ANALYSIS AND MODELING OF SPATIALLY AND TEMPORALLY VARYING
METEOROLOGICAL PARAMETER: PRECIPITATION OVER TURKEY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

PINAR ASLANTAŞ BOSTAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
GEODETICAL AND GEOGRAPHICAL INFORMATION TECHNOLOGIES

FEBRUARY 2013

Approval of the thesis:

**ANALYSIS AND MODELING OF SPATIALLY AND TEMPORALLY VARYING
METEOROLOGICAL PARAMETER: PRECIPITATION OVER TURKEY**

submitted by **PINAR ASLANTAŞ BOSTAN** in partial fulfillment of the requirements for the degree
of **Doctor of Philosophy in Geodetic and Geographic Information Technologies Department,
Middle East Technical University** by,

Prof. Dr. Canan Özgen   _____
Dean, Graduate School of **Natural and Applied Sciences**

Assoc. Prof. Dr. Ahmet Coşar   _____
Head of Department, **Geodetic and Geographic Inf. Tech.**

Prof. Dr. Sevda Zuhal Akyürek   _____
Supervisor, **Civil Engineering Dept., METU**

Assoc. Prof. Dr. Gerard Heuvelink   _____
Co-Supervisor, **Env. Sci. Dept., Wageningen University of Netherlands**

Examining Committee Members:

Prof. Dr. Mehmet Lütfi Süzen   _____
Geological Engineering Dept., METU

Prof. Dr. Sevda Zuhal Akyürek   _____
Civil Engineering Dept., METU

Prof. Dr. Ali Ünal Şorman   _____
Civil Engineering Dept., METU

Assoc. Prof. Dr. İsmail Yücel   _____
Civil Engineering Dept., METU

Assist. Prof. Dr. Levent Tezcan   _____
Geological Engineering Dept., Hacettepe University

Date: 11. 02. 2013

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

Name, Last Name: PINAR ASLANTAŞ BOSTAN

Signature:

ABSTRACT

ANALYSIS AND MODELING OF SPATIALLY AND TEMPORALLY VARYING
METEOROLOGICAL PARAMETER: PRECIPITATION OVER TURKEY

Aslantaş Bostan, Pınar
Ph.D., Department of Geodetic and Geographic Information Technologies
Supervisor: Prof. Dr. Sevda Zuhal Akyürek
Co-Supervisor: Assoc. Prof. Dr. Gerard Heuvelink

February 2013, 80 pages

As precipitation is the very important parameter of climate and hydrology, exploring spatial and temporal distribution and variation of this variable can give an idea about climate conditions and water resources in the future. Therefore accurate mapping of the temporal, spatial and space-time distributions of precipitation is important for many applications in hydrology, climatology, agronomy, ecology and other environmental sciences. In this thesis, temporal, spatial and space-time distributions and variations of total annual and long term annual precipitation of Turkey are analyzed. Main data source of thesis is point observations of monthly precipitation at meteorological stations and spatially exhaustive covariate data sets. These are elevation, surface roughness, distance to coast, river density, aspect, land use and eco-region. T-Test and Mann-Kendal tests are used to infer temporal trend of seasonal and annual precipitation observations of Turkey. Multiple linear regression (MLR), Geographically Weighted Regression (GWR), Ordinary Kriging (OK), Regression Kriging (RK) and Universal Kriging (UK) are applied to define spatial distribution and variation of long term annual precipitation observations of Turkey. For the spatio-temporal part of the study Space-time Ordinary Kriging and Space-time Universal Kriging methods are applied to total annual precipitation observations of Euphrates Basin, which is the largest basin in Turkey. Comparison of interpolation methods are made with ten-fold cross-validation methodology. Accuracy assessment is done by calculating the Root Mean Squared Error (*RMSE*), R-square (r^2) and Standardized MSE (*SMSE*) for spatial interpolation. According to these criteria, Universal Kriging is the most accurate with an *RMSE* of 178 mm, an *R-square* of 0.61 and an *SMSE* of 1.06, while Multiple Linear Regression performed worst (*RMSE* of 222 mm, *R-square* of 0.39, and *SMSE* of 1.44). Ordinary Kriging, UK using only elevation and Geographically Weighted Regression are intermediate with *RMSE* values of 201 mm, 212 mm and 211 mm, and an *R-square* of 0.50, 0.44 and 0.45, respectively. The RK results are close to those of UK with an *RMSE* of 186 mm and *R-square* of 0.57. For space-time interpolation *R-square*, *RMSE* and *ME* methods are used for accuracy assessment. Space-time Ordinary kriging has yielded more accurate prediction results than Space-time Universal kriging with *R-square* of 0.86, *RMSE* of 75 mm and *ME* of 57 mm.

Keywords: climate, geostatistics, precipitation, temporal variation, spatial interpolation, space-time interpolation.

ÖZ

MEKANDA VE ZAMANDA DEĞİŞİKLİK GÖSTEREN METEOROLOJİK YAĞIŞ
PARAMETRESİNİN TÜRKİYE ÜZERİNDE ANALİZ VE MODELLEMESİ

Aslantaş Bostan, Pınar
Doktora, Jeodezi ve Coğrafi Bilgi Teknolojileri Bölümü
Tez Yöneticisi: Prof. Dr. Sevda Zuhal Akyürek
Ortak Tez Yöneticisi: Doç. Dr. Gerard Heuvelink

Şubat 2013, 80 sayfa

İklim ve hidroloji açısından yağışın çok önemli bir parametre olduğu düşünüldüğünde, bu parametrenin mekãnsal ve zamansal dağılımının ve değişiminin incelenmesi gelecekteki iklim koşulları ve su kaynakları hakkında faydalı bilgiler verebilir. Bu nedenle yağışın zamansal, mekãnsal ve mekãnsal-zamansal dağılımlarının doğru bir şekilde haritalanması hidrolojide, iklim biliminde, tarım biliminde, ekolojide ve diğer çevre bilimleri gibi bir çok uygulamada önemlidir. Bu tezde Türkiye'nin toplam yıllık ve uzun yıllar ortalama yağış değerlerinin zamansal, mekãnsal ve mekãnsal-zamansal dağılımları ve değişimleri analiz edilmiştir. Tezin ana veri kaynağı meteorolojik istasyonlarda ölçülmüş aylık yağış değerleri ve bununla mekãnsal olarak ilişkili geniş kapsamlı veri setleridir. Bunlar yükseklik, yüzey pürüzlülüğü, deniz kıyısına mesafe, akarsu yoğunluğu, bakı, arazi kullanımı ve ekolojik bölge olarak belirlenmiştir. T-Test ve Mann-Kendal testleri Türkiye'nin mevsimsel ve yıllık yağış değerlerinin zamansal değişimini belirlemek için kullanılmıştır. Türkiye'nin uzun dönem yıllık yağış ölçümlerinin dağılım ve değişimini belirlemek için Çoklu Doğrusal Regresyon, Coğrafi Ağırlıklı Regresyon, Normal kriging, Regresyon kriging ve Genel kriging yöntemleri uygulanmıştır. Tezin mekãnsal-zamansal analiz kısmında Fırat havzasının yıllık yağış değerlerine mekãn-zaman Normal kriging ve mekãn-zaman Genel kriging yöntemleri kullanılmıştır. Enterpolasyon yöntemlerinin karşılaştırılması 10 gruplu çapraz sağlama yöntemi ile yapılmıştır. Mekãnsal enterpolasyon analizinde doğruluk değerlendirmesi işlemi Kare Kök Ortalama Hata (*RMSE*), R-kare (*R-square*) ve Standart Ortalama Kare Hata (*SMSE*) istatistiksel ölçütler kullanılarak yapılmıştır. Bu kriterlere göre *RMSE* değeri 178 mm, *R-square* değeri 0.61 ve *SMSE* değeri 1.06 olan Genel kriging yöntemi en doğru yöntem olarak belirlenmiştir. Çoklu Doğrusal Regresyon yöntemi ise *RMSE* değeri 222 mm, *R-square* değeri 0.39 ve *SMSE* değeri 1.44 ile en başarısız yöntem olarak ifade edilmiştir. *RMSE* değerleri sırasıyla 201 mm, 212 mm ve 211 mm olan ve *R-square* değerleri yine aynı sırayla 0.50, 0.44 ve 0.45 olan Normal kriging, sadece yükseklik verisi kullanılarak yapılan Genel kriging ve Coğrafi Ağırlıklı Regresyon yöntemleri orta düzeyde sonuçlar vermişlerdir. Regresyon kriging yönteminin sonuçları Genel kriging yöntemi ile yakın olup değerler şu şekildedir: *RMSE* değeri 186 mm ve *R-square* değeri 0.57'dir. Mekãnsal-zamansal enterpolasyon analizinde *R-square*, *RMSE* ve *ME* ölçütleri doğruluk testi için kullanılmıştır. Mekãn-zaman Normal kriging yöntemi 0.86 *R-square*, 75 mm *RMSE* ve 57 mm *ME* ile en doğru tahmin değerlerini üretmiştir.

Anahtar kelimeler: iklim, jeoistatistik, yağış, zamansal değişim, mekãnsal enterpolasyon, mekãn-zaman enterpolasyonu.

To My Family

ACKNOWLEDGEMENTS

ABREVIATIONS

3D	Three Dimensional
BKED	Block Kriging with External Drift
CORINE	Coordination of Information on the Environment
GWR	Geographically Weighted Regression
GWRK	Geographically Weighted Regression Kriging Hybrid
IDW	Inverse Distance Weighting
KED	Kriging with External Drift
KEDl	Kriging with External Drift in a Local Neighborhood
LST	Land Surface Temperature
ME	Mean Error
MKV	Mean Kriging Variance
MLR	Multiple Linear Regression
MODIS	Moderate Resolution Imaging Spectro-radiometer
MWR	Moving Window Regression
NDVI	Normalized Difference Vegetation Index
OK	Ordinary Kriging
OKl	Ordinary Kriging in a Local Neighborhood
OLS	Ordinary Least Squares
r^2	R-square
RK	Regression Kriging
RMSE	Root Mean Square Error
Sklm	Simple Kriging with a Locally Varying Mean
SMSE	Standardized Mean Square Error
SRTM	Shuttle Radar Topography Mission
ST	Space-Time
ST-KED	Space-Time Kriging with External Drift
ST-OK	Space-Time Ordinary Kriging
ST-UK	Space-Time Universal Kriging
TFPW	Trend Free Pre-whitening
UK	Universal Kriging

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

xiv

CHAPTER 1

INTRODUCTION

1.1 Motivation

Water is life! As we all know, without water habitats and all eco-systems over the earth cannot sustain life. Moreover water is very important for soil. Insufficient water resource makes it difficult or impossible to utilize the soil. Water is important also for all industrial activities since almost all production processes need water. Water requirement is a must everywhere on the Earth. Therefore water resources are very important for the human population and also for agronomy, hydrology, environmental science, industry in Turkey and all over the world. The single source of available and usable water resources is precipitation. In that respect many scientific disciplines are interested in and work with precipitation in different aspects like quantity, density, frequency, content, characteristics and spatial and temporal distribution. The most important motivation of this thesis research is the requirement of an understanding of the behavior of precipitation in time and over space for Turkey. Defining the variation of precipitation amount that occur in time and through space can support water resources policies, and agricultural and hydrological planning. A specified probable trend existence over an area may point to future drought. Describing the spatial variation and distribution of precipitation is required at locations where observations are lacking. Also precipitation distribution maps at small scale may be more useful for some studies or environmental plans than analyzing the meteorological observations from stations one by one. Identifying space-time variation and distribution of precipitation may provide valuable information as it takes advantage of time variation by including it into interpolation. Environmental variables like precipitation, temperature, air quality, soil characteristics, etc. are temporally variable parameters which mean that they may have different values at the same place through consecutive time periods. So rather than defining the distribution of environmental variables at a unique date, if sufficient amount of data are available at different time periods and different locations, space-time interpolation may be more beneficial than spatial techniques that use only location information. As Tobler (1970) said "in geography close things are more related with each other", observations at consecutive time periods are more related with each other than observations that are more distant in time.

1.2 Scope and Aim of Thesis

The scope of this thesis can be thought of as analyzing precipitation in three different contexts. These contexts include time, space and space-time issues. According to time issue, annual and seasonal precipitation totals are analyzed with statistical tests to expose a probable variation in time over Turkey. According to space issue, distribution and predictions of precipitation are obtained and compared over whole Turkey. At space-time issue, temporal and spatial information of precipitation are combined and processed in an integrated form to obtain predictions over the Euphrates Basin in Turkey.

Temporal precipitation analysis consists of applying trend tests which reveal probable trend existence over Turkey. The aim of temporal analysis is to define increasing or decreasing trend existence for annual and seasonal precipitation values of Turkey. Spatial precipitation analysis consists of applying interpolation techniques to define spatial distribution and obtain predictions at areas without measurements. The second aim of spatial precipitation analysis is to evaluate and compare interpolation techniques on precipitation distribution. Spatio-temporal precipitation analysis consists of applying space-time interpolation techniques to define precipitation distribution in space and also in time for the Euphrates Basin. Another aim of the space-time interpolation is to compare two methods, where one uses secondary information and the other does not. Finally, assessment of spatial and

space-time interpolation methods are performed mutually. This study may be helpful and be beneficial for environmental-climatic studies, government policies, agricultural plans and programs.

1.2.1 Temporal Precipitation Analysis

Precipitation, like many other environmental variables, may have some kind of increasing or decreasing tendency through time. Defining this increasing or decreasing tendency or trend in precipitation values is very important for human population and agricultural purposes. Also exploring changes in precipitation over the long term can be helpful for future studies related to drought. According to Love (1999), trends in water that are likely to have the greatest influence on the future situation include population growth, economic expansion and, in the longer term, climate change. Analyzing the precipitation in time is the first part of this study. Annual and seasonal precipitation totals are analyzed in temporal basis in order to check if there is an increasing or decreasing trend present over Turkey. In order to define the temporal trend, T-Test and Mann-Kendal tests are applied on annual average and seasonal precipitation values that were measured at 225 meteorological stations between 1970-2003. The concern is to assess trend tendency over precipitation values for this 34 year period. According to Türkeş (1999), there has been a general tendency from humid conditions of around the 1960's towards dry sub-humid climatic conditions in the aridity index values of many stations in general. At some stations in the Aegean Region, there has been a significant change from humid conditions to dry sub-humid or semi-arid climatic conditions. With regard to climatic factors, south-eastern Anatolia and the continental interiors of Turkey appear to be arid lands that are prone to desertification. From this perspective, variation of precipitation through time is very important for Turkey as there are areas which have drought stress. In this study the T-Test and Mann-Kendal test are applied to total annual precipitation values, whereas only the Mann-Kendal test is applied to seasonal values.

1.2.2 Interpolation of Precipitation in Space

In order to obtain a prediction map of a certain variable over an area, a sufficiently large number of measurements that are approximately regularly distributed in the study area of interest are required. However, direct measurement of environmental variables at every point on the Earth or at every point within a region on the Earth is often an impossible task. Indirect measurement using remote sensing instruments is a viable alternative that yields spatially exhaustive information, but the accuracy and resolution of the information may be insufficient for the intended use. Therefore, in many practical cases the scientist still has to create spatially exhaustive information from a limited set of direct measurements. For this reason many simple and complex spatial interpolation methods have been developed to estimate the value of environmental variables at unmeasured locations. In the past, these techniques only made use of direct measurements at point locations, but more recently many of these techniques have been extended such that spatially exhaustive information can be used as a covariate in spatial interpolation (e.g Bostan et al. 2012; Knotters et al. 1995; Phillips et al. 1997; Hengl et al. 2004; Carrera-Hernández and Gaskin, 2006; Heuvelink, 2006; Grimes and Pardo-Iguzquiza, 2010; Wang et al. 2010). This also holds for precipitation mapping. Precipitation is an important environmental variable for which spatial interpolation has been applied on daily (e.g. Carrera-Hernández and Gaskin, 2006; Kyriakidis et al. 2001; Symeonakis et al. 2009), monthly (e.g. Lloyd, 2005) and annual averages (e.g. Hofierka et al. 2002; Goovaerts, 2000; Martínez-Cob, 1996). In many studies it has been shown that using the relationship between precipitation and secondary information, such as radar imagery, elevation or land use, provides more accurate estimates than using only precipitation measurements (e.g. Bostan et al. 2012, Lloyd, 2005; Hofierka et al. 2002; Boer et al. 2001). Exploring the spatial distribution and variation of precipitation is the second issue of this study. In this part, secondary information is used to improve interpolation of precipitation in a spatial context. Seven sources of secondary information which are relevant for precipitation are used during spatial interpolation to obtain more reliable predictions. These are elevation, surface roughness, distance to coast, river density, aspect, land use class and eco-region. The methods used in this second part are Multiple Linear Regression (MLR), Geographically Weighted Regression (GWR), Ordinary kriging (OK), Regression kriging (RK) and Universal kriging (UK). In order to be sure those

secondary variables are useful and necessary for precipitation mapping, Ordinary kriging, which does not use secondary information, is added to the analyses for comparison. Also Universal kriging is done twice by using all significant secondary information sources and by using only elevation to check the requirement of other variables. The methods are applied to long-term mean annual precipitation values of Turkey and the performance of each is evaluated according to *RMSE* (Root Mean Square Error), *R-square* and *SMSE* (Standardized Mean Square Error) accuracy indicators.

1.2.3   Interpolation of Precipitation in Space-time

Methods to define and estimate the spatial variability of hydrologic, climatic and other environmental variables and perform spatial interpolation using the quantified spatial variability are abundant in the environmental sciences. Recently, the extension of these methods to variables that vary both in space and time has received increasing attention. Snepvangers et al. (2003) compared two Space-time kriging techniques: Space-time Ordinary kriging (ST-OK) and ST kriging with external drift (ST-KED) on soil water content interpolation. They found out that predictions are more realistically obtained from ST-KED, and prediction uncertainty of this method is lower compared to ST-OK. Jost et al. (2005) performed a study about spatio-temporal distribution of soil water storage by using space-time kriging methods in a forest ecosystem. Hengl et al. (2012) used ST Regression kriging to predict daily temperatures for 2008 obtained from 159 meteorological stations in Croatia. Precipitation is an environmental parameter which can be analyzed in space-time context since it has variability in time and space. In space-time (ST) kriging all observations in the past, present and future are used to predict the present situation because of temporal correlation as quantified by the space-time variogram. So space-time kriging makes use of all observations from all years and locations and if there is indeed temporal correlation then the observations from other times (other years) will be included in making the prediction. Space-time analysis of precipitation is included in this thesis to check if this new technique yields improvements on precipitation interpolation for Turkey and to put a new perspective to spatial interpolation. In this third part of the thesis, before applying this new method to the whole of Turkey, a small study area is selected in order to test the usefulness and efficiency of this technique since data preparation and statistical modeling take much time and effort for a space-time analysis. Therefore the Euphrates Basin which is the biggest and one of the most important basin of Turkey is selected as a study area to implement space-time interpolation techniques. For this purpose, ST-OK and ST-UK methods are applied to total annual observations for the period of 1970-2008. The former uses only observed values, the latter uses observed values and secondary information as well. The secondary information is composed of the same data used in spatial kriging of precipitation.

1.3   Thesis Organization

The thesis includes 6 main chapters and subchapters which explain mainly the study area, literature review, techniques and results. Chapters are briefly described below:

Chapter 1 makes introduction to the study by describing the aim and reasons to implement this study. In addition, scope of thesis is summarized in three subtitles.

Chapter 2 focuses on literature review. Studies which were performed about temporal, spatial and space-time analysis by using environmental variables are searched and summarized.

Chapter 3 describes the study area and data. The primary data source precipitation is explained by statistical information of it and characteristic features. The secondary data sources: elevation, surface roughness, distance to coast, river density, aspect, land use classes and eco-region are described in terms of spatial and space-time content.

Chapter 4 deals with methodology in four sub-sections. In temporal methodology section two methods: T-Test and Mann-Kendal tests are described in detail. In spatial interpolation methodology part kriging and regression techniques and their methodology are described. These are: Multiple

Linear Regression (MLR), Geographically Weighted Regression (GWR), Ordinary kriging (OK), Regression kriging (RK), and Universal kriging. In space-time methodology part, space-time interpolation techniques are explained. Two methods are used in space-time part; Space-time Ordinary kriging (ST-OK) and Space-time Universal kriging (ST-UK). Last section of methodology chapter describes cross-validation methodology.

Chapter 5 makes assessment of results from the general perspective and discusses the results. Temporal, spatial and space-time analysis results are explained in detail.

Chapter 6 describes the conclusions of the study with recommendations for future studies.

CHAPTER 2

LITERATURE REVIEW

Novotny and Stefan (2007) studied stream flow records for 2002 from 36 gauging stations in five major river basins of Minnesota. Seven annual stream flow statistics were extracted and analyzed: mean annual flow, 7-day low flow in winter, 7-day low flow in summer, peak flow due to snow melt runoff, peak flow due to rainfall as well as high and extreme flow days. The Mann–Kendal non-parametric test was used to detect significant trends over time windows from 90 to 10 years in combination with the Trend Free Pre-Whitening (TFPW) method for correcting time series data for serial correlation. Stream flows in Minnesota reflected changes in precipitation with increases in mean annual precipitation, a larger number of intense rainfall events, and earlier and more frequent snowmelt events.

Colombo et al. (2007) studied the climatic behavior of two principal observables, temperature and precipitation as obtained from 50 meteorological stations located in Italy for 1961-2000. Stations were classified to different classes according to their geographic location: mountain (11 stations), continental (17 stations) and coastal areas (21 stations). They checked for trends in temperature and precipitation during a 10 year period (1991-2000), using a reference World Meteorological Organization (WMO) standard data set. This reference data set is the Climatic Normal that contains the long-term means for the period 1961-1990 for Italy. Summer temperatures showed a sharp significant increase starting from 1980, especially for mountain stations. The trend analysis of precipitation showed evident increase during autumn and winter for mountain stations, but for the rest of Italy, precipitation had decreased during early spring.

Jiang et al. (2007) studied the effects of climate warming on precipitation variation in the Yangtze River Basin, China. Their primary data sources are daily precipitation observations measured at 147 stations for 1961-2000 and monthly discharge data observations measured at three stations. They analyzed the temporal trend on precipitation with the Mann-Kendal test, and the spatial trend with simple regression. They found out significant increasing trend on precipitation and rainstorm frequency during the summer season. Additionally they found out significantly increasing trend in flood discharges for the time period of their study. They concluded that trends in precipitation and rainstorms are probably caused by climate warming.

Chen et al. (2007) studied the temporal trends of climatic variables: runoff, annual and seasonal precipitation and temperature with the Mann-Kendal test and linear regression methods for 1951-2003 in the Danjiangkou reservoir basin, China. The spatial distribution of precipitation and temperature were investigated with the inverse distance weighting interpolation method. Also a two-parameter water balance model was used to define effects of climate change on hydrological parameters and make predictions for runoff change in reservoir basin for 2021-2050. According to their results, there was no trend in precipitation values for most parts of the basin whereas a significantly increasing trend was obtained for temperature. A significantly decreasing trend, which is the mutual effect of precipitation and temperature, was obtained in spring, winter and annual runoff values. According to the hydrological two-parameter model, precipitation and temperature would increase for 2021-2050 and runoff for all seasons would increase also. Increasing or decreasing the trend of monthly precipitation of 10% would increase or decrease the mean annual runoff with about 15%.

Aziz and Burn (2006) studied the temporal trend of 19 hydrological and six meteorological variables measured in the Mackenzie River basin, Canada. They used the Mann-Kendal test to reveal probable trend and the Trend Free Pre-Whitening approach to remove autocorrelation from data series. The hydrologic data were measured at 54 stations and meteorological data were measured at 10 stations. The time periods of study were 26, 31, 36 and 41 years duration and ends in 2000. According to their

results, a significantly increasing trend was found out at annual minimum flow over the winter months. Monthly temperature had an increasing trend in winter and spring months. Monthly precipitation had a decreasing trend in fall and winter seasons and an increasing trend in the spring season.

Cannarozzo et al. (2006) studied the temporal trends and spatial distribution of annual, seasonal and monthly precipitation in Sicily, Italy for 1921-2000. In order to determine a possible trend in precipitation, the Mann-Kendal test was applied to data from 250 meteorological stations. Local significance levels of each meteorological station were interpolated using the Inverse Distance Weighting method to analyze trend variation over the Sicily Island. According to their results, increasing precipitation trend was less clear and was found out only at few stations in summer period. A significantly decreasing precipitation trend was more frequently found annually and in the winter season.

The correlation between annual rainfall and elevation over Great Britain was examined using GWR by Brundson et al. (2001). The rate of increase of precipitation with height varies from around 4.5 mm/month in the northwest to almost zero in the southeast regions of Great Britain. As a conclusion, the GWR provides a useful method for incorporating the varying relationship between rainfall and altitude across the country with changes in predicted rainfall amounts.

Hengl et al. (2004) described Regression kriging as a spatial prediction method and compared it with ordinary kriging and plain regression. Dependent variables which were organic matter of soil, pH in topsoil and topsoil thickness were interpolated using six topographic variables and nine soil mapping units. The data set contained measurements at 135 locations in Croatia, which were divided into training and test datasets. Accuracy assessment was made by comparing the mean error (*ME*) and root mean square error (*RMSE*) at prediction points. According to their findings, the *RMSE* of organic matter and topsoil thickness predictions obtained by Regression kriging was smaller than for ordinary kriging. However, topsoil pH was difficult to predict with all three methods due to its weak correlation with the secondary variables.

Lloyd (2005) performed a study related to mapping monthly precipitation in Great Britain from point data using several interpolation methods. The study used the relationship between precipitation and secondary variables, such as elevation, and the aim was to assess if they provide more accurate estimates than methods that do not make use of secondary variables. The techniques applied were: (i) moving window regression (MWR), (ii) inverse distance weighting (IDW); (iii) OK, (iv) simple kriging with a locally varying mean (SKlm) and (v) kriging with an external drift (KED). MWR, SKlm and KED techniques made use of elevation as secondary variable. The relationship between precipitation and elevation was examined for each month of 1999. The performance of each interpolation method was assessed through examination of mapped predictions of precipitation and using cross-validation. It was concluded that KED provides the most accurate estimates of precipitation for all months from March to December whereas for January and February OK provided the most accurate estimates.

Heuvelink (2006) performed a study related with two approaches that incorporate ancillary information and process knowledge in spatial interpolation: Regression kriging and Space-time Kalman filtering. These techniques were compared and their application in practice was illustrated with examples. Incorporating process knowledge in spatial interpolation was advantageous not only because using more information resulted in more accurate maps, but also because it gave insight into how processes affect the state of the environment and because it was better suited to make extrapolations.

Propastin et al. (2006) performed a study concerned with the spatial relationship between vegetation patterns and rainfall and its trend over the period 1985-2001 in desert, semi-desert and steppe grassland of Middle Kazakhstan. The Normalized Difference Vegetation Index (NDVI) images (1985-2001) and measured rainfall data were used in the analyses. Ordinary Least Squares (OLS) regression

technique and GWR techniques were used to understand the relationship between NDVI and rainfall during the growing season. Regression models between NDVI and precipitation for every analysis year (1985-2001) were calculated by using both statistical approaches. The ordinary least squares regression model that had been applied to the whole study area was strong ($r^2 = 0.63$), however it gave no local description of the relationship. The strength of the relationship between NDVI and rainfall increased from desert ($r^2 = 0.36$), to semi-desert ($r^2 = 0.52$), and to steppe grassland ($r^2 = 0.67$). The approach of GWR provided stronger relationships from the same data sets (mean value of $r^2 = 0.88$), as well as highlighted local variations within the land cover classes. The study found out that the relationship between vegetation patterns and rainfall varied significantly in space and time. The application of GWR may reveal local patterns of relationship and significantly reduced the uncertainties of calculations.

Carrera-Hernández and Gaskin (2007) studied the temporal variation of minimum and maximum temperature and rainfall and their relationship with elevation. The aim of the study was to reveal if the relationship between them should be used when daily data are interpolated. The methods used for interpolation were OK, kriging with External Drift (KED), Block kriging with External Drift (BKED), Ordinary kriging in a local neighborhood (OKl) and kriging with External Drift in a local neighborhood (KEDl). The data were obtained from 200 meteorological stations located in the Basin of Mexico for 1978-1985. The result of the study was; the improvement of interpolation of daily events by the use of elevation as a secondary variable even when these variables show a low correlation with precipitation and temperature.

Haberlandt (2007) applied geo-statistical interpolation of hourly precipitation measured from rain gauges using secondary information from radar, a daily precipitation network and elevation. Comparison of KED, indicator KED, nearest neighbor, IDW, OK and ordinary indicator kriging showed that KED performed best when all secondary information was used simultaneously.

Harris et al. (2010) compared MLR, GWR, GWR kriging hybrid (GWRK), OK, UK and UK in a local neighborhood using simulated datasets based on different classes of spatial heterogeneity and autocorrelation. They found out that UK with local neighborhood performed best.

Hiemstra et al. (2010) used weighted average rainfall intensity measured from radar images of rainfall to improve interpolation of dose rates (radiation levels) of the Netherlands. Dose rate was measured every ten minutes at 153 stations in the Netherlands. They used one-day data for this study (20 July 2007) and compared UK with OK. They noticed that precipitation intensity is closely related with dose rates at short time scales. Therefore their hypothesis is that precipitation intensity can be used as a predictor to get a more reliable prediction for dose rates. According to their results, RMSE, ME, MKV (mean kriging variance) obtained from cross-validated data were similar for UK and OK. However, prediction maps of UK had more detail than OK prediction maps.

Bogaert (1996) compared simple and ordinary space-time kriging with simple and ordinary cokriging on simulated data. In order to obtain reliable results, 50 simulations with the same data configuration were performed. Performance assessment was achieved by comparing prediction errors of cross-validated data. Bogaert (1996) concluded that space-time kriging produced smaller prediction error variances and errors than usual kriging and cokriging. Moreover, the space-time variogram was beneficial because it gave the prediction for any point in time and space.

Kyriakidis and Journel (1999) studied and prepared a framework about geostatistical space-time models. They attempted to make a survey of models referred to in the literature by adding contributions and identifying limitations. They described and compared two conceptual approaches: space-time random function models and models that involve vectors of space random functions or vectors of time series. They described the links between these two approaches and their advantages and disadvantages. There are many techniques that handle modeling time series data and spatial modeling; the important challenge is to involve these techniques into space-time environment.

Huerta and et al. (2004) generated a spatio-temporal model for hourly ozone levels of Mexico City. The model was applied to data from September 1997. Air temperature was used as an independent variable. The model contains a set of time-varying Fourier coefficients. Their goal was to generate a model and by means of this model, obtain predictions temporally and spatially for hourly ozone levels. They used Markov Chain Monte Carlo to obtain spatio-temporal predictions. Eventually their results showed that the model produced logical predictions for spatial interpolation for short-term temporal cases.

Jost et al. (2005) analyzed soil water storage by using space-time kriging methods in a forest eco-system in Lower Austria. A physical-deterministic model about evapotranspiration was included as a deterministic trend to improve the results obtained with space-time interpolation methods. The dependent variable was measured at 198 locations every two weeks during growing the seasons of 2000 and 2001. ST-OK was used as a geostatistical interpolation method. According to their findings, using process knowledge of variable in ST geostatistical interpolation improves the accuracy.

Liu and Koike (2007) studied the chemical concentration of the Ariake Sea in Japan. They compared single variable ordinary kriging and multivariable space-time ordinary cokriging to predict four nutritive salts. The input data were measured at 38 stations over 25 years. In cokriging they used a sea-color image obtained from color composite of the visible wavelength band of Landsat-5 Thematic Mapper as a secondary variable. According to their results, multivariate space-time cokriging resulted in better predictions than single variable ordinary kriging. In addition, cokriging provided a more detailed concentration map whereas ordinary kriging resulted with smoother prediction maps.

Heuvelink and Griffith (2010) performed a case study about space-time interpolation of gamma dose rates observed at four states of Germany for a 5-year period (2003-2007). This study emphasizes the main concepts of space-time geostatistics. The main difficulties of space-time kriging are to define a suitable and realistic stochastic model according to data and to estimate the space-time correlation structure of that model. According to their case study results, the prediction error of gamma dose rates was small by using even a fairly simple space-time model. Although space-time interpolation is more demanding and difficult than spatial kriging, it is beneficial to model variations of phenomena that have different characteristics in time and space.

Hengl et al. (2012) performed a space-time interpolation to predict daily temperatures using ancillary variables such as MODIS LST images, coordinates, distance from sea, elevation, time and insolation. The dependent data set consisted of 57,282 meteorological station observations of daily temperatures for 2008. According to their findings, spatio-temporal interpolation could explain 84% part of the variation. In addition, cross-validation results confirmed that using space-time kriging methods and including time-series data sets produce more accurate maps than using only spatial methods.

In order to make a conclusion to literature review part, it can be said that non-parametric Mann-Kendal test has been frequently used by scientists to detect trend at hydrological, environmental and climatic records. It is important issue that serial correlation between time series should not be ignored.

Related with spatial and space-time interpolation, most studies show that including secondary information can improve the accuracy of the interpolation. OK and UK are the most commonly used kriging techniques at spatial and space-time studies by scientists. If the data varies spatially and temporally, it is useful to use space-time interpolation techniques.

8

CHAPTER 3

STUDY AREA AND DATA

3.1   Study Area

The bounding coordinates of Turkey are 25˚E - 42˚N and 45˚E - 36˚N. Turkey is a mountainous country with a mean elevation of around 500 m. Almost one-third (32%) of the area is arable land, 26% is covered with forest and woodland and 16% of the land is permanent pasture. Turkey's varied landscapes are the product of a wide variety of tectonic processes that have shaped Anatolia over millions of years and continue today as evidenced by frequent earthquakes and occasional volcanic eruptions. Except for a relatively small portion of its territory along the Syrian border that is a continuation of the Arabian Platform, Turkey geologically is part of the great Alpine belt that extends from the Atlantic Ocean to the Himalaya Mountains. The coastal areas of Turkey bordering the Aegean Sea and the Mediterranean Sea have a temperate Mediterranean climate, with hot, dry summers and mild to cool, wet winters. The coastal areas of Turkey bordering the Black Sea have a temperate Oceanic climate with warm, wet summers and cool to cold, wet winters. The Turkish Black Sea coast receives the greatest amount of precipitation and is the only region of Turkey that receives high precipitation throughout the year. The eastern part of that coast averages 2,200 mm precipitation annually which is the highest precipitation in the country.

The precipitation data used in this study were obtained from the Turkish State Meteorological Service. The primary dependent data source is monthly precipitation measured at 225 meteorological stations between 1970 and 2008. From this dataset, all 225 stations are used in the temporal precipitation analysis for 1970-2003, all 225 stations are used in spatial precipitation analyses for 1975-2006, and 47 numbers of stations are used for 1970-2008 years in the space-time precipitation analysis. Figure 3.1 shows the locations of the stations and the long–term average annual precipitation values. The stations are fairly uniformly distributed across the country, with a somewhat higher spatial density along the coastlines of the Mediterranean and Black Sea. Precipitation is high at the north and south coasts of Turkey. The highest precipitation (2209.3 mm) was observed in the north-east, where Turkey borders the Black Sea. At the centre of Turkey, very low precipitation values were observed (about 255-400 mm). Somewhat higher precipitation values were measured near the Van Lake, Ataturk and Keban Dams in the east. Summary statistics of the long-term annual precipitation are given in Table 3.1.The average annual precipitation of Turkey is 628.2 mm.

Table 3.1. Summary statistics of average annual precipitation of Turkey.

Mean	Std. dev.	Min.	Median	Max.
628.2 mm	285.9 mm	255.1 mm	567.6 mm	2209.3 mm
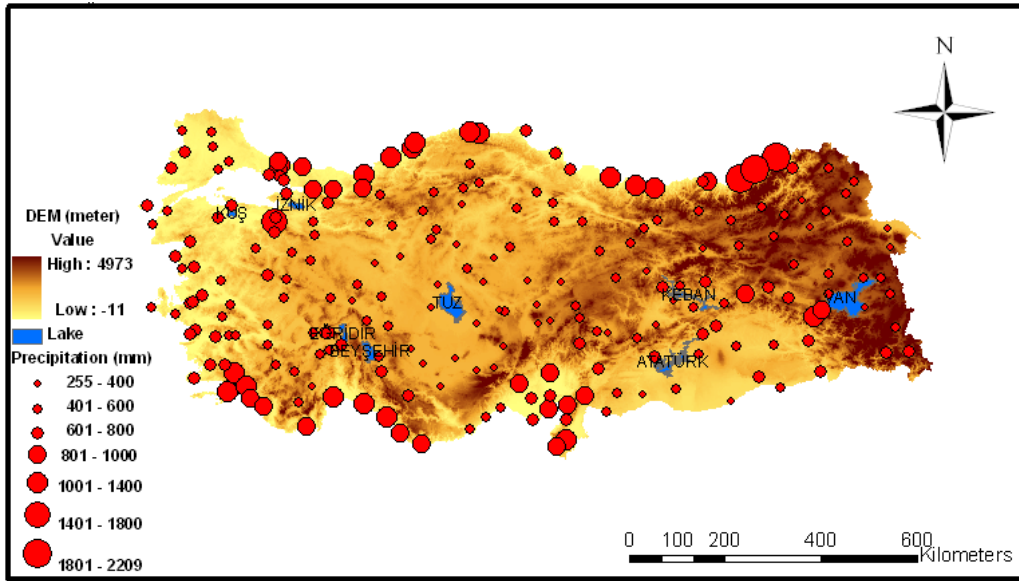
Figure 3.1. Average annual precipitation of Turkey, measured at 225 meteorological stations from 1970 to 2006.

Figure 3.2 gives the histogram of the annual precipitation and its log-transform. Log-transformation removed the skewness and the transformed data are reasonably symmetrically distributed and have no suspicious outliers. It was therefore decided to apply the spatial and space-time interpolation methods to the log-transformed precipitation data. The interpolation results were back-transformed to the original scale of measurement. Back-transform for predictions was performed as follows:

Mean predictions = 10^(lognormal kriging predictions+ (0.5×lognormal kriging variances))

Back-transform for kriging variance was performed as follows:

Kriging variance = (mean predictions^2) × (10^(sill of the variogram of the transformed variable)) × (1-(10^(-kriging variance of the transformed variable)))
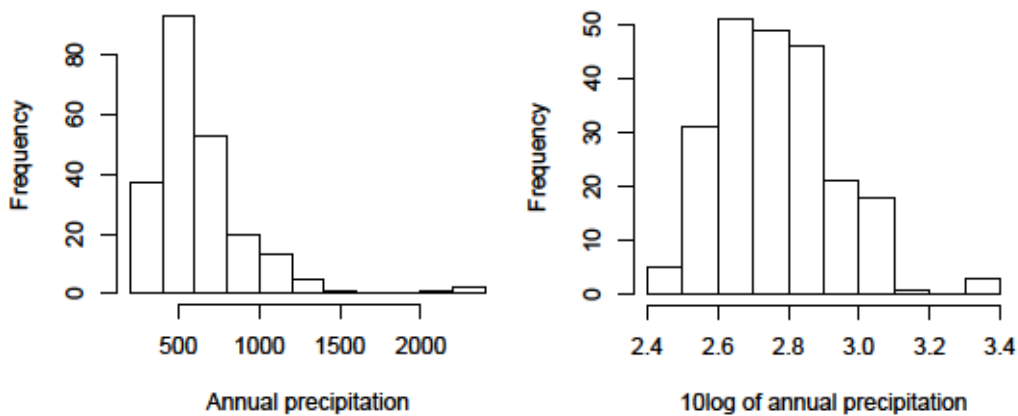
Figure 3.2. Histogram of original (left) and log-transformed (right) annual precipitation data (mm).

3.2   Temporal Data

In temporal precipitation analysis, the mean monthly observations taken from 225 meteorological stations for 1970-2003 were used. Trend tests were applied to seasonal and annual precipitation values. T-Test and Mann-Kendal tests were applied to annual precipitation values. For seasonal values, only the Mann-Kendal test was applied. The number of meteorological stations in each river basin is listed in Table 3.2.

Table 3.2. Number of stations in each river basin.

Number of stations	River basin
11	Marmara
10	Batı Karadeniz
5	Meriç
17	Kızılırmak
9	Yeşilırmak
7	Aras
9	Doğu Karadeniz
5	Çoruh
14	Sakarya
8	Kuzey Ege
7	Susurluk
29	Fırat
12	Konya Kapalı
6	Van Gölü
4	Gediz
7	Seyhan
3	Akarçay
9	Büyük Menderes
9	Dicle
7	Ceyhan
6	Küçük Menderes
8	Antalya
2	Burdur Göller
10	Batı Akdeniz
6	Doğu Akdeniz
5	Asi
Total 225 station	

It is useful to discuss the river basins of Turkey as these are frequently used in the trend analysis part of this thesis. The Ministry of Forestry and Water divides Turkey in 26 big river basins as can be seen in Figure 3.3.

11

Figure 3.3. Big river basins of Turkey (URL 1).

The productivities of the river basins differ from each other. The most efficient ones are Firat (Euphrates) and Dicle (Tigris) that have 28.5% of the total country water potential. Some descriptive information about the river basins is presented in Table 3.3.

Table 3.3. Summary statistics about the big river basins in Turkey
(Arslan-Alaton et al. 2005).

Batı Akdeniz:		Antalya:	
Population:	1.066.630	Population:	1.882.851
Area:	20.95 km^2	Area:	19.577 km^2
Population density:	51	Population density:	96
(# person/area)		(# person/area)	
Mean height:	894 m	Mean height:	1021 m
Max. height:	2953 m	Max. height:	2805 m
Mean annual runoff:	8.93 km^3	Mean annual runoff:	11.06 km^3
Poten. participation ratio:	4.8 %	Poten. participation ratio:	5.9 %
Doğu Akdeniz:		**Seyhan:**	
Population:	1.768.047	Population:	1.544.830
Area:	22.048 km^2	Area:	20.450 km^2
Population density:	80	Population density:	76
(# person/area)		(# person/area)	
Mean height:	2265 m	Mean height:	1329 m
Max. height:	3351 m	Max. height:	3609 m
Mean annual runoff:	11.07 km^3	Mean annual runoff:	8.01 km^3
Poten. participation ratio:	6.0 %	Poten. participation ratio:	4.3 %
Ceyhan:		**Asi:**	
Population:	2.286.178	Population:	1.332.73
Area:	21.98 km^2	Area:	7.796 km^2
Population density:	104	Population density:	171
(# person/area)		(# person/area)	
Mean height:	974 m	Mean height:	556 m
Max. height:	2955 m	Max. height:	2127 m
Mean annual runoff:	7.18 km^3	Mean annual runoff:	1.17 km^3
Poten. participation ratio:	3.9 %	Poten. participation ratio:	0.6 %

Table 3.3. Summary statistics about the big river basins in Turkey
(Arslan-Alaton et al. 2005) (continued).

Kuzey Ege:		**Gediz:**	
Population:	751.113	Population:	1.581.398
Area:	10.00 km^2	Area:	18.00 km^2
Population density:	75	Population density:	88
(# person/area)		(# person/area)	
Mean height:	302 m	Mean height:	577 m
Max. height:	1746 m	Max. height:	2273 m
Mean annual runoff:	2.09 km^3	Mean annual runoff:	1.19 km^3
Poten. participation ratio:	1.1 %	Poten. participation ratio:	1.1 %
Küçük Menderes:		**Büyük Menderes:**	
Population:	3.142.603	Population:	1.929.397
Area:	6907 km^2	Area:	24.97 km^2
Population density:	455	Population density:	77
(# person/area)		(# person/area)	
Mean height:	289 m	Mean height:	807 m
Max. height:	2057 m	Max. height:	2309 m
Mean annual runoff:	2.09 km^3	Mean annual runoff:	3.03 km^3
Poten. participation ratio:	1.1 %	Poten. participation ratio:	1.6 %
Meriç-Ergene		**Marmara**	
Population:	980.905	Population:	12.481.311
Area:	14.56 km^2	Area:	24.100 km^2
Population density:	67	Population density:	518
(# person/area)		(# person/area)	
Mean height:	154 m	Mean height:	213 m
Max. height:	909 m	Max. height:	894 m
Mean annual runoff:	1.33 km^3	Mean annual runoff:	8.33 km^3
Poten. participation ratio:	0.7 %	Poten. participation ratio:	4.5 %
Susurluk		**Burdur Göller**	
Population:	2.637.131	Population:	292.840
Area:	22.39 km^2	Area:	6374 km^2
Population density:	118	Population density:	46
(# person/area)		(# person/area)	
Mean height:	638 m	Mean height:	1182 m
Max. height:	2500 m	Max. height:	2620 m
Mean annual runoff:	5.43 km^3	Mean annual runoff:	0.50 km^3
Poten. participation ratio:	2.9 %	Poten. participation ratio:	0.3 %
Akarçay		**Sakarya**	
Population:	500.979	Population:	6.101.234
Area:	7605 km^2	Area:	58.160 km^2
Population density:	66	Population density:	105
(# person/area)		(# person/area)	
Mean height:	1212 m	Mean height:	966 m
Max. height:	2453 m	Max. height:	2396 m
Mean annual runoff:	0.49 km^3	Mean annual runoff:	6.40 km^3
Poten. participation ratio:	0.3 %	Poten. participation ratio:	3.4 %

Table 3.3. Summary statistics about the big river basins in Turkey
(Arslan-Alaton et al. 2005) (continued).

Batı Karadeniz		**Yeşilırmak**	
Population:	1.959.308	Population:	3.003.142
Area:	29.59 km^2	Area:	36.114 km^2
Population density:	66	Population density:	83
(# person/area)		(# person/area)	
Mean height:	809 m	Mean height:	1142 m
Max. height:	2250 m	Max. height:	3292 m
Mean annual runoff:	9.93 km^3	Mean annual runoff:	5.80 km^3
Poten. participation ratio:	5.3 %	Poten. participation ratio:	3.1 %
Kızılırmak		**Konya Kapalı**	
Population:	4.167.766	Population:	3.048.395
Area:	78.2 km^2	Area:	53.850 km^2
Population density:	53	Population density:	57
(# person/area)		(# person/area)	
Mean height:	1178 m	Mean height:	1205 m
Max. height:	3761 m	Max. height:	3296 m
Mean annual runoff:	6.48 km^3	Mean annual runoff:	4.52 km^3
Poten. participation ratio:	3.5 %	Poten. participation ratio:	2.4 %
Fırat		**Dicle**	
Population:	6.910.866	Population:	3.349.716
Area:	127.3km^2	Area:	57.614 km^2
Population density:	54	Population density:	58
(# person/area)		(# person/area)	
Mean height:	1388 m	Mean height:	1468 m
Max. height:	3596 m	Max. height:	3643 m
Mean annual runoff:	31.61 km^3	Mean annual runoff:	21.33 km^3
Poten. participation ratio:	17.0 %	Poten. participation ratio:	11.5 %
Doğu Karadeniz		**Çoruh**	
Population:	2.882.208	Population:	432.259
Area:	24.07km^2	Area:	19.872 km^2
Population density:	120	Population density:	22
(# person/area)		(# person/area)	
Mean height:	1126 m	Mean height:	1890 m
Max. height:	3702 m	Max. height:	3687 m
Mean annual runoff:	14.9 km^3	Mean annual runoff:	6.30 km^3
Poten. participation ratio:	8.0 %	Poten. participation ratio:	3.4 %
Aras		**Van Gölü Kapalı**	
Population:	808.570	Population:	3.349.716
Area:	27.5 km^2	Area:	57.614 km^2
Population density:	29	Population density:	58
(# person/area)		(# person/area)	
Mean height:	1973 m	Mean height:	2083 m
Max. height:	5054 m	Max. height:	4019 m
Mean annual runoff:	4.63 km^3	Mean annual runoff:	2.39 km^3
Poten. participation ratio:	2.5 %	Poten. participation ratio:	1.3 %

3.3   Spatial Data

In the spatial interpolation part of this thesis, an elevation map with 5 km spatial resolution was used (Figure 3.1). It was obtained by resampling the 3 arc second SRTM (the Shuttle Radar Topography Mission) (approximately 90 m spatial resolution) to 1 km spatial resolution first and then resampling again to a 5 km SRTM using the Nearest Neighbor algorithm.

As mentioned, secondary information can often improve the spatial interpolation of environmental variables. Secondary variables used in this study are given in Table 3.4, with the sub-classes of the categorical secondary variables specified in Table 3.5. In the implementation of the spatial prediction methods the categorical variables V4, V5 and V6 were converted to dummy variables which were presented as 0 or 1 to present the presence or absence of each category.

Table 3.4. Secondary variables used in the spatial interpolation.

Variable	Code	Description
Elevation map	Z	5 km spatial resolution elevation map of Turkey
Surface roughness	V1	Standard deviation of 1 km spatial resolution SRTM elevations in each 5×5 km grid cell
Distance to nearest coast	V2	Euclidean distance from each point to the nearest coast
River density	V3	5 km spatial resolution river density
Aspect	V4	four directional classes were defined over Turkey
Land cover	V5	Corine land cover database was used, aggregated to six classes
Eco-region	V6	Terrestrial eco-region classes were used, aggregated to eight classes

Table 3.5. Categorical variables and their class values.

V4-Aspect	V41- North-east
	V42- South-east
	V43- South-west
	V44- North-west
V5- Land cover	V50- Artificial surfaces
	V51- Agricultural
	V52- Wetlands
	V53- Open space
	V54- Vegetation
	V55- Forest
V6- Eco-region	V61- Anatolian conifer and deciduous mixed forests
	V62- Balkan mixed forests, Aegean and Western Turkey sclerophyllous and mixed forests
	V63- Caucasus mixed forests
	V64- Central Anatolian steppe and woodlands
	V65- Conifer-sclerophyllous-broadleaf forests, steppe, montane conifer and deciduous forests
	V66- Eastern Anatolian deciduous forests
	V67- Eastern Anatolian montane steppe, Zagros Mountains forest steppe
	V68- Euxine-Colchic broadleaf forests, Northern Anatolian conifer and deciduous forests

Maps of the secondary variables are given in Figure 3.4. Surface roughness is obtained from the 3 arc-second SRTM data of Turkey. Distance to nearest coast variable is obtained by calculating the Euclidean distances of each SRTM pixel to the nearest boundary of the sea coast vector. River density variable is calculated using the "kernel density" function of ArcMap 9.3 (http://www.esri.com/software/arcgis/arcgis-for-desktop). A 30 km search function and 5 km output resolution were used while calculating this variable. The aspect variable is calculated from the SRTM elevation. The original aspect is re-classified and grouped in four classes: north-east, south-east, south-west and north-west. The land-cover map is obtained from the CORINE database of Turkey. The Eco-region variable is extracted from the terrestrial eco-region database of the Earth created by http://www.worldwildlife.org.
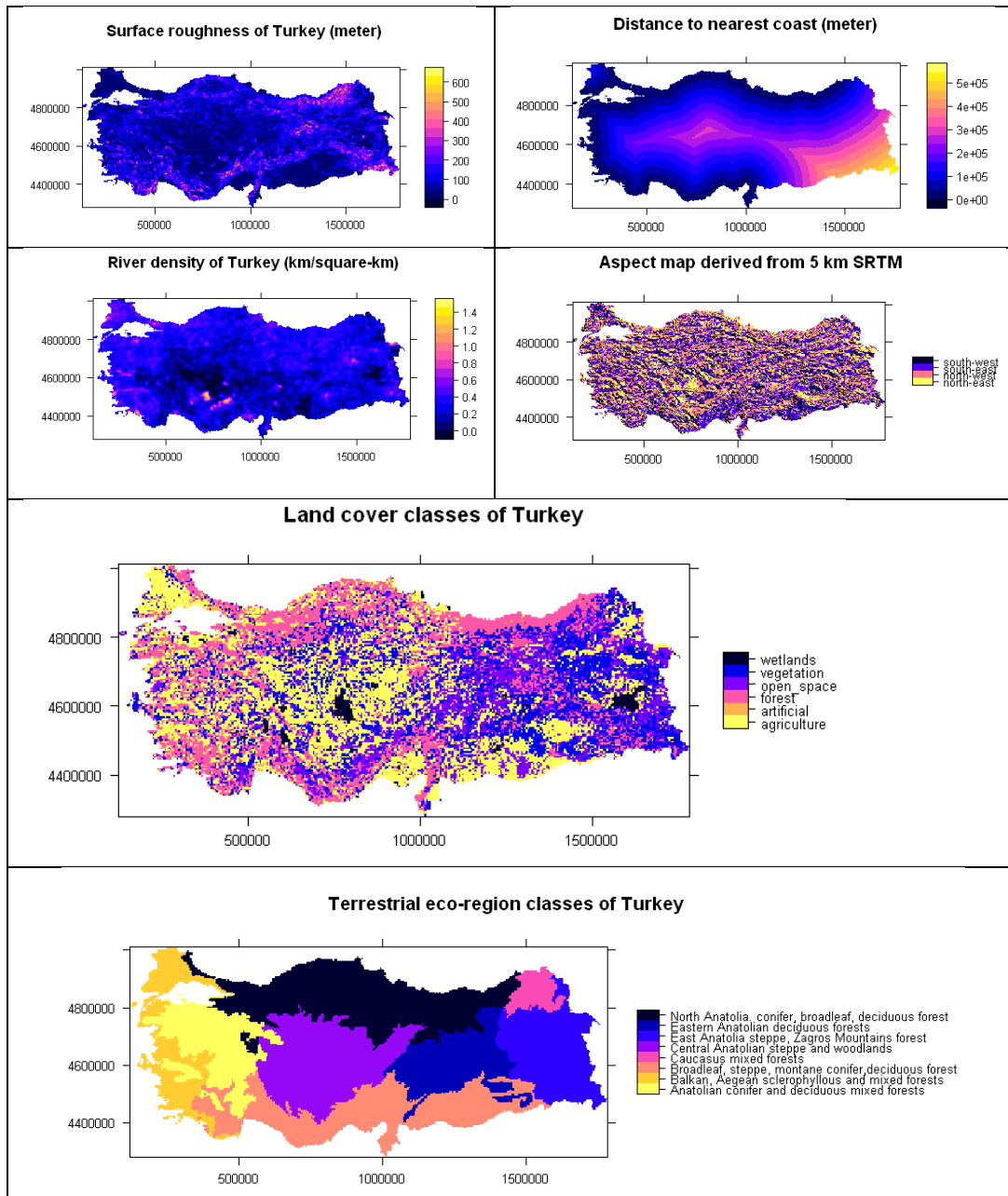
Figure 3.4. Maps of the secondary variables.

16

3.4 Space-time Data

The study area of space-time interpolation is the Euphrates Basin which is the biggest basin of Turkey (Figure 3.5). The basin is located in the south-east of Turkey and it has 17 % of the water potential of the country. Atatürk, Karakaya and Keban dams which are located in this basin are important with respect to water supply, irrigation and hydro-electric power generation (Yilmaz et al. 2011). The bounding coordinates of the basin are 36.7˚W – 43.9˚E – 40.4˚N – 36.3˚S. The northern parts of the basin are surrounded with high mountains whereas the southern parts are flatter. The area of the basin is about 127,000 km^2. The maximum elevation of the basin is 3747 m; minimum elevation is 284 m and mean elevation is 1380 m. The space-time data comprises annual precipitation values between 1970 and 2008. Annual precipitation values were calculated from observed monthly values obtained from the Turkish State Meteorological Service. Data were observed at 47 meteorological stations and in total 906 observations were used in space-time interpolation.
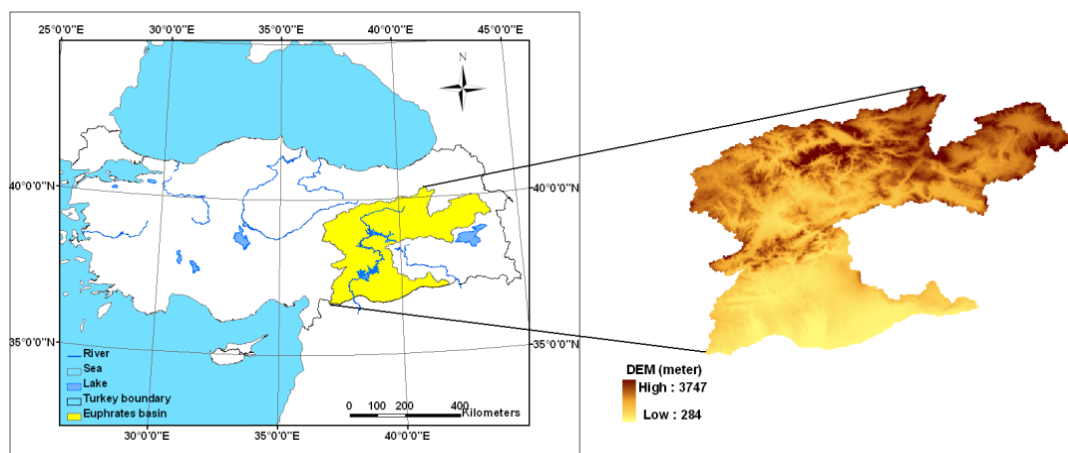
Figure 3.5. Turkey and delineation "Euphrates Basin" from SRTM.

The spatial distribution of the stations is fairly uniform over the basin but a bit condensed placement can be seen near the dams (Figure 3.6). The average long-term annual precipitations of the 47 meteorological stations are presented in Figure 3.6. According to this figure, the long-term averages of stations are higher at stations near the dams than for other stations. The highest total annual precipitation (about 1580 mm) is observed in the central-eastern parts near the Keban dam at the Bingol meteorological station in 1987.
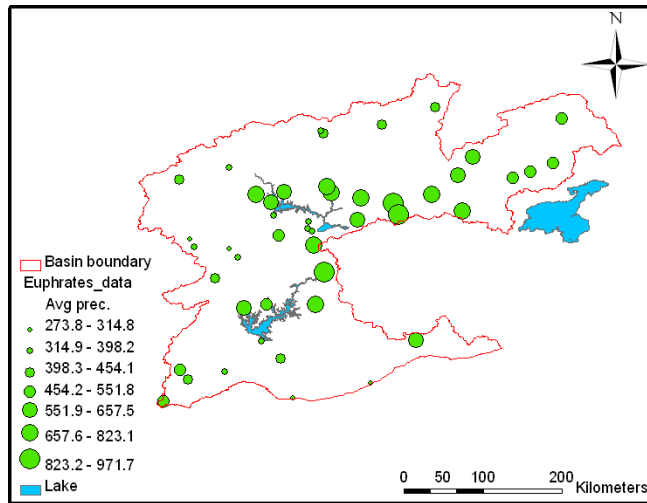
Figure 3.6. Distribution of meteorological stations and their proportional precipitation values (mm) over the Euphrates Basin.

The space-time dataset has missing values for some years for some stations. 21 stations have data for nine years or less. Five stations have observations between 10 and 20 years. The other 21 stations have more than 20 years with observations. The total space-time dataset comprises 906 observations. The same logarithmic transformation that was used for the dataset from whole Turkey is applied to the space-time precipitation data from the Euphrates Basin (Figure 3.7).
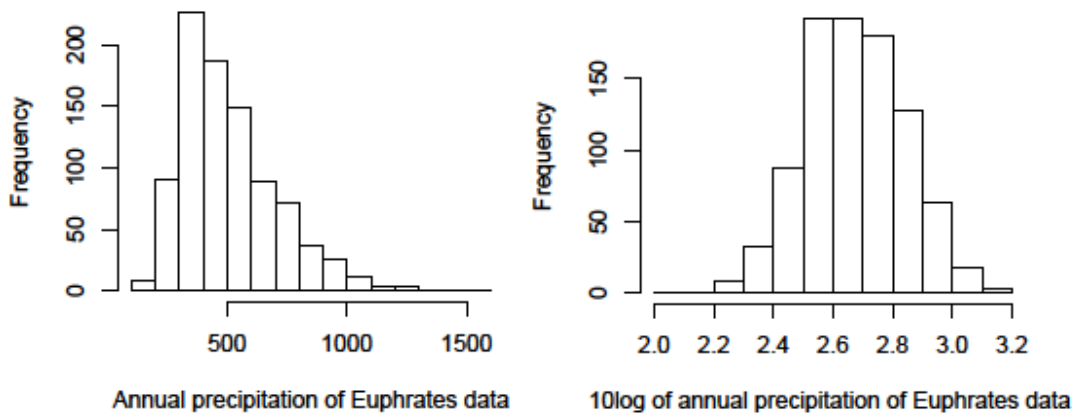
Figure 3.7. Histogram of original (left) and log-transformed (right) annual precipitation data of the Euphrates Basin (mm).

The secondary variables used in space-time interpolation are the same as those used with spatial kriging. Only the "Year" variable is added (Table 3.6). This variable expresses the observation time of each record in the database. The variable "eco-region" has only three classes (V65, V66 and V67) over the Euphrates Basin. So this variable does not have full characteristics for the study area. Therefore it is decided to exclude this variable from space-time interpolation.

18

Table 3.6. Explanatory variables used in the space-time geostatistical analysis.

Variable	Code	Description
Elevation	Z	5 km spatial resolution elevation map of Turkey
Surface roughness	V1	Standard deviation of 1km spatial resolution SRTM elevations in each 5×5 km grid cell
Distance to nearest coast	V2	Euclidean distance
River density	V3	5 km spatial resolution river density
Aspect	V4	four directional classes were defined over Turkey
Land cover	V5	Corine land cover database was used, aggregated to six classes
Year	Y	Measurement time of each data

Histograms of continuous secondary variables of Euphrates Basin are presented in Figure 3.8.
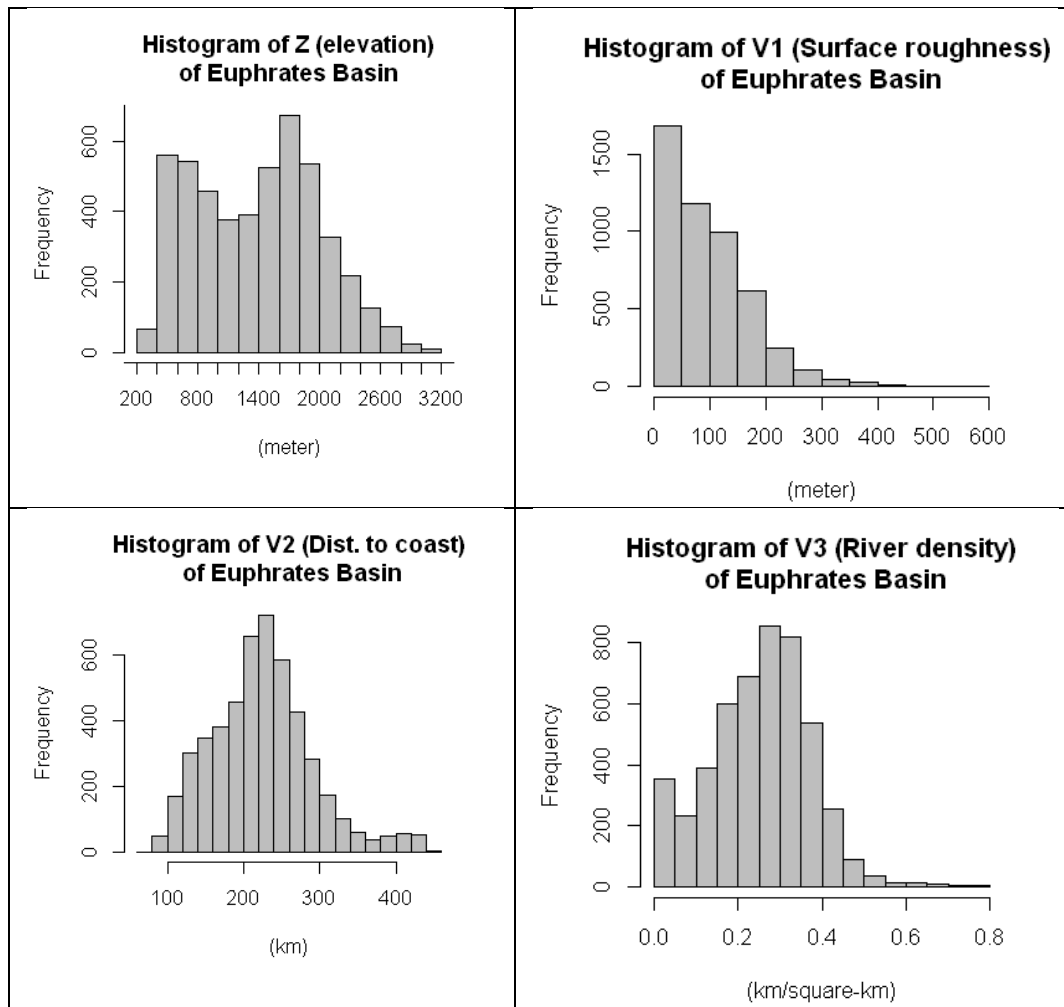
Figure 3.8. Histograms of continuous secondary variables of Euphrates Basin

The secondary variables that are used in space-time interpolation are given in Figure 3.9. All variables have 5 km spatial resolution.
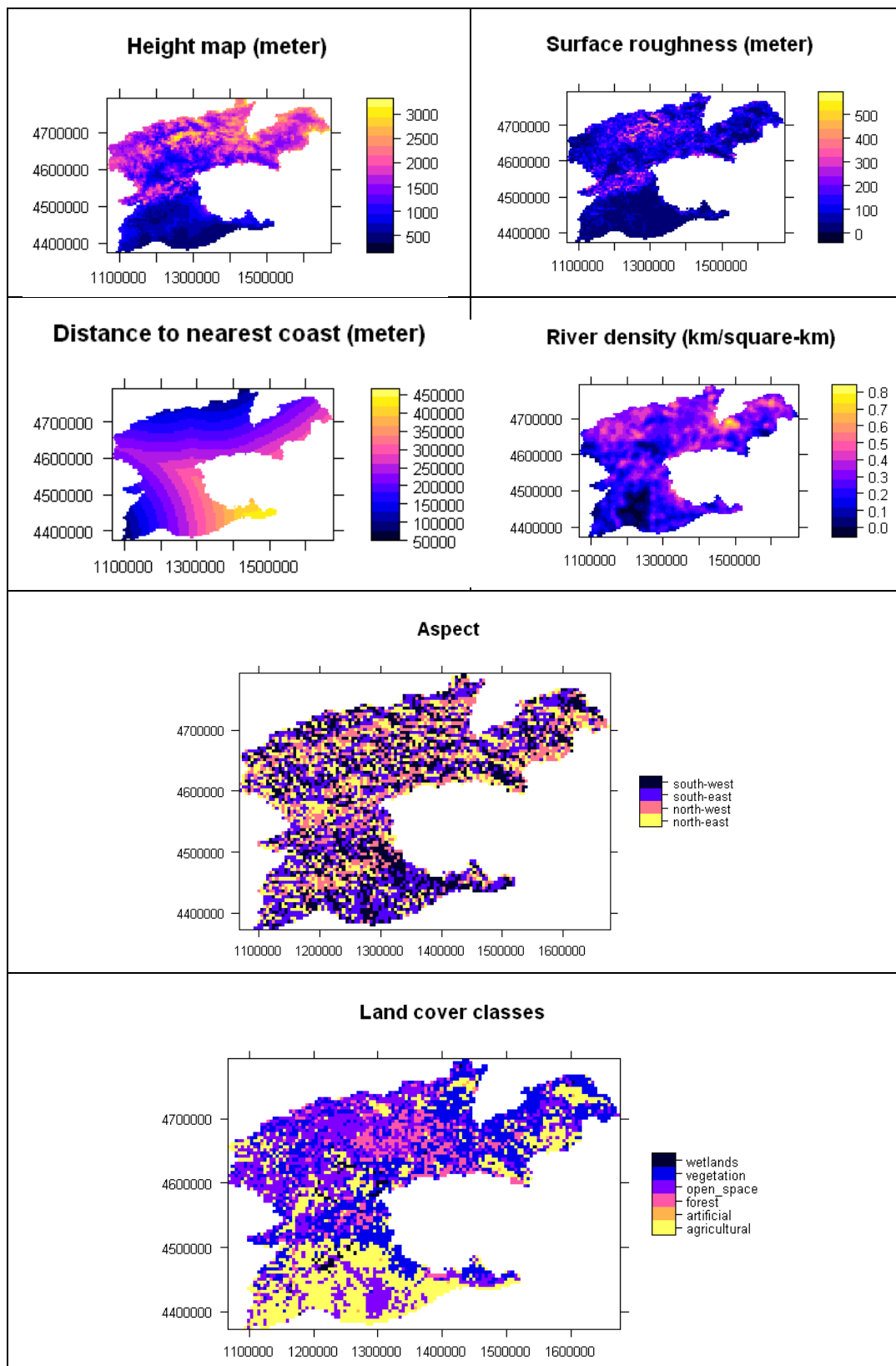
Figure 3.9. Secondary variables of space-time analysis for the Euphrates Basin.

In order to understand if the stations identify the properties of the basin, attribute table is prepared and presented in Table 3.7.

Table 3.7. Secondary variables of each meteorological station used in space-time analyses at Euphrates Basin.

Stat_no	Z	V1	V2	V3	V4	V5	# of data years
4553	1650	15.6125	269.06	0.397133	3	0	20
4896	1500	76.0358	256.969	0.289347	4	1	8
5711	1086	88.3791	221.839	0.176949	3	3	4
9002	1300	129.901	252.469	0.025423	4	4	8
17092	1154	20.8943	140.97	0.378972	3	1	5
17094	1216	20.8943	137.775	0.311366	3	0	39
17096	1758	9.82887	131.72	0.367959	2	0	37
17099	1646	12.705	231.607	0.539598	4	0	32
17165	981	151.533	206.781	0.235813	1	0	39
17199	950	36.3536	245.549	0.250067	1	0	39
17200	849	28.6686	244.879	0.370777	2	0	4
17201	976	98.8473	260.537	0.060667	1	1	39
17202	881	44.1156	265.389	0.183377	3	0	5
17203	1177	78.676	227.584	0.359429	1	0	21
17204	1322	26.3415	265.72	0.322744	1	0	18
17260	700	13.6519	113.328	0.165217	2	1	5
17261	859	28.7244	106.353	0.089346	2	0	39
17262	641	44.518	80.3018	0.066744	2	1	39
17265	627	23.1609	212.748	0.138945	3	0	8
17270	553	35.2361	233.433	0.001104	3	0	39
17275	1040	131.519	403.413	0.241796	2	3	39
17718	1425	94.2311	127.409	0.325278	3	0	37
17734	1225	178.485	172.763	0.441005	4	0	25
17736	1400	156.009	213.501	0.198954	4	3	7
17740	1715	32.5818	208.863	0.73451	1	0	29
17762	1512	31.225	199.377	0.297526	4	1	22
17764	1200	90.0572	206.984	0.386491	3	4	16
17766	900	78.6145	219.363	0.035082	3	3	7
17768	953	133.775	208.637	0.351807	3	1	7
17774	1090	40.6151	216.382	0.325554	3	1	8
17776	1366	98.6064	229.527	0.257921	4	1	24
17778	1650	67.9799	221.572	0.504202	4	5	30
17780	1565	39.182	262.034	0.200047	4	1	25
17804	808	82.9646	236.207	0.15357	4	1	14
17806	1000	97.4667	244.35	0.352476	4	1	9
17808	1250	47.929	243.19	0.410297	4	4	9

Table 3.7. Secondary variables of each meteorological station used in space-time analyses (continued).

Stat_no	Z	V1	V2	V3	V4	V5	# of data years
17842	1123	77.6293	217.851	0.265494	4	3	12
17843	1225	183.659	262.556	0.311092	2	1	5
17844	1240	281.292	282.103	0.159616	2	2	5
17872	1280	49.4395	206.255	0.362791	2	2	5
17874	700	80.5823	314.035	0.219979	3	5	3
17910	675	24.7479	240.594	0.153297	2	0	1
17912	801	22.4683	296.459	0.230325	3	0	4
17944	600	40.7911	215.271	2.48E-06	3	1	9
17966	347	25.4232	160.699	0.324594	3	0	39
17968	365	6.94393	343.003	0.219383	4	0	34
17980	363	4.57347	244.474	0.076092	1	0	33

"Stat_no" presents the number of meteorological station, "Z" presents elevation of that station with meter measurement unit, "V1" presents surface roughness in kilometer, "V2" presents distance to coast in kilometer, "V3" presents river density with measurement unit kilometers per square kilometer, "V4" presents aspect, "V5" presents land cover, "# of data years" presents how many years that station has observations. Average elevation of stations is 1077 m, average surface roughness is 67 km, average of distance to coast is 223 km, and average river density is 0.26 km /square km. Mode of land cover variable is artificial surfaces (V50). Maximum elevation of stations is 1758 m. 37 out of 47 meteorological stations are located at lower elevation than average elevation of basin (1380 m).

CHAPTER 4

METHODOLOGY

4.1 Trend Analysis Methodology

Time series data are observations measured regularly over time. Time series analysis or trend analysis investigates the behavior of time series data over time. From the meteorological variables, precipitation is one of the time series data as it is observed at the same meteorological stations over a long period of time. The main aim of the trend analysis part of the thesis is to determine whether there are increasing or decreasing trends in the annual and seasonal precipitation values.

Trend analysis is performed with parametric T-Test and non-parametric Mann-Kendal tests. The T-Test assumes that the variable is normally distributed. The Mann-Kendal test is non-parametric and does not make any assumption about the distribution of variable.

For mean annual precipitation values both the T-Test and Mann-Kendal tests were applied. For seasonal values only Mann-Kendal test was applied. Serial correlation at time series data has been removed and trend tests were applied to residual terms again. The significance level of 5 % is used to accept or reject the null hypothesis.

4.1.1 T-Test

T-Test is a parametric statistical trend test that assumes that the data has a normal distribution (Onoz and Bayazit, 2003, Novotny and Stefan, 2007). By measuring the Pearson correlation coefficient between observations and time, the t statistic is computed as follows:

$$t = \frac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}}$$
(4.1)

Where;
r is correlation coefficient, n is sample size.

The correlation coefficient r is a measure of dependence between two variables: in this case observations and time. It can take any value between -1 and 1. The t value computed using r is compared with the t-distribution and is used to test the null hypothesis (H_0) which states that there is no temporal trend.

4.1.2 Mann-Kendal Test

The Mann-Kendal test is a rank-based non-parametric test to evaluate the trend existence and its significance and has been frequently used to detect trend in time series data (Yue et al. 2002). The Mann–Kendal test is based on the null hypothesis that the sampled data are independent and identically distributed, which means that there is neither trend nor serial correlation among the data points. The alternative hypothesis is that a trend exists in the data. The first step in the Mann-Kendal method is to calculate a statistic defined by the variable S, which is the sum of the difference between the data points shown in the Equation (4.2) below:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Sgn(x_j - x_i)$$ (4.2)

where n is the number of values in the data, and x_j and x_i are the sequential data values. The sign of the value is determined as follows:

$$Sgn(\theta) = \begin{cases} +1, \theta > 0 \\ 0, \theta = 0 \\ -1, \theta < 0 \end{cases}$$ (4.3)

When n ≥ 8 the statistic S is approximately normally distributed with the mean and variance (corrected for ties) as follows in Equation (4.4):

$$E(S) = 0$$

$$V(S) = \frac{n(n-1)(2n+5) - \sum_{m=1}^{n} t_m m(m-1)(2m+5)}{18}$$ (4.4)

Where t_m is the number of ties of extent m. Tie (t) presents the group of observations in a time series have the same value. For example there is one tie in a time series with extending three. This means that three observations have same value in dataset. The normally distributed S statistic allows for the computation of the standardized test statistic (Z) and the corresponding ρ value of the Mann–Kendal test. The Z statistics is calculated as presented in Equation (4.5):

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{V(S)}}, S > 0 \\ 0, S = 0 \\ \dfrac{S+1}{\sqrt{V(S)}}, S < 0 \end{cases}$$ (4.5)

The standardized Z statistic has normal distribution, its mean is zero and variance is one. A positive or negative value of Z presents an upward or downward trend, respectively (Novotny, Stefan, 2007). The probability value P of Mann-Kendal statistic is calculated by using normal cumulative distribution function as follows:

$$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2 / 2} dt$$

(4.6)

If the probability value P is close to 0.5 then this means that the data do not have a trend. If the data have a positive trend then P is close to 1.0, in case of a negative trend P is close 0.0 (Yue et al. 2002).

4.1.3 Serial Correlation Analysis

The time series data may have correlation between consecutive values over time. This is referred to as serial correlation. Values or residuals of time series at a particular time may be related with those of a previous time. Examples from environmental science are stream flow, temperature and precipitation.

Yue et al. (2002) state that if there is a positive significant serial correlation in time series, then the probability to find a significant positive trend of Mann-Kendal method increases. Therefore in order to calculate trend existence accurately in time series the serial correlation should first be removed. The autocorrelation or serial correlation coefficients are the most commonly used parameters in order to define dependency between time series data. The lag-k serial correlation coefficient r_k of the sample data x_i (i=1,...n) is calculated as follows in Equation (4.7):

$$r_k = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(x_{i+1} - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{4.7}$$

Where;

x_i and x_{i+1} are observation values at times i and $i+1$, and $\bar{x}$ is the mean of the time series data.

The $x_i - (r_1 \times x_{i-1})$ operation is applied to the original observations x_i by using the lag-1 serial correlation coefficient "r_1". The obtained data series is referred to as residual terms and the trend significance tests are applied to these residual terms (Yucel et al. 1999).

4.2 Spatial Interpolation Methodology

As mentioned in Chapter 1, measurements of environmental, hydrological, agricultural and similar studies are based on point observations over the Earth. Precipitation values are measured from meteorological stations, soil characteristics are measured from soil samples, and pollution of a lake is measured by taking samples from lake. These are some examples from spatial point measurements. It is impossible to measure a variable at all parts of globe. Instead of this scientists prefer to make some interpolation to map spatial distributions of that variable.

Most spatial prediction methods have been based on mathematical, geometry and some approval of the physical nature of the phenomena. Almost all kriging methods calculate predictions based on weighted averages of data. The general kriging prediction formula is given in Equation 4.8:

$$\hat{z}(s_0) = \sum_{\alpha=1}^{n} \lambda_\alpha z (s_\alpha) \tag{4.8}$$

Where $\hat{z}(s_0)$ is target point which we want to get prediction on it, the $z(s_\alpha)$ is the observation at location α, n is the number of observations and λ_α are the weights for each observation (Webster&Oliver 2007, pp 37).

The underlying statistical theory of kriging enables to quantify accuracy of predictions by means of kriging variance. Kriging variance is a measure of uncertainty about true values (Knotters et al., 2010).

In this section methodologies of five regression and kriging techniques were described. MLR and GWR are regression based interpolation techniques. OK, RK and UK methods are based on geostatistical kriging. Secondary variables were used at MLR, GWR, RK and UK methods. OK

method does not require secondary variables. For the selection of secondary variables to be included in MLR, a step-wise forward selection approach was adopted. First all variables were incorporated in MLR one by one and the most significant included. Next variables were added in a similar procedure as long as these were significant at the 5% significance level. Two-way interactions were also included when significant.

4.2.1   Multiple Linear Regression

MLR is a regression model that uses linear relationships between response and explanatory variables. The response variable is also called the dependent while the explanatory or secondary variables are the independent variables. The model equation is defined as follows in Equation (4.9):

$$Z(s) = \beta_0 + \sum_{k=1}^{p} \beta_k . X_k(s) + \varepsilon(s) \qquad (4.9)$$

where s is location, $Z(s)$ is the dependent variable at location s, β_0 is the intercept term, β_1 to β_n are regression coefficients of the independent variables, X_1 to X_n are the values of 1 to p independent variables, and where $\varepsilon(s)$ is the residual term at location s, which is assumed normally distributed and uncorrelated.

The regression model is estimated by least squares so that the sum of squares of the differences between observed and predicted values are minimized. For detailed information, see e.g. Sheather (2009).

4.2.2   Geographically Weighted Regression

In GWR the multiple linear regression model is retained, but the difference with MLR is that the intercept and regression coefficients can take different values for each prediction point. In other words, they are not constant but vary in geographic space. To make the model written in explicit way as:

$$Z(s) = \beta_0(s) + \sum_{k=1}^{p} \beta_k(s) . X_k(s) + \varepsilon(s) \qquad (4.10)$$

With this revised model, local variability of the coefficients can be handled over space. For example, the effect of elevation on precipitation may not be the same everywhere over the study area. By applying GWR, local differences can be included (Fotheringham et al. 2002).

4.2.3   Modeling of Variogram

The variogram model of data is estimated by taking half of the squared distance between observations at given spatial lag as presented in Equation (4.11) (Snepvangers et al. 2003, Heuvelink and Griffith, 2010, Gething et al. 2007):

$$\gamma(x_i, x_j) = \frac{1}{2}[(z(x_i) - z(x_j))^2] \qquad (4.11)$$

These values can be plotted against spatial lag distance, and this plot is called "variogram cloud" which shows the spread of values at different lag. In principle, it is possible to fit a model by using variogram cloud, but in practice it may be completely impossible because if there is spatial correlation, it is difficult to evaluate from variogram cloud. Instead of using all variables, it is

26

preferable to take averages of semi-variances at each lags. For a set of data $z(x_i)$, $i=1,2,\ldots$, semi-variances can be computed as:

$$\hat{\gamma}(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} [(z(x_i) - z(x_i + h))^2]$$ (4.12)

Where $m(h)$ is the number of pairs of data points separated by lag vector h. By changing h an ordered set of semi-variances are obtained which constitute the experimental or sample variogram. The average semi-variance for any lag can be obtained by grouping individual lag distances between point pairs into bins. The averaging is performed by choosing a set of lags, h_j, $j=1,2,\ldots$, at constant increments d, and then correlating each h_j with a bin of width d and bounded by $h_j - d/2$ and $h_j + d/2$. Each pair of points separated by $h_j \pm d/2$ is used to estimate $\hat{\gamma}(h_j)$. The lag distance and increment is important as it affects resulting variogram. The right decision depends on the number of data, distribution of it and form of the underlying variogram. The starting point may be using the average separation distance between nearest neighbors as d. (Webster& Oliver 2007, pp 68). The curve of a variogram flattens out at a certain lag distance which is referred as "range". Point pairs further apart than range are spatially independent and autocorrelation becomes zero. The variogram has the maximum value on the y axis is called in geostatistics as "sill". Theoretically, at zero separation distance the variogram value should be zero. However, at small separation distance, the difference between measurements often does not tend to zero and called as the "nugget effect".

4.2.4 Ordinary kriging

The Ordinary kriging (OK) estimates are weighted averages of neighboring data attributes (Lloyd, 2005). This model assumes no trend in the data. For OK data should have these three requirements:

- trend function should be constant,

- variogram should be constant over the whole study area, and

- data variable should have approximately normal distribution (Hengl, 2009).

The OK estimate is a linear weighted moving average of the available n observations defined in Equation (4.13) as:

$$\hat{Z}_{OK}(s_0) = \sum_{\alpha=1}^{n} \lambda_\alpha^{OK} z(s_\alpha)$$ (4.13)

$\hat{Z}_{OK}(s_0)$ is the OK estimation at location s_0, λ_α^{OK} are the OK weights, s_α is the observation locations and n is the number of observations. The sum of the OK weights should be equal to 1 as presented Equation 4.14,

$$\sum_{\alpha=1}^{n} \lambda_\alpha^{OK} = 1$$ (4.14)

And the expected error is: $E[\hat{Z}(s_0) - Z(s_0)] = 0.$ (4.15)

The important part of OK is to define weights which are obtained such that the estimation error is unbiased and estimation variance is minimized. Lagrange multipliers are used to achieve this. For this

purpose a secondary function is defined as $f(\lambda_\alpha,\psi)$ which contains variance to be minimized and Lagrange multiplier term ψ. For kriging it is defined as:

$$T(\lambda_\alpha,\psi)=Var[\hat{Z}(s_0)-Z(s_0)]-2\psi\left\{\sum_{\alpha=1}^{n}\lambda_\alpha-1\right\} \tag{4.16}$$

Partial derivatives of the function are set with respect to the weights to 0:

$$\frac{\partial f(\lambda_\alpha,\psi)}{\partial\lambda_\alpha}=0, \quad \frac{\partial f(\lambda_\alpha,\psi)}{\partial\psi}=0 \tag{4.17}$$

For $\alpha=1,2,...n$. This leads to a set of $n+1$ equations in $n+1$ unknowns:

$$\sum_{\alpha=1}^{n}\lambda_\alpha\gamma(s_\alpha,s_\beta)+\psi(s_0)=\gamma(s_\beta,s_0) \tag{4.18}$$

for all β,

$$\sum_{\alpha=1}^{n}\lambda_\alpha=1 \tag{4.19}$$

For more detailed information Webster&Oliver (2007) can be checked.

The estimation variance of ordinary kriging $(Var(s_0))$ is defined in Equation (4.20) as:

$$Var(s_0)=-\gamma(0)-\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_\alpha^{OK}\lambda_\beta^{OK}\gamma(s_\alpha-s_\beta)+2\sum_{\alpha=1}^{n}\lambda_\alpha^{OK}\gamma(s_\alpha-s_0) \tag{4.20}$$

Where $\gamma(s_\alpha-s_\beta)$ is the variogram value of Z between locations s_α and s_β, and $\gamma(s_\alpha-s_0)$ is the variogram value between s_α and target location s_0. For detailed information, see e.g. Llyod (2005).

4.2.5 Regression kriging

The other method that was implemented is Regression kriging. The method combines regression and kriging by treating these as two separate, consecutive steps. The regression part applies MLR as described before. Next, a kriging step is done in which the regression residual is no longer treated as uncorrelated but allowed to be spatially correlated. Thus, simple kriging is applied to the residuals (i.e., the differences between the observations and the predicted values with MLR). Simple kriging is used instead of ordinary kriging because it can be assumed that the residual has a known mean (namely zero). Finally, the kriged residual is added to the regression result. This method can thus be seen as an extension of MLR because by adding residual kriging to regression one has the ability to include additional information and gain more accurate predictions. The RK prediction formula is given by Equation (4.21):

$$\hat{Z}(s) = \hat{\beta}_0 + \sum_{k=1}^{p} \hat{\beta}_k \cdot X_k(s) + \sum_{i=1}^{n} \lambda_i \cdot \varepsilon(s_i) \qquad (4.21)$$

where $\hat{Z}(s)$ is the prediction at location s, $\hat{\beta}_0$ is the estimated intercept, the $\hat{\beta}_k$ are estimated regression model coefficients, $X_k(s)$ are the values of independent variables, n is the number of observations, the λ_i are simple kriging weights derived from the spatial dependence structure of the residual and where $\varepsilon(s_i)$ is the (observed) regression residual at measurement locations s_i. For details, see Hengl et al. (2004, 2007).

The accuracy of the RK prediction is quantified with the simple kriging prediction error variance, which is given by:

$$Var(Z(s) - \hat{Z}(s)) = C(s,s) - \sum_{i=1}^{n} \lambda_i \cdot C(s,s_i) \qquad (4.22)$$

where $C(s,s)$ is the variance of $\varepsilon(s)$ and $C(s,s_i)$ is the covariance of $\varepsilon(s)$ and $\varepsilon(s_i)$. Note that this equation also holds for MLR, in which case the covariances are by definition zero (assuming that the prediction location s_i does not coincide with a measurement location) implying that the prediction error variance of MLR is just the variance of the regression residual. Note also that uncertainty in estimation of the regression coefficients β_k is not taken into account, because this is typically ignored in RK (but see UK below).

4.2.6 Universal kriging

Universal kriging uses the same underlying statistical model as RK, but unlike RK, estimation of the trend and kriging of the residuals are integrated. This is more attractive from a theoretical point of view because when residuals are correlated, then this influences optimal estimation of the regression coefficients, which is ignored in RK. The difficulty is that UK requires that the spatial correlation structure of the residual is known prior to estimation of the regression coefficients, while estimated regression coefficients are needed to be able to compute the residuals at observation locations and estimate the spatial correlation structure of the residuals. Solutions that involve iterations are used for this, which works well but increases complexity (Hengl et al. 2007). The UK prediction at an unobserved location is also given by Eq. (4.21), but now with (slightly) modified values for the estimated regression coefficients and kriging weights.

An additional advantage of UK over RK is that computation of the prediction error variance also takes the estimation error of the regression coefficients into account as well as the correlation between these errors and the residual interpolation error. It is most easily presented in matrix notation and given by Equation (4.23) (Brus and Heuvelink, 2007):

$$Var(Z(s) - \hat{Z}(s)) = C(s,s) - c^T Cc + (x - X^T C^{-1} c)^T (X^T C^{-1} X)^{-1} (x - X^T C^{-1} c) \qquad (4.23)$$

where X is the $n \times (p+1)$ matrix of covariates at the observation locations (the first covariate is constant unity and corresponds to the intercept), x is the vector of covariates at the prediction location s, C is the variance-covariance matrix of the n residuals, c is the vector of covariances between the residuals at the observation and prediction locations and where T denotes transpose and -1 inverse (Brus and Heuvelink, 2007).

Apart from a direct influence, secondary variables can also aid the interpolation of the target variable through interactions (i.e. through a combination of explanatory variables). Therefore two-way interactions between continuous variables were also included and supplied as explanatory variables. This was restricted to continuous variables only because for categorical variables it would yield a too large number of covariates.

4.3   Space-Time Interpolation Methodology

Consider a variable z which varies in the spatial (s) and time (t) domain. Let z be observed at n space-time points (s_α, t_α), $\alpha = 1, \ldots, n$. These measurements constitute a space-time network of observations. However it is practically impossible to measure z at each spatial and temporal point. In order to obtain complete space-time coverage, interpolation of z is required. The aim of space-time interpolation is to predict $z(s_\alpha, t_\alpha)$ at a point (s_α, t_α) where z is not measured, where the unmeasured point typically is a node of a space-time grid. It is assumed that z is the realization of random function Z which holds full statistical model and space-time dependence structure. Next $z(s_\alpha, t_\alpha)$ is predicted from the space-time observations by using the assumed space-time model. The random function Z can be decomposed into deterministic and stochastic parts as follows presented in Equation (4.24):

$$Z(s,t) = m(s,t) + V(s,t) \tag{4.24}$$

The m defines the deterministic part of the random function and presents the large-scale space-time variation. V defines the stochastic part and comprises a zero-mean residual, which represents small-scale space-time variation (Heuvelink and Griffith, 2010, Kyriakidis and Journel, 1999). In addition, the trend m can be part of Z that can be explained by using secondary information. The decomposition of Z into the trend and residual is a one choice and there are other choices to make by a modeler. When the trend and residual components are obtained space-time kriging is performed in same way with spatial kriging methods (Section 4.2). In space-time kiriging future measurements affect past and present measurements as they are weighted by using same variogram (Snepvangers et al., 2003).

Sample space-time variogram (Section 4.2.3) is fitted to the data by using sum-metric model, and exponential functions were used for spatial, temporal and spatio-temporal components. Each of them has nugget, sill and range parameters (Section 4.2.3). The space-time variogram has an additional parameter that referred as anisotropy ratio. By using space-time anisotropy ratio α, distances in space and time can be reduced to a single space-time distance (Heuvelink and Griffith 2010, Snepvangers et al. 2003).

4.3.1   Space-time Ordinary kriging

Space-time ordinary kriging ST-OK is the same with pure spatial OK (Section 4.2.4); it predicts $Z(s,t)$ as a linear combination of n number of space-time observations as follows:

$$z_{STOK}(s,t) = \sum_{\alpha=1}^{n(s,t)} \lambda_a(s,t)z(s_\alpha, t_\alpha) \tag{4.25}$$

$$\sum_{\alpha=1}^{n(s,t)} \lambda_a(s,t) = 1 \tag{4.26}$$

The important part of kriging is to define weights λ_α accurately, so kriging prediction variance is minimized as follows (Gething et al. 2007):

$$Var\ (\ s,t\ ) = Var[\ \hat{Z}(\ s,t\ ) - Z(\ s,t\ )]$$ (4.27)

4.3.2 Space-time Universal kriging

Space-time Universal kriging (ST-UK) is not very different from spatial UK (Section 4.2.6). The trend m which is the deterministic part of the random function Z is thought to be explained physically or empirically by using secondary variables. The simplest method is to assume that the trend m is composed of a linear relationship between dependent and independent or secondary variables (Equation 4.8). After the trend is specified, for example with multiple linear regression, it is subtracted from Z, so that the space-time stochastic residual V is obtained. The trend m in ST-UK is written as follows:

$$m(\ s,t\ ) = \sum_{i=1}^{p} \beta_i f_i(\ s,t\ )$$ (4.28)

Where the β_i are regression coefficients obtained from multiple linear regression, the f_i are the values of independent variables which are known over space-time network, and p is the number of independent variables (Heuvelink and Griffith, 2010).

Once the trend and space-time variogram are specified, then space-time kriging is performed in the usual way, as in spatial kriging (Section 4.2). The ST-UK prediction of $Z(s_\alpha, t_\alpha)$ is given in matrix notation by:

$$\hat{z}(\ s_\alpha, t_\alpha\ ) = m_0^T \hat{\beta} + c_0^T C_n^{-1}(\ \bar{z} - M\hat{\beta}\ )$$ (4.29)

Where M is an $n * p$ design matrix of independent variables at the observation locations, m_0 is the vector of independent variables at the target location, C_n is the $n \times n$ variance-covariance matrix for the n residuals at the observation locations, c_0 is the vector of covariances between the residuals at the observation and target locations, and $\bar{z}$ is the vector of space-time observations (Heuvelink and Griffith, 2010).

Space-time Universal kriging prediction variance is given as follows:

$$Var(\ s_\alpha, t_\alpha\ ) = var(\ Z(\ s_\alpha, t_\alpha\ ) - \hat{Z}(\ s_\alpha, t_\alpha\ ))$$
$$Var(\ s_\alpha, t_\alpha\ ) = C(\ s_\alpha, t_\alpha, s_\alpha, t_\alpha\ ) - c_0^T C_n^{-1} c_0$$
$$+ (\ m_0 - M^T C_n^{-1} c_0\ )^T (\ M^T C_n^{-1} M\ )^{-1} (\ m_0 - M^T C_n^{-1} c_0\ )$$ (4.30)

4.4 Cross-validation Methodology

Ten-fold cross-validation was used to evaluate the performances of the spatial and space-time interpolation techniques (e.g. Gilardi and Bengio, 2000; Rigol-Sanchez et al. 2003). For this purpose, the total dataset comprising all measurements was randomly split in ten (approximately) equally sized sub-datasets. For each sub-dataset, the remaining 90% of the data was used as a training set to calibrate the spatial and space-time prediction model and make predictions of average long-term annual and mean annual precipitation at the sub-datasets that was set aside, and which comprises the test or validation dataset. In this way, predictions at the test dataset locations were compared with the

observed data for each of ten test datasets. Performance assessment was done by comparing the Root Mean Squared Error (*RMSE*), standardized MSE values (*SMSE*) and *R-square* values for spatial interpolation. For space-time interpolation *RMSE*, *R-square* and Mean Error (*ME*) were compared.

The *RMSE* is a common accuracy performance measure that is frequently used as a measure of magnitude of errors (Equation 4.31) (Lloyd, 2005; Karl, 2010; Schuurmans et al. 2007; Spadavecchia and Williams, 2009):

$$RMSE = \sqrt{\sum_{s=1}^{n} \frac{(Z_s - \hat{Z}_s)^2}{n}} \qquad (4.31)$$

The value Z_S is observation at location *s*, $\hat{Z}_S$ is the prediction value, and *n* is the number of observations.

The *SMSE* is a measure of the goodness of the assessment of the prediction error. It is defined as the average ratio of the squared prediction error at validation points and the corresponding prediction error variance as indicated in Equation 4.32. This prediction error variance allows assessment of the estimation uncertainty (Martínez-Cob, 1996). If the prediction error variance is correctly assessed, then the ratio should be close to 1 (Equation 4.32).

$$SMSE = \frac{1}{n} \sum_{s=1}^{n} \frac{(Z_s - \hat{Z}_s)^2}{Var} \qquad (4.32)$$

"*Var*" is the prediction error variance.

The *R-square* (r^2) indicates the amount of variance explained by the model (Hengl et al. 2004). It is calculated from the sum of squares of residuals (*SSerr*) and total sum of squares (*SStot*) as shown in Equation (4.33), (4.34) and (4.35):

$$r^2 = 1 - \frac{SSerr}{SStot} \qquad (4.33)$$

$$SSerr = \sum_{k=1}^{m} \sum_{s=1}^{n_k} (Z(s_{ik}) - \hat{Z}(s_{ik}))^2 \qquad (4.34)$$

$$SStot = \sum_{k=1}^{m} \sum_{s=1}^{n_k} (Z(s_{ik}) - \bar{Z})^2 \qquad (4.35)$$

where *m* is the number of training data sets, n_k is the number of observations in the *k-th* training set, s_{ik} is the i-th observation location in the k-th training set and $\bar{Z}$ is the mean of the observed dataset. *R-square* gives information about goodness of fit of model and values of it only ever be a number from zero and 1.0. If all points lie along the regression line and it has a slope different from zero, the unexplained component (*SSerr*) will be a very small number and *R-square* will be close to 1. If the explained sum of squares (*SStot*) is small in relation to the unexplained (*SSerr*), *R-square* will be a small number (McKillup&Dyar 2010, pp 217).

The *ME* can be presented also as mean deviation and it is calculated as shown in Equation 4.36.

$$ME = \frac{1}{n}\sum_{s=1}^{n}\left\{Z_s - \hat{Z}_s\right\}$$

<div align="right">4.36</div>

In ideal case, the *ME* should be zero because kriging is unbiased. The *ME* is a weak indicator since kriging is insensitive to inaccuracies in the variogram (Webster&Oliver 2007, pp 192).

CHAPTER 5

RESULTS AND DISCUSSIONS

5.1   Trend Analysis Results

The aim of the trend analysis part of this thesis is to test for temporal trends in precipitation values which is the important parameter of climate for Turkey. Annual and seasonal precipitation values of Turkey measured from 225 meteorological stations for a 34 year period (1970-2003) were used in the trend analyses. Because it is thought that seasonal precipitation totals are more efficient to show variation in time, visual trends of seasonal values were analyzed first. Next statistical trend analyses were conducted on annual and seasonal values. Results are discussed for river basins of Turkey.

5.1.1   Visual Trend Analysis

As a first step of trend analysis, a possible trend existence was observed visually to get an opinion about variation in time. The graphs are created by using seasonal total precipitation versus year according to each basin (Figure 5.1).

In spring, seasonal precipitation was increased in Antalya, Aras, Batı Akdeniz, Burdur and Doğu Karadeniz, B. Menderes and Van Golu basins, decreased in Kuzey Ege and Meriç basins.

In summer, decreasing precipitation was observed in Akarçay, Asi, Büyük Menderes, Küçük Menderes, Ceyhan, Gediz, Kuzey Ege, Meriç and Susurluk basins. Increasing precipitation was observed in Aras, B. Karadeniz, Burdur and Coruh basins.

In autumn, unclear precipitation increment was observed for some basins (Akarcay and D. Karadeniz) but generally variation does not seem to have a clear trend.

In winter, precipitation was increased in Akarçay, Ceyhan and Seyhan basins and decreased in Susurluk basin.
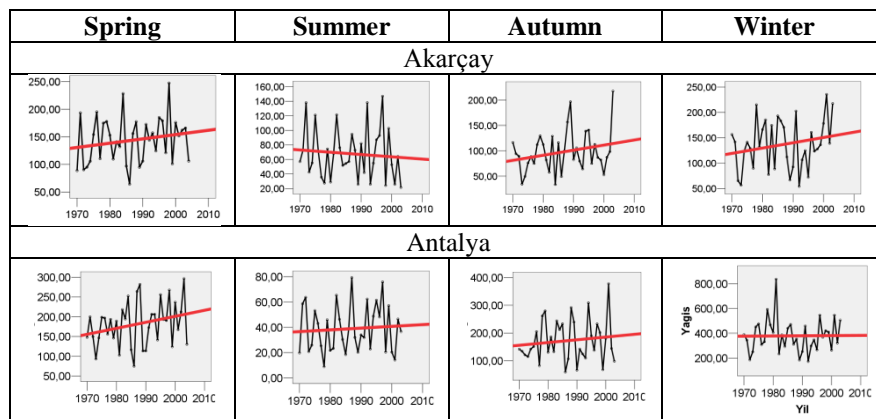
Figure 5.1. Variation of seasonal precipitation values by time according to each basin; x axis presents precipitation and y axis presents year. Red lines are fitted linear regression lines.

Figure 5.1. Variation of seasonal precipitation values by time according to each basin; x axis presents precipitation and y axis presents year. Red lines are fitted linear regression lines (continued).

36

	Spring	**Summer**	**Autumn**	**Winter**

Dicle

D. Akdeniz

D.Karadeniz

B.Karadeniz

Burdur

Ceyhan

Coruh

Dicle

D.Akdeniz

Figure 5.1. Variation of seasonal precipitation values by time according to each basin; x axis presents precipitation and y axis presents year. Red lines are fitted linear regression lines (continued).

37

Figure 5.1. Variation of seasonal precipitation values by time according to each basin; x axis presents precipitation and y axis presents year. Red lines are fitted linear regression lines (continued).

38

	Spring	Summer	Autumn	Winter
Sakarya				
Seyhan				
Susurluk				
Van Gölü				
Yeşilırmak				

Figure 5.1. Variation of seasonal precipitation values by time according to each basin; x axis presents precipitation and y axis presents year. Red lines are fitted linear regression lines (continued).

5.1.2 Annual Trend Analysis

T-Test and Mann-Kendal tests were applied to original annual precipitation measurements of Turkey. The H_0 hypothesis, which states that there is no trend through time, is accepted or rejected according to 5 % significance level. T statistic and Z statistic are used to reject or accept the T-Test and Mann-Kendal test results, respectively. In a second phase serial correlation in time-series data is removed (pre-whitening method) and tests were then applied to the residual terms again. Results were discussed by comparing the first outcomes with maps and tables.

5.1.2.1 T-Test

According to the T-Test, 21 meteorological stations had trend in annual precipitation values. From these, two stations had a decreasing trend and 19 had an increasing annual precipitation trend. In Table 5.1, T-Test results are presented with necessary statistics and basin information. In Figure 5.2, results are presented on a map. Generally, increasing trend appeared at the middle, north and north-east Anatolia in 14 different basins. A decreasing trend was seen in Kizilirmak and Van Golu basins at two meteorological stations. From a general perspective, annual precipitation has an increasing trend over Turkey according to the outcome of the T-Test.

Table 5.1. T-Test results obtained with the original data.

	Stat. no	Basin	Obs. Period	T statistic	t a=0.05	No Trend
T-Test Results						
Stations had decreasing trend	17648	Kızılırmak	34	-2.79	2.03	Reject
	17852	Van Gölü	30	-4.21	2.045	Reject
Total	**2 Stations**					
Stations had increasing trend	17026	B.Karadeniz	34	3.00	2.03	Reject
	17037	D. Karadeniz	34	3.72	2.03	Reject
	17190	Akarçay	33	2.75	2.037	Reject
	17238	Burdur Göller	34	2.38	2.03	Reject
	17370	Asi	34	2.18	2.03	Reject
	17606	B. Karadeniz	34	2.21	2.03	Reject
	17612	B. Karadeniz	34	2.32	2.03	Reject
	17630	Aras	34	2.38	2.03	Reject
	17656	Aras	34	2.19	2.03	Reject
	17684	Yeşilırmak	34	2.19	2.03	Reject
	17702	Sakarya	34	2.03	2.03	Reject
	17732	Kızılırmak	33	2.04	2.037	Reject
	17796	Akarçay	34	2.33	2.03	Reject
	17812	Van Gölü	34	2.44	2.03	Reject
	17820	K. Menderes	34	2.33	2.03	Reject
	17832	Konya Kapalı	34	4.14	2.03	Reject
	17898	Konya Kapalı	34	2.00	2.03	Reject
	17906	Seyhan	34	2.00	2.03	Reject
	17960	Ceyhan	34	2.03	2.03	Reject
Total	**19 Stations**					
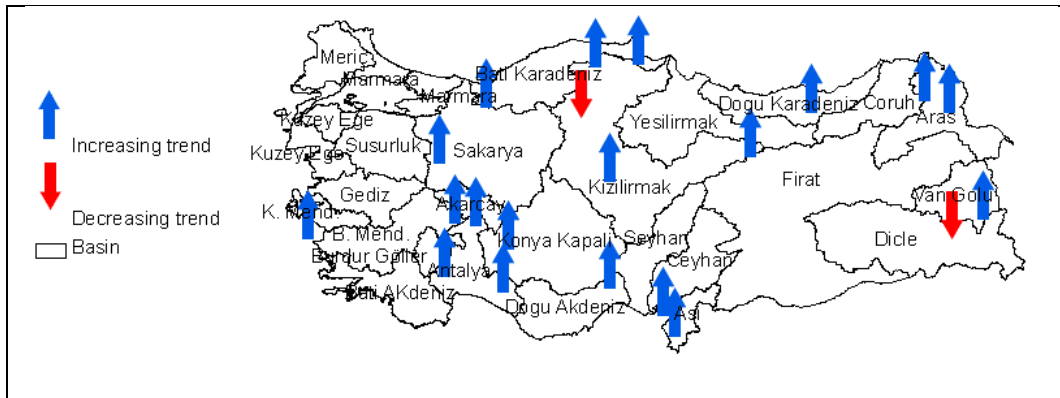
Figure 5.2. T-Test results.

5.1.2.2 Mann-Kendal Test

According to Mann-Kendal test, in total 15 stations had a significant trend on annual precipitation values. Among them, three stations had decreasing and 12 stations had an increasing trend (Table 5.2). Stations that had a significant trend are presented in Figure 5.3. The three stations with a decreasing trend are located in Sakarya and Van Golu basins. The 12 stations with an increasing trend are located

in the south, north and east parts of Turkey. Only station 17852 had a decreasing trend for both trend tests. Stations: 17026, 17037, 17190, 17238, 17370, 17612, 17812, 17820, 17832, 17898, and 17960 had an increasing trend for both tests.

Table 5.2. Mann-Kendal test results obtained with the original data.

	Mann-Kendal Test Results					
	Stat. no	Basin	Obs. Period	Z Statistic	Z=0,05	No Trend
Stations had decreasing trend	17679	Sakarya	34	-2.30	1.96	Reject
	17706	Sakarya	23	-2.01	1.96	Reject
	17852	Van Gölü	30	-3.09	1.96	Reject
Total	3 Stations					
Stations had increasing trend	17026	B. Karadeniz	34	2.46	1.96	Reject
	17037	D. Karadeniz	34	3.07	1.96	Reject
	17190	Akarçay	33	2.25	1.96	Reject
	17238	Burdur Göller	34	2.24	1.96	Reject
	17370	Asi	34	2.15	1.96	Reject
	17612	B. Karadeniz	34	1.97	1.96	Reject
	17774	Fırat	33	2.06	1.96	Reject
	17812	Van Gölü	34	3.02	1.96	Reject
	17820	Küçük Mend.	34	2.15	1.96	Reject
	17832	Konya Kapalı	34	3.07	1.96	Reject
	17898	Konya Kapalı	34	1.99	1.96	Reject
	17960	Ceyhan	34	2.09	1.96	Reject
Total	12 Stations					
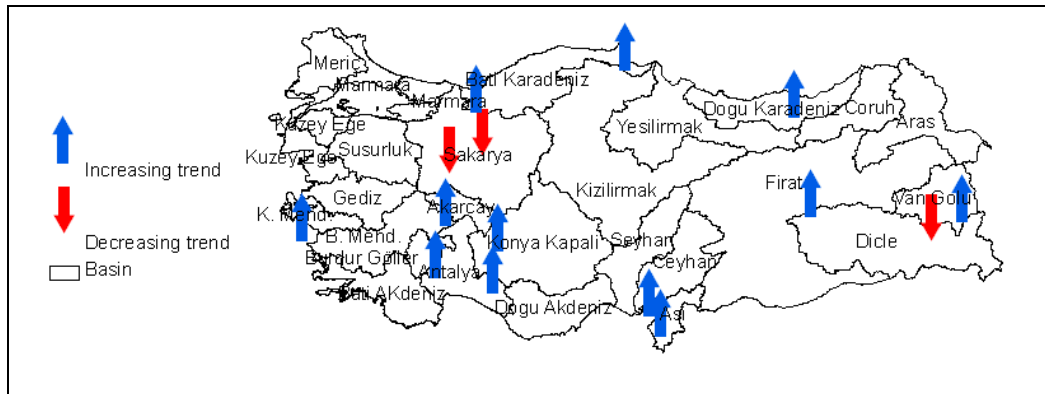
Figure 5.3. Mann-Kendal test results.

The serial correlation between time series data was removed and trend tests were then applied to residual terms again. The results are presented in Table 5.3 and Figure 5.4.

This time a significant trend was observed at nine stations for the T-Test and at six stations for Mann-Kendal test. The same two stations that had a decreasing trend for T-Test before had again a decreasing trend. However, for the Mann-Kendal test only one station (17648) had a decreasing trend after removal of serial correlation. Seven stations and five stations had an increasing trend for the T-Test and Mann-Kendal test, respectively.

Table 5.3. Results after eliminating "1" interval serial correlation coefficient.

Station_No	T-Test			Mann-Kendal Test		
	T statistic	t a=0.05	No Trend	Z statistic	Z 0,05	No Trend
17190	2.75	2.037	Reject	2.25	1.96	Reject
17238	1.97	2.03	Reject	2.02	1.96	Reject
17606	2.04	2.03	Reject	1.82	1.96	Accept
17648	-2.09	2.03	Reject	-2.05	1.96	Reject
17684	2.83	2.03	Reject	2.55	1.96	Reject
17702	2.02	2.03	Reject	1.57	1.96	Accept
17732	2.41	2.037	Reject	2.34	1.96	Reject
17852	-2.10	2.045	Reject	-0.98	1.96	Accept
17960	2.06	2.03	Reject	2.06	1.96	Reject
Total	9 Reject			6 Reject		

Generally the middle parts of Turkey had an increasing annual precipitation trend for both tests (Figure 5.4). A decreasing trend was observed in Kizilirmak and Van Golu basins. Increasing trend was observed in Akarcay, Burdur Goller, B. Karadeniz, Yesilirmak, Kizilirmak, Sakarya and Ceyhan basins.
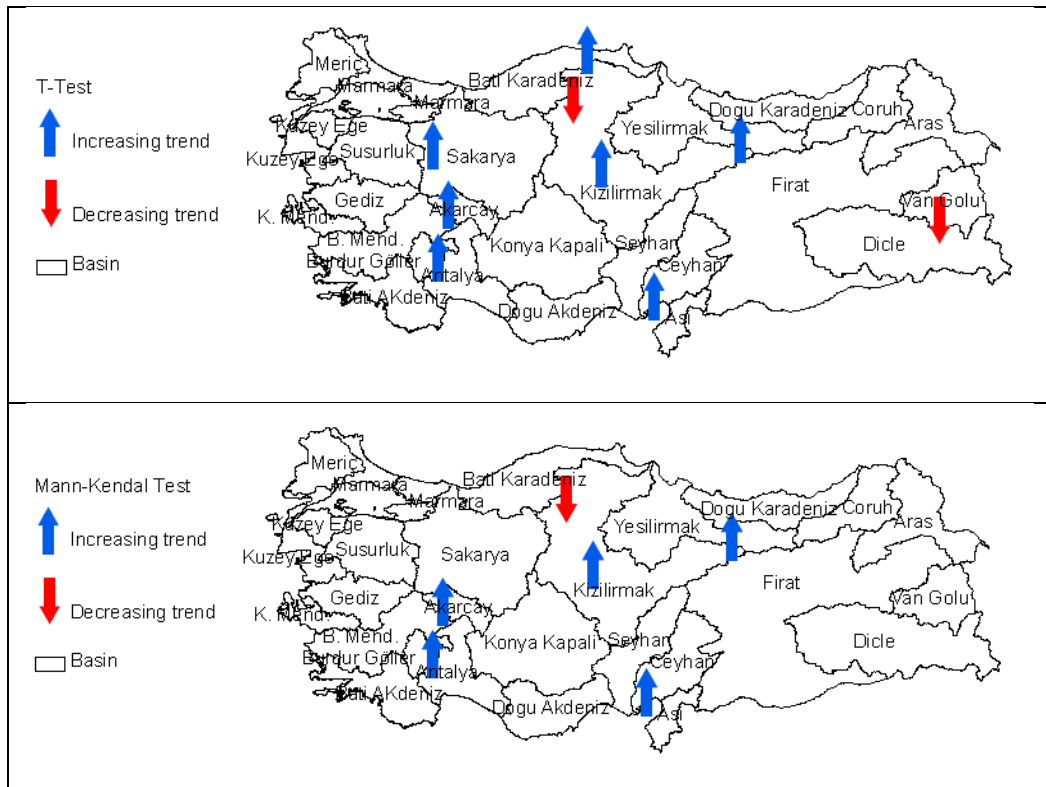
Figure 5.4. T-Test and Mann-Kendal test results after eliminating serial correlation.

42

5.1.3 Seasonal Trend Analysis

Mann-Kendal test was also applied to seasonal total precipitation values. This test was applied to observations twice, first to the original seasonal values and second to summer and winter values after removal of serial correlation. It is interesting that 19 stations had a significantly increasing precipitation trend in winter (Table 5.4) before eliminating serial correlation. After removing serial correlation, 11 stations had a significant increasing winter precipitation trend. For spring and autumn, there are again increasing precipitation trends at twelve and eight stations, respectively. It is interesting that there is a decreasing precipitation trend in summer, unlike for other seasons.

Table 5.4. Number of stations that show trends in seasonal averages.

Seasons	Increasing prec.	Decreasing prec.
Summer	2	18
Summer- after serial correlation	2	13
Autumn	8	0
Winter	19	1
Winter- after serial correlation	11	1
Spring	12	3

In order to see how the magnitude of the trend changes spatially, the Mann-Kendal coefficients of meteorological stations are interpolated using Inverse Distance Weighting (IDW) (Figure 5.5).

The results show that in spring, increasing precipitation existence appeared at south-east and north-east parts of Turkey (Figure 5.5a). In the Aegean region a weak decreasing trend draws attention. In summer, the weak decreasing trend became a little bit clearer and the magnitude of the Z statistic increased (Figure 5.5b). In autumn, an increasing precipitation trend shows up in the west and east parts of the Black Sea region and on a small area in southern part of Turkey (Fig 5.5c). In winter, an increasing precipitation trend appears in the south, south-east and east parts of Turkey (Figure 5.5d).

5.2 Spatial Interpolation Results

Five interpolation methods were applied to long-term annual average precipitation values of Turkey. Prediction maps and standard deviations of predictions were compared and interpreted on maps. Performance assessments of methods were made by comparing *RMSE, R-square* and *SMSE* values. In addition, prediction performances of methods were tested on extrapolation analysis.

The MLR method firstly was applied by using all secondary variables. From the results significant ones were selected according to 5 % level. Table 5.5 gives the coefficients of the MLR application of the first training dataset by using significant secondary variables. For this dataset the intercept, elevation, surface roughness, distance to nearest coast, land cover type, and the interaction between elevation and distance to coast were selected. For application of RK and UK the same explanatory variables and interactions were used. Additionally UK was applied with using only elevation as an explanatory variable. For GWR, as it is a local method that can take advantage of local effects, all variables and only the most significant interaction (elevation-distance to nearest coast) were used.
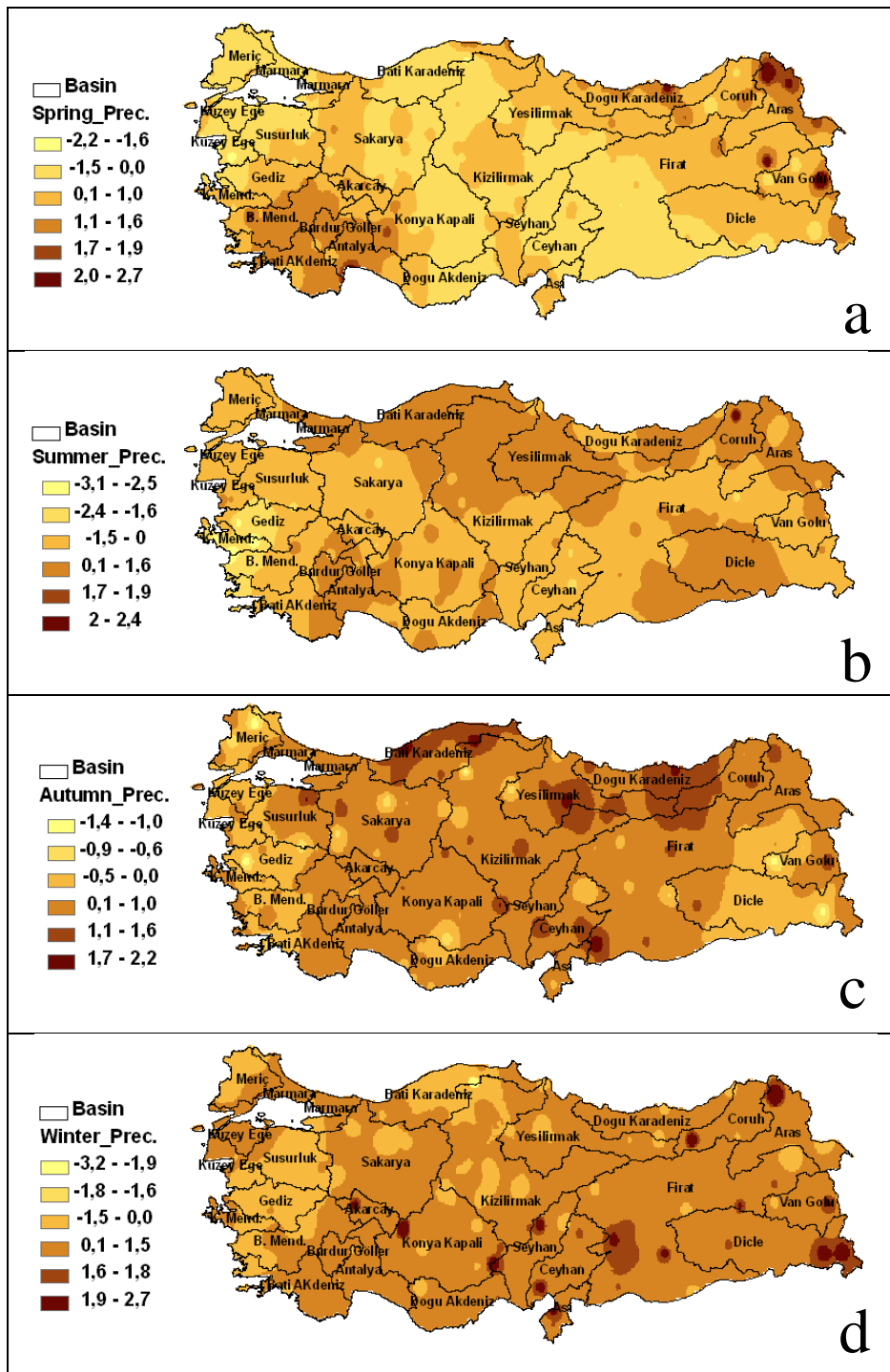
Figure 5.5. Distribution of Mann-Kendal coefficients according to seasons by IDW method, "a" presents spring, "b" presents summer, "c" presents autumn and "d" presents winter seasons.

Table 5.5. Model coefficients of the MLR application using only significant secondary variables and interactions.

Coefficients					
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**	**Signif. code**
(Intercept)	2.85E+00	1.68E-02	169.462	< 2e-16	***
Z-elevation	-2.1E-04	2.41E-05	-8.792	8.3E-16	***
V1- Surface roughness	5.15E-04	1.31E-04	3.935	0.00012	***
V2- Distance to nearest coast	-8.9E-07	1.52E-07	-5.903	1.6E-08	***
V51- Land cover class: agricultural	4.14E-02	2.14E-02	1.935	0.0544	.
V52- Land cover class: wetlands	9.22E-02	8.46E-02	1.09	0.2770	
V53- Land cover class: open-space	3.18E-02	3.65E-02	0.871	0.3847	
V54- Land cover class: vegetation	5.63E-02	3.66E-02	1.538	0.1257	
V55- Land cover class: forest	1.14E-01	3.27E-02	3.501	0.0005	***
Z:V2- interaction between elevation and dist. to coast	7.99E-10	1.25E-10	6.384	1.2E-09	***
Signif. code	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' ' 1

The residual variogram of the first training dataset is given in Figure 5.6. A spherical model was fitted to the sample variogram. There was a strong spatial correlation in the residuals, with zero nugget, partial sill of 38500 mm^2 and a spatial range of 130 km.
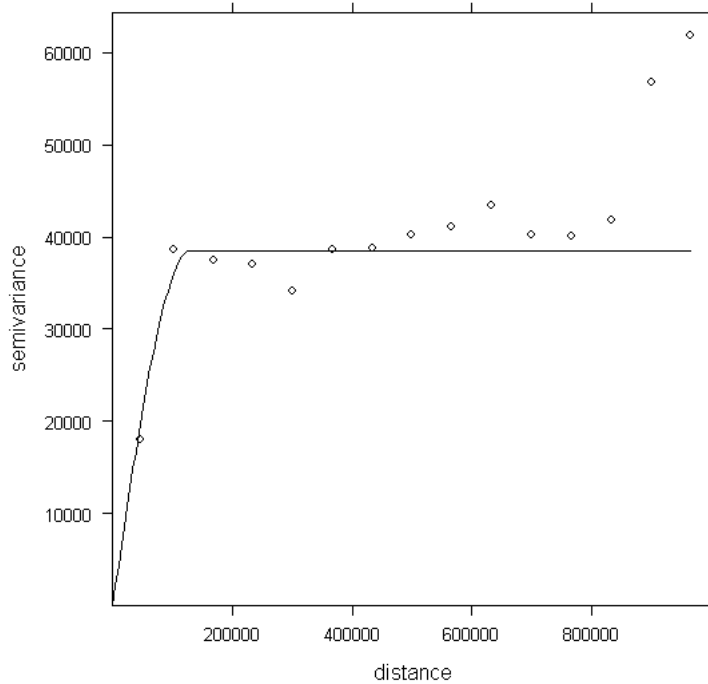
Figure 5.6. Semi-variogram of regression residual of the first training dataset. Open circles are the sample variogram values, solid line the fitted model.

The interpolation results given in Figure 5.7 show that the average annual precipitation estimates are large in the north-east, north-west, south-east, south and west of Anatolia when using MLR, GWR,

RK and UK. OK and UK with elevation methods predicted large precipitation in the north-east, north-west, south-east and south-west parts of Turkey. At the centre and in some parts of East Turkey, annual precipitation estimates were small (approximately 300 mm) by OK and UK with elevation. Around Lake Van, Keban and the Ataturk Dams a fairly high precipitation was estimated. Generally, the positive coastal effect on precipitation amount can be observed since the estimated values are high along the coasts and near the two big dams and the biggest lake of Turkey. OK and UK with elevation prediction maps show little detail and are fairly smooth. MLR, GWR, RK and UK prediction maps were fairly similar, but note that the observed high precipitation in the north-east of Anatolia was identified better with the kriging models than with the regression models.
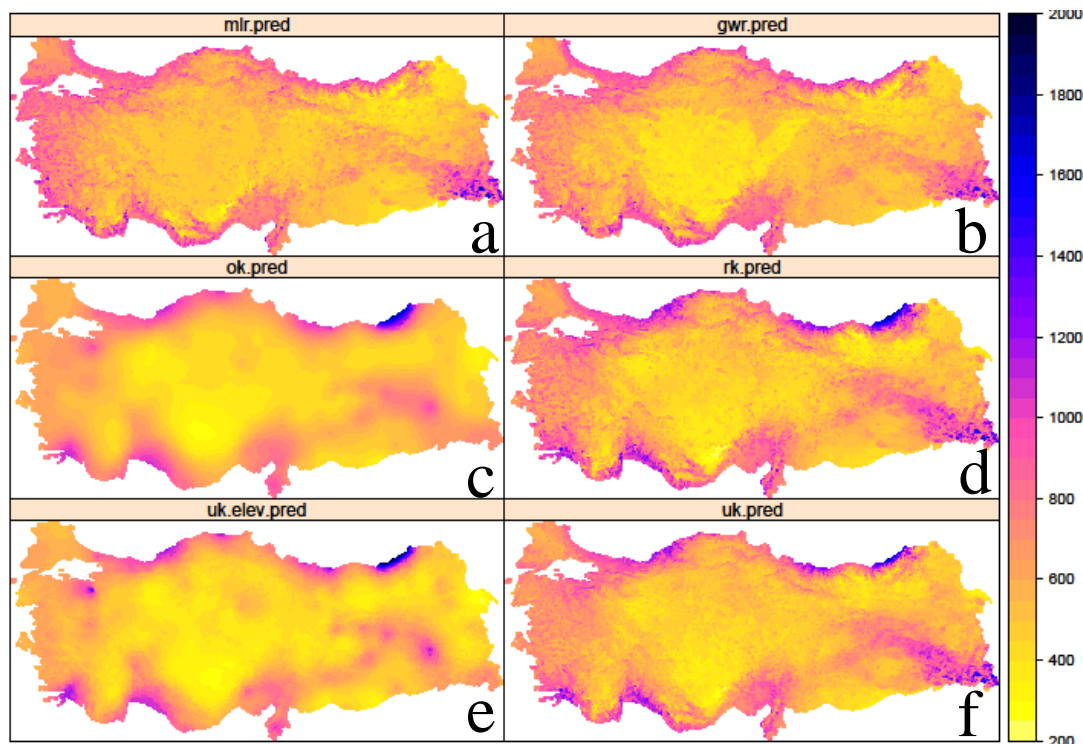
Figure 5.7. Predicted precipitation maps (mm) obtained from interpolation methods using the entire dataset, "a" presents MLR, "b" presents GWR, "c" presents OK, "d" presents RK, "e" presents UK with using elevation, and "f" presents UK method.

Prediction variances were converted to standard deviations by taking their square root. For MLR, OK, RK and UK prediction standard deviations were mapped in this way, but for GWR the prediction variance could not be calculated because it is unclear what the degrees of freedom of this method are (Roger Bivand, personal communication). This is necessary information to calculate an unbiased estimate of the prediction variance. Since we could not calculate GWR prediction variances, the standard deviations and *SMSE* could not be derived for this method. Figure 5.8 presents the prediction standard deviations of MLR, OK, RK and UK. The standard deviation is small near observation locations and greater far away from them. Generally it is expected that the standard deviation is large at the border of the study area, but in this case it is not because of the high spatial density of stations at the Turkish border. In fact the standard deviation is fairly constant over Turkey, because the distribution of meteorological stations is fairly homogeneous. The small values for all methods are located generally in regions which have the highest density of stations. For the MLR and UK methods the standard deviation is higher in the east and south-east parts of Turkey and overall these standard deviations are larger than those of OK, RK and UK with elevation.
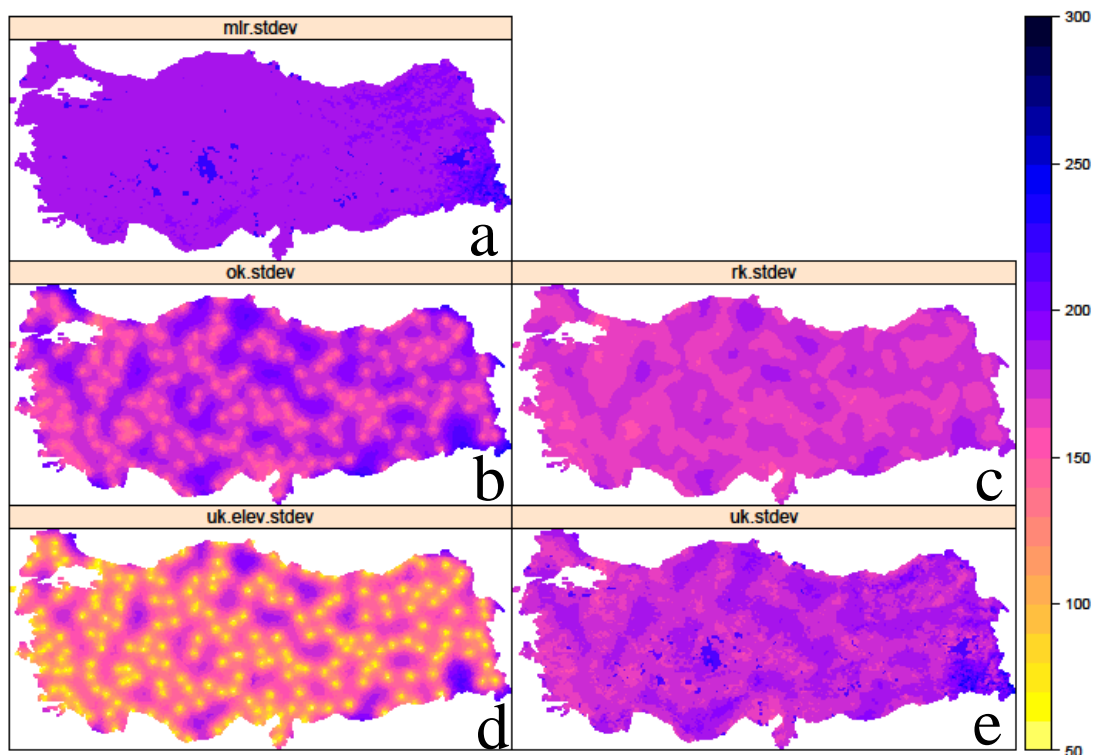
Figure 5.8. Prediction standard deviation maps of annual precipitation (mm) of MLR, OK, RK, and UK models, "a" presents MLR, "b" presents OK, "c" presents RK, "d" presents UK with using elevation, and "e" presents UK method.

5.2.1 Cross-validation results

In Figure 5.9 the observed versus predicted values are plotted for all models. As expected, the three stations that had the highest observed annual precipitation were underestimated with all methods. In general, if the stations had a larger than average precipitation (app. 1000 mm and higher) models under-predicted and yielded positive residuals. The models usually provided more accurate predictions at stations with precipitation below 1000 mm.
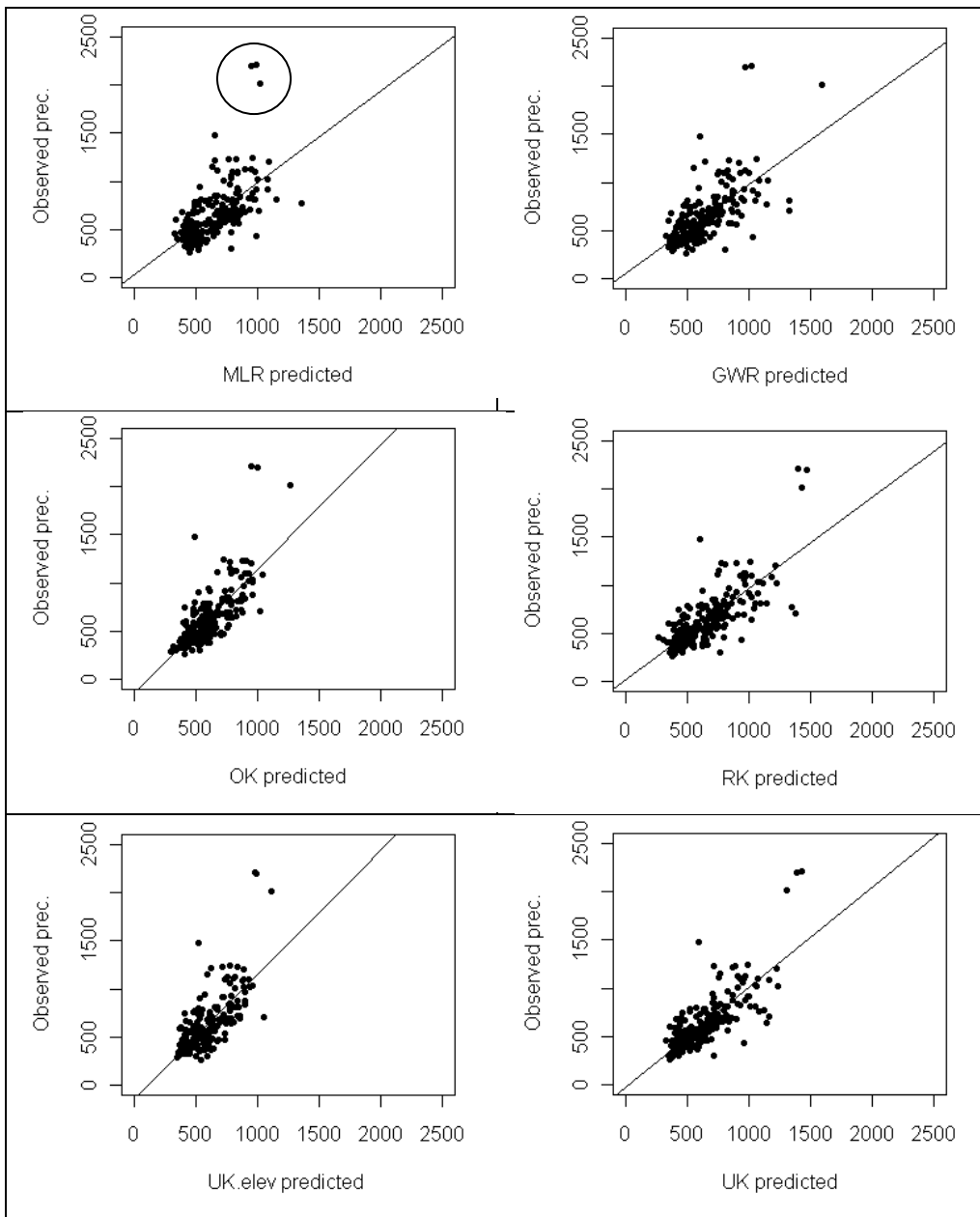
Figure 5.9. Observed versus predicted precipitation for each of the evaluated interpolation methods (circle in MLR plot indicates the three stations that had the highest precipitation over Turkey).

Figure 5.10 shows box-plots of interpolation errors for all interpolation models. The narrowest distributions are obtained for UK and RK, but the UK model seems to have fewer outliers than RK. GWR and OK have approximately the same plot width but GWR has more outliers. MLR and UK with elevation have the widest box-plots and MLR has more outliers than all other methods.
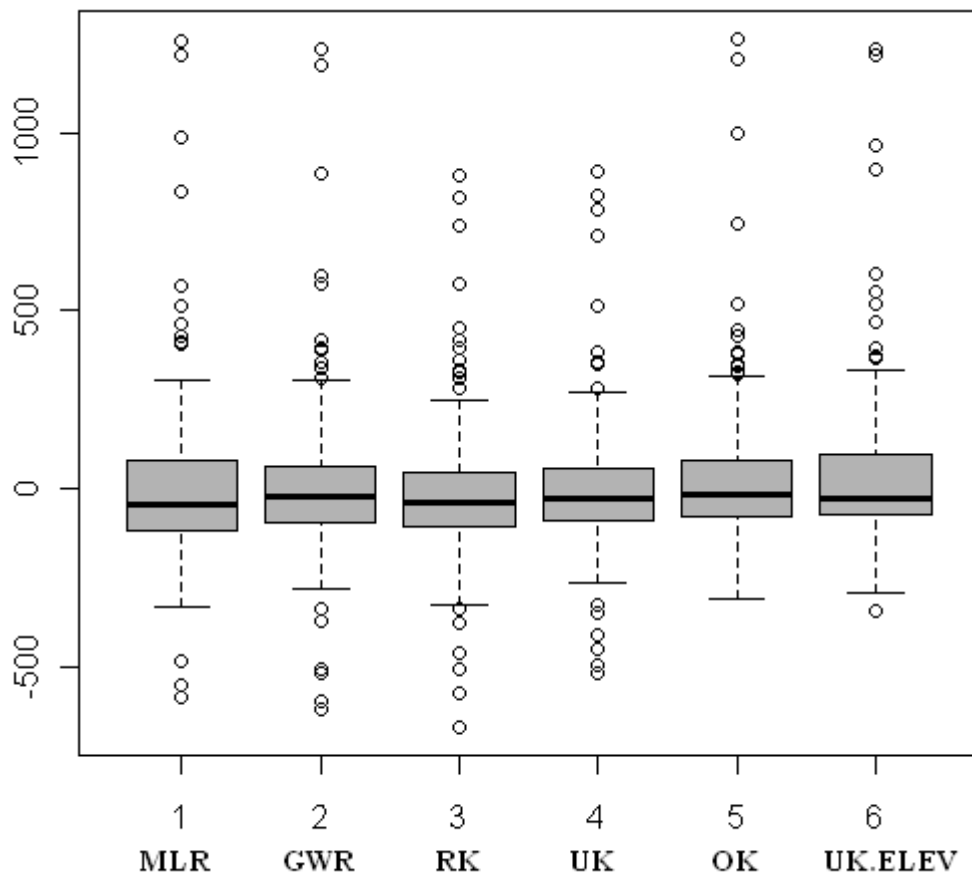
Figure 5.10. Box-plots of 10-fold cross-validation errors for the six interpolation methods.

Figure 5.11 shows a spatial plot of the cross-validation errors. Prediction errors at measurement locations were mostly between -200 and +200 mm. Overestimations and underestimations were observed generally at stations with higher or lower observed values than average. The three stations that have the highest observed precipitation are underestimated with all methods but RK and UK predictions are closer to the observed values than the other methods.
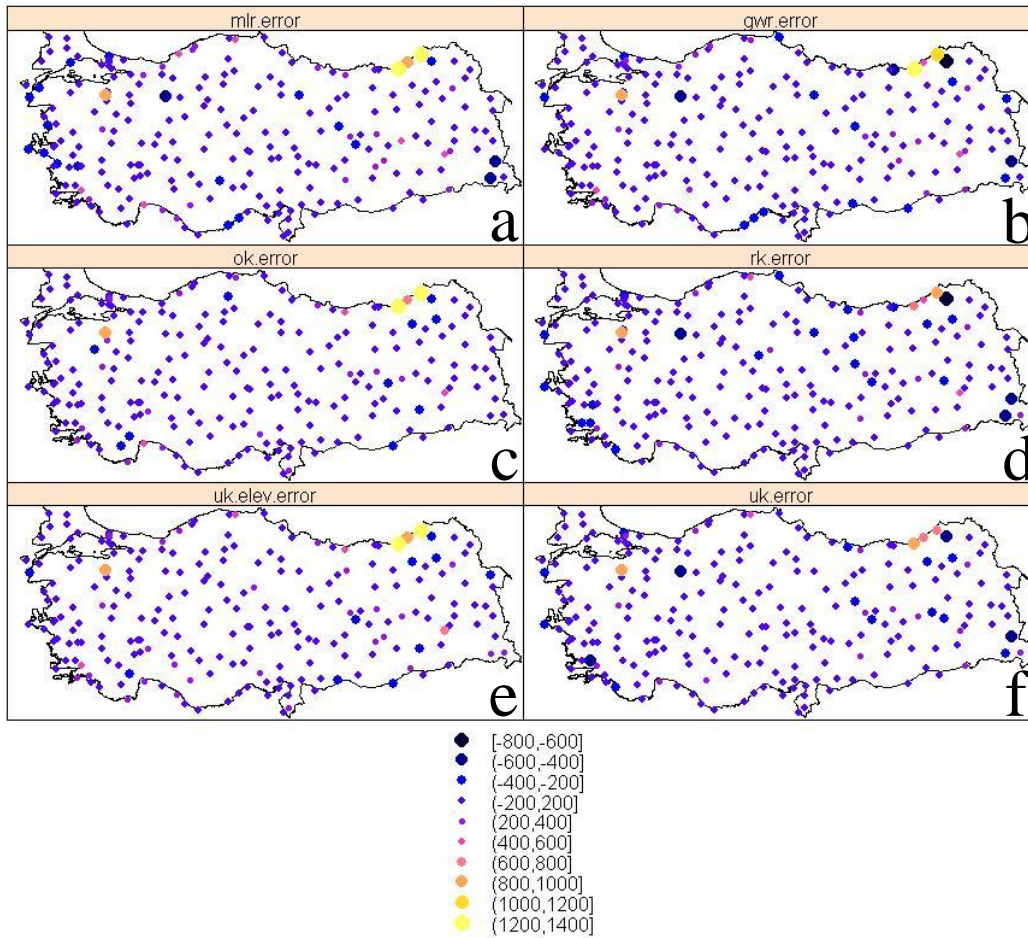
Figure 5.11. Prediction errors of interpolation methods, "a" presents MLR, "b" presents GWR, "c" presents OK, "d" presents RK, "e" presents UK with using elevation, and "f" presents UK method.

Table 5.6 shows the performance measures of each of the interpolation methods. UK is the most accurate method with an *RMSE* of 178 mm, *R-square* of 0.61 and *SMSE* of 1.06. MLR is the worst (*RMSE* of 222 mm, *R-square* of 0.39 and *SMSE* of 1.44). OK, UK with elevation and GWR are intermediate with performance values in between those of UK and MLR. RK is the second most accurate method with an *RMSE* of 186 mm, *R-square* of 0.57 and *SMSE* of 1.22. As expected, UK has an SMSE that is closer to 1 better than that of RK and hence yields a more realistic assessment of the interpolation error.

Table 5.6. Performance comparison of interpolation methods obtained with 10-fold cross-validation.

	MLR	**GWR**	**OK**	**RK**	**UK with only elevation**	**UK**
R-square	0.39	0.45	0.50	0.57	0.44	0.61
RMSE (mm)	222.2	210.8	200.8	185.9	211.9	177.7
SMSE	1.44	-	1.2	1.22	1.17	1.06

50

5.2.2   Testing Interpolation Methods on Extrapolation

The regression and kriging methods were tested on spatial interpolation but it is interesting to also look at how these techniques perform in spatial extrapolation. Journel and Rossi (1989) investigated and compared OK and UK performances for interpolation and extrapolation. In their study, the trend was a low-order polynomial of the geographic coordinates. According to their results, OK and UK performed equally well for interpolation, while including a trend was important for extrapolation. In our study the trend was a function of external covariates and found significant for interpolation. In order to analyze the importance of including a trend and evaluate the performance of the various methods an extrapolation analysis was added to the study.

For the extrapolation analysis, Turkey was divided into a western and eastern part. The data at the meteorological stations in the western part (125 stations) were used for training the models and prediction, while the stations in the eastern part (100 stations) were used for testing the models.

The same models that were used for interpolation were used in the extrapolation. As before, the accuracy assessment was performed by calculating the *RMSE*, *R-square* and *SMSE*, this time on the observations from the eastern part of Turkey.

The OK prediction map is different from those of all other methods (Figure 5.12). This is because OK is the only method that does not use covariate information and predicts the mean when observations are far away from the prediction location. MLR, GWR, RK, and UK used more or less the same regression formula and coefficients and hence gave similar results. UK with elevation had fewer details because it could only make use of elevation as covariate. All models that included the full regression predicted unreasonably high values in the south-east of Turkey. This region has very high topography which is outside the range of elevations of the western part. The unreasonably high values show the risk of extrapolation in feature space. On a similar note none of the models could reveal the high precipitation values in the north-east of Turkey.

UK with elevation was the worst extrapolator and had an *RMSE* of 331 mm, an *R-square* of 0.02 and *SMSE* of 2.67 (Table 5.7). OK was the second-worst extrapolator with similar performance measures. The small *R-square* s of UK with elevation and OK that were close to zero show that these methods perform equally poor as simply taking the mean of the observations. MLR, GWR and RK were the best extrapolators with *R-square*'s varying between 0.32 and 0.33. UK performed worse than anticipated. The *SMSE* was too large for all cases, indicating that the model assumptions are not realistic and underestimate the true spatial variability. This is partly due to three extreme values in the north-eastern part of the country.

Table 5.7. Performance comparison of extrapolation methods using data from western Turkey to predict the eastern part.

	MLR	GWR	OK	RK	UK with only elevation	UK
R-square	0.33	0.32	0.02	0.33	0.02	0.25
RMSE (mm)	272.6	275.7	330.6	273.0	331.3	289.8
SMSE	2.18	*	1.76	2.79	2.67	2.45
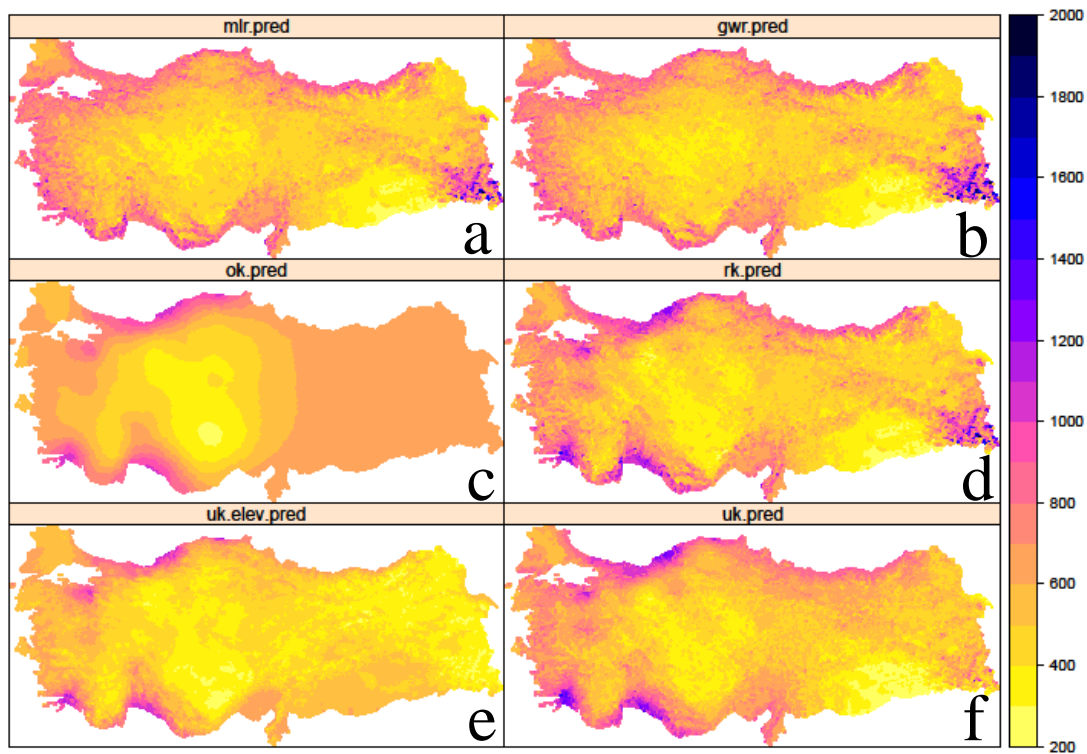
Figure 5.12. Prediction maps of annual precipitation (mm) obtained with the extrapolation analysis, "a" presents MLR, "b" presents GWR, "c" presents OK, "d" presents RK, "e" presents UK with using elevation, and "f" presents UK method.

The associated standard deviation maps are given in Figure 5.13. Except for MLR the east-west division can be easily recognized. Small values are obtained near observation locations whereas larger values occur in the extrapolation part. The similarities between the MLR and UK standard deviation maps show that uncertainty about the regression coefficients contributes importantly to the UK standard deviation map. Indeed the largest uncertainties are obtained in the south-east of Turkey, where extrapolation in feature space occurs due to the high elevation. RK does not have large values in this region because it ignores the estimation errors of the regression coefficients.
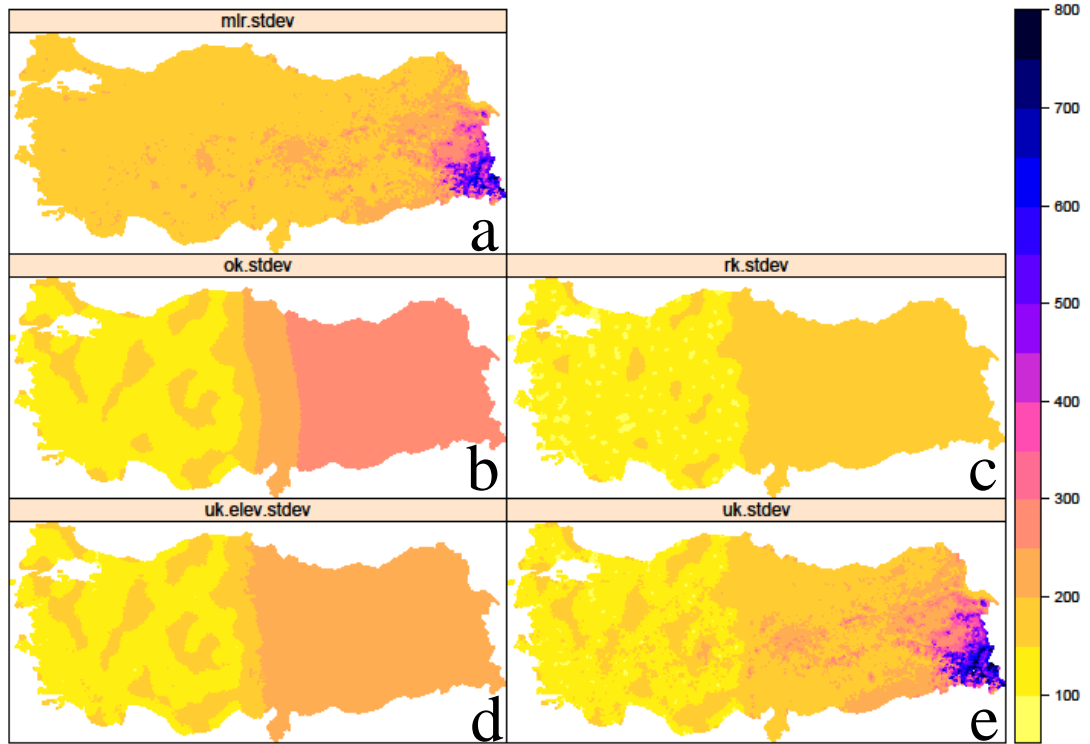
Figure 5.13. Prediction standard deviations of annual precipitation (mm) for various methods as obtained in the extrapolation analysis, "a" presents MLR, "b" presents OK, "c" presents RK, "d" presents UK with using elevation, and "e" presents UK method.

In order to examine the odd extrapolation results obtained with UK, the extrapolation was repeated by using the data from the eastern part of Turkey to predict the western part. The results are given in Table 5.8. This time UK had the largest *R-square* and smallest *RMSE*. MLR and RK are equally accurate, with similar values for the *R-square* and *RMSE*. These results are as expected because all three methods take covariate information into account which is particularly advantageous in extrapolation mode. GWR has a smaller *R-square* and hence larger *RMSE* than UK, MLR and RK but the cause of this it is not obvious. The *SMSE* values are smaller than 1 in all cases, which is in strong contrast with the *SMSE* results presented in Table 9.3. Apparently, the spatial variability in the eastern part of Turkey is greater than that in the western part, which leads to an underestimation of the variability when data from the West are used to predict the East and vice versa. Indeed the standard deviation of the annual precipitation dataset is substantially smaller for the western part than for the eastern part dataset (239 mm and 336 mm, respectively). This shows that extrapolation may give unreliable and inconsistent results if the statistics of the observations do not match those of the target variable in the prediction area. Indeed the eastern part of Turkey is much more mountainous than the western part, which explains the larger spatial variability of annual rainfall in the East.

Table 5.8. Performance comparison of extrapolation methods using data from eastern Turkey to predict the western part.

	MLR	GWR	OK	RK	UK with only elevation	UK
R-square	0.27	0.11	-0.19	0.28	0.01	0.29
RMSE (mm)	192.9	214.2	248.6	192.8	226.4	191.3
SMSE	0.97	*	0.57	0.77	0.59	0.69

5.3   Space-Time Interpolation Results

As mentioned before, methods defining and estimating the spatial variability of hydrologic, climatic and other environmental variables and performing spatial interpolation are abundant in the environmental sciences. Recently, the extension of these methods to variables that vary both in space and time has received increasing attention as shown in Section 4.3. In this part of the thesis the results of applying spatial universal kriging, space-time ordinary kriging and space-time universal kriging to annual precipitations of the Euphrates Basin, Turkey for a 39 year period from 1970 to 2008 are presented. The primary variable precipitation varies in time while all secondary variables are constant in time, which implies that the temporal variability is presented in the stochastic residual of the geostatistical models.   The results of space-time universal kriging are compared with space-time ordinary kriging and with spatial universal kriging for each separate year, using the *RMSE, R-square* and *ME* as performance measures.

For spatial universal kriging only continues variables; elevation, surface roughness, distance to coast and river density were used. The reason that secondary variables; aspect, eco-region and land cover were not used is that datasets that have few numbers of observations for a specific year may not yield reliable model parameters. Therefore it is thought that using continuous variables is more reliable even when the data have few observations.

In space-time ordinary kriging secondary variables are not used. The only data source is the observed precipitation values. The space-time variogram is created from the log-transformed of original measurements.

At space-time universal kriging, among the secondary variables; elevation, surface roughness, distance to coast, river density, aspect, land cover, Year and elevation-distance to coast interaction were used. In this case the space-time variogram was generated from the residuals. In order to obtain the residuals, multiple linear regression method was applied to each training datasets. Obtained residuals were then used to compute sample space-time variograms and fit variogram models.

Precipitation measurements for three arbitrary locations through observation period are presented in Figure 5.14. The locations of arbitrary stations are Stat.17165: x=1306119.9, y= 4650757.5; Stat.17270: x=1247446.1, y=4432773.9; Stat. 17203: x=1390248.6, y= 4629873.7.
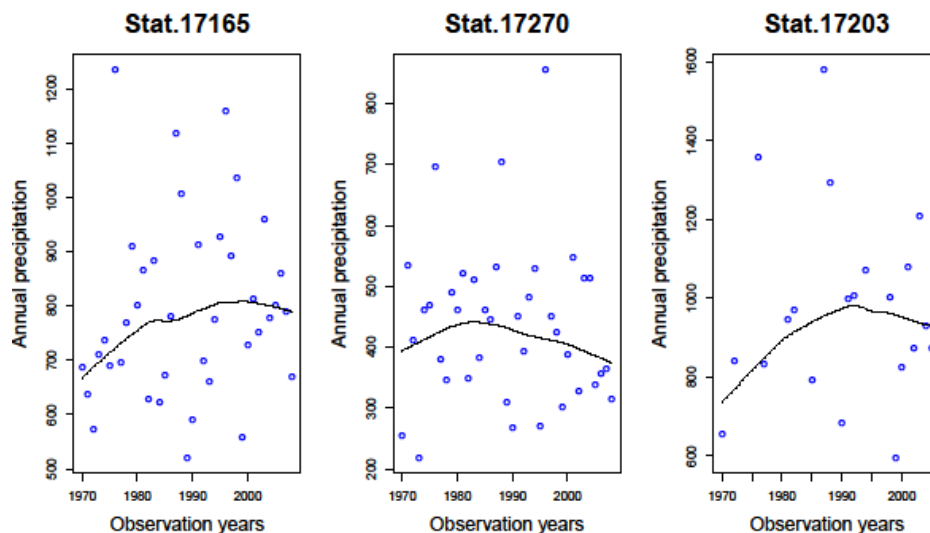

Figure 5.14. Observed annual precipitation over time at three locations. Blue dots present measurements; solid lines present smoothed curved observation lines.

Secondary variables of selected three meteorological stations are presented in Table 5.9. According to table, station 17203 has the highest elevation and the highest observed mean annual precipitation among them.

Table 5.9. Secondary variables of selected three arbitrary locations.

Stat. no	Z	V1	V2	V3	V4	V5	Mean prec.
17203	1177 m	78 m	227 km	0.36 km/sq.km	1 (north-east)	0 (artificial)	971 mm
17165	981 m	151 m	151 km	0.23 km/sq.km	1 (north-east)	0 (artificial)	791 mm
17270	553 m	35 m	35 km	0.001 km/sq.km	3 (south-west)	0 (artificial)	436 mm

5.3.1 Space-time Ordinary kriging

The space-time sample variogram and sample variograms for each time lag are given in Figure 5.15. The sample space-time variogram of observations are created for 0-10 time lags, and spatial lags with 10 km width up to 200 km. In Table 5.10, values for the first two time-lags are given. There are 20 spatial lags for the first time lag. Each spatial lag at zero-time lag has 10 km width. Lag-zero considers the records that contain data observed in the same year. So 18 pairs of records have observations measured in the same year at 5 to 8 km spatial distance. Lag-one searches records that have one year difference among them, and so on.

Table 5.10. Sample space-time variogram parameters with time lags 0 and 1 year.

np	dist	gamma	id	timelag	spacelag
18	7950.557	0.003132	lag0	0	5000
22	15423.61	0.008011	lag0	0	15000
41	28393.95	0.003343	lag0	0	25000
48	34144.18	0.013204	lag0	0	35000
166	45192.41	0.009655	lag0	0	45000
189	55590.67	0.026426	lag0	0	55000
79	65118.54	0.009134	lag0	0	65000
338	75025.56	0.02362	lag0	0	75000
148	85250.75	0.016104	lag0	0	85000
244	95342.2	0.010646	lag0	0	95000
183	105288.8	0.027788	lag0	0	105000
248	115068.1	0.022906	lag0	0	115000
348	125246.6	0.01296	lag0	0	125000
167	135621.1	0.024794	lag0	0	135000
286	146321.6	0.02454	lag0	0	145000
150	156198.6	0.012816	lag0	0	155000
312	163914.5	0.014643	lag0	0	165000
310	173891.2	0.025821	lag0	0	175000

Table 5.10. Sample space-time variogram parameters with time lags 0 and 1 year (continued).

238	185569.1	0.020037	lag0	0	185000
254	194538.5	0.017718	lag0	0	195000
580	0	0.004808	lag1	365	0
37	8390.718	0.002257	lag1	365	5000
31	15348.27	0.009912	lag1	365	15000
70	28276	0.005077	lag1	365	25000
77	33568.97	0.010726	lag1	365	35000
286	45218.53	0.010648	lag1	365	45000
362	55642.38	0.026241	lag1	365	55000
139	65418.69	0.011577	lag1	365	65000
608	75107.18	0.025154	lag1	365	75000
278	85364.3	0.016312	lag1	365	85000
452	94907.68	0.011251	lag1	365	95000
326	105195	0.030842	lag1	365	105000
435	114959.3	0.0233	lag1	365	115000
622	125131.8	0.014017	lag1	365	125000
312	135197	0.024396	lag1	365	135000
527	146379.7	0.02291	lag1	365	145000
279	156081.6	0.015667	lag1	365	155000
605	163948.4	0.015179	lag1	365	165000
583	174006.9	0.025101	lag1	365	175000
444	185584.2	0.022017	lag1	365	185000
481	194663.6	0.018573	lag1	365	195000

As seen from the sample variograms of observations, the variation in time is smaller than variation in space (Figure 5.15). The spatial range is about 50-100 km. At distances up to 50 km, differences between variograms among time lags can be seen easily. After 50 km there is no variation between variograms for different time lags. Therefore it can be said that time creates a difference within 50 km for precipitation prediction. After checking the sample variograms, sum-metric model is fitted and the space-time variogram was created. Parameters of the fitted space-time variogram are given in Table 5.11.

Table 5.11. Fitted parameters of the space-time OK variogram.

C_0 (time)	C_s (time)	C_r (time)	C_0 (space)	C_s (space)	C_r (space)	C_0 (space-time)	C_s (space-time)	C_r (space-time)	α
0 days	1e-3 days	1.5e+3 days	0 meter	2e-2 meter	5e+4 meter	0 meter	1.03e-2 meter	7e+4 meter	10 km/year

The wire-frame plot (Figure 5.16) gives a 3D presentation of the sample space-time variogram. This plot provides a different perspective to interpret spatial and temporal variation.
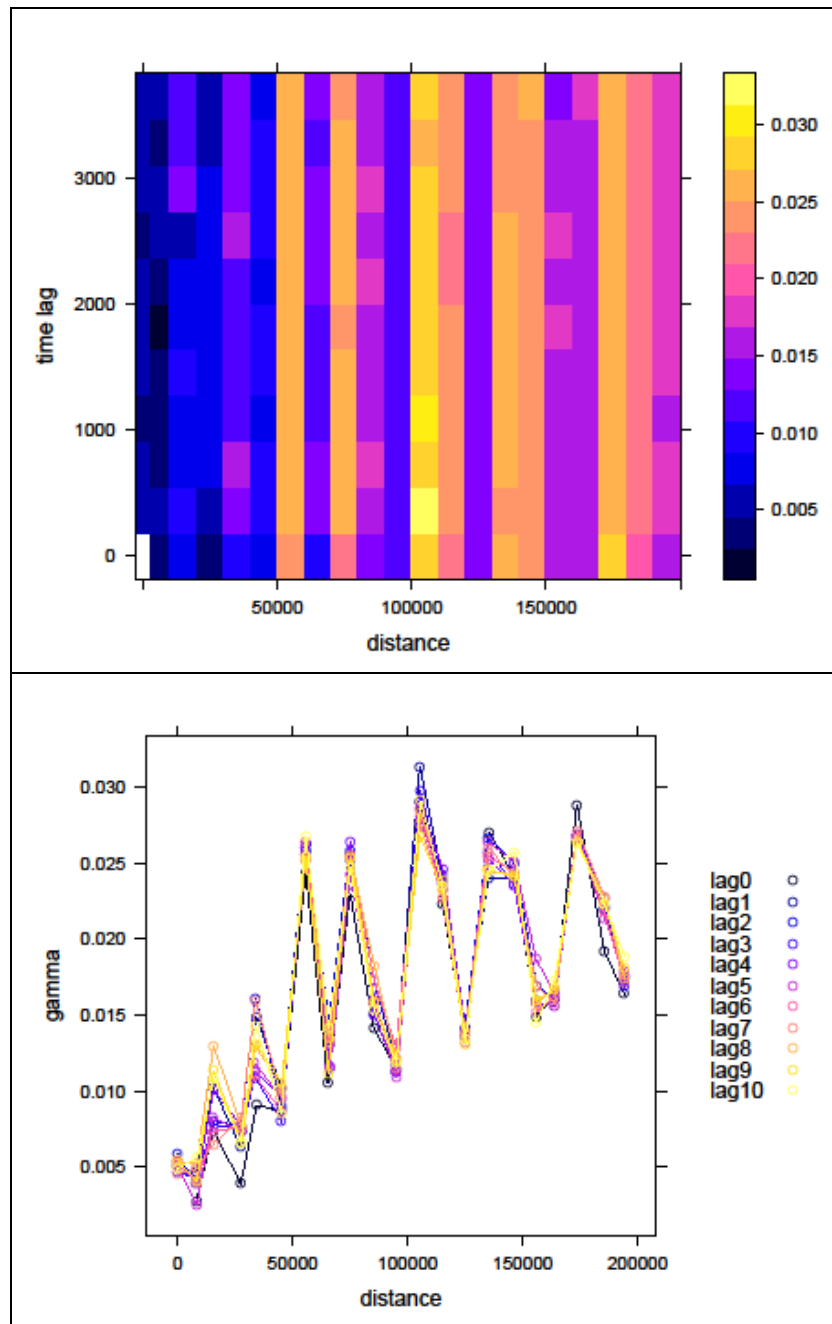
Figure 5.15. Space-time sample semi-variogram (top) and sample variograms for each time lag (bottom) of precipitation observations of a training dataset. The "time lag" expresses cumulative days, "space lag" express the spatial distance (m), "gamma" presents the semi-variance of variogram.
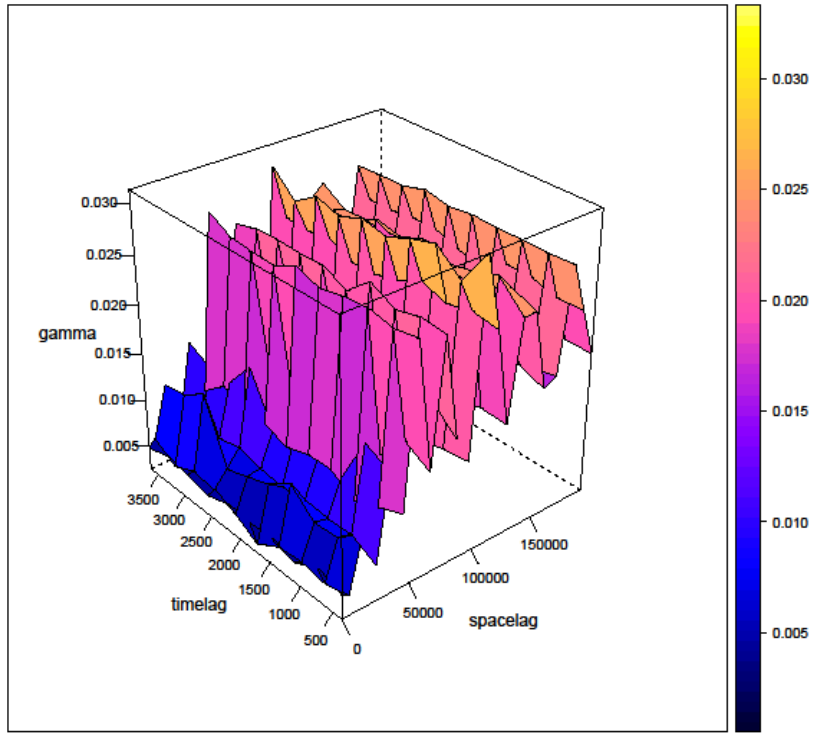
Figure 5.16. Wire-plot of sample space-time OK variogram.
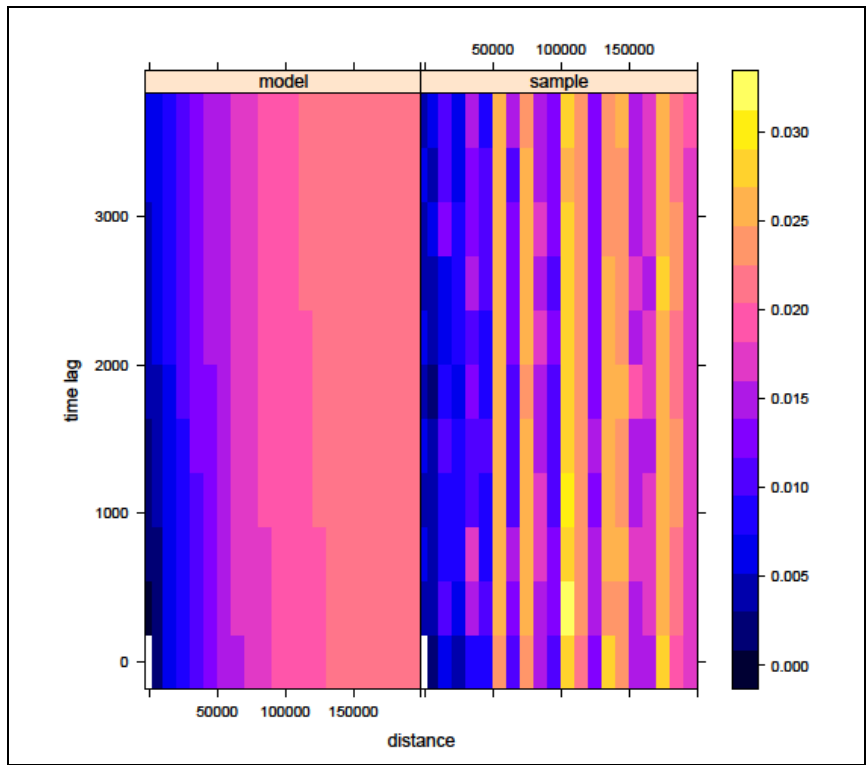

Figure 5.17. Sample (right) and modeled (left) variograms of observations of a training dataset.

Sample and modeled variogram of observations is presented in Figure 5.17.

In order to compare measurements with predictions, the three stations that were selected before as arbitrary locations are used for visual comparison. In Figure 5.18, measurements and predictions are indicated on a graph with a smooth line. Precipitation prediction lines are generally well-fitted to observation lines at stations 17270 and 17203, but predictions are not fitted to observations. Especially for station 17270, two lines are almost matched.
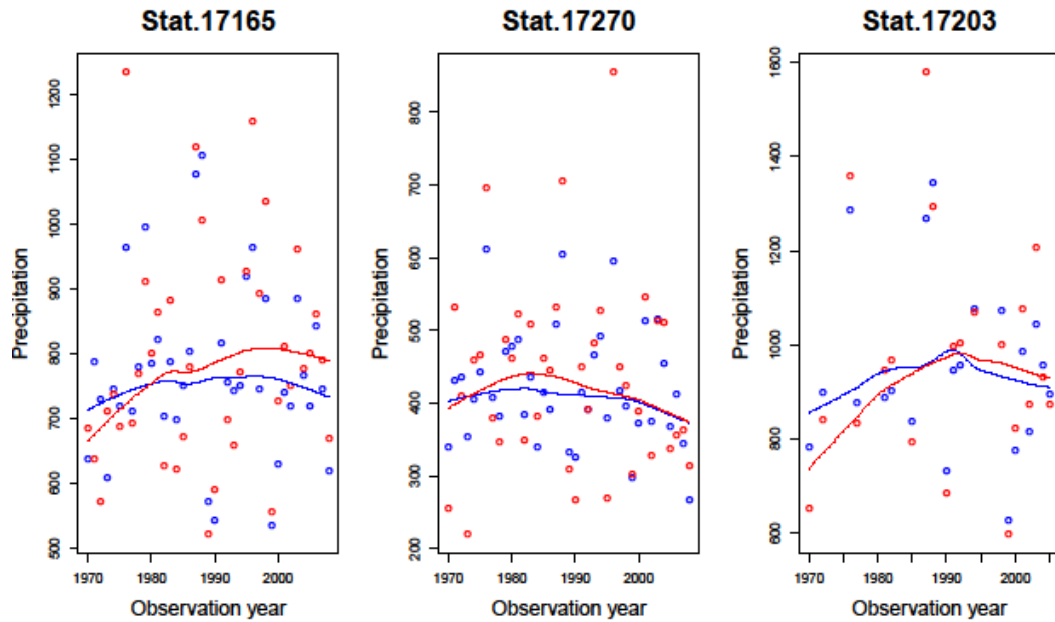
Figure 5.18. Predicted precipitation values versus predictions obtained from ST-OK for three arbitrary measurement locations; red dots are observations, blue dots are predictions, red line is smoothed curved observation line, blue line is smoothed curved prediction line.

Predictions according to each observation period are presented in Figure 5.19. As seen from the maps, ST-OK created smooth prediction maps with less detail over the area. The distribution of predictions is slightly changed and boundaries of different prediction classes are distinctive. High prediction values appear at the middle and east parts of the basin. Generally predictions are between 300-700 mm over the whole area. The maximum prediction value is about 1600 mm and predicted in 1976, 1987 and 1988. The minimum prediction value is about 120 mm and predicted in 1990, 1999 and 2008.

59
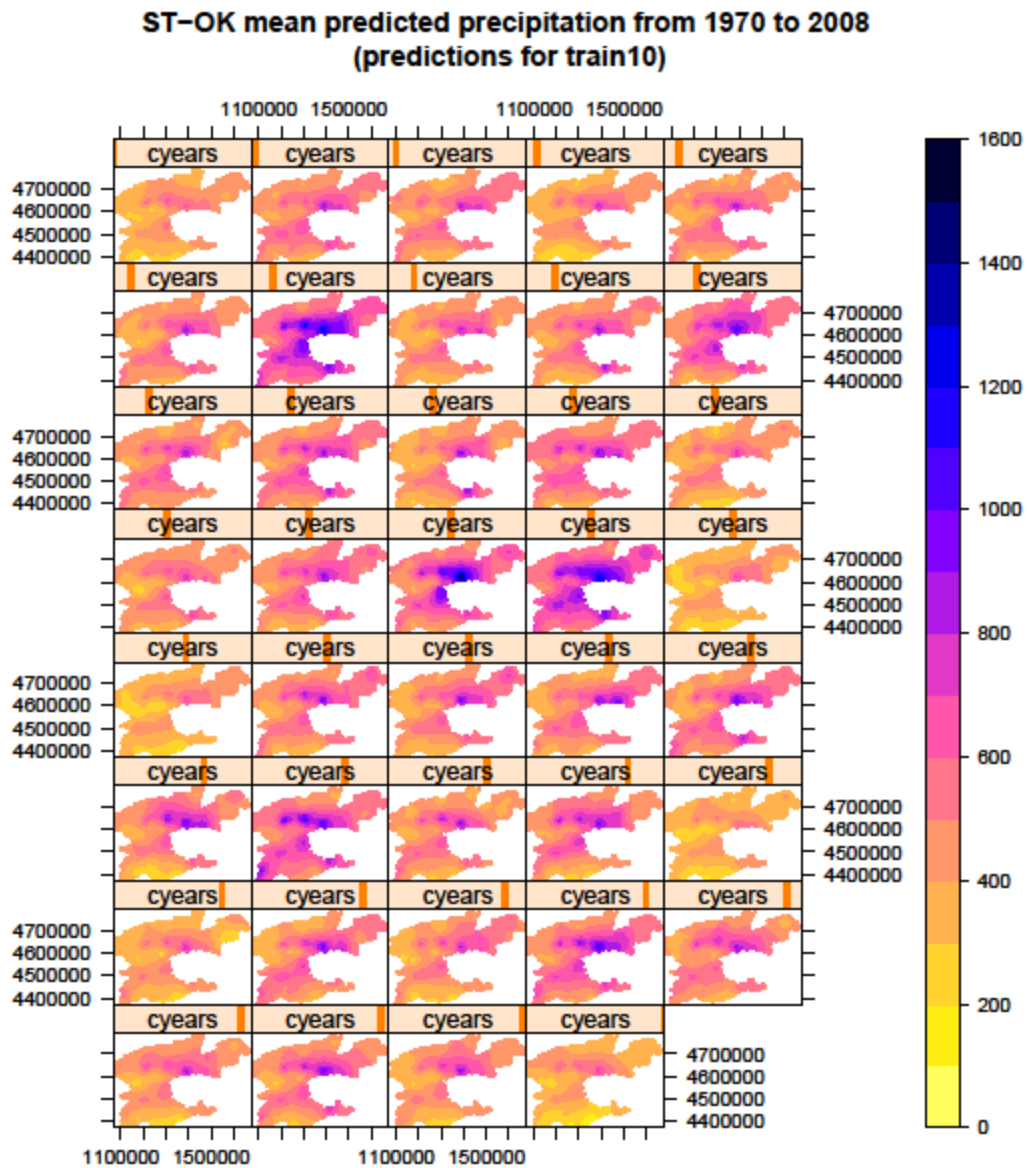
Figure 5.19. ST-OK mean predicted precipitation maps according to each observation year, "cyears" represent observation years which starts in 1970 and ends in 2008.

Standard deviations of ST-OK predictions are presented in Figure 5.20. Generally standard deviations are small around meteorological stations and large at greater distances from the stations. Standard deviations vary over the area between 0 and 200 mm. Standard deviations around stations are small in 1989, 1990, 1999 and 2008. Distribution of stations and number of measurements are affecting the standard deviation predictions.
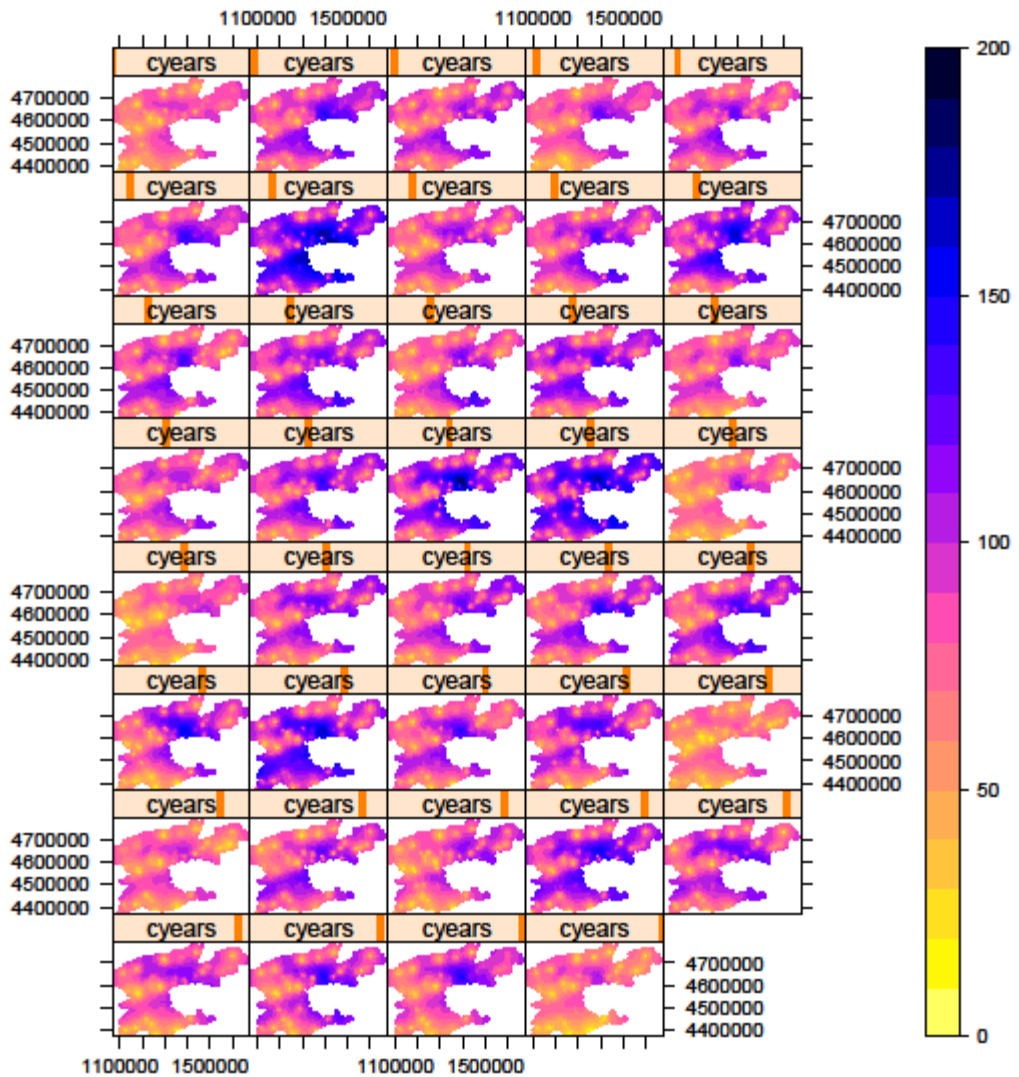
Figure 5.20. Space-time OK standard deviations of annual precipitation predictions (mm).

5.3.2   Space-time Universal kriging

For universal kriging the variogram is created from the residuals after trend removal. For this purpose MLR analysis was first performed to the log-transformed precipitation data. From the model outputs log-transformed residuals were obtained. The coefficients of the fitted regression model are given in Table 5.12. All of the continuous variables are significant but from the categorical variables some are not.

Table 5.12. Model coefficients of the MLR application of Euphrates dataset using secondary variables and interactions.

Coefficients					
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**	**Signif. code**
(Intercept)	5.278e+00	9.3e-01	5.67	2.0e-08	***
Z-elevation	-2.39e-04	4.63e-05	-5.17	2.9e-07	***
V1- Surface roughness	8.98e-04	1.18e-04	7.57	1.0e-13	***
V2- Distance to nearest coast	-1.25e-03	2.11e-04	-5.95	3.9e-09	***
V3- River density	9.74e-02	4.76e-02	2.04	0.04	*
Year	-1.2e-03	4.67e-04	-2.67	0.001	**
V42- South-east	2.54e-02	1.88e-02	1.35	0.17	
V43- South-west	-1.99e-02	1.54e-02	-1.29	0.19	
V44- North-west	-6.35e-02	1.5e-02	-4.05	5.6e-05	***
V51- Land cover class: agricultural	-2.66e-02	1.5e-02	-1.68	0.092	.
V52- Land cover class: wetlands	-5.47e-02	5.85e-02	-0.93	0.35	
V53- Land cover class: open-space	-4.5e-02	2.82e-02	-1.61	0.10	
V54- Land cover class: vegetation	1.01e-01	2.8e-02	3.57	0.001	***
V55- Land cover class: forest	9.39e-02	2.89e-02	3.24	0.001	**
Z:V2- interaction between elevation and dist. to coast	1.564e-06	2.10e-07	7.43	2.6e-13	***
Signif. code	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' ' 1

After obtaining residuals, the sample space-time variogram are created for 0:10 time lags, 10 km width up to 200 km as done for ST-OK. Space-time sample variogram and sample variograms for each time lag are presented in Figure 5.21. According to these figures, the variogram reaches to maximum value on *Y* axis at about 100 km. So it can be said that the spatial range is about 100 km. In this case the spatial sill is around 0.015. Spatial and temporal nugget can be accepted as zero. It is difficult to decide on the temporal range and sill from these figures since the variation of the variogram values is not so evident. The wire-frame plot (Figure 5.22) is somewhat helpful to resolve the temporal variogram parameters. According to this the temporal range is about 1500-2000 days and sill value is about 0.008. From these figures by using sample variogram parameters the model is fitted by using sum-metric method. Modeled variogram parameters are given in Table 5.13. Fitted variogram of residuals is shown in Figure 5.23.

Table 5.13. Fitted parameters of the space-time UK variogram.

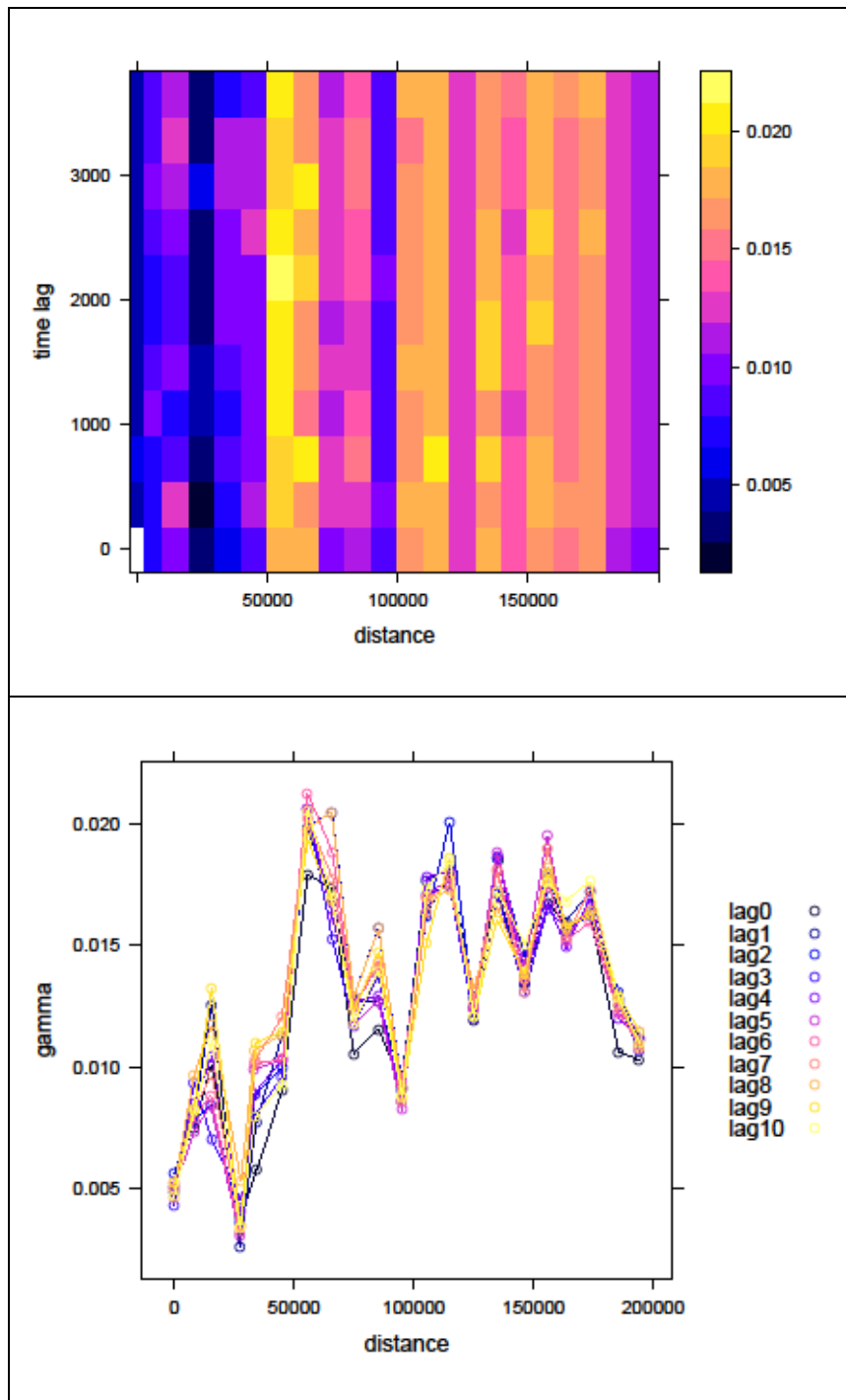| C_0 (time)	C_s (time)	C_r (time)	C_0 (space)	C_s (space)	C_r (space)	C_0 (space-time)	C_s (space-time)	C_r (space-time)	α
0 days	1.0e-4 days	2.0e+3 days	3.7e-3 meter	5.0e-3 meter	1.0e+5 meter	3.6e-3 meter	5.0e-3 meter	1.0e+5 meter	10 km/year

Figure 5.21. Space-time sample semi-variogram (top) and sample variograms for each time lag (bottom) of regression residuals of a training dataset.
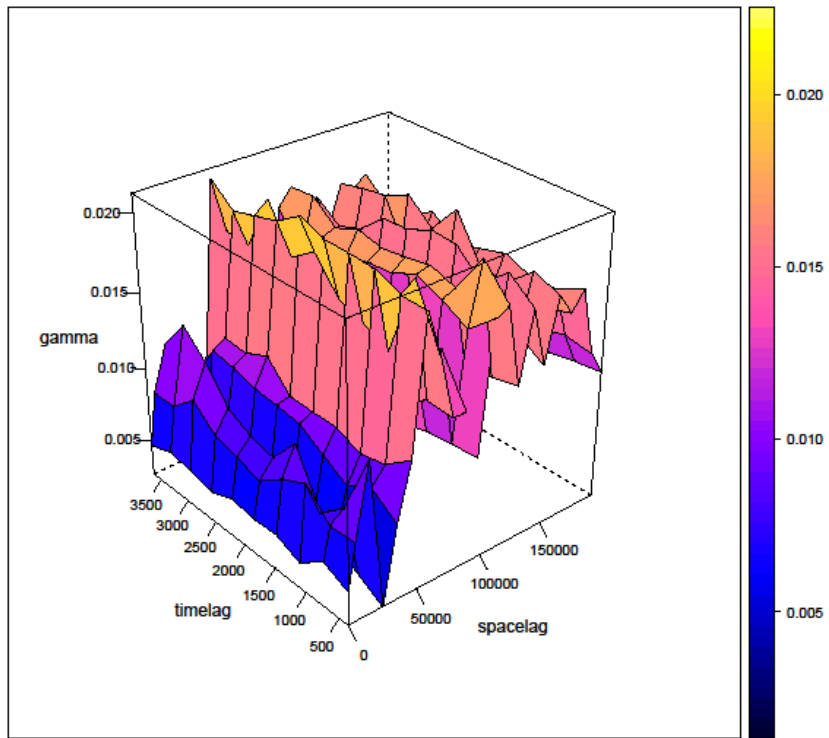
Figure 5.22. Wire-plot of sample space-time UK variogram.
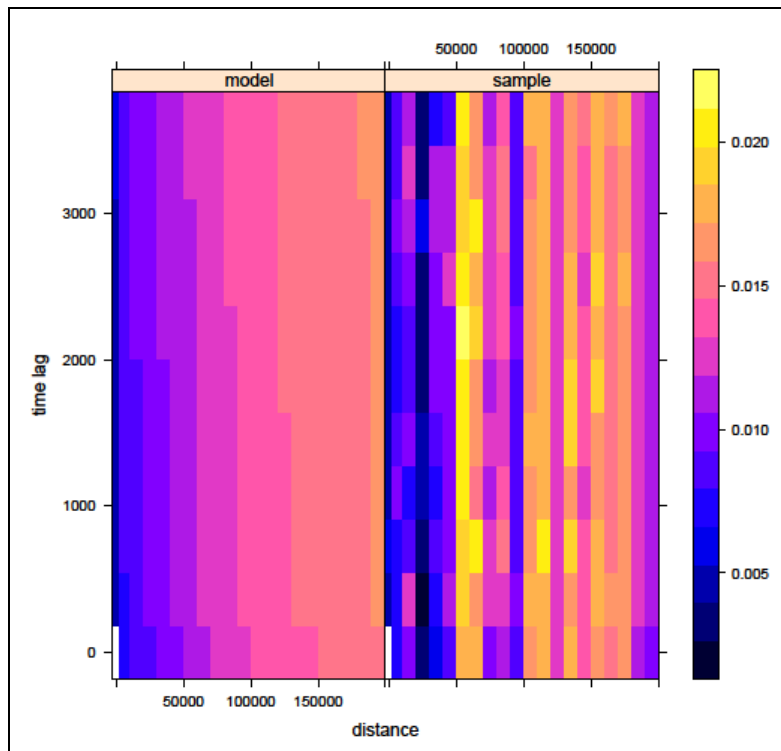
Figure 5.23. Sampled (right) and modelled (left) variograms of regression residuals of a training dataset.

Observations and predictions of three measurement locations are shown in a graph (Figure 5.24). Generally the fit line of predictions does not accurately match the observation line. At station 17270 a decreasing precipitation tendency is predicted and at station 17203 increasing precipitation tendency is predicted with ST-UK. For the station 17165 the prediction line is not very close to the observation line.
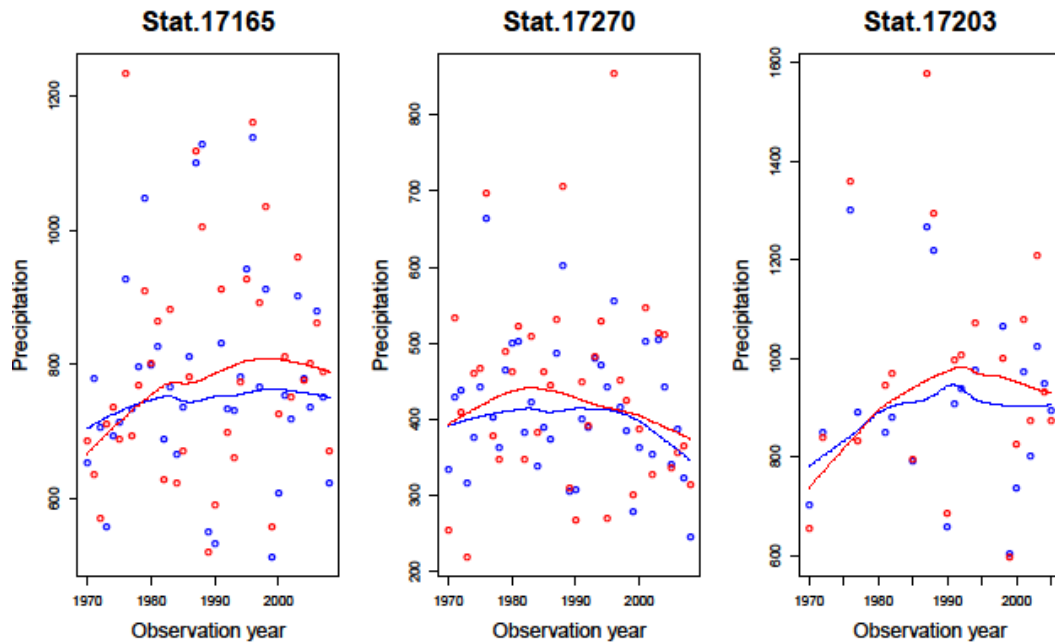
Figure 5.24. Predicted precipitation values versus predictions obtained from ST-UK at arbitrary three measurement locations; red dots are observations, blue dots are predictions, red line is smoothed curved observation line, blue line is smoothed curved prediction line.

Mean predicted precipitation values for each observation time are presented in Figure 5.25. Prediction maps have more detail than those of the ST-OK method. The distinction between classification groups is not so evident.

Maximum precipitation is predicted about 2000 mm and minimum prediction is about 160 mm. According to these maps, rainy years are 1976, 1987, 1988 and 1996. The dry years are 1989, 1990, 1999 and 2008. In general middle and east parts of the basin have higher precipitation than other regions. The highest precipitation predictions are located around the dams in the basin. The southern parts of basin are the driest areas.
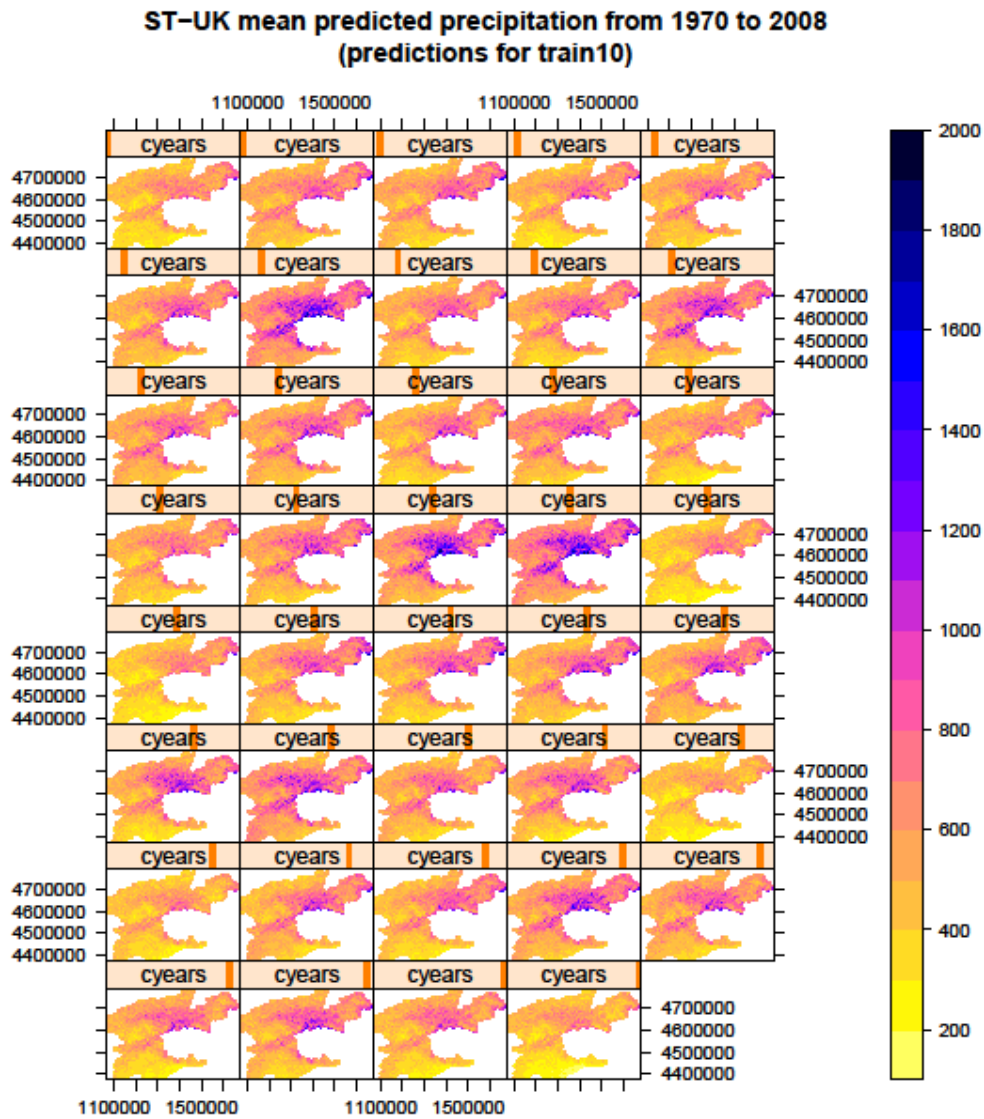
Figure 5.25. Space-time UK mean precipitation predictions for each observation period.

Prediction standard deviations for each year are presented in Figure 5.26. ST-UK standard deviations are predicted higher than ST-OK standard deviations (Figure 5.20). Maximum standard deviation is about 500 mm and minimum standard deviation is about 30 mm.

It is interesting that the standard deviations are not directly connected only to the number of observations of each year. It is also related with distribution of meteorological stations as ST-OK. In general, the values are continuously changing and there are no clear boundaries among standard deviation classes (Figure 5.26).
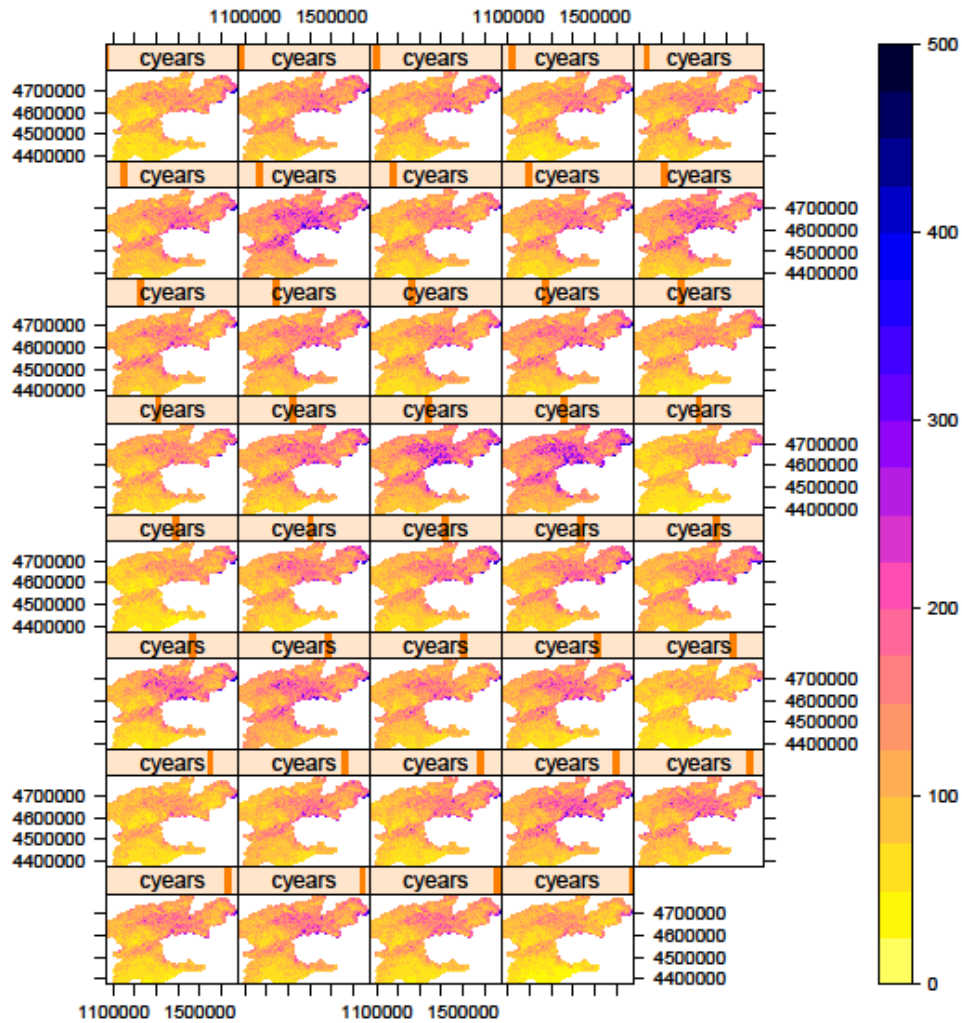
Figure 5.26. Space-time UK standard deviations of predictions.

Accuracy assessment of space-time kriging methods were performed with *RMSE*, *R-square* and *ME*. Methods were applied ten times to each training datasets to obtain predictions at all test locations. The predictions for the ten test datasets were then combined into a single file. Next the *RMSE*, *R-square* and *ME* were calculated. The results are given in Table 5.14. Mean annual precipitation of Euphrates Basin is 508 mm. *R-square* values are 0.86 and 0.73; *RMSE* values are 75 mm and 107 mm; and *ME* values are 57.3 mm and 81.7 mm for ST-OK and ST-UK, respectively. Contrary to expectations, ST-UK accuracy measures are worse than ST-OK. The reason for this can be because of secondary variables. ST-UK results are obtained using continuous, categorical and "Year" variable. In order to test effects of secondary variables, a second trial was performed for ST-UK. In this experiment only continuous environmental variables namely elevation, surface roughness, distance to coast, river density and elevation-distance to coast interaction were used. The obtained accuracy assessment results are more encouraging (Table 5.15). Possibly the difference between these two results is related to the categorical variables. The meteorological stations at the basin may not capture the whole characteristics of categorical variables. For land cover 60% of all stations have the artificial surface sub-class, approximately 25% of stations have agricultural sub-class, 1% of stations have wetlands

67

sub-class, 8% of stations have open-spaces sub-class, 3% of stations have vegetation sub-class, and finally 3% of stations have forest sub-class. Since the sub-classes of land cover could not be expressed accurately by meteorological stations, kriging weights may not be suitable to all data in the trend estimation part of the first ST-UK case. In the second case obtained accuracy measures are better because of probably using only continuous and deterministic variables. For sure it is just a hypothesis that using only continuous variables may help to obtain more reliable predictions. It needs to be researched in more detail to get a definite answer.

Table 5.14. Cross-validation results of space-time kriging methods.

Cross-Validation Results	
Mean prec.	508.4 mm
ST-UK.*R-square*	0.73
ST-UK.RMSE	107.1 mm
ST-UK.ME	81.7 mm
ST-OK.*R-square*	0.86
ST-OK.RMSE	75.1 mm
ST.OK.ME	57.3 mm

Table 5.15. Cross-validation results of space-time kriging methods by using different dataset in ST-UK.

Cross-Validation Results	
ST-UK.*R-square*	0.85
ST-UK.RMSE	78 mm
ST.UK.ME	58.5 mm
ST-OK.*R-square*	0.86
ST-OK.RMSE	75.1 mm
ST.OK.ME	57.3 mm

The results obtained in the second experiment are very close to each other (Table 5.15). ST-OK resulted with quite reliable predictions over the Euphrates Basin for precipitation prediction. Although ST-UK is not much better than ST-OK according to the cross-validation results, the resulting maps of ST-UK have more realistic patterns than ST-OK. Prediction maps of ST-OK are quite smooth (Figure 5.19), but prediction maps of ST-UK (Figure 5.25) have more spatial detail.

5.3.3   Comparison of Spatial and Space-time Universal Kriging

Spatial universal kriging and space-time universal kriging are selected because these are accurate interpolation methods to compare spatial and space-time methods. To evaluate the methods the same data conditions are used in both frameworks. Consequently the space-time data is separated into groups according to each observation year. In this way each year has its own database and is interpolated with spatial UK independently. Cross-validation is made in each dataset to assess the performances. 10% of each data set is selected for testing and the remaining part is allocated for training, and each data point is used in a test dataset exactly once.

Since the dataset is created from sparse space-time data, each data-year has a different number of observations. The minimum observation density was obtained for 1973; only eight stations have

observations for that year. Maximum observation density was obtained for 2002; 38 stations have data for this year. Continuous variables elevation, surface roughness, distance to coast and river density were used for the kriging applications. In Figure 5.27, predicted precipitation maps are presented according to each observation year. Generally the north parts of the basin have more precipitation than the south parts for all years. The prediction varies between 150 and 1800 mm. From a general perspective, there are some noises at prediction maps and sharp boundaries between predictions.
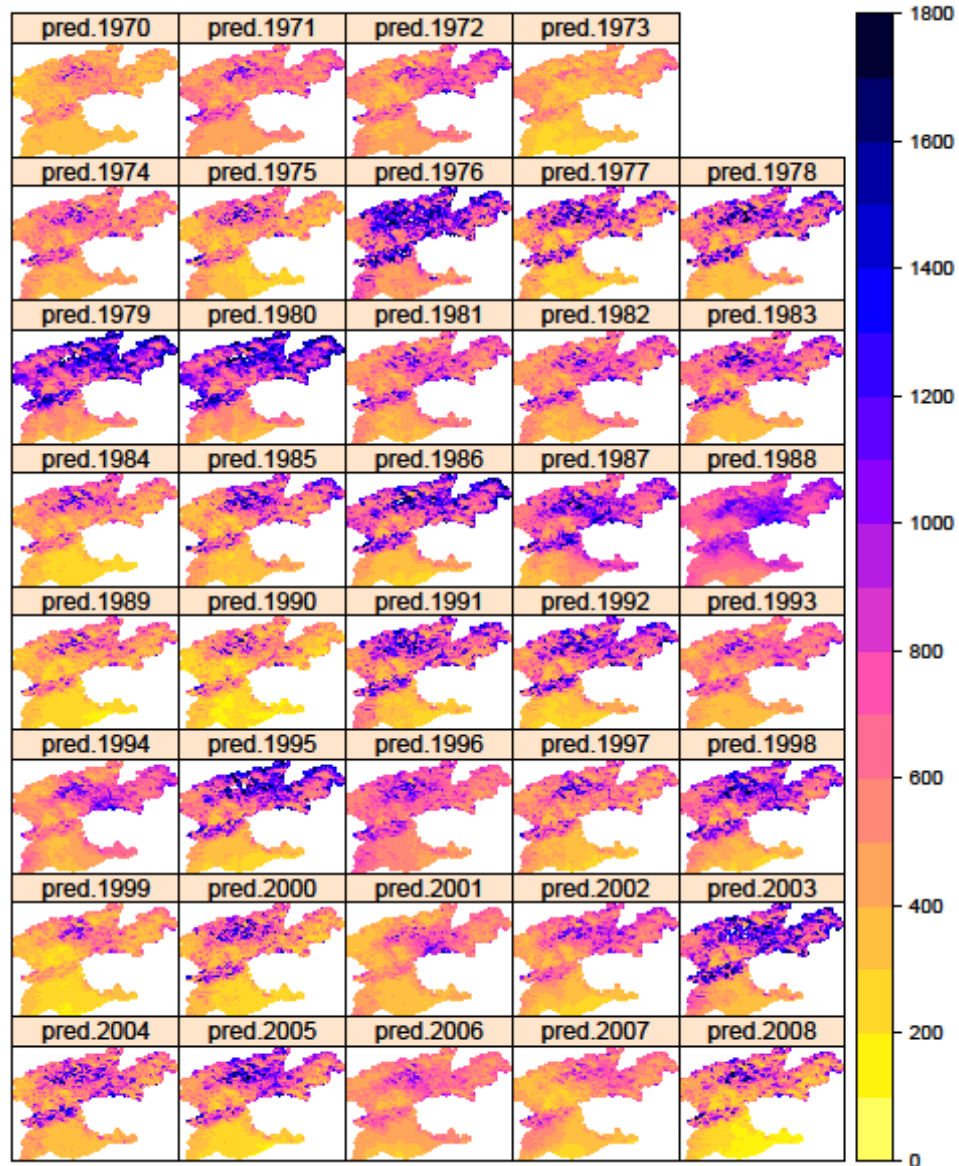
Figure 5.27 Spatial UK precipitation predictions for each year of the Euphrates Basin.

The prediction maps have reasonable accuracy, but when compared with the results of space-time kriging these are not sufficient because the data are analyzed without using interaction between consecutive years. The 39-year average *R-square* is 0.13 and the *RMSE* is 190 mm.

Standard deviations of spatial UK predictions are presented in Figure 5.28. Values differentiate according to number of observations and their spatial distribution. The standard deviation values are considerably larger than compared with those of space-time kriging.
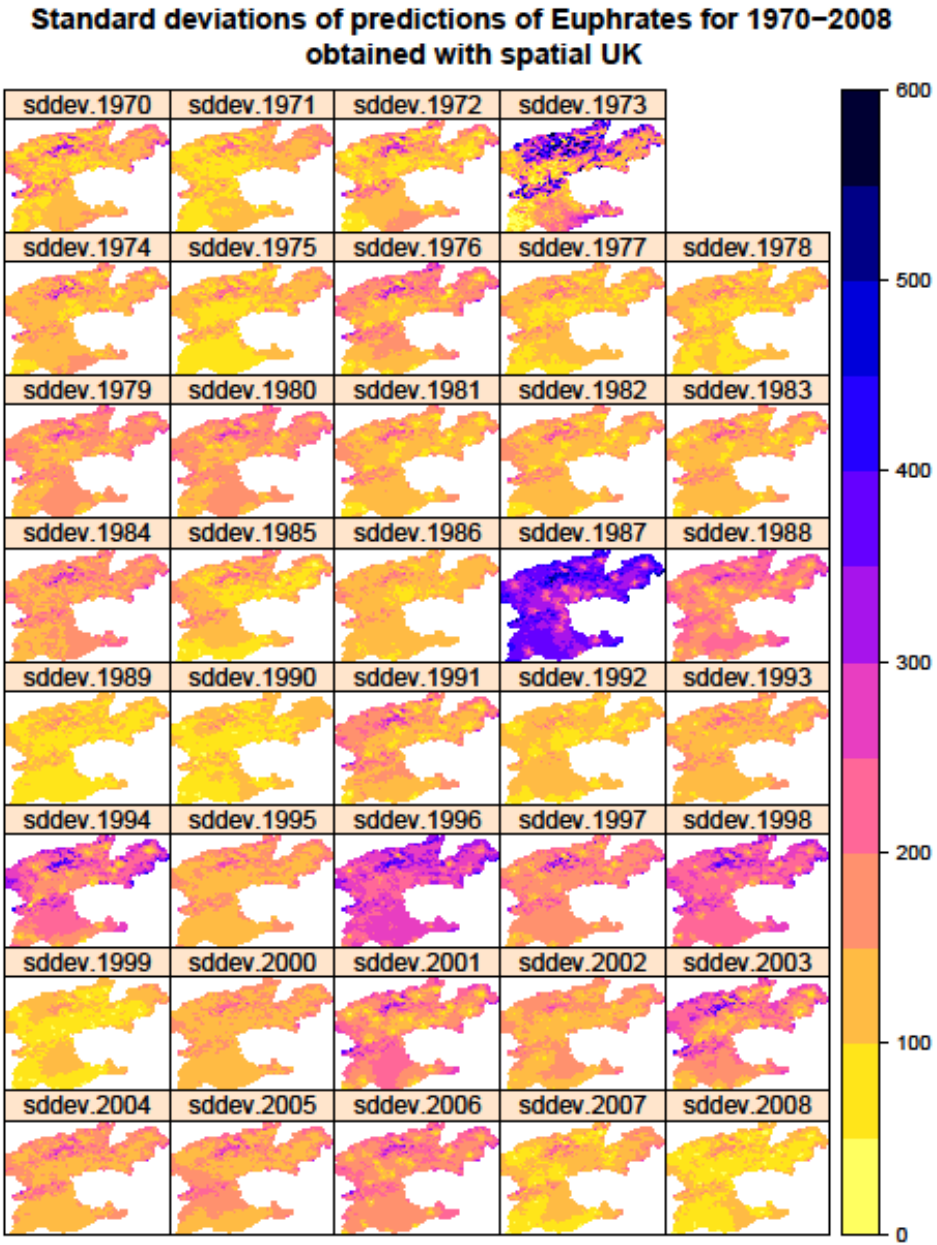
Figure 5.28. Spatial UK predicted standard deviations (mm) for each year of Euphrates Basin.

CHAPTER 6

CONCLUSIONS

In this thesis, temporal, spatial and space-time variation and distribution of precipitation were analyzed. Annual and seasonal precipitation values of Turkey measured at 225 meteorological stations for 1970-2003 years were used for temporal analyses. Long-term annual average precipitation of Turkey measured at 225 stations for 1970-2006 were used for spatial analyses. Annual totals of the Euphrates Basin measured at 47 meteorological stations for 1970-2008 were investigated with a space-time kriging procedure.

The aim of the temporal analysis of precipitation is to understand if there are systematic fluctuations in any direction through time. Defining fluctuations at precipitation in positive and/or negative direction may be helpful for water management studies, and agricultural and climatic planning. Before application of statistical tests, visual variations in time were examined. Visual analysis was performed on seasonal values instead of annual values because seasonal values can be a more accurate indicator of trend existence. Increasing and/or decreasing trends were realized at some basins with visual analysis. In addition, annual and seasonal totals were analyzed with the T-Test and Mann-Kendal test. Trend tests were applied firstly to original data and secondly to data without serial correlation. The main findings are:

- Generally increasing precipitation trend was experienced for annual precipitation in middle and northern parts of Turkey for both tests but with few meteorological stations.
- A decreasing annual precipitation trend was observed at three stations at Van Golu, Sakarya and Kizilirmak basins.
- Increasing precipitation trend was observed in spring, autumn and winter seasons.
- For winter season, before removing serial correlation, 19 meteorological stations and after removal of serial correlation 11 meteorological stations had a significantly increasing precipitation trend in the winter season. This trend is generally observed in East Anatolia, Black Sea, East of Mediterranean and Aegean regions.
- A decreasing precipitation trend is significant in the Aegean region in the summer season and should be examined with up-to-date data for future studies.
- Comparison of the T-Test and the Mann-Kendal test and make a decision about which one is better are difficult issues since both of these gave similar results about trend presence over Turkey. If data has normal or near normal distribution, T-Test otherwise Mann-Kendal test can be used for trend detection.

The aims of the spatial analysis part of this thesis were to reveal spatial distribution and variation of long-term annual precipitation over Turkey and to compare and discuss regression and kriging techniques in spatial interpolation and extrapolation. The MLR, GWR, OK, RK and UK were used to reveal spatial distribution of precipitation. Mean monthly precipitation values from 1970 to 2006 were measured from 225 meteorological stations distributed fairly regularly over Turkey, with an increased density of stations near the Turkish border. Long term averages were computed from these monthly values by averaging over the total period of 37 years. The original precipitation values had a skewed distribution, and log-transformation was therefore applied to convert these to a near-normal distribution. Results were back-transformed to the original scale of measurement. Instead of using only precipitation data, explanatory variables that could be related with precipitation distribution were used during model implementations to obtain higher prediction performance. The explanatory variables elevation, distance to coast, land cover type and interaction between elevation and distance to coast were statistically significant predictors of the annual rainfall as shown in a multiple linear regression analysis. The main results can be summarized as follows:

71

- UK was the most reliable method for spatial interpolation of the precipitation distribution of Turkey according to the *RMSE*, *R-square* and *SMSE* performance assessment methods. This model was followed by RK, OK, GWR, UK with elevation and finally MLR.
- Adding residual kriging improved the prediction performance as observed with the RK and UK models.
- Prediction maps of UK, RK, GWR and MLR models were similar, but high precipitation values in north-east Anatolia were identified better with the RK and UK models.
- Prediction maps of OK and UK with elevation show similarities and are smoother than maps obtained with models that use all significant covariates.
- Although the OK prediction map does not show spatial detail, cross-validation results show that it performs better than MLR and GWR. UK with elevation also has a higher performance than MLR. This implies that in this study kriging methods are superior to MLR for interpolation.
- Although MLR has the highest *RMSE*, it was a useful method to identify which environmental covariates had a significant effect on long-term average precipitation.
- Local estimation of regression coefficients did not improve spatial interpolation as GWR was outperformed by the kriging models.
- For the GWR model, prediction standard deviations could not be calculated. This is a serious drawback of this method.
- Estimation errors were high at stations that had very high or low observed values. Under-estimations were detected at stations that had very high measurements (especially the north-east parts of Turkey). Similarly precipitation was overestimated at stations that have low observations. This is a typical result of the smoothing effect of kriging.
- Unlike the interpolation analysis, at extrapolation MLR performed best according to *R-square* and *RMSE*. GWR, RK, and UK are second-best methods. OK and UK with elevation obtained very low accuracies when used for extrapolation.
- The similarity between the training and testing datasets of statistics directly affects the performance of interpolation and extrapolation methods.
- The lowest *RMSE* obtained with interpolation (178 mm) is still high when considering the average annual precipitation (628 mm) of Turkey. Apparently, covariates can only explain part of the variation and spatial correlation of the residual variation is not sufficiently strong to dramatically reduce the kriging error. While spatial extrapolation benefits most from covariate information as shown by an *RMSE* reduction of about 60 mm, in this study covariate information was also valuable for spatial interpolation because on average it reduced the *RMSE* with 30 mm.

In future studies to improve spatial distribution of precipitation other relevant explanatory variables can be included to further decrease the *RMSE* and obtain a higher *R-square*. For instance, remote sensing-based imagery such as MODIS data appear promising. Platnick et al. (2003) performed a study using MODIS cloud products, algorithms and examples from the Terra satellite. They investigated and made an overview of MODIS cloud products such as cloud mask, cloud-top properties, cloud thermodynamic phase, cloud optical thickness and cloud microphysical properties which have been used in many application areas like climatology, weather prediction and atmospheric research. Hengl et al. (2012) performed space-time kriging interpolation to predict daily temperatures using ancillary variables such as MODIS LST images, coordinates, distance from the sea, elevation, time and insolation. According to their findings using space-time regression kriging and including time-series data such as remote sensing images produce more accurate maps than using only spatial methods. In light of these studies, MODIS products can be used for precipitation prediction to obtain more reliable results. Topography has also large effect on precipitation but not linearly. In this thesis the effect of topography is tried to be represented linearly, but spatially variable non-linear relationship may give better interpolation results.

The last analysis part of thesis was the application of space-time kriging methods to annual precipitation values to a river basin in Turkey. The aims were to obtain more accurate prediction results than spatial kriging and to compare space-time ordinary (ST-OK) and universal kriging (ST-

UK) methods. In addition spatial and space-time kriging methods were compared using the same dataset of the Euphrates Basin, Turkey. The reason of changing the dataset is, to test more recent technique on smaller area as space-time interpolation techniques require more data and computation time. The Euphrates Basin is the biggest and most productive river basin of Turkey. Also it has the largest number of meteorological stations. 47 meteorological stations which composed of big and small climate stations are used at space-time analyses. Generally stations have uniform distribution over basin but a little bit high concentration can be seen near dams. In addition, most of them are located at lower altitudes compared to average altitude of basin. This means that areas have high altitudes may not be described accurately by meteorological stations. This situation is also valid for land cover and eco-region secondary variables. Therefore it can be said that distribution and number of stations over the area and selecting efficient secondary variables are important issues for general kriging purposes.

The most difficult part of space-time interpolation is to constitute the space-time variogram, its interpretation and modeling. Also this is the main difference from spatial interpolation. Environmental variables generally behave differently in time and space. To handle this variation accurately and reflect that variation in the space-time variogram are challenging parts of space-time geostatistical interpolation. Secondary variables that vary in space but are static in time (elevation, surface roughness, distance to coast, river density, aspect and land cover) were used by the spatial and space-time UK methods. ST-UK was performed twice using the full dataset and with only continuous variables, to test the usefulness of categorical variables. ST-OK does not use secondary variables. The data set comprised of 906 space-time observations measured from the period of 1970-2008. 26 meteorological stations have fewer than 20 years of observations and the other 21 meteorological stations have more than 20 years observations. Only nine meteorological stations have full (39-year observation) data. Annual precipitation values which were summed from monthly measurements were used. Ten-fold cross validation was applied to the whole dataset to make performance comparisons. *RMSE, R-square* and *ME* were computed for testing data sets. Spatial kriging to predict precipitation for each observation year was also performed by using continuous variables to compare spatial and space-time interpolation techniques. Spatial universal kriging was selected as it was thought that it is the most reliable spatial prediction method.

The main findings of space-time interpolation are:

- At first ST-UK application elevation, surface roughness, distance to coast, river density, land cover, Year and elevation-distance to coast interaction were used. According to performance assessment results of cross-validation, *R-square* is calculated as 0.73 and *RMSE* is 107 mm and *ME* is 81.7 mm.
- In the second application of ST-UK elevation, surface roughness, distance to coast, river density and elevation-distance to coast interaction were used. The obtaining results are more reliable and accurate. This time *R-square* is calculated as 0.85 and *RMSE* is 78 mm, *ME* is 58.5 mm.
- For ST-OK the results of *R-square* is 0.86 and *RMSE* is 75 mm and *ME* is 57.3 mm.
- Contrary to expectations, ST-OK method resulted in more accurate prediction values than ST-UK according to *R-square, RMSE* and *ME* values. Since most of the meteorological stations are located at lower elevations compared to basin's mean elevation, the secondary variables may not be representative parameters to precipitation prediction in the basin.
- However prediction maps of ST-UK can be regarded as more realistic than ST-OK since maps are not so smooth.
- The prediction maps of ST-OK have smooth appearance as details have disappeared during interpolation.
- The kriging standard deviations of ST-OK are substantially smaller than ST-UK standard deviations. In this study, this means that error tolerance is high when using secondary variables in ST kriging to predict annual precipitation.

- The requirement of secondary variables and selecting the most effective and useful ones for space-time interpolation are important issues although using only primary variable in space-time kriging is also promising as demonstrated in this thesis.
- If the aim of space-time interpolation is to obtain the highest *R-square* and lowest *RMSE* than the analyst can use ST-OK for precipitation prediction without including related secondary variables. If the aim is to obtain maps that have more details then one can use ST-UK. In this situation the analyst should think carefully about selecting secondary variables.
- When comparing space-time and purely spatial interpolation techniques, ST-UK results are more accurate than spatial UK method. This is expected because more data is used in space-time interpolation techniques. In space-time interpolation, time information with spatial information are integrated and analyzed jointly in kriging. The information of previous, current and following years are used to predict at the present time. This improves the prediction performances of interpolation methods.
- It can be concluded that space-time kriging may help to obtain more reliable predictions if the variable of interest varies in space and time.

In order to improve the space-time interpolation study conducted here, the data set can be increased. Using a full space-time data set without deficiencies (47 meteorological stations × 39 years of observations) would yield more accurate prediction maps. Instead of using annual values, monthly precipitation observations can also be used in space-time interpolation. Secondary variables that have a relationship with precipitation should be carefully selected. Temporally changeable variables such as remote-sensing derived temperature, atmospheric moisture, a combination of drainage density and geology indicating the infiltration capacity of soil can be used to improve the geostatistical prediction of precipitation as well.

All the analysis performed in this thesis depends on the number of meteorological stations. In order to obtain better results the number of stations must be increased and location of stations must be selected according to the representativeness of the stations for the area.

REFERENCES

Arslan-Alaton, İ., Eremektar, G., Torunoğlu, P.O., Gürel M., Övez S., Tanik A., Orhon D. (2005). "Türkiye'nin Havza Bazında Su- Atıksu Kaynakları ve Kentsel Atıksu Arıtma Potansiyeli". İTU Dergisi, Cilt: 4, Sayı:3,13-21.

Aziz, O.A., Burn, D.H., (2006). "Trends and variability in the hydrological regime of the Mackenzie River Basin", Journal of Hydrology 319, 282-294.

Boer, E. P. J., Beurs, K. M., Hartkamp, A. D., (2001). 'Kriging and thin plate splines for mapping climate variables' JAG l Volume 3 - Issue 2, 146-154.

Bogaert, P., (1996). "Comparison of Kriging Techniques in a Space-time Context", Mathematical Geology, volume 28, No: 1.

Bostan, P.A., Heuvelink, G.B.M., Akyürek, S.Z., (2012). "Comparison of Regression and Kriging Techniques for Mapping the Average Annual Precipitation of Turkey", International Journal of Applied Earth Observation and Geoinformation 19, 115-126.

Brundson, C., McClatchey, J., Unwin, D.J., (2001). "Spatial Variations In The Average Rainfall– Altitude Relationship In Great Britain: An Approach Using Geographically Weighted Regression", Int. J. Climatol. 21: p.455–466.

Brus, D.J., Heuvelink, G. B. M., (2007). 'Optimization of sample patterns for universal kriging of environmental variables', Goederma 138, 86-95.

Cannarozzo, M, Noto, L.V., Viola, F., (2006). "Spatial distribution of rainfall trends in Sicily (1921-2000)", Physics and Chemistry of the Earth 31, 1201-1211.

Carrera-Hernández, J.J, Gaskin, S.J., (2007). "Spatio temporal analysis of daily precipitation and temperature in the Basin of Mexico", Journal of Hydrology 336: p.231-249.

Chen, H., Guo, S., Xu, C., Singh, V.P., (2007). "Historical temporal trends of hydro-climatic variables and runoff response to climate variability and their relevance in water resource management in the Hanjiang basin", Journal of Hydrology 344, 171-184.

Colombo, T., Pelino V., Vergari S., Cristofanelli P., Bonasoni, P., (2007). "Study of temperature and precipitation variations in Italy based on surface instrumental observations", Global and Planetary Change 57 (2007) 308–318.

Fotheringham, A.S., Brunsdon, C., Charlton, M. (2002). 'Geographically weighted regression—the analysis of spatially varying relationships', Wiley, Chichester.

Gething, P.W., Atkinson, P.M., Noor, A.M., Gikandi, P.W., Hay, S.I., Nixon, M.S., (2007). "A local space-time kriging approach applied to a national outpatient malaria data set". Computers & Geosicences 33, 1337-1350.

Gilardi, N., Bengio, S., (2000). 'Local machine learning models for spatial data analysis', Journal of Geographic Information and Decision Analysis, volume 4, number 1, 11-28.

Goovaerts, P., (2000). 'Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall', Journal of Hydrology 228, 113–129.

Grimes, I. F. D., Pardo-Iguzquiza, E., (2010). 'Geo-statistical Analysis of Rainfall', Geographical Analysis 42, 136-160.

Haberlandt, U., (2007). 'Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event', Journal of Hydrology 332, 144– 157.

Harris, P., Fotheringham, A.S., Crespo, R., Charlton, M., (2010). 'The Use of Geographically Weighted Regression for Spatial Prediction: An Evaluation of Models Using Simulated Data Sets', Math Geosci, DOI 10.1007/s11004-010-9284-7.

Hengl, T., Heuvelink, G.B.M., Stein, A., (2004). "A generic framework for spatial prediction of soil variables based on regression kriging", Geoderma 120, 75-93.

Hengl, T., Heuvelink, G. B. M., Rossiter, D. G., (2007). 'About regression-kriging: From equations to case studies', Computers & Geosciences 33, 1301–1315.

Hengl, T., (2009). 'A Practical Guide to Geostatistical Mapping', ISBN: 978-92-79-06904-8.

Hengl, T., Heuvelink, G.B.M., Tadić, M.P., Pebesma, E.J., (2012). "Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images", Theor Appl Climatol, DOI 10.1007/s00704-011-0464-2

Heuvelink, G.B.M. (2006), "Incorporating process knowledge in spatial interpolation of environmental variables", In: Proceedings of Accuracy 2006 (Eds. M. Caetano and M. Painho), Lisbon: Instituto Geográfico Portugués, pp. 32-47.

Heuvelink, G.B.M., Griffith, D.A., (2010). "Space–Time Geostatistics for Geography: A Case Study of Radiation Monitoring Across Parts of Germany", Geographical Analysis, ISSN 0016-7363.

Hiemstra, P.H., Pebesma, E.J., Heuvelink, G.B.M., Twenhöfel, C.J.W., (2010). "Using rainfall radar data to improve interpolated maps of dose rate in the Netherlands", Science of the Total Environment 409, 123-133.

Hirsch, R.M., Alexander, R.B., Smith, R.A., (1991). "Selection of methods for the detection and estimation of trends in water quality". Water Resources Res., 27, 803-814.

Hofierka, J., Parajka, J., Mitasova, H., Mitas, L., (2002). 'Multivariate interpolation of precipitation using regularized spline with tension', Transactions in GIS 6 (2), 135–150.

Huerta, G., Sansó, B., Stroud, J.R., (2004). "A Spatiotemporal Model for Mexico City Ozone Levels", Journal of the Royal Statistical Society, Series C (Applied Statistics), Vol. 53, No. 2 (2004), pp. 231-248.

Jiang, T, Su, B., Hartmann, H., (2007). "Temporal and spatial trends of precipitation and river flow in the Yangtze River Basin, 1961-2000", Geomorphology 85, 143-154.

Jost, G., Heuvelink, G.B.M., A. Papritz, (2005). "Analysing the space–time distribution of soil water storage of a forest ecosystem using spatio-temporal kriging", Geoderma 128 (2005) 258– 273.

Journel, A.G., Rossi, M.E., 1989. 'When do we need a trend model in kriging', Mathematical Geology, Vol. 21, No.7, pp 715-739.

Karl, J.W., (2010). 'Spatial Predictions of Cover Attributes of Rangeland Ecosystems Using Regression Kriging and Remote Sensing', Rangeland Ecol Manage 63, 335–349.

Knotters, M., Brus, D.J., Voshaar, J.H.O., (1995). 'A comparison of kriging, co-kriging, and kriging combined with regression for spatial interpolation of horizon depth with censored observations, Geoderma 67, 227–246.

Knotters, M., Heuvelink, G.B.M., Hoogland, T., Walvoort, D.J.J., (2010). "A disposition of interpolation techniques", Wageningen Statutory Research Tasks Unit for Nature and the Environment, WOT-werkdocument 190.

Kyriakidis, P.C., Journel, A.G., (1999). "Geostatistical Space-time Models: A Review", Mathematical Geology, Vol: 31, No: 6, 651-684.

Kyriakidis, P.C., Kim, J., Miller, N.L., (2001). 'Geostatistical mapping of precipitation from rain gauge data using atmospheric and terrain characteristics', Journal of Applied Meteorology 40, 855– 1877.

Liu, C., Koike, K., (2007). "Extending Multivariate Space-time Geostatistics for Environmental Data analysis", Math Geol 39, 289-305.

Lloyd, C.D, (2005). "Assessing the Effect of Integrating Elevation Data into the Estimation of Monthly Precipitation in Great Britain", Journal of Hydrology 308: p.128–150.

Love, P., (1999). "Future trends: water resources: meeting future demand", The Journal of Futures Studies, Strategic Thinking and Policy, vol.01, no.03, p.275-278.

Martínez-Cob, A., (1996). 'Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain', Journal of Hydrology 174, 9–35.

McKillup, S., Dyar, M.D., (2010). "Geostatistics Explained, An Introductory Guide for Earth Scientists", Cambridge University Press, ISBN:978-0-521-74656-4.

Novotny, E.V., Stefan, H.G., (2007). "Stream flow in Minnesota: Indicator of Climate Change", Journal of Hydrology, 334, 319-333.

Onoz, B., Bayazit, M., (2003). "The power of statistical tests for trend detection". Turkish Journal of Engineering and Environmental Sciences 27 (4), 247-251.

Phillips, D.L., Lee, E.H., Herstrom, A.A., Hogsett, W.E., Tingey, D.T., (1997). 'Use of auxiliary data for spatial interpolation of ozone exposure in southeastern forests', Environmetrics 8, 43–61.

Propastin, P., Muratova, N., Kappas, M., (2006). "Reducing uncertainty in analysis of relationship between vegetation patterns and precipitation", 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences.

Rigol-Sanchez. J. P., Chica-Olmo, M., Abarca-Hernandez, F., (2003). 'Artificial neural networks as a tool for mineral potential mapping with GIS', International Journal of Remote Sensing, volume 24, number 5, 1151-1156.

Sheather, S.J., (2009). 'A Modern Approach to Regression with R, Chapter 5: Multiple Linear Regression', Springer Science - Business Media, 125-149.

Schuurmans, J. M., Bierkens, M. F. P., Pebesma, E. J., (2007). 'Automatic Prediction of High-Resolution Daily Rainfall Fields for Multiple Extents: The Potential of Operational Radar', Journal Of Hydrometeorology 8, 1204-1224.

Snepvangers, J.J.J.C., Heuvelink, G.B.M., Huisman, J.A., (2003). "Soil water content interpolation using spatio-temporal kriging with external drift", Geoderma 112, 253– 271.

Spadavecchia, L., Williams, M., (2009). 'Can spatio-temporal geostatistical methods improve high resolution regionalization of meteorological variables?', Agricultural and Forest Meteorology 149, 1105–1117.

Symeonakis, E., Bonifacio, R., Drake, N., (2009). 'A comparison of rainfall estimation techniques for sub-Saharan Africa', International Journal of Applied Earth Observation and Geoinformation 11, 15–26.

Tobler, W.R., A computer movie simulating urban growth in the Detroit region, Economic Geography 46: p.234-240 (1970).

Türkeş, M.,(1999). "Vulnerability o Turkey to Desertification With Respect to Precipitation and Aridity Conditions", Tr. J. of Engineering and Environmental Science 23, 363 – 380.

Yilmaz, A.G., Imteaz, M.A., Jenkins, G., (2011). "Catchment flow estimation using Artificial Neural Networks in the mountainous Euphrates Basin", Journal of Hydrology 410, 134-140.

Yucel, A., Topaloğlu, F., Tülücü, K., (1999). "Adana İlinin Standart Sürelerdeki Yağış Şiddetlerinin İstatistiksel Olarak Kullanılabilirliklerinin İncelenmesi", Tr. J. of Agriculture and Forestry, 23, Ek Sayı 1, 179-185.

Yue, S., Pilon, P., Phinney, B., Cavadias, G., (2002). "The influence of autocorrelation on the ability to detect trend in hydrological series". Hydrological Processes 16, 1807-1829.

Wang, J., He, T., Lv, C.Y., Chen, Y.Q., Jian, W., (2010). 'Mapping soil organic matter based on land degradation spectral response units using Hyperion images', International Journal of Applied Earth Observation and Geoinformation 12, S171–S180.

Webster, R., Oliver, M., (2007). "Geostatistics for Environmental Scientistis". John Wiley & Sons, Ltd, ISBN-13:978-0-470-02858-2.

URL 1: http://www.haritadunyasi.com/turkiye-haritalari/535-turkiyenin-buyuk-akarsu-havzalari-haritasi.html (last accessed date: 28.01.2013).

CURRICULUM VITAE

PERSONAL DETAILS

Name, Surname	: Pınar ASLANTAŞ BOSTAN
E-Mail	: pabostan@gmail.com
Title	: Research Assistant
Nationality	: Turk
Date of Birth	: 26 September 1979
Marital Status	: Married, one children
Address	: ODTU Insaat Muh. Bol. Su Kaynaklari Lab. K4-123, 06800, Dumllupınar Bulvarı, No:1 Cankaya- Ankara- TURKEY
Telephone	: +90- 312-3220721 (Home)
	: +90- 312-2107455 (Work)

EDUCATION

Ph.D. 2004-2013 Middle East Technical University, Institute of Natural and Applied Sciences, Geodetic and Geographic Information Technologies Department, Ankara, Turkey

B.Sc. 1997-2001 Landscape Architecture Department, Faculty of Agriculture, Ankara University, Ankara, Turkey

PROFESSIONAL EXPERIENCE

METU-
Middle East Tech. Unv. 2004 – 2013 Research Assistant in Geodetic and Geographic Information Technologies (GGIT), Middle East Technical University, Ankara Turkey

Wageningen University May 2010- November 2010 Working at the Wageningen University and Research Center as a visiting research assistant under supervision of Dr. Gerard Heuvelink

LANGUAGES

Turkish (Native language), good command of English

PUBLICATIONS

INTERNATIONAL JOURNAL PAPER:

1. **Bostan, P.A**., Heuvelink, G.B.M., Akyürek, S. Z., 2012. "Comparison of Regression and Kriging Techniques for Mapping the Average Annual Precipitation of Turkey", International Journal of Applied Earth Observation and Geoinformation 19, 115–126.

NATIONAL CONFERENCES

2. **Bostan, P.A.,** Akyürek, Z., Trend Analysis of Turkey Annual and Seasonal Mean Precipitation and Temperature Values, V. Ulusal Hidroloji Kongresi, Orta Doğu Teknik Üniversitesi, Ankara, 5 – 7 Eylül 2007. (Oral Presentation) (Full Paper).

3. **Bostan, P.A.,** Akyürek, Z., İkincil Veriler Kullanılarak Türkiye Ortalama Yıllık Yağış Değerlerinin Mekãnsal Dağılımının Modellenmesi, Ulusal Coğrafi Bilgi Sistemleri Kongresi, 30 Ekim- 2 Kasım 2007, Trabzon. (Oral Presentation) (Full Paper)

INTERNATIONAL CONFERENCES

4. **Bostan, P.A.,** Akyürek, Z., Exploring the Mean Annual Precipitation and Temperature Values over Turkey by Using Environmental Variables, ISPRS Workshop; "Visualization & Exploration of Geospatial Data", Stuttgart, 27-29 June 2007. (Poster Presentation) (Full Paper)

5. **Bostan, P.A.,** Heuvelink, G.B.M., Akyürek, Z., Comparison of Regression and Kriging Techniques for Mapping the Average Annual Precipitation of Turkey, Spatial Statistics Conference, 23-25 March 2011, The Netherlands. (Poster Presentation) (Full Paper)

6. **Bostan, P. A.,** Akyürek, Z., Heuvelink, G. B. M, 2012, "Space-time universal kriging of precipitation in the Euphrates Basin, Turkey", Proceeding of the 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Florianopolis-SC, Brazil, July 10-13, 2012 (oral presentation) (full paper).

ACADEMIC AWARDS

1. Poster secondary prize awarded by Department of Statistics, METU. (Graduate Summer School on New Advances in Statistics, 11-24 August 2007, METU, Ankara, Turkey)

STATEMENT OF RESEARCH INTERESTS

1. Geographic Information Technologies
2. Spatio-temporal Data Analysis
3. Precipitation Modeling