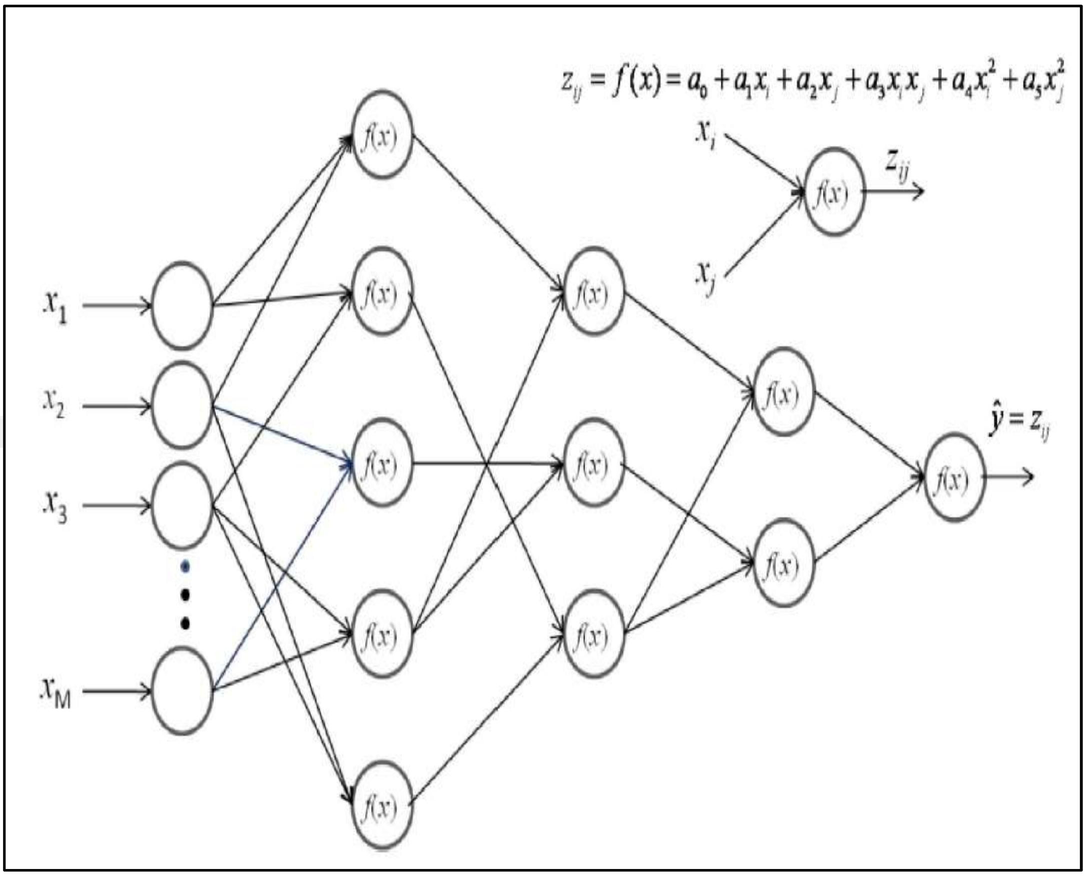

Figure 3.5 Network architecture of GMDH

GMDH automatically gains the links that dominate the system variables when the training process is running. That is to say, the optimal network structure is automatically selected to minimize the difference between the network output and the desired output. GMDH is therefore generalizable and can be adapted to the complexity of non-linear systems with a relatively simple and numerically stable network (Assaleh et al., 2013).

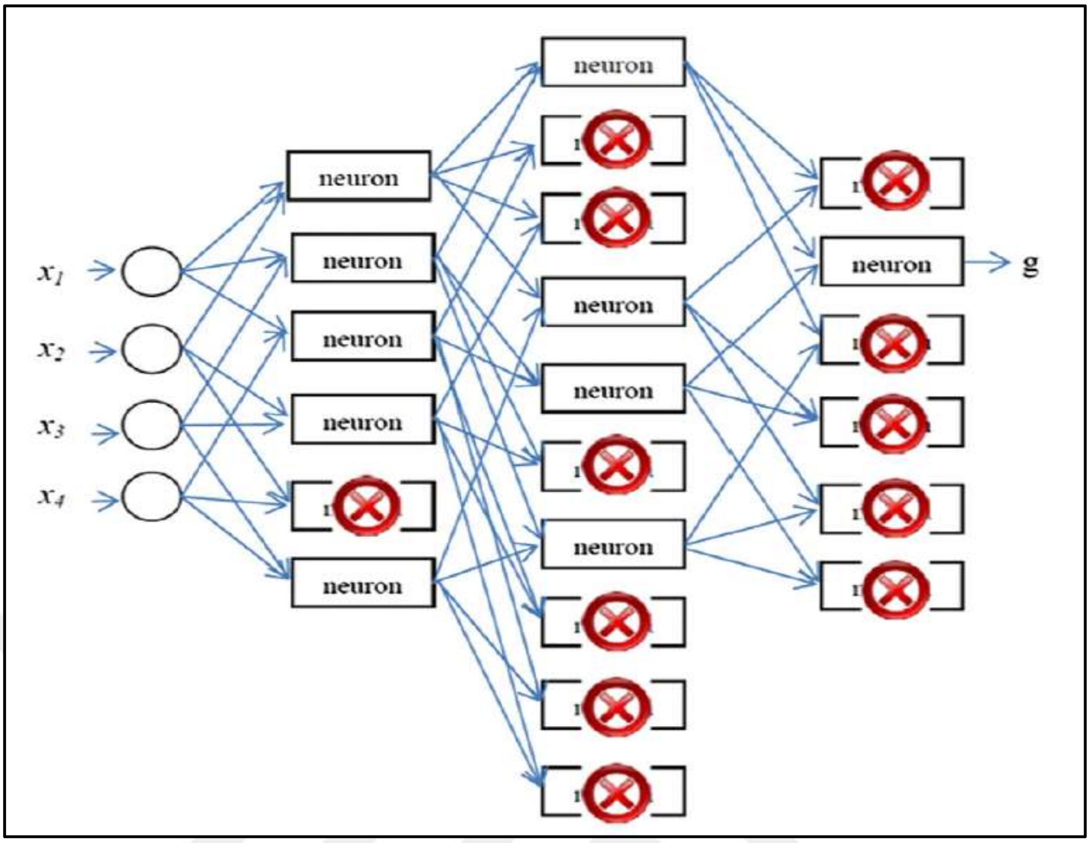

Figure 3.6 Network architecture of GMDH

The main step is to determine the set of network parameters in order to obtain a solution from GMDH model. In this thesis, the maximum network layer is set to fifteen (15), the maximum polynomial order is twelve (12), and the convergence tolerance factor 0,001 is used for GMDH model. Network layer connection is selected as the previous layer and original input variables. The number of neurons per layer is fixed as fifteen (15) neurons. The final main step is to decide suitable functions that will be used in the GMDH network. We got the best results by using Linear 1, 2, and 3 variables and Quadratic 1, and 2 variables.

3.5 Gene Expression Programming (GEP)

Gene expression programming (GEP), which is a new variant of Genetic Programming (Koza, 1992), was discovered by Ferreira (2001). GEP has capable of evolving advanced computer programs. This feature of GEP reflects the difference between GEP and GP. GEP imitates biological evolution to generate a function that places a set of data. GEP is a genotype/phenotype system that develops computer programs of variable size and shape encoded in fixed length linear chromosomes (genetically). GEP is an evolutionary algorithm, uses symbolic regression to fit the data to get an optimum form of mathematical function. The application of GEP is widely used in hydrologic engineering particularly in the last decade, especially it has been applied in the prediction of hydro-meteorological variables. Figure 3.7 represents the flow diagram of gene expression logarithm. The process is repeated until the best solution has been found for a certain number of tests.

The GEP has determined automatically the statements for the study solution by coding the expression as a tree formation with nodes and terminals. The link between symbols of function and the chromosome is stood for in the tress in one by one relation. Combination of genes or one gene create the chromosome and each gene produces a smaller sun-program. The linking function is utilized to participate in the genes when the chromosome has more than one gene. An example of a GEP chromosome with two genes and one linking function is shown in Figure 3.8 as a symbolic representation.

From Figure 3.8, ‘a’, ‘b’, ‘c’, and ‘d’ show the input variables and ‘-, / ’describe addition and division function. The mathematical expression of gene 1 is ‘–ab’ term, The mathematical expression of gene 2 is ‘/cd’ term. ‘+’ is the linking function which links the gene 1 and gene 2. The final mathematical expression with the linking function is given in equation 3.16.

$$(a-b)+\left({}^{c}/_{d}\right) \tag{3.16}$$

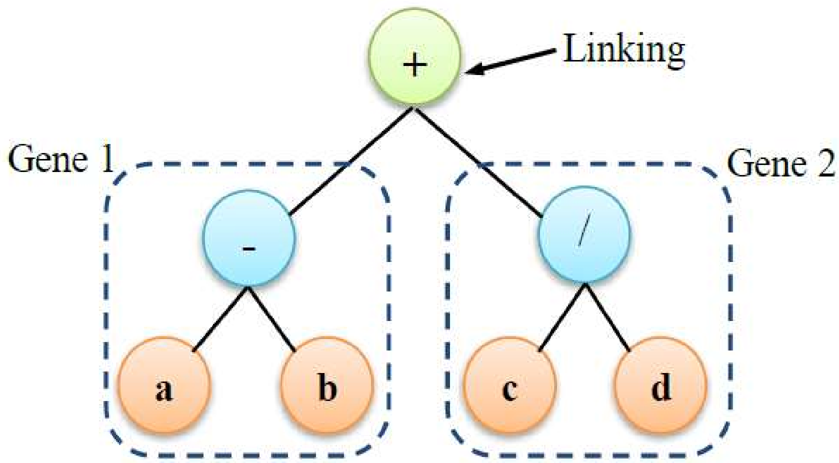

Figure 3.8 Expression tree of GEP chromosome with the linking function

The genetic character link genes have of two parts, namely tail and head. The tail contains terminal symbols while the head contains functions and constants in order to create mathematical expression.

There are four crucial steps to get the solution to the problem in GEP. In this thesis, the first critical step is to determine a set of function. A set of functions (+, -, *, /, power, Ln) were selected from the set of functions in the GEP program. The second critical step for GEP program is to determine the chromosome structure which includes the number of genes per chromosome and the size of the gene. We got the best results by using four genes (4 genes) per chromosome for each GEP model. The third critical step is to select the linking function. In this study, the linking function was selected as '+' (addition). The final critical step is the fitness measure. In this thesis, the root relative squared error (RRSE) of the training set is employed as a fitness function.

3.6 Model evaluation Criteria

As mentioned in previous methodology parts, three downscaling models have been introduced to predict monthly average mean precipitation of the basin for calibration (1971-1990) and validation (1991-2000) periods. Also, linear regression is done for comparison. To understand each model performance and to determine the best model performance, model evaluation criteria is needed for both proposed periods.

Model evaluation criteria are composed of two sections: graphical and quantitative evaluation. These sections are described comprehensively in below. Eventually, according to graphical and quantitative measurements, only the best model performance will be selected for P_M prediction under future period by using future emission scenarios.

3.6.1 Graphical evaluation

There are two sections for the graphical evaluation which are the scatter plot and monthly mean precipitation of the basin graph.

3.6.1.1 Scatter plot

A scatter plot is a sequence of points displayed the values of two independent variables as dots. A scatter plot demonstrates the relation between predicted data positioning in

X-axis and observed data positioning in Y-axis (Piñeiro et al., 2008). It is substantial in the evaluation of models, as it can indicate the best-fit line and correlation equation to examine the correlation between observed and predicted data.

3.6.1.2 Monthly mean precipitation graph

Monthly mean precipitation graphs have been figured out during calibration and validation periods with observed and predicted PM for each month. Also, for future prediction, a monthly mean precipitation line is used for the baseline period to observe climate change effects on the basin.

3.6.2 Quantitative evaluation

In order to forecast the precision of statistical downscaling models, some quantitative evaluations are expressed below for this study. Consider O_i is the observed value and P_i is the predicted value. M is the total number of data and i is the subscript.

- Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(O_i - P_i)^2} \tag{3.17}$$

Which is the square root of mean square error between predicted and observed data.

- Mean Absolute Error (MAE)

$$MAE = \frac{1}{M}\sum_{i=1}^{M}|O_i - P_i| \tag{3.18}$$

Which is mean overall values of absolute error between predicted and observed data.

- Correlation Coefficient (R)

$$R = \frac{\sum_{i=1}^{M}(O_i - \overline{O_i})(P_i - \overline{P_i})}{\sqrt{\sum_{i=1}^{M}(O_i - \overline{O_i})^2 \sum_{i=1}^{M}(P_i - \overline{P_i})^2}} \tag{3.19}$$

Where $\overline{O_i}$ is the mean of observed value and P_i is the mean of the predicted value.

3.6.3 Akaike information criterion (AIC)

Akaike Information Criterion (AIC) introduced by Akaike (1974) is employed to evaluate the generalization performance of the proposed models.

$$AIC = N \times Ln(MSE) + 2k \tag{3.20}$$

Where N is the number of data, MSE is the mean square error, and k is the number of fitting parameters. AIC is used to evaluate the exchange between calibration performance and network size. The aim is to get smaller AIC to acquire a network with the best generalization.

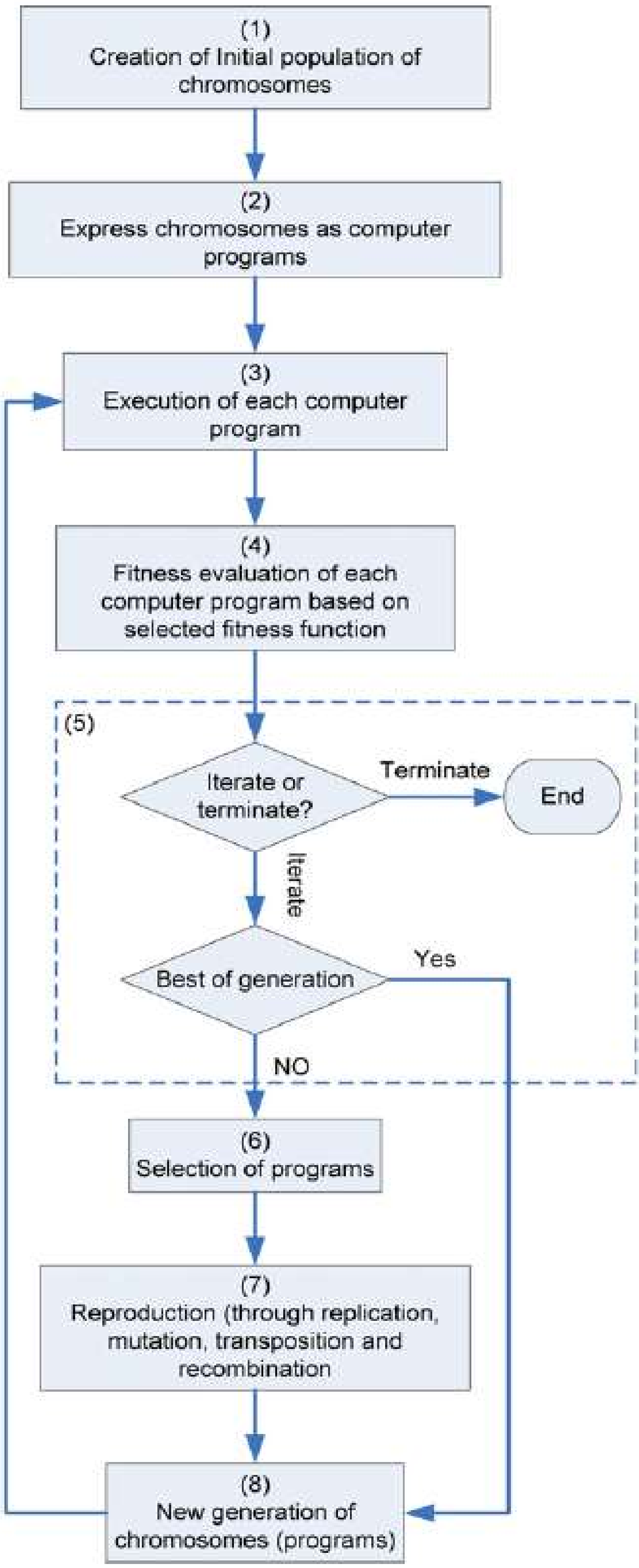

Figure 3.7 The flow diagram of gene expression logarithm (Hasmi,2011)

CHAPTER 4

CASE STUDY

4.1 Study Area and Watershed Delineation

In this study, Goksun River basin is used as a case study, which is the sub-basin of the Ceyhan River basin. This basin area is located 100 km from Kahramanmaraş on the boundary of Goksun region on the eastern Mediterranean, south of Turkey. Figure 4.1 shows the location of the basin according to Turkey and Kahramanmaraş province map. This basin surrounds an area of 2307 km^2, and its elevation ranges from 1170 m to 2800 m above sea level.

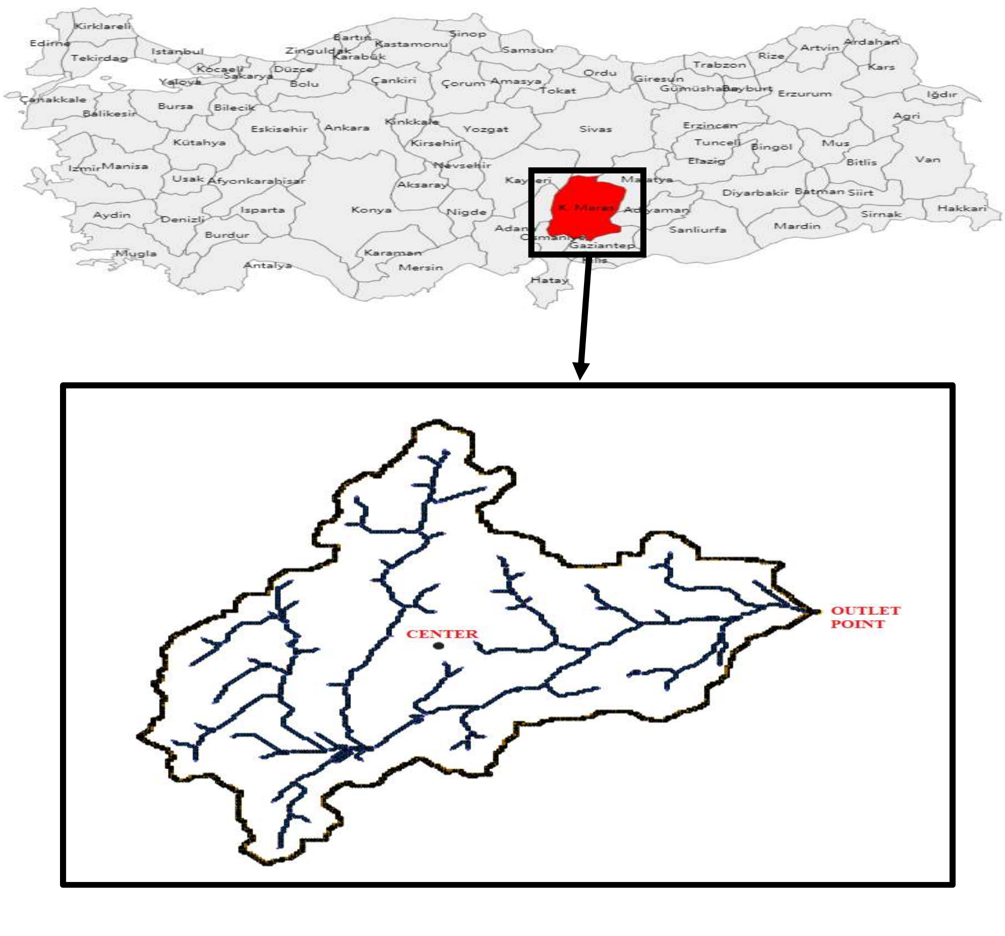

Figure 4.1 Location of the study area (Goksun River basin) according to the Turkey map

The center of the basin is located at coordinates 38° 07' 47" latitude 36° 36' 39" longitude as shown in Figure 4.2.

Figure 4.2 Location of the centroid of the basin

There are three local gauge stations near the study area. The locations and properties of the gauge stations are given in Table 4.1.

Table 4.1 Properties of the gauge stations

Station No	Station Name	Altitude (m)	Latitude (N)	Longitude (E)
17868	Afsin	1230	38° 14' 42"	37° 11' 15"
17870	Elbistan	1137	38° 01' 05"	37° 11' 54"
17866	Goksun	1344	38° 01' 26"	36° 28' 57"

The location of a flow gauge station is used for the outlet point of basin, name of gauge station is Poskoflu. The coordinates of the station is 38° 08' 55" latitude 37° 00' 04" longitude.

Watershed delineation is obtained by using the Watershed Modelling System (WMS) software, developed by Brigham Young University and U.S. Army Corps of Engineers. It is an exhaustive graphical modeling environment for all phases of watershed hydrology and hydraulics. WMS comprises strong tools to automate watershed modeling steps such as automated basin or sub-basin delineation, determination of geometric parameters, hydrological simulation, geographical information system overlay operations and stream cross-section extraction from terrain data (Brigham Yong University, 2010). WMS supports a lot of hydrological models contain (*HEC-1, HEC-HMS version 4, TR-20, TR-55, Rational Method, NFF, MODRAT, GSSHA, MODRAT, HEC-RAS, SMPDBK, SWMM, and HSPF*). WMS used to delineate the Goksun River watershed boundary by using the digital elevation model (DEM). DEM is a solid data structure that includes two-dimensional arrays of elevations where the spacing between elevations is constant in the x and y directions.

WMS operates the Topographic Parameterization Program (TOPAZ) for computing flow directions and flows accumulation. The flow direction describes the direction of flow from individual DEM cell to the neighboring cell that has the lowest elevation. The flow accumulation is evaluated from flow direction cell, which stands for the number of cells that drain into individual DEM cells. The high accumulation values illustrate cells in the stream, whereas low accumulation values represent areas of overland flow. TOPAZ is capable of further DEM elevation processing, watershed and stream network extracting and ordering. Drainage module in WMS is used to delineate the watershed boundary for Goksun river. Figure 4.3 illustrates the Goksun watershed boundary delineated by WMS and DEM contours with color fill. Figure 4.4 shows the flow distance contours that studied by TOPAZ in WMS and stream network of the basin. The physical characteristics of the drainage watershed contain drainage area, basin shape, basin slope, and centroid…etc. Table 4.2 displays the geomorphological parameters of the Goksun River watershed. Finally, Figure 4.5 indicates the shape of the basin located on google earth.

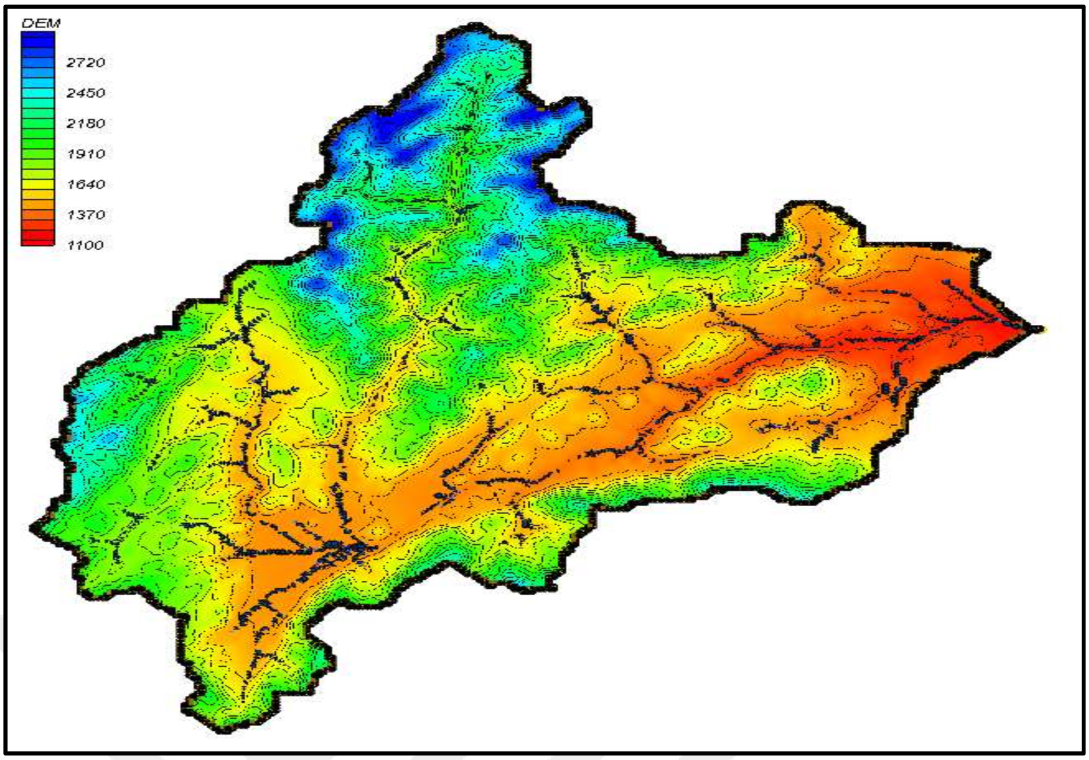

Figure 4.3 The boundary of the Goksun River basin and DEM contours

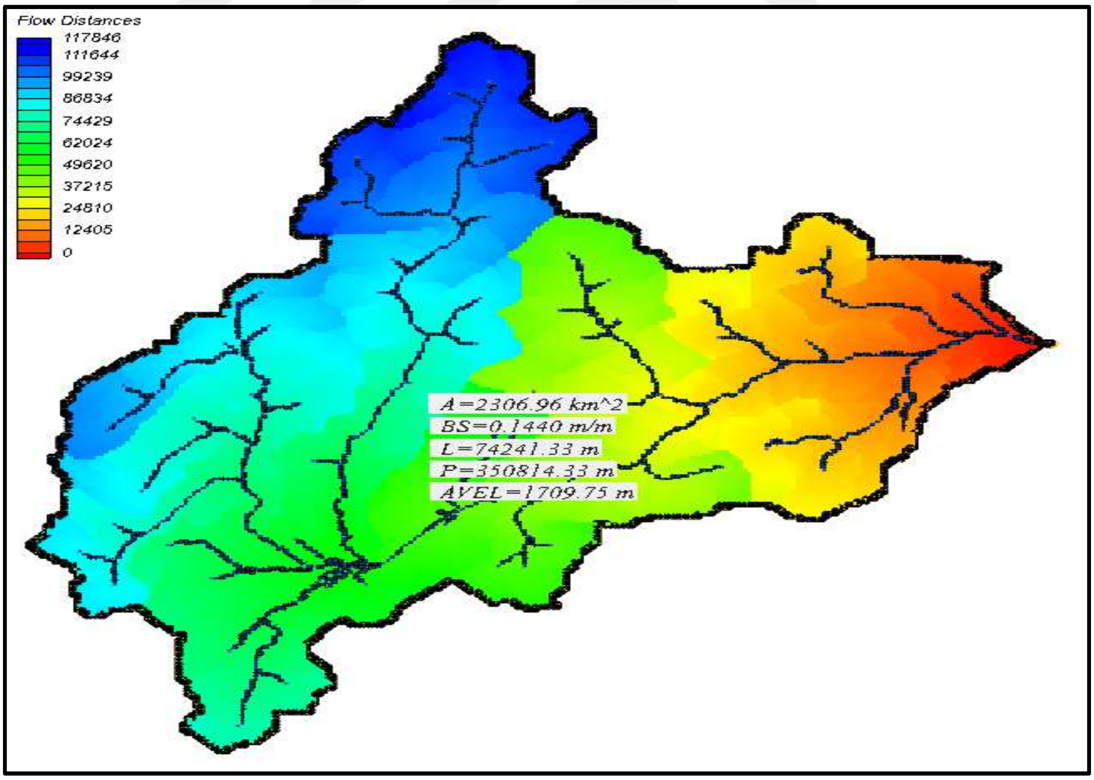

Figure 4.4 Flow distance contours and stream network for the Goksun River basin

Table 4.2 Geomorphological characteristics for Goksun River watershed.

Basin area (km^2)	2307
Basin Slope (m/m)	0.144
North aspect	0.46
South aspect	0.54
Perimeter (km)	350.8
Basin length (km)	74.24
Shape factor	2.39
Sinuosity factor	1.59
Mean basin elevation (m)	1709.75
Maximum flow distance (km)	124.05
Maximum flow slope (m)	0.013
Maximum stream length (km)	117.77
Maximum stream slope (m)	0.0091
Distance from centroid to stream (m)	2834.47
Centroid stream distance (km)	61.28
Centroid stream slope (m/m)	0.0053

Figure 4.5 Shape of basin getting from WMS program located in Google Earth

4.2 Precipitation Data Collection

All the meteorological data are taken from state meteorological works of Turkey (DMI) for three gauge stations listed in Table 4.1. The duration of available daily rainfall data for Afsin, Elbistan, and Goksun gauge stations are 1971-2011, 1970-2011, and 1970-2011, respectively. Maximum annual rainfall charts are shown in Figure 4.6,4.7, and 4,8 for each gauge station. All the data are listed in Appendix A.

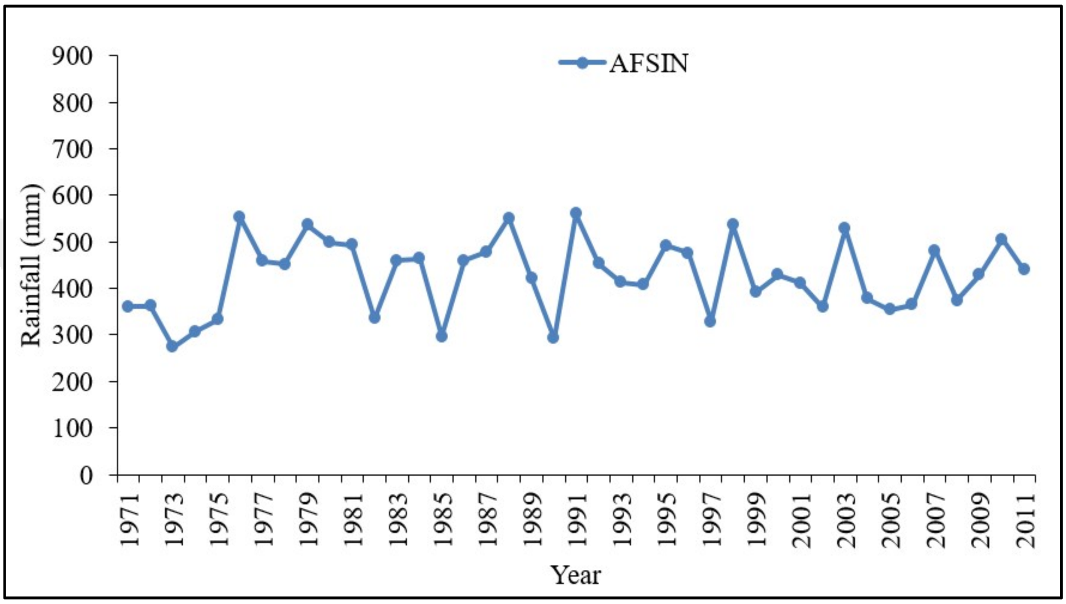

Figure 4.6 Maximum annual rainfalls in the gauge of Afsin

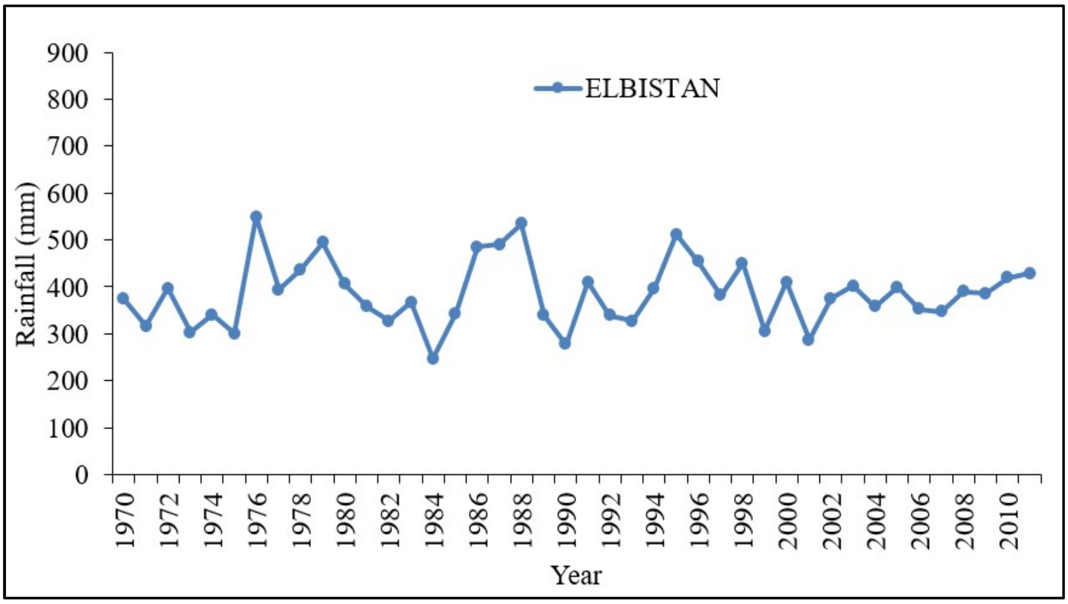

Figure 4.7 Maximum annual rainfalls in the gauge of Elbistan

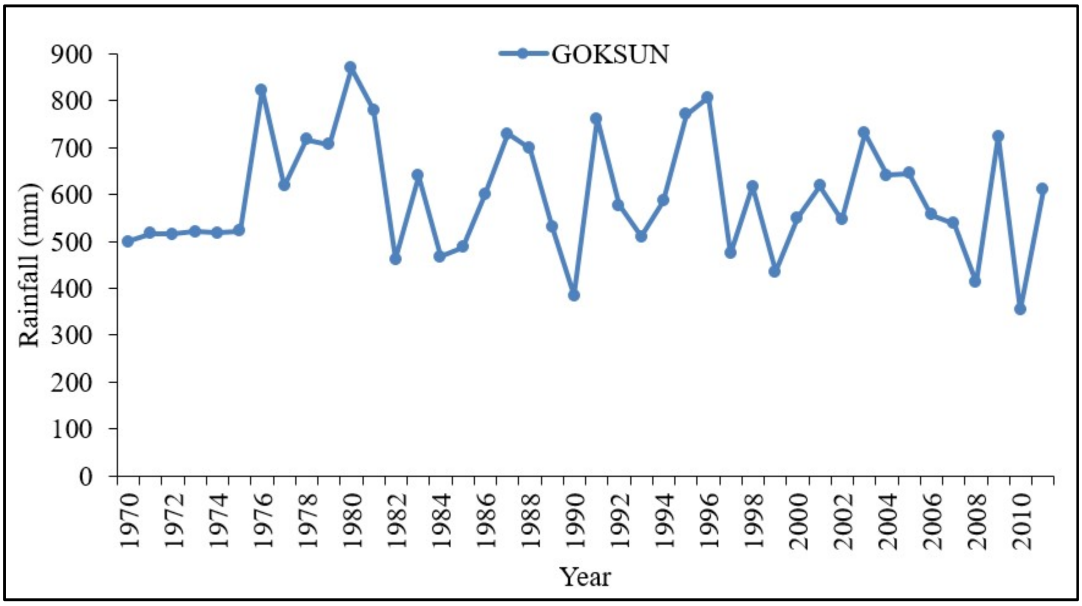

Figure 4.8 Maximum annual rainfalls in the gauge of Goksun

Table 4.3 shows the summary of the data used in this study for each station. As shown in the table, 41 year periods of data are used in this study which was enough to work with downscaling data.

Table 4.3 Summary of the data used in this study for each station

Item	Afsin	Goksun	Elbistan
Mean Annual rainfall (mm)	426.50	597.20	385.88
Mean Monthly rainfall (mm)	35.54	49.77	32.16
Maximum Annual Rainfal (mm)	560.30	869.40	548.50
Minimum Annual Rainfall (mm)	273.80	354.20	247.40
Maximum Daily Rainfall (mm)	65.50	85.40	46.10
Duration of Data (year)	41	42	42
Number of Data Set (days)	14,965	15,330	15,330

4.3 Monthly areal mean precipitation (Predictand)

Monthly areal mean precipitations on the basin were calculated by using the Thiessen polygons method from 3 gauge stations as shown in Figure 4.9.

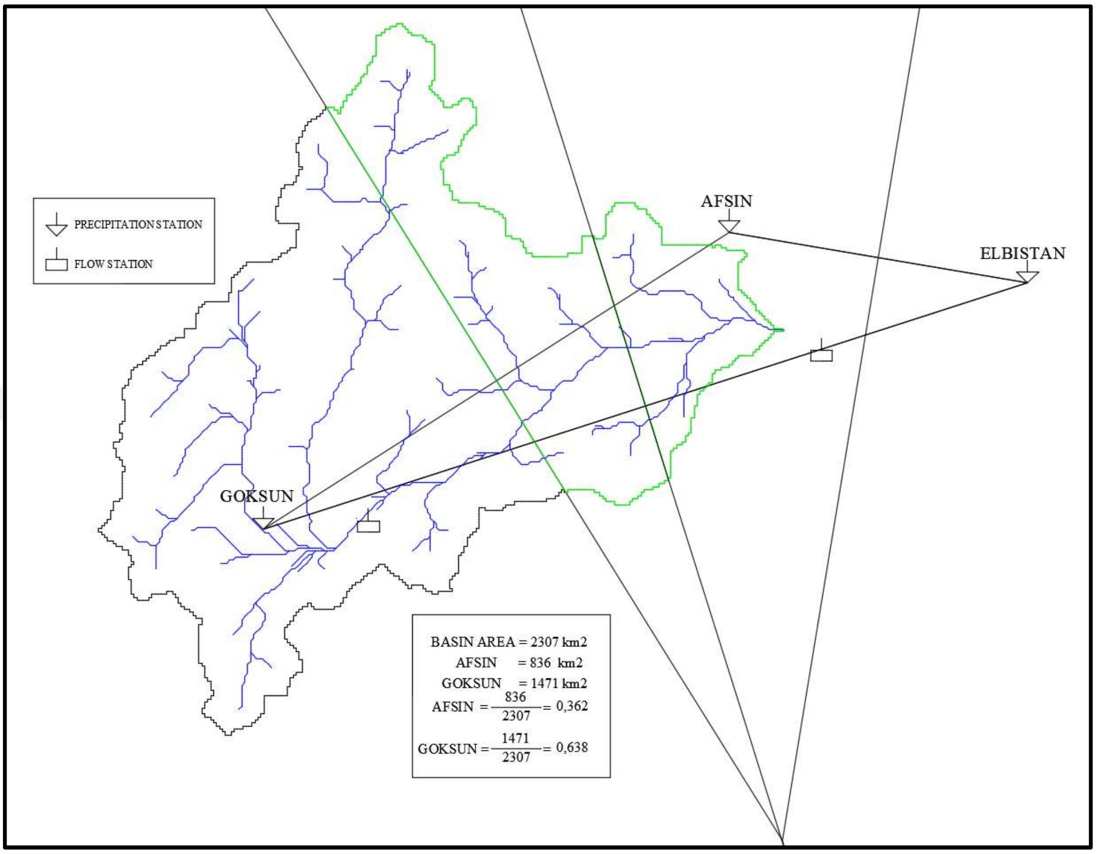

Figure 4.9 Thiessen polygons of the basin

The monthly areal mean precipitation was categorized into the calibration period (1971-1990), to evolve statistical downscaling models, and the validation period (1991-2000), to investigate the model's performance and compare downscaling results. The monthly areal mean precipitation will be used as predictand for the downscaling techniques.

Figures 4.10 and 4.11 present monthly areal mean rainfall and average monthly precipitation days (daily precipitation > 0 mm) for the 1971-2000 periods. Lastly, maximum annual rainfall is shown in Figure 4.12 with a baseline which is the mean annual rainfall for the 1971-2000 periods. Minimum, mean, and maximum annual precipitations are 362.8 mm, 542 mm, and 744.5 mm, respectively for the 1971-2000 periods.

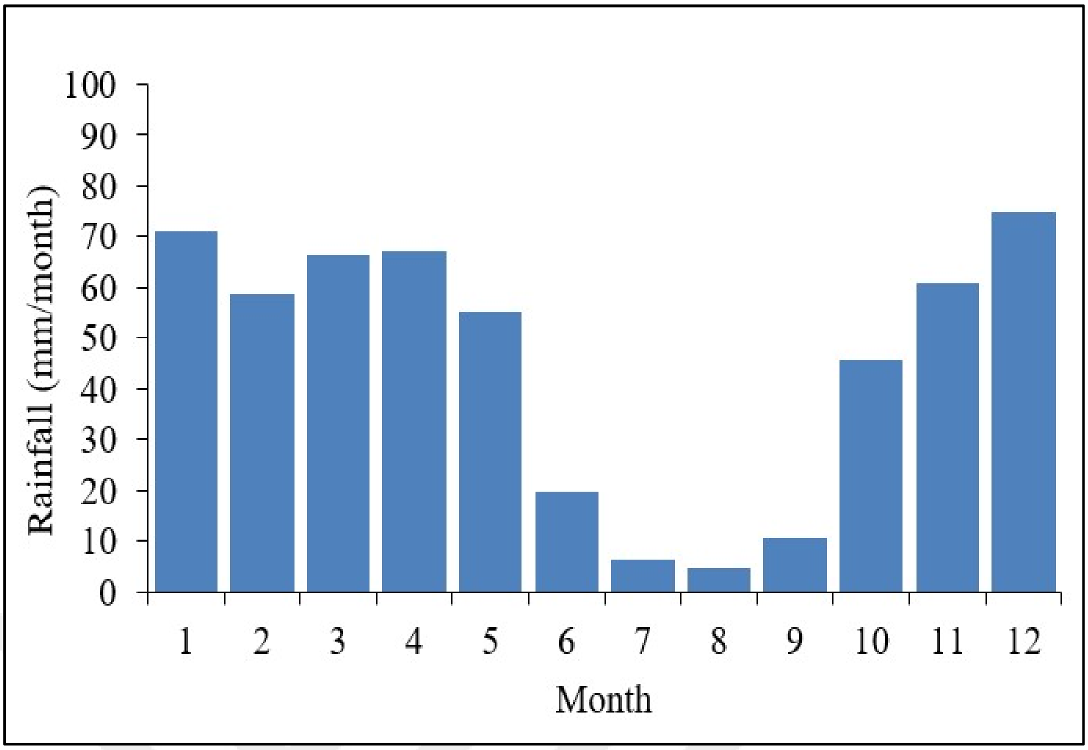

Figure 4.10 Monthly areal mean precipitation for the periods 1971-2000

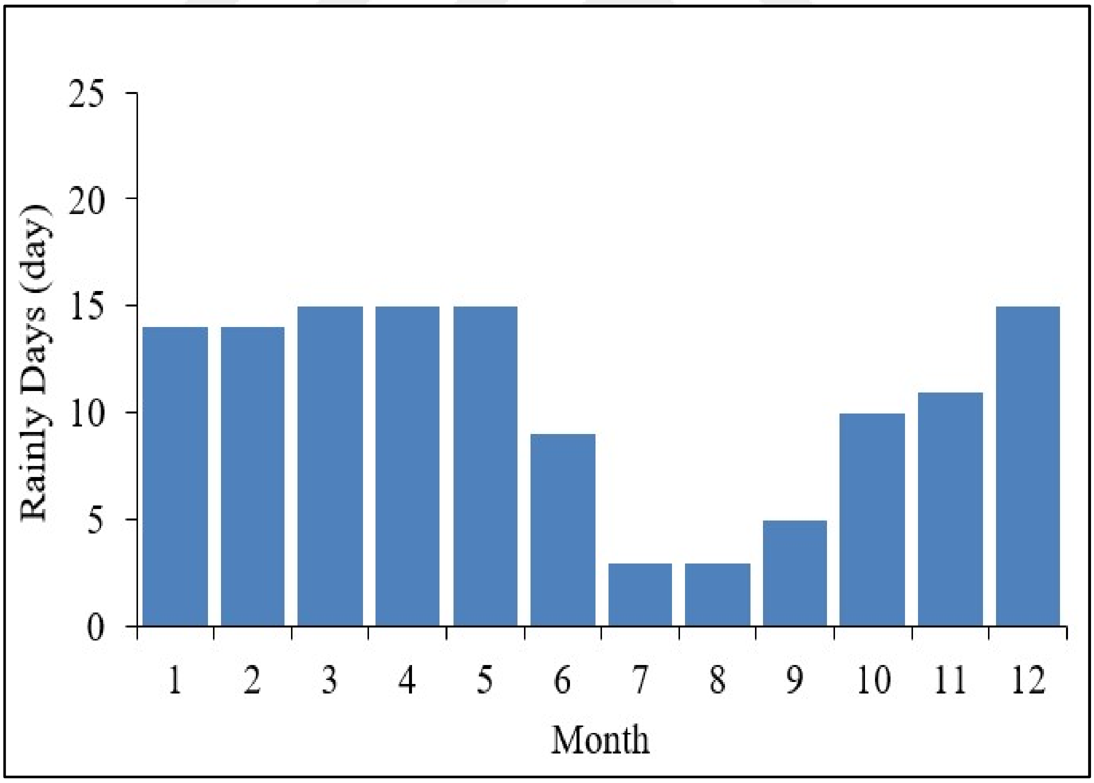

Figure 4.11 Average monthly precipitation days for the periods 1971-2000

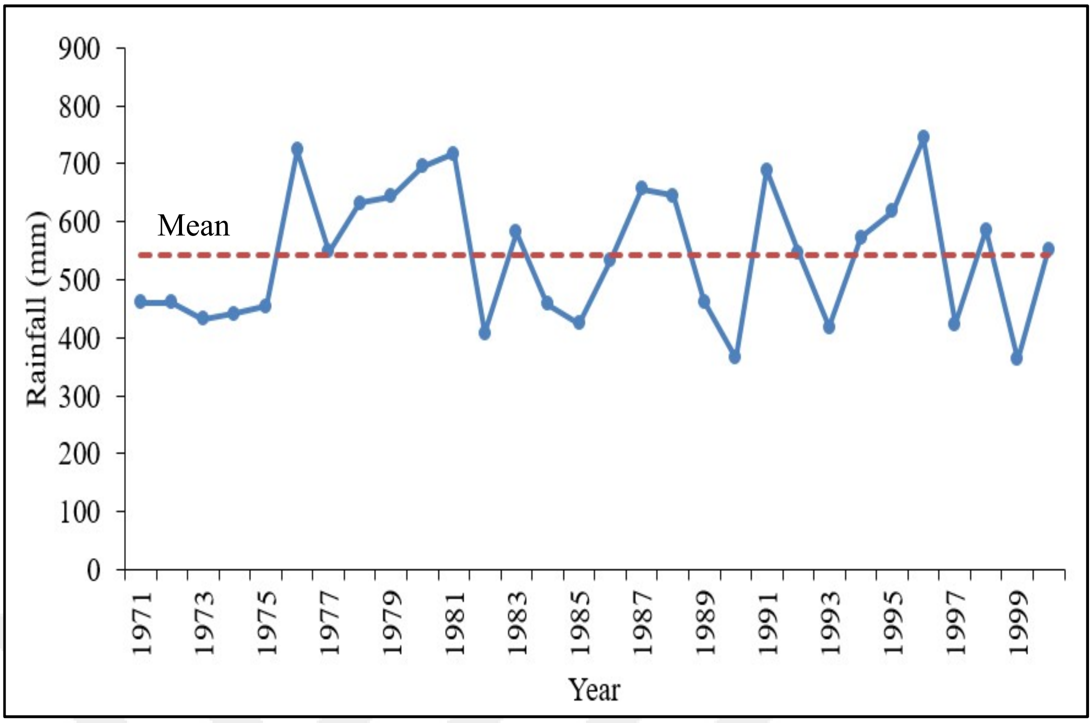

Figure 4.12 Maximum annual rainfall with the mean baseline for the periods 1971-2000

4.4 Acquisition of large-scale variables (predictors)

The data set contains both local precipitation on the basin and large-scale factors. In this study, the local precipitation is the output in the downscaling techniques. The large-scale factors (input data sets) can be downloaded as a zip file from the Canadian Center for Climate Modeling and Analysis scenarios (CCCma) (http://www.ec.gc.ca/ccmac) presents GCM daily time scale information for various atmospheric and surface variables for the CGCM3 T47 version. The combination of CGCM3 consists of data for two future scenarios A2 and A1B, which can be obtained from the Canadian Climate data and scenarios official website (http://www.csn.ec.gc.ca). The available data from the website presented according to grid cells. Each grid cell stands for a mesh surrounding the corresponding model grid points. The dimension of each grid cell is approximately 3.75° latitude and 3.75° longitude (Gaussian grid). This process takes place by input decimal latitude and longitudinal coordinate of any location, and in this study coordinate of the centroid of the basin is used, as shown in Figure 4.13. The CGCM3 variables are used as predictors in this study, they are broadly used in several climate changes and downscaling studies, such as Tofiq and Guven (2014), Singh et al., (2015) and Tofiq and Güven, (2015).

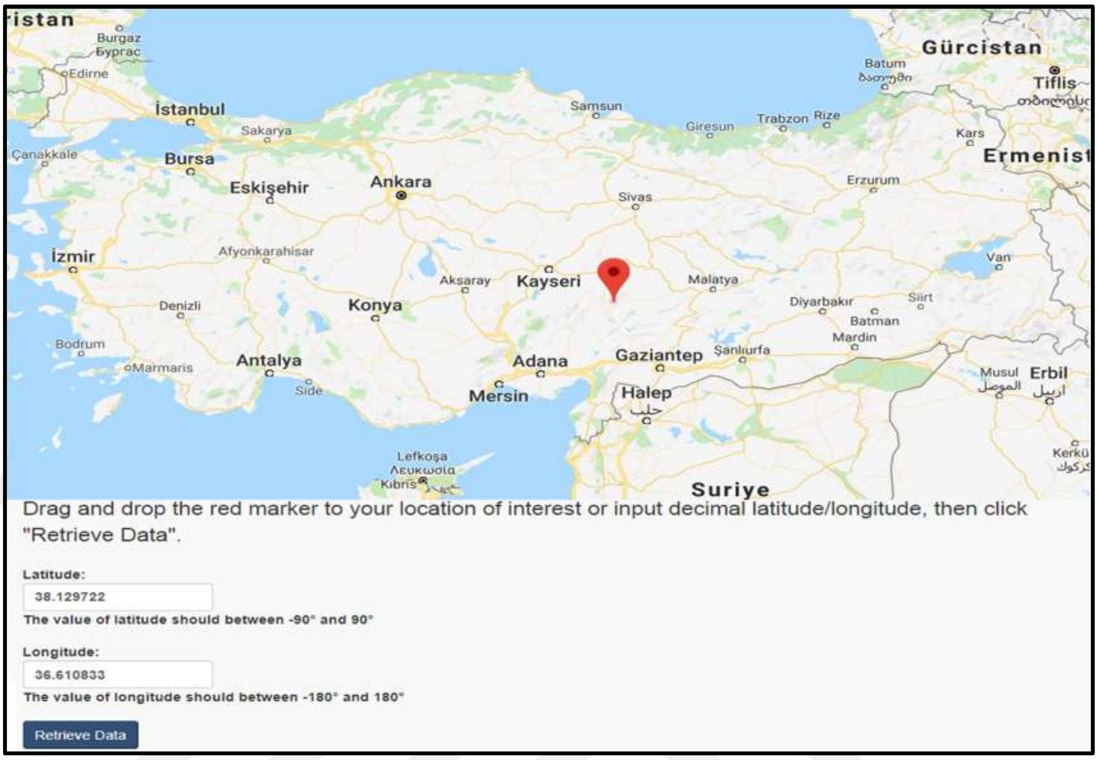

Figure 4.13 Location of the centroid of the basin in Environmental Canadian Global Climate Model

When entered retrieve data, a zipped file includes four folders and one-word file as shown in Figure 4.14, downloaded according to grid cell 11X-14Y.

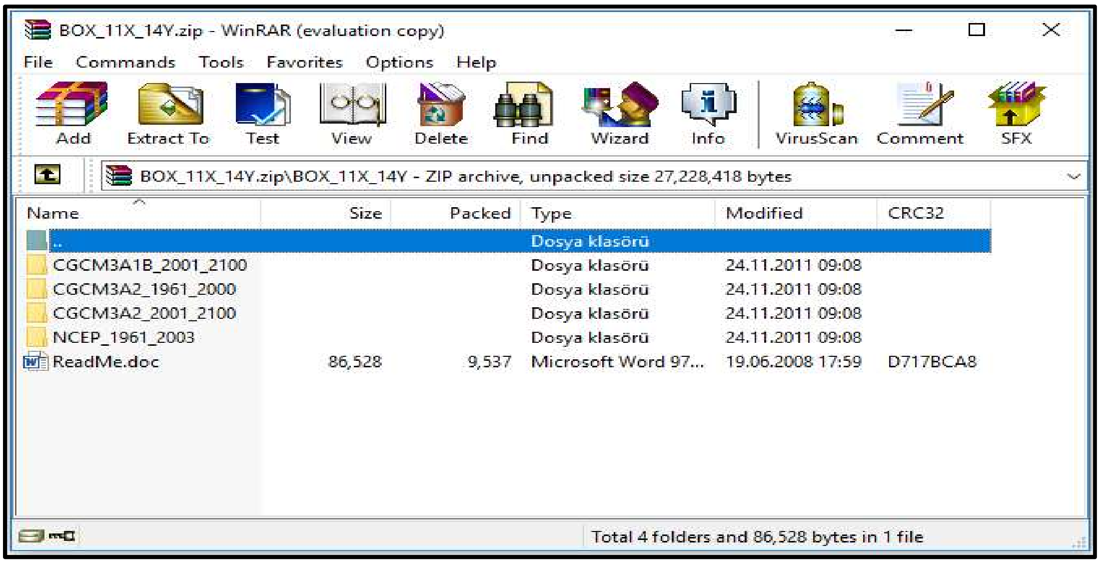

Figure 4.14 Zipped file which includes CGM3 and NCEP large-scale factors

These CGCM3 folders include 26 large-scale factors as input variable sets (predictors) listed in Table 3. CGCM3A2_1961_2000 folder's variables are used as predictors for the calibration (1971-1990) and validation (1991-2000) periods. CGCM3A1B_2001_2100 and CGCM3A2_2001_2100 folders include A1B and A2 scenario outputs were used to project future precipitation. All predictor variables are available as a daily time scale data, so they are converted to the corresponding monthly mean data as a predictand. Table 4.4 utilizes the description of CGCM3 variables which contain 26 variable data sets.

Table 4.4 Description of the 26 CGCM3 variables (predictors)

NO	CGCM3 Predictors	Description	Short Name
1	c3a2mslpgl	Mean sea level pressure	v1
2	c3a2p__fgl	1000 hPa Wind speed	v2
3	c3a2p__ugl	1000 hPa Zonal velocity	v3
4	c3a2p__vgl	1000 hPa Meridional velocity	v4
5	c3a2p__zgl	1000 hPa Vorticity	v5
6	c3a2p_thgl	1000 hPa Wind direction	v6
7	c3a2p_zhgl	1000 hPa Divergence	v7
8	c3a2p5_fgl	500 hPa Wind speed	v8
9	c3a2p5_ugl	500 hPa Zonal velocity	v9
10	c3a2p5_vgl	500 hPa Meridional velocity	v10
11	c3a2p5_zgl	500 hPa Vorticity	v11
12	c3a2p5thgl	500 hPa Geopotential	v12
13	c3a2p5zhgl	500 hPa Wind direction	v13
14	c3a2p8_fgl	500 hPa Divergence	v14
15	c3a2p8_ugl	850hPa Wind speed	v15
16	c3a2p8_vgl	850 hPa Zonal velocity	v16
17	c3a2p8_zgl	850 hPa Meridional velocity	v17
18	c3a2p8thgl	850 hPa Vorticity	v18
19	c3a2p8zhgl	850 hPa Geopotential	v19
20	c3a2p500gl	850 hPa Wind direction	v20
21	c3a2p850gl	850 hPa Divergence	v21
22	c3a2prcpgl	Accumulated precipitation	v22
23	c3a2s500gl	500 hPa Specific humidity	v23
24	c3a2s850gl	850 hPa Specific humidity	v24
25	c3a2shumgl	1000 hPa Specific humidity	v25
26	c3a2tempgl	Screen air temperature (2m)	v26

CHAPTER 5

RESULTS AND DISCUSSION

5.1 General

In this thesis, large-scale weather factors (26 input sets) were used as predictors which were obtained The Third Generation Coupled Global Circulation Model (CGCM3). The areal mean precipitation on the basin was used as a predictand (output). All data sets are based on a monthly time scale. Both data sets divided into two subsets (calibration and validation). The first one is the calibration periods range between 1971 and 1990 year. The second one is the validation periods range between 1991 and 2000 year. Group Method of Data Handling (GMDH), Gene Expression Programming (GEP), and Support Vector Machine (SVM) downscaling techniques were utilized to predict the downscaled monthly areal mean precipitation on the basin. Furthermore, the prediction results of downscaling techniques were found and compared with each other to examine the best model performance. Finally, Linear Regression was analyzed for comparison.

In this chapter, statistical results related to method sections described in the previous chapter for the downscaling process are presented and discussed. This includes the selection of the most effective predictors set, statistical results of calibration and validation periods, and future projection results of monthly areal mean precipitation (P_M). The graphical and statistical results of the four downscaling models, namely GMDH, GEP, SVM, and Linear regression models are displayed and analyzed for calibration and validation periods throughout this chapter. Moreover, the projections of annual areal mean precipitation are compared with the baseline annual mean precipitation from 1971 to 2000 period to investigate the expected climatic changes in the study area.

5.2 Selecting the most effective predictors

Selecting the large-scale weather GCM factors (predictors), 26 variables, is an important aspect in the evolution of the statistical downscaling approaches. To improve the correlation with predictor variables and predictands, variables were modified based on some normalization methods. Each data sets for predictors and predictand were converted to Ln (taking the natural logarithm of each data) variable, standardized (z value) variable, min/max (Mn) variable which the data range between 0 and 1. To select the most effective predictors and to see the linear relationship between inputs and outputs, correlation analysis was employed by Pearson rank correlation coefficient method. Large-scale weather factors were statically correlated with local precipitation under the confidence level of 99%. Factors with a higher correlation coefficient were selected as the predictors for downscaling models. The correlation coefficient (R) between the predictors and predictands were evaluated and listed in order of array as presented in Table 5.1, Table 5.2, Table 5.3 and Table 5.4 for each normalized data sets.

Table 5.1 The correlation coefficient (R) of Ln P_M (natural logarithm variables) between P_M and 26 GCM predictors.

No	GCM Predictors	R	No	GCM Predictors	R
1	V 1	0.538	14	V 14	0.297
2	V 2	0.286	15	V 15	0.544
3	V 3	0.475	16	V 16	0.569
4	V 4	0.577	17	V 17	-0.248
5	V 5	-0.273	18	V 18	-0.551
6	V 6	-0.323	19	V 19	-0.582
7	V 7	-0.593	20	V 20	-0.650
8	V 8	0.327	21	V 21	-0.082
9	V 9	0.202	22	V 22	0.065
10	V 10	-0.103	23	V 23	-0.151
11	V 11	0.016	24	V 24	-0.340
12	V 12	0.199	25	V 25	-0.666
13	V 13	0.118	26	V 26	-0.680

Table 5.2 The correlation coefficient (R) of standardized Z P_M variables (z value) between P_M and 26 GCM predictors.

No	GCM Predictors	R	No	GCM Predictors	R
1	V 1	0.461	14	V 14	0.344
2	V 2	0.312	15	V 15	0.532
3	V 3	0.460	16	V 16	0.514
4	V 4	0.533	17	V 17	-0.193
5	V 5	-0.142	18	V 18	-0.512
6	V 6	-0.341	19	V 19	-0.520
7	V 7	-0.552	20	V 20	-0.614
8	V 8	0.326	21	V 21	-0.143
9	V 9	0.235	22	V 22	0.058
10	V 10	-0.090	23	V 23	-0.199
11	V 11	-0.014	24	V 24	-0.319
12	V 12	0.100	25	V 25	-0.587
13	V 13	0.107	26	V 26	-0.614

Table 5.3 The correlation coefficient (R) of min/max M P_M variables between P_M and 26 GCM predictors.

No	GCM Predictors	R	No	GCM Predictors	R
1	V 1	0.461	14	V 14	0.344
2	V 2	0.312	15	V 15	0.532
3	V 3	0.460	16	V 16	0.514
4	V 4	0.533	17	V 17	-0.193
5	V 5	-0.142	18	V 18	-0.512
6	V 6	-0.341	19	V 19	-0.520
7	V 7	-0.552	20	V 20	-0.614
8	V 8	0.326	21	V 21	-0.143
9	V 9	0.235	22	V 22	0.058
10	V 10	-0.090	23	V 23	-0.199
11	V 11	-0.014	24	V 24	-0.319
12	V 12	0.100	25	V 25	-0.587
13	V 13	0.107	26	V 26	-0.614

Table 5.4 The correlation coefficient (R) of the non-normalized P_M variables between P_M and 26 GCM predictors.

No	GCM Predictors	R	No	GCM Predictors	R
1	V 1	0.461	14	V 14	0.344
2	V 2	0.312	15	V 15	0.532
3	V 3	0.460	16	V 16	0.514
4	V 4	0.533	17	V 17	-0.193
5	V 5	-0.142	18	V 18	-0.512
6	V 6	-0.341	19	V 19	-0.520
7	V 7	-0.552	20	V 20	-0.614
8	V 8	0.326	21	V 21	-0.143
9	V 9	0.235	22	V 22	0.058
10	V 10	-0.090	23	V 23	-0.199
11	V 11	-0.014	24	V 24	-0.319
12	V 12	0.100	25	V 25	-0.587
13	V 13	0.107	26	V 26	-0.614

The set of the predictors are divided into two categories according to the value of correlation coefficient (R) between predictand and predictors. These categories are: R > 0.4 and R > 0.5. Four cases for the correlation coefficient between predictand and predictors variables were provided in this study. The first case is a normal case of R, the second case is the correlation coefficient between Ln P_M predictand and predictor variables, the third case is the correlation coefficient between the standardized Z P_M predictand and predictor variables, and the fourth case is the correlation coefficient between max/min M P_M predictand and predictor variables. R > 0.5 performed the best performance for each normalized methods. As shown in the upper tables, normalized taking the natural logarithm of each variable Ln P_M has the highest correlation coefficient (R) than the other normalized variables. For this reason in this study, predictors with a correlation of R > 0.5 were used for natural logarithm normalization of predictor and predictand variables. The highest correlating GCM variables, R>0.5 (10 in this case) were selected and presented in Table 5.5. The input correlation bigger than 0.5 were Mean sea level pressure (V1), 1000 hPa Meridional velocity (V4), 1000 hPa Divergence (V7), 850 hPa Wind speed (V15), 850 hPa Meridional velocity (V16), 850 hPa Meridional velocity (V17), 850 hPa Vorticity (V18), 850 hPa Geopotential (V19), 850 hPa Wind direction (V20), 1000 hPa Specific humidity (V25), Screen air temperature (2m) (V26).

Table 5.5 Summary of final most effective GCM set predictor

No	GCM Predictors	R	No	GCM Predictors	R
1	V 1	0.538	14	V 18	-0.551
2	V 4	0.577	15	V 19	-0.582
3	V 7	-0.593	16	V 20	-0.650
4	V 15	0.544	17	V 25	-0.666
5	V 16	0.569	18	V 26	-0.680

The magnitude of the correlation coefficient displays the degree of association between variables and the sign indicates whether the relationship is direct or inverse. For example, -0.666 shows a strong inverse relationship, meaning that, as the value of one variable increases, the value of the other variable decreases. In Table 5.5, it is clear that the V26 (R=-0.680) is the super predictor for this study area. Although several large-scale variables like V1 (R=0.461), do not show a greater correlation with local precipitation, they are selected. That is because the set of GCM variables would describe the conditional process of the local-scale variable, which relies upon many intermediate processes.

5.3 Statistical results of model calibration

The local precipitation (monthly areal mean precipitation) amount and large-scale whether data of the calibration period were used to calibrate the parameters and equations of downscaling models which are GMDH, SVM, and GEP.

The SVM model has three parameters which are C, γ, and ε. Table 5.6 shows parameters, which are acquired automatically via the SVM to predict the monthly areal mean precipitation.

Table 5.6 Parameters of SVM model

C	γ	ε
9.03	3.39	0.01

Figure 5.1 displays the relative importance of predictors used as predictors in the SVM model. V20 (850 hPa Wind direction) predictor is the most effective predictor while V4 (1000 hPa Meridional velocity) predictor has the lowest relative importance.

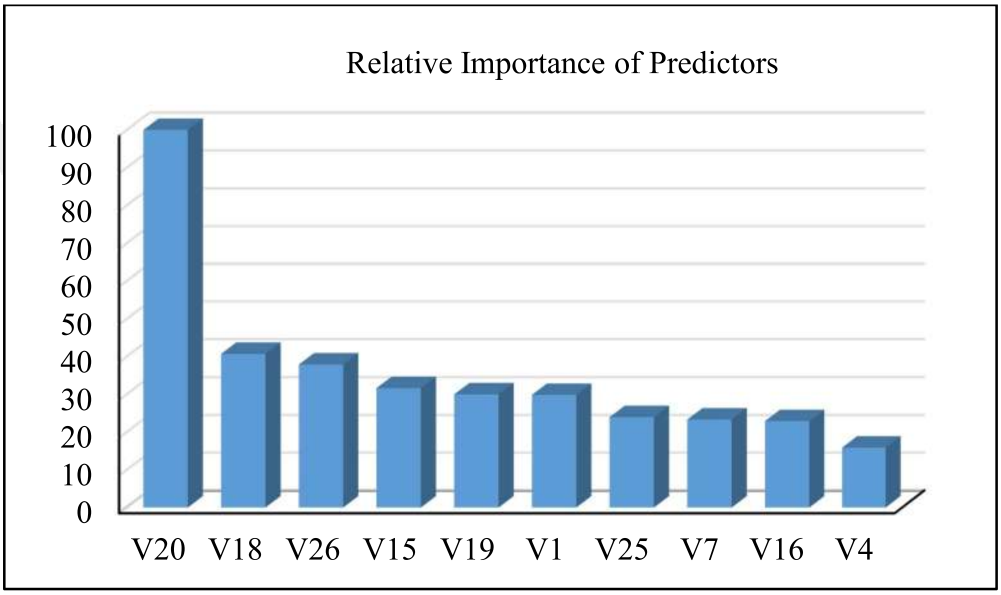

Figure 5.1 Relative importance of predictors of SVM model

The GMDH model developed equations for the calibration process to generate the validation process and to predict the future projections. Equations are given in Appendix B, generated by the GMDH model. The GMDH model chooses the most effective predictors automatically among 10 selected predictors. GMDH selected Mean sea level pressure (V1), 850 hPa Wind speed (V15), 850 hPa Meridional velocity (V16), 850 hPa Geopotential (V19), 850 hPa Wind direction (V20), Specific humidity (V25), Screen air temperature (2m) (V26) predictors to generate equation. Figure 5.2 displays the relative importance of predictors used as predictors in the GMDH model. V25 predictor is the most effective predictor while V20, V19, V1, V15, and V26 predictors have small relative importance.

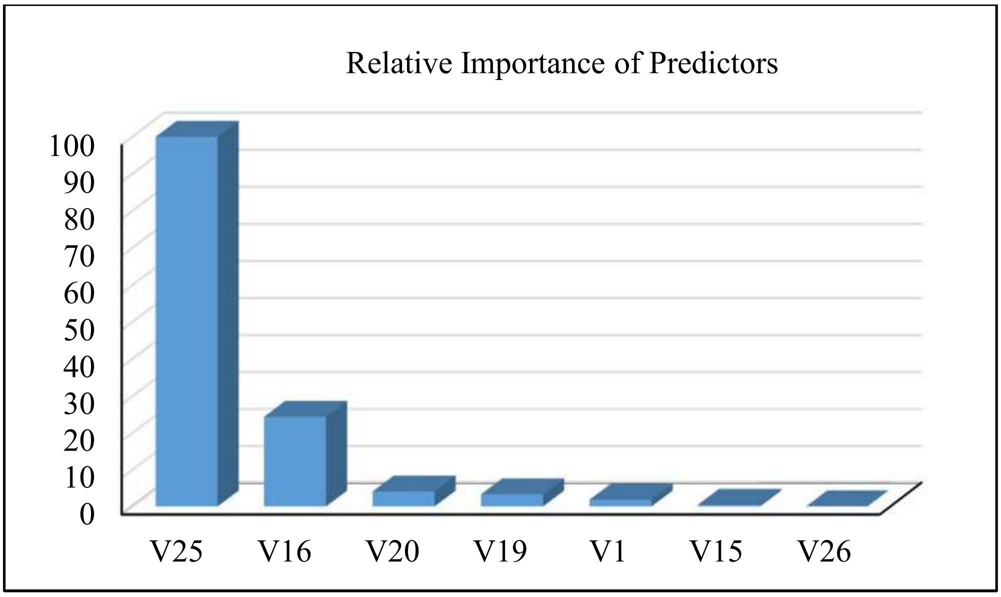

Figure 5.2 Relative importance of predictors of GMDH model

The GEP model also developed equations for the calibration process to generate the validation process and to predict the future projections. The GEP model chooses the most effective predictors automatically among 10 selected predictors. Generated equation of the GEP model is given equation 5.1. In equation 5.1, it is clearly seen that GEP selected Mean sea level pressure (V1), 1000 hPa Meridional velocity (V4), 850 hPa Wind speed (V15), 850 hPa Vorticity (V18), 850 hPa Wind direction (V20), Screen air temperature (2m) (V26) predictors to produce equation.

$$\mathrm{LnP}_M = \left(\frac{\left(\mathrm{V1} \times \left(\left(\frac{\mathrm{V26}}{\mathrm{V15}}\right) + \left(\frac{\mathrm{V4}}{\mathrm{V20}}\right)\right)\right)}{\mathrm{V18}}\right) + 2.4824 + \left(\frac{\left(\left(\mathrm{V1} - (\mathrm{V26} + 2.4924)\right) \times \mathrm{V26}\right)}{\mathrm{V18}}\right) + \left(0.0365 \times (27.383 - \mathrm{V26})\right) \quad (5.1)$$

Figure 5.3 displays the relative importance of predictors used as predictors in the GEP model. V26 predictor is the most effective predictor while V15, V1, V20, V18, and V4 predictors have small relative importance.

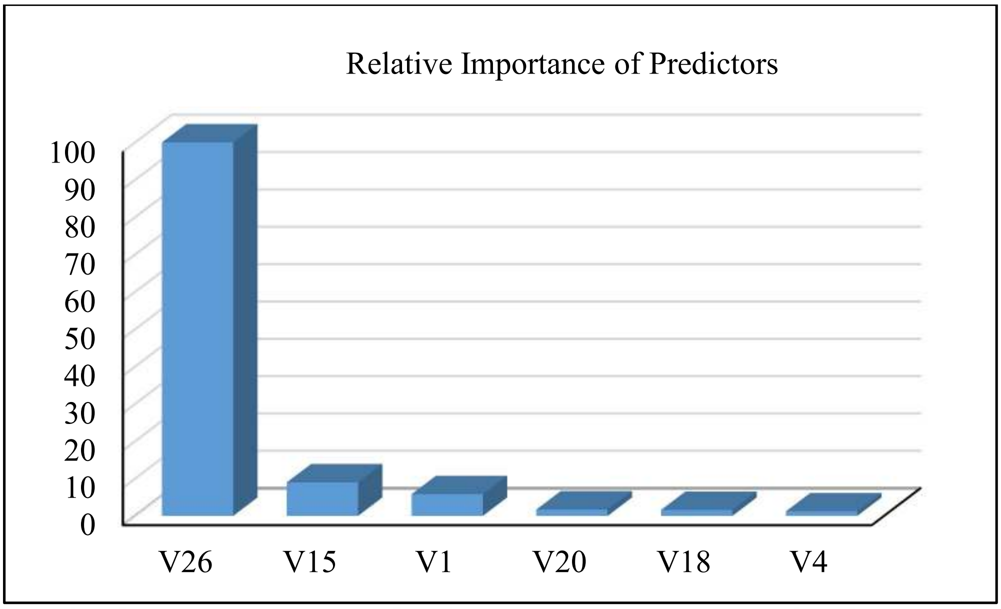

Figure 5.3 Relative importance of predictors of GEP model

The result of GMDH, GEP, SVM, and linear regression models for calibrations were expressed in correlation factor (R) for each model. A scatter plot was made between the observed (vertical axis) and the predicted P_M (horizontal axis) as presented in Figures 5.4-5.7 for the normalized variable (natural logarithm). The scatter diagram shows us the main vision of the relationship between the variables (observed and predicted). In Figures 5.4-5.7, it is seen that all models have an almost uphill pattern when moved from left to right. It means the models were run perfectly and made a good prediction according to observed data. Furthermore, points pattern in the SVM model (R =0.863) crowded along with the fit line as shown in Figure 5.6. While the points pattern in the linear regression (LR) model presented in Figure 5.7 is dispersed around the fit line with lower R = 0.746. The results also show that GMDH and GEP models are almost the same R-value (0.796-0.770). As a result, the SVM model related better to predict the monthly areal mean precipitation P_M.

The linear equations comprising the relation between the observed and the predicted P_M for the four models (GMDH,GEP,SVM,LR) were obtained to be:

$$y = 0.9319x + 0.2574 \quad \text{(GMDH model)} \tag{5.2}$$

$$y = 0.945x + 0.2204 \quad \text{(GEP model)} \tag{5.3}$$

$$y = 0.9666x + 0.1512 \quad \text{(SVM model)} \tag{5.4}$$

$$y = 0.9544x + 0.176 \qquad \text{(LR model)} \qquad (5.5)$$

Where y represents the observed P_M and x represents the predicted P_M.

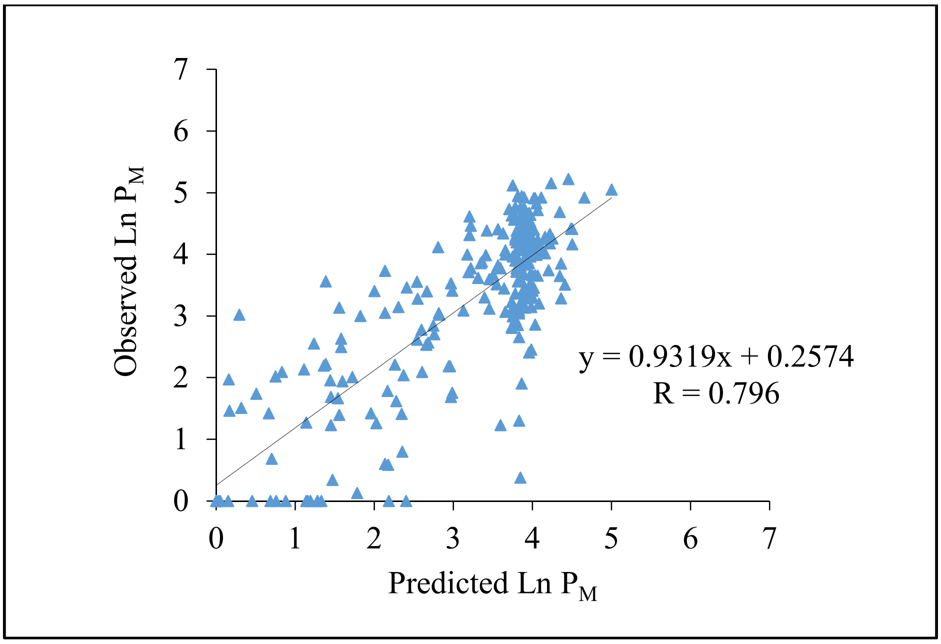

Figure 5.4 Scatter plot of the observed and predicted Ln PM values for calibration period 1971-1990 by GMDH model

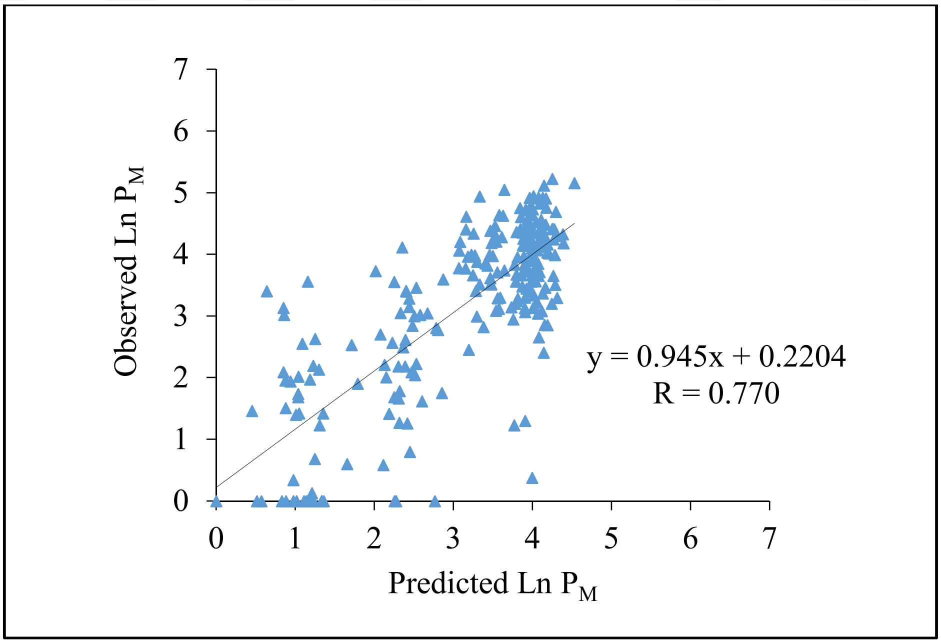

Figure 5.5 Scatter plot of the observed and predicted Ln P_M values for calibration period 1971-1990 by GEP model

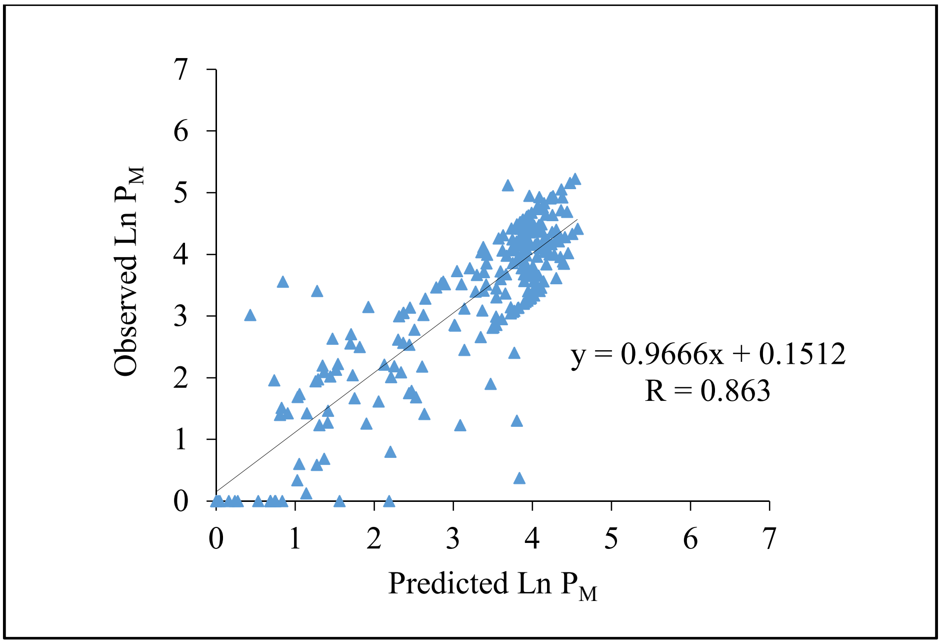

Figure 5.6 Scatter plot of the observed and predicted Ln P_M values for calibration period 1971-1990 by SVM model

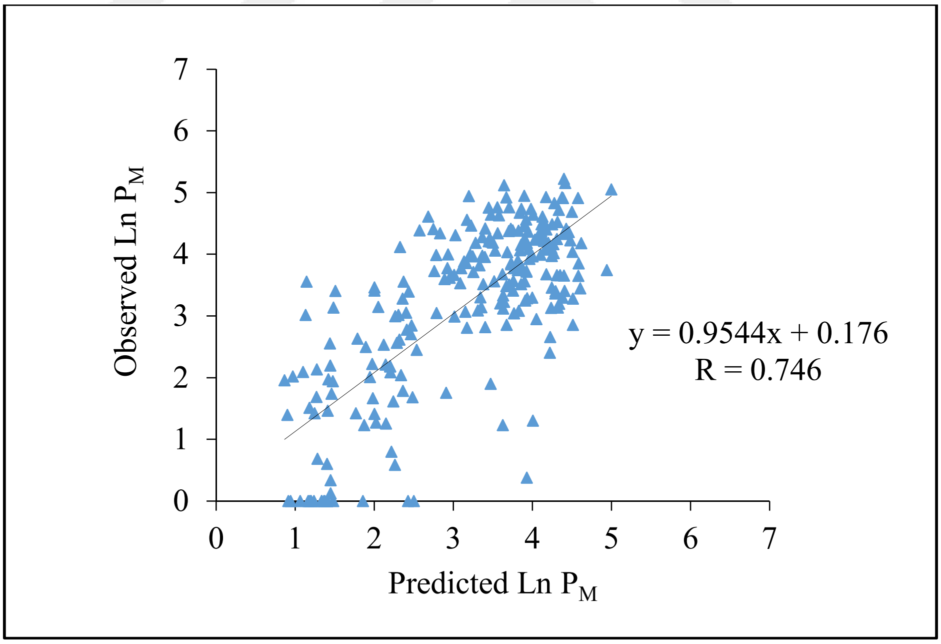

Figure 5.7 Scatter plot of the observed and predicted Ln P_M values for calibration period 1971-1990 by Linear Regression model

The outputs value (predicted P_M) of downscaling models can be negative values. Under this situation, negative values were adjusted to zero. The percentage of negative values over the total number of monthly areal mean precipitations were 0.83 % in the GMDH and GEP, 0.28 % in SVM, and 2.22 % in linear regression models. Linear regression has the biggest percentage correction of negative values adjusted to zero, while GMDH, GEP, and SVM need less correction. These corrections affect the statistical properties of models and cause the tendency to right of predicted value when the observed value is equal to 0.

Table 5.7 presents the statistical comparison between the simulated P_M for the four models and the observed P_M under the same calibration period (1971-1990).

From table 5.7, the simulated mean monthly areal mean precipitation by SVM model, LR model, GMDH model, and GEP model are 7.5, 8.1, 8.8, and 9,4 mm respectively, lower than the mean of observation. It means the means were underestimated by all models. However, the SVM model yielded better results than the other models to simulate mean. The SVM model performs well to simulate minimum precipitation while linear regression model performs poorly to simulate minimum precipitation. Moreover, there is a significant difference in the maximum value of simulated P_M compared to observed data. GMDH model record the minimum difference of 35.3% for maximum precipitation, and GEP model record minimum difference of 91.3% for maximum precipitation. Like the mean, all models underestimated the standard deviation, but linear regression model has the minimum underestimating difference of 26.4% among the other models. The standard deviation specifies the dispersion of the precipitation from their mean. Accordingly, downscaled precipitations with smaller standard deviation distribute closer around the mean value than the observations.

As seen by the values of statistical indicators given in Table 5.7, the SVM model has a lesser MAE (18.83 mm) and RMSE (27.28 mm) than the other proposed models. The MAE is approximately 12.6%, 15.6% and 26.8% for the GMDH model, GEP model, and linear regression model, respectively, greater than that of the SVM model. The RMSE is approximately 13%, 16.5%, and 26.9% for the GEP model, GMDH model, and linear regression model, respectively, greater than that of the SVM model.

Table 5.7 Comparision of statistical results of observed and simulated P_M during calibration (1971-1990)

Data	Mean	Min	Max	Std. Deviation	MAE	RMSE	R
Observed	44.771	0.07	185.58	39.184	-	-	-
GMDH model	35.984	1.03	148.41	24.031	21.21	30.83	0.796
GEP model	35.416	1.57	92.83	23.615	21.77	31.78	0.770
SVM model	37.282	1.02	96.33	25.070	18.83	27.28	0.863
Linear Regression	36.654	2.37	147.89	28.851	23.87	34.61	0.746

Also, the AIC values of each proposed models are calculated and given in Table 5.8. The number of data used in the calibration period is 240. It can be seen that AIC increases when the number of the fitting parameter (k) increases. Table 5.8 indicates, the SVM model has the lowest MSE value (744.2 mm^2), however SVM model has the biggest AIC value (3831) because of having high number of fitting parameter. Consequently, the lowest AIC with k values of the GEP model presents its best robustness and the generalization capacity.

Table 5.8 AIC values of proposed models during calibration (1971-1990)

Model	MSE (mm^2)	k	AIC
GMDH	950.2	490	2625
GEP	1009.7	4	1668
SVM	744.2	1122	3831
LR	1197.7	11	1723

Statistical results and scatter plots indicate, the SVM model performs well in the calibration period and simulated better results than the other downscaling models but AIC values show the GEP model has the lowest AIC value.

Figures 5.8-5.11 show the simulated averaged monthly mean precipitation P_M of GMDH, GEP, SVM, and LR models and observed data during the calibration period. Figure 5.12 illustrates the simulated averaged monthly mean precipitation P_M of all models and observed data in one figure. It is clear that the SVM model has slightly underestimated the P_M each month. In January, February, and March linear regression model overestimated the P_M while other models underestimated the P_M as shown in figure 5.12. In September, the SVM model underestimated the P_M while other models overestimated the P_M. In April, May, Jun, July, August, October, November, and December all models underestimated the P_M. In April, May, Jun, September, and October the SVM simulated the best averaged mean precipitation compared to observed data. The mean underestimated percentage in May which are 12.4%, 42.5%, and 47% for the GMDH model, GEP model, and linear regression model, respectively, smaller than that of the SVM model as shown in figure 5.12. In March, the SVM and GEP models simulated the same amount of mean precipitation and perform the best prediction for this month. In November and December, the GEP model performs best to predict mean precipitation. The mean underestimated percentage in December which are 7.3%, 8.9%, and 13% for the linear regression model, SVM model, and GMDH model, respectively, smaller than that of the GEP model. In January, February, July, and August, the linear regression model performs best to predict P_M. Finally, as shown in Figures 5.8-5.12 and Table 5.7, the SVM model performs well in predicting the variation in monthly areal mean precipitation data when compared to the rest models for the calibration period. Tables in Appendix C1, D1, E1, and F1 present the numerical results of observed and predicted averaged monthly mean precipitation values during the calibration period. The percent absolute errors are displayed in Figures C1, D1, E1, and F1 during the same calibration period.

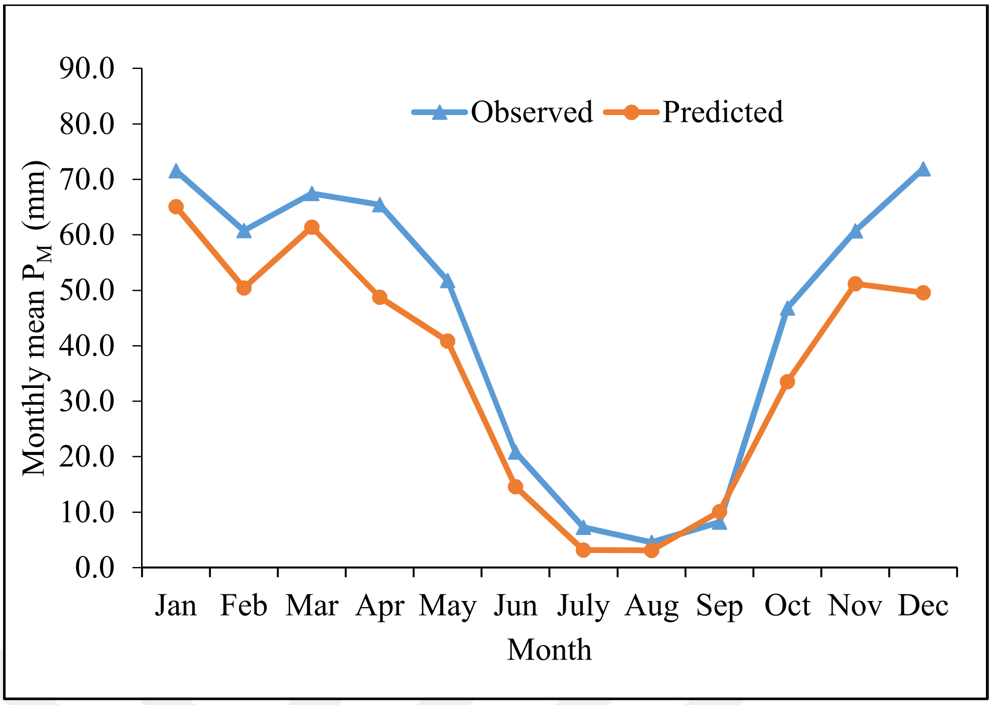

Figure 5.8 Observed and simulated averaged monthly P_M for the calibration period 1971-1990 by GMDH model

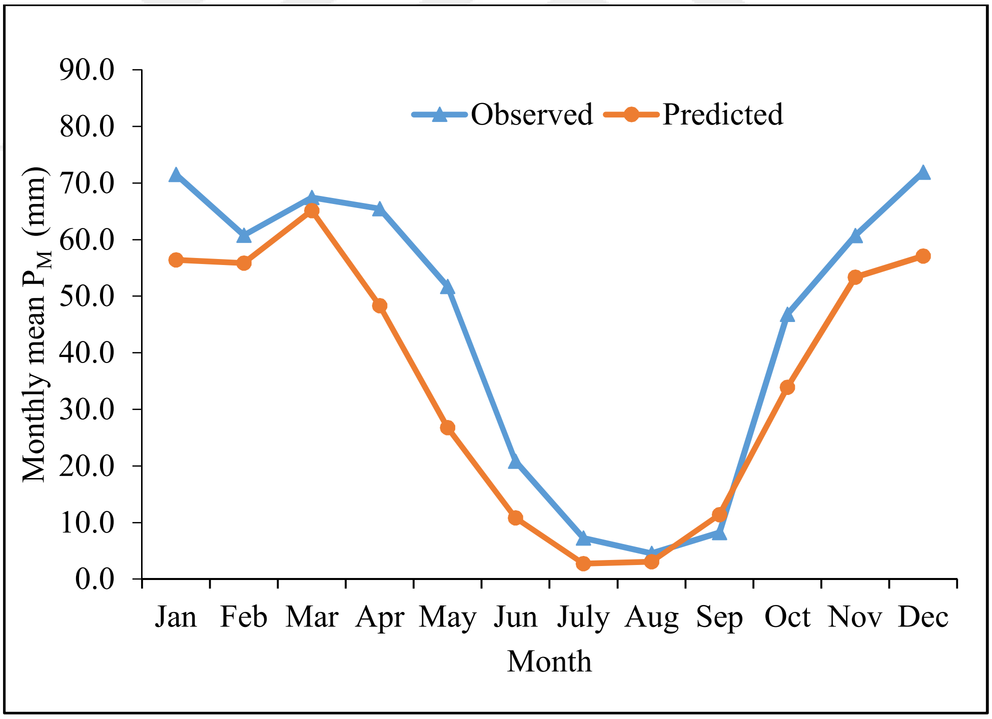

Figure 5.9 Observed and simulated averaged monthly P_M for the calibration period 1971-1990 by GEP model

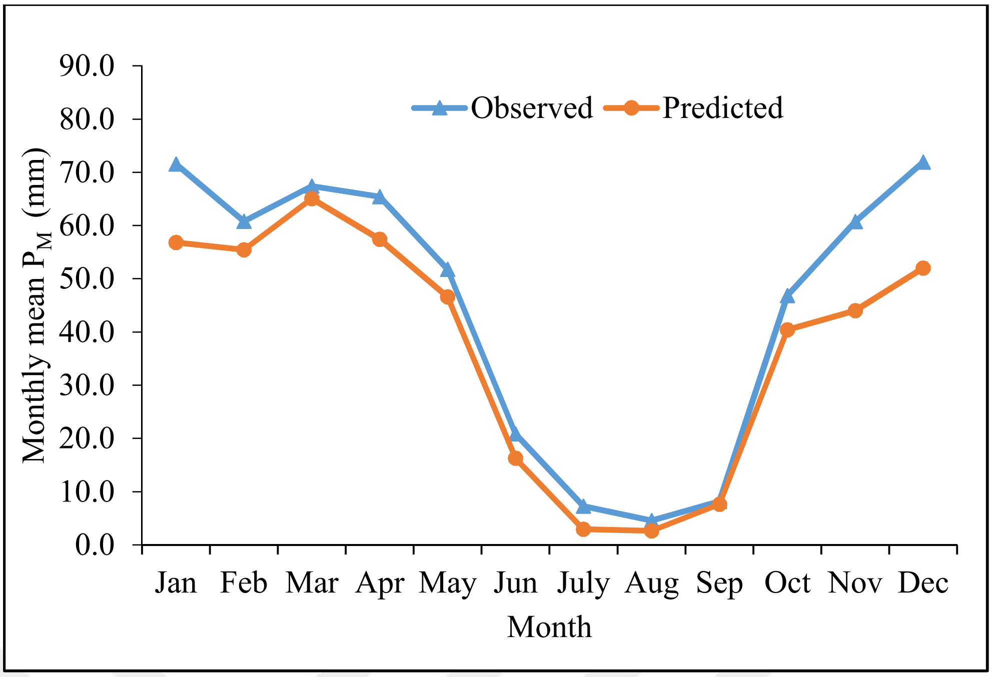

Figure 5.10 Observed and simulated averaged monthly P_M for the calibration period 1971-1990 by SVM model

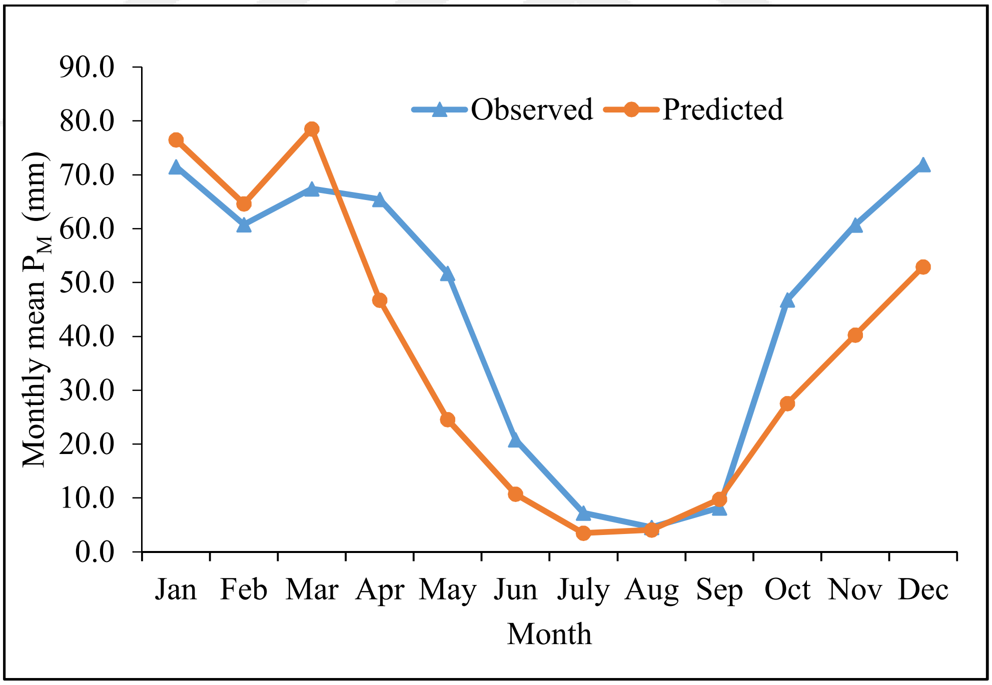

Figure 5.11 Observed and simulated averaged monthly P_M for the calibration period 1971-1990 by Linear Regression model

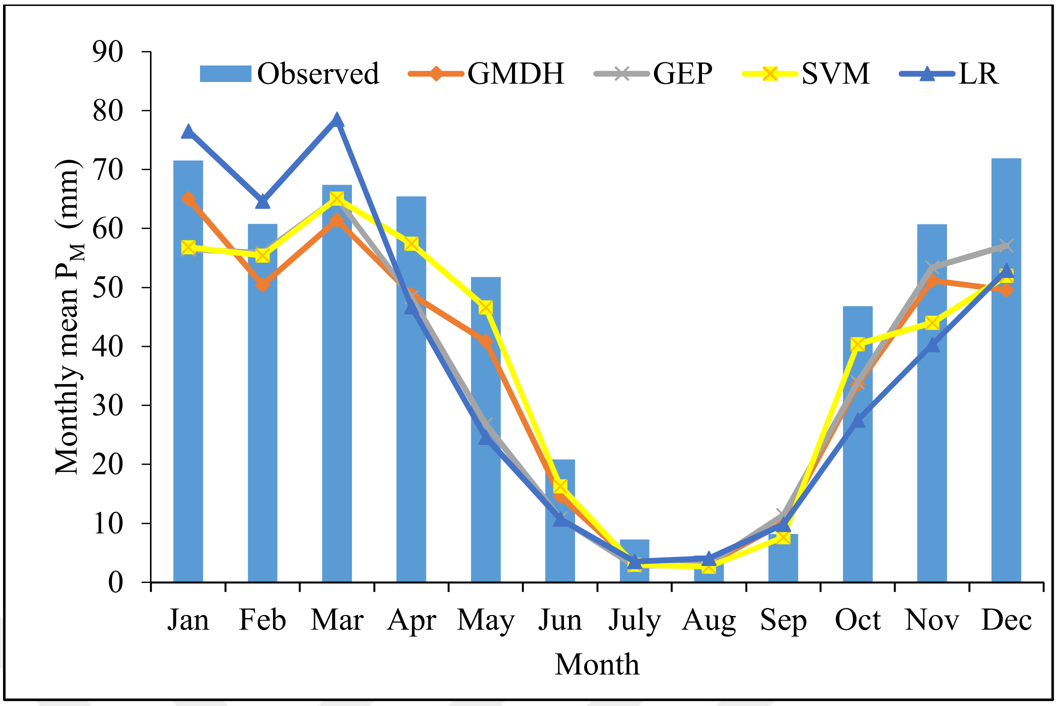

Figure 5.12 Observed and simulated averaged monthly P_M for the calibration period 1971-1990 by all models

5.4 Statistical results of model validation

According to the calibrated performance of proposed models, the SVM model simulated the best results. The validation process is done using the simplified formulation for each model obtained from calibration periods. For the SVM model, model parameters as presented in Table 5.6, are used to generate the validation process. For GEP model, equation 5.1 is used to generate the validation process. For the GMDH model, the equation given in the appendix is used to generate the validation process. The result of GMDH, GEP, SVM, and linear regression models for validations were expressed in correlation factor (R) for each model. A scatter plot was made between the observed (vertical axis) and the predicted P_M (horizontal axis) as presented in Figures 5.13-5.16. In Figures 5.13-5.16, it is seen that all models have an almost uphill pattern when moved from left to right. It means the models were run perfectly and made a good validation according to observed data. The SVM (R=0.788) and GMDH (R=0.782) model have the almost same correlation coefficient (R) and GEP model's R-value (0.771) are very close to these two models. However, the SVM model has the highest R-value (0.788) while the linear regression model has the lowest R-value (0.730).

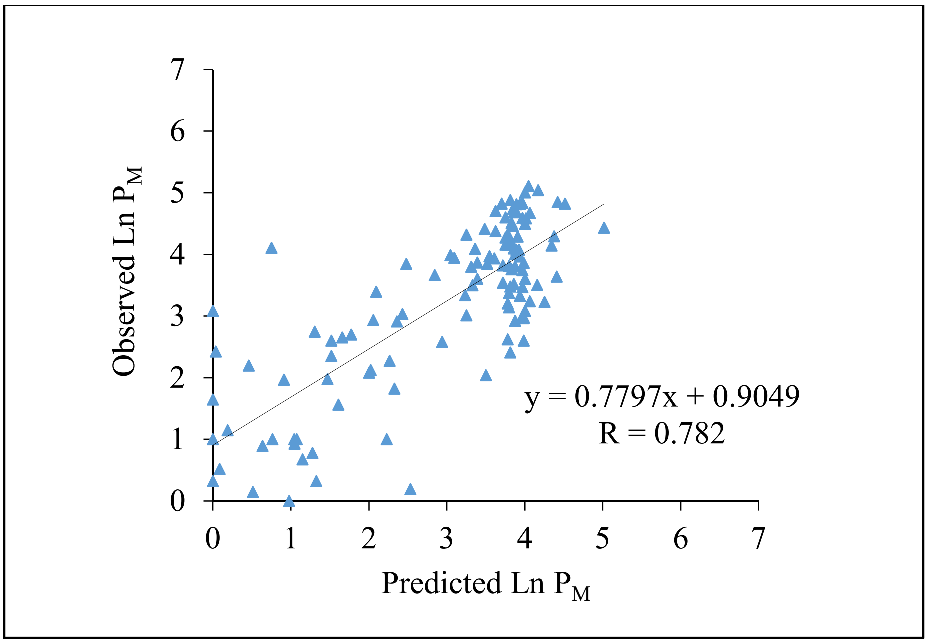

Figure 5.13 Scatter plot of the observed and predicted Ln P_M values for the validation period 1991-2000 by GMDH model

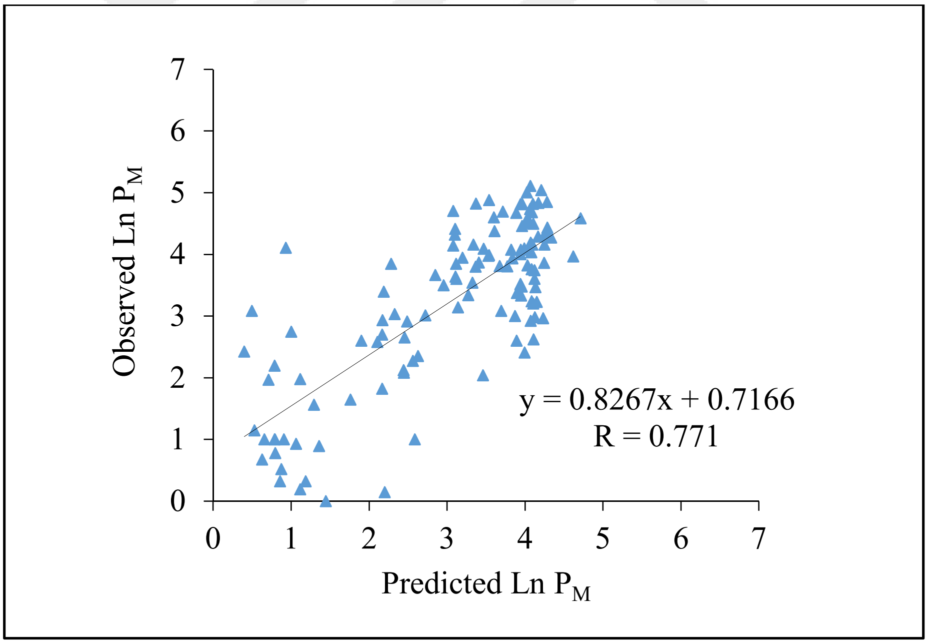

Figure 5.14 Scatter plot of the observed and predicted Ln P_M values for the validation period 1991-2000 by GEP model

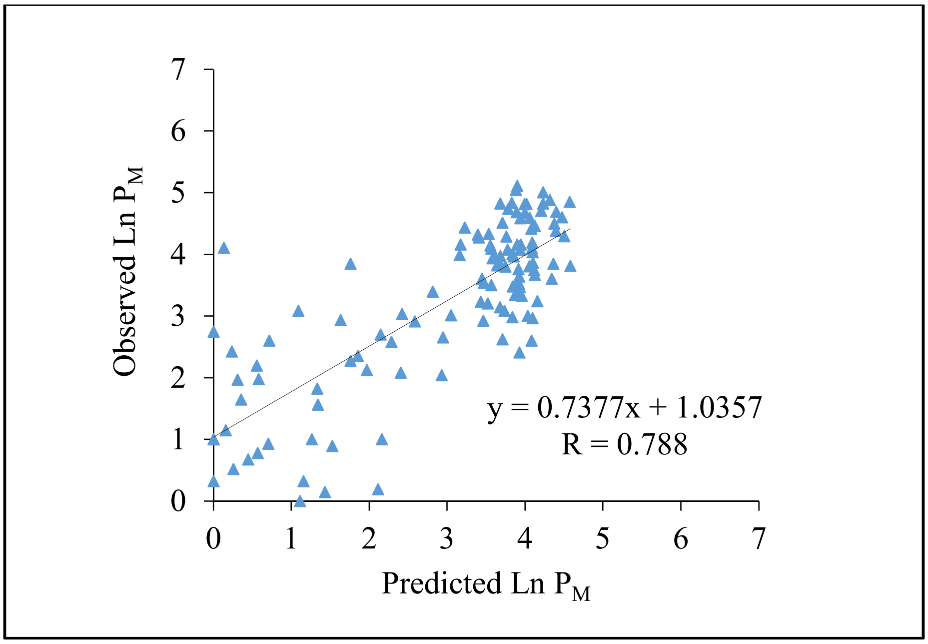

Figure 5.15 Scatter plot of the observed and predicted Ln P_M values for the validation period 1991-2000 by SVM model

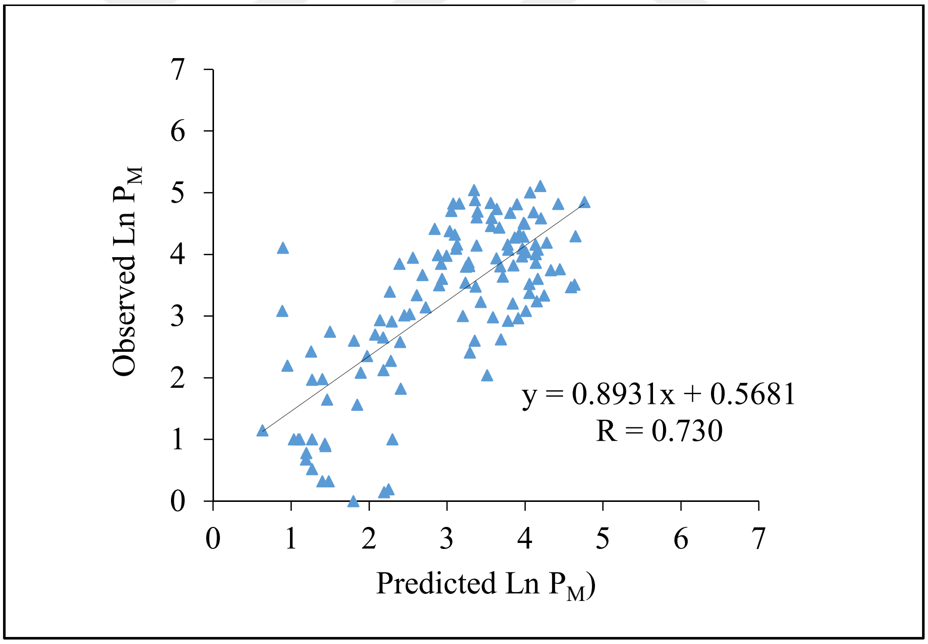

Figure 5.16 Scatter plot of the observed and predicted Ln P_M values for the validation period 1991-2000 by Linear Regression model

The linear equations comprising the relation between the observed and the predicted P_M for the four models (GMDH, GEP, SVM, LR) were obtained to be:

$$y = 0.7797x + 0.9049 \qquad \text{(GMDH model)} \qquad (5.6)$$

$$y = 0.8267x + 0.7166 \qquad \text{(GEP model)} \qquad (5.7)$$

$$y = 0.7377x + 1.0357 \qquad \text{(SVM model)} \qquad (5.8)$$

$$y = 0.8931x + 0.5681 \qquad \text{(LR model)} \qquad (5.9)$$

Where y represents the observed P_M and x represents the predicted P_M.

The outputs value (predicted P_M) of downscaling models can be negative values. Under this situation, those negative values were adjusted to zero. The percentage of negative values over the total number of monthly areal mean precipitations were 3.32 % in the GMDH, 2.49 % in the GEP, 0.83 % in the SVM, and 10 % in linear regression models. Linear regression has the biggest percentage correction of negative values adjusted to zero, while GMDH, GEP, and SVM need less correction. These corrections affect the statistical properties of models and cause the tendency to right of predicted value when the observed value is equal to 0.

Table 5.9 presents the statistical comparison between the simulated P_M for the four models and the observed P_M under the same validation period (1991-2000).

From table 5.9, the simulated mean monthly areal mean precipitation by SVM model, GEP model, GMDH model, and LR model are 10, 10.9, 11.6, and 14.5 mm respectively, lower than the mean of observation. It means the means were underestimated by all models. But, the SVM model yielded better results than the other models to simulate mean. The SVM model and the GMDH model perform well to simulate minimum precipitation while linear regression model performs poorly to simulate minimum precipitation. Moreover, there is a significant difference in the maximum value of simulated P_M compared to observed data. GMDH model record the minimum difference of 9% for maximum precipitation, and the SVM model record minimum difference of 41% for maximum precipitation. Like the mean, all models underestimated the standard deviation, but the SVM model has the minimum underestimating difference of 36% among the other models.

As seen by the values of statistical indicators given in Table 5.9, the SVM model has a lesser MAE (22.26 mm) and RMSE (32.36 mm) than the other proposed models. The MAE is approximately 8%, 8.8% and 19.5% for the GEP model, GMDH model, and linear regression model, respectively, greater than that of the SVM model. The RMSE is approximately 5.5%, 7.5%, and 19.5% for the GEP model, GMDH model, and linear regression model, respectively, greater than that of the SVM model.

Table 5.9 Comparision of statistical results of observed and simulated P_M during the validation (1991-2000)

Data	Mean	Min	Max	Std. Deviation	MAE	RMSE	R
Observed	45.935	0.07	165.47	40.527	-	-	-
GMDH model	34.372	1.00	150.75	25.253	24.23	34.14	0.782
GEP model	35.070	1.49	111.16	25.593	24.05	34.78	0.771
SVM model	35.944	1.00	97.61	25.943	22.26	32.36	0.788
Linear Regression	31.430	1.88	116.90	25.863	26.60	38.67	0.730

Also, the AIC values of each proposed models are calculated and given in Table 5.10. The number of data used in the validation period is 120. It can be seen that AIC increases when the number of the fitting parameter (k) increases. Table 5.8 indicates, the SVM model has the lowest MSE value (1047.3 mm^2), however SVM model has the biggest AIC value (3078) because of having high number of fitting parameter. Consequently, the lowest AIC (860) with k (4) values of the GEP model present its best robustness and the generalization capacity.

Table 5.10 AIC values of proposed models during validation (1971-1990)

Model	MSE (mm^2)	k	AIC
GMDH	1165.7	490	1827
GEP	1209.7	4	860
SVM	1047.3	1122	3078
LR	1495.1	11	899

Statistical results and scatter plots indicate, the SVM model performs well also in the validation period and simulated better results than the other downscaling models, however, AIC values show the GEP model has the lowest AIC value.

Figures 5.17-5.20 show the simulated averaged monthly mean precipitation P_M of GMDH, GEP, SVM, and LR models and observed data during the validation period. Figure 5.21 illustrates the simulated averaged monthly mean precipitation P_M of all models and observed data in one figure. It is clear that the SVM model has slightly underestimated the P_M each month. In January, the linear regression model overestimated the P_M while other models underestimated the P_M as shown in figure 5.21. In February, the linear regression model and the GEP model overestimated the P_M while other models underestimated the P_M. In July and October, the GMDH model overestimated the P_M while other models underestimated the P_M. In March, April, May, Jun, August, September, and October all models underestimated the P_M. In April, May, and Jun the SVM simulated the best averaged mean precipitation compared to observed data. The mean underestimated percentage in April which are 33%, 38%, and 41.8% for the GMDH model, GEP model, and linear regression model, respectively, smaller than that of the SVM model as shown in figure 5.21. In March, September, November, and December the GEP model performs best to predict mean precipitation. The mean underestimated percentage in March which are 7.5%, 10.4%, and 16.2% for the GMDH model, the linear regression model, and SVM model, respectively, smaller than that of the GEP model. In July and October, the GMDH model performs best to predict mean precipitation. The mean underestimated percentage in October which are

3%, 21.4%, and 34.6% for the SVM model, the GEP model, and linear regression model, respectively, smaller than that of the GMDH model. In January, February, July, and August, the linear regression model performs best to predict P_M (almost the same in the calibration period). In February, the SVM and GMDH models simulated the same amount of mean precipitation.

Finally, as shown in Figures 5.17-5.21 and Table 5.8, the SVM model performs well in predicting the variation in monthly areal mean precipitation data when compared to the rest models for the validation and calibration periods. Tables in Appendix C2, D2, E2, and F2 present the numerical results of observed and predicted averaged monthly mean precipitation values during the calibration period. The percent absolute errors are displayed in Figures C2, D2, E2, and F2 during the same calibration period.

Consequently, the SVM model outperformed over the other models in the calibration and validation period with the highest correlation coefficient and least RMSE and MAE values as presented in Table 5.7 and 5.9. Therefore, the SVM model has the ability for predicting the monthly areal mean precipitation P_M under future emission scenarios (A1B and A2) for the period of 2021-2100 years.

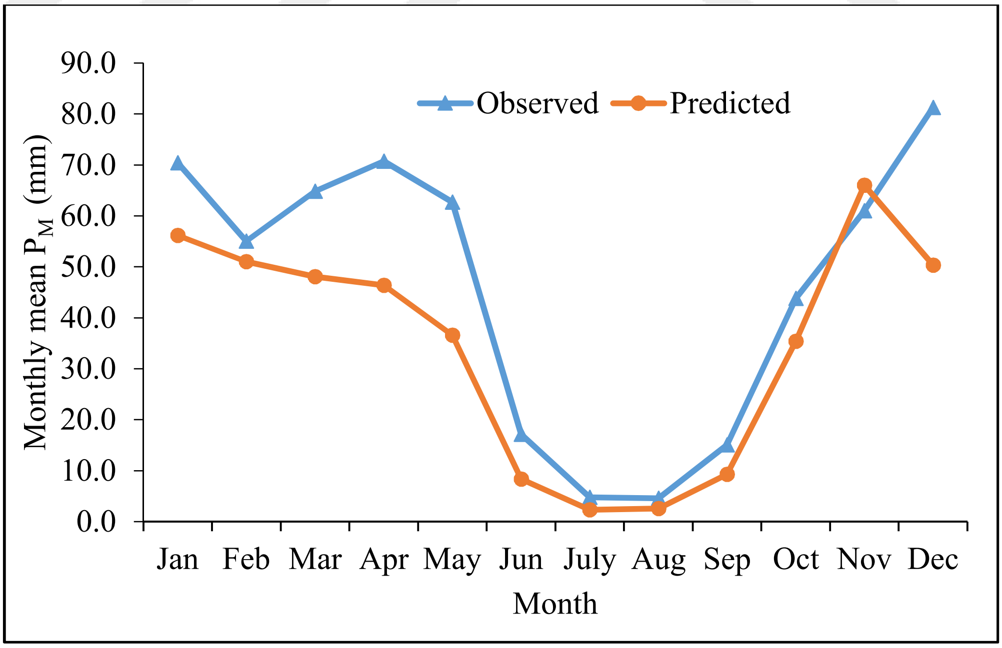

Figure 5.17 Observed and simulated averaged monthly P_M for the validation period 1991-2000 by GMDH model

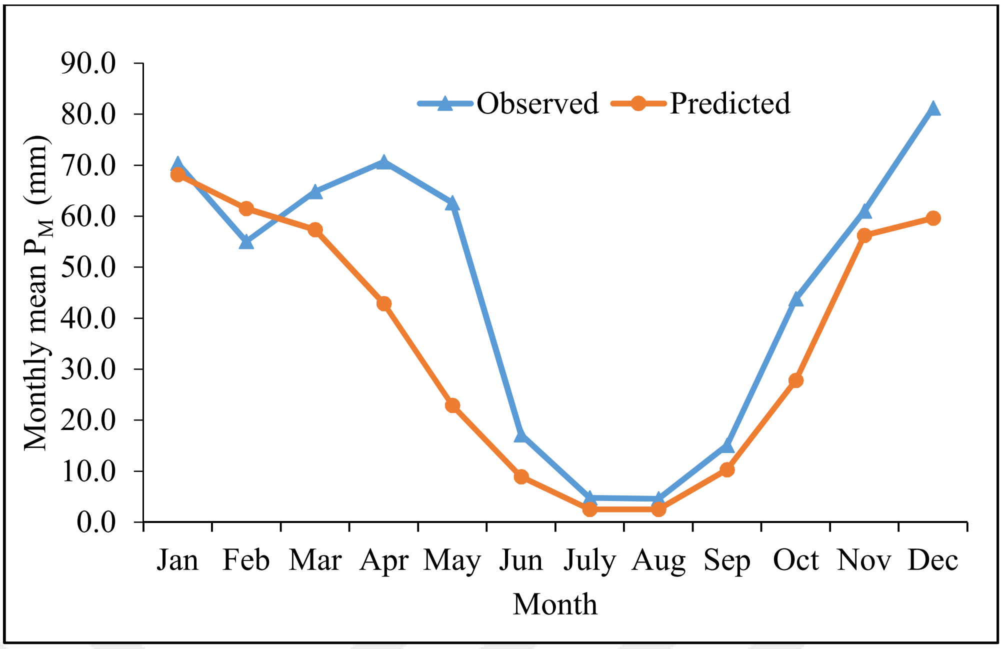

Figure 5.18 Observed and simulated averaged monthly P_M for the validation period 1991-2000 by GEP model

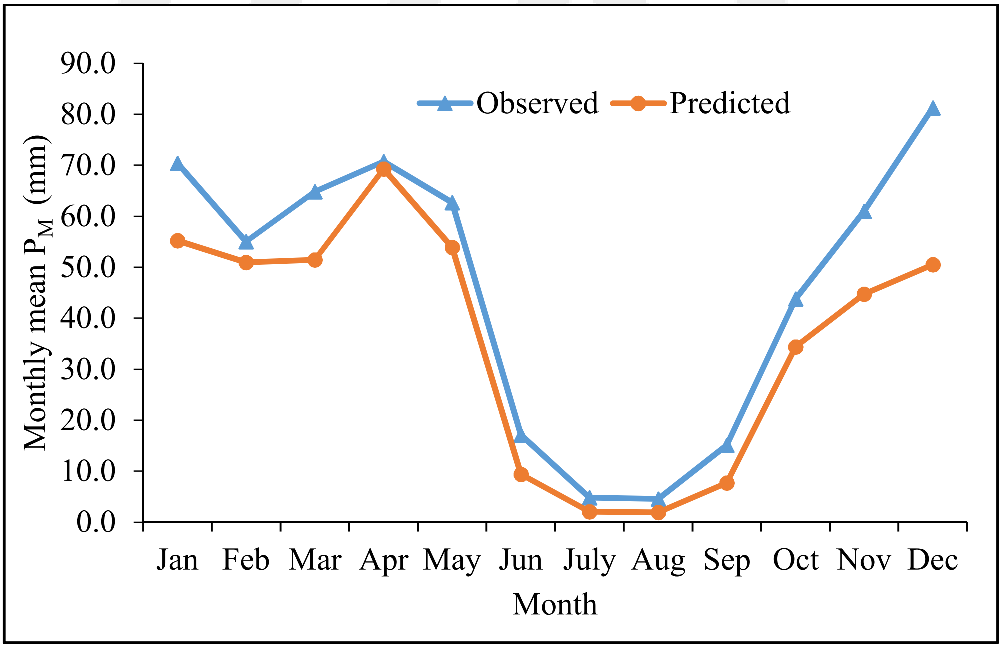

Figure 5.19 Observed and simulated averaged monthly P_M for the validation period 1991-2000 by SVM model

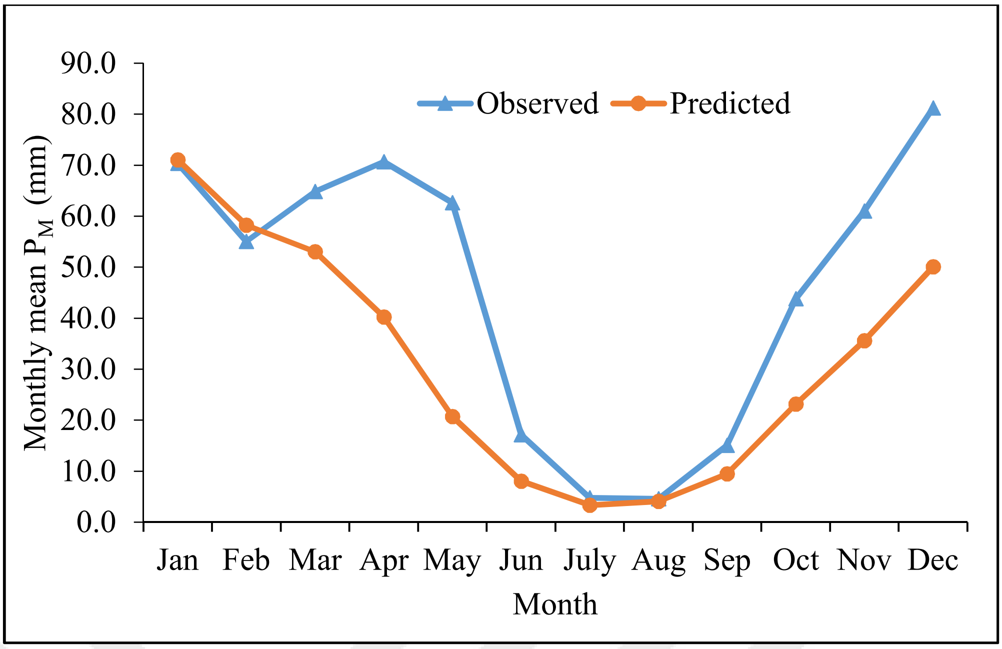

Figure 5.20 Observed and simulated averaged monthly P_M for the validation period 1991-2000 by Linear Regression model

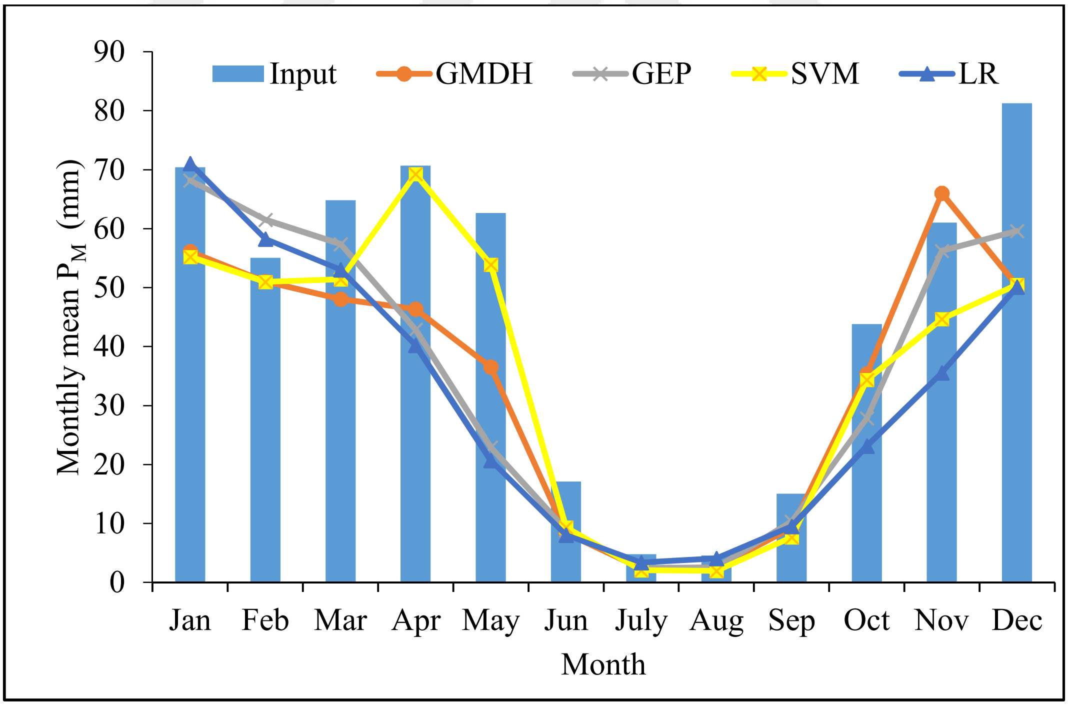

Figure 5.21 Observed and simulated averaged monthly P_M for the validation period 1991-2000 by all models

5.4 Future projection of mean precipitation under different emission scenarios

The downscaled results from both scenarios (A1B and A2) were divided into four time periods with 20 years range, namely: 2020s (2021-2040), 2040s (2041-2060), 2060s (2061-2080), and 2080s (2081-2100), and compared with the baseline period (1971-2000) to examine the future change in P_M for the basin area.

Figure 5.22 displays the projection of averaged monthly P_M change under CGCM3A1B scenario for different time periods by the SVM model. From Figure 5.23, it is clear that the SVM model projected the amount of precipitation for each different time periods will be decreased when compared with baseline mean precipitation. A noticeable decrease occurs in May and it projected a decrease in the averaged monthly mean precipitation P_M of about 18%, 31%, 35.5%, and 48.2% for the 2020s, 2040s, 2060s, and 2080s, respectively, under the A1B scenario. From figure 5.23, it is observed that a constant decrease in P_M occurs during the four periods (2020s, 2040s, 2060s, and 2080s) for the May, Jun, August, and September.

Figure 5.23 shows the projection of averaged monthly PM change under CGCM3A2 scenario for different time periods by the SVM model. From Figure 5.23, it is clear that the SVM model projected the amount of precipitation for each different time periods will be decreased when compared with baseline mean precipitation. However, in April the projected mean precipitation is so close to the observed baseline mean precipitation. A noticeable decrease occurs in May and it projected a decrease in the averaged monthly mean precipitation P_M of about 27.3%, 51.3%, 37.3%, and 59.1% for the 2020s, 2040s, 2060s, and 2080s, respectively, under the A2 scenario. From figure 5.23, it is observed that a constant decrease in P_M does not occur during the four periods (2020s, 2040s, 2060s, and 2080s).

The changes in max precipitation predicted by both scenarios are different from each other in magnitude, but almost similar in their patterns as seen in Figure 5.22 and 5.23. Both scenarios show an average annual decrement with respect to the baseline period (1971-2000) in these figures. Under the A1B scenario, the SVM model predicts a decrease in the averaged annual P_M by 24.7%, 28%, 30.7, and 32.5% for the 2020s, 2040s, 2060s, and 2080s, respectively. It also projects a decrease in the averaged annual P_M of about 27.5%, 32.3%, 33.1, and 31.1% for the 2020s, 2040s, 2060s, and

2080s, respectively, under the A2 scenario.

Figure 5.24 and 5.25 represent the projected annual mean precipitation P_M under A1B and A2 scenarios for different time periods by proposed downscaled models. From figure 5.24, the GMDH model projected the lowest annual precipitation, and the SVM model projected the highest. Hollow symbols in Figure 5.24 and 5.25 represent the mean annual precipitation downscaled by the models in the baseline period 1971-2000. From figure 5.25, the GMDH model projected the highest annual precipitation, and the SVM projected the lowest from 2021-2049 to 2055-2084 periods while the SVM model projected the highest annual precipitation, and the GMDH projected the lowest from 2055-2084 to 2070-2099 periods.

The monthly and yearly future P_M numerical results are presented in Appendix G.

Future projection of averaged monthly P_M under climate change scenarios by GMDH and GEP models presented in Appendix H and I.

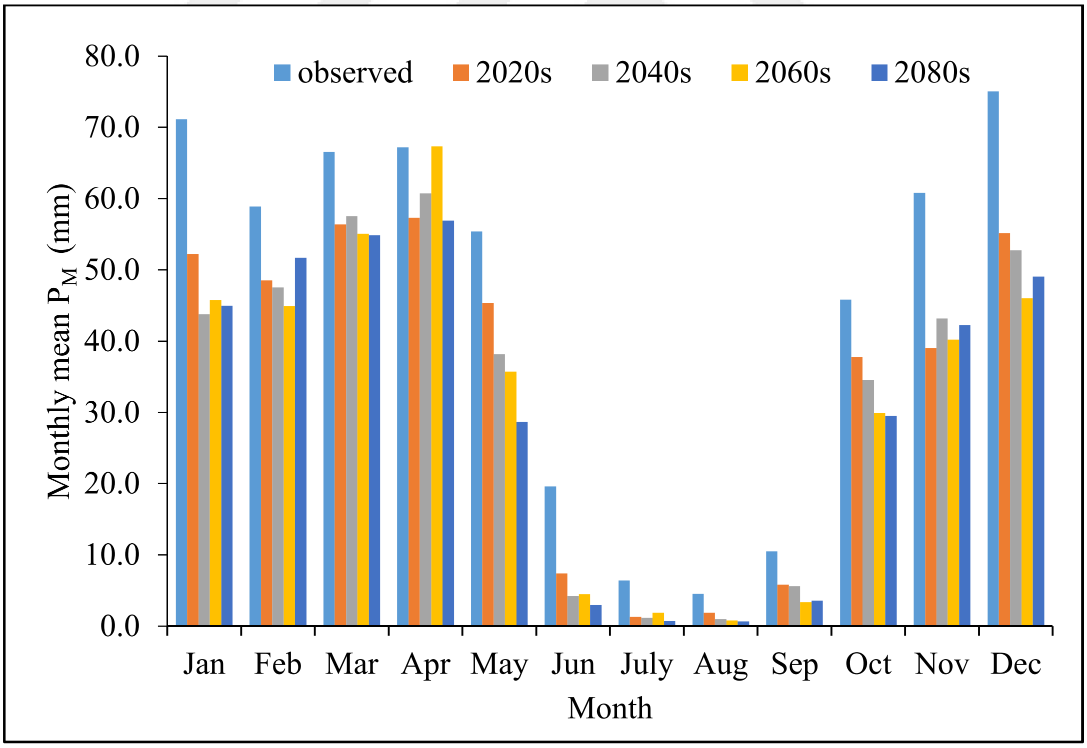

Figure 5.22 The projection of averaged monthly P_M change under CGCM3A1B scenario for different time periods by the SVM model

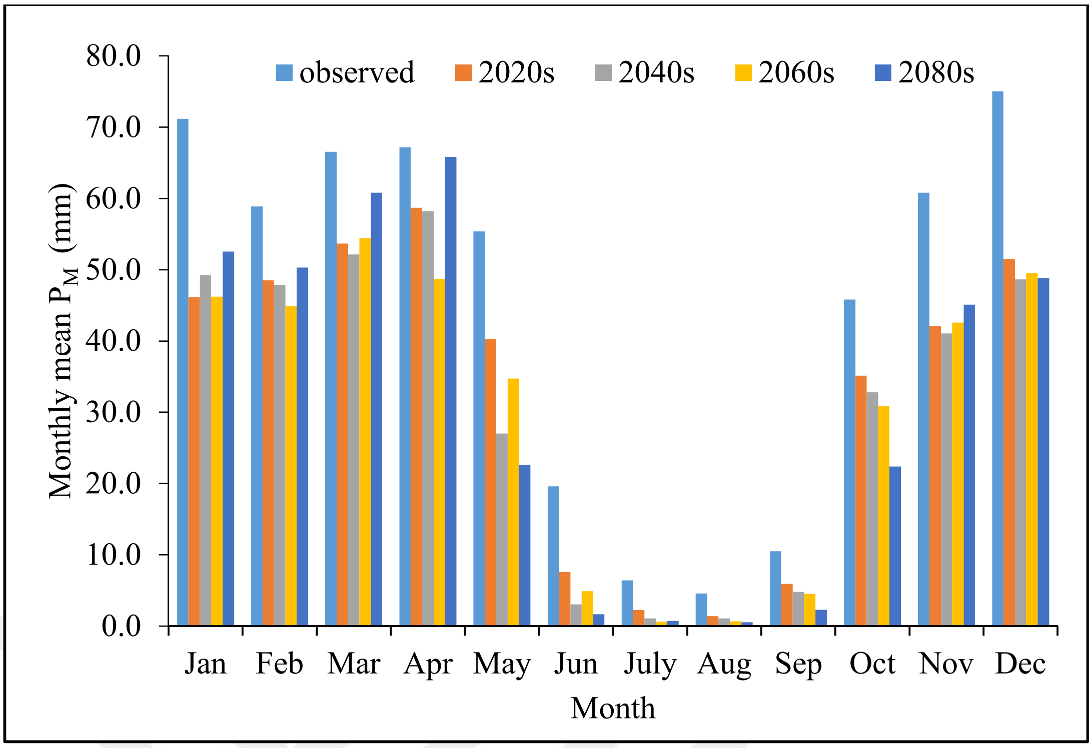

Figure 5.23 The projection of averaged monthly P_M change under CGCM3A2 scenario for different time periods by the SVM model

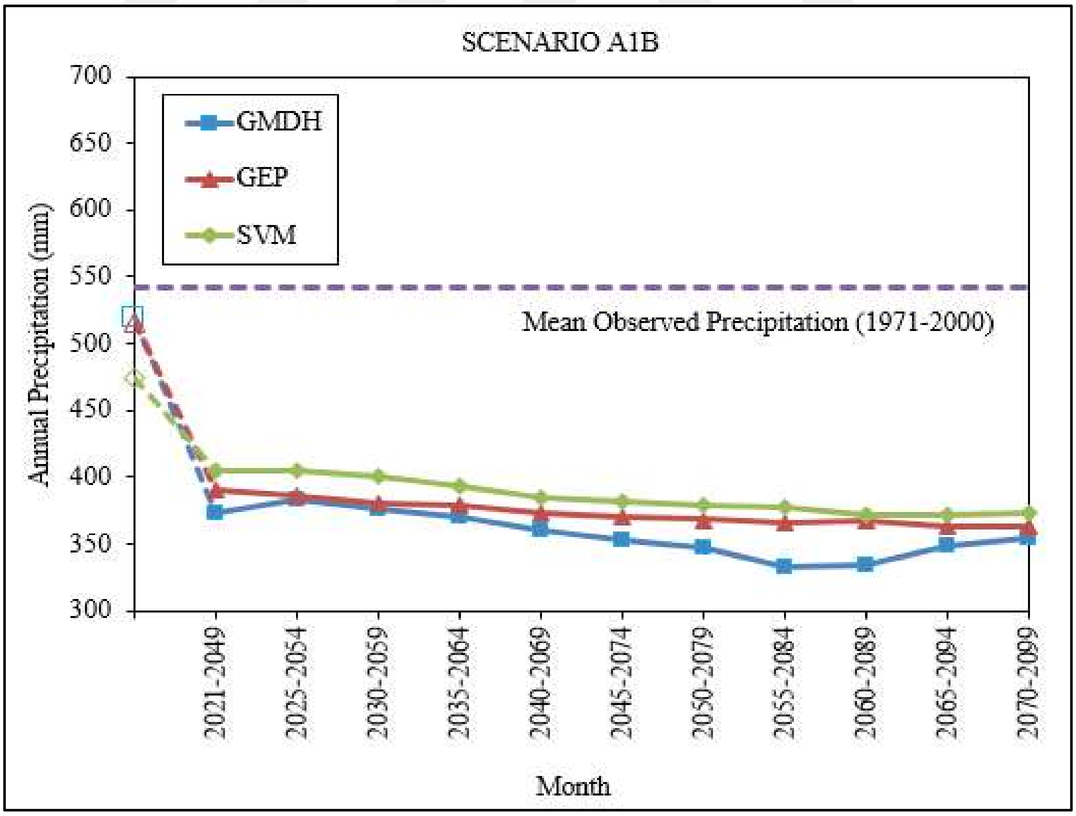

Figure 5.24 The projection of annual P_M change under CGCM3A1B scenario for different time periods by the SVM, GMDH, and GEP models

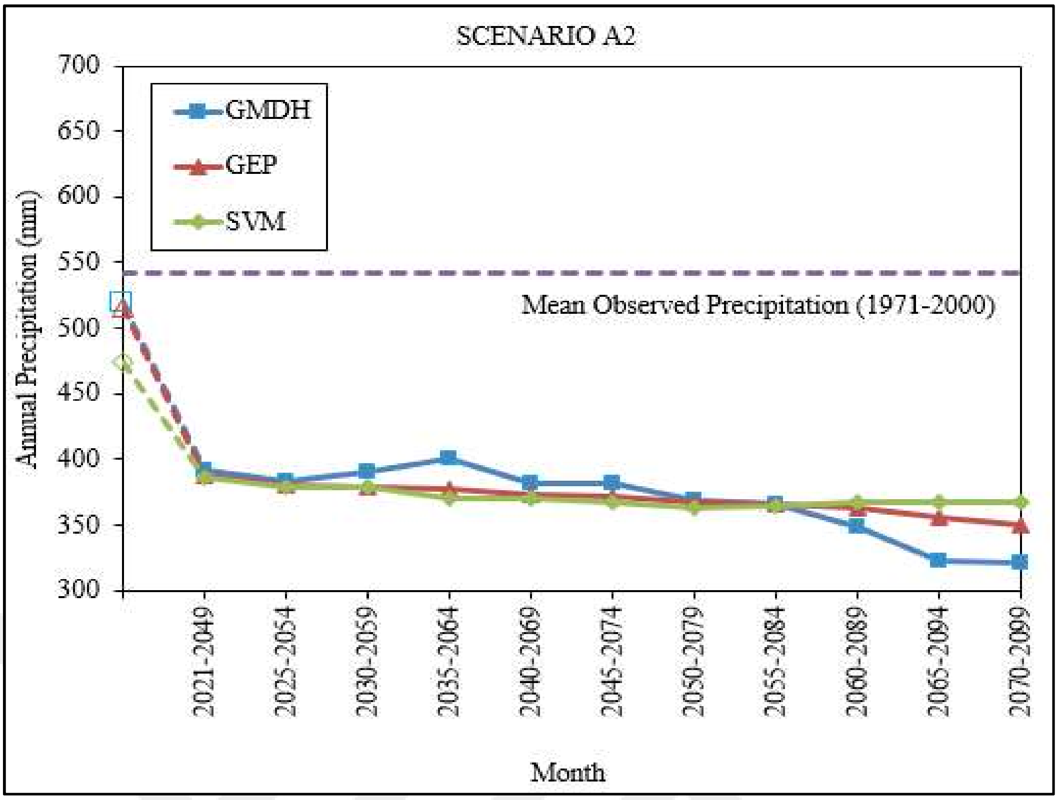

Figure 5.25 The projection of annual P_M change under CGCM3A2 scenario for different time periods by the SVM, GMDH, and GEP models

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

Climate change and its impact, increasing concentration of 'greenhouse gases' in the atmosphere causes to climate changes, have gained serious consideration in hydrology. This thesis examines and explores the effect of climate change on Goksun River basin. Watershed Modeling System (WMS) was used to delineate the Goksun River basin with three gauge stations. By using Thiessen areal mean method, mean basin precipitation was calculated as the output of downscaling models.

In this thesis, the statistical downscaling and the General Circulation Model (GCM) are widely utilized to estimate the climate change effects on the basin. GCM outputs cannot generate local climate details in the finer spatial resolution because of the uncertainty in the spatial resolution between the GCM and hydrological models. Subsequently, downscaling methods have been employed to transform the GCM outputs from coarse spatial resolution to a finer spatial resolution that can be used directly for forecasting the climate change at the local scale output. General Circulation Model has explained and for the use of downscaling methods, CGCM3 model was chosen to get future projection at the local scale.

Location of the centroid of the basin was used to download large-scale weather factors (26 variables) as input from CGCM3 grid boxes. To select the most effective predictors and to see the linear relationship between inputs (large-scale variables) and outputs (monthly areal mean precipitation), correlation analysis was employed by Pearson rank correlation coefficient method. After correlation analysis is done, 26 input variables were decreased to 10 input variables. Also, to improve the correlation with predictor variables and predictands, variables were modified based on some normalization methods. Taking the natural logarithm of each variable normalization method gave the best correlation.

Group Method of Data Handling (GMDH), Gene Expression Programming (GEP), and Support Vector Machine (SVM) downscaling techniques were utilized to estimate the downscaled monthly areal mean precipitation on the basin. Furthermore, the prediction results of downscaling techniques were found and compared with each other to examine the best model performance. Finally, Linear Regression was analyzed for comparison. The downscaling result of the GCM scale projected under both GCM A2 and A1B future emission scenarios for P_M in Goksun River basin has been investigated.

Results based on statistical indicator (model evaluation criteria) between observed and downscaled data (predicted) show the SVM model can estimate P_M better than other models during calibration (1971-1990) and validation (1991-2000) periods. The correlation coefficient (R) of the SVM produced for the calibration period is 0.863 and for the validation period, R is 0.788. Consequently, the SVM model was selected for predicting the monthly areal mean precipitation P_M under future emission scenarios (A1B and A2) for the period of 2021-2100 years.

The projected results for both scenarios indicate closeness in their trend; however, they are different in P_M amount. Generally, both scenarios forecast a decrease in the average annual P_M during the current decade. The downscaled results from both scenarios (A1B and A2) were divided into four time periods with 20 years range, namely: 2020s (2021-2040), 2040s (2041-2060), 2060s (2061-2080), and 2080s (2081-2100), and compared with the baseline period (1971-2000) to estimate the future change in P_M for the basin area.

A noticeable decrease occurs in May and it projected a decrease in the averaged monthly mean precipitation P_M of about 18%, 31%, 35.5%, and 48.2% for the 2020s, 2040s, 2060s, and 2080s, respectively, under the A1B scenario. It is observed that a constant decrease in P_M occurs during the four periods (2020s, 2040s, 2060s, and 2080s) for the May, Jun, August, and September. However for the A2 scenario, in April the projected mean precipitation is so close to the observed baseline mean precipitation. A noticeable decrease occurs in May and it projected a decrease in the average monthly mean precipitation P_M of about 27.3%, 51.3%, 37.3%, and 59.1% for the 2020s, 2040s, 2060s, and 2080s, respectively, under the A2 scenario. It is observed that a constant decrease in P_M does not occur during the four periods (2020s, 2040s, 2060s, and 2080s) for the May, Jun, August, and September.

The changes in max precipitation predicted by both scenarios are different from each other in magnitude, but almost similar in their patterns. Both scenarios display an average annual decrement with respect to the baseline period (1971-2000). Under the A1B scenario, the SVM model predicts a decrease in the averaged annual P_M by 24.7%, 28%, 30.7, and 32.5% for the 2020s, 2040s, 2060s, and 2080s, respectively. It also projects a decrease in the averaged annual P_M of about 27.5%, 32.3%, 33.1, and 31.1% for the 2020s, 2040s, 2060s, and 2080s, respectively, under the A2 scenario

6.2 Recommendations

1. During this research, only CGCM3 predictors were used to simulate mean precipitation. It is suggested to use other GCM model types such as HadCM3 and NCEP variables. Results can be compared between different types of GCM models.

2. Both observed and downscaled (predicted data) can be used to evaluate the effect of climate change on hydropower potential to assist the engineer while designing the hydraulic structures.

3. Monthly areal mean precipitation may be divided into two division, dry-month and wet-month according to the amount of precipitation in those months. This could affect the downscaling model performance and could get a better result.

4. For the predictand (output), observed precipitation data were used in this thesis. However, any other variables can be used for the future study, for example, temperature, humidity, and solar radiation, on the side of precipitation, to estimate the climate change effects on other fields.

REFERENCES

Aggarwal, S. K., Goel, A., & Singh, V. P. (2012). Stage and Discharge Forecasting by SVM and ANN Techniques, *Water resources management,* **26 (13),** 3705–3724.

Ahmadzadeh Araji, H., Wayayok, A., Massah Bavani, A., Amiri, E., Abdullah, A. F., Daneshian, J., & Teh, C. B. S. (2018). Impacts of climate change on soybean production under different treatments of field experiments considering the uncertainty of general circulation models. *Agricultural Water Management*, ***205***, 63–71.

Alam, J., & Danish, M. (2017). Flood frequency analysis using gene expression, *Symposium on Hydrology*, 22-23 December 2015, *New Delhi, India.*

Ali, M., Hojat, G., Zadeh, A., & Isazadeh, M. (2016). A comparative study of artificial neural network (MLP , RBF) and support vector machine models for river flow prediction. *Environmental Earth Sciences*, ***75*(6)**, 1–14.

Anandhi, A., & Bentley, C. (2018). Predicted 21st century climate variability in southeastern U.S. using downscaled CMIP5 and meta-analysis. *Catena*, **170**, 409–420.

Andersen, H. E., Brian, K., Søren E. L., Carl, C. H., Torben S. J., Erik K. R. (2006). Climate change impacts on hydrology and nutrients in a Danish lowland river basin. *Science of the Total Environment*, **365**, 223–237.

Arnell, N. W. (1999). Climate change and global water resources, *Global environmental change*, **9**, S31-S49.

Azamathulla, H., M., Ghani, A., A., Leow C., S., Chang C., K., Zakaria N., A. (2011). Gene- Expression Programming for the Development of a Stage Discharge Curve of the Pahang River, *Water Resour. Manage*, **25**, 2901-2916.

Behzad, M., Asghari, K., & Jr, E. A. C. (2010). Comparative Study of SVMs and

ANNs in Aquifer Water Level Prediction, *Journal of Computing in Civil Engineering*, **24(5),** 408–413.

Bernstein L, Bosch P, Canziani O, Chen Z, Christ R, & Riahi K (2008). IPCC, 2007: Climate Change 2007: Synthesis Report.

Bevilacqua, V., Intini, F., & Kühtz, S. (2008). A model of Artificial Neural Network for the analysis of climate change. *Proceeding of the 29th Annual International Sympsium on Forecasting*, **89**.

Brigham Yong University. (1997). Watershed Modeling System 7.1 user manual, Engineering Computer Graphics Laboratory.

Buonomo E., Jones R., Huntingford C. and Hannaford J. (2007). On the robustness of changes in extreme precipitation over Europe from two high resolution climate change simulations, Quarterly Journal of the Royal Meteorological Society, **133**, 65-81.

Carpenter, T. M., & Georgakakos, K. P. (2006). Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales. *Journal of Hydrology*, ***329*(1–2)**, 174–185.

Carter, T., Hulme, M., Lal, M. (2007). General guidelines on the use of scenario data for climate impact and adaptation assessment, *Journal of Hydrology*, **355**, 142-170.

Chen, D., Achberger, C., Raisanen, J., Hellstrom, C. (2006). Using Statistical Downscaling to Quantify the GCM-Related Uncertinaty in Regional Climate Change Scenarios: A case Study of Swedish Precipitation. *Advances in Atmospheric Sciences*, **23(1)**, 54-60.

Chen, S-T., Yu, Pao-Shan, Tang, Yi-Hsuan (2010). Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *Journal of Hydrology,* **385**, 13–22.

Christensen, J. H., and O. B. Christensen, (2007): A summary of the PRUDENCE model projections of changes in European climate by the end of this century. *Climatic Change,* **81,** 7–30.

Christensen, N. S., Wood, A. W., Voisin, N, Lettenmaier, D. P., Palmer, R. N. (2004).

The Effects of Climate Change on the Hydrology and Water Resources of the Colorado River Basin. *Climatic Change,* **62,** 337–363.

Christensen, Clayton M., Michael B. Horn, and Curtis W. Johnson. (2008). Disrupting Class: How Disruptive Innovation Will Change the Way the World Learns. *McGraw-Hill.*

Cimen, M. (2010). Estimation of daily suspended sediments using support vector machines Estimation of daily suspended sediments using support vector machines, *Hydrological Sciences Journal*, **53(3)**, 656-666.

Dawson, C., Wilby, R. (2007). Statistical Downscaling Model SDSM, version 4.1, *Department of Geography, Lancaster University, UK.*

Demircan, M., Demir, Ö., Atay, H. Eskioğlu, O., Tüvan A., Akçakaya, A. (2014). Climate change projections for Turkey with new scenarios. *The Climate Change and Climate Dynamics Conference*, 2014–CCCD2014.

Dibike, Y. B.,P. Coulibaly. (2005). Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models, *Journal of Hydrology*, **307**, 145-163.

Durman, C. F., Gregory, J. M., Hassell, D. C., Jones, R. G., & Murphy, J. M. (2001). A comparison of extreme European daily precipitation simulated by a global and a regional model for present and future climates. *Quarterly Journal of the Royal Meteorological Society*, **127(573**), 1005–1015.

Ebtehaj, I., Bonakdari, H., Hossein, A., & Azimi, H. (2015). Engineering Science and Technology , an International Journal GMDH-type neural network approach for modeling the discharge coef fi cient of rectangular sharp-crested side weirs. *Engineering Science and Technology, an International Journal*, **18(4)**, 746–757.

EPA (Environmental Protection Agency), (2014). *Climate Change: Basic Information.* 18 March. Accessed: September 05, 2014.

Farlow, S. J. (1981). The GMDH algorithm of Ivakhnenko. *The American Statistican,* **35 (4),** 210-215.

Fernando, A., Shamseldin, A., Abrahart, R. (2012). Use of gene expression programming for multimodel combination of rainfall–runoff models. *Journal of Hydrologic Engineering*, **17(9)**, 975–985.

Ferreira, C. (2001). Gene expression programming: a new adaptive algorithm for solving problems, *Complex Syst*, **13(2)**, 7–129.

Fowler, H., Blenkinsop, S., Tebaldi, C. (2007). Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, *International Journal of Climatology*, **27**, 1547-1578.

Fujihara, Y., Tanaka, K., Watanabe, T., Nagano, T., Kojiri, T. (2008). Assessing the impacts of climate change on the water resources of the Seyhan River Basin in Turkey: Use of dynamically downscaled data for hydrologic simulations, *Journal of Hydrology*, **353**, 33-48.

Garg, V. (2013). Inductive Group Method of Data Handling Neural Network Approach to Model Basin Sediment Yield, *Journal of Hydrologic Engineering,* **20 (6)**, C6014002-1 - C6014002-8.

Gholami, A., Bonakdari, H., Ebtehaj, I., Shaghaghi, S., & Khoshbin, F. (2017). Developing an expert group method of data handling system for predicting the geometry of a stable channel with a gravel bed, **1471**, 1460–1471.

Ghosh, S., and P. P. Mujumdar (2008), Statistical downscaling of GCM simulations to streamflow using relevance vector machine, Adv. Water Resour., **31(1),** 132– 146,

Goyal, M. K., Sharma, A., & Katsifarakis, K. L. (2017). Prediction of flow rate of karstic springs using support vector machines. *Hydrological Sciences Journal*, **62(13),** 2175-2186.

Giorgi, F., Hewitson, B., Christensen, J., Hulme, M., Von Storch, H., Whetton, P., Jones, R., Mearns, L., Fu, C., Arritt, R. (2001). Emerging patterns of simulated regional climatic changes for the 21st century due to anthropogenic forcings. *Geophys. Res. Lett.*, **28(17),** 3317–3320.

Guo, J., Zhou, J., Qin, H., Zou, Q., & Li, Q. (2011). Expert Systems with Applications

Monthly streamflow forecasting based on improved support vector machine model. *Expert Systems With Applications*, ***38*****(10),** 13073–13081.

Guven, A. (2009). Linear genetic programming for time-series modeling of Daily flow rate, *J. Earth Syst*, **118(2)**, 137-146.

Guven, A., Talu, N., E. (2010). Gene Expression Programming for Estimating Suspended Sediment Yield in Middle Euphrates Basin, Turkey. *Clean-Soil, Air, Water*, **38(12)**, 1159-1168.

Guven, A., Aytek, A. (2009). New approach for stage discharge relationship: gene expression programming, *Journal of Hydrologic Engineering*, **14**, 812-820.

H.Akaike 1974, A new look at the statistical model identification I*EEE Transactions on Automatic Control*, **19 (6),** 716-723.

Hashmi, M., Shamseldin, A., Melville, B., (2011). Statistical downscaling of watershed precipitation using Gene Expression programming, *Environmental Modeling & Software*, **26**, 1639-1646.

Hua, C. (2010). Downscaling GCMs Using the Smooth Support Vector Machine Method to Predict Daily Precipitation in the Hanjiang Basin, **27(2),** 274–284.

Huth, R. (1999). Statistical downscaling in central Europe: evaluation of methods and potential predictors, *Climate Research*, **13**, 91-101.

IPCC (2000). IPCC special report: emission scenarios. *Cambridge University Press*, Cambridge, United Kingdom and New York, NY, USA.

IPCC, (2001): Climate Change 2001: Synthesis Report. A Contribution of Working Groups I, II, and III to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Watson, R.T. and the Core Writing Team (eds.)]. Cambridge University Press, Cambridge, United Kingdom, and New York, NY, USA, 398 pp.

IPCC, (2007). Climate change 2007-impacts, adaptation and vulnerability: Working group II contribution to the fourth assessment report of the IPCC: Cambridge University Press.

IPCC, (2013). Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, Kiehl, J. T., Trenberth, K. E. (1997). Earth's annual global mean energy budget, *Bulletin of the American Meteorological Society*, **78**, 197-208.

Ivakhnenko, A.G., (1971). Polynomial theory of complex systems. *IEEE Trans. Syst., Man, Cybern.* **1 (4),** 364–378.

Jones, P. D., Jonsson, T., and Wheeler, D. (1997). Extension to the North Atlantic Oscillation using early instrumental pressure observations from gibraltar and south-west iceland. International Journal of Climatology, **17(13),** 1433–1450.

Kiehl, J. T., Trenberth, K. E. (1997). Earth's annual global mean energy budget, *Bulletin of the American Meteorological Society*, **78**, 197-208.

Kim, K., (2003). Financial time series forecasting using support vectormachines. *Neurocomputing*, **55**, 307–319.

Kisi, O. (2015). Streamflow Forecasting and Estimation Using Least Square Support Vector Regression and Adaptive Neuro-Fuzzy Embedded Fuzzy c-means Clustering, *Water Resources Management*, **29(14),** 5109–5127.

Koning, J. De, Bixio, D., Karabelas, A., Salgot, M., & Schäfer, A. (2008). Characterisation and assessment of water treatment technologies for reuse, **218,** 92–104.

Koza, J. R. (1992). *Genetic Programming: On the Programming of Computers by Means of* Natural Selection. Cambridge, MA: The MIT Press.

Kumar, M., & Ojha, G. C. S. P. (2011). Estimation of Scour Downstream of a Ski-Jump Bucket Using Support Vector and M5 Model Tree, *Water Resources Management,* **25,** 2177–2195.

Kusangaya, S., Warburton, M., van Garderen, E.A., and Jewitt, G.P.W. (2014). Impacts of climate change on water resources in southern Africa: A review. *Physics and Chemistry of the Earth, Parts A/B/C,* **67-69**: 47-54.

Lespinas, F., Ludwig, W., and Heussner, S. (2014). Hydrological and climatic uncertainties associated with modeling the impact of climate change on water resources of small Mediterranean coastal rivers. *Journal of Hydrology,* **511**, 403-422.

Lin, J., Cheng, C., Chau, K., Lin, J., Cheng, C., & Chau, K. (2010). Using support vector machines for long-term discharge prediction, *Hydrological Sciences Journal*, **51(4),** 599-612.

Liu, J., Chang, J., & Zhang, W. (2009). Groundwater Level Dynamic Prediction Based on Chaos Optimization and Support Vector Machine. *2009 Third International Conference on Genetic and Evolutionary Computing*, October 14-17, 39–43.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Venema, V. K. C., Chun, K. P., Goodess, C. M. , Jones, R. G., Onof, C., Vrac, M., Thiele-Eich, I. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.,* **48 (3)**, 1–38.

McAvaney, B., Covey, C., Joussaume, S., Kattsov, V., Kitoh, A., Ogana, W., Pitman, A., Weaver, A., Wood, R., Zhao, Z.-C. (2001). Model evaluation. Climate Change 2001: The scientific basis. Contribution of WG1 to the Third Assessment Report of the IPCC (TAR): Cambridge University, 471-523.

Mohammed, S., Faridah, O., & Ahmed, E. (2017). Prediction of Stream Flow in Humid Tropical Rivers by Support Vector Machines, *MATEC Web of Conferences,* ***111,*** 01007.

Moosavi, V., Talebi, A., & Hadian, M. R. (2017). Development of a Hybrid Wavelet Packet- Group Method of Data Handling (WPGMDH) Model for Runoff Forecasting. *Water Resources Management*, **31(1),** 43–59.

N, S. R., & Deka, P. C. (2014). Support vector machine applications in the field of hydrology : A review. *Applied Soft Computing Journal*, **19**, 372–386.

NASA. (National Aeronautics and Space Administration) (2014). *What Is Climate Change?* May 14. Accessed: September 05, 2014.

Najafzadeh, M., & Azamathulla, H. M. (2013). Group method of data handling to predict scour depth around bridge piers, *Neural Computing and Applications,* **23(7-8),** 2107–2112.

Najafzadeh, M., & Barani, G. (2011). Sharif University of Technology Comparison of group method of data handling based genetic programming and back propagation systems to predict scour depth around bridge piers. *Scientia Iranica*, **18(6),** 1207–1213.

Najafzadeh, M., Barani, G. Kermani, MR. (2013). Abutment scour in clear-water and live-bed conditions by GMDH network, *Water Science & Technology*, **67(5),** 1121–1129.

Najafzadeh, M., & Zahiri, A. (2014). Neuro-Fuzzy GMDH-Based Evolutionary Algorithms to Predict Flow Discharge in Straight Compound Channels, *Journal of Hydrologic Engineering,* **20 (12),** 04015035.

Nakicenovic, N., Alcamo, J., Davis, G., de Vries, B., Fenhann, J., Gaffin, S., Gregory, K., Grübler, A., Jung, T.Y., Kram, T., La Rovere, E.L., Michaelis, L., Mori, S., Morita, T., Pepper, W., Pitcher, H., Price, L., Raihi, K., Roehrl, A., Rogner, H.-H., Sankovski, A., Schlesinger, M., Shukla, P., Smith, S., Swart, R., van Rooijen, S., Victor, N., Dadi, Z. (2000). Emissions Scenarios. A Special Report of Working Group III of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK and New York, NY, USA, 599p.

Nguyen, V. (2002). A state-of-the-art review of downscaling methods for evaluating climate change impacts on the hydrologic cycle, *Water resources management and engineering series. Research Report. No. WRME03/1, McGill University, Montreal, Quebec, Canada*.

Okkan, U., Inan, G. (2015). Statistical downscaling of monthly reservoir inflows for Kemer watershed in Turkey: use of machine learning methods ,multiple GCMs and emission scenarios, *International Journal of Climatology* , **35**, 3274-3295.

Okkan, U. (2015). Assessing the effects of climate change on monthly precipitation proposing of a downscaling strategy through a case study in Turkey, *KSCE Journal of Civil Engineering*, **19**, 1150-1156.

Pacala, S. and Socolow, R. (2004) Stabilization Wedges: Solving the Climate Problem for the Next 50 Years with Current Technologies. Science **305**, 968-972.

Parsaie, A., Azamathulla, H. M., & Haghiabi, A. H. (2018). Prediction of discharge coefficient of cylindrical weir – gate using GMDH-PSO. *ISH Journal of Hydraulic Engineering*, **24(2),** 116-123.

Prudhomme, C., Reynard, N., Crooks, S. (2002). Downscaling of global climate models for flood frequency analysis, *Hydrological Processes*, **16**, 1137-1150.

Raghavendra, S.N. and Deka, P.C., (2014). Support vector machine applications in the field of hydrology: a review. *Applied Soft Computing*, **19**, 372–386.

Safavi, H. R., & Esmikhani, M. (2013). Conjunctive Use of Surface Water and Groundwater : Application of Support Vector Machines (SVMs) and Genetic Algorithms, *Water Resources Management*, **27(7),** 2623–2644.

Samsudin, R., Saad, P., & Shabri, A. (2011). River flow time series using least squares support vector machines, *Hydrology and Earth System Sciences,* **15,** 1835–1852.

Seckin, N. Guven, A., (2012). Estimation of Peak Flood Discharges at Ungauged Sites Across Turkey, *Water Resour Manage*, **26**, 2569–2581.

Shabri, A. (2012). Streamflow forecasting using least-squares support vector machines Streamflow forecasting using least-squares support vector machines, *Hydrological Sciences Journal*, **57(7),** 1275-1293.

Shaghaghi, S., Bonakdari, H., Gholami, A., & Ebtehaj, I. (2017). Comparative analysis of GMDH neural network based on genetic algorithm and particle swarm optimization in stable channel design, **313,** 271–286.

Singh, D., Jain, S. K., Gupta, R. (2015). Statistical downscaling and projection of future temperature and precipitation change in middle catchment of Sutlej River Basin, India, *Journal of Earth System Science*, **124**, 843-860.

Sivapragasam, C., Liong, S., & Pasha, M. F. K. (2001). Rainfall and runoff forecasting with SSA – SVM approach, *Journal of Hydroinformatics,* **3(3),** 141–152.

S.J. Farlow (1984) Self-organizing methods in modeling: GMDH type algorithms, vol. **54** CRC Press.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H. L. (eds.), (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 (IPCC). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp.

Tofiq, F., Guven, A. (2014). Prediction of design flood discharge by statistical downscaling and General Circulation Models, *Journal of Hydrology*, **517**, 1145- 1153.

Tofiq, F., Güven, A. (2015). Potential changes in inflow design flood under future climate projections for Darbandikhan Dam, *Journal of Hydrology*, **528**, 45-51.

Traore, S., Guven, A. (2013). New algebraic formulations of evapotranspiration extracted from gene-expression programming in the tropical seasonally dry regions of West Africa, *Irrig Sci*, **31**, 1–10.

Tripathi, S., Srinivas, V. V, & Nanjundiah, R. S. (2006). Downscaling of precipitation for climate change scenarios : A support vector machine approach, *Journal of Hydrology*, **330 (3-4),** 621–640.

Vapnik, V., (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, **10 (5),** 988–999.

Wang, W., Chau, K., Cheng, C., & Qiu, L. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*, ***374*****(3–4),** 294–306.

Weimin, C. Z., Zhou, C., Xiao, W., Tirpak, T. M., & Nelson, P. C. (2003). Evolving Classification Rules with Gene Expression Programming. *IEEE Transactions on Evolutionary Computation*, **7(6),** 519–531.

Wilby, R. L., Hay, L. E., Leavesley, G. H. (1999). A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River basin, Colorado, *Journal of Hydrology*, **225**, 67-91

Wilby, R. L., Wigley, T. (1997). Downscaling general circulation model output: a review of methods and limitations, *Progress in physical geography*, **21**, 530-548.

Wilby, R. L., Dawson, C. W., Barrow, E. M. (2002). SDSM—a decision support tool for the assessment of regional climate change impacts, *Environmental Modelling & Software*, **17**, 145-157.

Wilby, R.L., Charles, S.P., Zorita, E., Timbal, B., Whetton, P., Mearns, L.O., (2004). Guidelines for use of climate scenarios developed from statistical downscaling methods. Published on-line, supporting material to the IPCC, pp. 27.

Yano, T., Aydin, M., Haraguchi, T. (2007). Impact of climate change on irrigation demand and crop growth in a Mediterranean environment of Turkey, *Sensors*, **7**, 2297-2315.

Yu, P., Chen, S., & Chang, I. (2006). Support vector regression for real-time flood stage forecasting, *Journal of Hydrology*, **328,** 704–716.

Zakaria, Z. A. (2012). Streamflow Forecasting at Ungaged Sites Using Support Vector Machines, *Applied Mathematical Sciences,* **6(60),** 3003–3014.

Zorn, C., R., Shameldin A., Y. (2015). Peak flood estimation using gene expression programming, *Journal of Hydrology*, **531**, 1122-1128.

APPENDICES

APPENDIX A

Table A1. Tables of monthly areal mean rainfall data (predictand) of Goksun Rıver basin

Year	January	February	March	April	May	Jun	July	August	September	October	November	December	Annual Max
1971	14.3	60.3	92.2	103.6	30.3	8.9	4.0	20.5	1.8	22.6	49.3	52.8	103.6
1972	22.9	31.4	31.8	102.7	67.1	61.3	23.0	8.1	12.6	54.3	43.6	1.5	102.7
1973	24.5	72.2	55.6	79.5	43.5	17.2	5.4	0.0	7.7	19.1	56.5	50.5	79.5
1974	56.7	30.1	35.4	62.0	16.8	5.4	1.4	13.9	21.0	45.7	28.1	124.9	124.9
1975	86.0	81.4	29.9	88.9	43.6	8.9	0.5	12.8	5.0	3.4	35.2	58.4	88.9
1976	137.9	69.2	38.5	76.5	100.7	20.0	7.2	7.5	20.4	86.7	59.0	100.4	137.9
1977	39.5	73.7	108.7	116.8	67.4	13.7	7.1	0.0	16.0	21.5	6.7	77.8	116.8
1978	155.9	78.2	75.9	78.5	39.0	16.6	0.0	0.0	8.1	80.3	17.3	82.8	155.9
1979	137.1	88.7	27.0	46.4	58.1	29.9	30.2	7.0	3.5	40.9	95.5	79.9	137.1
1980	113.5	59.4	135.4	88.0	139.9	2.2	0.7	4.1	0.2	37.2	47.4	67.8	139.9
1981	185.6	88.5	82.4	35.2	53.2	41.6	8.4	3.4	5.8	27.1	44.9	140.6	185.6
1982	26.6	41.5	69.8	72.2	48.6	4.1	35.1	0.6	12.2	34.1	21.8	41.1	72.2
1983	65.5	70.2	68.5	80.4	65.7	31.9	0.0	1.1	5.3	47.2	117.0	28.9	117.0
1984	64.2	53.1	65.1	82.3	36.5	21.1	5.7	4.3	0.0	21.9	65.7	38.6	82.3
1985	42.1	69.1	47.0	39.4	23.2	30.3	4.1	4.5	3.6	103.4	32.6	25.2	103.4
1986	69.5	63.7	11.1	25.8	81.8	34.9	0.3	0.4	9.1	52.6	70.5	113.9	113.9
1987	111.8	38.9	135.8	27.1	19.9	23.3	0.7	1.8	0.2	53.7	106.8	137.3	137.3
1988	20.9	63.0	173.5	52.2	33.6	26.6	9.0	0.1	7.4	76.7	116.2	66.1	173.5
1989	17.4	3.7	42.2	26.8	11.6	13.0	0.3	0.1	9.2	74.5	166.8	95.9	166.8
1990	38.9	79.4	23.0	24.6	54.4	6.0	2.0	0.6	15.0	33.5	33.4	54.1	79.4

Table A.1 Continued…

Year	January	February	March	April	May	Jun	July	August	September	October	November	December	Annual Max
1991	47.5	75.8	56.5	89.8	51.8	18.4	0.0	0.0	9.7	59.9	124.3	154.7	154.7
1992	25.5	107.1	32.5	32.0	110.2	14.2	7.2	0.0	5.2	7.7	91.2	114.0	114.0
1993	33.8	29.2	72.7	44.9	124.6	20.3	0.1	1.4	0.0	23.1	13.5	54.7	124.6
1994	127.6	73.2	19.4	45.1	46.8	1.2	21.8	2.5	10.5	53.8	84.4	86.3	127.6
1995	71.8	27.9	58.9	108.9	82.7	60.8	9.0	1.4	8.3	44.7	123.3	19.7	123.3
1996	98.0	45.7	149.1	131.9	38.0	8.0	2.4	4.8	46.9	75.3	18.6	125.8	149.1
1997	33.3	52.8	24.6	99.6	36.7	13.5	3.2	1.7	20.8	47.8	25.3	64.0	99.6
1998	36.7	13.7	123.4	79.6	63.9	18.8	2.0	7.2	6.2	28.2	98.2	108.1	123.4
1999	64.1	58.9	51.1	21.9	33.1	14.8	2.2	15.6	13.2	34.6	11.1	42.3	64.1
2000	165.5	65.9	60.0	53.2	39.0	1.2	0.0	11.3	29.9	63.0	20.1	42.8	165.5

APPENDIX B

The equation of GMDH generated to predict mean precipitation P_M

N(16) = 3.453332-0.01791*(V26)+0.000721*(V26)^2-
0.019384*(V20)+0.00069*(V20)^2-0.002027*(V26)*(V20)

N(19) = 3.609049-0.029897*(V20)-
0.000367*(V20)^2+0.019012*(V16)+0.000354*(V16)^2+0.001116*(V20)*(V16)

N(21) = 3.589917-0.029924*(V20)-0.000329*(V20)^2-
0.018946*(V19)+0.000486*(V19)^2-0.001247*(V20)*(V19)

N(15) = -0.032398+0.543152*N(16)-3.107203*N(19)+3.574388*N(21)

N(13) = -1.280065-0.027373*(V15)-0.00007*(V15)^2+1.998859*N(15)-
0.175181*N(15)^2+0.008477*(V15)*N(15)

N(24) = 4.232115-0.129219*(V20)+0.000639*(V20)^2-
1.1908*N(19)+0.286453*N(19)^2+0.037791*(V20)*N(19)

N(23) = -1.221302-0.046625*(V15)-0.000238*(V15)^2+1.730651*N(24)-
0.106958*N(24)^2+0.013017*(V15)*N(24)

N(26) = 3.998968-0.12415*(V20)+0.000611*(V20)^2-
1.019035*N(21)+0.256524*N(21)^2+0.03605*(V20)*N(21)

N(25) = -1.281575-0.049644*(V15)-0.00024*(V15)^2+1.751412*N(26)-
0.108208*N(26)^2+0.013727*(V15)*N(26)

N(12) = 0.014558+0.28988*N(13)-2.078403*N(23)+2.783878*N(25)

N(27) = -0.04929-0.002612*(V16)-2.395334*N(23)+3.411312*N(25)

N(31) = 3.522135-0.010793*(V25)+0.000381*(V25)^2-0.029939*(V20)-
0.000103*(V20)^2-0.000994*(V25)*(V20)

N(33) = 3.547323-0.029759*(V20)-
0.000371*(V20)^2+0.012619*(V1)+0.000055*(V1)^2+0.000537*(V20)*(V1)

N(30) = -0.21781+1.271464*N(31)+6.76075*N(31)^2-
0.218599*N(33)+7.20511*N(33)^2-13.96947*N(31)*N(33)

N(29) = -0.020238+0.396603*N(30)-3.529498*N(24)+4.139352*N(26)

N(28) = -1.037118-0.043514*(V15)-0.000177*(V15)^2+1.570747*N(29)-0.078699*N(29)^2+0.012197*(V15)*N(29)

N(11) = -0.003966-0.704132*N(12)+0.589732*N(27)+1.115666*N(28)

N(36) = -0.069948+0.003242*(V19)-2.443335*N(23)+3.466044*N(25)

N(35) = -0.004177-0.738146*N(12)+0.619338*N(36)+1.12014*N(28)

N(40) = -0.037414-3.331391*N(19)+0.520594*N(33)+3.822735*N(21)

N(39) = -2.780193+0.000479*(V18)-1.583011e-008*(V18)^2+2.172428*N(40)-0.132757*N(40)^2-0.000092*(V18)*N(40)

N(38) = 0.014139+0.265011*N(39)-2.177498*N(23)+2.907976*N(25)

N(41) = -0.163098+0.000031*(V18)-2.391358*N(23)+3.402577*N(25)

N(37) = -0.004941-0.911871*N(38)+0.736899*N(41)+1.176548*N(28)

N(10) = 0.002167-8.424331*N(11)+7.827206*N(35)+1.596434*N(37)

N(45) = -1.180108-0.025848*(V15)-0.000041*(V15)^2+1.896971*N(40)-0.15522*N(40)^2+0.007667*(V15)*N(40)

N(44) = 0.014016-2.130738*N(23)+0.267909*N(45)+2.858357*N(25)

N(43) = -0.004017-0.824756*N(44)+0.673546*N(27)+1.152492*N(28)

N(46) = -0.004192-0.857684*N(44)+0.70143*N(36)+1.157592*N(28)

N(42) = 0.002366-9.669016*N(43)+8.616692*N(46)+2.05157*N(37)

N(49) = 0.036743-2.434034*N(23)+3.42231*N(25)

N(48) = -0.003777-0.700154*N(38)+0.57709*N(49)+1.124269*N(28)

N(47) = 0.001083+3.257943*N(48)-19.53855*N(43)+17.28026*N(46)

N(9) = 0.00025-2.354258*N(10)+2.641205*N(42)+0.712973*N(47)

N(55) = -0.035248+0.440182*N(31)-3.06864*N(19)+3.639703*N(21)

N(54) = -1.17521-0.020954*(V15)-0.000004*(V15)^2+1.939096*N(55)-0.166905*N(55)^2+0.006442*(V15)*N(55)

N(53) = 0.018002+0.216243*N(54)-2.176056*N(23)+2.954069*N(25)

N(52) = -0.003576-0.842584*N(53)+0.769766*N(27)+1.073959*N(28)

N(56) = -0.00367-0.866423*N(53)+0.79187*N(36)+1.075724*N(28)

N(57) = -0.00413-0.812479*N(38)+0.672233*N(27)+1.141564*N(28)

N(51) = 0.001486-6.889447*N(52)+6.509656*N(56)+1.379317*N(57)

N(50) = 0.000237+0.578103*N(51)-3.622002*N(10)+4.043823*N(42)

N(7) = -0.200237+0.000032*(V18)+2.270696*N(9)-1.248623*N(50)

N(6) = 1.078027e-013+1*N(7)

N(62) = -0.004333-0.855756*N(38)+0.707969*N(36)+1.149169*N(28)

N(61) = -0.230218+0.000036*(V18)-1.014433*N(43)+2.040251*N(62)

N(60) = 0.000827-2.246018*N(10)+0.037822*N(61)+3.207932*N(42)

N(59) = -0.092199+0.000015*(V18)-0.272109*N(60)+1.282291*N(9)

N(58) = 0.000344-4.267949e-008*(V18)-0.226763*N(59)+1.226709*N(7)

N(65) = 0.123697-0.000019*(V18)-3.071114*N(10)+4.057141*N(42)

N(66) = -0.241008+0.000038*(V18)-0.30976*N(61)+1.336586*N(47)

N(67) = -0.245942+0.000039*(V18)-0.636824*N(51)+1.664097*N(47)

N(64) = 0.000152-0.017978*N(65)-0.344311*N(66)+1.362241*N(67)

N(75) = 3.469581-0.030796*(V26)-
0.000351*(V26)^2+0.007199*(V16)+0.000072*(V16)^2+0.000202*(V26)*(V16)

N(76) = 3.462127-0.030135*(V26)-0.000322*(V26)^2-
0.008493*(V19)+0.000127*(V19)^2-0.000291*(V26)*(V19)

N(74) = -0.007922-9.074339*N(75)+8.922529*N(76)+1.154338*N(33)

N(73) = -1.049528-0.021635*(V15)-0.000091*(V15)^2+1.849424*N(74)-
0.15164*N(74)^2+0.006914*(V15)*N(74)

N(72) = 0.010721+0.233643*N(73)-2.4539*N(23)+3.216836*N(25)

N(71) = -0.003896-0.784457*N(72)+0.622955*N(27)+1.162745*N(28)

N(77) = -0.004021-0.789573*N(72)+0.630841*N(36)+1.160015*N(28)

N(70) = 0.000607-21.24416*N(71)+2.381558*N(48)+19.86241*N(77)

N(69) = 0.000397-2.701737*N(10)+2.904714*N(42)+0.796896*N(70)

N(68) = -0.000153-0.622736*N(66)+0.432703*N(69)+1.190082*N(67)

N(80) = 0.001581+2.686324*N(48)-17.04015*N(11)+15.35332*N(35)

N(79) = 0.00064-3.463031*N(10)+0.896587*N(80)+3.56624*N(42)

N(78) = 0.000407-0.325572*N(66)-0.419894*N(79)+1.745337*N(67)

N(63) = 0.000001-2.488767*N(64)+1.422192*N(68)+2.066575*N(78)

N(5) = 0.000123-4.260404*N(6)+3.886435*N(58)+1.373929*N(63)

N(82) = -3.330669e-015+1.411157e-017*(V18)+1*N(7)

N(85) = 0.000871-2.184183*N(10)+3.183906*N(42)

N(84) = -0.089708+0.000014*(V18)-0.255449*N(85)+1.265359*N(9)

N(83) = 0.00007-0.20562*N(84)+1.205598*N(7)

N(81) = 0.000122-4.37934*N(82)+4.018034*N(83)+1.361267*N(63)

N(86) = 0.000122-4.37934*N(6)+4.018034*N(83)+1.361267*N(63)

N(4) = 0.001522+36.38787*N(5)-1.809709e+011*N(81)+1.809709e+011*N(86)

N(89) = 0.000075-0.226727*N(59)+1.226703*N(7)

N(88) = 0.000123-4.219981*N(82)+3.850014*N(89)+1.369927*N(63)

N(87) = 0.000567+27.10875*N(88)-5.995405e+010*N(81)+5.995405e+010*N(86)

N(92) = 0.000321-3.982631e-008*(V18)-0.205654*N(84)+1.205603*N(7)

N(91) = 0.000123-4.422174*N(82)+4.056957*N(92)+1.365178*N(63)

N(90) = 0.000077-18.84226*N(91)+19.84223*N(88)

N(3) = -0.000583+29.40244*N(4)-60.70205*N(87)+32.29896*N(90)

N(95) = 0.000123-4.422174*N(6)+4.056957*N(92)+1.365178*N(63)

N(96) = 0.000123-4.219981*N(6)+3.850014*N(89)+1.369927*N(63)

N(94) = 0.000077-18.84226*N(95)+19.84223*N(96)

N(93) = -0.000583+29.40244*N(4)-60.70205*N(87)+32.29896*N(94)

N(2) = 0.000678-3.021329e+009*N(3)+3.021329e+009*N(93)

N(104) = -0.032246+0.000575*(V20)-4.38505*N(57)+5.395291*N(62)

N(103) = 0.000328+0.501704*N(104)-3.506497*N(10)+4.004688*N(42)

N(102) = -0.170321+0.000027*(V18)+1.97684*N(9)-0.958084*N(103)

N(101) = 0.000095-0.420957*N(59)+0.55333*N(102)+0.867597*N(7)

N(100) = 0.000109-1.857303*N(82)+1.642009*N(101)+1.21526*N(63)

N(99) = -9.878956e-007+0.009784*N(100)+0.990216*N(96)

N(105) = -6.439294e-015+1*N(81)

N(98) = -0.000196-16.46437*N(99)+16.0453*N(105)+1.419079*N(4)

N(97) = 0.000231+1.502507*N(98)-7.080428e+009*N(3)+7.080428e+009*N(93)

N(109) = 0.000109-1.857303*N(6)+1.642009*N(101)+1.21526*N(63)

N(108) = -9.878956e-007+0.009784*N(109)+0.990216*N(96)

N(107) = -0.000196-16.46437*N(108)+16.0453*N(105)+1.419079*N(4)

N(106) = 0.000231+1.502665*N(107)-7.081137e+009*N(3)+7.081137e+009*N(93)

N(1) = -0.000728+0.470885*N(2)-4587.998*N(97)+4588.527*N(106)

N(113) = 0.000077-18.84226*N(95)+19.84223*N(88)

N(112) = -0.000583+29.40244*N(4)-60.70205*N(87)+32.29896*N(113)

N(115) = 0.000077-18.84226*N(91)+19.84223*N(96)

N(114) = -0.000583+29.40244*N(4)-60.70205*N(87)+32.29896*N(115)

N(111) = 0.000933+2.141583e+009*N(112)-
3.804822e+009*N(3)+1.663239e+009*N(114)

N(110) = -0.001271+0.772258*N(111)-6487.459*N(97)+6487.687*N(106)

N(119) = -1.421085e-014+1*N(88)

N(120) = -5.162537e-015+1*N(86)

N(118) = -0.000377-39.7737*N(119)+38.60192*N(120)+2.171818*N(4)

N(117) = 0.000395+1.074177*N(118)-5.320199e+009*N(3)+5.320199e+009*N(93)

N(123) = -5.77316e-015+1*N(96)

N(122) = -0.000377-39.7737*N(123)+38.60192*N(120)+2.171818*N(4)

N(121) = 0.000395+1.074121*N(122)-5.319973e+009*N(3)+5.319973e+009*N(93)

N(125) = 0.00015+5.839474*N(87)-4.229822e+009*N(90)+4.229822e+009*N(115)

N(124) = 0.000199+1.66585*N(125)-7.640794e+009*N(3)+7.640794e+009*N(93)

N(116) = 0.000416-19001.25*N(117)+18999*N(121)+3.185901*N(124)

InT = -0.000008-3.179779*N(1)+3.106413*N(110)+1.073368*N(116)

APPENDIX C

Tables and Absolute E% results of PM during calibration and validation period by GMDH model

Calibration period 1971-1990

Table C1. Numerical results of the observed and predicted P_M, and absolute errors %

Month	Observed	Output	Absolute E%
Jan	71.54	65.08	0.090331
Feb	60.78	50.43	0.170359
Mar	67.43	61.41	0.089204
Apr	65.44	48.78	0.254631
May	51.75	40.85	0.210568
Jun	20.85	14.60	0.299606
July	7.25	3.16	0.564232
Aug	4.55	3.13	0.311489
Sep	8.21	10.10	0.230173
Oct	46.82	33.52	0.284036
Nov	60.71	51.18	0.156926
Dec	71.92	49.56	0.310883
Winter	68.08	55.02	0.191813
Spring	61.54	50.35	0.181860
Summer	10.88	6.96	0.360043
Autumn	38.58	31.60	0.180893
Annual	44.77	35.98	0.196264

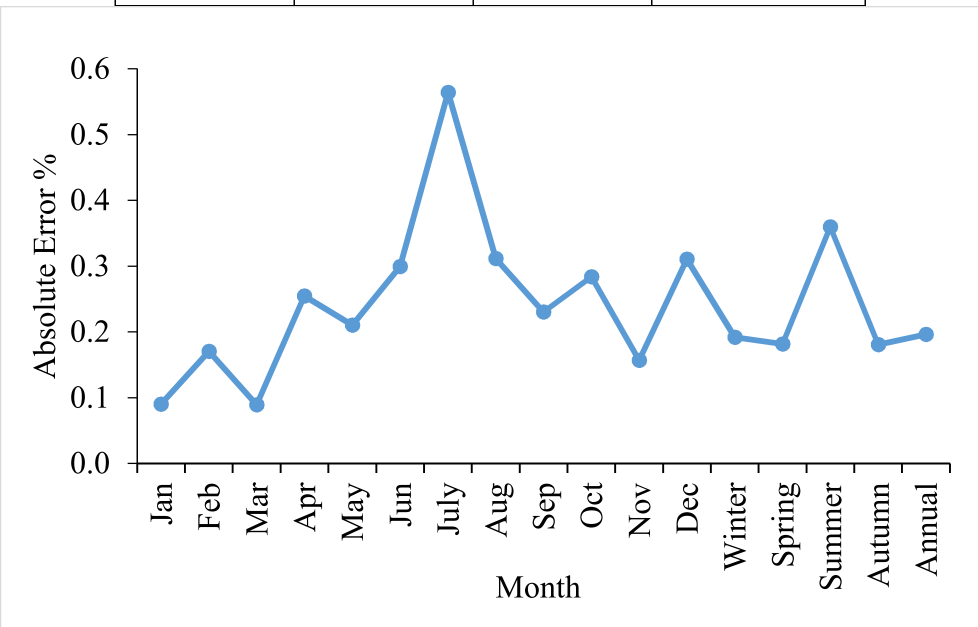

Figure C1. Absolute error percentages of observed and predicted P_M

Validation period 1991-2000

Table C2. Numerical results of the observed and predicted P_M, and absolute errors %

Month	Observed	Output	Absolute E%
Jan	70.39	56.14	0.202325
Feb	55.03	51.01	0.073053
Mar	64.81	48.07	0.258248
Apr	70.70	46.38	0.343993
May	62.67	36.56	0.416537
Jun	17.13	8.36	0.511922
July	4.79	2.33	0.512566
Aug	4.57	2.59	0.434181
Sep	15.07	9.25	0.386151
Oct	43.80	35.40	0.191764
Nov	61.01	66.02	0.082246
Dec	81.25	50.33	0.380572
Winter	68.89	52.50	0.237979
Spring	66.06	43.67	0.338890
Summer	8.83	4.43	0.498613
Autumn	39.96	36.89	0.076761
Annual	45.94	34.37	0.251722

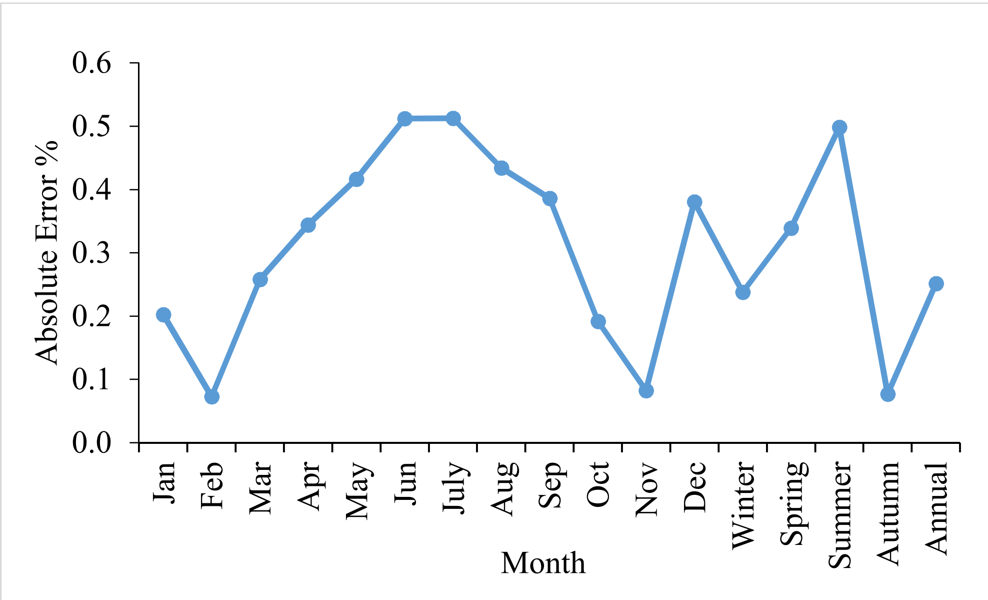

Figure C2. Absolute error percentages of observed and predicted P_M

APPENDIX D

Tables and Absolute E% results of PM during calibration and validation period by GEP model

Calibration period 1971-1990

Table D1. Numerical results of the observed and predicted P_M, and absolute errors %

Month	Observed	Output	Absolute E%
Jan	71.54	56.44	0.211042
Feb	60.78	55.88	0.080731
Mar	67.43	65.10	0.034543
Apr	65.44	48.33	0.261441
May	51.75	26.80	0.482058
Jun	20.85	10.85	0.479715
July	7.25	2.75	0.620923
Aug	4.55	3.07	0.325551
Sep	8.21	11.38	0.386194
Oct	46.82	33.92	0.275632
Nov	60.71	53.38	0.120654
Dec	71.92	57.09	0.206194
Winter	68.08	56.47	0.170554
Spring	61.54	46.74	0.240405
Summer	10.88	5.55	0.489598
Autumn	38.58	32.89	0.147402
Annual	44.77	35.42	0.208957

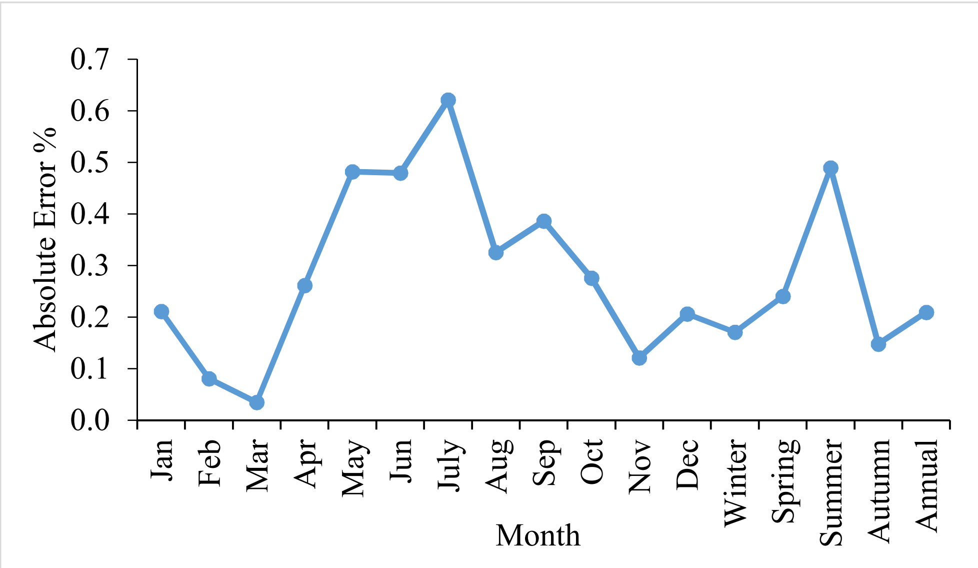

Figure D1. Absolute error percentages of observed and predicted P_M

Validation period 1991-2000

Table D2. Numerical results of the observed and predicted P_M, and absolute errors %

Month	Observed	Output	Absolute E%
Jan	70.39	68.16	0.031644
Feb	55.03	61.52	0.117862
Mar	64.81	57.37	0.114825
Apr	70.70	42.90	0.393226
May	62.67	22.94	0.633876
Jun	17.13	8.91	0.479633
July	4.79	2.52	0.473368
Aug	4.57	2.52	0.449402
Sep	15.07	10.32	0.315564
Oct	43.80	27.83	0.364566
Nov	61.01	56.22	0.078518
Dec	81.25	59.63	0.266082
Winter	68.89	63.10	0.084001
Spring	66.06	41.07	0.378273
Summer	8.83	4.65	0.473281
Autumn	39.96	31.46	0.212837
Annual	45.94	35.07	0.236527

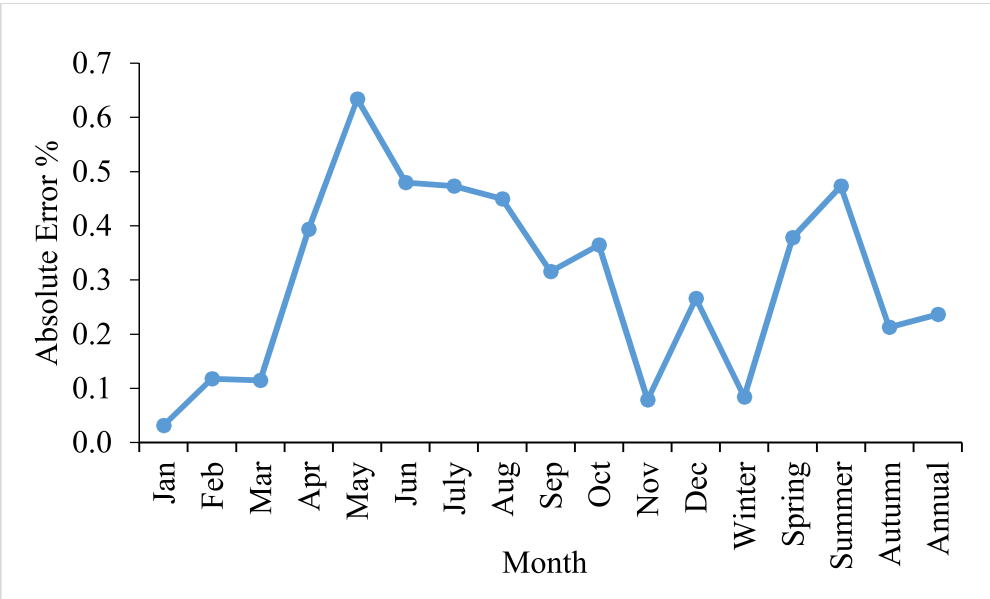

Figure D2. Absolute error percentages of observed and predicted P_M

APPENDIX E

Tables and Absolute E% results of PM during calibration and validation period by SVM model

Calibration period 1971-1990

Table E1. Numerical results of the observed and predicted P_M, and absolute errors %

Month	Observed	Output	Absolute E%
Jan	71.54	56.84	0.205467
Feb	60.78	55.41	0.088331
Mar	67.43	65.08	0.034902
Apr	65.44	57.43	0.122454
May	51.75	46.63	0.098862
Jun	20.85	16.28	0.218980
July	7.25	2.98	0.588945
Aug	4.55	2.65	0.417106
Sep	8.21	7.64	0.069596
Oct	46.82	40.40	0.137289
Nov	60.71	44.03	0.274680
Dec	71.92	52.01	0.276810
Winter	68.08	54.76	0.195730
Spring	61.54	56.38	0.083864
Summer	10.88	7.30	0.328770
Autumn	38.58	30.69	0.204549
Annual	44.77	37.28	0.167274

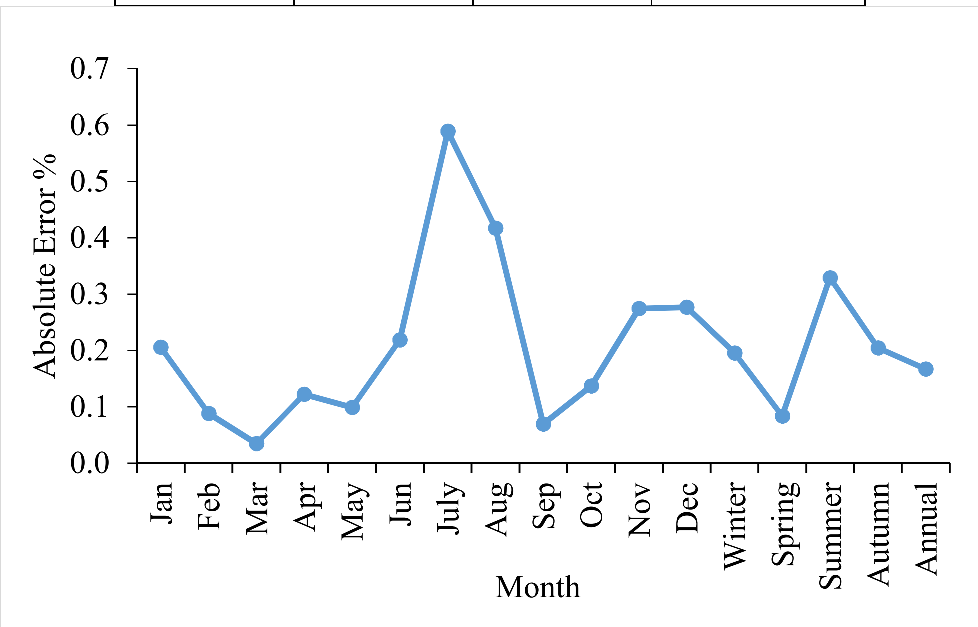

Figure E1. Absolute error percentages of observed and predicted P_M

Validation period 1991-2000

Table E2. Numerical results of the observed and predicted P_M, and absolute errors %

Month	Observed	Output	Absolute E%
Jan	70.39	55.21	0.215545
Feb	55.03	50.96	0.073982
Mar	64.81	51.43	0.206551
Apr	70.70	69.24	0.020650
May	62.67	53.90	0.139836
Jun	17.13	9.37	0.453059
July	4.79	2.06	0.569006
Aug	4.57	1.94	0.574929
Sep	15.07	7.68	0.490624
Oct	43.80	34.35	0.215713
Nov	61.01	44.70	0.267314
Dec	81.25	50.48	0.378750
Winter	68.89	52.22	0.242012
Spring	66.06	58.19	0.119133
Summer	8.83	4.46	0.495051
Autumn	39.96	28.91	0.276536
Annual	45.94	35.94	0.217501

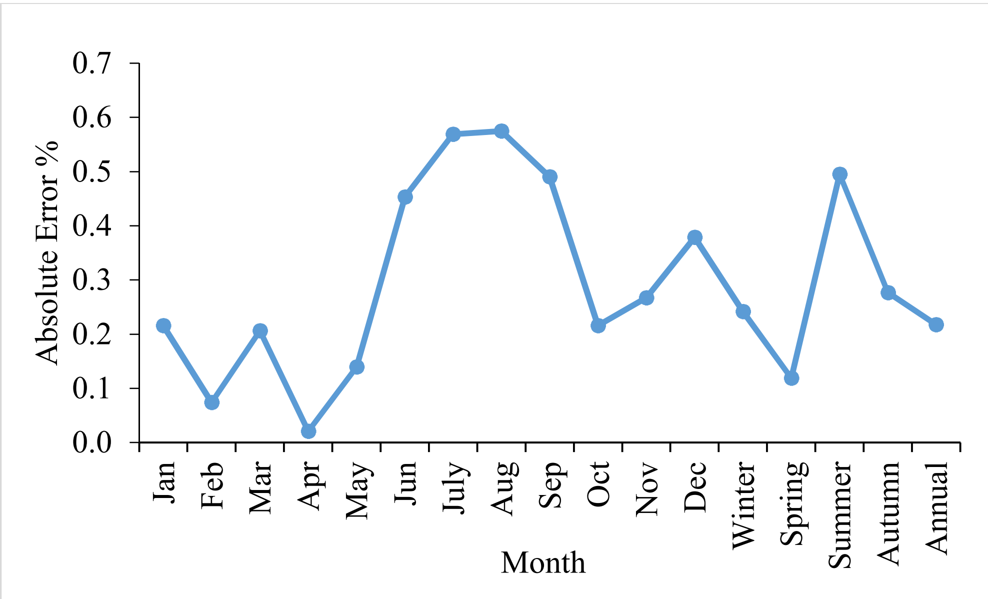

Figure E2. Absolute error percentages of observed and predicted P_M

APPENDIX F

Tables and Absolute E% results of PM during calibration and validation period by LR model

Calibration period 1971-1990

Table F1. Numerical results of the observed and predicted P_M, and absolute errors %

Month	Observed	Output	Absolute E%
Jan	71.54	76.50	0.069371
Feb	60.78	64.64	0.063442
Mar	67.43	78.53	0.164669
Apr	65.44	46.75	0.285619
May	51.75	24.57	0.525111
Jun	20.85	10.71	0.486050
July	7.25	3.52	0.515077
Aug	4.55	4.06	0.108178
Sep	8.21	9.82	0.195982
Oct	46.82	27.53	0.411992
Nov	60.71	40.30	0.336198
Dec	71.92	52.91	0.264359
Winter	68.08	64.68	0.049915
Spring	61.54	49.95	0.188282
Summer	10.88	6.10	0.439840
Autumn	38.58	25.88	0.329114
Annual	44.77	36.65	0.181305

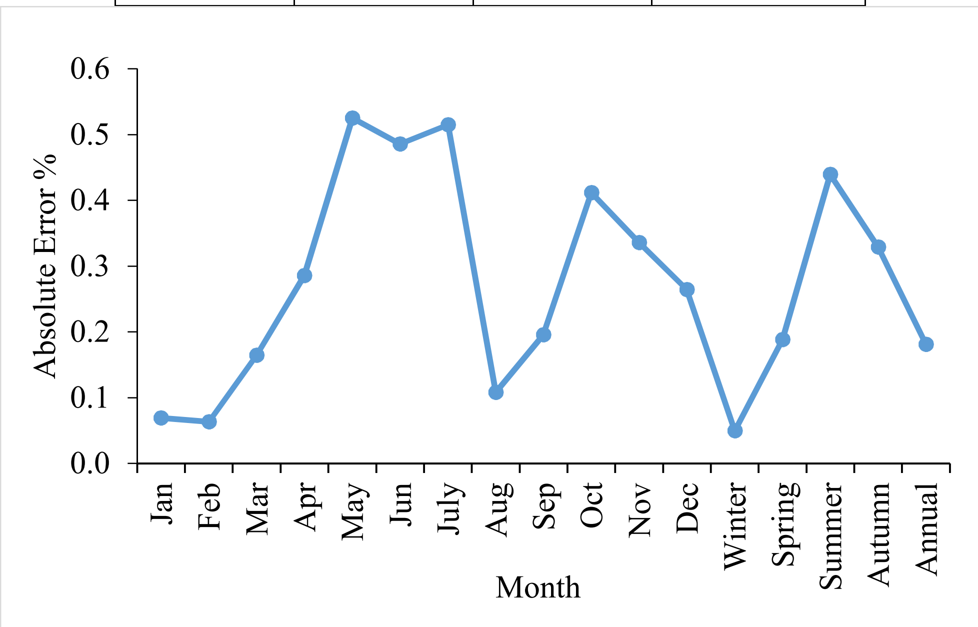

Figure F1. Absolute error percentages of observed and predicted P_M

Validation period 1991-2000

Table F2. Numerical results of the observed and predicted P_M, and absolute errors %

Month	Observed	Output	Absolute E%
Jan	70.39	71.08	0.009887
Feb	55.03	58.26	0.058616
Mar	64.81	53.05	0.181431
Apr	70.70	40.26	0.430610
May	62.67	20.70	0.669723
Jun	17.13	8.03	0.531232
July	4.79	3.36	0.298365
Aug	4.57	4.06	0.111848
Sep	15.07	9.51	0.369223
Oct	43.80	23.16	0.471219
Nov	61.01	35.58	0.416841
Dec	81.25	50.12	0.383124
Winter	68.89	59.82	0.131647
Spring	66.06	38.00	0.424728
Summer	8.83	5.15	0.416739
Autumn	39.96	22.75	0.430723
Annual	45.94	31.43	0.315763

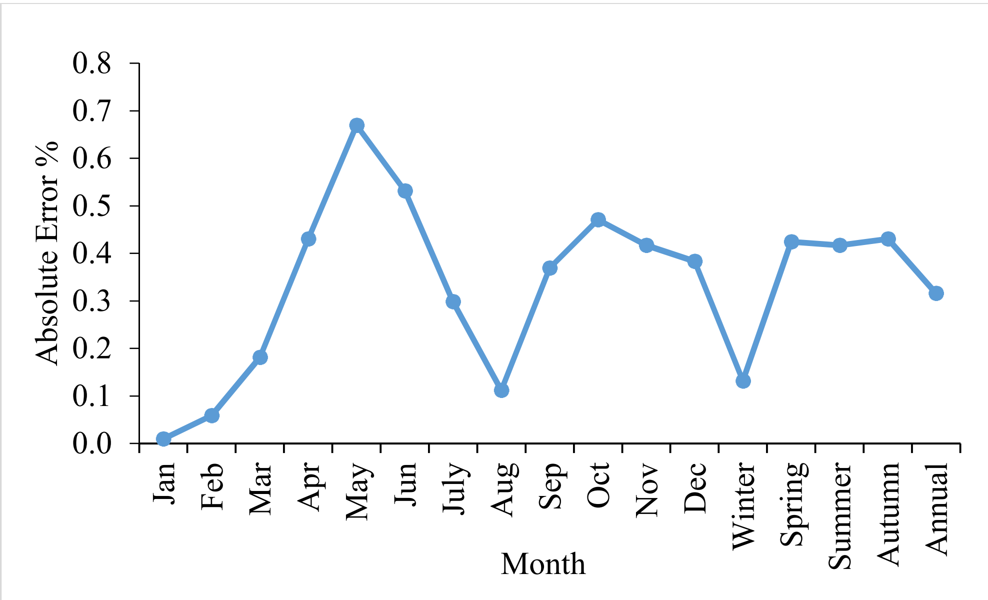

Figure F2. Absolute error percentages of observed and predicted P_M

APPENDIX G

Tables of averaged monthly mean P_M results under different CGCM3 scenarios for different time periods by SVM model

Table G1. Monthly mean P_M results under CGCM3A1B scenario for four future periods by SVM model

Month	Baseline	2020s	2040s	2060s	2080s
Jan	71.15	52.23	43.75	45.79	44.98
Feb	58.87	48.51	47.53	44.93	51.69
Mar	66.56	56.36	57.55	55.09	54.86
Apr	67.19	57.31	60.73	67.33	56.91
May	55.39	45.36	38.14	35.73	28.70
Jun	19.61	7.41	4.24	4.50	2.97
July	6.43	1.29	1.15	1.89	0.72
Aug	4.56	1.91	0.97	0.82	0.67
Sep	10.50	5.84	5.61	3.38	3.61
Oct	45.82	37.76	34.53	29.89	29.53
Nov	60.81	39.02	43.18	40.21	42.24
Dec	75.03	55.16	52.73	46.03	49.05

Table G2. Monthly mean P_M results under CGCM3A2 scenario for four future periods by SVM model

Month	Baseline	2020s	2040s	2060s	2080s
Jan	71.15	46.14	49.23	46.20	52.53
Feb	58.87	48.52	47.86	44.86	50.28
Mar	66.56	53.66	52.16	54.44	60.79
Apr	67.19	58.70	58.20	48.69	65.82
May	55.39	40.26	26.99	34.71	22.63
Jun	19.61	7.56	3.03	4.87	1.65
July	6.43	2.22	1.08	0.61	0.69
Aug	4.56	1.39	1.09	0.65	0.55
Sep	10.50	5.92	4.79	4.51	2.29
Oct	45.82	35.13	32.78	30.90	22.39
Nov	60.81	42.07	41.05	42.60	45.08
Dec	75.03	51.50	48.63	49.48	48.83

APPENDIX H

Tables and Figures of averaged monthly mean P_M results under different CGCM3 scenarios for different time periods by GMDH model

Table H1. Monthly mean P_M results under CGCM3A1B scenario for four future periods by GMDH model

Month	Baseline	2020s	2040s	2060s	2080s
Jan	71.15	50.85	52.17	51.62	52.63
Feb	58.87	48.95	46.90	44.99	47.28
Mar	66.56	51.76	48.11	45.20	45.15
Apr	67.19	46.20	40.91	39.28	38.24
May	55.39	28.17	23.74	22.55	24.57
Jun	19.61	8.52	4.03	3.51	2.92
July	6.43	1.08	0.76	0.29	0.18
Aug	4.56	1.06	0.53	0.34	0.15
Sep	10.50	6.69	5.17	3.67	2.68
Oct	45.82	29.91	34.28	25.41	31.08
Nov	60.81	48.11	64.89	50.82	68.37
Dec	75.03	49.43	53.52	45.14	52.78

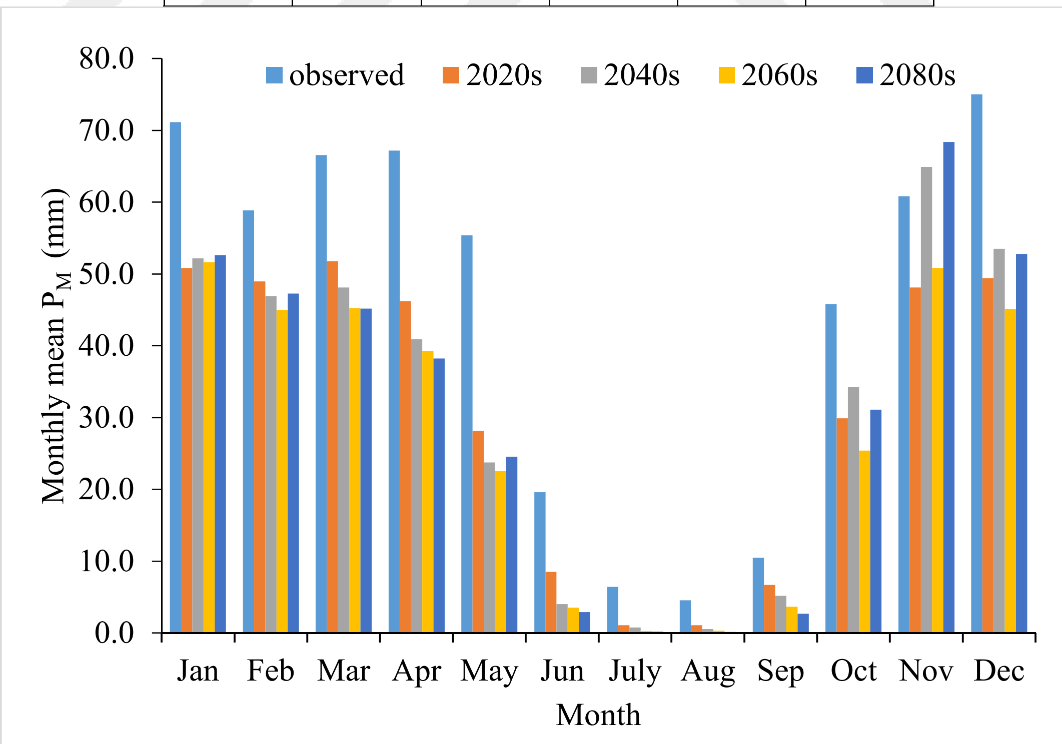

Figure H1. Monthly mean P_M under CGCM3A1B scenario for four future periods by GMDH model

Table H2. Monthly mean P_M results under CGCM3A2 scenario for four future periods by GMDH model

Month	Baseline	2020s	2040s	2060s	2080s
Jan	71.15	52.96	50.47	50.94	66.66
Feb	58.87	48.24	46.24	43.99	44.15
Mar	66.56	48.67	43.51	44.69	44.06
Apr	67.19	39.96	37.36	39.17	35.11
May	55.39	29.52	20.72	23.21	16.89
Jun	19.61	9.70	3.61	3.32	1.57
July	6.43	1.50	0.43	0.18	0.08
Aug	4.56	1.05	0.40	0.18	0.05
Sep	10.50	6.43	5.43	3.13	1.40
Oct	45.82	51.30	60.87	29.59	21.84
Nov	60.81	63.89	67.12	45.12	58.31
Dec	75.03	52.09	47.45	62.46	42.88

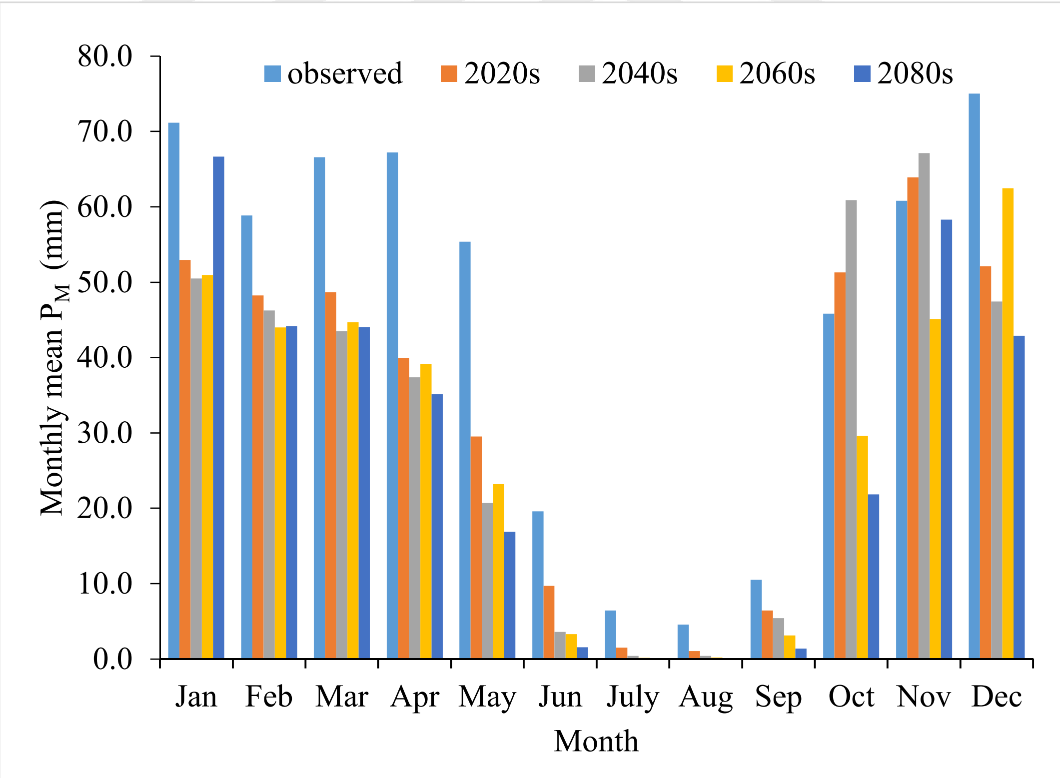

Figure H2. Monthly mean P_M under CGCM3A2 scenario for four future periods by GMDH model

APPENDIX I

Tables and Figures of averaged monthly mean P_M results under different CGCM3 scenarios for different time periods by GEP model

Table I1. Monthly mean P_M results under CGCM3A1B scenario for four future periods by GEP model

Month	Baseline	2020s	2040s	2060s	2080s
Jan	71.15	58.48	61.94	55.24	58.48
Feb	58.87	55.98	52.69	59.08	52.82
Mar	66.56	57.71	54.23	49.68	48.09
Apr	67.19	40.64	36.79	33.60	28.92
May	55.39	20.12	18.68	15.65	12.68
Jun	19.61	7.42	6.35	5.24	3.07
July	6.43	1.98	1.45	0.92	0.73
Aug	4.56	1.99	1.78	1.21	0.75
Sep	10.50	8.87	7.44	6.32	4.77
Oct	45.82	27.18	23.76	24.14	17.97
Nov	60.81	54.67	51.47	52.21	47.48
Dec	75.03	58.81	57.48	63.47	61.97

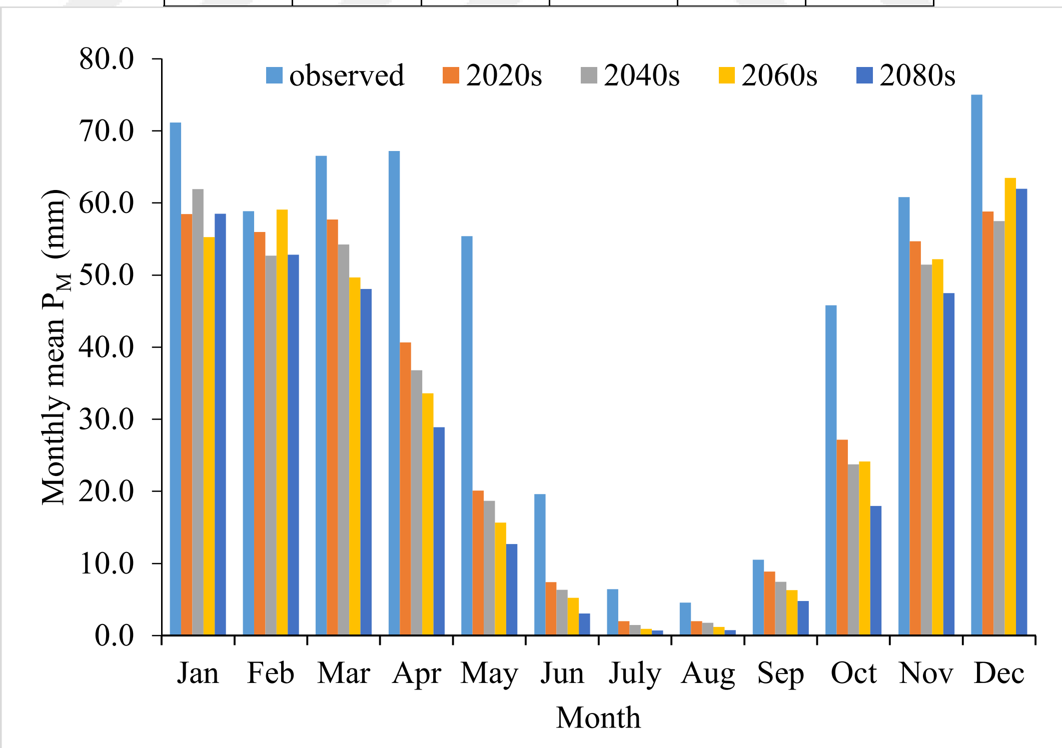

Figure I1. Monthly mean P_M under CGCM3A1B scenario for four future periods by GEP model

Table I2. Monthly mean P_M results under CGCM3A2 scenario for four future periods by GEP model

Month	Baseline	2020s	2040s	2060s	2080s
Jan	71.15	60.76	57.88	45.79	44.98
Feb	58.87	54.77	54.83	44.93	51.69
Mar	66.56	53.90	54.54	55.09	54.86
Apr	67.19	36.51	35.91	67.33	56.91
May	55.39	19.05	16.38	35.73	28.70
Jun	19.61	7.73	4.93	4.50	2.97
July	6.43	2.25	1.24	1.89	0.72
Aug	4.56	2.17	1.28	0.82	0.67
Sep	10.50	9.04	7.60	3.38	3.61
Oct	45.82	26.56	25.90	29.89	29.53
Nov	60.81	56.07	55.18	40.21	42.24
Dec	75.03	57.89	60.87	46.03	49.05

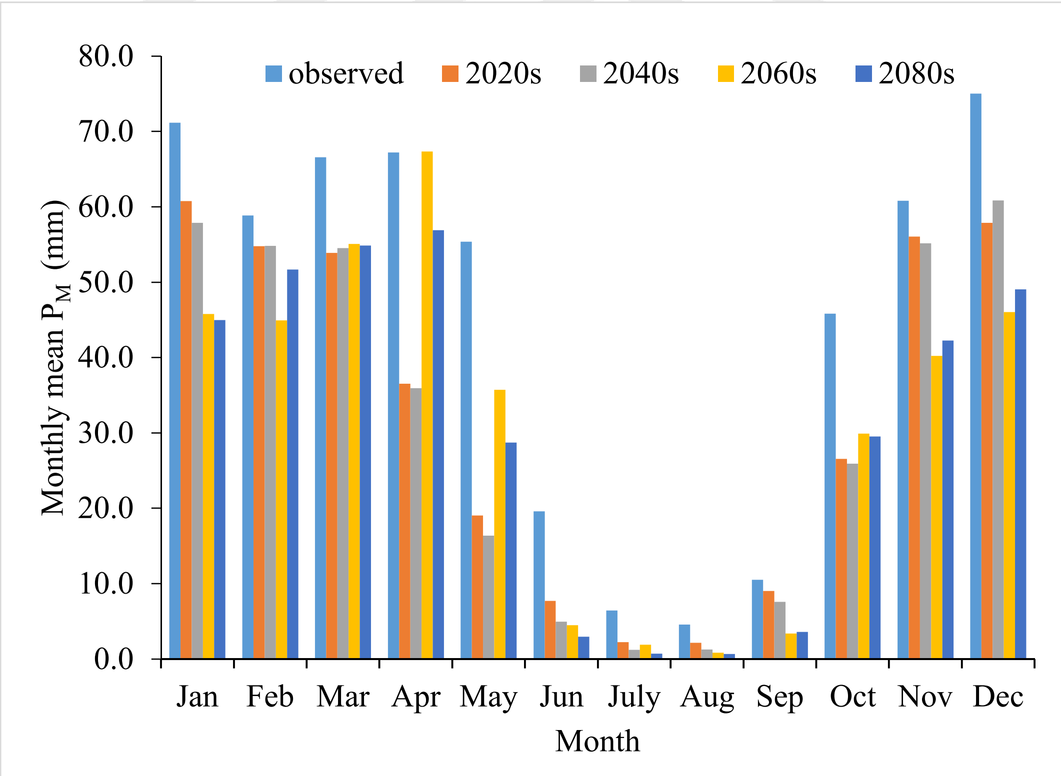

Figure I2. Monthly mean P_M under CGCM3A2 scenario for four future periods by GEP model