MODELLING PRECIPITATION DATA OF CERTAIN REGIONS FOR TURKEY VIA HIDDEN MARKOV MODELS

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

BY

NEVİN YAMAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

DECEMBER 2014

Approval of the Thesis

MODELLING PRECIPITATION DATA OF CERTAIN REGIONS FOR TURKEY VIA HIDDEN MARKOV MODELS

submitted by **NEVİN YAMAN** in partial fulfillment of the requirements for the degree of **Master of Science in Statistics Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural
Dean, Graduate School of **Natural and Applied Sciences** _______________

Prof. Dr. İnci Batmaz
Supervisor, Head of Department, **Statistics** _______________

Assist. Prof. Dr. Ceylan Talu Yozgatlıgil
Co-advisor, **Department of Statistics, METU** _______________

Examining Committee Members:

Assoc. Prof. Dr. Özlem İlk Dağ
Department of Statistics, METU _______________

Prof. Dr. İnci Batmaz
Supervisor, Head of Department, Statistics _______________

Assist. Prof. Dr. Ceylan Talu Yozgatlıgil
Department of Statistics, METU _______________

Assoc. Prof. Dr. Yeliz Yolcu Okur
Institute of Applied Mathematics, METU _______________

Assist. Prof. Dr. Berna Burçak Başbuğ Erkan
Department of Statistics, METU _______________

Date: **18.12.2014**

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 : Nevin YAMAN

Signature :

ABSTRACT

MODELLING PRECIPITATION DATA OF CERTAIN REGIONS FOR TURKEY VIA HIDDEN MARKOV MODELS

Yaman, Nevin
M.S, Department of Statistics
Supervisor: Prof. Dr. İnci Batmaz
Co-Supervisor: Assist.Prof.Dr. Ceylan Yozgatlıgil

December 2014, 164 pages

Estimation methods on climate changes have become increasingly popular in the world over the recent years. They are useful for making comments about the future by using the past data related to temperature and precipitation. Especially, precipitation models, which are usefull for forecasting and simulation purposes, play a crucial role in forecasting climate changes. Estimations of daily rainfall amounts and occurrences found by using precipitation models are commonly used to generate scenarios of runoff, drought, flood, and so on.

The main purpose of this study is to estimate the daily occurrence of rainfall and the daily amount of rainfall. For this purpose, daily amount of rainfall data from nine stations located at East Black Sea Region, one of the wettest regions of Turkey; located at Central Anatolian Region, one of driest regions of Turkey and

Aegean Region, having a normal moisture climate in Turkey are modelled separately by using Hidden Markov Models (HMMs). HMMs are based on Markov Chains (MCs). The most suitable models are decided by comparing Akaike ınformation criterion (AIC), Bayesian ınformation criterion (BIC), mean square error (MSE) and misclassification (Error) rate (MR). It is observed that HMMs give good results for regions that has normal moisture climate compared with the wettest and driest region to estimate the daily precipitation occurrence. On the other hand, they give good results for the wettest region compared with the driest region or with normal moisture climate region to estimate the daily precipitation amount. Also, they successfully predict the most probable states that represents the daily precipitation occurrence by using Viterbi algorithm, when a sequence of observations and the model parameters are known.

In this context, by using HMMs which is thought to be more effective than other precipitation models, the precipitation occurrence and precipitation amount are estimated in this thesis study. This work is the first phase to make estimations related to precipitation, providing very fast and less costly computations, and it gives general weather forecast and information about the unknown state of precipitation occurrences.

Keywords: Hidden Markov Model, Daily Rainfall Occurrence, Daily Rainfall Amount, Viterbi Algorithm

ÖZ

SAKLI MARKOV MODELİ (SMM) İLE TÜRKİYE'NİN BELLİ BÖLGELERİNE İLİŞKİN YAĞIŞ VERİSİNİN MODELLENMESİ

Yaman, Nevin
Yüksek Lisans, İstatistik Bölümü
Tez Yöneticisi: Prof. Dr. İnci Batmaz
Ortak Tez Yöneticisi: Yrd. Doç. Dr. Ceylan Yozgatlıgil

Aralık 2014, 164 sayfa

Son yıllarda dünyadaki iklim değişikleri üzerine yapılan geleceğe dönük tahmin yöntemleri oldukça yaygınlaşmıştır. Bu tahmin yöntemleri geçmiş sıcaklık ve yağış verilerini kullarak geleceğe yönelik yorum yapmak için oldukça kullanışlıdır. Özellikle yağış modelleri iklim değişiklikleri konusunda geleceğe dönük tahminler yapabilmek için önemli bir yere sahiptir. Bu modellerin sağlamış olduğu günlük yağış miktarı ve yağış olup olmama durumu gibi tahminler; taşkın, kuraklık, sel, vb. senaryoları oluşturmak için yaygın olarak kullanılmaktadır.

Bu çalışmanın amacı günlük yağış olup olmama durumu ve günlük yağış miktarını tahmin etmektir. Bu amaçla, Türkiye' nin en çok yağış alan bölgelerinden biri olan Doğu Karadeniz Bölgesi, Türkiye'nin en az yağış alan bölgelerinden biri olan İç

Anadolu Bölgesi ve normal yağışlı bir iklime sahip olan Ege Bölgesinde kaydedilen 9 istasyonun günlük yağış miktarları verisi Markov zincirlerini temel alan Saklı Markov Modelleri (SMM) ile ayrı ayrı modellenmiştir. En uygun modellere Akaike bilgi kriteri, Bayes bilgi kriteri, ortalama karesel hata ve yanlış sınıflandırma oranı kullanılarak karar verilmiştir. Seçilen en uygun SMM'lerin normal iklime sahip bölgelerde günlük yağış olup olmama durumunu tahmin etmede, kurak ve yağışlı bölgelere kıyasla daha iyi sonuçlar verdiği görülmüştür. Yağış bölgelerde ise SMM'ler günlük yağış miktarını tahmin etmede, diğer iki bölgeye kıyasla daha iyi sonuçlar vermiştir. Ayrıca, model parametreleri ve gözlem dizisi bilindiğinde, Viterbi algoritması kullanılarak günlük yağış varlığını temsil eden en olası durumlar başarılı şekilde tahmin edilmiştir.

Bu kapsamda, bu tez çalışmasında etkinliği diğer yağış modellerine göre daha fazla olacağı düşünülen, SMM'ler geliştirilerek yağış varlığı ve yağış miktarı tahminleri konusunda çalışılmıştır. Bu çalışma yağış konusunda hızlı ve kolay hesaplanan tahminler yapabilmek için bir aşama oluşturmuş ve bilinmeyen yağış durumu hakkında genel tahminler vermiştir.

Anahtar Kelimeler: Saklı Markov Modelleri, Günlük Yağış Varlığı Tahmini, Günlük Yağış Miktarı Tahmini, Viterbi Algoritması

To My Husband Hüseyin Yaman and My Commander Sezgin Şakrucu

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Prof. Dr. İnci Batmaz for her guidance throughout the graduate study.

I wish to express my deepest gratitude to my co-supervisor Assist.Prof. Dr. Ceylan Yozgatlıgil for her insight, suggestions, advice, patience and encouragements throughout the research.

I owe my special thanks to Prof. Dr. Walter Zucchini for explaining the main idea behind the theory of my thesis topic, listening my problems, guiding me patiently and providing support for maturing my ideas and opinions.

I am thankful to my examining committee members, Assoc. Prof. Dr. Özlem İlk Dağ, Assoc. Prof. Dr. Yeliz Yolcu Okur and Assist. Prof. Dr. Berna Burçak Başbuğ Erkan for their suggestions, comments and relevant discussions. I would also like to thank Prof. Dr. Murat Türkeş who kindly reply to my questions.

Sezgin Şakrucu and Volkan Acar who are commanders of my work place supported throughout the research. I express my special appreciation and I will remember them as the ones who enable me to conclude the study.

I would like to thank my friends Burcu Yılmaz Önder, Kezik Alkan Kotan, Derya Tatlılıoğlu and Gül İnan for their full support, motivation and nice friendship.

I would also like to give special thanks to my mother and my sister for their full support and motivation.

My deepest thank is to my husband Hüseyin Yaman for his charitably support, encouragements, patience and love.

TABLE OF CONTENTS

APPENDICES

LIST OF FIGURES

FIGURES

LIST OF TABLES

TABLES

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
AR	Autoregressive Model
ARIMA	Autoregressive Integrated Moving Average Model
ARMA	Autoregressive Moving Average Model
BIC	Bayesian Information Criterion
EM	Expectation-Maximization
GAP	Southeast Anatolia Project Area
GCM	Global Climate Model
GP	Generalized Pareto
HMM	Hidden Markov Model
MA	Moving Average Model
MC	Markov Chain
ML	Maximum Likelihood
MR	Misclassification Error Rate
MSE	Mean Square Error
NHMM	Nonhomogeneous Hidden Markov Model
PNI	Percent of Normal Index
SMM	Saklı Markov Modeli
SPI	Standardized Precipitation Index

CHAPTER 1

INTRODUCTION

The importance of water is increasing rapidly every day with the rapid increase in population of the world. Like many other regions of the world, also in some regions of Turkey, irregular precipitation occurs because of reasons like increasing world population, increasing civilization, climate change due to global warming, desertification, destruction of forests and etc. While this irregularity sometimes results in excess precipitation which also causes natural disasters like flood and mudslide, sometimes it may result in long lasting drought periods. For this reason precipitation models has an important role about understanding the probabilistic structure of rainfall and give precipitation simulation. These simulations are used for modelling data sets related to climate, hydrological and environmental system to take some precaution for disasters such as runoff, droughts and floods.

There are different precipitation models. However, not all of them use synoptic atmospheric information such as temperature, solar radiation, and other climatic factors (Bellone et al., 2000). Also, such precipitation models that do not include synoptic atmospheric information can only produce simulations under the current climate systems. They are not suitable for predictions of global climate to local precipitation patterns. On the other hand, weather-state models such as Global Climate Models (GCMs) use synoptic atmospheric information to categorize each day into a weather-state and then precipitation is modeled by using multivariate distributions (He and Kundu, 1991; Bardossy and Plate, 1992; Hughes et al., 1993;

Bardossy, 1994; Bartholy et al., 1995). These models help to understand regional and local effects of global climate change (Bellone et al., 2000). GCMs have been well understood and modeled; however, GCMs cannot get small-scale atmospheric patterns and specify the impact of changes in the atmosphere due to their grid scale. In other words, GCMs are not suitable for deriving small local and regional rainfall (Giorgi and Mearns, 1991; Hughes and Guttorp, 1994).

In addition to GCMs, moving averaging models (MA), autoregressive models (AR), combined autoregressive moving average models (ARMA) and autoregressive–integrated-moving average models (ARIMA) could be used as precipitation models. However, precipitation time series data which have many zero values due to persistence of dry periods and storms with short durations in arid and semi-arid areas prevents from using traditional time series approaches that is MA, AR, ARMA and ARIMA. Also, they can only be used when time series are Gaussian. Time series data which are not Gaussian need to be transformed (Yevjevich, 1991).

Other precipitation models are Markov Chain models. This theory is explored and enhanced by Gabriel and Neumann (1992). Markov Chain that has homogeneous transition matrix is used for modeling daily wet and dry occurrences at a one rain station which take place in Israel (Gabriel and Neumann, 1992). This model is extended to show the seasonal differences by managing time-varying parameters (Stern and Coe, 1984; Woolhiser, 1992). Markov model has been used to simulate the Kenyan longest dry and wet spells and largest rain-sums (Sharma, 1996). Also, monthly rainfall records in arid zones in Saudi Arabia have been modelled using by Markov Chain (Elfeki and Al-Amiri, 2011).

Nonparametric models can also be used to model precipitation (Young, 1994; Lall et al., 1996; Moron et al., 2008). They do not require assumption for dependence on parametric distributions and describe nonlinear relationships between variables. However, nonparametric models can only generate values which have already been found and this prevents incorporate effects of long term climate changes into the precipitation process.

Another stochastic precipitation model is that HMMs which are firstly used by Zucchini and Guttorp (1991) in the estimation of rainfall occurrence. Basic HMMs which are stationary in time are developed by incorporating time-varying covariates such as seasonal forecasts, temperature, etc. (Hughes and Guttorp, 1994a; Hughes and Guttorp, 1994b; Hughes et al., 1999; Bellone et al., 2000; Robertson et al., 2004; Robertson, 2005; Robertson et al., 2007). In other words, nonhomogenous HMMs (NHMMs) are derived since global climate change causes nonstationary patterns in precipitation. Today, many researchers tries to find the effect of the climate change in precipitation models with improving the future predictions. Therefore, HMMs are one of the best models for precipitation models since they have powerful and quick algorithms to solve problems about nonstationary patterns in precipitation.

When the literature is investigated for Turkey, it can be seen that HMMs are not used for precipitation modelling. The ARIMA models which are mixture of AR and MA models are widely used. These models are used for estimating changes about temperature and precipitation in the Southeast Anatolia Project Area (GAP) (Bahadır, 2011). Also, they are used for producing synthetic series to estimate the rainfall potential in Gediz Basin (Topçuoğlu, 2005) and modelling precipitation data in Manisa (Topçuoğlu, 2010). In addition, non-stationary temporal climate series consisting of temperature, vaporization and precipitation series for Denizli which is located in Aegean Region of Turkey is analyzed with ARIMA models (Özdemir and Bahadır, 2010). Drought analysis is done in Eastern Mediterranean Region (Fidan, 2011), drought characteristic of Regions are determined in Central

Anatolian Region (Yeğnidemir, 2005) and the state of drought is studied in Trakya Region (Çaldağ et al., 2004) by using Standardized Precipitation Index (SPI). Drought occurrence probability is found through by using MCs (Fidan, 2011), total daily rainfall amount data from Göztepe station is analyzed by second order MC to define the distribution of daily rainfall amount (Koçak and Şen, 1997). In addition, MC is used to estimate the yearly precipitation probabilities (Özgürel and Kılıç, 2003) and it is used to find drought occurrence probability in GAP Region (Tonkaz, 2008). Projected changes in future air temperature and precipitation climatology of Turkey are analyzed with RegCM4.3.5 Climate simulations for the period of 2070-2100 (Öztürk et al., 2002). Also, nonparametric tests (such as Mann-Kendall), linear regression and coefficient of variation techniques are used to describe rainfall trends for Kahramanmaraş which is located in the Southeast Anatolia Region of Turkey (Karabulut and Cosun, 2009).

It is known that climate change studies concentrate on temperature and precipitation (Türkeş, 1996; Türkeş et al., 2002; Türkeş et al., 2008; Özdemir and Bahadır, 2010). Turkey is one of the countries that has occurrence risk of short-time and long-time climate changes (Türkeş et al., 2002). In addition, precipitation is the most unstable parameter within the climate variables in terms of time and space.Various models are studied for Turkey's precipitation data (Aykan et al., 2012; Aksoy et al., 2013). When studies related to precipitation changes in Turkey is investigated, precipitation per year tends to decline and the number of dry periods tends to increase after 1970 (Türkeş, 1996). Therefore, the problems that the world and Turkey face due to the irregular precipitation makes it very critical to make precipitation prediction and taking necessary precautions. With this scope, by using the previously recorded precipitation data from some Regions of Turkey, in order to prevent the financial damage that can occur because of flood and mudslide by the excess precipitation or to prevent the adverse effects of long lasting drought, daily statistical precipitation prediction will be done by using

Hidden Markov Model (HMM). Since HMM has a more general and flexible structure and more useful algorithms than other precipitation models, the daily precipitation amounts of certain Regions of Turkey are modelled by HMMs. By the model, daily precipitation occurrence and daily precipitation amount are predicted and precipitation scenarios are constructed, whose results can help people that are in charge at the government and the farmers, so that they will have the chance to take the necessary precautions against the disasters.

In this study, the main purpose is estimating the probability of rainfall occurrence with amount of rainfall for some specific Regions of Turkey and comparing their results. Firstly, the probability of rainfall occurrence is estimated for three stations from Eastern Black Sea Region which is one of the wettest Regions of Turkey and for three stations from Central Anatolian Region which is one of the driest Regions of Turkey and for three stations from Aegean Region which is a normal moisture climate Region of Turkey separately and results are compared. Secondly, the amount of rainfall is estimated for one station from each three Regions separately and results are compared.

The thesis study is organized as follows. In chapter 2, the history of HMM and precipitation models which are developed by HMMs are explained. In chapter 3, firstly the brief information related MC is given. Secondly, the definition of HMMs is clarified. In addition, the description of the daily precipitation HMM and the parameter estimation method are explained. Thirdly, model selection criterion is described. Lastly, the estimation of hidden states is explained. In chapter 4, the results of the simulation cases defined in the previous chapter are presented and discussed. The graphs are prepared from the results obtained by the simulations according to the objectives of the cases. According to the models defined in the previous chapter, different types of graphs are presented in order to observe future prediction related to rainfall. Finally, the main conclusions reached throughout the study are stated and the work for future investigations is summarized.

CHAPTER 2

LITERATURE REVIEW

2.1. Introduction

HMMs have been mainly used for many areas for three decades especially for signal-processing implementation and speech recognition. Also, they have been extended to other fields such as all kinds of recognition problems (face, gesture, handwriting, and signature), bioinformatics (biological sequence analysis), environment (wind direction, rainfall, and earthquakes), finance (series of daily returns), and biophysics (ion channel modelling).

HMMs are simple, versatile and their results are mathematically observable since likelihood function can be computed in an uncomplex manner. Generally HMMs could be used as general-purpose models for time series (Zucchini and MacDonald, 2009). Basic HMM is univariate and roots from a homogeneous MC which do not have trend and seasonal variation, observations can be discrete or continuous and without including attainable covariates information. HMMs have many possible extensions of the basic HMMs. They can be multivariate and can be used for analyzing time series data that has trend and seasonal variation, and they can use covariate information.

2.2. History of Hidden Markov Models

The theory of HMMs were introduced in 1966 (Baum and Petrie, 1966) and they are referred as probabilistic functions of MCs. They studied statistical properties of HMMs and developed ergodic theorem for almost-sure convergence (Baum and Petrie, 1966). In 1969, some assumptions related to HMMs were relaxed (Petrie, 1969). Forward-backward algorithm developed for calculating the conditional probability of a state gives an observation sequence from an HMM and this algorithm was used for computing maximum likelihood (ML) estimation of HMM parameters efficiently (Baum et al., 1967; Baum et al., 1970; Baum, 1972). This parameter estimation procedure is defined as expectation-maximization (EM) algorithm and it was applied in HMMs (Dempster et al., 1977). Also, Baum-Welch algorithm which is referred as local convergence was studied (Baum et al., 1970; Baum, 1972) and forward-backward algorithms were developed. (Chang and Hancock, 1966). In addition to these studies, there are many studies which contribute to improve HMMs in recent years (Finesso, 1990; Merhav, 1991; Robert et al., 1993; Elliott et al., 1994; Ryden, 1995; Macdonald and Zucchini, 1997; Lapidoth and Ziv, 1998; Charles et al., 1999b; Scott, 2002; Charles et al., 2004; Ailliot, 2009).

HMMs are defined as deterministic function of MC with augmented state space (Baum and Petrie, 1966; Petrie, 1969; Finesso, 1990). There is a relation between HMMs and the mixture processes since each observation generated by an HMM that has a mixture distribution and need not to be statistically independent (Everitt and Hand, 1981; Redner and Walker, 1984; Titterington et al., 1985 MacLachlan and Basfard, 1988; Leroux and Puterman, 1992). Also, HMMs are special cases of switching autoregressive processes whose dynamics at each time instant depend on the state of an MC at that time (Hamilton, 1994). HMMs are used commonly in random process such as engineering, statistics, and econometrics. Automatic character recognition and speech recognition were the earliest applications of

HMMs (Burke and Rosenblatt, 1958; Raviv, 1967; Baum et al., 1970). It has been developed a new recursion for conditional probability of a state which is defined by MC given the observations considering minimum character error rate sense. In the mid-1970s, a phonetic speech recognition system that relies on hidden Markov modeling of speech signals was developed by using Baum algorithm (Baker, 1975; Jelinek et al., 1975; Jelinek, 1976). In the early 1980s and 1990s, speech recognition applications were developed and this area become leading application of HMMs. (Poritz, 1988; Rabiner, 1989; Lee, 1989a; Lee, 1989b; Huang and Jack, 1989; Huang et al., 1990; Lee, 1990; Charniak, 1993). Therefore, HMMs became popular and they began to be used in many applications. HMMs was used for solving problems related to economics, financial mathematics, banking and assurance (Ryden et al., 1998; Hamilton, 1989; Knab, 2000; Wichern, 2001; Knab et al., 2003; İnce et al., 2005). They also was used in biosciences, biology, bioinformatics and genetics (Thompson, 1983; Guttorp et al., 1990; Krogh et al., 1994; Yada et al., 1994; Yada and Hirosawa, 1996; Yada 1998; Durbin et al.,1998; Schliep et al., 2003; Won et al., 2004). Gene expression time course data has been analysed to predict the behavior of gene data by using HMMs (Schliep et al., 2003). Also, HMMs were used for explaining or predicting the decisions of persons in the area of social sciences (Schrodt, 1998) and it was used for simulating data about environment issues to predict the future (Zucchini and Guttorp, 1991; Hughes et al., 1999; Greene at al., 2008; Zucchini and MacDonald, 2009). Brands choices and their reasons were also studied by using HMMs (Can and Öz, 2009). HMMs have been used to model time series of epileptic seizure counts (Albert, 1991; Le et al., 1992). In a similar way, HMMs have been used to determine pattern movement of a fetal lamb (Leroux and Puterman, 1992).

2.3. Modelling Precipitation with HMMs

Precipitation models help us to understand the probabilistic structure of rainfall and give precipitation simulations. These simulations are used for modelling data sets related to climate, hydrological and environmental system to take some precaution for disasters such as runoff, droughts and floods.

It is known that GCMs, ARMA, ARIMA, MC and nonparametric models are used to model precipitation. Another precipitation model is HMM. It is firstly used by Zucchini and Guttorp (1991) in the estimation of rainfall occurrence. MC assumption holds in the climate process and unobserved states. This HMMs are then extended by Hughes and Guttorp (1994a) by describing a non-homogeneous HMM (NHMM) which links the synoptic atmospheric information. Also, Hughes and Guttorp (1994b) use autologistic model for the transition probability of rainfall data given the weather state to extend the NHMM.

Chain-dependent models assuming that precipitation amounts are conditionally independent given the precipitation occurrences which follow a first-order MC are developed by Katz (1977) and Katz and Parlange (1996).

A 15-year sequence south-western Australia winter data which includes 30 rain stations is used by Hughes et al. (1999). Their model produces accurate rainfall statistics and gives important prediction related to rainfall process in the South-Western Australia.

Charles et al. (1999a) introduces a new NHMM as extending NHMM to observe climate change in the South-Western Australia and describes that NHMM can be placed against the criteria of Wilby et al. (1998) for a useful downscaling model. This model simulates the survival curves related to dry (wet) spell lengths, wet day probabilities, daily rainfall amount, and correlation between daily rainfall amounts. Bellone et al. (2000) simulate precipitation data for 24 rain gauge stations in Washington as extending NHMMS for precipitation occurrences which include precipitation amount.

Robertson et al. (2004) explore if NHMM enables to understand frequency of daily occurrence in terms of large-scale atmospheric patterns at the station level, and to get scenarios related to daily rainfall sequence in Northeast Brazil.

Neytchev et al. (2008) find that NHMM helps to give insights specific information related to inter-annual climate variability and detection of climate change for Bulgaria's precipitation data.

It is confirmed that the NHMMs are beneficial appliance to search the connections between large-scale climatic process and local climate variables by Neykov et al. (2008). NHMMs are successfully given statistics of daily precipitation results related to 32 stations of Bulgaria.

Robertson et al. (2009) have been recently used NHMMs and their simulation results related to precipitation occurrences are very accurate. It is observed that such downscaling methods have become significant on climatic researchers. Hence, HMMs are successfully used in prediction of precipitation occurrence in this research.

Neykov et al. (2014) has been compared stochastic daily precipitation models and realized that these models tend to underestimate the occurrence of storms. Hybrid gamma-generalized Pareto (GP) and hybrid Weibull-GP have been used to develope new precipitation models for daily rainfall data. They found that the underestimate problems of extreme weather conditions would be solved, if NHMMs had been developed by using such distributions.

CHAPTER 3

METHODS

3.1. Markov Chain (MC)

In this chapter, firstly the brief information related to MC is given. Secondly, the definition of HMMs, the description of the daily precipitation HMMs and the parameter estimation method are explained. Thirdly, model selection criterion are described.

An MC is defined as a discrete-time process for which the future behavior only depends on the present when it is given the past and the present values. On the other hand, a Markov process is defined as the continuous-time version of an MC (Zucchini and MacDonald, 2009).

An MC analyzes discrete time points defined as *"0, 1, 2…"* A set of states denoted *S* and the transition probabilities denoted p_{ij} which are the probabilities that the MC is at the next time point in state j, given that it is at the present time point at state *i* (Şahin and Şen, 2001).

$\{S_t : t \epsilon \mathbb{N}\}$, which is a sequence of discrete random variables, is defined as an MC, if it satisfies the following Markov property for all $t \epsilon \mathbb{N}$:

$$Pr(S_{t+1}|S_t, \dots, S_1) = \Pr(S_{t+1}|S_t). \quad (1)$$

The process "S_{t+1}" depends only on the most recent value "S_t" (Cinemre, 2003). $S^{(t)}$ is defined as

$$S^{(t)} = (S_1, S_2, \dots, S_t). \quad (2)$$

Then Markov property can be written as:

$$\Pr\left(S_{t+1}\middle|S^{(t)}\right) = \Pr(S_{t+1}|S_t). \quad (3)$$

Figure 1 shows that past and future are dependent only through present.

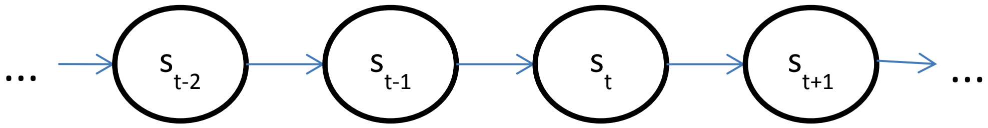

Figure 1 Markov Chain

An MC includes the following conditional probabilities called transition probabilities (Tijms, 2003)

$$\Pr(S_{s+t} = j|S_s = i). \quad (4)$$

The MC is homogeneous when transition probabilities do not depend on "s." It is assumed that MC is homogeneous, if there is no expression related to homogeneity. If MC is homogeneous, the transition probabilities are defined as

$$\gamma_{ij}(t) = \Pr(S_{s+t} = j|S_s = i), \quad (5)$$

where "γ_{ij}" is the probability that the MC is at next time point in state j, given that this chain is at the present time point at state i. The row sum of "γ_{ij}"s are equal to 1.

The transition probabilities $(\gamma_{ij}(t))$ constitute the transition matrix $(\Gamma(t))$ (Chung and Walsh, 2005).

$$\Gamma = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{pmatrix}. \tag{6}$$

All finite state space homogeneous MCs satisfy the following Chapman-Kolmogorov equation (Zucchini and MacDonald, 2009):

$$\Gamma(t+u) = \Gamma(t)\Gamma(u). \tag{7}$$

This equation imply that,

$$\Gamma(t) = \Gamma(1)^t, \tag{8}$$

for all t. In other words, it is said that the matrix of t-step transition probabilities equals to the t^{th} power of $\Gamma(1)$ which is the matrix of one-step transition probabilities.

$\Pr(S_t = j)$ is defined as unconditional probabilities of an MC in a given state at time t. These probabilities are denoted by the following row vector:

$$u(t) = (\Pr(S_t = 1), \ldots, \Pr(S_t = m)), \tag{9}$$

for $t \epsilon \mathbb{N}$. Here, $u(1)$ is defined as the initial distribution of MC, and

$$u(t+1) = u(t)\Gamma. \tag{10}$$

An MC is defined as "stationary" if $\delta\Gamma = \delta$ and $\delta 1' = 1$ hold. First expression refers "stationarity" and second expression refers that δ is indeed a probability distribution. Also, Γ represents transition probability matrix (Gabriel and Neuman, 1992; Zucchini and Guttorp, 2009).

3.2. Hidden Markov Models

3.2.1. The Definition of Hidden Markov Models

An HMM is an extension of Markov models to the case where the observation is a probabilistic function of the state. HMMs are mixture models and they include mixture component which generates observation described by the state of a hidden Markov process instead of a static mixing distribution. It is known that a mixture distribution is the marginal distribution of an HMM. Mixture components can be represented by the known probability distribution and first or higher-order Markov process (Zucchini and MacDonald, 2009). Hidden Markov Process, Latent Markov Models, Markov-Switching Models, Markov Dependent Mixture are other names of the HMM used in the literature (Leroux and Puterman, 1992; Eprahaim and Merhav, 2002). Hidden-Semi Markov Models, State-Space Models and Markov-Switching Models are also related models of HMM in literature.

It is known that independent mixture model is not useful when the serial dependence is observed in the observations. In this situation, MC assumption holds and HMMs could be constructed. HMMs consist of two stochastic processes. The first one is an MC that is characterized by the states and transition probabilities. The states of the chain are externally not visible, therefore it is defined as "Hidden." The second stochastic process, on the other hand, produces observations at each moment, depending on a state-dependent probability distribution. Each

HMM is defined by states, state probabilities, transition probabilities, emission probabilities and initial probabilities.

In order to define an HMM completely, the following five elements have to be defined (Rabiner, 1989):

1. The states of model are shown as S and there are N states of the model. The individual states are denoted as follows:

$$S = \{S_1, \ldots, S_N\}.$$

The state at time *t* is denoted as q_t for *t=1,2,…,N*.

2. There are M distinct observations which are the physical output of the system and they represent discrete output for per state. It is denoted as a set of individual observation symbols as V= $\{v_1, v_2, \ldots, v_m)$.

3. The state transition probability distribution is shown as

$$A = \{a_{ij}\},$$

where a_{ij} is the probability that the state at time $t + 1$ is S_j given that the state at time t is S_i . These transition probabilities constitute the transition probability matrix. The transition probabilities should satisfy the normal stochastic constraints:

$$a_{ij} \geq 0, 1 \leq i, j \leq N \sum_{j=1}^{N} a_{ij} = 1, 1 \leq i \leq N. \tag{11}$$

4. The observation symbol probability distribution in each state is shown as

$$B = \{b_j(k)\},$$

where $b_j(k)$ is the probability that symbol X_k is emitted in state S_j and

$$b_j(k) = p\{O_t = X_k | S_t = j\}, \tag{12}$$

where X_k denotes the k^{th} observation symbol in the alphabet, and O_t is one of the symbols from *V,* for $1 \leq j \leq N, 1 \leq k \leq M$.

5. The HMM has the initial state distribution which is shown as $\pi = \{\pi_i\}$, where π_i is the probability that the model is in state S_i at the same time $t = 0$ with

$$\pi_i = p\{S_1 = i\}, \tag{13}$$

for $1 \leq i \leq N$.

The HMM can be used as a generator to find an observation sequence, such an

$$O = O_1, O_2, \ldots, O_K,$$

where O_t is one of the symbols from *V*, and *K* denotes the number of observations in the sequence when it is given suitable values of *N, M, A, B*, and π as follows,

1) Choose an initial state $q_1 = S_i$ by using π which is initial state distribution.

2) Define t=1.

3) Choose O_t=v_k by using b_i(k) which is the symbol probability distribution in state S_i.

4) Skip next state q_{t+1}=S_j by using the state transition probability distribution for S_i.

5) Skip t=$t+1$ and return step 3 if $t<T$, otherwise end the procedure.

The representation of a basic HMM is shown in Figure 2. Orange circles show "Hidden States". Hidden states are dependent only on the previous state. The past is independent of the future given the present (Markov assumption). Blue circles show "Observations." Observations depend only on their corresponding hidden state.

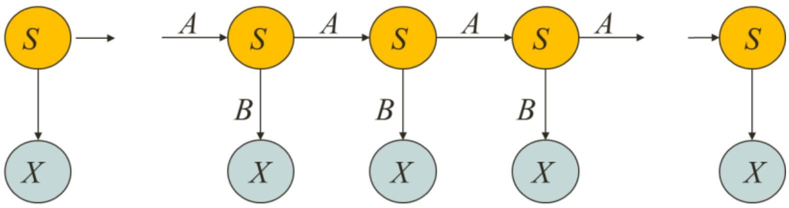

Figure 2 Representation of basic HMM

S: {S_1,…,S_N } represents the values for the hidden states and *X : {$X_{1,}$…,X_M}* represents the values for the observations.

An HMM (X_t: t ∈ N) is defined as a particular dependent mixture. $X^{(t)}$ which denotes the observations and $S^{(t)}$ which denotes the states represent the time histories between *1* and *t*. The simplest model can be summarized as two dependency structures:

1st order of Markov assumption of transition:

$$P(S_t|S_1, S_2, \dots, S_{t-1}) = P(S_t, S_{t-1}). \tag{14}$$

Conditional independency of observation parameters:

$$\mathrm{P}(\mathrm{X}_t|\mathrm{S}_t, \mathrm{X}_1, \dots, \mathrm{X}_{t-1}, \mathrm{S}_1, \dots, \mathrm{S}_{t-1}) = \mathrm{P}(\mathrm{X}_t|\mathrm{S}_t). \tag{15}$$

The model includes two processes. First process is unobserved "parameter process" *{S_t : t=1,2,…}* which satisfies "Markov assumption." Second process is "state-dependent process" *{X_t : t=1,2,….}*. When S_t is known, the distribution of

X_t depends only on the current state S_t, it does not depend on the other states and observations (Zucchini and Macdonald, 2009). These structures represent in Figure 3, $\{X_t\}$ represents the *m*-state HMM if the MC S_t has *m* states.

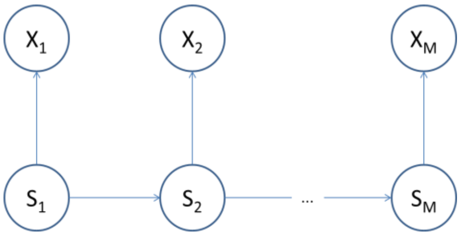

Figure 3 Basic representation of m-state HMM

3.2.2. Description of the Daily Precipitation Model

In this study, the development of daily precipitation models for three Regions which are the wet, dry and normal dry in Turkey has been considered. For this reason, HMMs are used to describe daily rainfall occurrence and daily amount of precipitation for these Regions. The precipitation process is defined as a two-state first order MC and ıt has been discovered as an adequate model in many different Regions of the world (Katz, 1977; Coe and Stern, 1982; Stern and Coe, 1984 and Zucchini et al., 2001a).

Daily total precipitation data from 1964 to 2005 for East Black Sea Region, 1977 to 2006 for Central Anatolian Region and 1972 to 2005 for Aegean Region which are obtained from Turkish State Meteorological Service are used. Because of missing values in the daily series, these time periods are selected. Data include amount of total precipitation for a day. Also, we observe that amount of precipitation is zero for many days in Central Anatolian Region and Aegean Region. Therefore, we define “dry day” and “wet day” before applying daily precipitation occurrence process which is used for estimating daily precipitation occurrence. If the total amount of precipitation is less than 0.1 *mm*, it is defined as "dry day," otherwise, it is defined as "wet day." This criterion value is chosen according to the meteorological expert opinion (Türkeş, 1996). In intensity process which is used for estimating daily precipitation amount, the daily precipitation series is used directly.

We use the following notations to describe our model. Let "$R_t = (R_t^1, \dots, R_t^K)$" is a nonnegative multivariate vector of precipitation amounts at a network on K stations. Observed values on day *t*, where *t=1,2,…, N* at station *i*, where *i =1,2,…,K* is shown as r_t^i . Also, S_t is defined as hidden rainfall state for day *t*. In addition daily sequences of precipitation, total amounts are represented by $R_{1:T}$ and hidden rainfall states are represented by $S_{1:T}$ (Hughes et al., 1999).

There are two conditional independence assumptions to construct an HMM for rainfall (Hughes and Guttorp, 1994a; Zucchini and MacDonald, 2009). The first independence assumption is that R_t (multivariate precipitation observations at time *t*) is independent of all other variables which take place in the model up to time *t*. That is,

$$P(R_t|S_{1:t}, R_{1:t-1}) = P(R_t \mid S_t). \tag{16}$$

The second independence assumption is that the hidden state are the first-order Markov process. That is,

$$P(S_t|S_{1:t-1}) = P(S_t \mid S_{t-1}). \tag{17}$$

We try to find density of R_t to construct an HMM for the rainfall data. In this contex, so-called *occurrence* and *intensity* processes are used to find the density of R_t in an explicit form (Katz, 1977; Stern and Coe, 1984; Neykov et al., 2003).

i. Occurrence Process

In this study, $R_t = (R_t^1, \dots, R_t^K)$ is a nonnegative multivariate vector of precipitation amounts for a network of K=3 stations. Observed values are denoted by "r_t^i" are either "0" or "1" for the occurrence process. Observed values are defined as follows:

r_t^i=1, if the rainfall amount observed on day t at station i is greater than 0.1 *mm,*

r_t^i=0, if the rainfall amount observed on day t at station i is less than 0.1 *mm.*

Therefore R_t becomes a multivariate random vector of rainfall occurrences for a network of K stations and we assume that the distribution of R_t is Bernoulli (Zucchini, 1991; Neykov et al., 2003; Neykov et al., 2007). Also, $R_{1:T}$ denotes the daily sequence of precipitation occurences and $S_{1:T}$ denotes the sequence of hidden states. An HMM for the rainfall data includes two conditional independence assumptions which are already explained.

The first assumption is that:

$$P(R_t|S_{1:t}, R_{1:t-1}) = P(R_t \mid S_t). \tag{18}$$

And, the second assumption is that:

$$P(S_t|S_{1:t-1}) = P(S_t \mid S_{t-1}). \tag{19}$$

In words, it is a first-order Markov process and the Markov process is homogeneous in time. In other words, transition probability matrix which is part of the Markov process does not change within time. If transition probability matrix changes over the time, nonhomogeneous model is extended by using the homogenous model.

In order to model these transition probabilities, we have needed to use "logit link function."

"Logit link function" is defined as follows:

$$\pi(x_t) = l\big(u(x_t)\big) = \exp\big(u(x_t)\big)\,/(1 + \exp\big(u(x_t)\big)). \qquad (20)$$

The function "$u(x_t)$" provides connection and explains the different temporal and seasonal effects. Also, it should be periodic and parametric, and its shape is sinusoidal (Stern and Coe, 1984; Neykov et al., 2003).

The following function is used as a logit link function to capture the seasonal behavior:

$$u(x_t) = \alpha_0 + \alpha_1 sin\left(\frac{2\pi tk}{365}\right) + \alpha_2 \cos\left(\frac{2\pi tk}{365}\right). \qquad (21)$$

This function includes seasonal terms that repeat each year and remainder term that provides information related to deviation from the regular pattern. Also, instead of square root function, logarithm, cubic root or power transformation could also be used (Neykov et al., 2003).

ii. Intensity (Precipitation Amount) Process

R_t is a multivariate vector of precipitation amounts for a network of K stations for an intensity process, and observed values are denoted by X_t. In this process, we define "W_t" as follows:

$$W_t = \begin{cases} R_t, & if\ R_t > 0\ , where\ t = 1,2,\dots,n \\ 0, & o.w \end{cases}. \quad (22)$$

The distribution of rainfall amount on the wet days is assumed as positively skewed. Because, it is known that larger amounts occur less than smaller amounts. Therefore, the distribution could be exponential, lognormal, Weibull or gamma. (Stern and Coe, 1984; Grunwald and Jones, 2000; Zucchini et al., 2001a). Also, it can be seen that seasonal variability exists in this distribution. To model such precipitation data, a single family distribution whose parameters change smoothly over the year is chosen and they are represented as a Fourier series (Neykov et al., 2003).

In order to model our data, gamma distribution has been chosen. Gamma probability density function ɣ(z,μ,β) is given:

$$y(z,\mu,\beta) = \begin{cases} (\beta/\mu)^{\beta} z^{\beta-1} \exp(-\beta z/\mu)\ / \Gamma(\beta) & for \quad z > 0 \\ 0 & for \quad z = 0, \end{cases} \quad (23)$$

where Γ(β) denotes the gamma function, μ>0 denotes the mean and $\beta > 0$ denotes the shape parameter.

In the precipitation amount model, a “log link function” has been used again. This function is:

$$\log(\mu_t(x_t)) = \theta_0 + \theta_1 sin\left(\frac{2\pi tk}{365}\right) + \theta_2 \cos\left(\frac{2\pi tk}{365}\right), \quad (24)$$

where $(\theta_0\ , \theta_1, \theta_2\ )^T$ are the unknown parameters.

3.2.3. The Estimation of Parameters

In general, HMMs have many parameters. The number of parameter equals to "$m(m-1)+\ q$" for a stationary m-state HMM and the number of parameter equals to "m^2-1+q" for a homogeneous and stationary HMM where "m" denotes the number of states, "q" denotes the total number of parameters in the state dependent models. These parameters are estimated by using the distribution of X_t and higher-order marginal distributions of (X_t,X_{t+k}) (Zucchini and MacDonald, 2009).

The cases of discrete and continuous observations are defined as follows, for *i=1,2,…,m :*

$$p_i(x) = \Pr(X_t = x|\ S_t = i\ ), \tag{25}$$

where $p_i(x)$ is defined as the probability mass function for X_t, if the MC is in state "i" at time t for discrete case; $p_i(x)$ is defined as the probability density function of X_t, if the MC is in state "i" at time t for continuous case.

We assume that X_t denotes the discrete-valued observations and $u_i(t) = \Pr(S_t = i)$ denotes that the MC is in state "i" at time "t." The univariate distribution of X_t is:

$$\Pr(X_t = x) = \sum_{i=1}^{m} \Pr(S_t = i) \Pr(X_t = x|S_t = i) = \sum_{i=1}^{m} u_i(t) p_i(t). \tag{26}$$

This expression can be written in matrix notation as follows:

$$\begin{aligned}\Pr(X_t = x) &= \big(u_1(t), \dots, u_m(t)\big) \begin{pmatrix} p_1(x) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & p_m(x) \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \\ &= u(t)P(x)1'.\end{aligned} \tag{27}$$

where $P(x)$ denotes the diagonal matrix and i th diagonal element is $p_i(x)$.

$$u(t) = u(1)\Gamma^{t-1} \tag{28}$$

$$\Pr(X_t = x) = u(1)\Gamma^{t-1}\mathrm{P(x)}1'. \tag{29}$$

MC is merely homogeneous and it is not necessarily stationary, if equation (28) holds, otherwise not. If MC is homogeneous and stationary, i.e., MC includes stationary distribution δ, $\delta\Gamma^{(t-1)} = \delta$ for t=1,2,…,T, the equation (29) become more simpler.

$$\Pr(X_t = x) = \delta P(x)1'\,. \tag{30}$$

The bivariate distribution of X_t and X_{t+k} which are discrete-valued observations can be witten by using $u_i(t) = \Pr(S_t = i)$ which denotes the MC is in state 'i' at time "t" for i=1,2,…,T as follows:

$$\begin{aligned}
&\Pr(X_t = v, X_{t+k} = w) \\
&\qquad = \sum_{i=1}^{m}\sum_{j=1}^{m}\Pr(X_t = v, X_{t+k} = w, S_t = i, S_{t+k} = j) \\
&\qquad = \sum_{i=1}^{m}\sum_{j=1}^{m}\Pr(\delta_t = 1)\, p_i(v)\Pr(\delta_{t+k} = j|\delta_t = i)p_j(w) \\
&\qquad = \sum_{i=1}^{m}\sum_{j=1}^{m}\mathrm{v}_i(t)p_i(v)\gamma_{ij}(k)p_j(w),
\end{aligned} \tag{31}$$

where $\gamma_{ij}(k)$ denotes the i,j-th element of $\Gamma(k)$. This expression can be written in the matrix notation as:

$$\Pr(X_t = v, X_{t+k} = w) = u(t)P(v)\Gamma^k P(w)1'. \tag{32}$$

If an MC is stationary, then:

$$\Pr(X_t = v, X_{t+k} = w) = \delta P(v)\Gamma^k P(w)1'. \tag{33}$$

In a similar way, higher-order marginal distributions can be obtained. For our data set, we have three rain stations and three observations X_t, X_{t+k} and X_{t+k+l}. The marginal distributions of X_t, X_{t+k} and X_{t+k+l} is:

$$\Pr(X_t = v, X_{t+k} = z, X_{t+k+l} = w) = \delta P(v)\Gamma^k P(z)\Gamma^l P(w)1'. \tag{34}$$

To estimate the unknown parameters, the likelihood function has to be calculated. However, calculation of likelihood function requires many calculations which makes it diffucult to find. On the other hand, likelihood function can be computed for the consecutive observations x_1, x_2, ..., x_T, which are generated by HMMs in a simple way (Baum, 1972; Lange and Boehnke, 1983; Zucchini and MacDonald, 2009). Therefore, unknown parameters can be estimated by maximizing the likelihood function. Here, $X_t=\{x_1, x_2, \ldots, x_T\}$ is an observation sequence generated by the HMM. The likelihood function for the observation sequence can be calculated as follows:

$$L_T = \delta P(x_1)\Gamma P(x_2)\Gamma P(x_3) \ldots \Gamma P(x_T)1', \tag{35}$$

where "δ" denotes the initial distribution of the MC, "Γ" denotes the transition probability matrix and "p_i" denotes the state dependent probability function for the *m*-state HMM.

If it is known that δ is the stationary distribution of MC, i.e., $\delta\Gamma = \delta$, then;

$$L_T = \delta\Gamma P(x_1)\Gamma P(x_2)\Gamma P(x_3) \dots \Gamma P(x_T)1'. \qquad (36)$$

To simply eq. (36), a new matrix is defined which is called D_t. It is equal to "$\Gamma P(x_t)$." Then, the equation becomes:

$$L_T = \delta P(x_1)D_2D_3 \dots D_T1' = \delta D_1D_2D_3 \dots D_T1'. \qquad (37)$$

Hence, the likelihood function is shown as:

$$L_T = L\left(\Gamma, \Lambda; X^{(T)}\right) = \delta D_1D_2 \dots D_T1', \qquad (38)$$

where Γ is the transition probability matrix for MC ($\delta = \delta\Gamma$, for stationary chain), Λ is the state dependent distributions, $X^{(T)}$ is an observation sequence generated by the HMM and D_t is equal to $\Gamma P(x_t)$, in other words, it is a function of Γ and Λ.

3.2.4. Application of Two-State Bernoulli HMM

Precipitation models are based on MCs (Dunn, 2004). The simplest precipitation model has two states which are "rain" and "no rain" and first order MC, i.e, the precipitation probability depends only on the previous precipitation probability. This model can be extended by increasing number of states and order of MC.

We constructed the simplest precipitation model for Turkey certain Region data and we decided 2-state and 3-state Bernoulli-HMM can be suitable for Turkish precipitation data set for analyzing the occurrence probability of rainfall. State-dependent distribution parameters are "p_1, p_2 and p_3" observations are "X_1, X_2 and X_3." The likelihood function can be calculated as follows:

$$L_x = \delta\Gamma P(x_1)\Gamma P(x_2)\Gamma P(x_3)1'. \qquad (39)$$

This expression can be rewritten in matrix notation as follows:

- For two state cases:

$$\Gamma P(x) = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} \Pr(x|state1) & 0 \\ 0 & \Pr(x|state2) \end{pmatrix}$$
$$= \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} p_1^x(1-p_1)^{1-x} & 0 \\ 0 & p_2^x(1-p_2)^{1-x} \end{pmatrix}. \tag{40}$$

- For three state cases:

$$\Gamma P(x)$$
$$= \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{pmatrix} \begin{pmatrix} \Pr(x|state1) & 0 & 0 \\ 0 & \Pr(x|state2) & 0 \\ 0 & 0 & \Pr(x|state3) \end{pmatrix} \tag{41}$$
$$= \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{pmatrix} \begin{pmatrix} p_1^x(1-p_1)^{1-x} & 0 & 0 \\ 0 & p_2^x(1-p_2)^{1-x} & 0 \\ 0 & 0 & p_3^x(1-p_3)^{1-x} \end{pmatrix},$$

where $\boldsymbol{f}(x;p) = p^x(1-p)^{1-x}\ for\ x \in \{0,1\}$ is a probability mass function for the Bernoulli distribution. However, there are parameter constraints which lead to problems to estimate the parameters for the Bernoulli-HMM. These constraints are:

1) $0 \le p_i \le 1\ , i = 1,2,\dots,m;$
   The occurrence probability p_i of the state dependent distribution is between 0 and 1, inclusive.

2) $0 \le \gamma_{ij} \le 1\ , i,j = 1,2,\dots,m\ ;$
   The transition probabilities are between 0 and 1, inclusive.

3) $\sum_{j=1}^{m} \gamma_{ij} = 1\ \ , i = 1,2,\dots,m.$
   The rows of the transition probability matrix "Γ" summation must be equal to 1.

In order to overcome these problems, which stem from constraints, that transformations can be done (Zucchini and Macdonald, 2009). It is defined $\eta_i = \log p_i$, for $i = 1, \dots, m$ ($\eta_i \in R$) for the transformation of the parameters"p_i." Firstly, maximized likelihood function by using unconstrained parameters (η_i's) has to be obtained; secondly it has been transformed back to the following constrained parameter estimates. Therefore, $\widehat{p_i}$ which is an estimate of constrained parameter is defined by using η_i which is an estimate of the unconstrained parameter expressed as follows:

$$\widehat{p_i} = \exp \eta_i. \tag{42}$$

In addition, we need to transform transition probability matrix Γ. However, transformation of it involves more work. It is known that "Γ" has *m*^{2} entries but only *m (m-1)* free parameters. There are *m* row sum constraints:

$$\gamma_{i1} + \gamma_{i2} + \cdots + \gamma_{im} = 1 , \tag{43}$$

where $i = 1,2, \dots, m$.

Assume that *m=2* and define a matrix as:

$$T = \begin{pmatrix} - & T_{12} \\ T_{21} & - \end{pmatrix}. \tag{44}$$

It has *m(m-1)*=2 entries and $T_{ij} \in \mathbb{R}$.

Define a $k: \mathbb{R} \rightarrow R^+$ be strictly increasing function, in other words:

$$k(x) = e^x \; or \; k(x) = \begin{cases} e^x , & x \leq 0 \\ x + 1 , & x \geq 0. \end{cases} \tag{45}$$

Define a g_{ij} function by using this increasing function:

$$g_{ij} = \begin{cases} k(\mathrm{T}_{ij}) \ , & for\ i \neq j \\ 1 \ , & for\ i = j. \end{cases} \tag{46}$$

Define $\gamma_{ij} = \frac{g_{ij}}{\sum_{k=1}^{2} g_{ij}}$, for $i, j = 1,2$ and the set transition probability matrix $\Gamma = (\gamma_{ij})$ which satisfies the constraints. "η_i" and "T_{ij}" are defined as working parameters; "p_i" and "γ_{ij}" are defined as natural parameters. The transformation of Γ and p_i are useful for computing the likelihood-maximizing parameters in two following steps:

Step 1: Likelihood function is maximized with respect to working parameters $\mathrm{T} = \{T_{ij}\}\ and\ \eta = (\eta_1, \dots, \eta_m)$.

Step 2: The estimates of working parameters are transformed to the estimates of the natural parameters:

$\hat{\mathrm{T}} \rightarrow \hat{\Gamma}$ (Estimate of transition probability matrix).

$\hat{\eta} \rightarrow \hat{\mathrm{p}}$ (Estimate of state-dependent distribution).

3.3. Model Selection

There are two problems when data are modelled with HMM. The first problem is selecting an appropriate number of states "*m*." The second one is selecting the state–dependent distributions such as Bernoulli, Poisson, and Geometric. For these reasons, we have to use some criteria to compare the performances of models developed (Zucchini and Macdonald, 2009).

In our model, four model selection criteria, which are Akaike information criterion (AIC), Bayesian information criterion (BIC), mean square error (MSE), and misclassification error rate (MR) are calculated and some plots which includes observed and predicted values are drawn. The model which has the minimum AIC, BIC, MSE and misclassification error rate values can be selected as the best model. However, in some cases different model selection criteria may not give the same model as the best. Therefore, the best model is defined by using overall performance of them and the observed versus predicted graphs.

3.3.1. Akaike Information Criterion (AIC)

AIC is calculated as follows (Akaike, 1973):

$$\text{AIC}= -2\log L+2p, \tag{47}$$

where L denotes the log-likelihood value observed from the fitted model, and p denotes the number of parameters. The measure of the fit part is "-2logL" term and it decreases when the number of states increases; the penalty part is "2*p*" term and it increases when the number of parameters increases.

3.3.2. Bayesian Information Criterion (BIC)

BIC is calculated as follows (Schwarz, 1978; Acquah, 2010):

$$\text{BIC}=-2\log L+p\log T, \tag{48}$$

where L denotes the log-likelihood value which is observed from the fitted model and p denotes the number of parameters, T denotes the number of observations. The measure of the fit part is the same as in AIC; the penalty term is different however. BIC often chooses models which have fewer parameters compared to AIC criterion.

3.3.3. Mean Square Error (MSE)

MSE is calculated as follows:

$$MSE = \frac{\sum_{1}^{365}\left(P_{obs} - P_{pre}\right)^2}{365}, \tag{49}$$

where P_{obs} denotes the observed values obtained from the data set, P_{pre} denotes the predicted values derived from two-state HMMs. MSE values are used to show the difference between the observed values and the predicted values. Minimum value of it indicate that predicted values are close to the observed ones.

3.3.4. Misclassification (Error) Rate (MR)

MR for a two-class classification problem is calculated by using the "Confusion Matrix" given in Table 1. It shows the results for a two-class classification problem.

Table 1 Confusion matrix

		Predicted Values	
		0	1
Observed Values	0	a	b
	1	c	d

Then, MR is calculated as follows:

$$MR = \frac{b+c}{a+b+c+d} \ . \tag{50}$$

MR represents the proportion of an observation being assigned to incorrect class.

3.4. Estimation of the Sequences of Hidden States

There are three problems which can be solved by using HMMs' algorithm in order to use HMMs in practical application (Rabiner, 1989). One of them is to find optimal state sequence or optimal path in the HMM that maximizes the observation probability of the given observation sequence. We want to find the state sequence that best describes the observation sequences among all possible state sequences (Yoon, 2009).

The Viterbi algorithm is a dynamical programming algorithm which helps us to compute the most probable path. When a sequence of observations $O= O_1, O_2$,...,

O_T, and the model parameters $\lambda = (A, B, \pi)$ are given, a sequence of optimal states $S = S_1, S_2, \dots, S_T$ can be obtained by using the algorithm of Viterbi (Rabiner, 1989; Can and Öz, 2009; Yoon, 2009).

However, there is a difficulty to describe the definition of optimal state sequence, because there are many possible optimality criteria. One of them is to choose the states q_t that are individually most likely (Rabiner, 1989).
We define the variable "$\gamma_t(i)$" as follows:

$$\gamma_t(i) = P(q_t = S_i \mid 0, \lambda), \tag{51}$$

and it is the probability of being in state S_i at time t, given an observation sequence, and the model parameters. The variable "$\alpha_t(i),$" which is forward variable is defined as follows:

$$\alpha_t(i) = P(O_1, O_2, \dots, O_t, q_t = S_i | \lambda), \tag{52}$$

and it is the probability of the partial observation sequence $O = O_1, \dots, O_t$ and the state S_i at time t, given the model parameters. "$\alpha_t(i)$" can be solved inductively as follows, according to Rabiner (1989):

1) Initialization:

$$\alpha_1(i) = \pi_i b_i(O_1), 1 \leq i \leq N. \tag{53}$$

2) Induction:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(O_{t+1}), 1 \leq t \leq T-1, 1 \leq j \leq N. \tag{54}$$

3) Termination:

$$P(0|\lambda) = \sum_{i=1}^{N} \alpha_t(i). \tag{55}$$

The forward variable "$\beta_t(i)$" is defined as follows:

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T | q_t = S_i, \lambda), \quad (56)$$

and it is the probability of the partial observation sequence from *t+1* to the end, given the state S_i at time *t* and the model parameters. "$\beta_t(i)$" can be solved inductively as follows:

1) Initialization:

$$\beta_T(i) = 1, 1 \leq i \leq N. \quad (57)$$

2) Induction:

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j\,(O_{t+1}) B_{t+1}(j), \quad (58)$$

where $t = T-1, T-2, \dots, 1;\ 1 \leq i \leq N;\ i \leq j \leq N.$

$\gamma_t(i)$ can be written in terms of $\alpha_t(i)$ which explains the partial observation sequence O_1, O_2, ..., O_T; state S_i at *t*, and $\beta_t(i)$ which explains the remainder of the observation sequence O_{t+1}, O_{t+2}, ..., O_T given state S_i at *t* as:

$$\gamma_t(i) = \frac{\boldsymbol{\alpha_t(i)\beta_t(i)}}{\boldsymbol{P(O|\lambda)}} = \frac{\boldsymbol{\alpha_t(i)\beta_t(i)}}{\boldsymbol{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)}}\,. \quad \textbf{(59)}$$

Here, $P(0|\lambda) = \sum_{i=1}^{N} \alpha_t(i)\beta_t(i)$ is a normalization factor that makes $\gamma_t(i)$ a probability measure. Then:

$$\sum_{i=1}^{N} \gamma_t(i) = 1. \quad (60)$$

Hence, we can find the individually most likely state q_t at time t by using $\gamma_t(i)$, as follows:

$$q_t = argmax[\gamma_t(i)], 1 \leq t \leq T, 1 \leq i \leq N. \quad (61)$$

CHAPTER 4

RESULTS

4.1. Introduction

In this chapter, the results of the precipitation models defined in the previous chapter are presented and discussed. The graphs are prepared from the results obtained by precipitation models according to the objectives of the cases. According to the models defined in the previous chapter, different types of graphs are presented in order to observe future prediction related to precipitation. For each model, the results are discussed in terms of some criteria listed in Section 3.3.

In the following sections, firstly brief information about data is given. Secondly, the probability of precipitation occurrence in three stations which take place in the East Black Sea Region, the Central Anatolian Region and the Aegean Region are found and HMMs are compared to choose the best model by using the model selection criteria considered. Thirdly, analyses of the precipitation amount estimation for one station which takes place in three regions are done separately and results are compared. Lastly, Viterbi algorithm is applied and some unknown states tried to be estimated.

4.2. Description of Data

Daily total precipitation series from 1964 to 2005 for the East Black Sea Region, from 1977 to 2006 for the Central Anatolian Region and from 1972 to 2005 for the Aegean Region are obtained from Turkish State Meteorological Service. Totally nine stations are chosen, three from each region. HMMs are applied to wet region, normal moisture Region and dry region for comparing the performance of them. These regions are selected according to yearly total precipitation amounts and drought map obtained from Meteorology Administration of Turkey. Figure 4 shows the meteorological drought map obtained by using Percent of Normal Index (PNI) method.

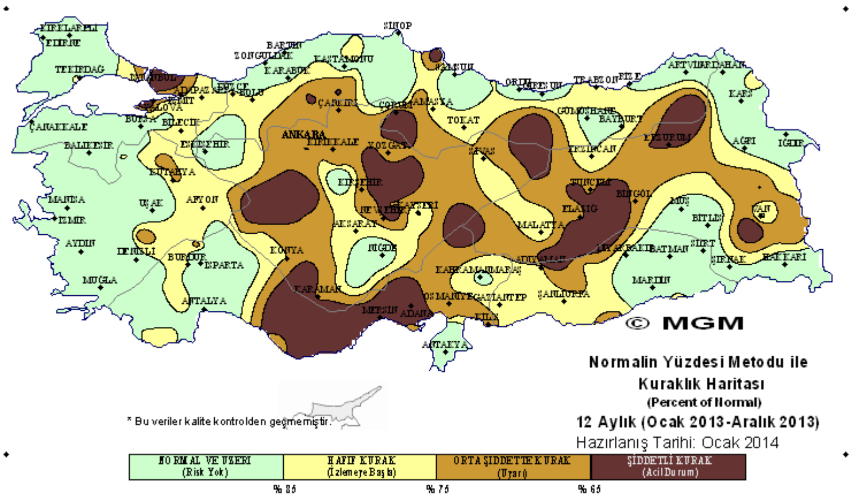

Figure 4 Meteorological drought map by using PNI

PNI method shows that drought cannot be observed in the East Black Sea Region and the Southwest Aegean Region. When two regions are compared, the yearly total precipitation amounts for the East Black Sea Region is higher than the South Aegean Region according to the data obtained from website of Meteorology Administration Turkey.

The stations are chosen according to their locations, completeness of the data and total precipitation amounts. The 29th of Februaries are omitted so that we have the same number of time points in data and then data are classified into two groups namely “wet day” and “dry day.” If the daily precipitation amount is greater than the threshold value that is 0.1 *mm*, then it is labeled as “1,” if it is smaller than the threshold value, then it is labeled “0.” This threshold value is determined as measurable precipitation amount. (Neykov et al., 2014).

4.2.1. Central Anatolian Region (Konya-Karaman)

The map of stations from Central Anatolian Region is shown in Figure 5 and the information belongs to them are given in Table 2.

Figure 5 The selected stations from Central Anatolian Region

Table 2 Information related to the selected stations from Konya and Karaman

Station Number	**Station Name**	**Longitude**	**Latitude**
17244	Konya_Centre	37˚.52'N	32˚.28'E
17246	Karaman	37˚.12'N	33˚.13'E
17900	Çumra	37˚.34'N	32˚.47'E

4.2.2 East Black Sea Region (Rize-Artvin)

The map of stations from East Black Sea Region is given in Figure 6 and the information belongs to them is shown in Table 3.

Figure 6 The selected stations from East Black Sea Region

Table 3 Information related to the selected stations from Rize and Artvin

Station Number	Station Name	Longitude	Latitude
17040	Rize_Centre	41˚.02'N	40˚.30'E
17042	Hopa	41˚.24'N	41˚.25'E
17628	Pazar	41˚.17'N	40˚.91'E

4.2.3 Aegean Region (Aydın-Muğla)

Figure 7 shows the map of stations from Aegean Region and Table 4 gives some information about them.

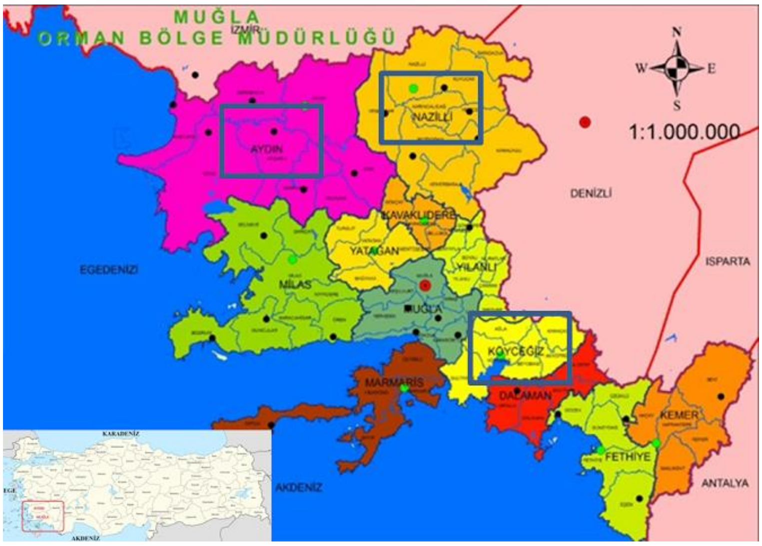

Figure 7 The selected stations from Aegean Region

Table 4 Information related to the selected stations from Aydın and Muğla

Station Number	Station Name	Longitude	Latitude
17860	Nazilli	37˚.54'N	28˚.20'E
17234	Aydın_Center	37˚.50'N	27˚.50'E
17924	Muğla	36˚.57'N	28˚.41'E

4.3 Results of Daily Precipitation Occurrence Analyses

In the following sections, the probability of daily precipitation occurrences in nine stations, which take place in the Aegean Region, Central Anatolian Region and East Black Sea Region, are estimated and evaluated separately.

4.3.1 Analyses of Daily Precipitation Occurrence in Three Stations from Aegean Region

To observe the daily occurrences of precipitation in the Aegean Region, a homogeneous, i.e, the transition probability matrix, which is part of Markov process, does not change with time, two-state HMM and three-state HMM have been developed. The observation probability distribution is chosen as Bernoulli distribution because this is the most suitable distribution for the occurrence analysis of precipitation. This distribution generates observations which are denoted by R_t.

$R_t = (R_t^1 \ldots R_t^K)$, where K denotes the number of stations, represents the multivariate random vector of precipitation occurrence for three stations. Observed values denoted by "r_t^i " are either “0” or “1.”

Three stations are selected from the Aegean Region to compare the wettest and the driest Regions from Turkey. These stations are also chosen according to closeness and correlation between them to make the local analyses successfully.

Data include the daily precipitation amounts between 1972 and 2005 years. Figure 8 shows the daily relative frequency of wet days for three stations from Aegean Region.

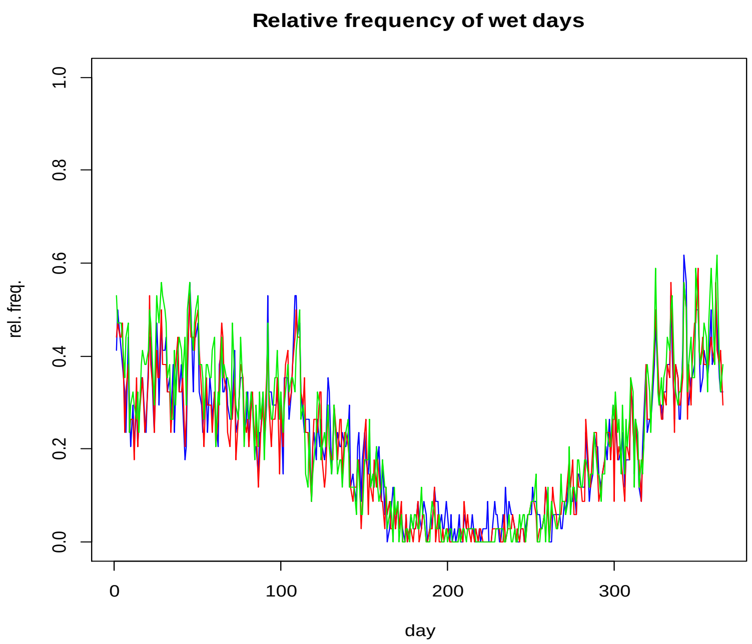

Figure 8 The daily relative frequency of wet days for three stations from Aegean Region

The daily relative frequency of wet days is used to compare stations. When Figure 8 is analyzed, it can be seen that there are three lines. The red line shows the corresponding relative frequency for Nazilli station, the green line shows the corresponding relative frequency for Köyceğiz station and the blue line shows the corresponding relative frequency for Aydın_Center station. When these frequencies are compared, it is observed that they are not the same for three stations. However there are similarities between them. Also, the summary statistics for three stations arecalculated in order to find whether the stations have the same characteristic or not.

Table 5 Summary statistics for stations from Aegean Region

Station Name	**Min**	**Max**	**Median**	**Mean**	**Variance**	**Standard Deviation**
Nazilli	0	88.40	0	1.56	29.83	5.46
Aydın_Center	0	92.0	0	1.69	33.13	5.76
Köyceğiz	0	239.2	0	2.92	105.21	10.26

Table 5 shows the summary statistics for stations from Aegean Region. It is observed that the minimum total amount of precipitation for three stations is the same and the maximum total amount of precipitation for three stations are different, especially the maximum total amount of precipitation in Köyceğiz station is much more than the others. Also, mean and variance of total precipitation amount are similar in Nazilli and Aydin_Center stations. However, their values are less than Köyceğiz station. Therefore, it can be said that the summary statistics of Nazilli and Aydın_Center are similar and clearly differ from Köyceğiz station. However, they show similar characteristics in terms of the relative frequency of wet days. In addition, they have parallel weather conditions because all of them are under the influence by Mediterranean climate. Therefore, these stations are used to develop two-state and three-state HMMs. Table 6 shows the models and model selection criteria which are AIC and BIC.

Table 6 Comparison of AIC and BIC values for stations from Aegean Region

	Model	**Likelihood Values**	**AIC Values**	**BIC Values**
1	Model 131	19330	38663	38678
2	Model 132	19330	38663	38678
3	Model 133	11062	22129	22143
4	Model 332	11061	22127	22142
5	Model 532	10959	21622	21637
6	Model 533	10840	21683	21698
7	Model 732	10931	21866	21881
8	Model 733	10777	21557	21572
9	Model 932	10925	21854	21869

There are nine models and each model is represented with three numbers. The first number is the number of seasonality term. This seasonal term that is observed in time series as periodic oscilliatons is represented with sine and cosine terms in Eq. (21). Also, it should be periodic and parametric, and its shape is sinusoidal (Stern and Coe, 1984; Neykov et al., 2003). The second number represents the number of stations and third number represents the number of states. For example, Model 332 has three seasonality terms, three stations and two-states.

Models are compared by using AIC and BIC values shown in Table 6. Also, Figure 9 is drawn to observe the increase or decrease in the values of AIC and BIC.

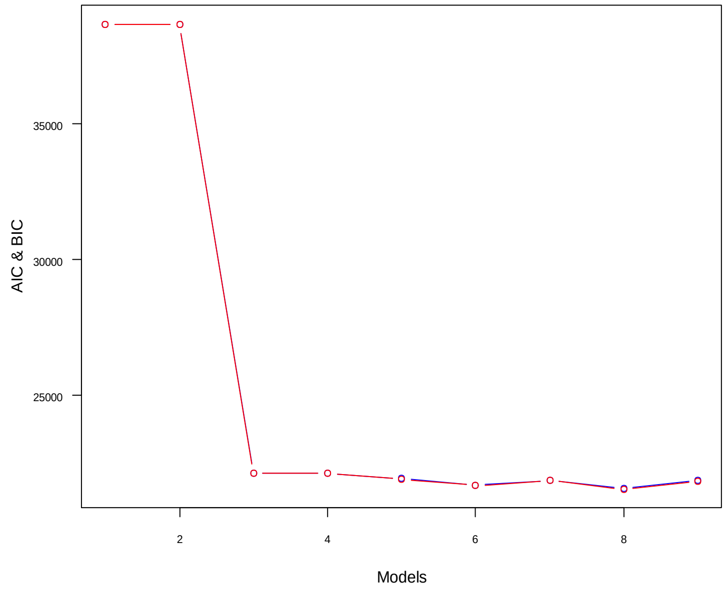

Figure 9 Comparison of AIC and BIC values for stations from the Aegean Region

It is clearly seen that there are nine dots which represent the models in order given in Figure 9. The model which has the minimum values of AIC and/or BIC is chosen as the best model. Therefore, it can be said that "Model 532" and Model "733" are the candidates to be the best model. Before deciding which model is better, other model selection criteria (MSE, MR and Observed versus Predicted Values Plots) should be calculated and analyzed. MSE values are calculated and displayed in Table 7.

Table 7 Comparison of MSE values for stations from the Aegean Region

	Model	**Station Number 17244**	**Station Number 17246**	**Station Number 17900**
1	Model 131	0.0195	0.0213	0.0257
2	Model 132	0.0195	0.0213	0.0257
3	Model 133	0.0182	0.0194	0.0241
4	Model 332	0.0094	0.0091	0.0080
5	Model 532	0.0082	0.0083	0.0079
6	Model 533	0.0083	0.0078	0.0085
7	Model 732	0.0072	0.0072	0.0067
8	Model 733	0.0055	0.0052	0.0051
9	Model 932	0.0007	0.0007	0.0007

It is observed that values of MSE are close to each other. The best model is defined as the model with the minimum value of MSE. When Table 7 is investigated, it can be seen that the minimum values of MSE belong to “Model 932” and “Model 733”. Therefore, one of them is chosen as the best model. Other model selection criterion which is misclassification error rate is calculated and shown in Table 8.

Table 8 Comparison of MR values for stations from the Aegean Region

	Model	**Station Number 17244**	**Station Number 17246**	**Station Number 17900**
1	Model 131	0.4767	0.4958	0.5205
2	Model 132	0.4767	0.4958	0.5205
3	Model 133	0.4685	0.4362	0.4673
4	Model 332	0.1096	0.1068	0.0877
5	Model 532	0.1288	0.1178	0.0986
6	Model 533	0.1276	0.1190	0.1078
7	Model 732	0.1342	0.1370	0.1315
8	Model 733	0.1178	0.1205	0.1150
9	Model 932	0.1342	0.1425	0.1315

In order to calculate MR, the estimated probability of precipitation occurrence is categorized according to a threshold value. It is not chosen as 0.5 since data sets lead to a considerably higher prediction error rate (Kutner et al., 2005). Instead, threshold value has been calculated as follows:

$$Threshold\ value = \frac{Total\ number\ of\ rainy\ days}{Total\ number\ of\ days}. \qquad (61)$$

Threshold value enables to classify the estimated probability values. These values are categorized as “1” if it is greater or equal to the threshold value; and they are categorized as “0” if it is smaller than the threshold value. Threshold values for each station are given in Table 9.

Table 9 Threshold values for stations from Aegean Region

Station Name	Threshold Value
Nazilli	0.2002
Aydın_Center	0.1970
Köyceğiz	0.2140

When Table 8 is examined, it can be said that the MR values of “Model 131” and “Model 132” are very higher than the other models. On the other hand, minimum values of it belong to the “Model 332” and “Model 532.” This means that the probability of precipitation occurrence is more accurate in “Model 332” and “Model 532” when compared with the other models. Since MR values of them are between 0.0877 and 0.1178.

After calculating model selection criteria, observed versus predicted probability of precipitation occurrence graphs are drawn for each station to observe the performance of the model. When all graphs are compared, it can be seen that the graph of “Model 532” is better than the others. The graphs of “Model 532” for each station are shown in Figure 10, Figure 11 and Figure 12. Other graphs are placed in the Appendix A.

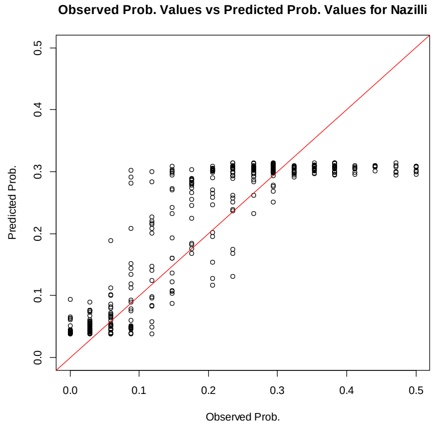

Figure 10 The observed versus predicted probability of precipitation occurrence for Nazilli station

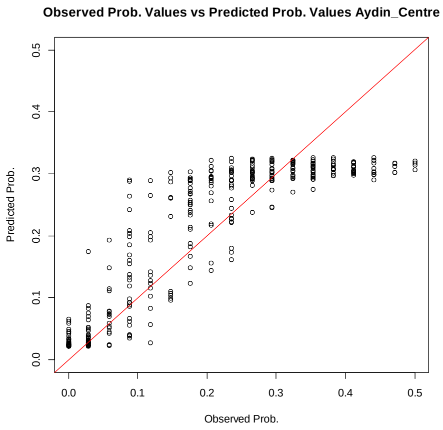

Figure 11 The observed versus predicted probability of precipitation occurrence for Aydın_Center station

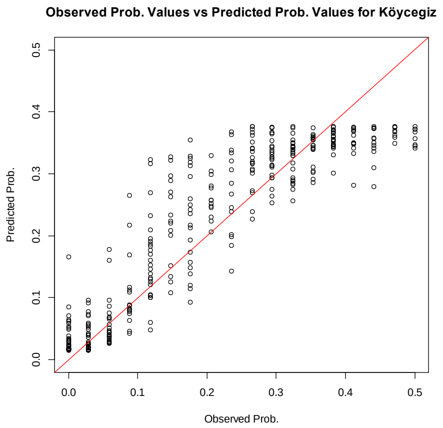

Figure 12 The observed versus predicted probability of precipitation occurrence for Köyceğiz station

The distribution of the observed versus predicted probability values around *"x=y"* line is similar in Figure 10, Figure 11 and Figure 12. This means that predicted probability values are close to observed probability values for stations. Especially, the deviation is significantly lower in Köyceğiz station. However, the maximum values show a small deviation around "*x=y*" line. This result shows that the model should be improved to predict extreme values.

When selecting the best model, several model selection criteria are considered, because a single model selection criterion may not be helpful to decide the best model. However all model selection criteria do not give the same results. Therefore, the best model is defined by using overall performance of them and the observed versus predicted graphs.

In conclusion, “Model 532” gives better results according to AIC, BIC, MR and the observed versus predicted graphs. Therefore, the result of this model is considered for estimation. The estimation graph is drawn and the results are shown in Figure 13.

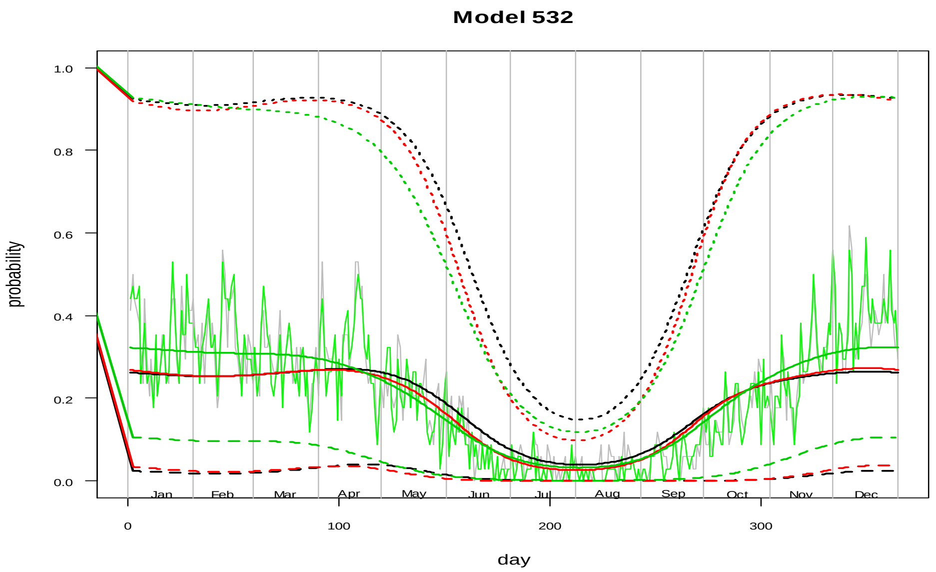

Figure 13 The estimated probability of precipitation occurrence for stations from Aegean Region

The estimation graph shows the probability of precipitation occurrence for three stations in the Aegean Region. According to Figure 13, there are two states and the overall estimated probability of precipitation occurrence which is between two states. The first state represents the minimum estimated probability of precipitation occurrence, which means precipitation will not occur and the second state represents the maximum estimated probability of precipitation occurrence, which means precipitation will occur. In addition, there are three lines which represent the probability of precipitation occurrence for stations in each state. Black line represents the probability of precipitation occurrence for Nazilli station, green line represents the probability of precipitation occurrence for Köyceğiz station and red line represents the probability of precipitation occurrence for Aydın_Center station.

The probability of precipitation occurrence is evaluated for each station according to its closeness to states. For example, the probability of precipitation occurrence for Köyceğiz station in the first day of January is close to state 2; this means that precipitation will occur for Köyceğiz station in the first day of January and each day for each station is evaluated similarly. Also, the overall probability of precipitation occurrence lines for Nazilli station and Aydın_Center station are parallel to each other and are close to state 1, between January and April. This means that there are more dry days in Nazilli station, and Aydın_Center station compare to Köyceğiz station between January and April. In addition, it is observed that the overall probability of precipitation occurrence lines for all stations are very close to first state between June and August. This means that precipitation will not occur between these months.

4.3.2 Analyses of Daily Precipitation Occurrence in Three Stations from Central Anatolian Region

Two-state and three-state HMMs have been developed to estimate the daily probability of precipitation occurrences in the Central Anatolian Region. Bernoulli distribution is chosen as the observation probability distribution which generates observations for predicting the probability of precipitation occurrence in stations.

Three stations which are among the driest stations in Turkey are chosen from the Central Anatolian Region to compare with other regions. In order to make local analyses successfully, they are selected by considering closeness and correlation between them.

Data include the daily amount of precipitation between 1964 and 2005 years. The daily relative frequency of wet days for three stations from the Central Anatolian Region is shown in Figure 14.

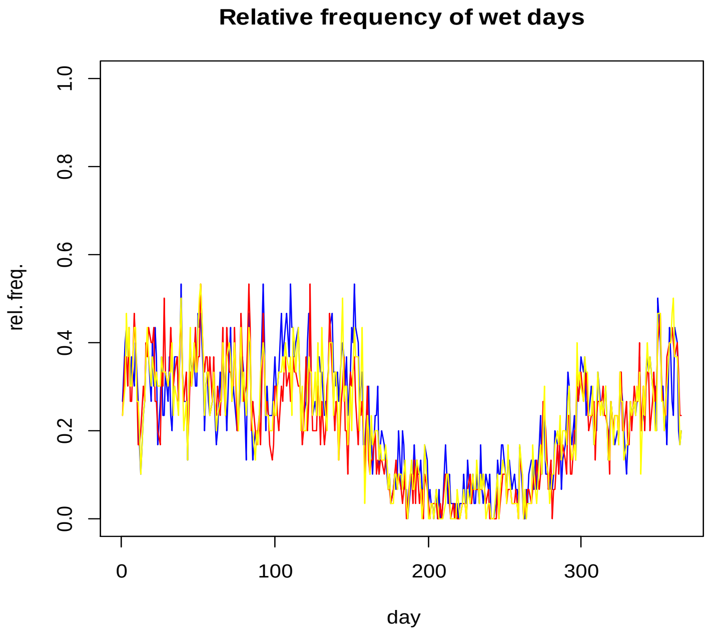

Figure 14 The daily relative frequency of wet days for three stations from Central Anatolian Region

It is observed that there are three lines in Figure 14. The red line shows the corresponding relative frequency for Konya_Center station, the yellow line shows the corresponding relative frequency for Karaman station and the blue line shows the corresponding relative frequency for Çumra station. These lines are not the same; however there are similarities between them. In addition, summary statistics are calculated to observe the addition similarities and are shown in Table 10.

Table 10 Summary statistics for stations from Central Anatolian Region

Station Name	Min	Max	Median	Mean	Variance	Standard Deviation
Konya_Centre	0	64.50	0	0.85	10.20	3.19
Karaman	0	60.70	0	0.88	10.14	3.18
Çumra	0	50.10	0	0.86	9.08	3.01

When Table 10 is analyzed, the minimum value of precipitation amount for three stations is the same. However, the maximum values of precipitation amount for three stations are different. Specially, the maximum value of precipitation amount in Çumra station is less than others. In addition, mean and variance of precipitation amount are similar for three stations. Therefore, it can be said that these stations show similar characteristics according to the relative frequency of wet days and summary statistics and they are used to constitute two-state and three-state HMMs. AIC and BIC values calculated for nine different HMMs are shown in Table 11.

Table 11 Comparison of AIC and BIC values for stations from Central Anatolian Region

	Model	Likelihood Values	AIC Values	BIC Values
1	Model 131	17297	34597	34612
2	Model 132	17297	34597	34612
3	Model 133	11336	22676	22690
4	Model 332	11393	22790	22805
5	Model 532	11307	22617	22632
6	Model 533	11200	22617	22632
7	Model 732	11290	22258	22272
8	Model 733	11127	22258	22272
9	Model 932	11281	22567	22581

Table 11 shows the models and their associated AIC and BIC values. The meaning of the numbers at the name of the models is as explained in section 4.3.1. Figure 15 is composed in order to see the changes in the values of AIC and BIC.

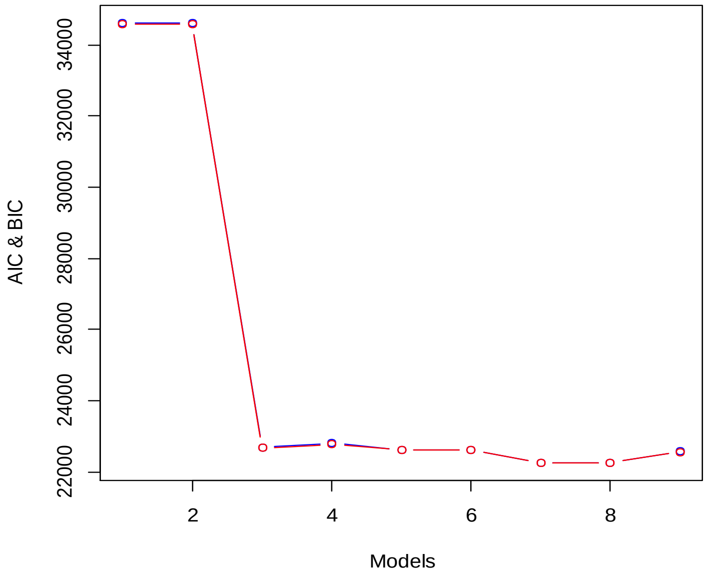

Figure 15 Comparison of AIC and BIC values for stations from Central Anatolian Region

There are nine dots that present the models in order given in Table 11. It can be said that there is no difference between the two-state and three-state models with respect to the AIC and BIC values for “Model 532” and “Model 732.” The minimum values of AIC and BIC are observed in “Model 732” and “Model 932.” Therefore, they can be best models according to AIC and BIC values. However, other model selection criteria should be calculated and examined, before deciding which model is the better. MSE values are shown in Table 12.

Table 12 Comparison of MSE values for stations from the Central Anatolian Region

	Model	**Station Number 17244**	**Station Number 17246**	**Station Number 17900**
1	Model 131	0.0164	0.0168	0.0169
2	Model 132	0.0164	0.0168	0.0169
3	Model 133	0.0158	0.0154	0.0163
4	Model 332	0.0117	0.0082	0.0103
5	Model 532	0.0080	0.0067	0.0078
6	Model 533	0.0081	0.0065	0.0079
7	Model 732	0.0077	0.0065	0.0078
8	Model 733	0.0081	0.0066	0.0079
9	Model 932	0.0076	0.0065	0.0076

Table 12 shows the MSE values are very small. In other words, the observed probability values are close to estimated probability values. The minimum value of MSE is used to describe the best model, and “Model 732” and “Model 932” have minimum values of MSE. This means that “Model 732” and “Model 932” are the best models according to MSE values. Another model selection criterion which is MR is calculated and shown in Table 13.

Table 13 Comparison of MR values for stations from the Central Anatolian Region

	Model	**Station Number 17244**	**Station Number 17246**	**Station Number 17900**
1	Model 131	0.4520	0.4931	0.4931
2	Model 132	0.5479	0.5068	0.5397
3	Model 133	0.4489	0.5043	0.4912
4	Model 332	0.2320	0.2109	0.2164
5	Model 532	0.1424	0.1753	0.1369
6	Model 533	0.1432	0.1618	0.1462
7	Model 732	0.1424	0.1616	0.1452
8	Model 733	0.1433	0.1620	0.1467
9	Model 932	0.1369	0.1534	0.1452

Table 13 displays MR values for each model. Estimated and observed probabilities are categorized as “0” and “1” by using threshold value as calculated in the Aegean Region. These values for each station are in Table 14.

Table 14 Threshold values for stations from Central Anatolian Region

Station Name	**Threshold Value**
17244	0.2284
17246	0.2111
17900	0.2199

According to Table 14, the MR values are still high in “Model 131” and “Model 132.” However, the value of misclassification error rates obtained from other models is less than them.

Also, it can be said that the probability of precipitation occurrence is more accurate in “Model 532” , “Model 732” and “Model 932” when compared with other models because the minimum values of MR are belong to them. In addition, MR values from Central Anatolian Region are higher than MR values from the Aegean Region. This shows that the results of HMMs develop for the Aegean Region seems to be better than for Central Anatolian Region.

In addition to model selection criteria, observed versus predicted probability of precipitation occurrence graphs are composed to observe the performance of the model. When graphs which are composed for each model are compared, the graph of “Model 932” is better than others. Other graphs are placed in the Appendix A.

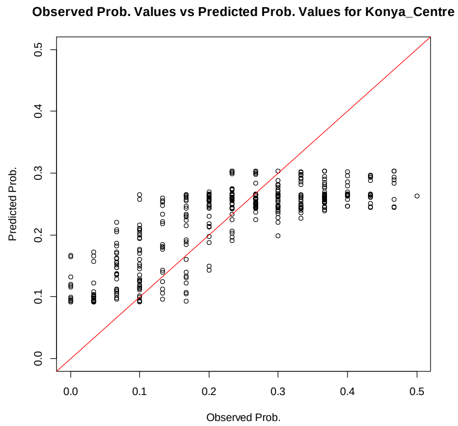

Figure 16 The observed versus predicted probability of precipitation occurrence for Konya_Center station

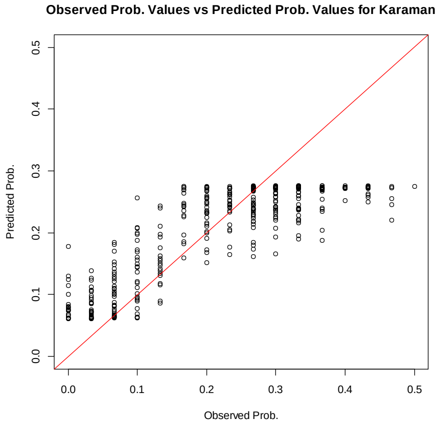

Figure 17 The observed versus predicted probability of precipitation occurrence values for Karaman station

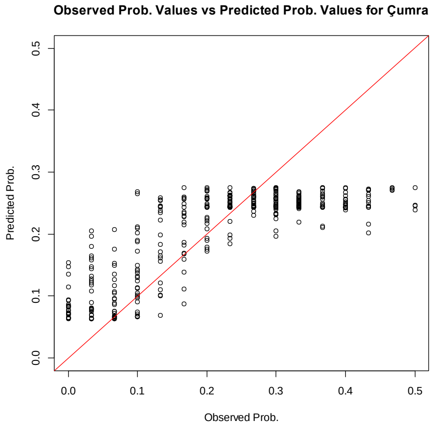

Figure 18 The observed versus predicted probability of precipitation occurrence values for Çumra station

It is said that the observed and predicted probability of precipitation occurrence values are similar for small probability values in Figure 16, Figure 17 and Figure 18. However, there is a deviation around "x=y" line for high probability values. This shows that model which is develop to estimate the occurrence probability for the Central Anatolian Region lack of estimating maximum probabilities.

Finally, "Model 732" is defined as best model by using model selection criteria and the observed versus predicted probability of precipitation occurrence graphs.

For this reason, the result of this model is used for estimation. The estimation graph is made up and the results are shown in Figure 19.

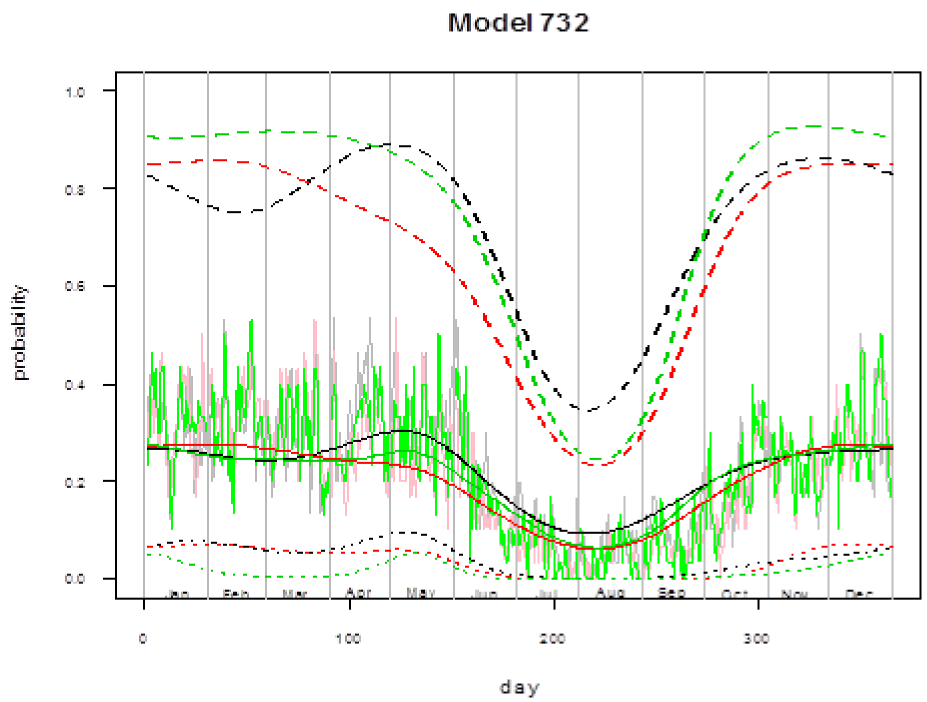

Figure 19 The estimated probability of precipitation occurrence for stations from Central Anatolian Region

Figure 19 shows two state and overall estimated probability of precipitation occurrence lines. The minimum estimated probability of precipitation occurrence is defined as first state and the maximum estimated probability of precipitation occurrence is defined as second state. Overall estimated probability of precipitation occurrence is between the first state and the second state. Also, it is observed that there are three lines in each state. They represent the probability of precipitation occurrence for stations. Black line represents the probability of precipitation occurrence for Konya_Center station, green line represents the probability of precipitation occurrence for Karaman station and red line represents the probability of precipitation occurrence for Çumra station.

Overall estimated probability of precipitation occurrence lines are used to predict the precipitation. The probability of precipitation occurrence is evaluated for each station according to their closeness to states. For example, the probability of precipitation occurrence for stations in the first day of January and the last day of December are close to state 2; this means that precipitation is likely to occur at Konya_Center station, Karaman station and Çumra station in the first day of January and last day of December. Similarly, the probability of precipitation occurrence for each day and each station can be evaluated. Also, it is observed that the overall estimated probability line for Konya_Center station is close to state 2 and the overall estimated probability lines for Karaman station and Çumra station are close to state 1 between April and September. This means that there are more rainy days in Konya_Center station compare with Karaman station and Çumra station in April through September.

4.3.3 Analyses of Daily Precipitation Occurrence in Three Stations from East Black Sea Region

The daily probability of precipitation occurrence is analyzed with two-state and three-state HMMs in East Black Sea Region. Observations are generated by Bernoulli distribution which is described as the observation probability distribution.

Three stations which are located near to each other and correlated are selected to make regional analyses accurately. Data have the daily precipitation amounts from 1964 to 2005. Time series plot of relative frequency of wet days for three stations from East Black Sea Region is shown in Figure 20.

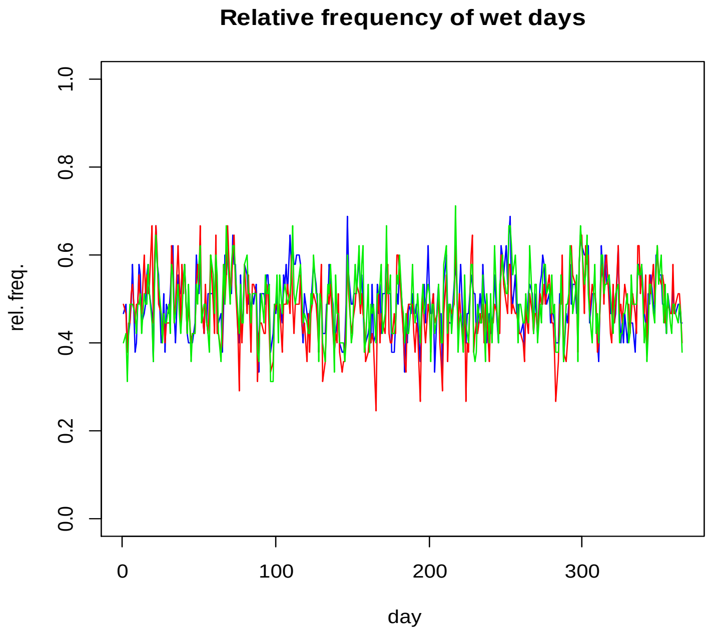

Figure 20 The relative frequency of wet days for stations from East Black Sea Region

Figure 20 shows the corresponding relative frequency lines for each station. These line shows similarity with each other. This result helps to make local analyses. In addition, it is calculated the summary statistics for three stations and it is given in Table 15.

Table 15 Summary statistics for stations from East Black Sea Region

Station Name	Min	Max	Median	Mean	Variance	Standard Deviation
Rize_Centre	0	178.7	0	6.08	178.9	13.38
Pazar	0	186.2	0	5.59	171.9	13.11
Hopa	0	209.8	0	6.09	187.5	13.69

Table 15 displays that the minimum value of precipitation for three stations is the same and maximum amount of precipitation for three stations are different; especially the amount of precipitation in Hopa station is greater than the others.

Mean of precipitation amount is similar in Rize_Centre station and Hopa station, and higher than Pazar station. On the other hand, variances of total precipitation amount are similar in Rize_Centre station and Pazar station, and less than Hopa station.

It can be said that Rize_Centre, Pazar and Hopa stations show similar characteristics according to the relative frequency of wet days and summary statistics. Therefore nine HMMs are developed and calculated AIC and BIC values to compare the models. Models and values of AIC and BIC are given in Table 16.

Table 16 Comparison of AIC and BIC values for station from East Black Sea Region

	Model	**Likelihood Values**	**AIC Values**	**BIC Values**
1	Model 131	34133	68720	68285
2	Model 132	34133	68270	68285
3	Model 133	22689	45832	45997
4	Model 332	22808	45621	45636
5	Model 532	22779	45562	45578
6	Model 533	22779	45562	45578
7	Model 732	22766	45536	45552
8	Model 733	22766	45536	45552
9	Model 932	22754	45511	45527

Table 16 gives the two-state and three-state HMMs with their AIC and BIC values. It is clearly seen that there are nine models and each model is represented with three numbers. These numbers represents the number of seasonality terms, stations and states. Detailed information about the models is given in section 4.3.1. The changes in the values of AIC and BIC can be observed in Figure 21.

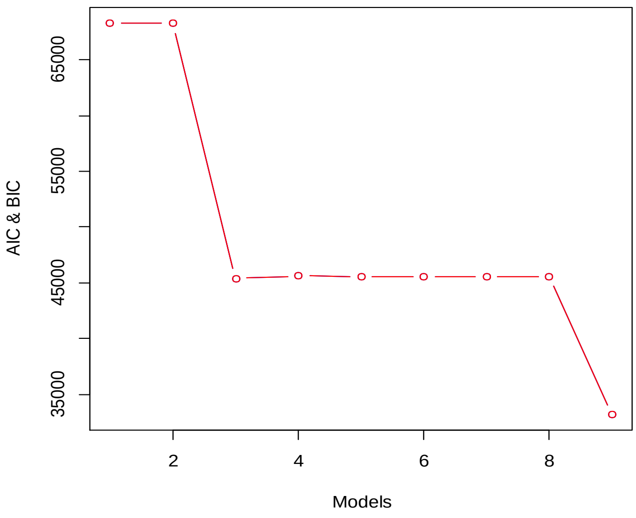

Figure 21 Comparison of HMMs by AIC and BIC

It is observed that AIC and BIC values are decreasing at the beginning, then there are little increase and decrease and then there is a continuous decrease. Also, there is no difference between the two-state and three-state models with respect to AIC and BIC values in Model 532 and Model 732. Model 932 can be chosen as the best model according to AIC and BIC values. Before deciding which model is the best, other criteria such as MSE, MR should be computed and the observed versus predicted values graphs should be examined. MSE values are shown in Table 17.

Table 17 Comparison of MSE values for stations from East Black Sea Region

	Model	**Station Number 17040**	**Station Number 17042**	**Station Number 17628**
1	Model 131	0.00471	0.00584	0.00542
2	Model 132	0.00471	0.00584	0.00543
3	Model 133	0.00471	0.00573	0.00542
4	Model 332	0.00470	0.00536	0.00544
5	Model 532	0.00468	0.00539	0.00541
6	Model 533	0.00468	0.00538	0.00538
7	Model 732	0.00467	0.00539	0.00537
8	Model 733	0.00477	0.00540	0.00536
9	Model 932	0.00481	0.00537	0.00535

When Table 17 is examined, it can be said that observed and predicted occurrence probabilities are close to each other, and hence, MSE values are very small. However, it is observed that the minimum values of MSE belong to the "Model 732" and "Model 932." Therefore, "Model 732" and "Model 932" are the best models according to MSE values. In addition, the MR values are calculated and given in Table 18.

Table 18 Comparison of MR values for stations from East Black Sea Region

	Model	**Station Number 17040**	**Station Number 17042**	**Station Number 17628**
1	Model 131	0.4547	0.4821	0.4438
2	Model 132	0.5095	0.4054	0.5013
3	Model 133	0.5082	0.4039	0.5011
4	Model 332	0.4657	0.4027	0.4328
5	Model 532	0.4767	0.3972	0.4383
6	Model 533	0.4689	0.4056	0.4231
7	Model 732	0.4739	0.4027	0.4109
8	Model 733	0.4487	0.4068	0.4039
9	Model 932	0.4465	0.4000	0.4027

It is observe that the MR values of HMMs are not very small although the threshold value is not defined as "0.5". It is calculated as in the Aegean and the Central Anatolian Regions. However, it can be observed that MR values are still high for stations from East Black Sea Region compared to results obtained from other regions. This might be due to the fact that East Black Sea Region is an extremely wet region and its distribution is highly skewed. Threshold values for each station are displayed in Table 19.

Table 19 Threshold values for stations from East Black Sea Region

Station Name	**Threshold Value**
Rize_Center	0.4923
Hopa	0.4741
Pazar	0.4913

It is observe that the MR values of HMMs are not very small. When models are compared according to MR, “Model 932” seems to be better than the other models, since it has smaller misclassification values than others.

In order to evaluate the performance of models, observed versus predicted probability of precipitation occurrence graphs are composed in addition to model selection criteria and it is concluded that the graph of “Model 932” is better than others. Other graphs are placed in the Appendix A.

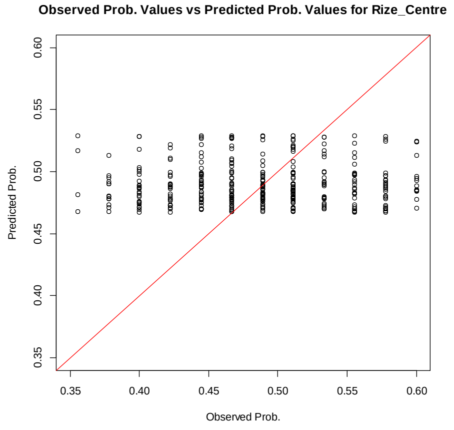

Figure 22 The observed versus predicted probability precipitation occurence for Rize_Center station

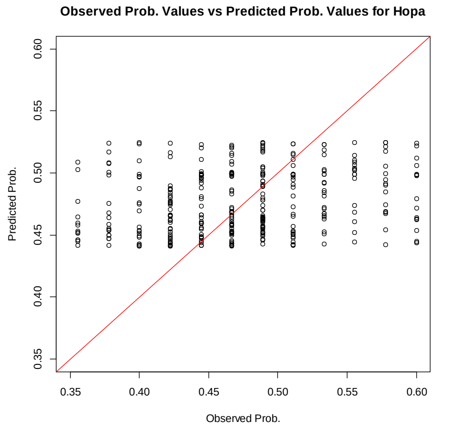

Figure 23 The observed versus predicted probability of precipitation occurrence for Hopa station

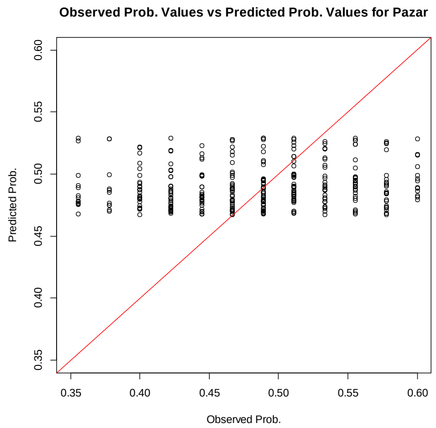

Figure 24 The observed versus predicted proability of precipitation occurrence for Pazar station

When Figures 22, 23 and 24 are analyzed, it can be said that the observed and predicted probability values do not look like each other. Especially, extreme values show great deviation around "x=y" line. This means that models for East Black Sea Region do not give good results when compare with other regions.

In conclusion, "Model 932" can be selected for estimation because model selection criteria and the graph of observed and predicted values of this model are considerably good. The estimation plot is drawn by using results of this model and it is shown in Figure 25.

Figure 25 Estimation values for East Black Sea Region

We can observe the probability of precipitation occurrence for three stations by using the estimation plots given in Figure 25. There are two states from which the two-state HMM is derived. The first state represents the minimum estimate of probability for precipitation occurrence, which means precipitation will not occur and the second state represents the maximum estimate of probability for precipitation occurrence, which means precipitation will occur. Also, there is overall estimate of probability for precipitation which is derived from our model. The probability of precipitation occurrence is evaluated for each station according to its closeness to states.

Two states which are minimum and maximum estimated probability of precipitation occurrence and overall estimated probability of precipitation occurrence lines can be seen in Figure 24. There are three lines which represent the probability of precipitation occurrence for stations in each station. Black line represents the probability of precipitation occurrence for Rize_Center station, green line represents the probability of precipitation occurrence for Hopa station and red line represents the probability of precipitation occurrence for Pazar station. Overall estimated probability of precipitation occurrence which is between the first state and second state is used for estimating daily precipitation occurrence for each station. If it is close to the first state, precipitation will not occur; if it is close to the second state, precipitation will occur.

The probability of precipitation occurrence is evaluated for each station according to it closeness to states. When Figure 25 is analyzed, it is said that overall estimated probability of precipitation lines for Rize_Center station and Hopa station are parallel to each other and they are close to second state between August and October. This means that there are more rainy days at these stations compare with Pazar station at these months. Therefore, each day for each station is evaluated similarly. In addition, it is observed that the overall estimated probability values are higher than other regions. In other words, East Black Sea Region has more rainy days compare the Aegean and Central Anatolian Region according to this estimation plot. However, it is observed that the misclassification error rates which are calculated from the Aegean and Central Anatolian Region are smaller than misclassification rates which are derived from the East Black Sea Region. Also, when the graph of observed versus predicted probability values is examined, it can be said that there is great deviation around "x=y" lines. This shows that HMMs which are develop for estimating occurrence probabilities to East Black Sea Region do not give good prediction for precipitation occurrence compare with the other regions.

In other words, two-state homogenous HMMs are useful tool for predicting overall probability of precipitation in regions like the Aegean and Central Anatolian.

4.4 Results of Daily Total Precipitation Amount Analyses

In the following sections, the daily total precipitation amount in one station which takes place in the East Black Sea Region, Central Anatolian Region and Aegean Region are estimated and evaluated separately.

4.4.1 Analyses of the Precipitation Amount Estimation for One Station from East Black Sea Region

The distribution of rainfall amount on the wet days is assumed as positively skewed. Because, it is known that larger amounts occur less than smaller amounts. In order to model daily precipitation amount data, gamma distribution is chosen. Daily precipitation amount data of Rize_Center station is used to estimate the total precipitation amount by using HMMs. This station is chosen among stations which are used to estimate precipitation occurrence.

Data include the daily precipitation amount between 1964 and 2005 years. Some graphics are composed to show the difference of daily total precipitation amount between days. The mean of daily total precipitation amount is shown in Figure 26.

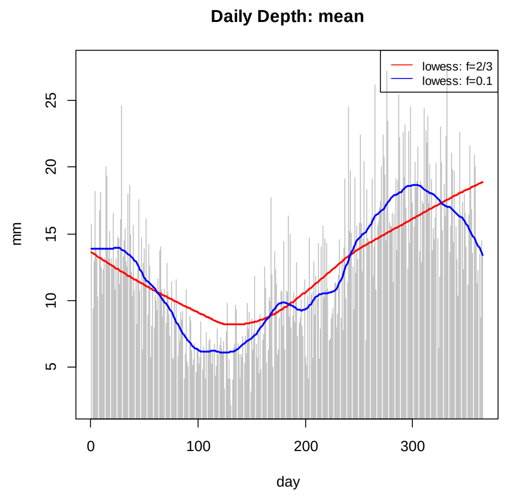

Figure 26 Mean of daily total precipitation amount for Rize_Center station from East Black Sea Region

Figure 26 shows the mean of precipitation and two smoothing line for Rize_Center station. It is said that mean of precipitation decreases until 100th day of the year and then it increases and reaches maximum value around the 300th day of the year. In general, overall mean of daily total precipitation amount is above the 5mm. Also, the standard deviation of daily total precipitation amount is shown in Figure 27.

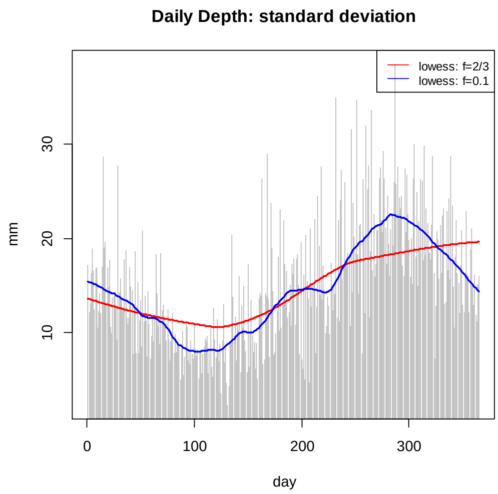

Figure 27 Standard deviation of daily precipitation amount for Rize_Center station from East Black Sea Region

Figure 27 shows the standard deviation of daily precipitation amount for Rize_Center station. This value is the smallest around 100th day of year which means that the precipitation amount of these days closes to mean and it reaches higher value around the 300th day of the year which means that the total precipitation amount of these days are spread out over a large range of total precipitation amounts. In addition, the coefficient of variation of daily total precipitation amount is calculated and it is displayed in Figure 28.

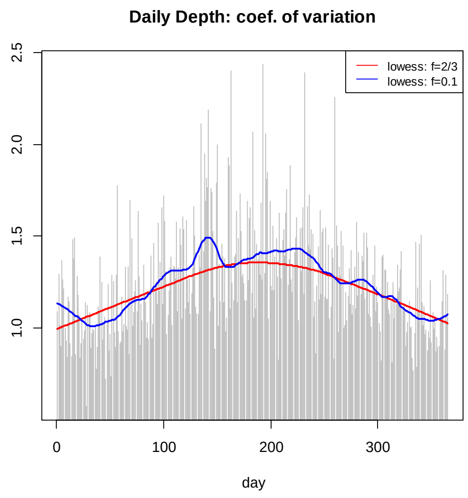

Figure 28 Coefficient of variation values of daily total precipitation amount for Rize_Center station from East Black Sea Region

Figure 28 shows that the coefficient of variation of daily precipitation amount is not the same for all days. This means that there is different ratio of standard deviation to the mean between the days.

In order to analyze the daily precipitation amount data, two-state seasonal gamma HMMs are developed. Information about the models is given in Table 20.

Table 20 Information about two-state seasonal gamma HMMs

Models	**Seasonal Components for Mean**	**Seasonal Components for Coefficient of Variation**	**The Number of States**
Model 1	3	3	2
Model 2	5	3	2
Model 3	5	5	2
Model 4	7	3	2
Model 5	7	5	2

It can be seen that there are five HMMs developed to observe the daily total precipitation amount for Rize_Center station. Each model contains estimates. For example, “Model 1” has two-state and each state has three parameters in the seasonal component for the mean and three parameters for the coefficient of variation, i.e., a total of 12 parameters. Hence, other models contain more than 12 parameters. In addition, the values of AIC and BIC are calculated to compare the models and they are given in Table 21.

Table 21 Comparison of AIC and BIC for Rize_Center Station from East Black Sea Region

Model	Likelihood Values	AIC Values	BIC Values
Model 1	26694	53416	53524
Model 2	26675	53386	53525
Model 3	26660	53365	53534
Model 4	26673	53391	53560
Model 5	26659	53370	53570

As can be seen from the Table 21, “Model 3” has minimum AIC value and “Model 1” has minimum BIC value. Therefore, they are candidates of the best model. Before deciding which model is better, MSE values are calculated and a graph which includes the observed versus predicted values is drawn. MSE values are shown in Table 22.

Table 22 MSE values for Rize_Center station from East Black Sea Region

Model	MSE Values
Model 1	43.8062
Model 2	43.2053
Model 3	43.6797
Model 4	43.1057
Model 5	42.9254

When Table 22 is analyzed, it can be seen that MSE values are close to each other. However, it is observed that the minimum values of MSE belong to the “Model 4” and “Model 5.” Hence one of these models could be the best model to observe the total precipitation amount for Rize_Center station according to MSE values. In addition to model selection criteria, graphs that contain the observed versus the predicted precipitation values are drawn for each model and it is observed that the result of “Model 4” is the best compare with other models. Other graphs are placed in the Appendix A.

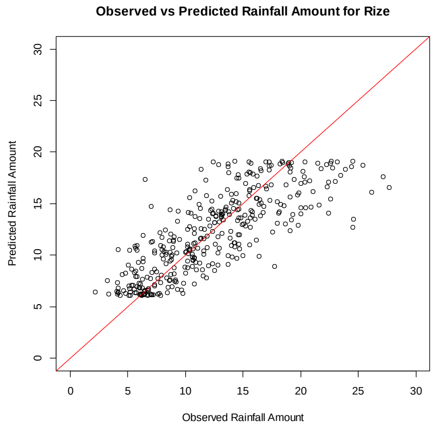

Figure 29 The observed versus predicted values of precipitation amount for Rize_Center station

Figure 29 displays the observed and predicted precipitation amount values look like each other. However, there is a deviation around “x=y” line especially in maximum probability values. In conclusion, it seems to logical choose the best model as Model 4 according to BIC, MSE values and graph of the observed and predicted precipitation amount values. Finally, the results of this model are used for estimation. Estimation plots are shown in Figure 30.

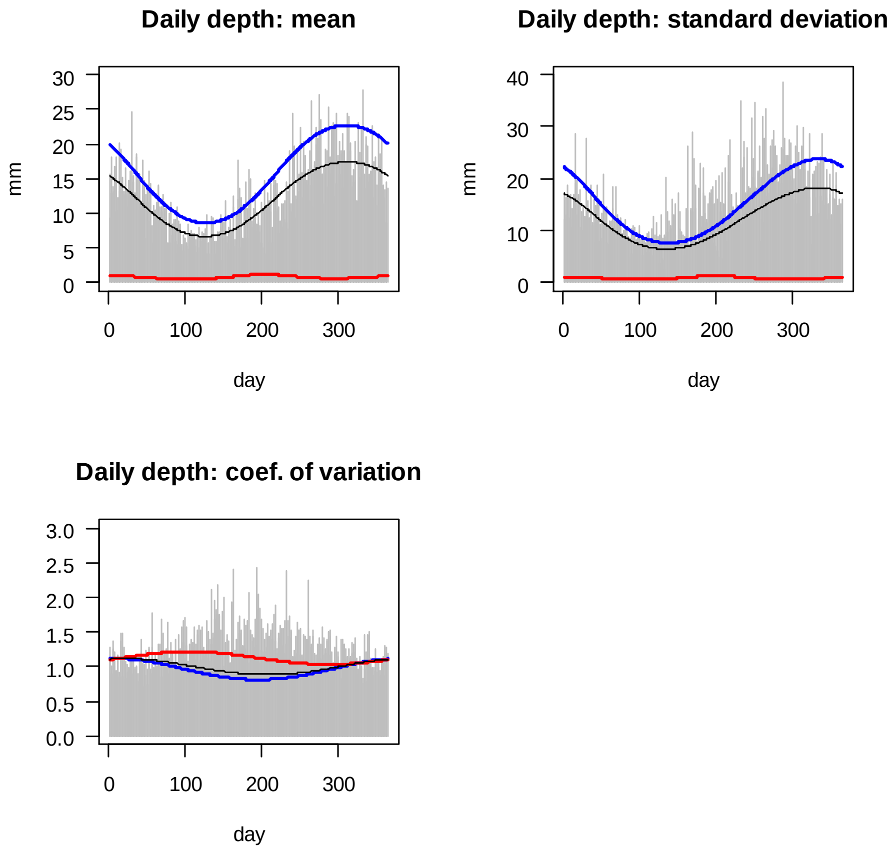

Figure 30 Estimation of the mean, standard deviation and coefficient of variation of total precipitation amount for Rize_Center station

The estimation plots of mean, standard deviation and coefficient of variation about precipitation amount for Rize_Center station are given in Figure 30. There are two states which are derived from two-state HMMs and overall value in each estimation plots. The first state (red line) represents the minimum estimated value of mean, standard deviation and coefficient of variation for precipitation amount. The second state (blue line) represents the maximum estimated value of mean, standard deviation and coefficient of variation for precipitation amount. However these states are unknown from definition of HMM and they can be defined according to goal of the study. Also, there is an overall estimate of precipitation amount (dark line) which is between the first state and the second state. When the precipitation amount is estimated, the overall estimate of precipitation amount help to evaluate the precipitation amount according to closeness to states. For example, estimation plot for mean indicates that overall estimate of precipitation amount line close to the second state. This gives an idea about the mean of precipitation amount to us. For Rize_Center station, it will be about 10mm. In addition to mean of precipitation amount, standard deviation of precipitation amount and coefficient of variation of precipitation can be evaluated day by day with these estimation plots. Also, predictions of 365 days related to mean, standard deviation and coefficient of variation of precipitation amount give general information about daily precipitation amount for people.

4.4.2 Analyses of the Precipitation Amount Estimation for One Station from Central Anatolian Region

In order to estimate daily precipitation amount at the Central Anatolian Region, Konya_Center station is selected. This station is one of the stations which are used to estimate precipitation occurrence. Data include the daily amount of total precipitation between 1977 and 2006 years. Some graphics are drawn to get general opinion about the difference of daily precipitation amount between days. The mean of daily precipitation amount is shown in Figure 31.

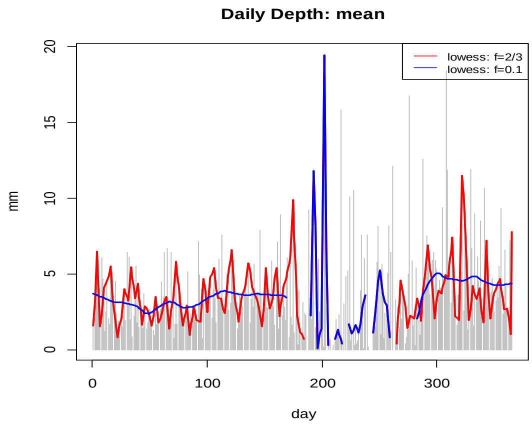

Figure 31 Mean of daily total precipitation amount for Konya_Center station

According to Figure 31, the mean of precipitation amount is low for Konya_Center station compare to Rize_Center station. There are few days for which mean of precipitation amount is more than 10mm. In general, overall mean of daily precipitation amount is below 5 mm. Also, the standard deviation of daily precipitation amount is displayed in Figure 32.

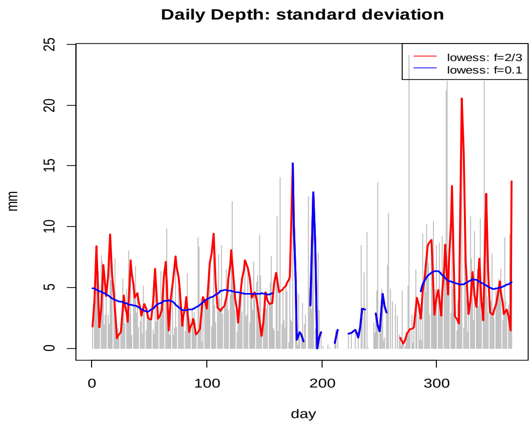

Figure 32 Standard deviation values of daily total precipitation amount for Konya_Center station

Figure 32 provides information about the standard deviation of total precipitation for Konya_Center station. This value is zero around 200th day of year which means that the precipitation amount of these days closes to mean. Also, it reaches high values around the 300th day of the year which means that the total precipitation amount of these days spread out over a large range of precipitation amounts. In general, overall standard deviation of daily total precipitation amount is between 0 mm and 5 mm. In addition to the mean and standard deviation of precipitation amount, the coefficient of variation of daily precipitation amount is calculated and it is given in Figure 33.

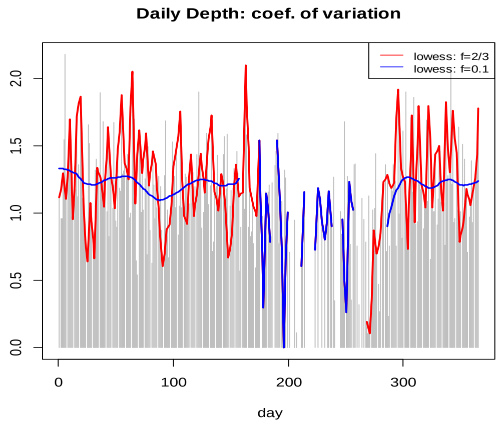

Figure 33 The coefficient of variation of daily precipitation amount for Konya_Center station

It is observed that the coefficient of variation is not the same for all days from Figure 33. This shows that there is a significant difference between the days in terms of mean and standard deviation. Also, there are some days that coefficient of variation values cannot be calculated. This happens because mean takes value of zero.

Two-state seasonal HMMs have been developed to estimate the daily precipitation amount at the Central Anatolian Region. Gamma distribution is selected as the observation probability distribution which generates observations for predicting the probability of precipitation amount at stations. To compare models, model selection criteria such as AIC, BIC and MSE are calculated and graphs which include the observed and predicted precipitation amount are drawn. It has been shown that AIC and BIC values for five different HMMs in Table 23.

Table 23 Comparison of AIC and BIC values for Konya_Center station

Model	Likelihood Values	AIC Values	BIC Values
Model 1	5433	10894	10996
Model 2	5419	10874	11005
Model 3	5418	10880	11041
Model 4	5412	10868	11028
Model 5	5410	10873	11063

Models and the values of AIC and BIC are given in Table 23. The explanation of models is as given in section 4.4.1. This table indicates that the minimum value of AIC belongs to "Model 4" and the minimum value of BIC belongs to "Model 1." However, MSE should be calculated and graphs which include the observed and predicted values should be analyzed to describe the best model. Table 24 displays the MSE values.

Table 24 Comparison of MSE values for Konya _Center station

Model	MSE Values
Model 1	10.9192
Model 2	9.1819
Model 3	9.1487
Model 4	9.2455
Model 5	9.2251

It is said that there is no big difference between the MSE values of models. However, it can be seen that the minimum MSE value is found with "Model 3." Hence it could be the best model to estimate the total precipitation amount for Konya_Center station according to MSE values. In addition to MSE values, some graphs are drawn to see the values of observed versus predicted precipitation amount together.

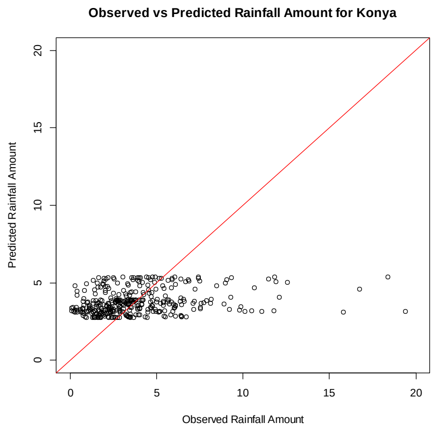

Figure 34 The observed versus predicted values of precipitation amount for Konya_Center station

It can be clearly seen that the values of observed and predicted precipitation amount do not seem to be similar. There is high deviation around "$x=y$" line. Because the predicted precipitation amount values are higher than observed precipitation amount. However, the graph which is drawn by using the result of "Model 3" is better than other models. Other graphs can be seen in Appendix A. At the end, "Model 3" is chosen as the best model according to the values of BIC, MSE and the graph of the observed versus predicted precipitation amount. Therefore, it is used for estimation. The plots of estimation are drawn by using the result of "Model 3."

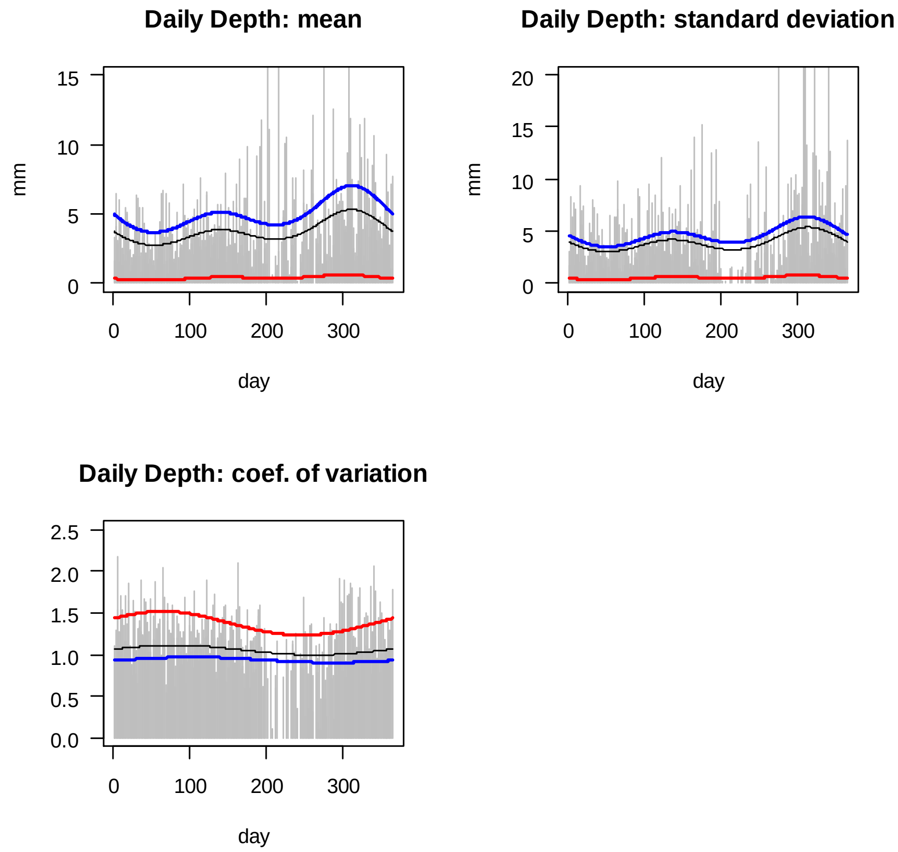

Figure 35 Estimation of precipitation amount for Konya_Center station

Figure 35 includes the estimation plots of mean, standard deviation and coefficient of variation about precipitation amount for Konya_Center station. There are two states which are derived from two-state HMM in each graph. The minimum estimate of precipitation amount is presented by the first state (red line). The maximum estimate of precipitation amount is presented by the second state (blue line). Also, there is overall estimate of total precipitation amount (dark line). Precipitation amount is predicted according to the position of the overall line. For example, overall line is close to the second state in the graphs of mean and standard deviation of precipitation amount. It means that the value of mean and precipitation amount will be around 5mm. On the other hand, overall line is close to the first state in the graph of coefficient of variation. It means that the value of coefficient variation will be around 1mm. Therefore; the value of mean, standard deviation and coefficient of variation for precipitation amount are estimated day by day in addition to overall evaluation for precipitation amount by using the plots of estimation.

4.4.3 Analyses of Total Precipitation Amount Estimation for One Rain Station from Aegean Region

In order to develop a model for Aegean Region, Aydın_Centre station is chosen. This station is chosen among stations which are used to estimate precipitation occurrence. Data include the daily amount of precipitation between 1972 and 2005 years. The graphs of mean, standard deviation and coefficient of variation are drawn to observe the difference of daily precipitation amount between days. The mean of daily precipitation amount is shown in Figure 36.

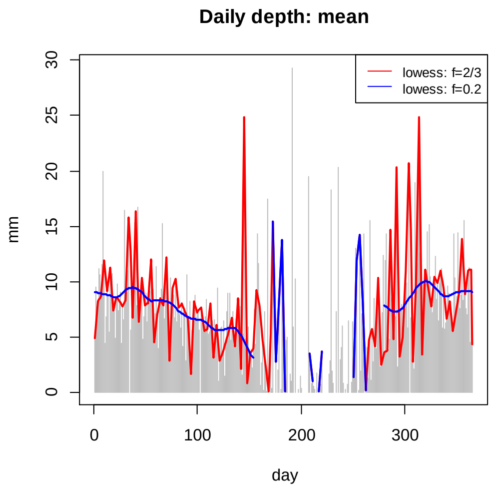

Figure 36 Mean of daily precipitation amount for Aydın_Center station

As can be seen from Figure 36, the mean of precipitation is between 5mm and 10mm. There are few days whose mean of precipitation amount is more than 20mm. In general, overall mean of daily total precipitation amount is below 10 mm. However, mean of daily total precipitation amount in Aydın_Center station is higher than Konya_Centre station from Central Anatolian Region and less than Rize_Center station from East Black Sea Region. Also, the standard deviation of daily precipitation amount is calculated and displayed in Figure 37.

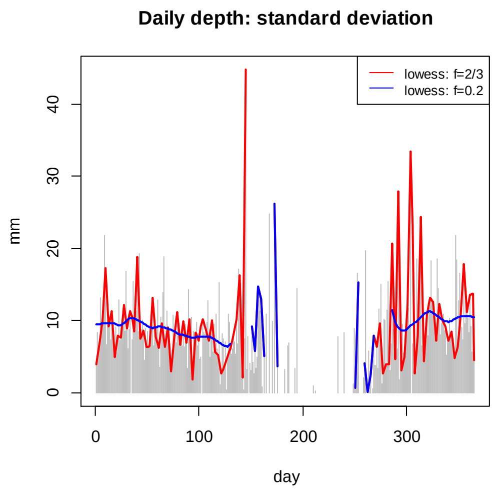

Figure 37 Standard deviation of daily precipitation amount for Aydın_Center station

Figure 37 reveals the standard deviation of precipitation amount for Aydın_Center station and it is zero around 200th day of year which means that the precipitation amount is equal to mean of it. In general, overall standard deviation of daily total precipitation amount is 0 mm and 10 mm. In addition, the coefficient of variation of daily precipitation amount is calculated to compare the ratio of mean and standard deviation for every day.

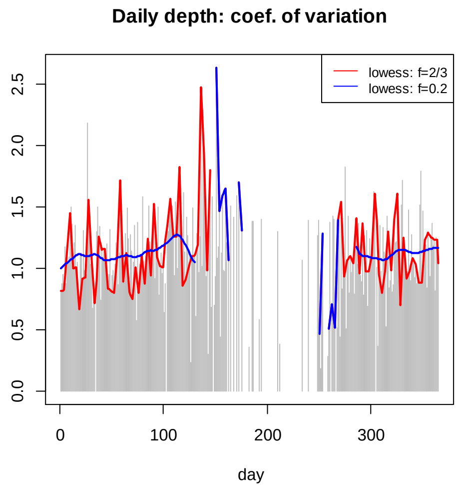

Figure 38 Coefficient of variation values of daily total precipitation amount for Aydın_Center station

The coefficient of variation of daily precipitation amount is revealed in Figure 38. It is not the same for all days. However, there is similarity between the first 100 and the last day of the year. Also, it cannot be calculated the values of coefficient of variations for many days since the mean of precipitation amount is zero.

In the following sections, AIC, BIC and MSE are calculated by using the results of the two-state seasonal gamma HMMs. Table 25 indicates the models and the values of AIC and BIC.

Table 25 Comparison of AIC and BIC values for Aydın_Center station

Model	Likelihood Values	AIC Values	BIC Values
Model 1	7638	15304	15408
Model 2	7634	15304	15438
Model 3	7628	15301	15464
Model 4	7629	15301	15465.
Model 5	7625	15303	15496

The values of AIC and BIC in Table 25 indicates that the minimum value of AIC are found in both “Model 3” and “Model 4” and the minimum value of BIC is found in “Model 2.” These models can be described as the best model. However, MSE and the graphs of observed versus predicted values should be controlled to decide the best model. The MSE values are given in Table 26.

Table 26 Comparison of MSE values for Aydın_Center station

Model	MSE Values
Model 1	103.5531
Model 2	103.2161
Model 3	84.1768
Model 4	101.9010
Model 5	101.6165

As it can be seen from Table 26, the values of MSE are similar except for the MSE value for “Model 3” which is the minimum value. Therefore, it could be the best model to estimate the precipitation amount for Aydın_Centre station according to MSE values.

In order to see the closeness of observed versus predicted precipitation amount, some graphs are drawn by using the results of two-state seasonal gamma HMMs. It is found that all graphs are similar to each other. A graph which is drawn by using the result of “Model 3” is shown in below.

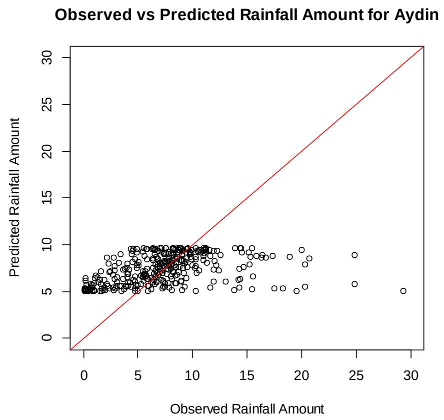

Figure 39 The observed versus predicted values of precipitation amount for Aydın_Center station

It is figured out that the values of observed and predicted precipitation amount values do not show similarity. There is a high deviation around “x=y” line as in Konya_Centre station from Central Anatolian Region and an underestimation of the precipitation amount.

Finally, “Model 3” seems to be better than the other models according to the values of BIC, MSE and the graph of the observed versus predicted values. Therefore, the result of this model is used for estimation. The plot of estimation is revealed in Figure 40.

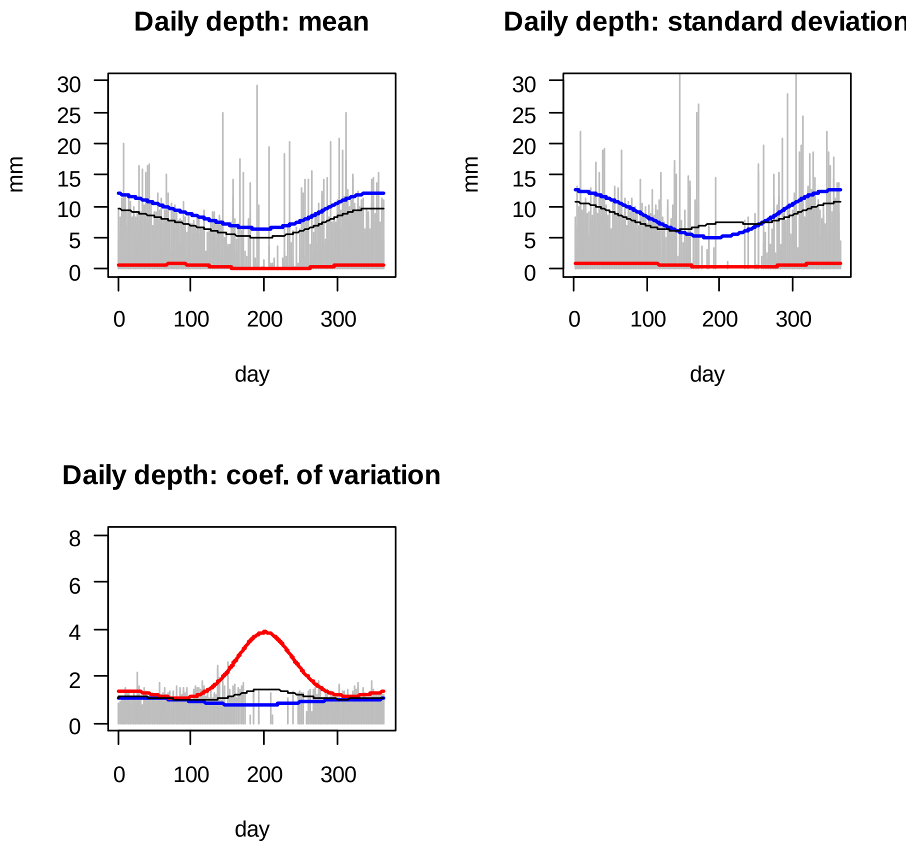

Figure 40 Estimation of total precipitation amount for Aydın_Center station

The graphs of estimation of mean, standard deviation and coefficient of variation for precipitation amount at the Aydın_Center station are displayed in Figure 40. It can be observed the first state (red line) which presents the minimum estimate of precipitation amount, the second state (blue line) which presents the maximum estimate of precipitation amount and also, overall estimate of precipitation amount (dark line) which is between the first state and the second state.

When the mean of precipitation amount is analyzed, it can be seen that overall line is close to the second state at the end of the year. This means that the precipitation amount will be around 10 mm. On the other hand, when coefficient of variation of precipitation amount is analyzed, the overall line is not close to the second state at the end of the year. Especially, the overall line is close to the first state around the 200th of the year. It can be said that the coefficient of variation of precipitation amount will be less than 2 mm in these days. Similarly, an evaluation can be done for standard deviation of precipitation amount. Also, daily predictions can be done by using the estimation plots of mean, standard deviation and coefficient of variation.

4.5 Application of Viterbi Algorithm

The Viterbi algorithm helps us to find the state sequence that best describes the observation sequences among all possible state sequences. In other words, when a sequence of observations and the model parameters (A, B, π) are known, a sequence of optimal states can be obtained by using this algorithm (Rabiner, 1989). It shows the the prediction power of HMMs. A plan is made for events which occur more likely at a future date. For example, a sequence of observation is considered for the next year, and the most probable sequences of states are estimated through the algorithm of Viterbi (Yoon, 2009).

In order to apply the algorithm of Viterbi, the examples of observation sequence are taken from the results of precipitation occurrence which is derived from the Aegean Region.

Observations are used to perform the algorithm of Viterbi to find the corresponding state. Dry day is defined as state “1”, wet day is defined as state “2”. Observation for dry day is defined as “D” and observation for wet day is defined as “W”. Also, the model parameters are required to apply this algorithm. For this reason, initial probability values, transition matrix and emission matrix are taken from “Model 532” described as the best model for Aegean Region.

The transition matrix is:

$$\begin{array}{c} \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{cc} \quad 1 \quad & \quad\quad 2 \quad \end{array} \\ \begin{bmatrix} 0.7569003 & 0.3009495 \\ 0.2439997 & 0.6990505 \end{bmatrix} \end{array}$$

The emission matrix is:

$$\begin{array}{cc} . & Observations \\ States & \begin{array}{cc} D & W \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \end{array}$$

In addition to model parameters, the sequences of observations are generated for "January and April" randomly. Each sequence of observations includes ten values and this number can be increased. The sequence of observations which are generated for the first ten days of "April" is *"”D, "D", "D", "D", "W", "D", "D", "W", "W", "W"* and the prediction of sequence of states is "1", "1", "1", "1", "1", "1", "1", "2", "2", "2". The real sequences of states are estimated by using the observed data set. Finally, results are compared to compute the error due to the model.

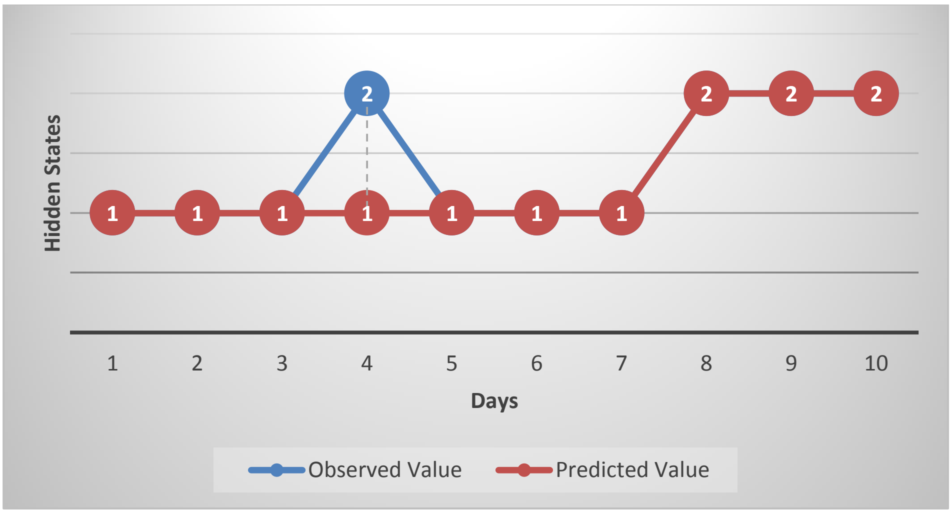

Figure 41 Comparison of sequences of real hidden and predicted states by the model

When drawing the graph, the sequence of observations is represented by the more likely for this sequence of states as shown in Figure 41. When analyzing the sequences of observations and states, it can be seen that there is only one observation which are estimated wrong. Therefore, error is calculated as 10 %. This means that this algorithm gives good prediction for future.

In order to compare the performance of the algorithm of Viterbi, the second sequence of observation is generated for the first ten days of “January” and the most probable states are predicted. The observation sequence of observations is *"W”,"W", "W", "D", "D", "W", "W", "W"*, *"W"*, *"D"* and the prediction of sequence of states is "2", "2", "2", "1", “1", “1", "2", "2", "2", "2").

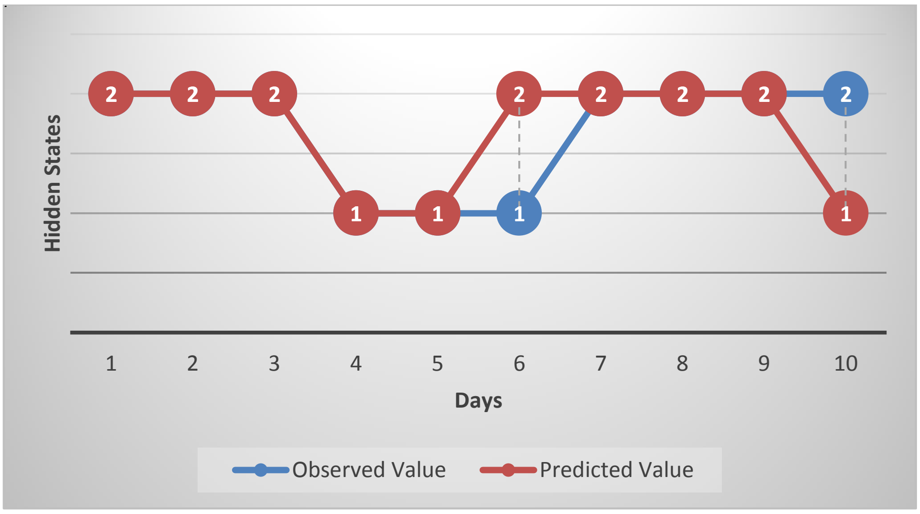

Figure 42 Comparison of sequences of real hidden and predicted states by the model

Figure 42 displays two different sequences of hidden states. One of them shows the predicted sequence which represents the sequence of observed values. The other one shows the predicted sequence which is estimated by the algorithm of Viterbi. It can be seen that there are two mispredicted observation. Hence, the error value can be found as 20 %. It can be said that the model estimates the reality 80%.

At the end, it can be concluded that the performance of algorithm of Viterbi is better at “April” when compared with the result of “January.” Since the number of true prediction is higher. Therefore, it can be said that the most probable path corresponding to a given sequence of observations can be determined by this algorithm easily.

Viterbi algorithm is used to find some sequence of states which is derived in Aegean Region. It provides information for unknown states. In our example, our observations are defined by using past data related to daily total precipitation amount and we want to know the unknown states which lead to occur these observations. In our cases, we have just two states and estimating these states may not seem to be beneficial. However, in many HMM applications there are more states and estimating states are very beneficial. For example, when trying to predict unknown states at four-state HMM which is developed for estimating precipitation amounts, the estimated results of algorithm of Viterbi becomes important.

CHAPTER 5

CONCLUSION AND FUTURE WORK

The main purpose of this study is modelling the daily precipitation occurrence and the amount of total precipitation observed in certain regions of Turkey by using HMMs. HMMs have been successfully applied in precipitation modelling. The main advantage of HMMs for this study is that they provide general information about understanding the probabilistic structure of precipitation and estimate the daily total precipitation amount.

The first chapter of the thesis starts with an introduction. The motivation of study is briefly explained and outline of thesis is given in this chapter. In the second chapter of the thesis, the history of HMMs and precipitation models are explained. In the third chapter, firstly the brief information related to MC is given. Secondly, the definition of HMMs, the description of the daily precipitation HMMs and the parameter estimation method are explained. Thirdly, model selection criteria are described. In the fourth part of the study, the results of the simulation cases defined in previous chapter are presented and discussed. The graphs are prepared from the results obtained from the simulations according to the objectives of the cases. According to the models defined in previous chapter, different types of graphs are presented in order to observe the future prediction related to precipitation.

In this study, homogeneous HMM is applied to daily total precipitation data from three stations in the East Black Sea Region, three stations in the Central Anatolian Region and three stations in the Aegean Region of Turkey. In the first part of the study, two-state and three-state Bernoulli HMM are developed to observe the probability of rainfall occurrence. The performance of HMMs is evaluated by comparing AIC, BIC, MSE, MR and plots which depict observed probability values versus predicted probability values. HMMs which are developed for the East Black Sea Region do not give better results compare to results of the other Regions. Because the MR values are very high and the plots which include observed and predicted probability values do not deviate from "*x=y.*" On the other hand, it is observed that the MR values are small and observed and predicted probability values are similar in the results of HMMs developed for the Aegean Region. This shows that the two-state homogeneous HMM is the most successful for regions that has normal moisture climate like Aegean Region compare to wettest region like East Black Sea and driest region like Central Anatolian.

In the second part of the study, two-state seasonal gamma HMM is developed for one station from the East Black Sea Region, one station from the Central Anatolian Region and one station from the Aegean Region to observe the amount of total precipitation. The performance of HMMs is evaluated by comparing AIC, BIC, MSE and plots of observed versus predicted probability values. HMMs developed for the East Black Sea Region give better results compare to the results of other regions. Because plots of observed versus predicted precipitation amounts scatter around "*x=y*." On the other hand, HMMs developed for the Central Anatolian Region and Aegean Region do not give good results. It is observed that the predicted amount of total precipitation is higher than the observed amount of total precipitation. This is mainly because data includes many zeros, and this leads to false predictions. In the third part of study, we apply Viterbi algorithm to find some sequence of states which are derived for the Aegean Region. It is observed that this algorithm can be used to estimate an unknown situation. In our cases, we have just two states and estimating states may not seem to be beneficial. However, in many HMM applications there are more states and estimating states are very beneficial to observe unknown states.

In conclusion, we see that two-state homogeneous HMM may not be a useful tool to find the probabilities of rainfall occurrence for wettest regions like East Black Sea Region. However it can be useful tool to find the amount of total precipitation these regions for preparing strategies and planning for the unpredicted disaster such as flood.

As the future study, we consider developing nonhomogeneous HMMs by using synoptic atmospheric information such as temperature, solar radiation, and other climatic factors. Because it is known that there are more factors that affect the weather conditions in addition to past data related to total rainfall. When the number of variables which affect the future weather conditions increase to estimate weather condition, the success of prediction of weather condition will increase. Therefore, better results can be observed for the occurrence probability of rainfall predictions and amount of rainfall predictions. Also, the number of states can be increased to observe different weather conditions. Since, we see that homogeneous HMMs are not successful to estimate extreme values. When the number of states for HMMs increases, it can estimate extreme values.

Finally, we believe that HMM is a very useful tool to simulate precipitation. The results obtained from the application of HMM encourage us to find possibility of realizing good local predictions of precipitation. This work would be first phase to make estimations related to precipitation, providing very fast and less costly computations and it gives general weather forecast and information about the state of regions. When a nonhomogeneous HMM is developed, extreme precipitation events could be estimated and an alert system could be constructed. Also, we think that HMMs can help us to understand the probabilistic structure of different application areas.

REFERENCES

Acquah, H.G. (2010). Comparison of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) in Selection of an Asymmetric Price Relationship. *Journal of Development and Agricultural Economics*, 2(1), 1-6.

Ailliot, P., Thompson, C., and Thomson, P. (2009). Space Time Modeling of Precipitation Using a Hidden Markov Model and Censored Gaussian Distributions. *Journal of the Royal Statistical Society,* 58(3), 405-426.

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. *2nd International Symposium on Information Theory, Budapest (Hungary),* 267-281.

Aksoy, B., Purutçuoglu Gazi, V., Batmaz, İ., and Yozgatligil, C. (2013). Modeling the Extreme Precipitation Data: Case Study from Turkey. *26th European Conference on Operational Research, Rome (Italy),* July 1-4, 221.

Albert, P.S. (1991). A Two-State Markov Mixture Model for Time Series of Epileptic Seizure Counts. *Biometrics,* 47, 1371-1381.

Aykan, F., Kartal Koç, E., Yozgatligil, C., İyigün, C., Purutcuoglu Gazi, V., and Batmaz, İ. (2012). Developing Precipitation Models for Continental Central Anatolia, Turkey. *25th European Conference on Operational Research, Vilnius (Lithuania),* July 8-11, 204.

Bahadır, M. (2011). Güneydoğu Anadolu Projesi (GAP) Alanında Sıcaklık ve Yağışın Trend Analizi. *The Journal of International Social Research*, 4(16), 46-59.

Bardossy, A. (1994). Downscaling from GCMs to Local Climate through Stochastic Linkages. *Paoli, G (Ed.), Climate Change, Uncertainty and Decision-Making, Waterloo, ON: Institute for Risk Research*, 33–46.

Bardossy, A., and Plate, E.J. (1992). Space-Time Models for Daily Rainfall Using Atmospheric Circulation Patterns. *Water Resources Research*, 28, 1247-1259.

Bartholy, J., Bogardi, I., and Matyasovszky, I. (1995). Effect of Climate Change on Regional Precipitaion in Lake Balaton Watershed. *Theor. Appl. Climatol.*, 51, 237-250.

Baum, L.E. and Petrie, T. (1966). Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *Ann. Math. Statist.*, 37, 1554–1563.

Baum, L.E. and Eagon, J.A. (1967). An Inequality with Applications to Statistical Estimation for Probabilistic Functions of Markov Processes and to a Model for Ecology. *Bull. Am. Math. Stat.*, 37, 360-363.

Baum, L.E., Petrie, T., Soules, G., and Weiss, N. (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The Annals of Mathematical Statistics*, 41(1), 164–171.

Baum, L.E. (1972). An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes. *Inequalities*, 3, 1-8.

Bellone, E., Hughes, J.P., and Guttorp, P. (2000). A Hidden Markov Model for Downscaling Synoptic Atmospheric Patterns to Precipitation Amounts. *Climate Research*, 15(1), 1–12.

Burke, C. J., and Rosenblatt, M. (1958). A Markovian Function of a Markov Chain. *The Annals of Mathematical Statistics*, 29(4), 1112-1122.

Çaldağ, B., Sırdaş, S., and Şaylan, L. (2003). Trakya Bölgesinde Kuraklık Durumunun Standart Yağış İndeksi ile Belirlenmesi. *III. Atmosfer Bilimleri Sempozyumu Bildirileri*, İstanbul, Turkey, March 19-21, 334-340.

Can, T., and Öz, E. (2009). Marka Tercihlerine ve Tercih Nedenlerine Gizli Markov Modellerinin Uygulanması. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 10(2), 167-186.

Chang, R.W., and Hancock, J.C. (1966). On Receiver Structures for Channels Having Memory. *IEEE Trans. Inform. Theory*, 12, 463–468.

Charles, S. P., Bates, B.C, Whetton, P.H., and Hughes, J.P. (1999a). Validation of Downscaling Models for Changed Climate Conditions: Case Study of Southwestern Australia. *Climate Research*, 12(1), 1–14.

Charles, S. P., Bates, B.C., and Hughes, J.P. (1999b). A Spatiotemporal Model for Downscaling Precipitation Occurrence and Amounts. *Journal of Geophysical Research*, 104(24), 31657–31669.

Charles, S. P., Bates, B.C., Smith, I.N., and Hughes, J.P. (2004). Statistical Downscaling of Daily Precipitation from Observed and Modelled Atmospheric Fields. *Hydrological Processes*, 18, 1373–1394.

Charniak, E. (1993). *Statistical Language Learning*. Cambridge: MIT.

Chung, K.L., and Walsh, J.B. (2005). *Markov Processes, Brownian Motion, and Time Symetry*. United States of America: Springer.

Cinemre, N. (2003). *Yöneylem Araştırması*. İstanbul: Beta.

Coe, R., and Stern, R. D. (1982). Fitting Models to Daily Rainfall. *J. Appl. Meteorol.*, 21, 1024-1031.

Dempster, A. P., Laird, N.M., and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. Roy. Statist. Soc. B*, 39(1), 1–38.

Dunn, K.P. (2004). Occurrence and Quantity of Precipitation can be Modelled Simultaneously. *International Journal of Climatology*, 24, 1231-1239.

Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. (1998). *Biological Sequence Analysis*. Cambridge: Cambridge University Press.

Elfeki, A., and Al-Amri, N. (2011). *International Journal of Water Resources and Arid Environments*, 1(6), 411-416.

Elliott, R.J., Aggoun, L., and Moore, J.B. (1994). *Hidden Markov Models: Estimation and Control*. New York: Springer Verlag.

Everitt, B.S., and Hand, D.J. (1981). *Finite Mixture Distributions*. New York: Chapman & Hall.

Fidan, İ.H. (2011). Doğu Akdeniz Bölgesinde Standardize Yağış İndeksi (SYİ) ile Kuraklık Analizi ve Markov Zinciri Yöntemi Kullanılarak Kurak Olma Olasılıklarının Belirlenmesi. Yüksek Lisans Tezi, Çukurova Üniversitesi, Adana, Türkiye.

Finesso, L. (1990). *Consistent Estimation of the Order for Markov and Hidden Markov Chains*. Ph.D. Dissertation, Univ. Maryland, College Park, USA.

Gabriel, K. R., and Neumann, J. (1992). A Markov Chain Model for Daily Rainfall Occurrences at Tel-Aviv. Q. J. R. *Meterol. Soc.*, 88, 85-90.

Giorgi, F., and Mearns, L.O. (1991). Approaches to Regional Climate Change Simulation: A Review. *Review of Geophysics*, 29, 191-216.

Greene, A.M., Robertson, A.W., and Kirshner, S. (2008). Analysis of Indian Monsoon Daily Rainfall on Subseasonal to Multidecadal Time-Scales Using a Hidden Markov Model. *Quarterly Journal of the Royal Meteorological Society*, 134, 875–887.

Grunwald, G.K., and Jones, R.J. (2000). Markov models for time series with mixed distribution. *Environmetrics*, 11, 327–339.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57, 357-384.

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton Univ. Press.

He, Y., and Kundu, A. (1991). 2-D Shape Classification Using Hidden Markov Models. *IEEE Trans. Pattern Anal. Machine Intell.*, 13, 1172-1184.

Huang, X. D., Ariki, Y., and Jack, M.A. (1990). *Hidden Markov Models for Speech Recognition*. Edinburg, U.K.: Edinburgh University Press.

Huang, X.D., and Jack, M.A. (1989). Semi-continuous Hidden Markov Models for Speech Recognition. *Computer Speech and Language*, 3(3), 239-252.

Hughes, J. P., and Guttorp, P. (1994a). A Class of Stochastic Models for Relating Synoptic Atmospheric Patterns to Regional Hydrologic Phenomena. *Water Resources Research*, 30(5), 1535–1546.

Hughes, J. P., and Guttorp, P. (1994b). Incorporating Spatial Dependence and Atmospheric Data in a Model of Precipitation. *Journal of Applied Meteorology*, 33(12), 1503–1515.

Hughes, J. P., Guttorp, P., and Charles, S.P. (1999). A Non-homogeneous Hidden Markov Model for Precipitation Occurrence. *Journal of the Royal Statistical Society*, 48(1), 15–30.

İnce, H. T., and Weber, G.W. (2005). *Analysis of Bauspar System and Model Based Clustering with Hidden Markov Models*. Term Project in MSc Program "Financial Mathematics – Life Insurance", Institute of Applied Mathematics METU, Ankara, Turkey.

Jelinek, F. (1976). Continuous Speech Recognition by Statistical Methods. *Proc. IEEE*, 64, 532–556.

Jelinek, F., Bahl, L.R., and Mercer, R.L. (1975). Design of a Linguistic Statistical Decoder for Recognition of Continuous Speech. *IEEE Trans. Inform. Theory*, 21(3), 250–256.

Katz, R.W (1977). An Application of Chain-dependent Processes to Meteorology. *Journal of Applied Probability*, 14, 598–603.

Katz, R.W., and Parlange, M.B. (1996). Mixtures of Stochastic Processes: Application to Statistical Downscaling. *Climate Research*, 7, 185–193.

Karabulut, M., and Cosun, F. (2009). Precipitation Trend Analyses in Kahramanmaraş (Kahramaraş İlinde Yağışların Trend Analizi). Coğrafi Bilimler Dergisi, 7(1), 65-83.?

Knab, B., Schliep, A., Steckemetz, B., and Wichern, B. (2003). Model-based Clustering with Hidden Markov Models and Its Application to Financial Time-series Data. *Between Data Science and Applied Data Analysis, Springer*, 561-569.

Koçak, K., and Şen, Z. (1997). Kurak ve Yağışlı Gün Oluşumlarının Markov Zinciri Yaklaşımı ile Uygulamalı İncelenmesi. *Tr. J. of Engineering and Environmental Science,* 22, 479- 487.

Krogh, A., Brown, M., Mian, I. S., Sjolander, K., and Haussler, D. (1994). Hidden Markov Models in Computational Biology: Applications to Protein Modeling. *J. Mol. Biol.,* 235, 1501-1531.

Kutner, M.H., Nachtsheim, C.J., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models*. Newyork: McGraw Hill.

Lall, U., Rajagopalan, B., and Tarboton, D.G. (1996). A Nonparametric Wet/Dry Spell Model for Precipitation Occurence. J. *Royal Stat. Soc.*, 48, 15-30.

Lapidoth, A., and Ziv, J. (1998). On the Universality of the LZ-based Decoding Algorithm. *IEEE Trans. Inform. Theory*, 44, 1746–1755.

Le, N.D., Leroux, B.G., and Puterman, M.L. (1992). Exact Likelihood Evaluation in a Markov Mixture Model for Time Series of Seizure Counts. *Biometrics*, 48, 317-323.

Lee, K.F. (1989a). *Large-vocabulary Speaker-independent Continuous Speech Recognition: The SPHINX System.* Ph.D. Thesis, Department of Computer Science, Carnegie-Mellon University, Pittsburgh, USA.

Lee, K.F. (1989b). *Automatic Speech Recognition. The Development of the SPHINX System*. Boston, MA: Kluwer Academic Publishers.

Lee, K.F. (1990). Context-dependent Phonetic Hidden Markov Models for Speaker Independent Continuous Speech Recognition. *IEEE Trans. Acoust., Speech Signal Processing*, 38(4), 599-609.

Leroux, B.G., and Puterman, M.L. (1992). Maximum-penalized-likelihood Estimation for Independent and Markov-dependent Mixture Models. *Biometrics*, 48, 545-558.

MacDonald, I., and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series Models.* London: Chapman & Hall.

Merhav, N. (1991). Universal Coding with Minimum Probability of Codeword Length Overflow. *IEEE Trans. Inform. Theory*, 37, 556–563.

Moron, V., Robertson, A.W., Ward, M.N., and Ndiaye, O. (2008). Weather Types and Rainfall over Senegal. *Journal of Climate*, 28, 288-307.

Neykov, N.M., Zucchini, W., Neytchev, P., and Hristov, H.P. (2007). Hidden Markov Models for Relating Atmospheric Circulation Patterns to Bulgaria Daily Precipitation Occurrence and Amounts. *4th European Conference on Severe Storms*, Trieste, Italy, September 10 – 14.

Neykov, N., Neytchev, P., Zuchini, W., and Hristov, H. (2008). Stochastic Downscaling of Daily Precipitation over the Territory of Bulgaria by Means of Hidden Markov Models. *3rd International Conference on Water Observation and Information System for Decision Support*, Ohrid, Republic of Macedonia, May 27-31.

Neykov, N., Neytchev, P., and Zucchini, W. (2003). Detecting Precipitation Climate Changes: An Approach Based on a Stochastic Daily Precipitation Model. Pliska Stud. Math. Bulgar., 29, 91-106.

Neykov, N., Neytchev, P., and Zucchini, W. (2014). Stochastic Daily Precipitation Model with a Heavy-tailed Component. *Natural Hazards and Earth System Sciences Discussions*, 2(2), 1223-1258.

Neytchev, P.N, Neykov, N.M., Zucchini, W., and Hristov, H. (2008). Statistical Linkage of Daily Precipitation in Bulgaria to Atmospheric Circulation. XXIV Conference of the Danubian Countries on the Hyrological Forecasting and Hydrological Bases of Water Management. Bled (Slovenia), June 2-4.?

Özdemir, M.A., and Bahadır, M. (2010). Denizli'de Box-Jenkins Tekniği ile Küresel İklim Değişikliği Öngörüleri. *The Journal of International Social Research*, 3(12), 352-362.

Özgürel, M., and Kılıç, M. (2003). İzmir için Geleceğe Yönelik Yağış Olasılıklarının Markov Zinciri Modeliyle Belirlenmesi. *Ege Üniv. Ziraat Fak. Derg.*, 40(3), 105-112.

Öztürk, K. (2002). Küresel İklim Değişikliği ve Türkiye'ye Olası Etkileri. *Gazi Üni. Gazi Eğitim Fakültesi Dergisi*, 22(1), 47-65.

Petrie, T. (1969). Probabilistic Functions of Finite State Markov Chains. *Ann. Math. Statist,*, 40(1), 97–115.

Poritz, A.B. (1988). Hidden Markov Models: A Guided Tour. *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 7–13.

Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2), 257–286.

Raviv, J. (1967). Decision Making in Markov Chains Applied to the Problem of Pattern Recognition. *IEEE Trans. Inform. Theory*, 3, 536–551.

Redner, R.A., and Walker, H.F. (1984). Mixture Densities, Maximum Likelihood and the EM Algorithm. *SIAM Rev.* 26(2), 195–239.

Robert, C.P., Celeux, G., and Diebolt, J. (1993). Bayesian Estimation of Hidden Markov Chains: Astochastic Implementation. *Statist. Probab. Lett.*, 16, 77–83.

Robertson, A. W., Kirshner, S. and Smyth, P. (2004). Downscaling of Daily Rainfall Occurrence over Northeast Brazil Using a Hidden Markov Model. *Journal of Climate*, 17(22), 4407–4424.

Robertson, A. W., Kirshner, S., Smyth, P., Charles, S.P., and Bates, B.C. (2005). Subseasonal-to-interdecadal Variability of the Australian Monsoon over North Queensland. *Quarterly Journal of the Royal Meteorological Society*, 132(615), 519–542.

Robertson, A. W., Ines, A.V.M., and Hansen, J.W. (2007). Downscaling of Seasonal Precipitation for Crop Simulation. *Journal of Applied Meteorology and Climatology*, 46, 677–693.

Robertson, A. W., Moron, V., and Swarinoto, Y. (2009). Seasonal Predictability of Daily Rainfall Statistics over Indramayu District, Indonesia. *International Journal of Climatology*, 29, 1449–1462.

Ryden, T. (1995). Estimating the Order of Hidden Markov Models. *Statistics*, 26, 345–354.

Ryden, T., Terasvirta, T., and Asbrink, S. (1998). Stylized Facts of Daily Return Series and the Hidden Markov Model of Absolute Return. *Journal of Applied Econometrics*, 13, 217-244.

Schliep, A., Schönhuth, A., and Steinhoff, C. (2003). Using Hidden Markov Models to Analyze Gene Expression Time Course Data. *Bioinformatics*, 19(1), 255–263.

Schrodt, P. A. (2000). Pattern Recognition of International Crises using Hidden Markov Models. *Diana Richards (ed.) Political Complexity: Nonlinear Models of Politics. Ann Arbor: University of Michigan Press*, 296-328.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6(2), 461–464.

Scott, S. L. (2002). Bayesian Methods for Hidden Markov Models: Recursive Computing in the 21st Century. *Journal of the American Statistical Association*, 97, 337-351.

Sharma, T.C. (1996). Simulation of the Kenyan Longest Dry and Wet Spells and the Largest Rain-sums using a Markov Model. *Journal of Hydrology*, 178, 55-67.

Stern, R. D., and Coe, R. (1984). A Model Fitting Analysis of Daily Rainfall Data. *Journal of the Royal Statistical Society*, 147, 1-34.

Şahin, A., and Şen, Z. (2001). First Order Markov Chain Approach Wind Speed Modeling. *Wind Engineering and Industrial Aerodynamic*, 89(3-4), 263-270.

Tijms, H.C. (2003). *A First Course in Stochastic Models.* England: John Wiley&Sons.

Titterington, D.M., Smith, A.F.M., and Makov, U.E. (1985). *Statistical Analysis of Finite Mixtures Distributions*. New York: Wiley.

Tonkaz, T. (2008). Birinci Dereceden Markov Zinciri ile Güneydoğu Anadolu Projesi Alanında Kuraklık Analizi. *HR.Ü.Z. Dergisi*, 12(1), 13-18.

Topçuoğlu, K., Pamuk, G., and Özgürel, M. (2005). Gediz Havzası Yağışlarının Stokastik Modellenmesi. *Ege Üniv.Ziraat.Fak.Derg.*, 42(3), 89-97.

Topçuoğlu, K. (2010). Manisa İstasyonu Yağışlarının Stokastik Modellenmesi. *Soma Meslek Yüksek Okulu Teknik Bilimler Dergisi*, 2(13), 55-62.

Türkeş, M. (1996). Spatial and Temporal Analysis of Annual Rainfall Variations in Turkey. *International Journal of Climatology*, 16, 1057-1076.

Türkeş, M., Sümer, U.M., and Demir, İ. (2002). Re-Evaluation of Trends and Changes Mean, Maximum and Minimum Temperature of Turkey for the period 1929-1999. *International Journal of Climatology*, 22, 947-977.

Türkeş, M., Koç, T., and Sarış, F. (2008). Türkiye'nin Yağış Toplamı ve Yoğunluğu Dizilerindeki Değişikliklerin ve Eğilimlerin Zamansal ve Alansal Çözümlenmesi. *Coğrafi Bilimler Dergisi*, 3, 57-73.

Wichern, B. (2001). Hidden Markov Model Analyze and Simulation. Ph.D. Thesis, Cologne University, Germany.

Wilby, R.L., Hassan, H., and Hanaki, K. (1998). Statistical Downscaling of Hydrometeorological Variables using General Circulation Model Output. *Journal of Hydrology*, 205, 1–19.

Won, K.J., Prügel-Bennett, A., and Krogh, A. (2004). Training HMM Structure with Genetic Algorithm for Biological Sequence Analysis. *Bioinformatics*, 20(18), 3613-3619.

Woolhiser, D.A. (1992). Modeling Daily Precipitation Progress and Problems. *Walden, A.T., Guttorp, P. (Eds.), Statistics in the Environmental and Earth Sciences, John Wiley*, 71-89.

Yada, T., Ishikawa, M., Tanaka, H., and Asai, K. (1994). DNA Sequence Analysis using Hidden Markov Model and Genetic Algorithm. *Genome Informatics*, 5, 178-179.

Yada, T., and Hirosawa, M. (1996). Gene Recognition in Cyanobacterium Genomic Sequence Data using the Hidden Markov Model. *Proceedings of the Fourth International Conference on Intelligent Systems for Molecular Biology*, Menlo Park, California, June 12-15, 252–260.

Yada, T. (1998). *Stochastic Models Representing DNA Sequence Data - Construction Algorithms and Their Applications to Prediction of Gene Structure and Function*. Ph.D. Thesis, University of Tokyo.

Yeğnidemir, M.K. (2005). *İç Anadolu Bölgesinin Standartlaştırılmış Yağış İndeksi Metodu ile Kuraklık Analizi*. Master Thesis, Kırıkkale, Turkey.

Yevjevich, V. (1991). Tendencies in Hydrology Research and Its Applications for 21st Century. *Water Resources Management*, 5(1), 1-23.

Yoon, B.J. (2009). Hidden Markov Models and Their Applications in Biological Sequence Analysis. *Curr Genomics*, 10(6), 402-415.

Young, K.C. (1994). A Multivariate Chain Model for Simulating Climatic Parameters from Daily Data. *Journal of Applied Meteorology*, 33, 661-671.

Zucchini, W., and Guttorp, P. (1991). A Hidden Markov Model for Space-time Precipitation. *Water Research,* 27(8), 1917-1923.

Zucchini, W., and MacDonald, I.L. (2009). *Hidden Markov Models for Time Series: An Introduction using R*. London: Chapman & Hall.

,

APPENDIX A

ESTIMATION PLOTS

A.1. Estimation Plots of Probability of Precipitation Occurence

Estimation Plots For East Black Sea Region

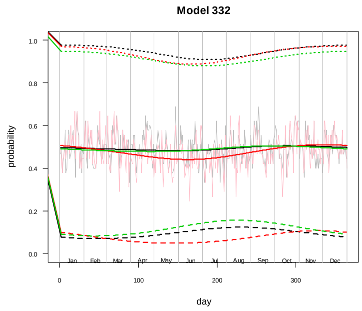

Figure 43 The estimated probability of precipitation occurrence for stations from East Black Sea Region (Model 332)

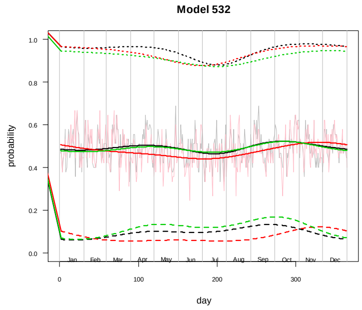

Figure 44 The estimated probability of precipitation occurrence for stations from East Black Sea Region (Model 532)

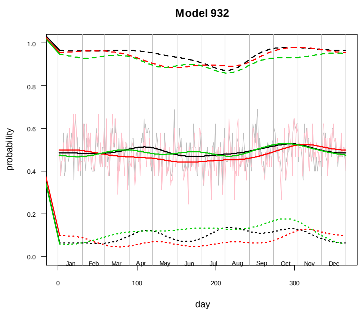

Figure 45 The estimated probability of precipitation occurrence for stations from East Black Sea Region (Model 932)

Observed Probability vs Predicted Probability Plots for East Black Sea Region

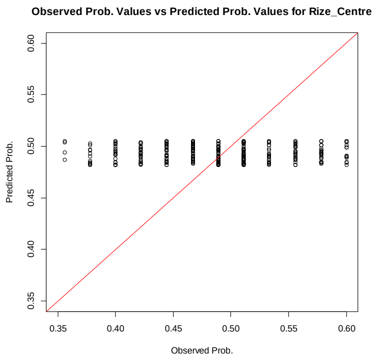

Figure 46 The observed versus predicted probability of precipitation occurrence for Rize_Center station (Model 332)

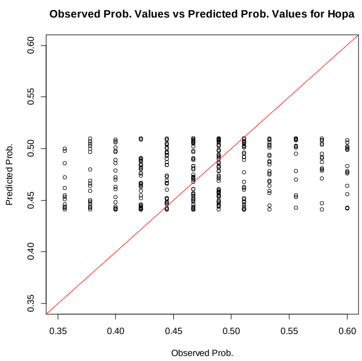

Figure 47 The observed versus predicted probability of precipitation occurrence for Hopa station (Model 332)

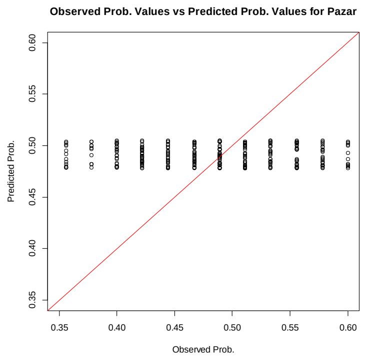

Figure 48 The observed versus predicted probability of precipitation occurrence for Pazar station (Model 332)

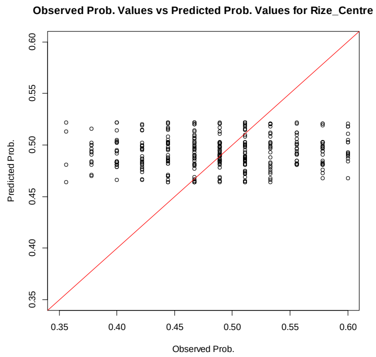

Figure 49 The observed versus predicted probability of precipitation occurrence for Rize_Center station (Model 532)

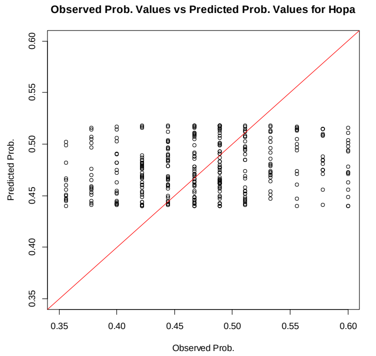

Figure 50 The observed versus predicted probability of precipitation occurrence for Hopa station (Model 532)

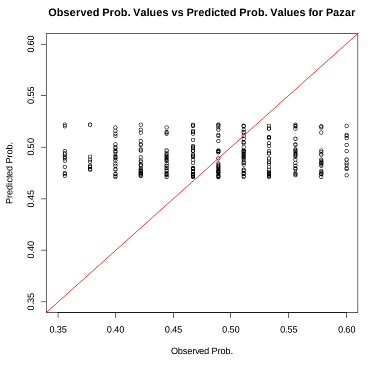

Figure 51 The observed versus predicted probability of precipitation occurrence for Pazar station (Model 532)

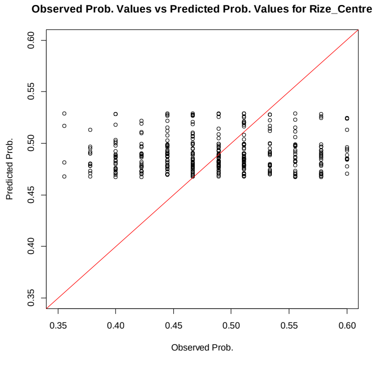

Figure 52 The observed versus predicted probability of precipitation occurrence for Rize_Center station (Model 932)

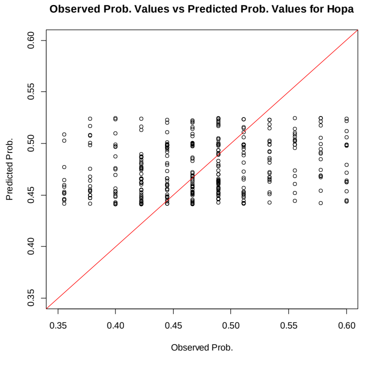

Figure 53 The observed versus predicted probability of precipitation occurrence for Hopa station (Model 932)

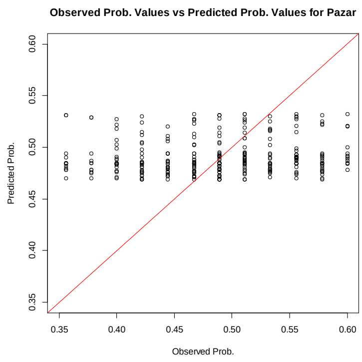

Figure 54 The observed versus predicted probability of precipitation occurrence for Pazar station (Model 932)

Estimation Plots for Central Anatolian Region

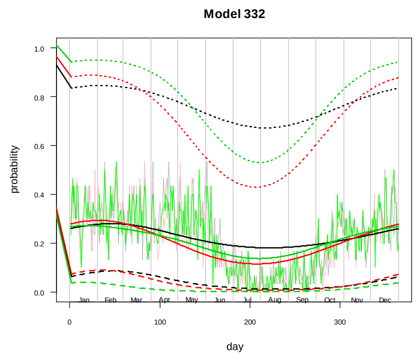

Figure 55 The estimated probability of precipitation occurrence for stations from Central Anatolian Region (Model 332)

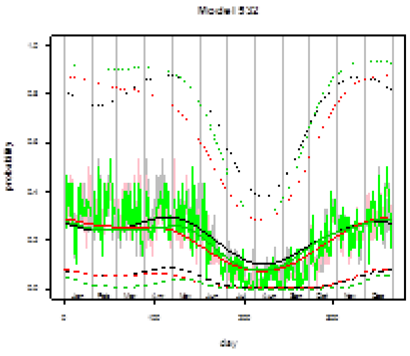

Figure 56 The estimated probability of precipitation occurrence for stations from Central Anatolian Region (Model 532)

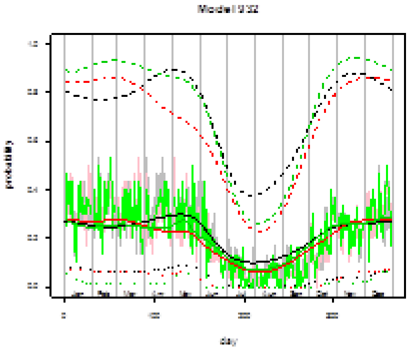

Figure 57 The estimated probability of precipitation occurrence for stations from Central Anatolian Region (Model 932)

Observed Probability vs Predicted Probability Plots for Central Anatolian Region

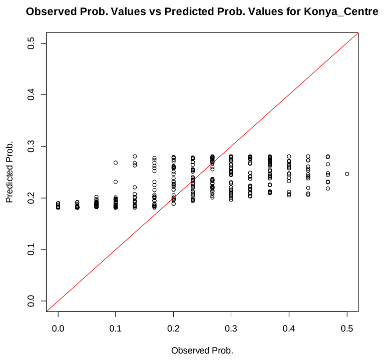

Figure 58 The observed versus predicted probability of precipitation occurrence for Konya_Center station (Model 332)

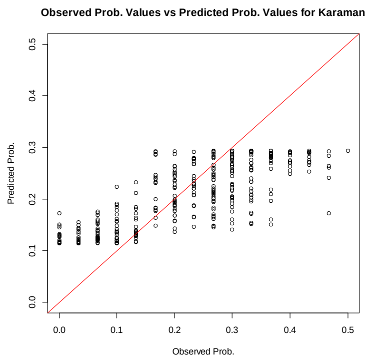

Figure 59 The observed versus predicted probability of precipitation occurrence for Karaman station (Model 332)

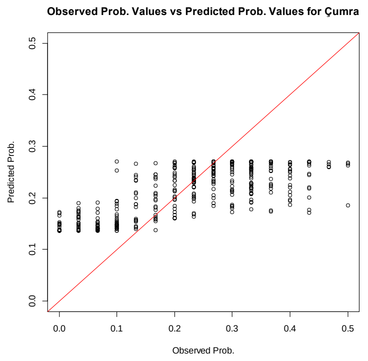

Figure 60 The observed versus predicted probability of precipitation occurrence for Çumra station (Model 332

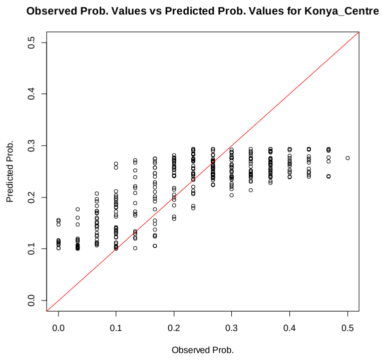

Figure 61 The observed versus predicted probability of precipitation occurrence for Konya_Center station (Model 532)

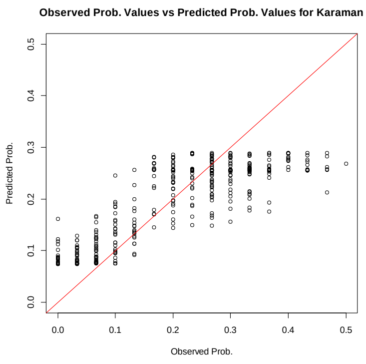

Figure 62 The observed versus predicted probability of precipitation occurrence for Karaman station (Model 532)

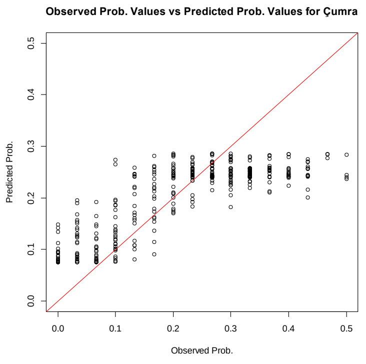

Figure 63 The observed versus predicted probability of precipitation occurrence for Çumra station (Model 532)

Estimation Plots for Aegean Region

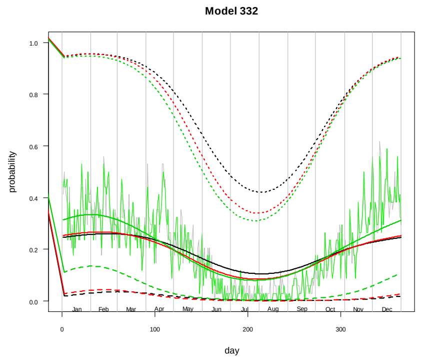

Figure 64 The estimated probability of precipitation occurrence for stations from Aegean Region (Model 332)

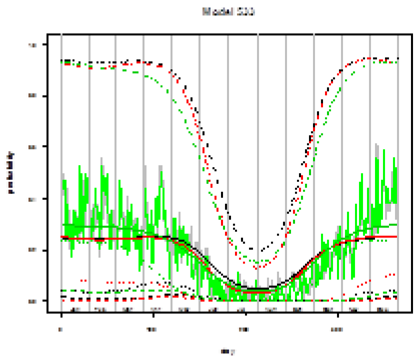

Figure 65 The estimated probability of precipitation occurrence for stations from Aegean Region (Model 533)

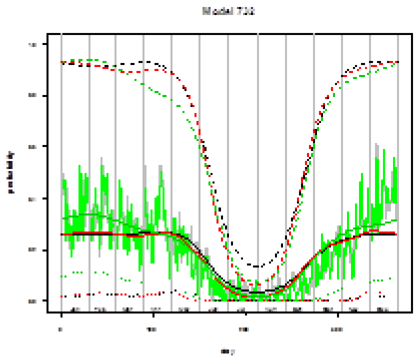

Figure 66 The estimated probability of precipitation occurrence for stations from Aegean Region (Model 732)

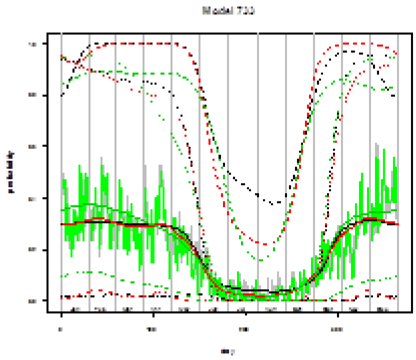

Figure 67 The estimated probability of precipitation occurrence for stations from Aegean Region (Model 733)

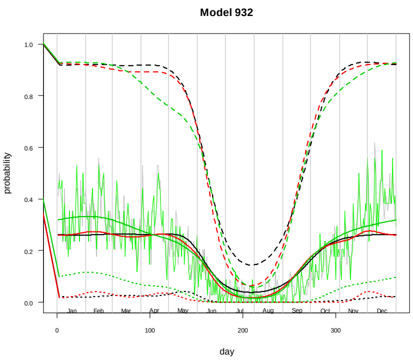

Figure 68 The estimated probability of precipitation occurrence for stations from Aegean Region (Model 932)

Observed Probability Values vs Predicted Probability Values Plots for Aegean Region

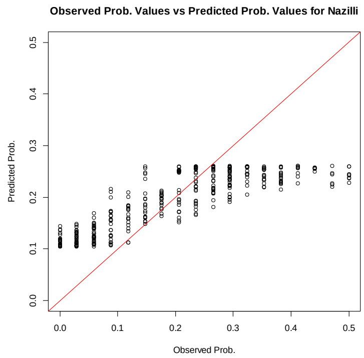

Figure 69 The observed versus predicted probability of precipitation occurrence for Nazilli station (Model 332)

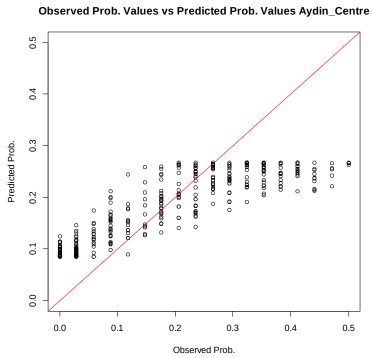

Figure 70 The observed versus predicted probability of precipitation occurrence for Aydın_Center station (Model 332)

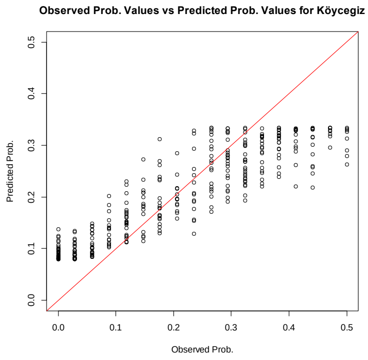

Figure 71 The observed versus predicted probability of precipitation occurrence for Köyceğiz station (Model 332)

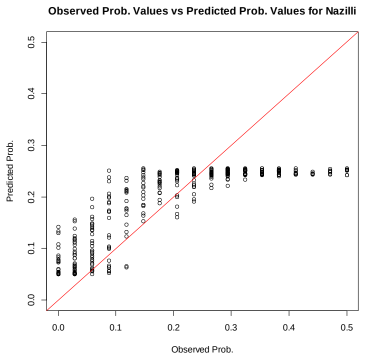

Figure 72 The observed versus predicted probability of precipitation occurrence for Nazilli station (Model 532)

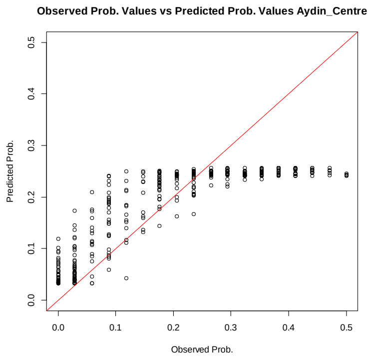

Figure 73 The observed versus predicted probability of precipitation occurrence for Aydın_Center station (Model 532)

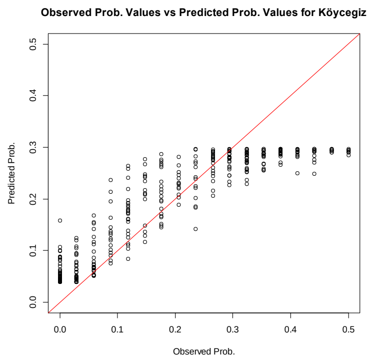

Figure 74 The observed versus predicted probability of precipitation occurrence for Köyceğiz station (Model 532)

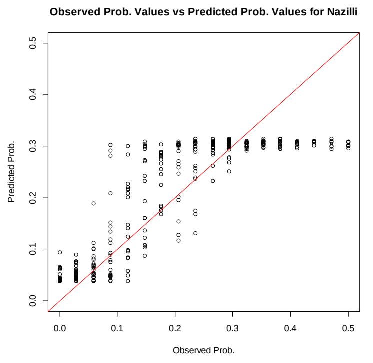

Figure 75 The observed versus predicted probability of precipitation occurrence for Nazilli station (Model 732)

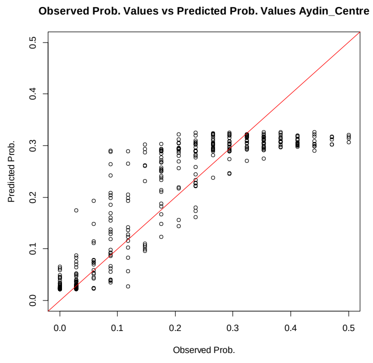

Figure 76 The observed versus predicted probability of precipitation occurrence for Aydın_Center station (Model 732)

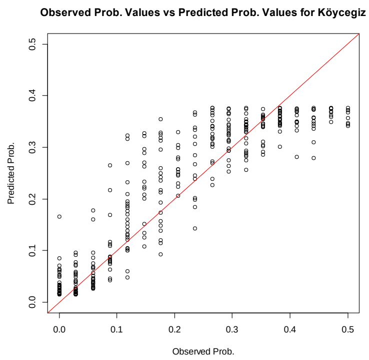

Figure 77 The observed versus predicted probability of precipitation occurrence for Köyceğiz station (Model 732)

A.2. Estimation Plots of Precipitation Amount

Estimation Plots of Precipitation Amount for East Black Sea Region

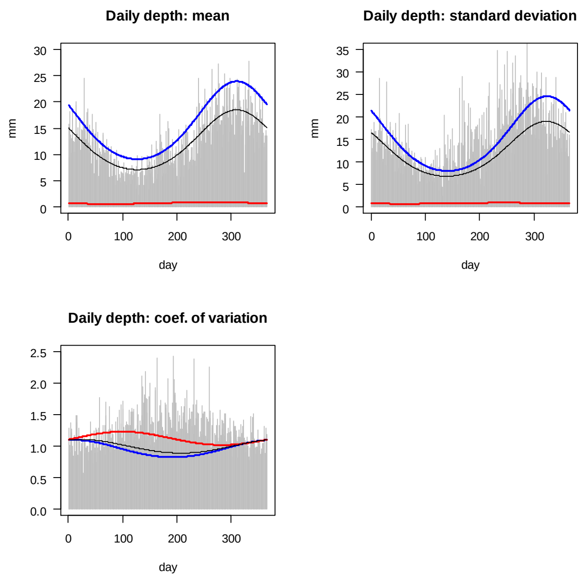

Figure 78 Estimation of mean, standard deviation and coefficient of variation of total precipitation amount for Rize_Center station (Model 1)

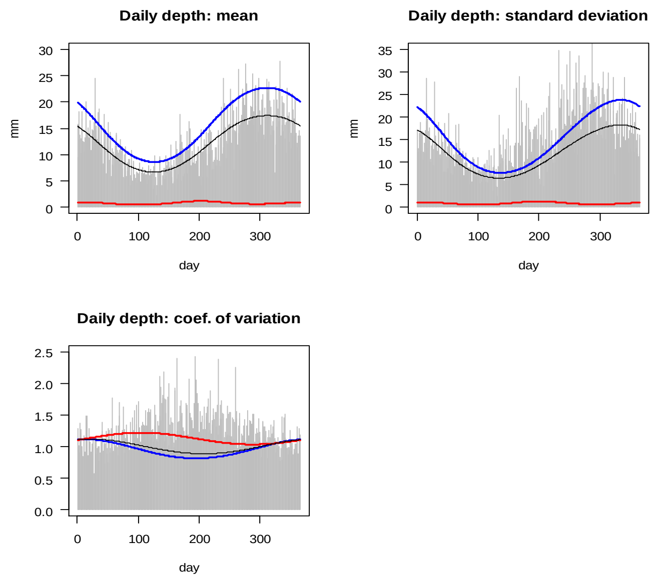

Figure 79 Estimation of mean, standard deviation and coefficient of variation of total precipitation amount for Rize_Center station (Model 2)

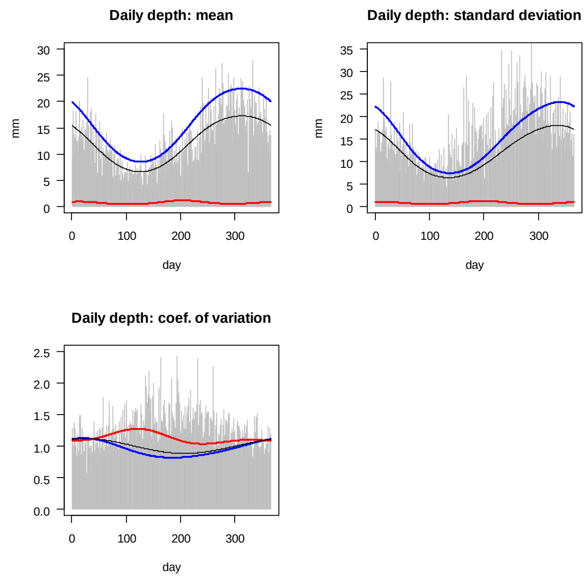

Figure 80 Estimation of mean, standard deviation and coefficient of variation of total precipitation amount for Rize_Center station (Model 3)

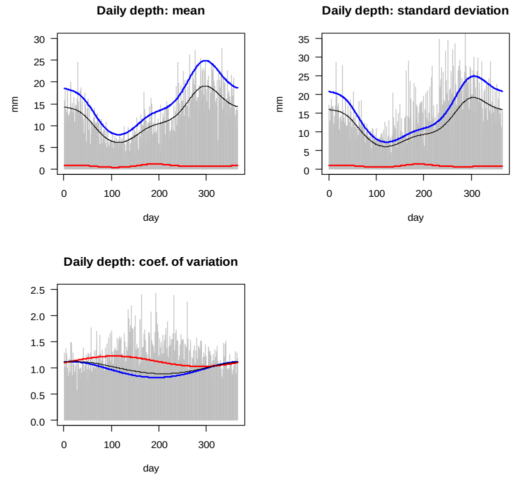

Figure 81 Estimation of mean, standard deviation and coefficient of variation of total precipitation amount for Rize_Center station (Model 4)

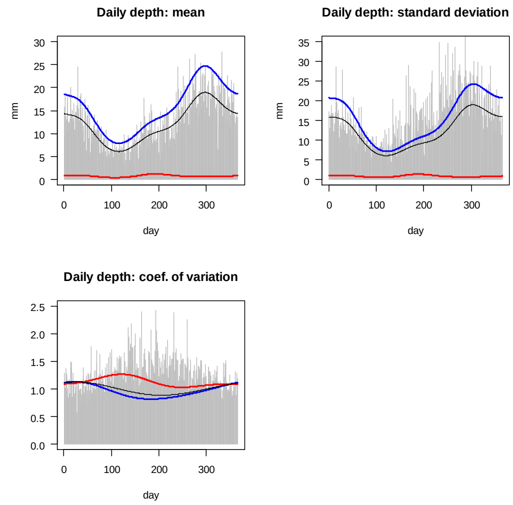

Figure 82 Estimation of mean, standard deviation and coefficient of variation of total precipitation amount for Rize_Center station (Model 5)

Observed Probability vs Predicted Amount of Precipitation Plots for East Black Sea Region

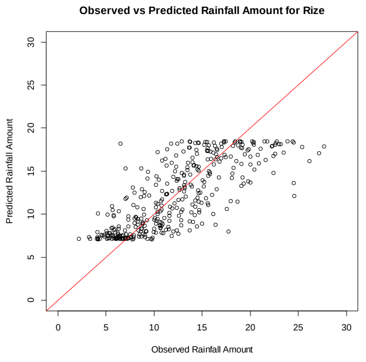

Figure 83 The observed versus predicted values of precipitation amount for Rize_Center station (Model 1)

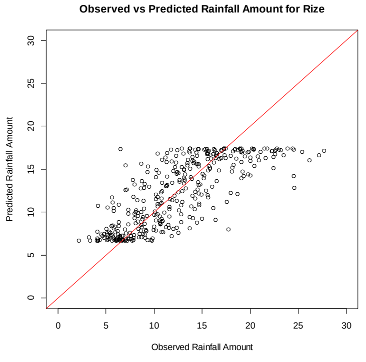

Figure 84 The observed versus predicted values of precipitation amount for Rize_Center station (Model 2)

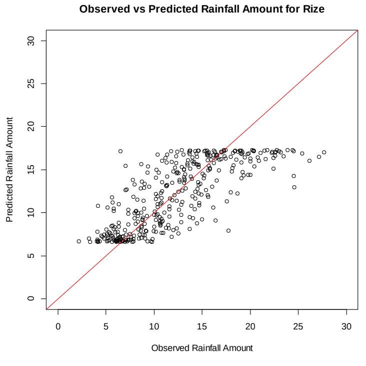

Figure 85 The observed versus predicted values of precipitation amount for Rize_Center station (Model 3)

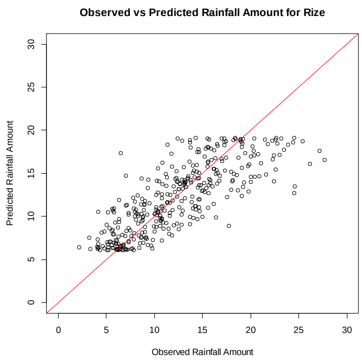

Figure 86 The observed versus predicted values of precipitation amount for Rize_Center station (Model 4)

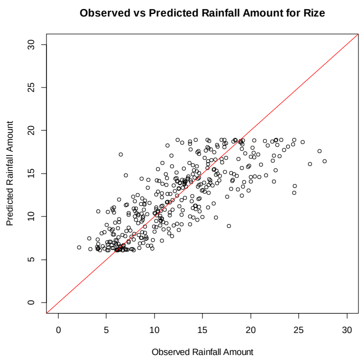

Figure 87 The observed versus predicted values of precipitation amount for Rize_Center station (Model 5)

Estimation Plots of Precipitation Amount for Central Anatolian Region

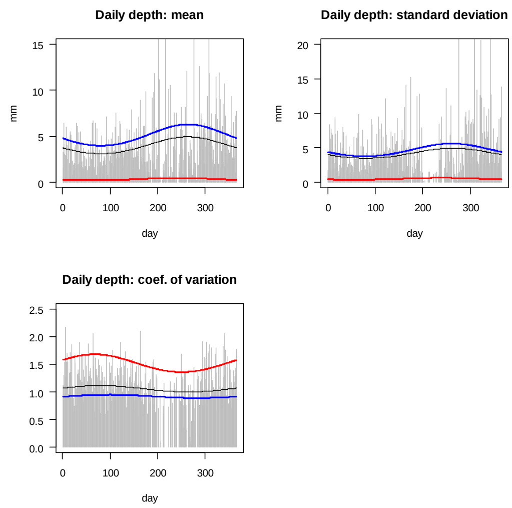

Figure 88 Estimation of mean, standard deviation and coefficient of variation of total precipitation amount for Konya_Center station (Model 1)

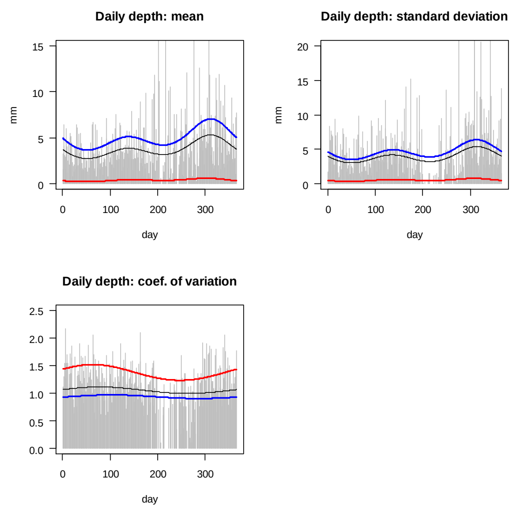

Figure 89 Estimation of mean, standard deviation and coefficient of variation of total precipitation amount for Konya_Center station (Model 2)

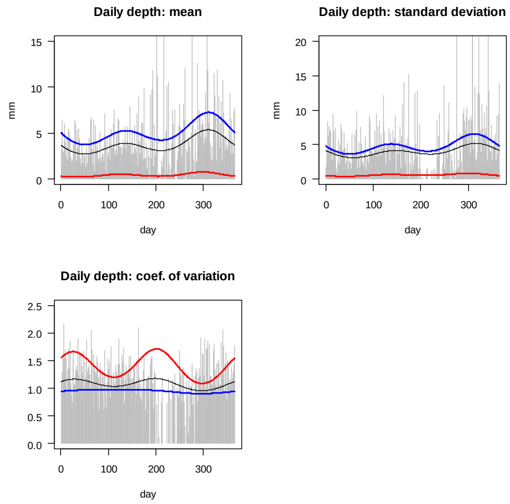

Figure 90 Estimation of mean, standard deviation and coefficient of variation of total precipitation amount for Konya_Center station (Model 3)

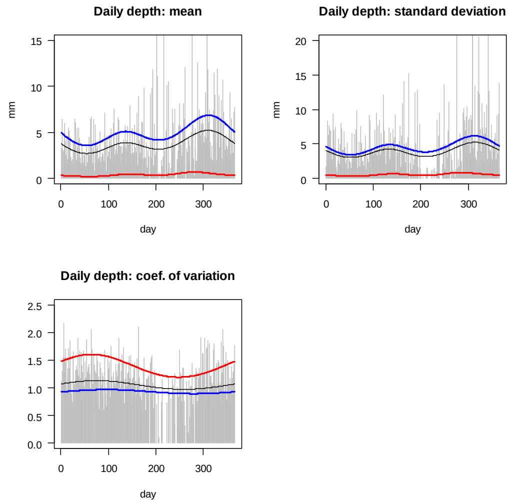

Figure 91 Estimation of mean, standard deviation and coefficient of variation of total precipitation amount for Konya_Center station (Model 4)

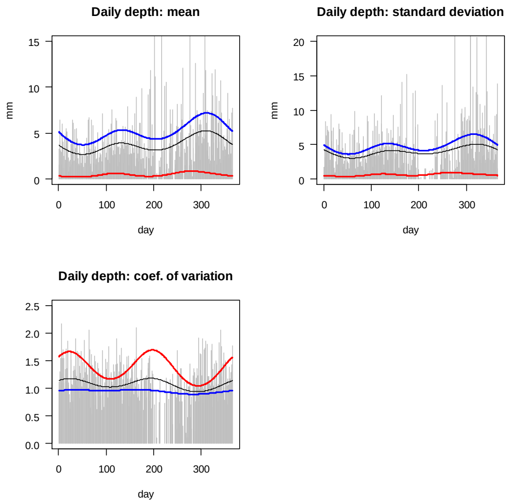

Figure 92 Estimation of mean, standard deviation and coefficient of variation of total precipitation amount for Konya_Center station (Model 5)

Observed Probability vs Predicted Amount of Precipitation Plots for Central Anatolian Region

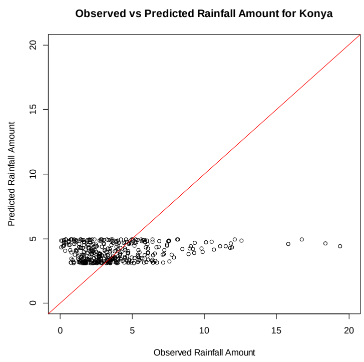

Figure 93 The observed versus predicted values of precipitation amount for Konya_Center station (Model 1)

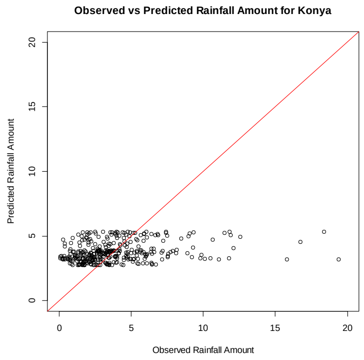

Figure 94 The observed versus predicted values of precipitation amount for Konya_Center station (Model 2)

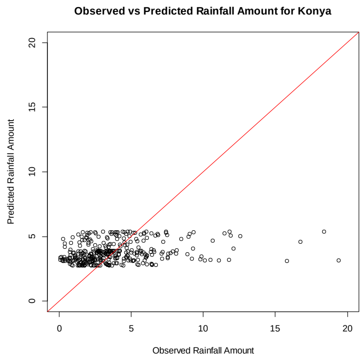

Figure 95 The observed versus predicted values of precipitation amount for Konya_Center station (Model 3)

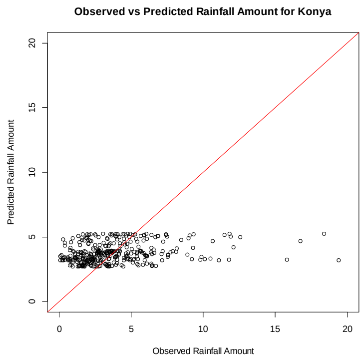

Figure 96 The observed versus predicted values of precipitation amount for Konya_Center station (Model 4)

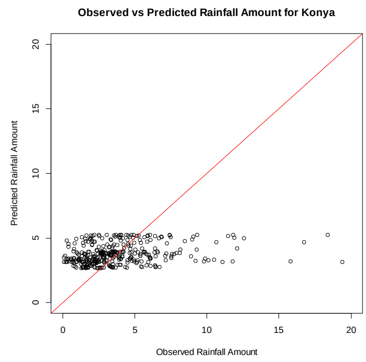

Figure 97 The observed versus predicted values of precipitation amount for Konya_Center station (Model 5)

Estimation Plots of Precipitation Amount for Aegean Region

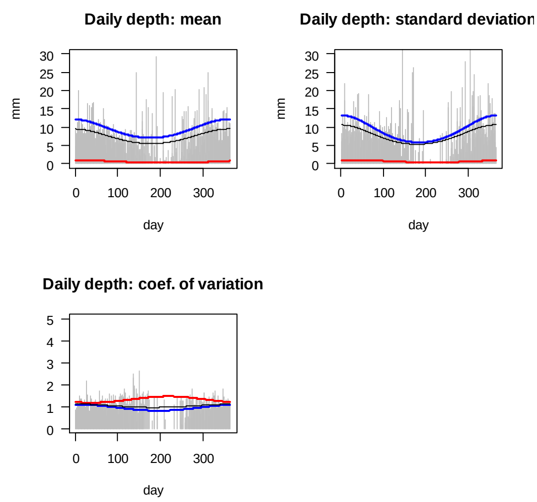

Figure 98 Estimation of mean, standard deviation and coefficient of variation of total precipitation amount for aydın_Center station (Model 1)

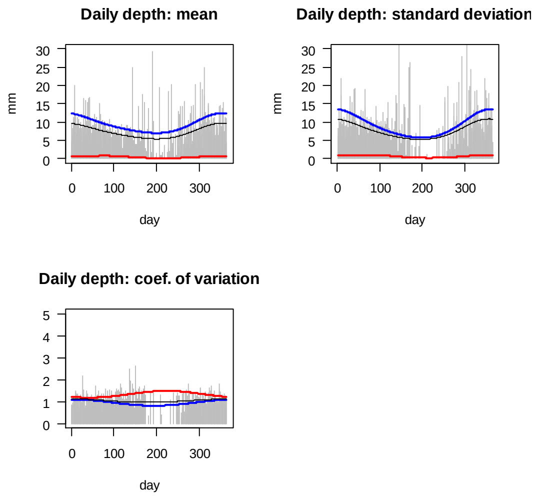

Figure 99 Estimation of mean, standard deviation and coefficient of variation of total precipitation amount for Aydın_Center station (Model 2)

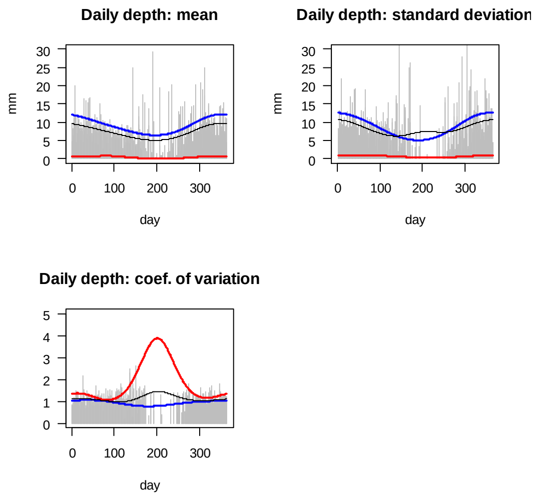

Figure 100 Estimation of mean, standard deviation and coefficient of variation of total precipitation amount for Aydın_Center station (Model 3)

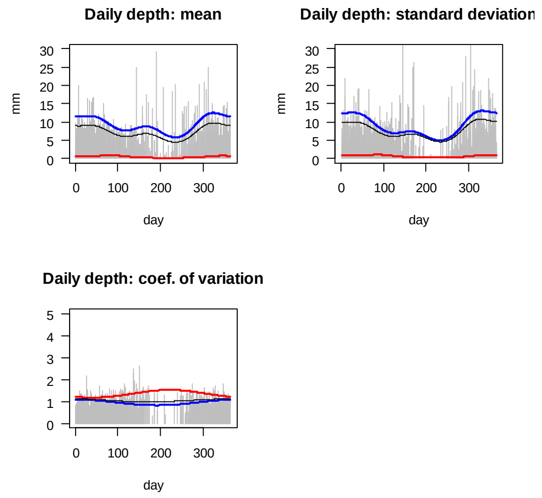

Figure 101 Estimation of mean, standard deviation and coefficient of variation of total precipitation amount for Aydın_Center station (Model 4)

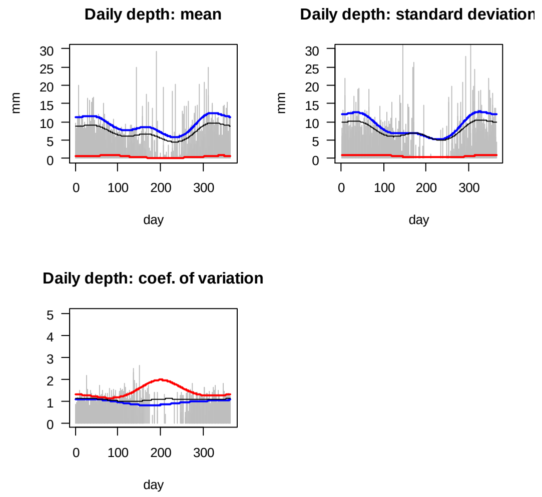

Figure 102 Estimation of mean, standard deviation and coefficient of variation of total precipitation amount for Aydın_Center station (Model 5)

Observed Probability vs Predicted Amount of Precipitation Plots for Aegean Region

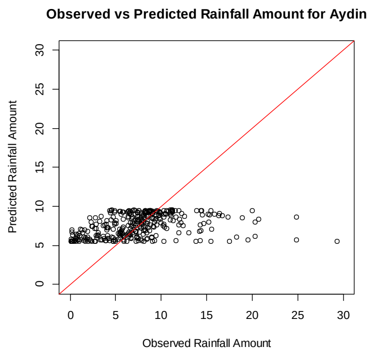

Figure 103 The observed versus predicted values of precipitation amount for Aydın_Center station (Model 1)

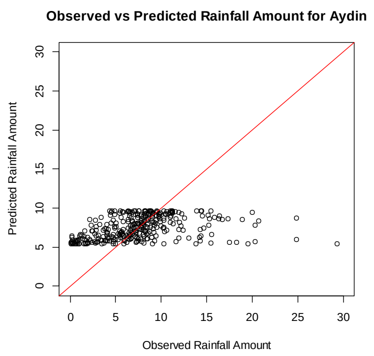

Figure 104 The observed versus predicted values of precipitation amount for Aydın_Center station(Model 2)

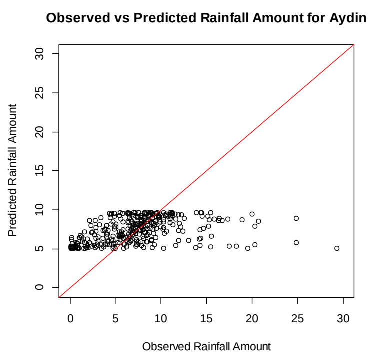

Figure 105 The observed versus predicted values of precipitation amount for Aydın_Center station (Model 3)

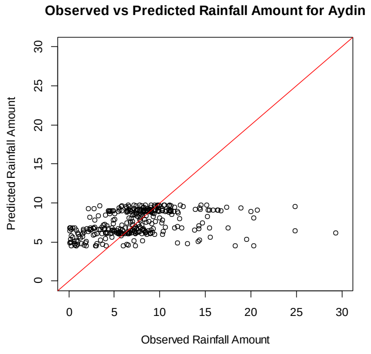

Figure 106 The observed versus predicted values of precipitation amount for Aydın_Center station (Model 4)

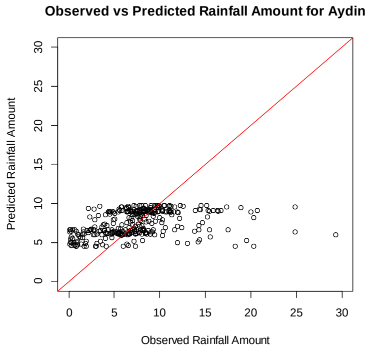

Figure 107 The observed versus predicted values of precipitation amount for Aydın_Center station (Model 5)

APPENDIX B

R CODES

B.1. R Codes for HMMs to Find Probability of Precipitation Occurrence

The following R codes are used to find probability of precipitation and estimation plots for each region seperately.

```
# =======Read and arrange data

chirpan<- read.csv("C:\\Users\\Lab_User\\Desktop\\17040.csv")

plovdiv<- read.csv("C:\\Users\\Lab_User\\Desktop\\17628.csv")

klovpan<-read.csv("C:\\Users\\Lab_User\\Desktop\\17042.csv")


dc<-unlist(t(chirpan))

dp<-unlist(t(plovdiv))

pc<-unlist(t(klovpan))


oc<-as.integer(dc>0)

op<-as.integer(dp>0)

ok<-as.integer(pc>0)


ocm<-t(matrix(oc,nrow=365))

opm<-t(matrix(op,nrow=365))
```

```
okm<-t(matrix(ok,nrow=365))

mean.c<-apply(ocm,2,mean)
mean.p<-apply(opm,2,mean)
mean.k<-apply(okm,2,mean)

plot(mean.c,type="l",col="blue",ylim=c(0,1),xlab="day",ylab="rel. freq.", main="Relative
frequency of wet days")
lines(mean.p,col="red")
lines(mean.k,col="green2")

rain<-cbind(oc,op,ok)
rm(dc,dp,pc,oc,op,ok,ocm,opm,okm)
dev.off()
#=====Descriptive statistics
summary(chirpan)
var(chirpan)
sqrt(var(chirpan))
summary(plovdiv)
var(plovdiv)
sqrt(var(plovdiv))
summary(klovpan)
var(klovpan)
sqrt(var(klovpan))
```

```
# =================================Set up seasonal factors
x0  <- rep(1,365)
x1s <- sin(1*2*pi*(1:365)/365);x1c <- cos(1*2*pi*(1:365)/365)
x2s <- sin(2*2*pi*(1:365)/365);x2c <- cos(2*2*pi*(1:365)/365)
x3s <- sin(3*2*pi*(1:365)/365);x3c <- cos(3*2*pi*(1:365)/365)
x4s <- sin(4*2*pi*(1:365)/365);x4c <- cos(4*2*pi*(1:365)/365)

X<-cbind(x0, x1s, x1c, x2s, x2c, x3s,x3c, x4s, x4c)
rm(x0,x1s,x1c, x2s,x2c, x3s,x3c, x4s, x4c)

# ================= Transform natural parameters to working parameters
multivariate.seasonal.bernoulli.HMM.pn2pw <- function(nfact,nsite,m,pipars,gamma)
{
 tpipars  <- as.vector(pipars)
 tgamma  <- NULL
 if(m>1)
   {
   foo   <- log(gamma/diag(gamma))
   tgamma<- as.vector(foo[!diag(m)])
   }
 parvect <- c(tpipars,tgamma)
 return(parvect)
}
```

```
# ================== Transform  working parameters to natural parameters

multivariate.seasonal.bernoulli.HMM.pw2pn <- function(nfact,nsite,m,parvect)

{

 npiterms<-nfact*nsite*m

 pipars  <- array(parvect[1:npiterms],dim=c(nfact,nsite,m))

 gamma<- diag(m)

 if(m>1)

   {

   gamma[!gamma] <- exp(parvect[(npiterms+1):length(parvect)])

   gamma        <- gamma/apply(gamma,1,sum)

   }

 delta  <- solve(t(diag(m)-gamma+1),rep(1,m))

 list(pipars=pipars,gamma=gamma,delta=delta)

}

# ===================== Function to compute the likelihood

multivariate.seasonal.bernoulli.HMM.mllk<-
function(parvect,rain,nfact,nsite,m,X,details=TRUE,...)

{

 n     <-length(rain[,1])

 pir    <-array(NA,dim=c(365,nsite,m))

 pn     <-multivariate.seasonal.bernoulli.HMM.pw2pn(nfact,nsite,m,parvect)

for (im in 1:m){

 for (isite in 1:nsite){

  pir[,isite,im]<-X[,1:nfact]%*%as.matrix(pn$pipars[,isite,im])
```

```
 }
}
pir<-exp(pir)/(1+exp(pir))
lscale <-0
foo<-pn$delta
j<-0
for (i in 1:n)
{
 j<-j+1; if(j==366) j<-1
 prob=rep(1,m)
 for (isite in 1:nsite){
  if (!is.na(rain[i,isite])) prob=prob*pir[j,isite,]^rain[i,isite]*(1-pir[j,isite,])^(1-rain[i,isite])
 }
foo<-foo%*%pn$gamma*prob
sumfoo<-sum(foo)
lscale<-lscale+log(sumfoo)
foo<-foo/sumfoo
}
mllk<- -lscale
if(details) cat(paste("-log(likelihood) =",mllk,"\n"))
return(mllk)
}
```

```
# ===============Function to maximize the likelihood

multivariate.seasonal.bernoulli.HMM.mle<-
function(rain,nfact,nsite,m,pipars0,gamma0,X=X,details=TRUE...)

{

n       <-length(rain[,1])

parvect0 <-multivariate.seasonal.bernoulli.HMM.pn2pw(nfact,nsite,m,pipars0,gamma0)

mod     <-
nlm(multivariate.seasonal.bernoulli.HMM.mllk,parvect0,rain=rain,nfact=nfact,nsite=nsite
,m=m,X=X,details=details,iterlim = 1000)

mllk    <-mod$minimum

pn      <-multivariate.seasonal.bernoulli.HMM.pw2pn(nfact,nsite,m,mod$estimate)

pir     <-array(NA,dim=c(365,nsite,m))

for (im in 1:m){for (isite in 1:nsite)  pir[,isite,im]<-
X[,1:nfact]%*%as.matrix(pn$pipars[,isite,im])}

pir<-exp(pir)/(1+exp(pir))

np      <-length(parvect0[13:14])

AIC     <-2*(mllk+np)

BIC     <-2*mllk+np*log(n)

return(list(nfact=nfact,nsite=nsite,m=m,pir=pir,gamma=pn$gamma,delta=pn$delta,pipar
s=pn$pipars,code=mod$code,mllk=mllk,AIC=AIC,BIC=BIC))

}

# ================= Function to plot & summarize model

multivariate.seasonal.bernoulli.HMM.plot<-function(mod,pdf=FALSE){

modname    <-paste("Model ",mod$nfact,mod$nsite,mod$m,sep="")

mpir      <-matrix(0,ncol=mod$nsite,nrow=365)

nmon      <-c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov","Dec")
```

```
 dmon     <-cumsum(c(0,31,28,31,30,31,30,31,31,30,31,30,31))

 mmon     <-(dmon[-1]+dmon[-13])/2

if(pdf) {pdf(paste(modname,".pdf",sep=""),width=8,height=7)} else
{windows(width=8,height=7)}

  par(las=1,cex.axis=0.7)

  plot(mod$pir[,1,1],type="n",xlim=c(0,365),ylim=c(0,1),xlab="day",ylab="probability",

  main=modname)

#lines((mod$pir[,1,1]+(mod$pir[,1,2])/2,col="yellow")

  lines(mean.c,col="gray")

  lines(mean.p,col="pink")

  abline(v=dmon,col="gray",lwd=1)

  text(mmon,rep(-0.03,12),nmon,cex=0.7,xpd=TRUE)

  for (isite in 1:mod$nsite) {for (im in 1:mod$m){

  mpir[,isite]<-mpir[,isite]+mod$delta[im]*mod$pir[,isite,im]

print(mpir[,isite])

  lines(mod$pir[,isite,im],lwd=2,col=isite,lty=im+1)

}

lines(mpir[,isite],lwd=2,col=isite,lty=1)

}

if(pdf) dev.off()

return(list(nfact=mod$nfact,nsite=mod$nsite,m=mod$m,gamma=mod$gamma,delta=m
od$delta,code=mod$code,mllk=mod$mllk,AIC=mod$AIC,BIC=mod$BIC))

}
```

```
# ====================mod131
nfact=1;nsite<-3;m=1
pipars0<-array(
c(
-1,-1
 ),dim=c(nfact,nsite,m))
gamma0<-matrix(c(1),m,m,byrow=TRUE)
mod<multivariate.seasonal.bernoulli.HMM.mle(rain,nfact,nsite,m,pipars0,gamma0,X=X,
details=TRUE)
mod131<-mod
rm(nfact,nsite,m,pipars0,gamma0)
mod131
# ========================mod132
nfact=1;nsite<-3;m=2
pipars0<-array(
c(
-0.85,-0.94,
-1.00 -1.00
  ),dim=c(nfact,nsite,m))
gamma0<-matrix(c(
 9,1,
 1,9
 )/10,m,m,byrow=TRUE)
mod<multivariate.seasonal.bernoulli.HMM.mle(rain,nfact,nsite,m,pipars0,gamma0,X=X,
details=TRUE)
```

```
mod132<-mod

rm(nfact,nsite,m,pipars0,gamma0)

mod132

# ============================================== mod133

nfact=1;nsite<-3;m<-3

pipars0<-array(

c(

 1.9, 1.5,

 0.0, 0.5,

-2.0, -2.5

  ),dim=c(nfact,nsite,m))

 gamma0<-matrix(c(

7,2,1,

3,6,1,

1,3,6

 )/10,m,m,byrow=TRUE)


mod<multivariate.seasonal.bernoulli.HMM.mle(rain,nfact,nsite,m,pipars0,gamma0,X=X,
details=TRUE)

mod133<-mod

rm(nfact,nsite,m,pipars0,gamma0)

mod133
```

```
# ==========================================mod331
nfact=3;nsite<-3;m<-1
pipars0<-array(
c(
 1.4, 1.0, 0.0,
 0.5,-0.1, 0.0
  ),dim=c(nfact,nsite,m))
gamma0<-matrix(c(
10
  )/10,m,m,byrow=TRUE)
mod<multivariate.seasonal.bernoulli.HMM.mle(rain,nfact,nsite,m,pipars0,gamma0,X=X,
details=TRUE)
mod331<-mod
rm(nfact,nsite,m,pipars0,gamma0)
mod331
# ============================================ mod332
nfact=3;nsite<-3;m<-2
pipars0<-array(
c(
 1.1,0.5,0.1,
 1.2,0.4,0.1,
-1.2,0.5,0.1,
-1.1,0.4,0.1
  ),dim=c(nfact,nsite,m))
```

```
gamma0<-matrix(c(
7,3,
4,6
)/10,m,m,byrow=TRUE)
mod<multivariate.seasonal.bernoulli.HMM.mle(rain,nfact,nsite,m,pipars0,gamma0,X=X,
details=TRUE)
mod332<-mod
rm(nfact,nsite,m,pipars0,gamma0)
mod332
#==============================mod333
nfact=3;nsite<-3;m<-3
pipars0<-array(
c(
 1.1,0.5,0.1,
 1.2,0.4,0.1,
-1.2,0.5,0.1,
-1.1,0.4,0.1
  ),dim=c(nfact,nsite,m))
gamma0<-matrix(c(
7,3,
4,6
)/10,m,m,byrow=TRUE)
mod<multivariate.seasonal.bernoulli.HMM.mle(rain,nfact,nsite,m,pipars0,gamma0,X=X,
details=TRUE)
mod333<-mod
```

```
rm(nfact,nsite,m,pipars0,gamma0)

mod333

# ============================================ mod532

nfact=5;nsite<-3;m<-2

pipars0<-array(

c(

 1.1,0.5,0.1,0.1,0.0,

 1.2,0.4,0.1,0.1,0.0,

-1.2,0.5,0.1,0.1,0.0,

-1.1,0.4,0.1,0.1,0.0

  ),dim=c(nfact,nsite,m))

gamma0<-matrix(c(

7,3,

4,6

)/10,m,m,byrow=TRUE)

mod<multivariate.seasonal.bernoulli.HMM.mle(rain,nfact,nsite,m,pipars0,gamma0,X=X,
details=TRUE)

mod532<-mod

rm(nfact,nsite,m,pipars0,gamma0)

mod532

# ============================================ mod533

nfact=5;nsite<-3;m<-3

pipars0<-array(0,dim=c(nfact,nsite,m))

pipars0[,,1]<-mod522$pipars[,,1]
```

```
pipars0[,,3]<-mod522$pipars[,,2]

pipars0[,,2]<-(pipars0[,,1]+pipars0[,,3])/2

gamma0<-mod123$gamma

mod<multivariate.seasonal.bernoulli.HMM.mle(rain,nfact,nsite,m,pipars0,gamma0,X=X,
details=TRUE)

mod533<-mod

rm(nfact,nsite,m,pipars0,gamma0)

mod533

# ============================================ mod732

nfact=7;nsite<-3;m<-2

pipars0<-array(

c(

 1.9,0.4,0.2,-0.2,0.1,0.0,0.0,

 1.7,0.3,0.3,-0.5,0.0,0.0,0.0,

-3.4,0.9,0.1,-0.8,0.6,0.0,0.0,

-4.0,1.5,-0.5,-0.4,0.7,0.0,0.0

 ),dim=c(nfact,nsite,m))

gamma0<-matrix(c(

6,4,

2,8

)/10,m,m,byrow=TRUE)

mod<-multivariate.seasonal.bernoulli.HMM.mle(rain,nfact,nsite,m,pipars0,
gamma0,X=X,details=TRUE)

mod732<-mod

rm(nfact,nsite,m,pipars0,gamma0)
```

```
mod732

#=============================================== mod733

nfact=7;nsite<-3;m<-3

pipars0<-array(0,dim=c(nfact,nsite,m))

pipars0[,,1]<-mod722$pipars[,,1]

pipars0[,,3]<-mod722$pipars[,,2]

pipars0[,,2]<-(pipars0[,,1]+pipars0[,,3])/2

gamma0<-mod523$gamma

mod<-multivariate.seasonal.bernoulli.HMM.mle(rain,nfact,nsite,m,pipars0,
gamma0,X=X,details=TRUE)

mod733<-mod

rm(nfact,nsite,m,pipars0,gamma0)

mod733

# ============================================== mod932

nfact=9;nsite<-3;m<-2

pipars0<-array(

c(

 1.9,0.4, 0.2,-0.2,0.2,0.3,0.1,0.0,0.0,

 1.6,0.3, 0.3,-0.4,0.0,0.2,0.0,0.0,0.0,

-3.4,0.8, 0.1,-0.7,0.6,0.1,0.0,0.0,0.0,

-4.2,1.7,-0.5,-0.4,0.9,0.1,0.1,0.0,0.0

 ),dim=c(nfact,nsite,m))

gamma0<-matrix(c(

6,4,
```

```
2,8

)/10,m,m,byrow=TRUE)

mod<-multivariate.seasonal.bernoulli.HMM.mle(rain,nfact,nsite,m,pipars0,
gamma0,X=X,details=TRUE)

mod932<-mod

rm(nfact,nsite,m,pipars0,gamma0)

mod932

# ================================================= plots

multivariate.seasonal.bernoulli.HMM.plot(mod131,pdf=FALSE)

multivariate.seasonal.bernoulli.HMM.plot(mod132,pdf=FALSE)

multivariate.seasonal.bernoulli.HMM.plot(mod133,pdf=FALSE)

multivariate.seasonal.bernoulli.HMM.plot(mod332,pdf=FALSE)

multivariate.seasonal.bernoulli.HMM.plot(mod532,pdf=FALSE)

multivariate.seasonal.bernoulli.HMM.plot(mod533,pdf=FALSE)

multivariate.seasonal.bernoulli.HMM.plot(mod732,pdf=FALSE)

multivariate.seasonal.bernoulli.HMM.plot(mod733,pdf=FALSE)

multivariate.seasonal.bernoulli.HMM.plot(mod932,pdf=FALSE)

# Summarize

round(c(mod131$mllk,mod132$mllk,mod133$mllk,mod332$mllk,mod532$mll,mod533$
mllk,mod732$mllk,mod733$mllk,mod932$mllk))

round(c(mod131$AIC,mod132$AIC,mod133$AIC,mod332$AIC,mod532$AIC,mod533$AIC
,mod732$AIC,mod733$AIC,mod932$AIC))

round(c(mod131$BIC,mod132$BIC,mod133$BIC,mod332$BIC,mod532$BIC,mod533$BIC,
mod732$BIC,mod733$BIC,mod932$BIC))
```

```
#======== Compare the AIC BIC values of models

BICs<c(mod131$AIC,mod132$AIC,mod133$AIC,mod332$AIC,mod532$AIC,mod533$AIC,
mod732$AIC,mod733$AIC,mod932$AIC)

AICs<c(mod131$BIC,mod132$BIC,mod133$BIC,mod332$BIC,mod532$BIC,mod533$BIC,
mod732$BIC,mod733$BIC,mod932$BIC)


plot(AICs,type="b",col="blue",ylab="AICs", xlab="Models")

plot(BICs,type="b",col="red",ylab="BICs", xlab="Models")


plot(AICs,type="b",col="blue",ylab="AIC & BIC", xlab="Models")

lines(BICs,type="b",col="red")
```

B.2. R Codes for HMMs to Find Amount of Precipitation

The following R codes are used to find the amount of precipitation and estimation plots for each region seperately.

```
# ==========================Import the dailydata

dailydata<- read.csv("C:\\Users\\Lab_User\\Desktop\\17234.csv")

# extract rainfall amounts

r<-dailydata[,1]

summary(r)

n<-length(r)


# remove the last January (for now)

n<-floor(n/365)*365

r<-r[1:n]
```

```
ny<-n/365

#================= The following exploratory bits work only for the complete years

# Extract wet days

r0<- r==0

r1<- r>0

r1g0<- as.numeric(c(TRUE,r0[-n] & r1[-1]))

r1g1<- as.numeric(c(TRUE,r1[-n] & r1[-1]))

rw<-r;rw[r0]<-NA


rm   <-matrix(r,nrow=365)

r0m  <-matrix(r0,nrow=365)

r1m  <-matrix(r1,nrow=365)

r1g0m<-matrix(r1g0,nrow=365)

r1g1m<-matrix(r1g1,nrow=365)

rwm  <-matrix(rw,nrow=365)


rmean   <-apply(rm,1,mean)

r0mean  <-apply(r0m,1,mean)

r1mean  <-apply(r1m,1,mean)

rwmean  <-apply(rwm,1,mean,na.rm=TRUE)

rwsd    <-apply(rwm,1,sd,na.rm=TRUE)

rwcv    <-rwsd/rwmean


r0sum   <-apply(r0m,1,sum)
```

```
r1sum   <-apply(r1m,1,sum)
r1g0sum <-apply(r1g0m,1,sum);r1g0sum[1]<-r1g0sum[1]-1
r1g1sum <-apply(r1g1m,1,sum);r1g1sum[1]<-r1g1sum[1]-1

n1     <-cbind(r1sum,ny-r1sum)
n1g0   <-cbind(r1g0sum,c(r0sum[365],r0sum[1:364])-r1g0sum)
n1g1   <-cbind(r1g1sum,c(r1sum[365],r1sum[1:364])-r1g1sum)

r1g0mean <-n1g0[,1]/(n1g0[,1]+n1g0[,2])
r1g1mean <-n1g1[,1]/(n1g1[,1]+n1g1[,2])

#===============================Plots for depths
par(mfrow=c(2,2),las=1)
plot(rmean,type="h",lwd=0.5,col="gray",xlab="day",ylab="mm",
 main="Daily rainfall: mean")
lines(lowess(rmean,f=2/3),col=2,lwd=2)
lines(lowess(rmean,f=0.2),col=4,lwd=2)
legend("topright",c("lowess: f=2/3","lowess: f=0.2"),lwd=c(1,1),col=c(2,4),cex=0.8)

plot(rwmean,type="h",lwd=0.5,col="gray",xlab="day",ylab="mm",
 main="Daily depth: mean")
lines(lowess(rwmean,f=2/3),col=2,lwd=2)
lines(lowess(rwmean,f=0.1),col=4,lwd=2)
legend("topright",c("lowess: f=2/3","lowess: f=0.2"),lwd=c(1,1),col=c(2,4),cex=0.8)
```

```
plot(rwsd,type="h",lwd=0.5,col="gray",xlab="day",ylab="mm",
 main="Daily depth: standard deviation")
lines(lowess(rwsd,f=2/3),col=2,lwd=2)
lines(lowess(rwsd,f=0.1),col=4,lwd=2)
legend("topright",c("lowess: f=2/3","lowess: f=0.2"),lwd=c(1,1),col=c(2,4),cex=0.8)

plot(rwcv,type="h",lwd=0.5,col="gray",xlab="day",ylab="",
 main="Daily depth: coef. of variation")
lines(lowess(rwcv,f=2/3),col=2,lwd=2)
lines(lowess(rwcv,f=0.1),col=4,lwd=2)
legend("topright",c("lowess: f=2/3","lowess: f=0.2"),lwd=c(1,1),col=c(2,4),cex=0.8)
dev.off()

#====================================================
# Model for depths
# Functions needed to fit a seasonal gamma-HMM model
# ====================================
# Function to transform seasonal_gamma-HMM
# Natural parameters to working parameters
seasonal_gamma.HMM.pn2pw<-function(parsm1,parsm2,parsc1,parsc2,gamma)
{
 tspars  <- c(parsm1,parsm2,parsc1,parsc2)
 tgamma   <- NULL
```

```
 foo      <- log(gamma/diag(gamma))
 tgamma   <- as.vector(foo[!diag(2)])
 parvect  <- c(tspars,tgamma)
 parvect
}
# =======================================
# Function to transform seasonal_gamma-HMM
# working parameters to natural parameters
seasonal_gamma.HMM.pw2pn<-function(parsn,parvect)
{
n1<-1;n2=n1+parsn[1]-1
parsm1<-parvect[n1:n2]
n1<-n2+1;n2=n2+parsn[2]
parsm2<-parvect[n1:n2]
n1<-n2+1;n2=n2+parsn[3]
parsc1<-parvect[n1:n2]
n1<-n2+1;n2=n2+parsn[4]
parsc2<-parvect[n1:n2]
n1<-n2+1;n2=n2+2
gamma       <- diag(2)
gamma[!gamma]<- exp(parvect[n1:n2])
gamma       <- gamma/apply(gamma,1,sum)
delta      <- solve(t(diag(2)-gamma+1),rep(1,2))
```

```
list(parsn=parsn, parsm1=parsm1, parsm2=parsm2, parsc1=parsc1, parsc2=parsc2,
gamma=gamma,delta=delta)

}

# ======================================

# Function to compute the likelihood of a gamma-HMM

seasonal_gamma.HMM.mllk<-function(parvect,rain,parsn,X,details=TRUE,...)

{

n     <-length(rain)

p     <-seasonal_gamma.HMM.pw2pn(parsn,parvect)

m1    <-exp(as.matrix(X[,1:parsn[1]])%*%p$parsm1)

m2    <-exp(as.matrix(X[,1:parsn[2]])%*%p$parsm2)

c1    <-exp(as.matrix(X[,1:parsn[3]])%*%p$parsc1)

c2    <-exp(as.matrix( X[,1:parsn[4]])%*%p$parsc2)

shape  <-cbind(c1,c2)^2

rate   <-cbind((c1^2)/m1,(c2^2)/m2)

lscale <-0

foo<-p$delta

j<-0

for (i in 1:n)

{

j<-j+1; if(j==366) j<-1

B<-p$gamma

if (!is.na(rain[i])) B<-B*dgamma(rain[i],shape=shape[j,],rate=rate[j,])

foo<-foo%*%B
```

```
sumfoo<-sum(foo)

lscale<-lscale+log(sumfoo)

foo<-foo/sumfoo

}

mllk<- -lscale

if(details) cat(paste("-log(likelihood) =",mllk,"\n"))

return(mllk)

}

# test it

rain<-rw

parsn=c(3,3,3,3)

parsm1  =c(1.38,-0.01,0.01)

parsm2  =c(1.38,0,0.01)

parsc1  = c(0,0.01,0.01)

parsc2  = c(0,-0.01,0.01)

gamma   = matrix(c(0.9,0.2,0.1,0.8),2,2)

parvect<-seasonal_gamma.HMM.pn2pw(parsm1, parsm2, parsc1,parsc2,gamma)

seasonal_gamma.HMM.pw2pn(parsn,parvect)

seasonal_gamma.HMM.mllk(parvect,rain=rain,parsn,X=X, details=TRUE)

# ====================================

# Function to minimize mllk=-log(likelihood) of a seasonal gamma-HMM

seasonal_gamma.HMM.mle<-function(rain,parsn,parsm10,
parsm20,parsc10,parsc20,gamma0,X=X,details=TRUE...)

{
```

```
n       <-length(rain)

parvect0 <-seasonal_gamma.HMM.pn2pw(parsm10, parsm20,
parsc10,parsc20,gamma0)

mod     <-nlm(seasonal_gamma.HMM.mllk,parvect0,rain=rain,parsn=parsn,X=X,

 details=details,iterlim = 1000)

mllk    <-mod$minimum

p       <-seasonal_gamma.HMM.pw2pn(parsn,mod$estimate)

np      <-length(parvect0)

AIC     <-2*(mllk+np)

BIC     <-2*mllk+np*log(n)

return(list(parsn=p$parsn,parsm1=p$parsm1,parsm2=p$parsm2,parsc1=p$parsc1,

 parsc2
=p$parsc2,gamma=p$gamma,delta=p$delta,code=mod$code,mllk=mllk,AIC=AIC,BIC=BIC
))

}

# ======================================== mod3333

parsn=parsn0=c(3,3,3,3)

parsm1=parsm10  =c(1.84,-0.06,0.33)

parsm2=parsm20  =c(-0.34,0.14,0.42)

parsc1=parsc10  =c(0.003,0.008,-0.008)

parsc2=parsc20  =c(0.25,-0.02,-0.10)

gamma=gamma0    =matrix(c(0.8,0.6,0.2,0.4),2,2)

parvect=parvect0=seasonal_gamma.HMM.pn2pw(parsm1, parsm2,
parsc1,parsc2,gamma)

seasonal_gamma.HMM.pw2pn(parsn,parvect)
```

```
seasonal_gamma.HMM.mllk(parvect0,rain=rain,parsn=parsn0,X=X, details=TRUE)


mod3333<-seasonal_gamma.HMM.mle(rain,parsn0,parsm10, parsm20,
parsc10,parsc20,gamma0,X=X,details=TRUE)

mod3333

# ========================================= mod5533

parsn=parsn0=c(5,5,3,3)

parsm1=parsm10  =c(mod3333$parsm1,0,0)

parsm2=parsm20  =c(mod3333$parsm2,0,0)

parsc1=parsc10  =mod3333$parsc1

parsc2=parsc20  =mod3333$parsc2

gamma=gamma0    =mod3333$gamma

parvect=parvect0=seasonal_gamma.HMM.pn2pw(parsm1, parsm2,
parsc1,parsc2,gamma)

seasonal_gamma.HMM.pw2pn(parsn,parvect)

seasonal_gamma.HMM.mllk(parvect0,rain=rain,parsn=parsn0,X=X, details=TRUE)


mod5533<-seasonal_gamma.HMM.mle(rain,parsn0,parsm10, parsm20,
parsc10,parsc20,gamma0,X=X,details=TRUE)

mod5533

# ========================================= mod5555

parsn=parsn0=c(5,5,5,5)

parsm1=parsm10  =mod5533$parsm1

parsm2=parsm20  =mod5533$parsm2

parsc1=parsc10  =c(mod5533$parsc1,0,0)
```

```
parsc2=parsc20  =c(mod5533$parsc2,0,0)

gamma=gamma0    =mod5533$gamma

parvect=parvect0=seasonal_gamma.HMM.pn2pw(parsm1, parsm2,
parsc1,parsc2,gamma)

seasonal_gamma.HMM.pw2pn(parsn,parvect)

seasonal_gamma.HMM.mllk(parvect0,rain=rain,parsn=parsn0,X=X, details=TRUE)


mod5555<-seasonal_gamma.HMM.mle(rain,parsn0,parsm10, parsm20,
parsc10,parsc20,gamma0,X=X,details=TRUE)

mod5555

# ========================================= mod7733

parsn=parsn0=c(7,7,3,3)

parsm1=parsm10  =c(mod5533$parsm1,0,0)

parsm2=parsm20  =c(mod5533$parsm2,0,0)

parsc1=parsc10  =mod5533$parsc1

parsc2=parsc20  =mod5533$parsc2

gamma=gamma0    =mod5533$gamma

parvect=parvect0=seasonal_gamma.HMM.pn2pw(parsm1, parsm2,
parsc1,parsc2,gamma)

seasonal_gamma.HMM.pw2pn(parsn,parvect)

seasonal_gamma.HMM.mllk(parvect0,rain=rain,parsn=parsn0,X=X, details=TRUE)

mod7733<-seasonal_gamma.HMM.mle(rain,parsn0,parsm10, parsm20,
parsc10,parsc20,gamma0,X=X,details=TRUE)

mod7733
```

```
# ======================================== mod7755

parsn=parsn0=c(7,7,5,5)

parsm1=parsm10  =mod7733$parsm1

parsm2=parsm20  =mod7733$parsm2

parsc1=parsc10  =mod5555$parsc1

parsc2=parsc20  =mod5555$parsc2

gamma=gamma0    =mod7733$gamma

parvect=parvect0=seasonal_gamma.HMM.pn2pw(parsm1, parsm2,
parsc1,parsc2,gamma)

seasonal_gamma.HMM.pw2pn(parsn,parvect)

seasonal_gamma.HMM.mllk(parvect0,rain=rain,parsn=parsn0,X=X, details=TRUE)


mod7755<-seasonal_gamma.HMM.mle(rain,parsn0,parsm10, parsm20,
parsc10,parsc20,gamma0,X=X,details=TRUE)

mod7755

# Select

mod3333$mllk;mod5533$mllk;mod7733$mllk;mod5555$mllk;mod7755$mllk

mod3333$AIC;mod5533$AIC;mod7733$AIC;mod5555$AIC;mod7755$AIC

mod3333$BIC;mod5533$BIC;mod7733$BIC;mod5555$BIC;mod7755$BIC


#----For each model repeat all following steps removing by "#"----

mod<-mod3333; modcaption<-"Model(3,3,3,3)"

#mod<-mod5533; modcaption<-"Model(5,5,3,3)"

#mod<-mod5555; modcaption<-"Model(5,5,5,5)"

#mod<-mod7733; modcaption<-"Model(7,7,3,3)"
```

```
#mod<-mod7755; modcaption<-"Model(7,7,5,5)"


# Repeat the plot for each model
filename=paste("Depth-",modcaption,".pdf",sep="")
m1fit<-exp(X[,1:mod$parsn[1]]%*%mod$parsm1)
m2fit<-exp(X[,1:mod$parsn[2]]%*%mod$parsm2)
mfit<-mod$delta[1]*m1fit+mod$delta[2]*m2fit
c1fit<-exp(X[,1:mod$parsn[3]]%*%mod$parsc1)
c2fit<-exp(X[,1:mod$parsn[4]]%*%mod$parsc2)
cfit<-mod$delta[1]*c1fit+mod$delta[2]*c2fit


# Plots for depth model
par(mfrow=c(2,2),las=1)


plot(rmean,type="h",ylim=c(0,5),col="gray",lwd=1,xlab="day",ylab="mm",
 main="Daily rainfall: mean")
text(365/2,5*0.95,paste(modcaption,": AIC = ",round(mod$AIC,1),sep=""),col=1,cex=1)
text(365/2,5*0.85,paste("P(State 1) =",round(mod$delta[1],2)),col=4)
text(365/2,5*0.75,paste("P(State 2) =",round(mod$delta[2],2)),col=2)


plot(rwmean,type="h",ylim=c(0,15),col="gray",lwd=1,,xlab="day",ylab="mm",
 main="Daily depth: mean")
lines(m1fit,col=4,lwd=2)
lines(m2fit,col=2,lwd=2)
```

```
lines(mfit,col=1,lwd=1)


plot(rwsd,type="h",ylim=c(0,20),col="gray",lwd=1,,xlab="day",ylab="mm",
 main="Daily depth: standard deviation")
lines(c1fit*m1fit,col=4,lwd=2)
lines(c2fit*m2fit,col=2,lwd=2)
lines(cfit*mfit,col=1,lwd=1)


plot(rwcv,type="h",,ylim=c(0,2.5),col="gray",lwd=1,,xlab="day",ylab="",
 main="Daily depth: coef. of variation")
lines(c1fit,col=4,lwd=2)
lines(c2fit,col=2,lwd=2)
lines(cfit,col=1,lwd=1)
dev.off()
```

B.3. Viterbi Algorithm for Hidden Markov Model

“HiddenMarkov” package from R is used to find the most likely state sequences. It contains for the analysis of “Discrete Time Hidden Markov Models” and functions for simulation, parameter estimation and the Viterbi algorithm.