REGIONAL FLOOD FREQUENCY ANALYSIS FOR CEYHAN BASIN

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

BY

MEHMET ALTUĞ ŞAHİN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
CIVIL ENGINEERING

JANUARY 2013

Approval of the thesis:

**REGIONAL FLOOD FREQUENCY ANALYSIS FOR CEYHAN BASIN**

submitted by **MEHMET ALTUĞ ŞAHİN** in partial fulfillment of the requirements for the degree of **Master of Science in Civil Engineering Department, Middle East Technical University** by,

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

Prof. Dr. Ahmet Cevdet Yalçıner
Head of Department, **Civil Engineering** ____________________

Assoc. Prof. Dr. Zuhal Akyürek
Supervisor, **Civil Engineering Dept., METU** ____________________

**Examining Committee Members:**

Assoc. Prof. Dr. Nuri Merzi
Civil Engineering Dept., METU ____________________

Assoc. Prof. Dr. Zuhal Akyürek
Civil Engineering Dept., METU ____________________

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

Assoc. Prof. Dr. Yakup Darama
State Hydraulic Works ____________________

Dr. Fatih Keskin
Palye Engineering and Consultancy Company ____________________

**Date:** 31.01.2013

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

Name, Last name: Mehmet Altuğ ŞAHİN

Signature :

# ABSTRACT


### REGIONAL FLOOD FREQUENCY ANALYSIS FOR CEYHAN



Şahin, Mehmet Altuğ
M.Sc., Department of Civil Engineering
Supervisor: Assoc. Prof. Dr. Zuhal Akyürek



January 2013, 97 pages



Regional flood frequency techniques are commonly used to estimate flood quantiles when flood data are unavailable or the record length at an individual gauging station is insufficient for reliable analyses. These methods compensate for limited or unavailable data by pooling data from nearby gauged sites. This requires the delineation of hydrologically homogeneous regions in which the flood regime is sufficiently similar to allow the spatial transfer of information. Therefore, several Regional Flood Frequency Analysis (RFFA) methods are applied to the Ceyhan Basin. Dalyrmple (1960) Method is applied as a common RFFA method used in Turkey. Multivariate statistical techniques which are Stepwise and Nonlinear Regression Analysis are also applied to flood statistics and basin characteristics for gauging stations. Rainfall, Perimeter, Length of Main River, Circularity, Relative Relief, Basin Relief, $H_{max}$, $H_{min}$, $H_{mean}$ and $H_{\Delta}$ are the simple additional basin characteristics. Moreover, before the analysis started, stations are clustered according to their basin characteristics by using the combination of Ward's and k-means clustering techniques. At the end of the study, the results are compared considering the Root Mean Squared Errors, Nash-Sutcliffe Efficiency Index and % difference of results. Using additional basin characteristics and making an analysis with multivariate statistical techniques have positive effect for getting accurate results compared to Dalyrmple (1960) Method in Ceyhan Basin. Clustered region data give more accurate results than non-clustered region data. Comparison between clustered region and non-clustered region $Q_{100}/Q_{2.33}$ reduced variate values for whole region is 3.53, for cluster-2 it is 3.43 and for cluster-3 it is 3.65. This show that clustering has positive effect in the results. Nonlinear Regression Analysis with three clusters give less errors which are 29.54 RMSE and 0.735 Nash-Sutcliffe Index, when compared to other methods in Ceyhan Basin.



**Keywords:** Regional Flood Frequency Analysis, Stepwise Regression Analysis, Nonlinear Regression Analysis.

# ÖZ

## CEYHAN HAVZASI İÇİN BÖLGESEL TAŞKIN FREKANS ANALİZİ

Şahin, Mehmet Altuğ
Yüksek Lisans, İnşaat Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. Zuhal Akyürek

Ocak 2013, 97 sayfa

Bölgesel Taşkın Frekans Analiz yöntemi genellikle ölçüm yapılmamış ya da yetersiz miktarda ölçümün bulunduğu havzalardaki taşkın debilerinin hesaplanmasında kullanılır. Bu methodun uygulaması esnasında yetersiz veri sayısını arttırmak için bölgede bulunan diğer istasyonlar kullanılır. Diğer istasyonların kullanılabilmesi ve verilerin ölçüm olmayan havzaya taşınabilmesi için bu havzaların aynı özellikleri taşıması önemlidir. Bu yüzden bazı Bölgesel Taşkın Frekans Analizi yöntemleri Ceyhan havzasına uygulanmıştır. Dalyrmple (1960) method ise Türkiyede genel geçer kullanılan BTFA yöntemi olarak uygulanmıştır. İlave olarak adım adım ve Doğrusal Olmayan Regresyon Analiz yöntemleri de Çok Değişkenli İstatistiksel Yöntemler olarak havzadaki taşkını hesaplamak adına havza parametreleri ile birlikte kullanılmıştır. Bu Çok Değişkenli İstatistiksel Yöntemler çalışılırken yağış, en uzun nehir kolu, yuvarlaklık, göreceli rölyef, havza rölyefi, $H_{maks}$, $H_{min}$, $H_{ort}$ ve $H_{\Delta}$ gibi havza parametreleri kullanılmıştır. Ayrıca, analizlere başlamadan önce havzadaki istasyonlar özelliklerine göre Ward ve K-ortalama kümeleme yöntemleri kullanılarak gruplandırılmıştır. Analiz sonuçları değerlendirilirken Ortalama Hata Karesinin Kökü, Nash-Sutclife Verim İndeksi ve % değişimi sonuçları kullanılmıştır. Çalışma sonucunda, Ceyhan havzası için ilave havza parametreleri ile birlikte Çok Değişkenli İstatistiksel Yöntemler kullanılarak yapılan hesapların sonucunun, klasik yöntem olarak bilinen Dalyrmple (1960) yöntemine göre daha doğru sonuçlar verdiği gözlenmiştir. Kümeleme yapılan veriler kümelendirilmemiş verilere göre daha doğru sonuçlar vermiştir. Kümelenmiş ve kümelendirilmemiş bölgelerdeki $Q_{100}/Q_{2.33}$ katsayılarını kıyasladığımızda, kümelendirilmemiş bölgede 3.52, 2. Kümelenmiş bölgede 3.43 ve 3. Kümelenmiş bölgede 3.65 olduğu gözlenmiştir. Üç bölgeye ayıracak şekilde kümeleme yapılarak Doğrusal Olmayan Regresyon Analizi yapıldığında Ortalama Hata Karesinin Kökü 29.54 ve Nash-Sutcliffe Verim İndeksi 0.735 çıkmış, kümelenme ile uygulanan metodun diğer methodlara göre en az hata verdiği görülmüştür.

**Anahtar Kelimeler:** Bölgesel Taşkın Frekans Analizi, Doğrusal Olmayan Regresyon Analizi, Adım Adım Regresyon Analizi

To My Family

# ACKNOWLEDGEMENTS

This research has been carried out under the supervision of Assoc. Prof. Dr. Zuhal Akyürek in the Water Resources Department of Civil Engineering in the Middle East Technical University.

The author wishes to express his sincere appreciation to Assoc. Prof. Dr Zuhal Akyürek for her guidance and helpful advice throughout the research.

I am also quite thankful to Serdar Sürer and Arzu Soytekin for their help and a friend to me. Without their guidance and help, I would be lost in this study.

I am in great debt to my mother Feryal Şahin and father Nusret Şahin for their patience and supports. In addition, I would like to thank specially to my brother Ahmet Aykut Şahin who always brings happiness and joy to my life.

Special thanks to my managers; Nihat Dilek, Rıfat Göksu and Ahmet Oktay Kavas in Kayı Energy, for their understanding and tolerance throughout the preparation of this thesis.

I also would give my special thanks to my ex-manager Nedret Ünel in DSİ, for her tolerance and motivation at the beginning of master education.

I would like to thank my lovely friends Ahmet Korhan Gözcü, Mete Kaymak, Ahmet Gökhan Coşkun, İrem Şafak and Aslı Selver, for their support, patience and endless love.

Last but not the least, sincere thanks to my colleagues Cemre Harzem Yardım, Melih Türkoğlu and Mert Levendoğlu for their endless help and support at even the most difficult times during the preparation of this thesis.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

## 1.1 Definition of the problem

Starting from the ancient times, man is building hydraulic structures for different purposes. In the modern age, the design of any water project consists of the following consequent steps: Hydrologic design, hydraulic design and structural design. Among these steps, hydrologic design has very important role because any mistake made at this point will result in the failure of design no matter how correct the other steps are carried out.

Hydrologists are dealing with nature. Hydrologic events appear as uncertainties of nature. Since there are numerous sources of uncertainty about the physical processes that give rise to observed events, a statistical approach is often desirable. For instance, it is not possible to predict stream flow and precipitation on a purely deterministic way in either the past (hindcasting) or future (forecasting) since it is impossible to know all their casual mechanisms quantitatively. Fortunately, methods of statistical analysis provide ways to reduce and summarize observed data, to present information in precise and meaningful form, to determine the underlying characteristics of the observed phenomena and to make predictions concerning future behavior. In other words, statistical methods acknowledge the existence of the uncertainty and enable its effects to be quantified.

Frequency analysis, being a statistical method, is the estimation of how often a specified event will occur. The goal of frequency analysis is to obtain a useful estimate of the quantile $Q_T$ for return period of T where Q is magnitude of the event that occurs at a given time at a given site. An estimate should not only be close to the true quantile but should also come with an assessment of how accurate it likely to be.

Flood frequency analysis has a significant role in social and economic assessment of water resources projects. The beneficial effects of frequency analysis may be stated as it helps to estimate the magnitudes of the extreme events that will occur in the future and thus will create a reasonable design criterion for the water resources projects. A frequency analysis is an efficient tool in design via forecasting which reduces the cost of projects by determining the values of extreme events in a rational way.

Frequency analysis is an information problem. If the length of the available data increases, the shape of the frequency distributions is determined more precisely and accurately. If an adequately long record of flood flows or rainfall is available then a frequency distribution for a site could be correctly calculated, so long as the relationships of concern are not changed externally, like change in the vegetal cover in the region or building a hydraulic structure on the river. Such changes may affect the relationships between the hydrologic elements like precipitation-runoff relationship.

Flood series at an individual site are seldom long enough to accurately estimate flood quantiles for return periods of interest. In other cases, flood data are unavailable at the site of interest, making at-site flood frequency analysis impossible. Regional flood frequency techniques which employ data from nearby sites have thus been developed to overcome the lack of flood data at a particular location.

Other major problems faced in the Regional Flood Frequency Analysis are the formation of a homogeneous region without any discordancy sites in the region, fitting a good enough distribution to the region and assessments of the accuracy of the estimated quantiles.

If data are available at the site of interest then the observed data provide a sample of realizations of Q. In many environmental applications the sample size is rarely sufficient to enable quantiles to be reliably estimated. It is generally held that a quantile of return period T can be reliably estimated from a data record of length n only if $T \leq n$. However, in many engineering applications based on annual data (e.g., annual maximum precipitation, streamflow, windspeed) this condition is rarely satisfied typically n< and T=100 or T=1000. To overcome this problem, several approaches have been devised that use alternative or additional sources of data.

Regional frequency analysis augments the data from the site of interest by using data from other sites that are judged to have frequency distributions similar to that of the site of interest. If a set of N sites each with n years of record can be found, then one might natively hope that the Nn data values will provide accurate estimates of quantiles as extreme as the Nn-year quantile $Q_{Nn}$. In practice this is not reasonable; problems arise because frequency distributions at different sites are not exactly identical and because event magnitudes at different sites may not be statistically independent.

### 1.2 Aim of the study

Accurate flood estimations are necessary for the development of floodplain management and flood warning systems, and the design and operation of water-control structures, such as reservoirs and culverts. Standard procedures for at-site flood frequency analysis involve assembling the annual maximum flood record at the site of interest and fitting an analytic probability distribution to the data (e.g., IACWD 1982). The fitted distribution is then used to estimate flood quantiles associated with a given return period, such as the flood magnitude expected to be probability of occurrence of flood is 1/100 which is equal to p=0.01 at any time. However, in most cases the at-site record length is too short to accurately estimate flood quantiles for return periods of interest: estimation of the 100-year event often requires extrapolation beyond the observed flood record. In other cases, flood data are unavailable at the site of interest, making at-site flood frequency analysis impossible. As the latter is often the case for watersheds throughout the world, particularly in data sparse developing countries, but also in data rich countries such as the United States (e.g., Mishra and Coulibaly 2009), the development of appropriate methods for flood quantile estimation in ungauged basins is a common research theme in hydrology. To compensate for limited or unavailable flood data, one solution is to “trade space for time” (Stedinger et al. 1993) using a regional flood frequency analysis, wherein the characterization of flood flows at the site of interest is derived using information pooled from nearby hydrologically similar gauged sites (NERC 1988). Regional flood frequency methods include the Index Flood method (e.g. Dalrymple 1960; Hosking and Wallis 1988, 1997; Stedinger and Lu 1995; Fill and Stedinger 1998; De Michele and Rosso 2001; Kjeldsen and Rosbjerg 2002), and regional regression procedures, such as weighted and generalized least squares regression (e.g. Tasker and Stedinger 1989; Tasker et al. 1996; Madsen and Rosbjerg 1997; Eng et al. 2005, 2007a, 2007b; Griffis and Stedinger 2007a; Jeong et al. 2007). Much of the recent research has focused on improving or comparing existing regional flood frequency techniques (e.g., Castellarin et al. 2001; Chiang et al. 2002a; Kjeldsen and Rosbjerg 2002; Eng et al. 2007b; Griffis and Stedinger 2007a; Neykov et al. 2007; Gruber and Stedinger 2008), and developing new methods for quantile estimation at ungauged basins located within the area used for model development (e.g., Chiang et al. 2002b; Eng et al. 2005, 2007b; Shu and Ouarda 2008; Saf 2009; Malekinezhad et al. 2011).

The research presented in this thesis draws on this base of knowledge to propose additional recommendations to improve regional flood estimation for ungauged basins. This research presents a method by which the flood regime in Ceyhan Basin can be derived based on knowledge of the relationships between flood statistics and additional basin characteristics. There are known methods for calculation Flood Frequency Analysis in Turkey. The research presented in this thesis is tried to extend the known methods and offer some guides to improve flood calculations in practical implementations.

## 1.3 Organization of Thesis

Chapter 1 is about the Introduction of the thesis. Literature about the used methodology is explained in Chapter 2. Ceyhan basin and basin data have been described in Chapter 3. Chapter 4 includes details of used processes and also includes the calculations. In Chapter 5, calculation results are discussed in details. Chapter 6 includes conclusions about the result and the research and give also recommendations about future studies.

# CHAPTER 2

# LITERATURE REVIEW

This part of the thesis reviews the literature on the subject of frequency analysis including at-site, regional analyses and method of estimation.

In this review, three aspects are considered namely single site analysis, regional analysis and method of estimation. Although the reviews on methods of estimation is included as a separate part for describing the background information used in the following sections.

On the other hand, at-site analysis is the one most commonly found in both research and practice. Although the topic of this study is mainly on Regional Frequency Analysis, review on at-site analysis is here since it is usually the first step in Regional Frequency Analysis.

### 2.1 Flood Frequency Analysis which are used by State Hydraulic Works(DSI)

There are many of Flood Frequency Analysis which are commonly used in the world. However, there are some standart methods which are well accepted by DSI in Flood Frequency Analysis. These methods are used for all of the hydroelectric, irrigation, flood protection etc. projects of DSI. All these methods are divided into two main groups, synthetic methods and natural methods.

**Synthetic methods:** In these methods, Flood Frequency Analysis is not only due to the former flood values in the basin. For example, these methods also consider the time period which the flood come to the basin mouth. Therefore, these analysis do not only consider meteorological values, they also consider Drainage Area, Mean Annual Rainfall, Length of Main River, Index Flow, Harmonic Slope, Curve Number etc. There are four common methods which are used as Synthetic methods in Turkey:

- Rational Method
- DSI Synthetic Method
- Mocus Method
- Snyder Method

**Natural Methods:** Only two paremeters are used in the analysis, one of them is station's drainage areas, other is the former flood values of stations. This method generally uses the stations which are on the same basin. However, these methods have only one parameter which is the flood frequency of the stations. Therefore, this method generally is applied by using closer station values. There are two common methods which are used as Natural Methods:

- Point Flood Frequency Analysis
- Regional Flood Frequency Analysis

### 2.2 At-Site Frequency Analysis

Many available methods of frequency analysis have been based on at-site probability distribution functions. Thus many univariate distribution functions, such as the Normal, Log-Normal, Gumbel, Pearson Type III and Log-Pearson Type III distributions, have been used for frequency analysis.

Gumbel is a special case of Extreme Value Family distribution. Gumbel described the genesis of the EV-1 distribution and the fitting method which was based on plotting the data on a double exponential probability scale such that they formed a nearly straight line.

Lowery and Nash (1970) compared a number of methods fitting EV-1 distribution, such as the method of moments, the method of regression, Gumbel's fitting method and the method of maximum likelihood, in terms of bias, mean square errors and relative efficiency using the same numerical data.

U.S. Water Recources Council (1976) recommended the use of the Log-Pearson Type III distribution for frequency analysis in the U.S.A., which was derived from the Pearson Type III. As a result, this distribution has been widely used. The Pearson Type III and Log-Pearson Type III distributions have been analyzed and criticized at length by Matalas and Wallis, (1973); Bobee, (1975); Bobee and Robitaille, (1977); Condie, (1977); Kite, (1977); Wallis and Wood, (1985).

Gumbel distribution is a statistical method often used for predicting extreme hydrological events such as floods (Zelenhasic, 1970; Haan, 1977; Shaw, 1983).
The Log-Pearson Type-3 distribution is used broadly in hydrologic applications and is right now used in Turkey. The properties of the LP3 distribution are rather complicated in that the distribution has two interacting shape parameters. LP3 distribution are used in different tasks (Bobee, 1975; Bobee and Ashkar, 1991; Griffis and Stedinger, 2007)

**2.3 Regional Flood Frequency Analysis**

The two most common regional flood frequency techniques are the Index Flood method (e.g. Dalrymple 1960; Hosking and Wallis 1988, 1997; Stedinger and Lu 1995; Fill and Stedinger 1998; De Michele and Rosso 2001; Kjeldsen and Rosbjerg 2002), and regional regression analyses (e.g. Tasker and Stedinger 1989; Tasker et al. 1996; Madsen and Rosbjerg 1997; Eng et al. 2005, 2007a, 2007b; Griffis and Stedinger 2007a; Jeong et al. 2007). Each of these procedures and their application in the context of the research herein are discussed below.

The Index Flood method is based on the premise that sites within a statistically (or hydrologically) homogeneous region share the same parent (or regional) flood frequency distribution with a common shape parameter, but each watershed has a site-specific scale parameter (a.k.a. the "index-flood") to represent possible changes in magnitude across the region. For application at gauged sites, the scale parameter is often given by the mean of the flood flows. For ungauged sites, however, this parameter must be related to physiographic characteristics of the watershed, the most important of which is drainage area. The basin characteristics at any point in the region (i.e. an ungauged site) can then be used to estimate the mean annual flood, which in turn can be used with the non-dimensional parent distribution to determine the flood magnitude corresponding to any return period at that location.

Index flood procedure is applied first by Dalyrmple (1960). In recent, this procedure is also used as Regional Flood Frequency Analysis for all hydropower, irrigation and flood projects in DSI. This procedure is explained in details in Chapter 4.

Early regional analysis based on the index flood method (Dalyrmple, 1960) was used for most of the regional frequency analysis made by U.S. Geological Survey prior to 1965. However there were some difficulties in defining the geographic regions in which all sites had similarly shaped frequency curves. These difficulties led to the use of regression method for regionalization, which was able to better represent the relationships between basin characteristics and frequency curves. In addition, the index flood method was renewed recently and various regional frequency analyses based on the index flood method have been studied.

Several developments have evolved from the original index flood method of regionalization. Wallis (1980) suggested the use of the method of probability weighted moments (PWMs) of Greenwood et al. (1979) to the regional flood frequency analysis based on the index flood method. This new method of estimation calculates the PWMs at each site in the region from the indexed annual floods, then the weighted regional average PWMs are used to compute the dimensionless average frequency curve.

Greis and Wood (1981) investigated regional flood frequency estimation and network design using the Gumbel distribution. They showed that regional estimates at gauged sites using PWM method were improved over more conventional methods such as the method of moments and the method of maximum likelihood and also proposed an improved method for ungauged basins, which combined PWM technique with more traditional mean peak flow estimation such as the index flood method.

Stedinger (1983) described some theoretical limitations of index flood method and suggested that the poor description of the true dimensionless flood distribution (caused by index flood method) could be overcome by using the logarithms of the peak flow values and unbiased moments or PWMs estimators.

Hosking and Wallis (1988) explored the effect of intersite dependence on regional flood frequency analysis based on the regional PWM procedures for both homogeneous and heterogeneous regions. Using Monte Carlo simulation, they concluded that there was no change in bias of quantile estimates by the presence of intersite dependence. Even though the accuracy of quantile estimates decreases when intersite dependence appears, the effect of intersite dependence is less important for practical applications than that of heterogeneity. They also mentioned that the Regional Frequency Analysis was preferable to at-site analysis even when both intersite dependence and heterogeneity appeared. Cunnane (1988) reviewed twelve different methods of Regional Frequency Analysis and rated the regional PWM as the best.

After this index flood approach, Wallis (1989) advised the approach based on theory of L-moments, then developed by Hosking and Wallis (1997). Probability distributions and data samples are summerized by L-moments. L- moments are determined from linear combinations of the ordered data values. They are related to ordinary moments, as a consequence sustaining measure of location, dispersion, skewness, kurtosis, and other aspects of the shape of probability distributions or data samples.

With increasing computing abilities and the availability of remotely sensed data, it may also be possible to improve quantile estimates in data limited areas by using available data more efficiently. Recent research demonstrates the ability to delineate climatic regions as a function of remotely sensed data, including land surface temperature, precipitation, and infiltration categories based on microtopography, surface crusting and soil cover (Corbane et al. 2008; Rhee et al. 2008). Remote sensing systems have also been used extensively to identify soil type, land use, land cover, geology and topography (Corbane et al. 2008; Brink and Eva 2009; Bertoldi et al. 2011; Inbar et al. 2011). Previous studies suggest that geology, land use, and land cover may help define the flood distribution of drainage basins (Chiang et al. 2002a, 2002b; Rao et al. 2006), and thus remotely sensed data could be used to infer the flood regime in areas with limited or unavailable flood data.

Regionalization of flood data using either the Index Flood method or regional regression assumes that the watershed processes governing the flood regime are sufficiently characterized by physical parameters aggregated at the watershed scale. Some may argue that spatially distributed parameters should be used to develop finer scale representations of hydrological processes (Beven and Kirkby 1979; Abbott et al. 1986; Boyle et al. 2001; Duffy 2004; Panday and Huyakorn 2004; Reed et al. 2007). However, there is a trade-off between characterizing the heterogeneity within and uniqueness of a single watershed using spatially distributed values as commonly practiced in hydrologic modeling, and characterizing the heterogeneity within a region using parameter values aggregated at the watershed scale as in regional flood frequency techniques. In the latter analyses, simple models are needed to infer the dominant processes governing extreme hydrologic response at the watershed scale, such that flood statistics can be successfully extrapolated from gauged basins for improved prediction in ungauged basins in data limited areas. (See for example, McDonnell et al. 2007; Tetzlaff et al. 2008; MacKinnon and Tetzlaff 2009, and citations therein.) Use of these simple models in conjunction with remotely sensed data would allow for the development of quantile estimators in data sparse countries such as Haiti by extrapolating the relationships developed for basins of similar physical composition in data rich countries such as the United States.

Hydrologic models such as rainfall-runoff models are another option for flood quantile estimation in ungauged basins. Unlike regional flood frequency analyses, however, rainfall-runoff models do not provide information pertaining to the flood distributions nor do they explain possible similarities in flood distributions among watersheds. Further, rainfall-runoff models pose additional problems because the critical storm duration and the spatial distribution of relevant storm events (and corresponding inflows) are unknown. In general, regional flood frequency analyses provide less accurate flood quantile estimates than at-site flood frequency analyses when sufficient gauged data is available (see for example, Griffis and Stedinger 2007b), but often provide more accurate flood quantiles than hydrologic modeling (USACE 1994).

Multivariate statistical techniques such as cluster analysis, principal component analysis, canonical correlation analysis, and linear discriminant analysis are commonly employed to delineate homogeneous regions, i.e. group sites with similar extreme hydrologic response, and subsequently classify ungauged sites (Zrinji and Burn 1994; Burn 1997; Burn et al. 1997; Chiang et al. 2002a, 2002b; Rao and Srinivas 2006; Srinivas et al. 2008; Mamun et al. 2012). Application of these methods requires the selection of appropriate similarity measures to characterize the extreme hydrologic response, or flood regime, at individual sites.

Several clustering techniques are available in the statistical literature, including hierarchical approaches such as single linkage, complete linkage, average linkage and Ward's method, as well as non-hierarchical approaches such as the k-means method (Johnson and Wichern, 2007)

Ward's minimum variance method is a special case of the objective function approach originally presented by Ward (1963). Ward suggested a general agglomerative hierarchical clustering procedure, where the criterion for choosing the pair of clusters to merge at each step is based on the optimal value of an objective function. This objective function could be "any function that reflects the investigator's purpose".

Nash and Shaw (1965) studied the degree of correlation among different catchment properties with mean annual flood(MAF). It was reported that a combination of drainage area and mean annual rainfall exhibited the highest coefficient of determination of 0.92 with MAF. Average catchment slope also provided a good coefficient of determination.

There are many methods are considered after Nash and Shaw (1965). For example Robson, A. J. & Reed, D. W. (1999) studied a new equation based on drainage area, soil index and vegetation in Flood Estimation Handbook. However, in Turkey there is limited knowledge about the vegetation and soil parameters to be considered in the calculations.

In addition, Ceyhan Basin is studied by Topaloglu (2005). In this study Dalyrmple (1960) method is used with Gumbel Distributions and an equation which is $\mathbf{Q_{2.33} = 0.585 * A^{0.727}}$ is derived for Ceyhan Basin. This equation is derived with only 15 stream gauging stations values. Results of this study could be used in comparison chapter.

### 2.4 Methodology

The motivation of the study is the flood calculation by using regional flood frequency analysis in Ceyhan Basin. All the steps of the study is shown in Figure 2.1.

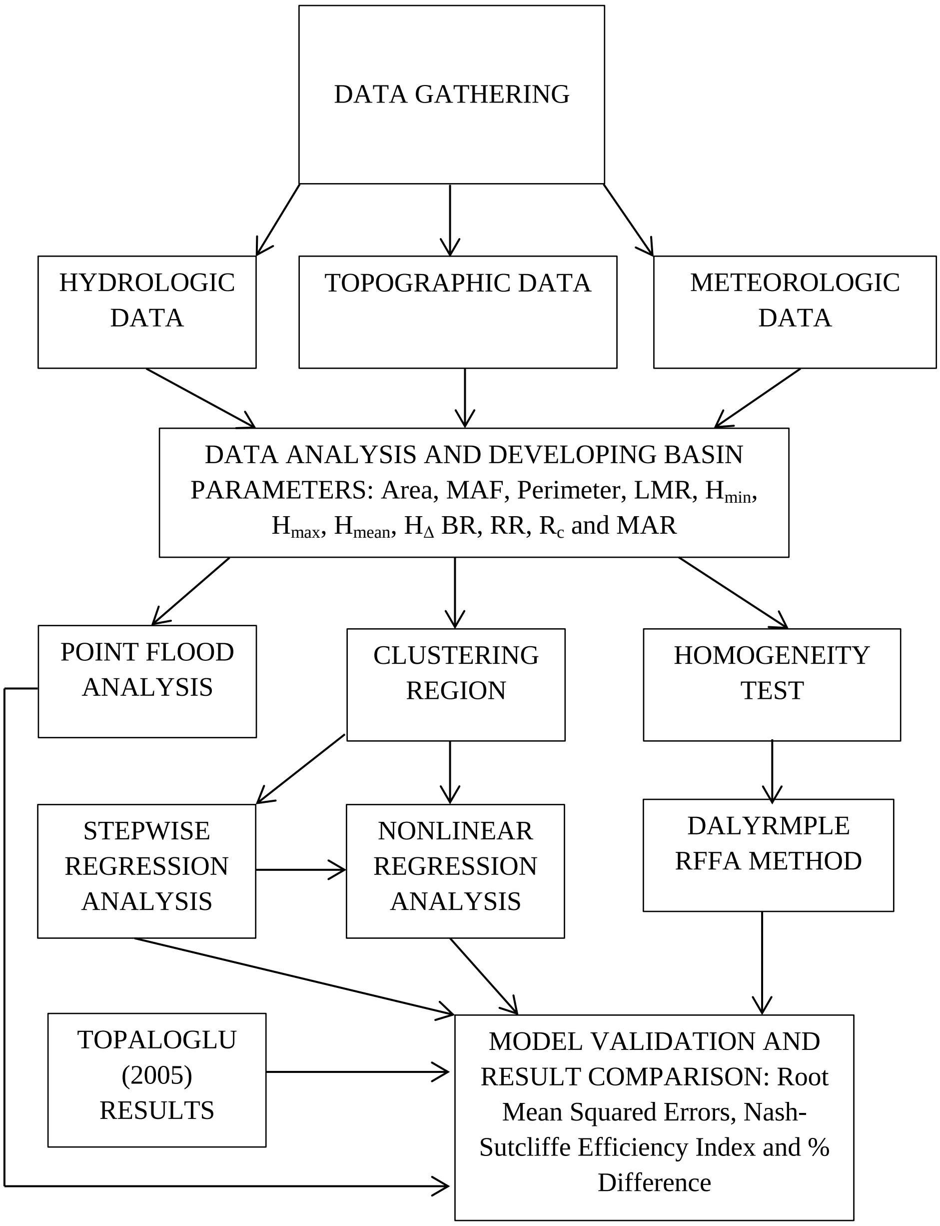


Figure 2.1 Steps of Methodology Chart.

# CHAPTER 3

# DESCRIPTION OF THE STUDY AREA AND DATA COLLECTION

## 3.1 Description of Study Area

Ceyhan Basin has drainage area of 21982.6 $km^2$, 7.18 $km^3$ average annual flow and 19 sub basins. Between 1940 and 2000 years daily discharge series were observed for these nineteen sub basins, Ceyhan Basin lies within 36.55° to 38.72° North latitudes, 35.45° to 37.81° East longitudes at the Southeast of Turkey (see Figure 3.1)

The study area has been chosen as Ceyhan basin, since the number of stream gauging stations in the basin is insufficient; many HEPP and irrigation projects exist in this area. Therefore, flood calculations of a new project is hard to calculate for ungauged catchment areas.

## 3.2 Data collection

### 3.2.1 Hydrologic Data

The first step of hydrologic modeling is to collect data by making observations. To discover hydrologic events, many gage stations that have sensible instruments should be built (limnigraph, etc) and observation network at gauging stations should be set up. Furthermore, at these observation networks that include many gauging stations, hydrometric measurements should be done carefully. Because hydrologic data change not only in time but also in location, measurements should be done regularly at closer points (Karahan, 2010).

There is one public organization in Turkey which is responsible for hydrologic observation and measurements: State Hydraulic Works (DSI). Daily, monthly, yearly and annual peak discharges are available for in these stream gauging stations. However, in the past years, Electrical Power Resources Survey and Development Administration (EIE) is also responsible for the observation of hydrologic events. Therefore, the stations which have measured by EIE are used with name of it.

Annual peak discharges are obtained from EIE and DSI annual flood books. The raw hydrologic data is taken from the annual flow observation books, which are published by DSI and EIE, presented in have been re-written to Excel and these hydrologic data were used in several analysis to conduct a well prepared Regional Flood Frequency Analysis solutions. Furthermore, all 73 stations data which are taken from DSI and EIE are given in Table 3.1, Figure 3.1, and the location of these stations are given in Figure 3.1

As one can see from Table 3.1, some stations have less than 10 year record duration. There are 26 stations which have a record duration less than 10 year are highlighted with yellow color in Table 3.1. Therefore, these stations are excluded from the list due to the inadequacy of record for flood calculations. In addition, 7 stations which are given with green color are installed downstream part of dam, weir, syphon etc. These stream gauging station records can be regulated and can have negative affect on the model results. To sum up, 33 stream gauging stations are not used due to lack of records and regulation. 40 stations which are shown in Table 3.2, Figure 3.2 & 3.3 with coordinates are used in model development and all other model parameters are determined for these 40 stream gauging stations.

Table 3.1 Stream gauging stations in Ceyhan Basin with their record periods.

| # | Station | Drainage Area (km$^2$) | Record Duration (yrs) | # | Station | Drainage Area (km$^2$) | Record Duration (yrs) |
|---|---|---|---|---|---|---|---|
| 1 | EIE/2001 | 8484 | 51 | 38 | DSI/20-17 | 1740 | 21 |
| 2 | EIE/2004 | 20466 | 30 | 39 | DSI/20-19 | 274.2 | 10 |
| 3 | EIE/2005 | 4219.08 | 35 | 40 | DSI/20-21 | 1103 | 5 |
| 4 | EIE/2006 | 733.2 | 47 | 41 | DSI/20-22 | 1056 | 5 |
| 5 | EIE/2007 | 623 | 41 | 42 | DSI/20-24 | 15700 | 8 |
| 6 | EIE/2008 | 444 | 32 | 43 | DSI/20-27 | 211.4 | 10 |
| 7 | EIE/2009 | 1387.2 | 46 | 44 | DSI/20-31 | 62 | 12 |
| 8 | EIE/2010 | 3498.8 | 30 | 45 | DSI/20-33 | 288.2 | 4 |
| 9 | EIE/2011 | 646 | 3 | 46 | DSI/20-34 | 727.7 | 1 |
| 10 | EIE/2012 | 19727.2 | 17 | 47 | DSI/20-35 | 13500 | 13 |
| 11 | EIE/2013 | 13840 | 8 | 48 | DSI/20-36 | 174.2 | 24 |
| 12 | EIE/2015 | 915.2 | 39 | 49 | DSI/20-40 | 79 | 18 |
| 13 | EIE/2016 | 846.8 | 15 | 50 | DSI/20-41 | 235 | 9 |
| 14 | EIE/2017 | 546.4 | 6 | 51 | DSI/20-42 | 274 | 14 |
| 15 | EIE/2018 | 245.2 | 3 | 52 | DSI/20-43 | 163 | 25 |
| 16 | EIE/2019 | 6248 | 9 | 53 | DSI/20-44 | 35 | 20 |
| 17 | EIE/2020 | 14708.4 | 35 | 54 | DSI/20-45 | 170 | 18 |
| 18 | EIE/2021 | 402 | 3 | 55 | DSI/20-46 | 477 | 26 |
| 19 | EIE/2022 | 428 | 28 | 56 | DSI/20-47 | 2787.3 | 6 |
| 20 | EIE/2024 | 434 | 2 | 57 | DSI/20-48 | | 12 |
| 21 | EIE/2025 | 914.7 | 5 | 58 | DSI/20-50 | 39.81 | 9 |
| 22 | EIE/2026 | 25.6 | 1 | 59 | DSI/20-51 | 131.4 | 15 |
| 23 | DSI/20-01 | 234.3 | 8 | 60 | DSI/20-52 | 23 | 17 |
| 24 | DSI/20-02 | 197.1 | 26 | 61 | DSI/20-53 | 178.5 | 12 |
| 25 | DSI/20-04 | 178 | 27 | 62 | DSI/20-54 | 207.5 | 14 |
| 26 | DSI/20-05 | 94 | 37 | 63 | DSI/20-55 | 111.6 | 14 |
| 27 | DSI/20-06 | 174.9 | 27 | 64 | DSI/20-56 | 238.4 | 11 |
| 28 | DSI/20-07 | 2084 | 35 | 65 | DSI/20-57 | 224.3 | 6 |
| 29 | DSI/20-08 | 131.1 | 35 | 66 | DSI/20-58 | 24.38 | 16 |
| 30 | DSI/20-09 | 635 | 3 | 67 | DSI/20-59 | 171.5 | 16 |
| 31 | DSI/20-10 | 217.3 | 19 | 68 | DSI/20-63 | 57.5 | 11 |
| 32 | DSI/20-11 | 407 | 3 | 69 | DSI/20-65 | 161 | 11 |
| 33 | DSI/20-12 | 356.7 | 6 | 70 | DSI/20-66 | 150 | 8 |
| 34 | DSI/20-13 | 105.1 | 39 | 71 | DSI/20-69 | 31.5 | 8 |
| 35 | DSI/20-14 | 310.5 | 11 | 72 | DSI/20-71 | 49.7 | 5 |
| 36 | DSI/20-15 | 189.7 | 20 | 73 | DSI/20-72 | 48.4 | 8 |
| 37 | DSI/20-16 | 291 | 34 | | | | |

| | |
|---|---|
| (yellow) | Stations which have less than 10 years record period are excluded |
| (green) | Stations which are on the downstream of a weir, dam, or small dam are excluded |

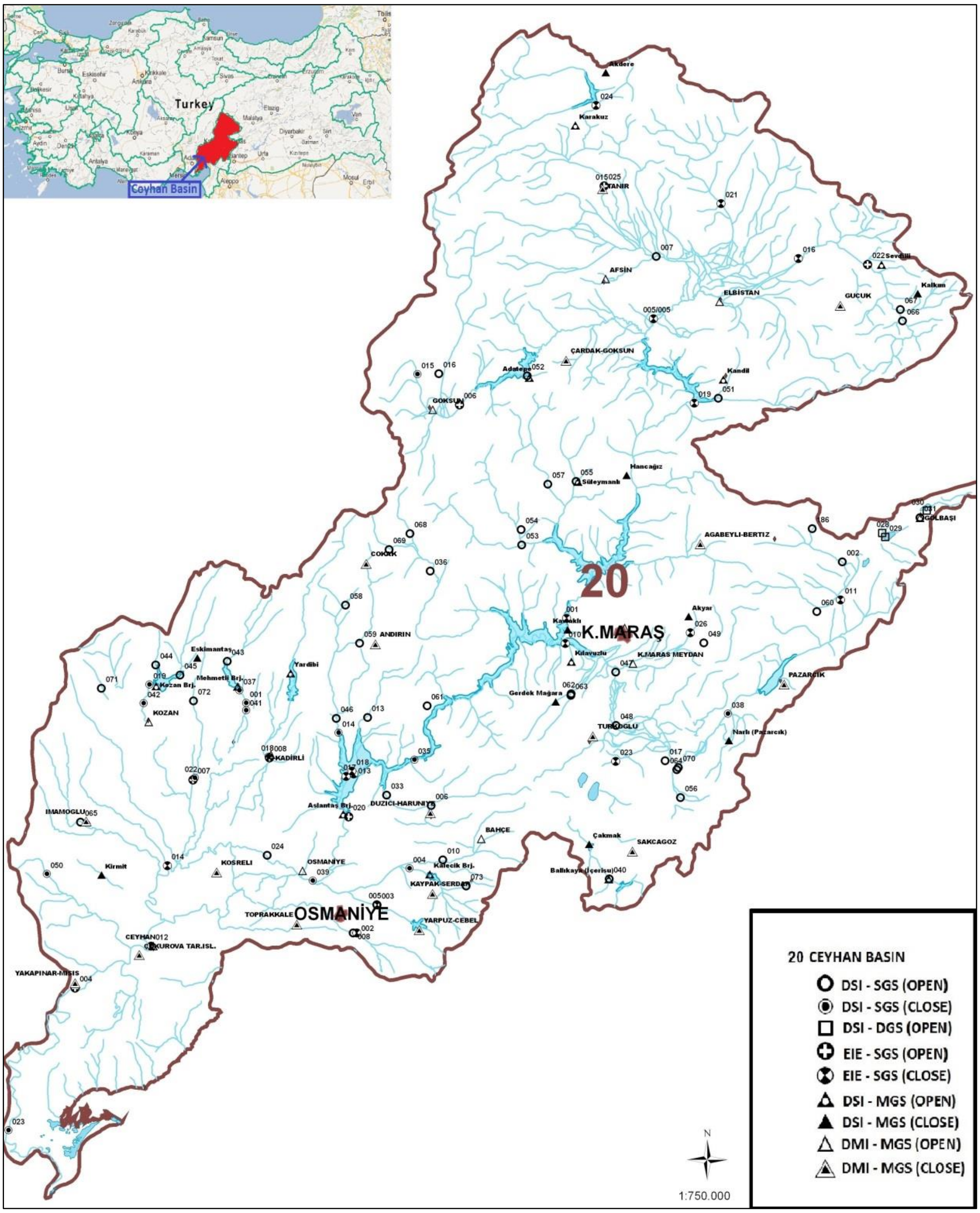


Figure 3.1 Hydrological and Meteorological stations in the Ceyhan Basin.

Table 3.2 Stations used in the model development with their drainage areas, record durations and coordinates.

| Station | Drainage Area ($km^2$) | Record Duration(yrs) | Coordinates | |
|---|---|---|---|---|
| | | | X | Y |
| **EIE/2001** | 8484 | 51 | 36.7983 | 37.6208 |
| **EIE/2004** | 20466 | 30 | 35.6336 | 36.9572 |
| **EIE/2005** | 4219.08 | 35 | 37.0264 | 38.1728 |
| **EIE/2006** | 733.2 | 47 | 36.5628 | 38.0239 |
| **EIE/2007** | 623 | 41 | 35.9510 | 37.3360 |
| **EIE/2008** | 444 | 32 | 36.0931 | 37.3722 |
| **EIE/2009** | 1387.2 | 46 | 36.9800 | 38.1650 |
| **EIE/2010** | 3498.8 | 30 | 36.7939 | 37.5744 |
| **EIE/2012** | 19727.2 | 17 | 35.8114 | 37.0328 |
| **EIE/2015** | 915.2 | 39 | 36.9206 | 38.4225 |
| **EIE/2016** | 846.8 | 15 | 37.3030 | 38.2430 |
| **EIE/2022** | 428 | 28 | 37.5356 | 38.2578 |
| **DSI/20-02** | 197.1 | 26 | 37.4516 | 37.7061 |
| **DSI/20-04** | 178 | 27 | 36.3966 | 37.1658 |
| **DSI/20-05** | 94 | 37 | 36.3364 | 37.0989 |
| **DSI/20-06** | 174.9 | 27 | 36.4701 | 37.2810 |
| **DSI/20-07** | 2084 | 35 | 37.1100 | 38.2690 |
| **DSI/20-08** | 131.1 | 35 | 36.2800 | 37.0485 |
| **DSI/20-10** | 217.3 | 19 | 36.4933 | 37.1778 |
| **DSI/20-13** | 105.1 | 39 | 36.3272 | 37.4478 |
| **DSI/20-14** | 310.5 | 11 | 36.2510 | 37.5620 |
| **DSI/20-15** | 189.7 | 20 | 36.4659 | 38.0678 |
| **DSI/20-16** | 291 | 34 | 36.5278 | 38.0842 |
| **DSI/20-17** | 1740 | 21 | 37.0238 | 37.3593 |
| **DSI/20-36** | 174.2 | 24 | 36.4885 | 37.7310 |
| **DSI/20-40** | 79 | 18 | 36.8770 | 37.1337 |
| **DSI/20-43** | 163 | 25 | 36.0033 | 37.5616 |
| **DSI/20-44** | 35 | 20 | 35.8363 | 37.5551 |
| **DSI/20-45** | 170 | 18 | 35.8934 | 37.5353 |
| **DSI/20-46** | 477 | 26 | 36.2541 | 37.4470 |
| **DSI/20-51** | 131.4 | 15 | 37.1640 | 38.0190 |
| **DSI/20-52** | 23 | 17 | 36.7246 | 38.0736 |
| **DSI/20-53** | 178.5 | 12 | 36.6070 | 37.7620 |
| **DSI/20-54** | 207.5 | 14 | 36.6560 | 37.8210 |
| **DSI/20-55** | 111.6 | 14 | 36.8305 | 37.8749 |
| **DSI/20-56** | 238.4 | 11 | 37.0440 | 37.3090 |
| **DSI/20-58** | 24.38 | 16 | 36.2710 | 37.6510 |
| **DSI/20-59** | 171.5 | 16 | 36.3133 | 37.5864 |
| **DSI/20-63** | 57.5 | 11 | 36.8000 | 37.4790 |
| **DSI/20-65** | 161 | 11 | 35.6210 | 37.2540 |

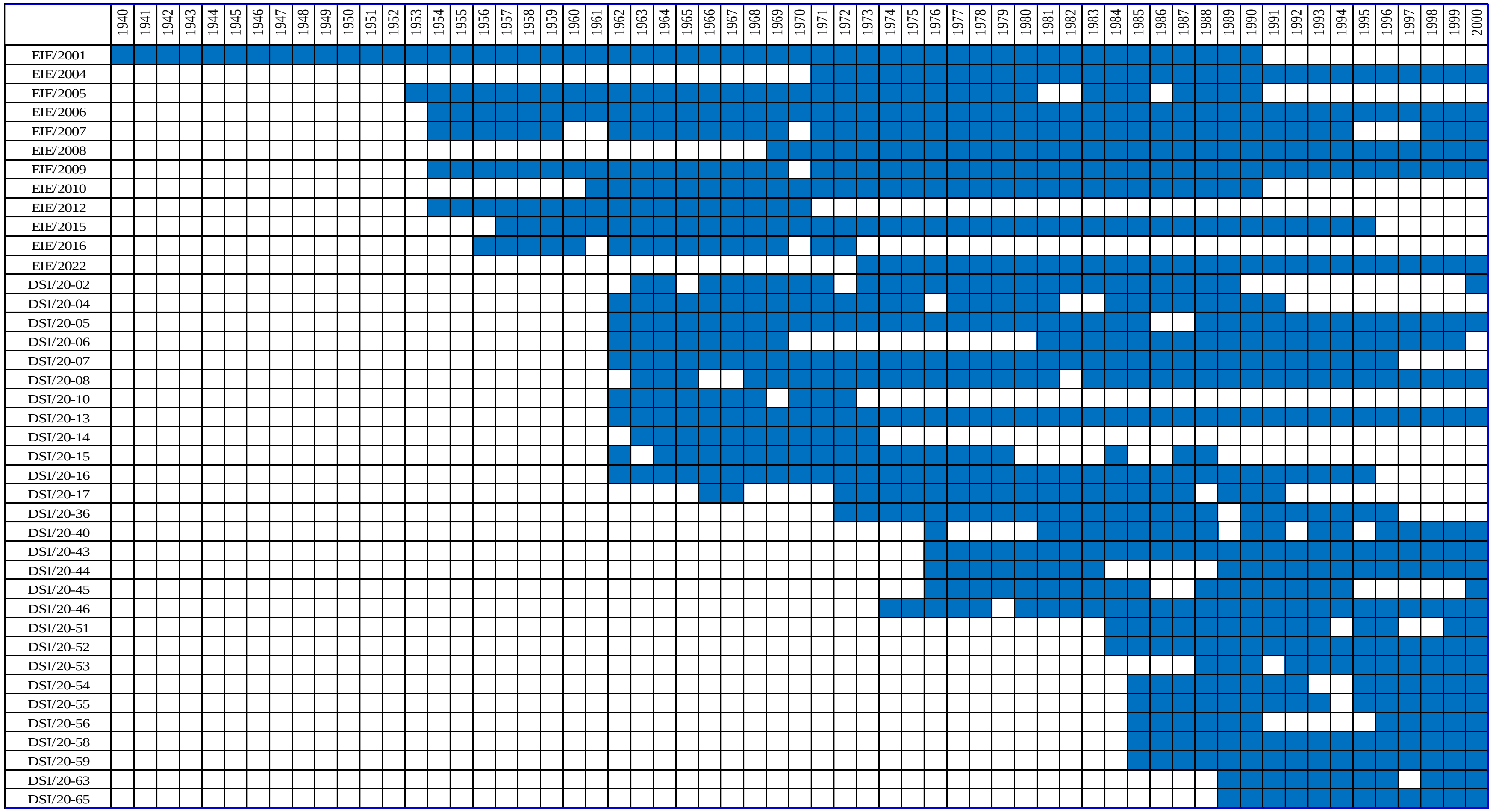


Figure 3.2 Record Period of Stations which are used in model development

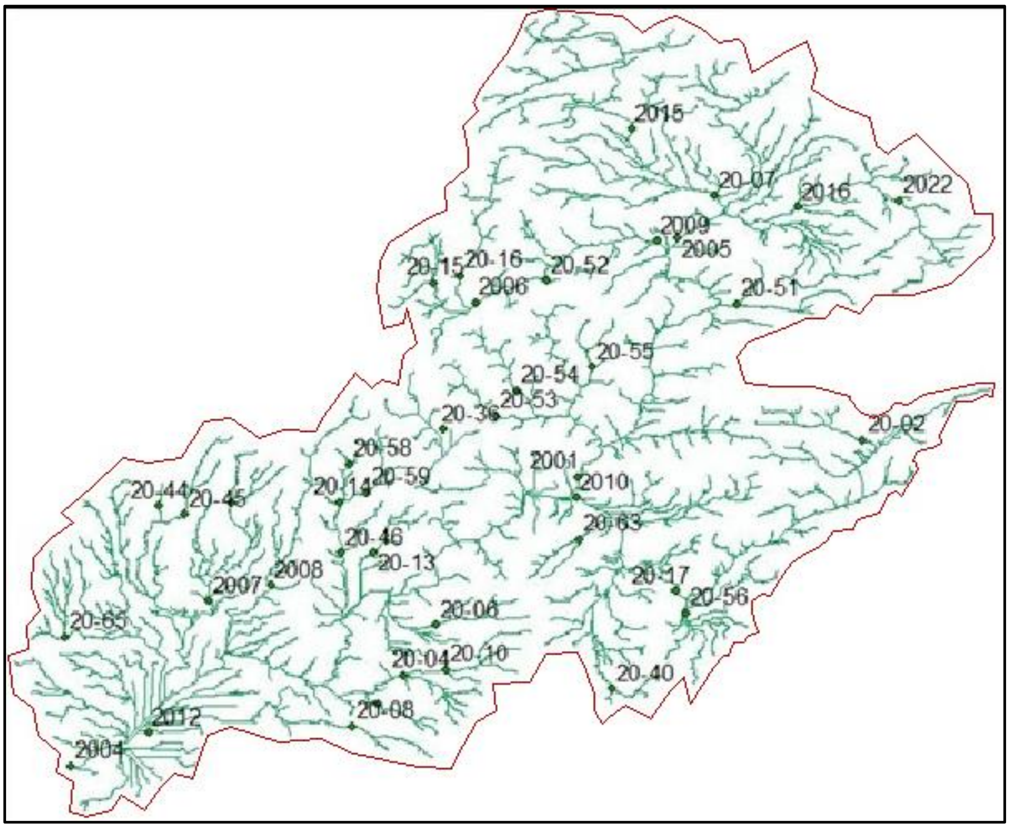


Figure 3.3 Station Locations and River Network in Ceyhan Basin.

### 3.2.2 Meteorological Data

Some properties of rainfall like duration and intensity must be known to do planning in water resources, agriculture, urbinatization, drainage, flood control and transportation. Moreover rainfall properties are needed to design/operate safe and economical engineering structures (Karahan, 2010). In this study, Mean Annual Rainfall calculations are gathered from Bostan et al,(2012). Universal kriging method is selected as the proper method in predicting the guided distributed precipitation values precipitation values for Ceyhan Basin. This map is shown in Figure 3.4.

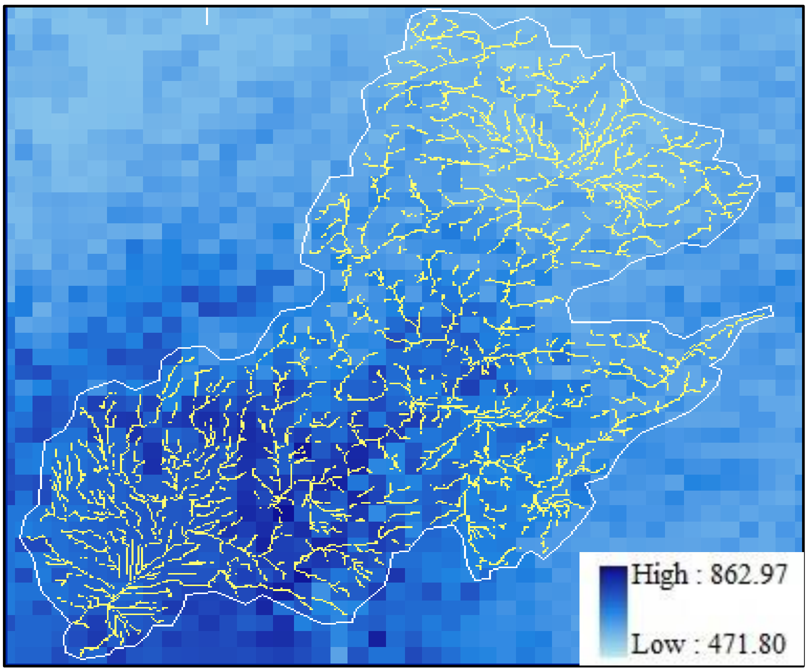


Figure 3.4 Mean Annual Rainfall Map of Ceyhan Basin which is developed by Kriging Method.(Bostan et all, 2012)

### 3.2.3 Topographic Data

The Shuttle Radar Topography Mission (SRTM) map is used in this study. DEM of the study area is shown in Figure 3.5.

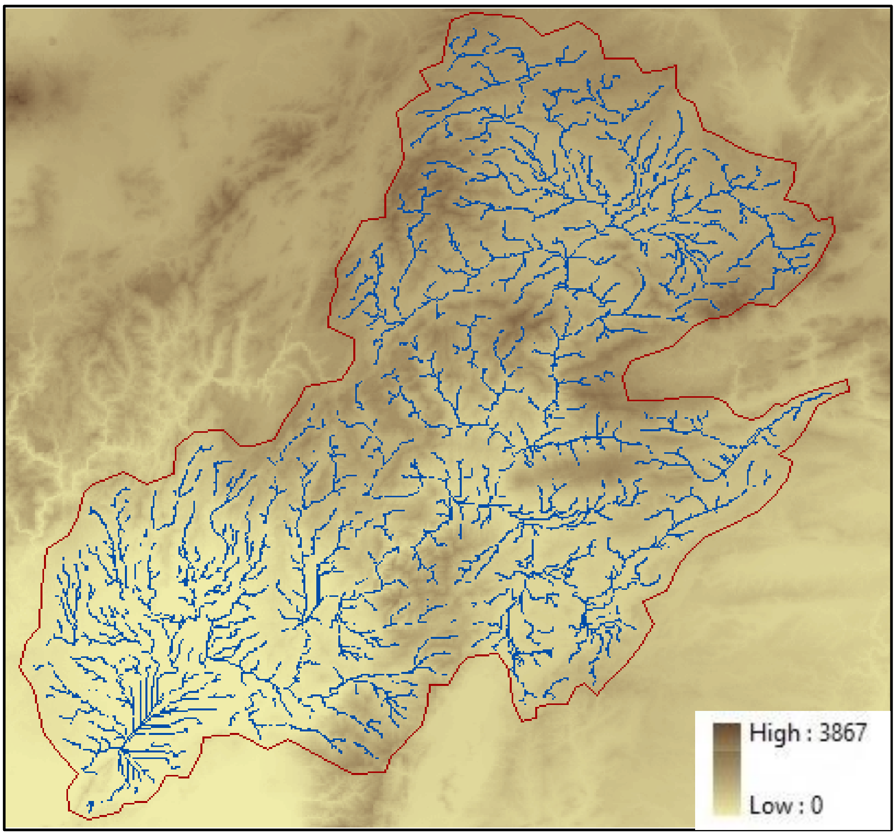


Figure 3.5 Digital Elevation Model and River Network of Ceyhan Basin.

By the help of Digital Elevation Model and River Network, one can determine the topographic values about the stations by using computer program tools. ArcHydro tool of ArcGIS has been used to create vector data of catchment and drainage lines. The following steps of obtaining catchment and drainage line are listed below:

1. ArcHydro (AH), Terrain Preprocessing (TP) → DEM Manipulation → Fill Sinks → Filled DEM
2. Filled DEM, AH → TP → Flow Direction
3. Flow Direction, AH → TP → Flow Accumulation
4. Flow Accumulation, AH → TP → Stream Definition → Stream (10000),10k
5. Stream, AH → TP → Stream Segmentation → Stream Link
6. Stream Link, AH → TP → Catchment Grid Delineation → Cat
7. Cat, AH → TP → Catchment Polygon Processing → Catchment Polygons (Catchment Boundaries)
8. AH → TP → Drainage Line Processing → Drainage Line

**3.3 Data Processing**

Excel, Autocad, ArcGIS, SPSS etc. were used to process and prepare the raw data for making analysis. According to the literature survey, these parameters are needed to perform the study:

- Drainage Area (A): gives an idea about the size of stations.
- Perimeter (P): gives the total measurement of drainage basin periphery length.
- Length of Main River (LMR): gives an idea about the shape of basin.
- Mean Annual Rainfall (MAR): gives the rainfall distribution in the stations.
- Minimum Elevation ($H_{min}$): gives an idea about location of stations.
- Maximum Elevation ($H_{max}$): gives an idea about the maximum point of stations
- Mean Elevation ($H_{mean}$): gives an idea about mean elevations of stations.
- Elevation Difference ($H_{\Delta}$): gives an idea about the elevation change in the stations.
- Basin Relief (BR): gives an idea about the minimum elevation and mean elevation of stations.
- Relative Relief (RR): gives an idea about the slope parameter about the stations.
- Circularity ($R_c$): gives an idea about the shape type of stations.
- Mean Annual Flood (MAF): gives an idea about the $Q_{2.33}$ flood value of stations.

Topographic, hydrologic and meteorological data are the 12 parameters which have been defined and generated by using appropriate software. The determined values are shown in Table 3.3.

Table 3.3 EIE-DSI Stream Gauging Stations of Ceyhan Basin and their parameters obtained through data processing

| Stream Gauging Station # | AREA ($km^2$) | Perimeter (km) | Length of Main River (km) | Annual Mean Rainfall (mm) | $H_{min}$ (m) | $H_{max}$ (m) | $H_{mean}$ (m) | $H_{\Delta}$ (m) | Basin Relief (m) | Relative Relief | Circularity | $Q_{2.33}$ ($m^3/s$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | Y |
| **EIE/2001** | 8484 | 742 | 209,83 | 540,72 | 420 | 3073 | 1554,32 | 2653 | 1134,32 | 0,1529 | 0,1936 | 552,95 |
| **EIE/2004** | 20466 | 1466 | 385,76 | 665,98 | 2 | 3073 | 1027,29 | 3071 | 1025,29 | 0,0699 | 0,1197 | 1027,82 |
| **EIE/2005** | 4219,08 | 516 | 110,55 | 473,06 | 1059 | 3073 | 1603,12 | 2014 | 544,12 | 0,1054 | 0,1991 | 111,42 |
| **EIE/2006** | 733,2 | 232 | 49,305 | 557,89 | 1313 | 2941 | 1828,41 | 1628 | 515,41 | 0,2222 | 0,1712 | 53,43 |
| **EIE/2007** | 623 | 196 | 59,58 | 830,54 | 7 | 1959 | 441,71 | 1952 | 434,71 | 0,2218 | 0,2038 | 59,86 |
| **EIE/2008** | 444 | 209 | 64,68 | 835,72 | 56 | 2244 | 908,44 | 2188 | 852,44 | 0,4079 | 0,1277 | 160,70 |
| **EIE/2009** | 1387,2 | 300 | 101,57 | 560,21 | 1102 | 2941 | 1672,16 | 1839 | 570,16 | 0,1901 | 0,1937 | 85,06 |
| **EIE/2010** | 3498,8 | 546 | 147,34 | 644,19 | 401 | 2436 | 871,73 | 2035 | 470,73 | 0,0862 | 0,1475 | 287,69 |
| **EIE/2012** | 19727,2 | 1417 | 371,98 | 662,04 | 2 | 3073 | 1048,80 | 3071 | 1046,80 | 0,0739 | 0,1235 | 1094,73 |
| **EIE/2015** | 915,2 | 277 | 62,62 | 471,80 | 1209 | 2851 | 1826,80 | 1642 | 617,80 | 0,2230 | 0,1499 | 43,95 |
| **EIE/2016** | 846,8 | 242 | 66,44 | 481,87 | 1153 | 3073 | 1640,36 | 1920 | 487,36 | 0,2014 | 0,1817 | 83,56 |
| **EIE/2022** | 428 | 149 | 37,98 | 517,23 | 1335 | 3073 | 1808,58 | 1738 | 473,58 | 0,3178 | 0,2423 | 68,93 |
| **DSI/20-02** | 197,1 | 99 | 27,93 | 555,61 | 889 | 2436 | 1473,67 | 1547 | 584,67 | 0,5906 | 0,2527 | 40,83 |
| **DSI/20-04** | 178 | 88 | 35,19 | 830,57 | 214 | 1953 | 903,84 | 1739 | 689,84 | 0,7839 | 0,2888 | 38,11 |
| **DSI/20-05** | 94 | 61 | 17,66 | 772,37 | 196 | 2070 | 1178,83 | 1874 | 982,83 | 1,6112 | 0,3175 | 45,57 |
| **DSI/20-06** | 174,9 | 80 | 21,32 | 800,25 | 340 | 2215 | 1184,68 | 1875 | 844,68 | 1,0558 | 0,3434 | 57,01 |
| **DSI/20-07** | 2084 | 340 | 92,92 | 474,08 | 1095 | 2887 | 1661,00 | 1792 | 566,00 | 0,1665 | 0,2265 | 54,54 |
| **DSI/20-08** | 131,1 | 72 | 21,03 | 774,20 | 211 | 2140 | 1283,14 | 1929 | 1072,14 | 1,4891 | 0,3178 | 39,33 |
| **DSI/20-10** | 217,3 | 95 | 25,77 | 790,71 | 486 | 1782 | 1045,02 | 1296 | 559,02 | 0,5884 | 0,3026 | 62,53 |
| **DSI/20-13** | 105,1 | 75 | 18,52 | 862,97 | 183 | 1602 | 840,45 | 1419 | 657,45 | 0,8766 | 0,2348 | 50,53 |

Table 3.3 EIE-DSI Stream Gauging Stations of Ceyhan Basin and their parameters obtained through data processing (Continued)

| Stream Gauging Station # | AREA ($km^2$) | Perimeter (km) | Length of Main River (km) | Annual Mean Rainfall (mm) | $H_{min}$ (m) | $H_{max}$ (m) | $H_{mean}$ (m) | $H_{\Delta}$ (m) | Basin Relief (m) | Relative Relief | Circularity | $Q_{2.33}$ ($m^3/s$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | Y |
| **DSI/20-14** | 310,5 | 126 | 27,93 | 725,30 | 417 | 2244 | 1269,95 | 1827 | 852,95 | 0,6769 | 0,2458 | 163,28 |
| **DSI/20-15** | 189,7 | 93 | 22,99 | 558,76 | 1350 | 2737 | 1785,93 | 1387 | 435,93 | 0,4687 | 0,2756 | 26,15 |
| **DSI/20-16** | 291 | 124 | 38,54 | 537,76 | 1363 | 2941 | 2092,48 | 1578 | 729,48 | 0,5883 | 0,2378 | 35,84 |
| **DSI/20-17** | 1740 | 479 | 103,63 | 594,89 | 478 | 2436 | 992,20 | 1958 | 514,20 | 0,1073 | 0,0953 | 121,32 |
| **DSI/20-36** | 174,2 | 102 | 19,52 | 626,09 | 1232 | 2340 | 1634,99 | 1108 | 402,99 | 0,3951 | 0,2104 | 62,79 |
| **DSI/20-40** | 79 | 63 | 14,51 | 716,34 | 617 | 1419 | 971,30 | 802 | 354,30 | 0,5624 | 0,2501 | 5,66 |
| **DSI/20-43** | 163 | 87 | 19,45 | 775,15 | 213 | 1959 | 1012,86 | 1746 | 799,86 | 0,9194 | 0,2706 | 68,16 |
| **DSI/20-44** | 35 | 33 | 5,94 | 735,88 | 300 | 1353 | 750,89 | 1053 | 450,89 | 1,3663 | 0,4039 | 28,96 |
| **DSI/20-45** | 170 | 90 | 25,75 | 750,63 | 274 | 1938 | 719,15 | 1664 | 445,15 | 0,4946 | 0,2637 | 67,52 |
| **DSI/20-46** | 477 | 146 | 40,65 | 758,38 | 171 | 2244 | 1086,26 | 2073 | 915,26 | 0,6269 | 0,2812 | 207,68 |
| **DSI/20-51** | 131,4 | 96 | 31,09 | 546,21 | 1099 | 3069 | 1824,14 | 1970 | 725,14 | 0,7554 | 0,1792 | 10,14 |
| **DSI/20-52** | 23 | 27 | 3,25 | 605,52 | 1275 | 2411 | 1629,32 | 1136 | 354,32 | 1,3123 | 0,3965 | 6,12 |
| **DSI/20-53** | 178,5 | 103 | 20,95 | 598,75 | 1060 | 2379 | 1655,54 | 1319 | 595,54 | 0,5782 | 0,2114 | 38,27 |
| **DSI/20-54** | 207,5 | 88 | 24,54 | 651,58 | 758 | 2470 | 1571,25 | 1712 | 813,25 | 0,9241 | 0,3367 | 58,87 |
| **DSI/20-55** | 111,6 | 74 | 17,11 | 601,50 | 830 | 3010 | 1906,45 | 2180 | 1076,45 | 1,4547 | 0,2561 | 21,06 |
| **DSI/20-56** | 238,4 | 145 | 25,47 | 645,65 | 488 | 1466 | 804,98 | 978 | 316,98 | 0,2186 | 0,1425 | 21,53 |
| **DSI/20-58** | 24,38 | 29 | 5,3 | 673,66 | 970 | 2171 | 1556,12 | 1201 | 586,12 | 2,0211 | 0,3643 | 18,38 |
| **DSI/20-59** | 171,5 | 97 | 20,12 | 697,23 | 1018 | 2244 | 1362,08 | 1226 | 344,08 | 0,3547 | 0,2291 | 62,49 |
| **DSI/20-63** | 57,5 | 60 | 14,62 | 694,28 | 530 | 1522 | 992,00 | 992 | 462,00 | 0,7700 | 0,2007 | 14,61 |
| **DSI/20-65** | 161 | 89 | 26,34 | 737,03 | 71 | 698 | 257,43 | 627 | 186,43 | 0,2095 | 0,2554 | 41,45 |

### 3.3.1. Drainage Area (A), Perimeter (P) and Length of Main River (LMR)

First one should obtain the coordinate values of all stream gauging stations in the basin. Than it is not a hard task to get area values of these basins because polygons represent the area values in ArcGIS. The polygon areas ($km^2$) have been calculated by using ArcGIS ArcHydro Tool which is explained in 3.2.3 Topographic Data section. DSI and EIEI stream gauging station area values are compared to the ones which have been obtained from GIS. By the results of this comparison, data processing handled in ArcGIS are meaningful because the observed data values are close to each other.

After the comparison of observed and real Drainage Area values, one can measure Perimeter (P) and Length of Main River (LMR) of all stations. All the results are also given in Table 3.3

### 3.3.2. Mean Annual Rainfall Parameter (MAR)

The annual rainfall map which is explained in 3.2.2 section presented in Figure 3.3, has been used to derive the average rainfall values in the subcatchments of Ceyhan Basin by using ArcGIS zonal statistics tools. The mean annual rainfall MAR (mm) is given in Table 3.3.

### 3.3.3. Mean Annual Flood (MAF)

The meaning of mean annual flood is clearly explained in Chapter 2. The mean annual flood correspond to the $Q_{2.33}$ flood value of each stream gauging station. $Q_{2.33}$ flood value is between $Q_2$ and $Q_5$ flood values. Therefore one could find $Q_{2.33}$ flood value by using the calculation method of each distribution. For example, $Q_{2.33}$ flood values calculation for Gumbel Distribution are calculated by using Gumbel Distribution by calculating for 2.33 years flood value. $Q_{2.33}$ values which are obtained by using Gumbel distribution are also given in Table 3.3. $Q_{2.33}$ flood values which are obtained by Different Distributions are also calculated its related distribution equation.

$Q_{2.33}$ flood values can be calculated by using Excel Forecast formula with is using related stream gauging station's $Q_2$ and $Q_5$ values. This method is also used to calculate $Q_{2.33}$ flood values. However, this method do not give accurate values, when comparing the results of this method with real distribution values.

### 3.3.4 Elevation Parameters ($H_{min}$, $H_{max}$, $H_{mean}$, $H_{\Delta}$, BR)

By using the Digital Elevation Map of Ceyhan Basin, all the elevation parameters can be calculated easily. Because, the drainage areas of stream gauging stations are already calculated and can be used to find Elevation Parameters by using Zonal statistic tool of ArcGIS. $H_{min}$, $H_{max}$ and $H_{mean}$ parameters can be found directly after the usage of Zonal statistics tool. $H_{\Delta}$ and BR are not found directly from the tool, but these parameters can also have an effect on modeling. Therefore, one can calculate $H_{\Delta}$ and BR by using these formulas:

$$H_{\Delta} = H_{max} - H_{min} \qquad \text{(Equation 3.2)}$$

$$BR = H_{mean} - H_{min} \qquad \text{(Equation 3.3)}$$

All the calculated elevation parameters are presented in Table 3.3.

### 3.3.4 Relative Relief (RR) and Circularity($R_c$)

Drainage areas can have infinite variety of shapes, and it is hard to explain the drainage area type with using formulas. However, one can use these two parameters to give an idea about the shapes of drainage areas. First, Relative Relief(RR) is a ratio between Basin Relief(BR) and Perimeter(P). If this ratio is more higher, the slope parameter of the basin is higher. Basin Relief(BR) is calculated as elevation parameters and also Perimeter(P) is determined by using ArcGIS. Therefore the Relative Relief(RR) can be calculated by according to that formula:

$$RR = BR*0.1/P \qquad \text{(Equation 3.4)}$$

Secondly, Circularity($R_c$) is also an important measure to learn something about the characteristics of basin. If the number is closer to 1, it means that basin has a circular characteristics. Potential of circular or elongated basins are very different from each other due to floods. Therefore, one can calculate Circularity($R_c$) by using Drainage Area(A) and Perimeter(P) according to that formula:

$$R_c = (4\pi * A)/P^2 \qquad \text{(Equation 3.5)}$$

These two parameters are also shown in the Table 3.3 for all stream gauging stations.

# CHAPTER 4

# MODEL DEVELOPMENT

To make RFFA, one should obtain Flood Frequencies for different distribution parameters. Gumbel Distribution is offered by Dalyrmple (1960), but other distribution parameters are also used in Turkey by DSI. Therefore, outline of this chapter starts with calculation of Flood Frequency for each station for different distribution parameters. Then Point Flood Frequency Analysis and Regional Flood Frequency Analysis is done with the method which is currently used in Turkey by DSI. Then, to obtain more reliable results for the whole basin, Twostep Clustering technique is applied to the basin for grouping homogeneous regions. After grouping the stations, Stepwise Linear Regression Analysis and Nonlinear Regression Analysis applied for all these stations. To compare the results one should obtain the Root Mean Square Errors, Nash-Sutcliffe Efficiency Index and percentage errors of the results.

**4.1 Distribution Parameters**

Suppose that observations are made at regular intervals at some site of interest. Let Q be the magnitude of the event that occurs at a given site. We regard Q as a random quantity (a random variable), potentially taking any value between zero and infinity. The fundamental quantity of statistical frequency analysis is the frequency distribution, which specifies how frequently the possible values of Q occur. Denote by F(x) the probability that the actual value of Q is at most x:

$$F(x) = Pr\ [\ Q \leq x\ ] \qquad \text{(Equation 4.1)}$$

F(x) is the cumulative distribution function of the frequency distribution. Its inverse function x(F), the quantile function of the frequency distribution, expresses the magnitude of an event in terms of its nonexceedance probability F. The quantile of return period T, $Q_T$, is an event magnitude so extreme that it has probability 1/T of being exceeded by any single event. For an extreme high event, in the upper tail of the frequency distribution, $Q_T$ is given by

$$Q_T = x(1 - 1/T) \qquad \text{(Equation 4.2)}$$

Or

$$F(Q_T) = 1 - 1/T \qquad \text{(Equation 4.3)}$$

For an extreme low event, in the lower tail of the frequency distribution the corresponding relations are $Q_T$= x(1/T) and F($Q_T$) = 1/ T. The goal of frequency analysis is to obtain a useful estimate of the quantile $Q_T$ for a return period of scientific relevance. This period may be the design life of a structure (T=50 years, say) or some legally mandated design period (e.g., T= 10000 years in some dam safety applications). More generally, the goal may be to estimate $Q_T$ for a range of return periods or to estimate the entire quantile function. To be “useful” an estimate should not only be close to the true quantile but should also come with an assessment of how accurate it is likely to be.

There are many distribution parameters, but in Regional Flood Frequency Analysis Method which is obtained by Dalyrmple (1960) only Gumbel Distribution parameter is used to obtain flood values. Therefore, Gumbel distribution parameters should be determined to make Frequency analysis by using the equation which is shown in Table 4.1.

Moreover, different distribution parameters are used to make Regional Flood Frequency Analysis in Turkey for DSI projects. Three Parameter Lognormal, Pearson Type III, Log-pearson Type III, Two Parameter Lognormal, Normal are the distributions which are used an option for Gumbel distribution. All these parameters have been tested by Kolmogorov-Smirnov Test and the best distribution fits to the related station is obtained. Parameters for all distribution functions of Probability Density Functions are given in Table 4.1.

### 4.1.1 Kolmogorov-Smirnov Test

In order to avoid the loss of information due to grouping suffered by the chi-square test, other tests of goodness of fit have been developed such as the Neyman-Barton "smooth" tests, and the Cramer-Von Mises $W^2$ test. The most important of these alternatives to chi-square is the Kolmogorov statistic, D, which is based on deviations of the sample distribution function P(x) from the completely specified continuous hypothetical distribution function $P_0(x)$, such that:

$$D_n = \text{Max} \mid F(x) - P_0(x) \mid \qquad \text{(Equation 4.4)}$$

Developed by Kolmogrov in 1933, the test requires that the value of $D_n$ computed from the sample distribution be less than the tabulated value of $D_n$ at the required confidence level.

In practice the values P(x) are obtained as nj/n where nj is cumulative number of sample events at class limit j and n is the total number of events. P0(x) is then 1/k, 2/k, etc. Where k is the number of class intervals. Given the %95 and %90 critical values of 0.18 and 0.16, for a sample size of 60, it can be seen that all distributions are well within the acceptance limits. On the basis of the Kolmogorov-Smirnov test the preferred order of distributions would be Three parameter Lognormal, Pearson Type III, Log-pearson Type III, Two Parameter Lognormal, Gumbel and Normal.

Table 4.1 Distributions and Probability Density Functions which are used in Turkey for Regional Flood Frequency Analysis.

| # | Distributions | Probability Density Functions |
|---|---|---|
| 1 | Gumbel | $f(x) = e^{-x} e^{-e^{(-x)}}$ |
| 2 | Normal | $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ |
| 3 | Log-Normal (Two Parameters) | $f_x(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(lnx - \mu)^2}{2\sigma^2}}$ |
| 4 | Log-Normal (Three Parameters) | $f(x; \mu, \sigma, \gamma) = \frac{1}{(x - \gamma)\sigma\sqrt{2\pi}} exp\left\{-\frac{[\ln(x-\gamma) - \mu]^2}{2\sigma^2}\right\}$ |
| 5 | Pearson Type III | $f(x) = \frac{1}{\beta\Gamma(p)} \left(\frac{x-a}{\beta}\right)^{p-1} exp\left(-\frac{x-a}{\beta}\right)$ |
| 6 | Log-pearson Type III | $f(x) = \frac{1}{\beta\Gamma(p)} \left(\frac{x-a}{\beta}\right)^{p-1} exp\left(-\frac{x-a}{\beta}\right)$ , a = 0 |

Calculation and testing of these different distribution functions are done by using a program which is generally used by DSI. The annual maximum flood events for related years are entered to the program and any desired calculation is done by using Macro Programming in the program. The data entering part and the result parts of the program is shown in Figure 4.1 and Figure 4.2, respectively.

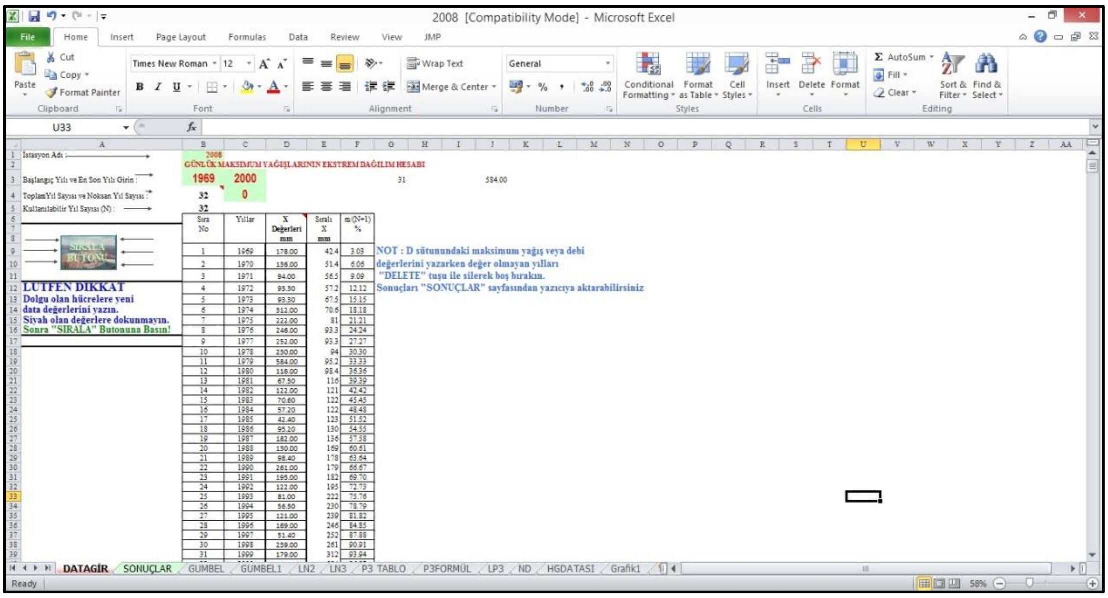

Figure 4.1 Data Enter Page of Distribution Parameters Program

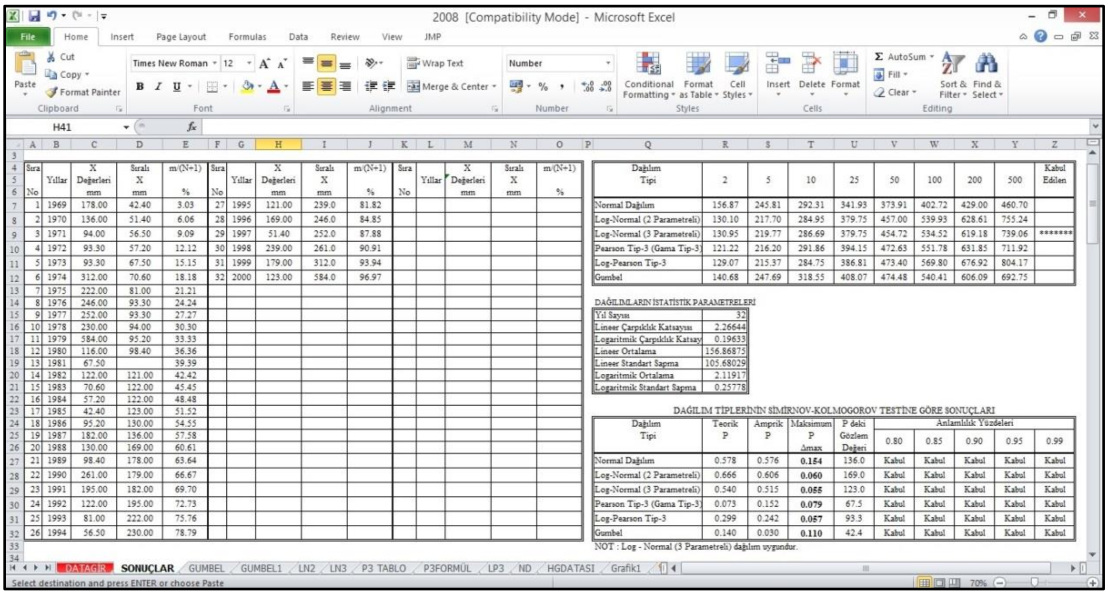

Figure 4.2 Results Page of Distribution Parameters Program.

### 4.1.2 Distribution Parameters Calculation

Gumbel Distribution is the main method for our research because of Dalyrmple (1960) offered to use this distribution in Regional Flood Frequency Analysis. However, Three Parameter Lognormal, Pearson Type III, Log-pearson Type III, Two Parameter Lognormal and Normal distributions are also used distributions in Turkey by DSI. Therefore, Gumbel Distribution results are found for all stream gauging stations and they are given in Table 4.2. Then, other distribution parameters are also calculated for all stations, but only the one has the best result in Kolmogorov-Smirnov test is selected as the resultant distributions and given in Table 4.3.

Table 4.2 Flood Frequency Values with Gumbel Distribution for Ceyhan Basin Stations

| Station | Drainage Area (km$^2$) | Record Duration (yrs) | Distribution Parameter Type | $Q_2$ | $Q_5$ | $Q_{10}$ | $Q_{25}$ | $Q_{50}$ | $Q_{100}$ |
|---|---|---|---|---|---|---|---|---|---|
| **EIE/2001** | 8484 | 51 | Gumbel | 502.65 | 771.46 | 949.44 | 1174.31 | 1341.13 | 1506.72 |
| **EIE/2004** | 20466 | 30 | Gumbel | 944.47 | 1389.91 | 1684.83 | 2057.46 | 2333.89 | 2608.29 |
| **EIE/2005** | 4219.08 | 35 | Gumbel | 102.41 | 150.56 | 182.44 | 222.72 | 252.60 | 282.26 |
| **EIE/2006** | 733.2 | 47 | Gumbel | 48.34 | 75.55 | 93.56 | 116.32 | 133.21 | 149.97 |
| **EIE/2007** | 623 | 41 | Gumbel | 55.50 | 78.83 | 94.27 | 113.79 | 128.27 | 142.64 |
| **EIE/2008** | 444 | 32 | Gumbel | 140.68 | 247.69 | 318.55 | 408.07 | 474.48 | 540.41 |
| **EIE/2009** | 1387.2 | 46 | Gumbel | 75.60 | 126.16 | 159.64 | 201.93 | 233.30 | 264.45 |
| **EIE/2010** | 3498.8 | 30 | Gumbel | 258.95 | 412.56 | 514.26 | 642.76 | 738.09 | 832.72 |
| **EIE/2012** | 19727.2 | 17 | Gumbel | 1024.20 | 1401.12 | 1650.68 | 1965.99 | 2199.91 | 2432.10 |
| **EIE/2015** | 915.2 | 39 | Gumbel | 37.76 | 70.84 | 92.74 | 120.41 | 140.93 | 161.31 |
| **EIE/2016** | 846.8 | 15 | Gumbel | 72.11 | 133.28 | 173.79 | 224.96 | 262.93 | 300.61 |
| **EIE/2022** | 428 | 28 | Gumbel | 59.17 | 111.32 | 145.84 | 189.47 | 221.83 | 253.95 |
| **DSI/20-02** | 197.1 | 26 | Gumbel | 35.55 | 63.72 | 82.37 | 105.94 | 123.42 | 140.77 |
| **DSI/20-04** | 178 | 27 | Gumbel | 31.16 | 68.32 | 92.93 | 124.02 | 147.08 | 169.97 |
| **DSI/20-05** | 94 | 37 | Gumbel | 38.77 | 75.11 | 99.16 | 129.55 | 152.10 | 174.48 |
| **DSI/20-06** | 174.9 | 27 | Gumbel | 51.57 | 80.62 | 99.86 | 124.16 | 142.19 | 160.09 |
| **DSI/20-07** | 2084 | 35 | Gumbel | 43.93 | 100.65 | 138.21 | 185.66 | 220.86 | 255.80 |
| **DSI/20-08** | 131.1 | 35 | Gumbel | 32.43 | 69.31 | 93.72 | 124.57 | 147.45 | 170.17 |
| **DSI/20-10** | 217.3 | 19 | Gumbel | 57.74 | 83.34 | 100.29 | 121.71 | 137.60 | 153.38 |
| **DSI/20-13** | 105.1 | 39 | Gumbel | 45.23 | 73.57 | 92.34 | 116.06 | 133.65 | 151.11 |
| **DSI/20-14** | 310.5 | 11 | Gumbel | 143.49 | 249.23 | 319.23 | 407.68 | 473.30 | 538.43 |
| **DSI/20-15** | 189.7 | 20 | Gumbel | 20.73 | 49.73 | 68.94 | 93.20 | 111.21 | 129.07 |
| **DSI/20-16** | 291 | 34 | Gumbel | 27.82 | 70.65 | 99.01 | 134.84 | 161.43 | 187.81 |
| **DSI/20-17** | 1740 | 21 | Gumbel | 99.85 | 214.58 | 290.53 | 386.51 | 457.70 | 528.38 |
| **DSI/20-36** | 174.2 | 24 | Gumbel | 55.45 | 94.66 | 120.62 | 153.42 | 177.76 | 201.91 |
| **DSI/20-40** | 79 | 18 | Gumbel | 4.89 | 9.04 | 11.79 | 15.27 | 17.85 | 20.41 |
| **DSI/20-43** | 163 | 25 | Gumbel | 57.94 | 112.54 | 148.69 | 194.36 | 228.24 | 261.87 |
| **DSI/20-44** | 35 | 20 | Gumbel | 23.56 | 52.46 | 71.59 | 95.77 | 113.70 | 131.51 |
| **DSI/20-45** | 170 | 18 | Gumbel | 58.88 | 105.04 | 135.61 | 174.23 | 202.88 | 231.31 |
| **DSI/20-46** | 477 | 26 | Gumbel | 184.21 | 309.62 | 392.65 | 497.56 | 575.38 | 652.63 |
| **DSI/20-51** | 131.4 | 15 | Gumbel | 8.07 | 19.16 | 26.51 | 35.79 | 42.67 | 49.51 |
| **DSI/20-52** | 23 | 17 | Gumbel | 5.06 | 10.72 | 14.46 | 19.20 | 22.71 | 26.20 |
| **DSI/20-53** | 178.5 | 12 | Gumbel | 34.64 | 54.06 | 66.93 | 83.18 | 95.24 | 107.21 |
| **DSI/20-54** | 207.5 | 14 | Gumbel | 51.67 | 90.17 | 115.67 | 147.88 | 171.78 | 195.50 |
| **DSI/20-55** | 111.6 | 14 | Gumbel | 19.37 | 28.40 | 34.39 | 41.94 | 47.55 | 53.12 |
| **DSI/20-56** | 238.4 | 11 | Gumbel | 19.81 | 29.04 | 35.16 | 42.89 | 48.62 | 54.31 |
| **DSI/20-58** | 24.38 | 16 | Gumbel | 16.31 | 27.38 | 34.70 | 43.96 | 50.83 | 57.65 |
| **DSI/20-59** | 171.5 | 16 | Gumbel | 53.21 | 102.84 | 135.69 | 177.21 | 208.01 | 238.58 |
| **DSI/20-63** | 57.5 | 11 | Gumbel | 12.25 | 24.88 | 33.25 | 43.81 | 51.65 | 59.43 |
| **DSI/20-65** | 161 | 11 | Gumbel | 37.37 | 59.15 | 73.58 | 91.80 | 105.32 | 118.74 |

Table 4.3 Selected Distribution Parameter Results After Kolmogorov-Smirnov Test.

| Station | Drainage Area ($km^2$) | Record Duration (yrs) | Distribution Parameter Type | $Q_2$ | $Q_5$ | $Q_{10}$ | $Q_{25}$ | $Q_{50}$ | $Q_{100}$ |
|---|---|---|---|---|---|---|---|---|---|
| **EIE/2001** | 8484 | 51 | Three Parameter Log-Normal | 492.03 | 735.23 | 901.94 | 1118.18 | 1282.53 | 1449.96 |
| **EIE/2004** | 20466 | 30 | Normal Distribution | 1011.17 | 1379.08 | 1571.43 | 1776.68 | 1909.01 | 2028.18 |
| **EIE/2005** | 4219.08 | 35 | Two Parameter Log-Normal | 100.62 | 143.01 | 171.87 | 209.11 | 237.29 | 265.91 |
| **EIE/2006** | 733.2 | 47 | Three Parameter Log-Normal | 48.78 | 73.60 | 89.23 | 108.29 | 122.04 | 135.51 |
| **EIE/2007** | 623 | 41 | Two Parameter Log-Normal | 54.96 | 75.89 | 89.84 | 107.56 | 120.79 | 134.10 |
| **EIE/2008** | 444 | 32 | Three Parameter Log-Normal | 130.95 | 219.77 | 286.69 | 379.75 | 454.72 | 534.52 |
| **EIE/2009** | 1387.2 | 46 | Three Parameter Log-Normal | 71.78 | 115.90 | 148.13 | 191.93 | 226.56 | 262.88 |
| **EIE/2010** | 3498.8 | 30 | Three Parameter Log-Normal | 260.03 | 394.91 | 480.45 | 585.31 | 661.31 | 736.02 |
| **EIE/2012** | 19727.2 | 17 | Pearson Type III | 1056.39 | 1358.98 | 1527.47 | 1715.31 | 1841.42 | 1957.59 |
| **EIE/2015** | 915.2 | 39 | Two Parameter Log-Normal | 33.93 | 60.41 | 81.68 | 112.69 | 138.67 | 167.17 |
| **EIE/2016** | 846.8 | 15 | Two Parameter Log-Normal | 65.90 | 111.31 | 146.41 | 196.14 | 236.84 | 280.68 |
| **EIE/2022** | 428 | 28 | Log-Pearson Type III | 50.51 | 91.64 | 128.90 | 189.61 | 246.48 | 314.74 |
| **DSI/20-02** | 197.1 | 26 | Two Parameter Log-Normal | 32.70 | 55.18 | 72.53 | 97.11 | 117.21 | 138.86 |
| **DSI/20-04** | 178 | 27 | Log-Pearson Type III | 26.25 | 56.25 | 80.31 | 113.76 | 140.16 | 167.32 |
| **DSI/20-05** | 94 | 37 | Pearson Type III | 33.16 | 66.42 | 91.67 | 125.11 | 150.42 | 175.73 |
| **DSI/20-06** | 174.9 | 27 | Log-Pearson Type III | 49.54 | 74.94 | 93.35 | 118.24 | 137.89 | 158.50 |
| **DSI/20-07** | 2084 | 35 | Log-Pearson Type III | 38.06 | 72.55 | 103.08 | 151.48 | 195.21 | 246.33 |
| **DSI/20-08** | 131.1 | 35 | Log-Pearson Type III | 26.67 | 52.37 | 76.65 | 117.50 | 156.77 | 204.80 |
| **DSI/20-10** | 217.3 | 19 | Gumbel Distribution | 57.74 | 83.34 | 100.29 | 121.71 | 137.60 | 153.38 |
| **DSI/20-13** | 105.1 | 39 | Log-Pearson Type III | 44.19 | 71.78 | 89.60 | 110.92 | 125.80 | 139.80 |
| **DSI/20-14** | 310.5 | 11 | Log-Pearson Type III | 141.05 | 234.54 | 292.17 | 357.48 | 400.63 | 439.29 |
| **DSI/20-15** | 189.7 | 20 | Log-Pearson Type III | 15.56 | 33.03 | 51.63 | 86.78 | 124.21 | 174.34 |
| **DSI/20-16** | 291 | 34 | Log-Pearson Type III | 22.09 | 43.49 | 66.03 | 108.40 | 153.70 | 214.49 |
| **DSI/20-17** | 1740 | 21 | Log-Pearson Type III | 78.81 | 181.84 | 270.65 | 402.10 | 511.37 | 627.99 |
| **DSI/20-36** | 174.2 | 24 | Gumbel Distribution | 55.45 | 94.66 | 120.62 | 153.42 | 177.76 | 201.91 |
| **DSI/20-40** | 79 | 18 | Log-Pearson Type III | 4.39 | 7.83 | 10.57 | 14.52 | 17.81 | 21.38 |
| **DSI/20-43** | 163 | 25 | Log-Pearson Type III | 47.10 | 93.51 | 136.69 | 208.46 | 276.46 | 358.41 |
| **DSI/20-44** | 35 | 20 | Log-Pearson Type III | 17.99 | 40.70 | 62.55 | 99.05 | 133.43 | 174.53 |
| **DSI/20-45** | 170 | 18 | Log-Pearson Type III | 51.86 | 90.21 | 123.00 | 173.91 | 219.55 | 272.26 |
| **DSI/20-46** | 477 | 26 | Three Parameter Log-Normal | 178.58 | 285.46 | 358.97 | 454.56 | 527.36 | 601.64 |
| **DSI/20-51** | 131.4 | 15 | Log-Pearson Type III | 8.19 | 14.40 | 20.26 | 30.25 | 40.01 | 52.17 |
| **DSI/20-52** | 23 | 17 | Two Parameter Log-Normal | 4.34 | 8.26 | 11.57 | 16.57 | 20.90 | 25.75 |
| **DSI/20-53** | 178.5 | 12 | Two Parameter Log-Normal | 33.65 | 48.51 | 58.73 | 72.01 | 82.13 | 92.46 |
| **DSI/20-54** | 207.5 | 14 | Log-Pearson Type III | 49.52 | 81.35 | 102.48 | 128.60 | 147.51 | 165.95 |
| **DSI/20-55** | 111.6 | 14 | Gumbel Distribution | 19.37 | 28.40 | 34.39 | 41.94 | 47.55 | 53.12 |
| **DSI/20-56** | 238.4 | 11 | Normal Distribution | 20.89 | 27.53 | 31.00 | 34.70 | 37.09 | 39.24 |
| **DSI/20-58** | 24.38 | 16 | Gumbel Distribution | 16.31 | 27.38 | 34.70 | 43.96 | 50.83 | 57.65 |
| **DSI/20-59** | 171.5 | 16 | Log-Pearson Type III | 49.10 | 89.04 | 118.13 | 156.47 | 185.55 | 214.64 |
| **DSI/20-63** | 57.5 | 11 | Two Parameter Log-Normal | 10.80 | 19.36 | 26.26 | 36.35 | 44.84 | 54.16 |
| **DSI/20-65** | 161 | 11 | Three Parameter Log-Normal | 36.44 | 52.91 | 64.06 | 78.38 | 89.18 | 100.11 |

### 4.2 Regional Flood Frequency Analysis with Dalyrmple Method

#### 4.2.1 Methodology of RFFA

The basic idea behind the index flood method is to increase the reliability of the frequency characteristics within a region. If, within a hydrologically homogeneous area, a number of hydrometric stations have been operating and recording the effects of the same meteorological factors then a combination of these records will provide, not a longer record, but a more reliable record. The following brief description of the index flood method does not include all the computational details of the procedure. These can be obtained, if required, from Dalrymple (1960).

Firstly, the data sets available within a region are listed, unsuitable stations eliminated, and a common period of record selected. Generally stations having less than 5 years of record of gauging and all regulated or controlled streams are excluded. A bar graph showing the period of record of each gauge is useful in determining which base period to use. The base period should be planned so as to include the maximum information content i.e. maximum number of station-years. Missing data points may be filled in by inter-station correlations. Data points filled in this way are not used directly but only as aids in assigning representative return periods to the recorded events.

The index-flood method next computes return periods, T, for each recorded event for each station in the region using the equation:

$$T = \frac{n+1}{m} \qquad \text{(Equation 4.5)}$$

Where n is the sample size and m is the order number of event; m=1 for the maximum event and m=n for the minimum event. For each station a graph of T versus event magnitude is plotted and a smooth curve drawn through the points. No attempt is made to force a straight line to fit any mathematical distribution. The mean annual event for the station is then picked from the smooth curve at the point T=2.33 . This is a theoretical result taken from the Gumbel distribution. Benson (1962) has confirmed experimentally that the mean annual event does occur with a return period of 2.33 years. It is preferred in the index flood method to derive the mean annual event graphically rather than arithmetically.

Dalrymple (1960) has described a test which should be used at this stage of index flood procedure to check for regional hydrologic homogeneity. The standard error of estimate of the reduced variable, y, in a type Gumbel distribution is given by:

$$\sigma_y = \frac{e^y}{\sqrt{n}}\sqrt{\frac{1}{T-1}} \qquad \text{(Equation 4.6)}$$

Then, assuming a normal distribution of the estimates, 95 % of the estimates will lie within $\pm\sigma_y$ of the most probable value. If T, the return period of the estimate, is taken as 10 years, then

$$2\sigma_y = \frac{0.666e^y}{\sqrt{n}}\pi r^2 \qquad \text{(Equation 4.7)}$$

Since for T=10 the reduced variable in a type Gumbel distribution is 2.25 then confidence limits are given by

$$2.25 \pm 6.33/\sqrt{n} \qquad \text{(Equation 4.8)}$$

Dalrymple (1960), gives the upper and lower confidence limits with the corresponding return periods for various values of n.

Table 4.4 Confidence Limits for Index Flood Homogeneity Test

| **Sample size n** | **Lower y-2 $\sigma_y$** | **Limit $T_L$** | **Upper y+2 $\sigma_y$** | **Limit $T_U$** |
|---|---|---|---|---|
| **5** | -0.59 | 1.2 | 5.09 | 160 |
| **10** | 0.25 | 1.8 | 4.25 | 70 |
| **20** | 0.83 | 2.8 | 3.67 | 40 |
| **50** | 1.35 | 4.4 | 3.15 | 24 |
| **100** | 1.62 | 5.6 | 2.88 | 18 |
| **200** | 1.80 | 6.5 | 2.70 | 15 |
| **500** | 1.97 | 7.7 | 2.53 | 13 |
| **1000** | 2.05 | 8.3 | 2.45 | 12 |

The procedure used for the test is to first of all plot $T_L$ and Tu from Table 4.4 versus n on probability scale graph paper. Then, for each station in the region to be tested, the ratio of the 10 year event to the mean annual event is computed and an average ratio for the region is multiplied by the mean annual event for each station to give a modified 10-year event magnitude for each station. The return periods corresponding to these modified 10-year events are then found for each station from the individual station frequency curves, say $T_E$. The effective period of record of each gauging station is determined as the number of recorded annual events plus one half the number of events computed for that station by inter-station correlation, say $N_E$. Next, the coordinate pairs ($T_E$, $N_E$) for each station are plotted on the test graph showing curves of $T_L$ and $T_U$. Any station for which the plotted point is outside the confidence limit curves is then excluded from the homogeneous region. The base graph for use of this test is shown in Figure 4.3.

For each station which remains in the hydrologically homogeneous region, ratios of events of different return periods to the mean annual event are computed for T values of say 2, 2,33, 5, 10, 25, 50 and 100 median values of these ratios are determined for the region. A plot of these median ratios versus return period is then the regional frequency curve and represent the most likely relationship for all parts of the region. In this step, T values are not directly put into the equation, Gumbel has reduced variate factor to put a number rather than T values. The formula to change T values to a related $y$ is given:

$$y = -ln\left(-ln\left(1 - \frac{1}{T}\right)\right) \qquad \text{(Equation 4.9)}$$

The next major step in the index flood analysis is to plot drainage area versus mean annual event for those stations within the homogeneous region and graphically fit a smooth curve through the points.

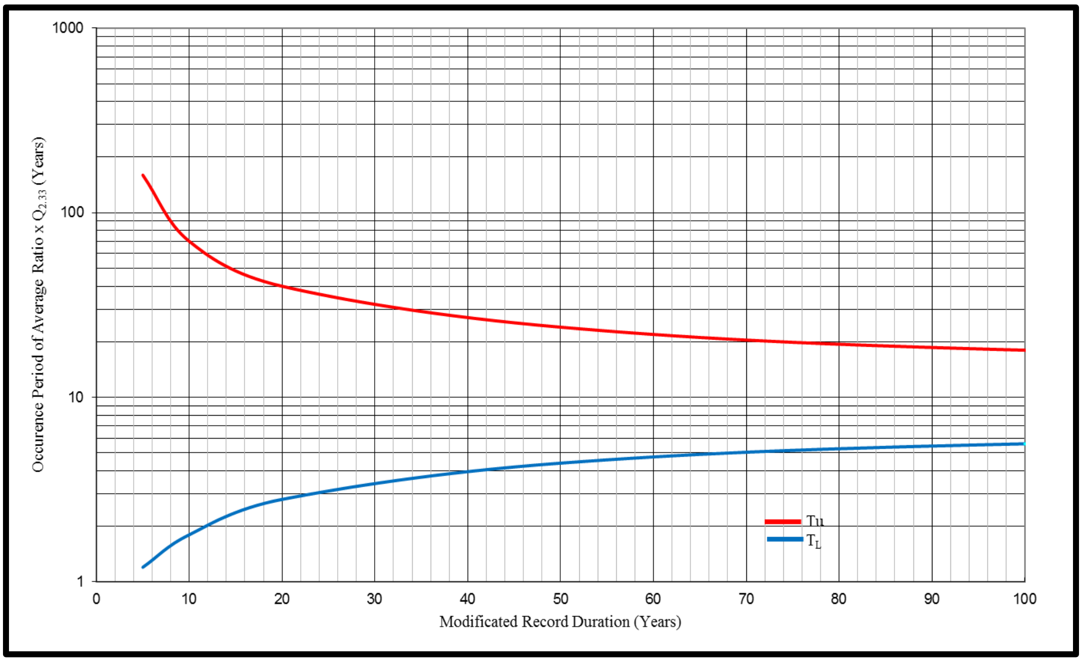


Figure 4.3 Example of Homogeneity Test Graph.

To define a frequency curve at any location within the homogeneous region the mean annual event is determined from curve relating this event to drainage area. The mean annual event is then multiplied by the median ratios for the various return periods required, as determined from the regional frequency curve.
Benson (1962) has noted three deficiencies found in the index flood method:

- The index flood (mean annual flood) for stations with short periods of records may not be typical. This means that the ratios of floods of different return periods to the index-flood may vary widely between stations.

- The homogeneity test is used to determine whether the difference in slopes of frequency curves are greater than may be attributed to chance alone. This test uses the ratio of the 10 year flood because many individual records are too short to adequately define the frequency curve at higher levels. It has been found in some studies that although homogeneity is apparently established at the 10 year level, the individual curves show wide and sometimes systematic differences at higher levels.

- In the use of the index flood method, it has been accepted that within a flood-frequency region frequency curves may be combined for all size of drainage areas, excluding only the largest. Although the variation in the slope of frequency curve with drainage area had been investigated at the time of each study, it was studied at the 10 year point where the effect is small. The error of neglecting this drainage area effect has been reduced by giving separate and special treatment to large streams. For which ratios of less frequent floods were used, have shown, in all regions where such data are available, that the ratios of any specified flood to the mean annual flood will vary inversely with the drainage area. In general, the larger the drainage area, the flatter the frequency curve. The effect of drainage area is relatively greater for floods of higher recurrence intervals.

### 4.2.2 Calculations of RFFA

Gumbel Distribution is the main distribution parameter for the method of Dalyrmple (1960) to make this analysis. However, Regional Flood Frequency Analysis in Turkey is done by using different parameters rather than Gumbel Distribution. Therefore, Regional Flood Frequency Analysis is done both only Gumbel Distribution and best distribution which were given in Table 4.2 and Table 4.3, respectively.

#### 4.2.2.1 Regional Flood Frequency Analysis with Dalyrmple Method

For the clearness of calculation, Regional Flood Frequency Analysis is done for only Gumbel Distributions which are given in Table 4.2. First, $Q_{2.33}$ flood values which mean the mean annual flood should be found out by applying the procedure which is explained in Chapter 3. All calculated $Q_{2.33}$ flood values are given in Table 4.5.

Then, homogeneity test procedure is applied for the reliable calculations. $Q_{10}$ and $Q_{2.33}$ values which are calculated before this process are divided each other. The average of these numbers get as a value and this value is multiplied by $Q_{2.33}$ flood values of all stations to get a ratio. By using this number, occurrence periods for related station is calculated. Then, the record values are adjusted by the base period which is 51 years for this research. There is no need to divide the region according to record years because most of the station records have the same time period. After getting the results is shown in Table 4.5, one could draw a graphic to get an idea about which stations are in homogeneous region and which are not in the region. Upper and Lower boundaries of graphic are given by Dalyrmple (1960). EIE/2007 and EIE/2012 are not in the homogeneous region which is shown on Table 4.5 and Figure 4.4.

Table 4.5 Calculation Data For Homogeneity Test in RFFA with Gumbel Distribution.

| Station | Drainage Area ($km^2$) | $Q_{2.33}$ ($m^3/s$) | $Q_{10}$ ($m^3/s$) | $Q_{10}$ / $Q_{2.33}$ | Avrg x $Q_{2.33}$ | Occurrence Period for Avrg x $Q_{2.33}$ | Record Duration (years) | Adjusted Record Duration (years) |
|---|---|---|---|---|---|---|---|---|
| **EIE/2001** | 8484.00 | 552.95 | 949.44 | 1.72 | 1118.10 | 21.25 | 51.0 | 51.0 |
| **EIE/2004** | 20466.00 | 1027.82 | 1684.83 | 1.64 | 2078.31 | 25.84 | 30.0 | 40.5 |
| **EIE/2005** | 4219.08 | 111.42 | 182.44 | 1.64 | 225.30 | 25.96 | 35.0 | 43.0 |
| **EIE/2006** | 733.20 | 53.43 | 93.56 | 1.75 | 108.03 | 19.54 | 47.0 | 49.0 |
| **EIE/2007** | 623.00 | 59.86 | 94.27 | 1.57 | 121.04 | 37.53 | 41.0 | 46.0 |
| **EIE/2008** | 444.00 | 160.70 | 318.55 | 1.98 | 324.95 | 11.07 | 32.0 | 41.5 |
| **EIE/2009** | 1387.20 | 85.06 | 159.64 | 1.88 | 172.01 | 14.39 | 46.0 | 48.5 |
| **EIE/2010** | 3498.80 | 287.69 | 514.26 | 1.79 | 581.73 | 17.88 | 30.0 | 40.5 |
| **EIE/2012** | 19727.20 | 1094.73 | 1650.68 | 1.51 | 2213.61 | 51.46 | 17.0 | 34.0 |
| **EIE/2015** | 915.20 | 43.95 | 92.74 | 2.11 | 88.87 | 9.12 | 39.0 | 45.0 |
| **EIE/2016** | 846.80 | 83.56 | 173.79 | 2.08 | 168.95 | 9.40 | 15.0 | 33.0 |
| **EIE/2022** | 428.00 | 68.93 | 145.84 | 2.12 | 139.38 | 9.06 | 28.0 | 39.5 |
| **DSI/20-02** | 197.10 | 40.83 | 82.37 | 2.02 | 82.55 | 10.05 | 26.0 | 38.5 |
| **DSI/20-04** | 178.00 | 38.11 | 92.93 | 2.44 | 77.06 | 6.78 | 27.0 | 39.0 |
| **DSI/20-05** | 94.00 | 45.57 | 99.16 | 2.18 | 92.15 | 8.54 | 37.0 | 44.0 |
| **DSI/20-06** | 174.90 | 57.01 | 99.86 | 1.75 | 115.27 | 19.51 | 27.0 | 39.0 |
| **DSI/20-07** | 2084.00 | 54.54 | 138.21 | 2.53 | 110.28 | 6.28 | 35.0 | 43.0 |
| **DSI/20-08** | 131.10 | 39.33 | 93.72 | 2.38 | 79.53 | 7.09 | 35.0 | 43.0 |
| **DSI/20-10** | 217.30 | 62.53 | 100.29 | 1.60 | 126.43 | 28.30 | 19.0 | 35.0 |
| **DSI/20-13** | 105.10 | 50.53 | 92.34 | 1.83 | 102.18 | 16.22 | 39.0 | 45.0 |
| **DSI/20-14** | 310.50 | 163.28 | 319.23 | 1.96 | 330.16 | 11.85 | 11.0 | 31.0 |
| **DSI/20-15** | 189.70 | 26.15 | 68.94 | 2.64 | 52.88 | 5.82 | 20.0 | 35.5 |
| **DSI/20-16** | 291.00 | 35.84 | 99.01 | 2.76 | 72.46 | 5.32 | 34.0 | 42.5 |
| **DSI/20-17** | 1740.00 | 121.32 | 290.53 | 2.39 | 245.32 | 5.88 | 21.0 | 36.0 |
| **DSI/20-36** | 174.20 | 62.79 | 120.62 | 1.92 | 126.97 | 12.90 | 24.0 | 37.5 |
| **DSI/20-40** | 79.00 | 5.66 | 11.79 | 2.08 | 11.45 | 9.38 | 18.0 | 34.5 |
| **DSI/20-43** | 163.00 | 68.16 | 148.69 | 2.18 | 137.82 | 8.50 | 25.0 | 38.0 |
| **DSI/20-44** | 35.00 | 28.96 | 71.59 | 2.47 | 58.57 | 6.60 | 20.0 | 35.5 |
| **DSI/20-45** | 170.00 | 67.52 | 135.61 | 2.01 | 136.52 | 10.36 | 18.0 | 34.5 |
| **DSI/20-46** | 477.00 | 207.68 | 392.65 | 1.89 | 419.94 | 13.90 | 26.0 | 38.5 |
| **DSI/20-51** | 131.40 | 10.14 | 26.51 | 2.61 | 20.51 | 5.92 | 15.0 | 33.0 |
| **DSI/20-52** | 23.00 | 6.12 | 14.46 | 2.36 | 12.37 | 7.20 | 17.0 | 34.0 |
| **DSI/20-53** | 178.50 | 38.27 | 66.93 | 1.75 | 77.39 | 19.65 | 12.0 | 31.5 |
| **DSI/20-54** | 207.50 | 58.87 | 115.67 | 1.96 | 119.04 | 11.57 | 14.0 | 32.5 |
| **DSI/20-55** | 111.60 | 21.06 | 34.39 | 1.63 | 42.59 | 26.27 | 14.0 | 32.5 |
| **DSI/20-56** | 238.40 | 21.53 | 35.16 | 1.63 | 43.54 | 26.28 | 11.0 | 31.0 |
| **DSI/20-58** | 24.38 | 18.38 | 34.70 | 1.89 | 37.17 | 13.99 | 16.0 | 33.5 |
| **DSI/20-59** | 171.50 | 62.49 | 135.69 | 2.17 | 126.37 | 8.58 | 16.0 | 33.5 |
| **DSI/20-63** | 57.50 | 14.61 | 33.25 | 2.27 | 29.55 | 7.79 | 11.0 | 31.0 |
| **DSI/20-65** | 161.00 | 41.45 | 73.58 | 1.78 | 83.80 | 18.42 | 11.0 | 31.0 |
| **Average** | | | | **2.02** | | | | |

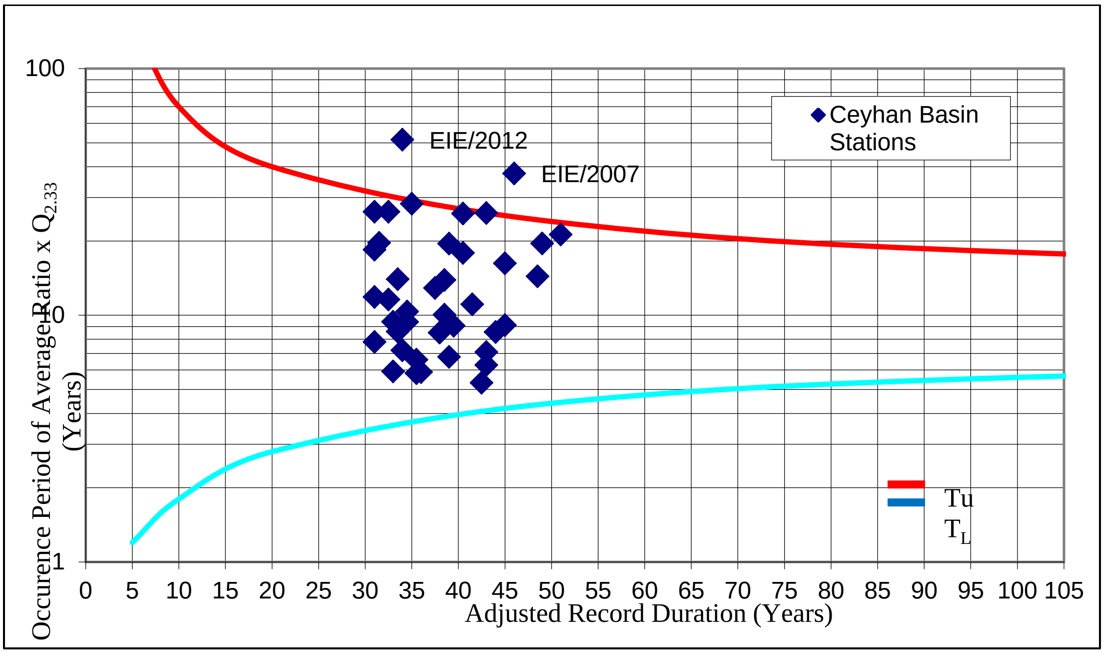


Figure 4.4 Homogeneity Test Graphic For RFFA in Ceyhan Basin with Gumbel Distribution.

Therefore, these two stations should not be used in the Regional Flood Frequency Analysis. Forward steps of this procedure are done without these two stations.

For each station which remains in the hydrologically homogeneous region, ratios of events of different return periods to the mean annual event are computed and shown in Table 4.6.

Table 4.6 Flood Frequency Ratios Over $Q_{2.33}$ Flood Values in RFFA with Gumbel Distribution

| Station | Drainage Area ($km^2$) | $Q_2$ | $Q_{2.33}$ | $Q_5$ | $Q_{10}$ | $Q_{25}$ | $Q_{50}$ | $Q_{100}$ |
|---|---|---|---|---|---|---|---|---|
| **EIE/2001** | 8484.00 | 0.91 | 1.00 | 1.40 | 1.72 | 2.12 | 2.43 | 2.72 |
| **EIE/2004** | 20466.00 | 0.92 | 1.00 | 1.35 | 1.64 | 2.00 | 2.27 | 2.54 |
| **EIE/2005** | 4219.08 | 0.92 | 1.00 | 1.35 | 1.64 | 2.00 | 2.27 | 2.53 |
| **EIE/2006** | 733.20 | 0.90 | 1.00 | 1.41 | 1.75 | 2.18 | 2.49 | 2.81 |
| **EIE/2008** | 444.00 | 0.88 | 1.00 | 1.54 | 1.98 | 2.54 | 2.95 | 3.36 |
| **EIE/2009** | 1387.20 | 0.89 | 1.00 | 1.48 | 1.88 | 2.37 | 2.74 | 3.11 |
| **EIE/2010** | 3498.80 | 0.90 | 1.00 | 1.43 | 1.79 | 2.23 | 2.57 | 2.89 |
| **EIE/2015** | 915.20 | 0.86 | 1.00 | 1.61 | 2.11 | 2.74 | 3.21 | 3.67 |
| **EIE/2016** | 846.80 | 0.86 | 1.00 | 1.60 | 2.08 | 2.69 | 3.15 | 3.60 |
| **EIE/2022** | 428.00 | 0.86 | 1.00 | 1.61 | 2.12 | 2.75 | 3.22 | 3.68 |
| **DSI/20-02** | 197.10 | 0.87 | 1.00 | 1.56 | 2.02 | 2.59 | 3.02 | 3.45 |
| **DSI/20-04** | 178.00 | 0.82 | 1.00 | 1.79 | 2.44 | 3.25 | 3.86 | 4.46 |
| **DSI/20-05** | 94.00 | 0.85 | 1.00 | 1.65 | 2.18 | 2.84 | 3.34 | 3.83 |
| **DSI/20-06** | 174.90 | 0.90 | 1.00 | 1.41 | 1.75 | 2.18 | 2.49 | 2.81 |
| **DSI/20-07** | 2084.00 | 0.81 | 1.00 | 1.85 | 2.53 | 3.40 | 4.05 | 4.69 |
| **DSI/20-08** | 131.10 | 0.82 | 1.00 | 1.76 | 2.38 | 3.17 | 3.75 | 4.33 |
| **DSI/20-10** | 217.30 | 0.92 | 1.00 | 1.33 | 1.60 | 1.95 | 2.20 | 2.45 |
| **DSI/20-13** | 105.10 | 0.90 | 1.00 | 1.46 | 1.83 | 2.30 | 2.64 | 2.99 |
| **DSI/20-14** | 310.50 | 0.88 | 1.00 | 1.53 | 1.96 | 2.50 | 2.90 | 3.30 |
| **DSI/20-15** | 189.70 | 0.79 | 1.00 | 1.90 | 2.64 | 3.56 | 4.25 | 4.94 |
| **DSI/20-16** | 291.00 | 0.78 | 1.00 | 1.97 | 2.76 | 3.76 | 4.50 | 5.24 |
| **DSI/20-17** | 1740.00 | 0.82 | 1.00 | 1.77 | 2.39 | 3.19 | 3.77 | 4.36 |
| **DSI/20-36** | 174.20 | 0.88 | 1.00 | 1.51 | 1.92 | 2.44 | 2.83 | 3.22 |
| **DSI/20-40** | 79.00 | 0.86 | 1.00 | 1.60 | 2.08 | 2.70 | 3.15 | 3.60 |
| **DSI/20-43** | 163.00 | 0.85 | 1.00 | 1.65 | 2.18 | 2.85 | 3.35 | 3.84 |
| **DSI/20-44** | 35.00 | 0.81 | 1.00 | 1.81 | 2.47 | 3.31 | 3.93 | 4.54 |
| **DSI/20-45** | 170.00 | 0.87 | 1.00 | 1.56 | 2.01 | 2.58 | 3.00 | 3.43 |
| **DSI/20-46** | 477.00 | 0.89 | 1.00 | 1.49 | 1.89 | 2.40 | 2.77 | 3.14 |
| **DSI/20-51** | 131.40 | 0.80 | 1.00 | 1.89 | 2.61 | 3.53 | 4.21 | 4.88 |
| **DSI/20-52** | 23.00 | 0.83 | 1.00 | 1.75 | 2.36 | 3.14 | 3.71 | 4.28 |
| **DSI/20-53** | 178.50 | 0.91 | 1.00 | 1.41 | 1.75 | 2.17 | 2.49 | 2.80 |
| **DSI/20-54** | 207.50 | 0.88 | 1.00 | 1.53 | 1.96 | 2.51 | 2.92 | 3.32 |
| **DSI/20-55** | 111.60 | 0.92 | 1.00 | 1.35 | 1.63 | 1.99 | 2.26 | 2.52 |
| **DSI/20-56** | 238.40 | 0.92 | 1.00 | 1.35 | 1.63 | 1.99 | 2.26 | 2.52 |
| **DSI/20-58** | 24.38 | 0.89 | 1.00 | 1.49 | 1.89 | 2.39 | 2.77 | 3.14 |
| **DSI/20-59** | 171.50 | 0.85 | 1.00 | 1.65 | 2.17 | 2.84 | 3.33 | 3.82 |
| **DSI/20-63** | 57.50 | 0.84 | 1.00 | 1.70 | 2.27 | 3.00 | 3.53 | 4.07 |
| **DSI/20-65** | 161.00 | 0.90 | 1.00 | 1.43 | 1.78 | 2.22 | 2.54 | 2.87 |
| | **AVERAGE** | 0.87 | 1.00 | 1.58 | 2.05 | 2.64 | 3.08 | 3.52 |

Reduced variate values which is explained in part 4.2.1 is calculated by using given formula and the results of reduced variate values are given in Table 4.7.

Table 4.7 Reduced Variate(y) Results for Related T.

| T (Years) | Reduced Variate(y) |
|---|---|
| 2 | 0.367 |
| 2.33 | 0.579 |
| 5 | 1.500 |
| 10 | 2.250 |
| 25 | 3.199 |
| 50 | 3.902 |
| 100 | 4.600 |

A graphic representation of these median ratios versus return period is then the regional frequency curve and represent the most likely relationship for all parts of the region and also this graphic shown in Figure 4.5. Reduced variate values are used rather than T values to see graphic more clearly.

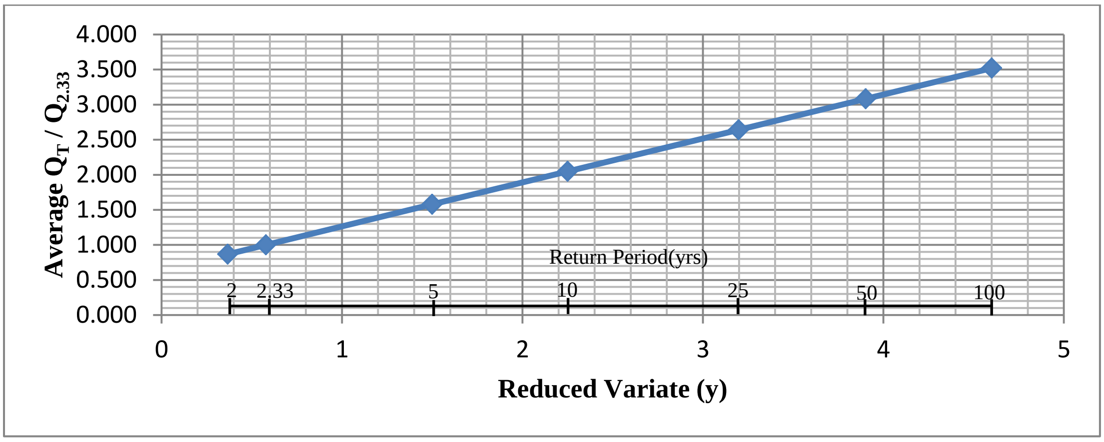


Figure 4.5 Regional Flood Frequency Curve For Ceyhan Basin with Gumbel Distribution.

To obtain a relationship between the drainage area and $Q_{2.33}$, two type of equation is drawn on this graphic. First equation which is drawn as an envelope equation is commonly used by DSI. The main idea behind this equation is drawing a line above all the station data. This is because the equation should include all the station values on the safe side. The first equation which is got from envelope line is shown here:

$$\mathbf{Q_{2.33} = 1.7284 * AREA^{0.8076}} \qquad \text{(Equation 4.10)}$$

Second, equation is found by fitting the best line on this graphic data. SPSS is used for making an analysis with variety of equations. The best equation is cubic equation for the results, but the graphic $Q_{2.33}$ result of cubic equation is started to decrease while the drainage area increasing. Therefore, best equation is linear equation for this data and the test results are given in Table 4.8.. The linear equation which is shown Figure 4.6 is as follows:

$$\mathbf{Q_{2.33} = 0.0492 * AREA + 39.713} \qquad \text{(Equation 4.11)}$$

Table 4.8 Best Equation fits over RFFA results with Gumbel Distribution Parameters.

| Equation | Model Summary | | | | | Parameter Estimates | | | |
|---|---|---|---|---|---|---|---|---|---|
| | R Square | F | df1 | df2 | Sig. | Constant | b1 | b2 | b3 |
| Linear | ,922 | 423,717 | 1 | 36 | ,000 | 39,713 | ,049 | | |
| Logarithmic | ,508 | 37,229 | 1 | 36 | ,000 | -387,657 | 85,843 | | |
| Inverse | ,083 | 3,269 | 1 | 36 | ,079 | 142,453 | -5291,540 | | |
| Quadratic | ,922 | 206,939 | 2 | 35 | ,000 | 38,065 | ,052 | -1,660E-007 | |
| Cubic | ,930 | 149,429 | 3 | 34 | ,000 | 47,751 | ,020 | 6,463E-006 | -2,502E-010 |
| Compound | ,426 | 26,698 | 1 | 36 | ,000 | 41,990 | 1,000 | | |
| Power | ,652 | 67,307 | 1 | 36 | ,000 | 2,111 | ,567 | | |
| S | ,356 | 19,906 | 1 | 36 | ,000 | 4,458 | -63,816 | | |
| Growth | ,426 | 26,698 | 1 | 36 | ,000 | 3,737 | ,000 | | |
| Exponential | ,426 | 26,698 | 1 | 36 | ,000 | 41,990 | ,000 | | |
| Logistic | ,426 | 26,698 | 1 | 36 | ,000 | ,024 | 1,000 | | |

Then the last part of this method is drawing a graph which represents drainage area versus $Q_{2.33}$ flood values for homogeneous stations. Figure 4.6 is shown an idea about the graph of $Q_{2.33}$ and drainage area.

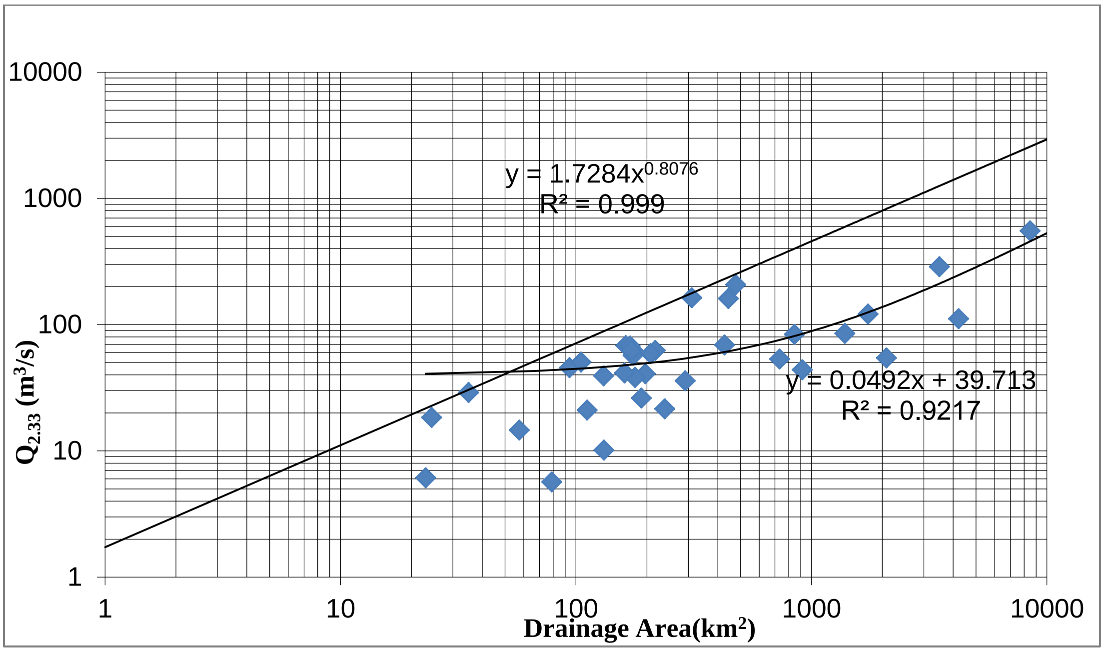


Figure 4.6 Drainage Area versus $Q_{2.33}$ Graph for RFFA with Gumbel Distribution.

After finding these equations, one can find any desired flood value for any occurrence period with using Drainage area of the ungauged basin. In the next chapters of this study, flood values for all stations are calculated by using these equations. The idea behind this calculation is to check the model reliability with observed station values. However, only two stations which are not in the homogeneous region are not taken into consideration to compare with other models.

#### 4.2.2.2 Regional Flood Frequency Analysis which is used in Turkey with Different Distributions

In Turkey, Regional Flood Frequency Analysis is taken an important role in the comparing of results. The procedure of Regional Flood Frequency Analysis is similar to the Dalyrmple's Regional Flood Frequency Analysis. However, these both methods are different from Distribution parameters. Dalyrmple Method is used only Gumbel Distribution, but in Turkey, Three Parameter Lognormal, Pearson Type III, Log-pearson Type III, Two Parameter Lognormal, Normal and Gumbel are the distributions which are used.

Therefore, all of the procedures explained in part 4.2.2.1 except for distributions, is applied. Best distribution results after Kolmogorov-Smirnov test for all stations which are given in Table 4.3 are used in the calculations.

Then the first procedure of Dalyrmple is applied and homogeneity test results are given in Table 4.9 and Figure 4.7. Therefore, there are totally 12 stations which are not in homogeneous region. Dalyrpmle(1960) offer us to use Gumbel Distribution in the procedure. However, different distributions are used in Turkey while Regional Flood Frequency Analysis. Therefore, these calculations are made to compare the results.

These 12 stations are omitted from data set to make the median ratios of 2, 2.33, 5, 10, 25, 50 and 100 occurrence year floods. These results are given in the Table 4.10.

After getting the median ratios from Table 4.10 and using the reduced variates which are given in Table 4.7, one can draw a Regional Frequency Curve which is shown in Figure 4.8.

Table 4.9 RFFA Homogeneity Test in Ceyhan Basin with Different Distribution Parameters

| Station | Drainage Area ($km^2$) | $Q_{2.33}$ ($m^3/s$) | $Q_{10}$ ($m^3/s$) | $Q_{10}$ / $Q_{2.33}$ | Avrg x $Q_{2.33}$ | Occurrence Period for Avrg x $Q_{2.33}$ | Record Duration (years) | Adjusted Record Duration (years) |
|---|---|---|---|---|---|---|---|---|
| EIE/2001 | 8484.00 | 536.54 | 901.94 | 1.68 | 1116.92 | 24.91 | 51.0 | 51.0 |
| EIE/2004 | 20466.00 | 1089.20 | 1571.43 | 1.44 | 2267.39 | 498.97 | 30.0 | 40.5 |
| EIE/2005 | 4219.08 | 108.41 | 171.87 | 1.59 | 225.69 | 39.71 | 35.0 | 43.0 |
| EIE/2006 | 733.20 | 53.55 | 89.23 | 1.67 | 111.47 | 30.78 | 47.0 | 49.0 |
| EIE/2007 | 623.00 | 58.86 | 89.84 | 1.53 | 122.52 | 56.49 | 41.0 | 46.0 |
| EIE/2008 | 444.00 | 146.34 | 286.69 | 1.96 | 304.63 | 12.89 | 32.0 | 41.5 |
| EIE/2009 | 1387.20 | 79.56 | 148.13 | 1.86 | 165.63 | 15.99 | 46.0 | 48.5 |
| EIE/2010 | 3498.80 | 285.82 | 480.45 | 1.68 | 594.99 | 28.19 | 30.0 | 40.5 |
| EIE/2012 | 19727.20 | 1089.68 | 1527.47 | 1.40 | 2268.39 | 670.47 | 17.0 | 34.0 |
| EIE/2015 | 915.20 | 38.34 | 81.68 | 2.13 | 79.82 | 9.56 | 39.0 | 45.0 |
| EIE/2016 | 846.80 | 73.65 | 146.41 | 1.99 | 153.31 | 12.08 | 15.0 | 33.0 |
| EIE/2022 | 428.00 | 53.93 | 128.90 | 2.39 | 112.26 | 7.77 | 28.0 | 39.5 |
| DSI/20-02 | 197.10 | 36.54 | 72.53 | 1.99 | 76.06 | 12.15 | 26.0 | 38.5 |
| DSI/20-04 | 178.00 | 28.54 | 80.31 | 2.81 | 59.42 | 5.66 | 27.0 | 39.0 |
| DSI/20-05 | 94.00 | 36.82 | 91.67 | 2.49 | 76.64 | 7.02 | 37.0 | 44.0 |
| DSI/20-06 | 174.90 | 51.84 | 93.35 | 1.80 | 107.92 | 18.78 | 27.0 | 39.0 |
| DSI/20-07 | 2084.00 | 40.86 | 103.08 | 2.52 | 85.06 | 7.05 | 35.0 | 43.0 |
| DSI/20-08 | 131.10 | 28.72 | 76.65 | 2.67 | 59.79 | 6.53 | 35.0 | 43.0 |
| DSI/20-10 | 217.30 | 62.53 | 100.29 | 1.60 | 130.16 | 38.29 | 19.0 | 35.0 |
| DSI/20-13 | 105.10 | 46.61 | 89.60 | 1.92 | 97.04 | 15.23 | 39.0 | 45.0 |
| DSI/20-14 | 310.50 | 149.17 | 292.17 | 1.96 | 310.52 | 14.22 | 11.0 | 31.0 |
| DSI/20-15 | 189.70 | 16.90 | 51.63 | 3.06 | 35.18 | 5.58 | 20.0 | 35.5 |
| DSI/20-16 | 291.00 | 23.80 | 66.03 | 2.77 | 49.55 | 6.35 | 34.0 | 42.5 |
| DSI/20-17 | 1740.00 | 86.40 | 270.65 | 3.13 | 179.86 | 4.94 | 21.0 | 36.0 |
| DSI/20-36 | 174.20 | 62.79 | 120.62 | 1.92 | 130.71 | 14.61 | 24.0 | 37.5 |
| DSI/20-40 | 79.00 | 4.68 | 10.57 | 2.26 | 9.75 | 8.50 | 18.0 | 34.5 |
| DSI/20-43 | 163.00 | 50.79 | 136.69 | 2.69 | 105.73 | 6.42 | 25.0 | 38.0 |
| DSI/20-44 | 35.00 | 19.68 | 62.55 | 3.18 | 40.96 | 5.06 | 20.0 | 35.5 |
| DSI/20-45 | 170.00 | 55.11 | 123.00 | 2.23 | 114.73 | 8.74 | 18.0 | 34.5 |
| DSI/20-46 | 477.00 | 198.10 | 358.97 | 1.81 | 412.39 | 18.38 | 26.0 | 38.5 |
| DSI/20-51 | 131.40 | 8.71 | 20.26 | 2.33 | 18.14 | 8.19 | 15.0 | 33.0 |
| DSI/20-52 | 23.00 | 4.97 | 11.57 | 2.33 | 10.35 | 8.15 | 17.0 | 34.0 |
| DSI/20-53 | 178.50 | 36.37 | 58.73 | 1.61 | 75.70 | 34.12 | 12.0 | 31.5 |
| DSI/20-54 | 207.50 | 53.02 | 102.48 | 1.93 | 110.37 | 14.53 | 14.0 | 32.5 |
| DSI/20-55 | 111.60 | 21.06 | 34.39 | 1.63 | 43.84 | 33.46 | 14.0 | 32.5 |
| DSI/20-56 | 238.40 | 22.30 | 31.00 | 1.39 | 46.42 | 764.08 | 11.0 | 31.0 |
| DSI/20-58 | 24.38 | 18.38 | 34.70 | 1.89 | 38.26 | 15.76 | 16.0 | 33.5 |
| DSI/20-59 | 171.50 | 52.43 | 118.13 | 2.25 | 109.14 | 8.45 | 16.0 | 33.5 |
| DSI/20-63 | 57.50 | 12.22 | 26.26 | 2.15 | 25.45 | 9.41 | 11.0 | 31.0 |
| DSI/20-65 | 161.00 | 39.47 | 64.06 | 1.62 | 82.17 | 33.79 | 11.0 | 31.0 |
| AVERAGE | | | | 2.08 | | | | |

(yellow) Stations which are in the non-homogeneous region which are shown on Figure 4.7

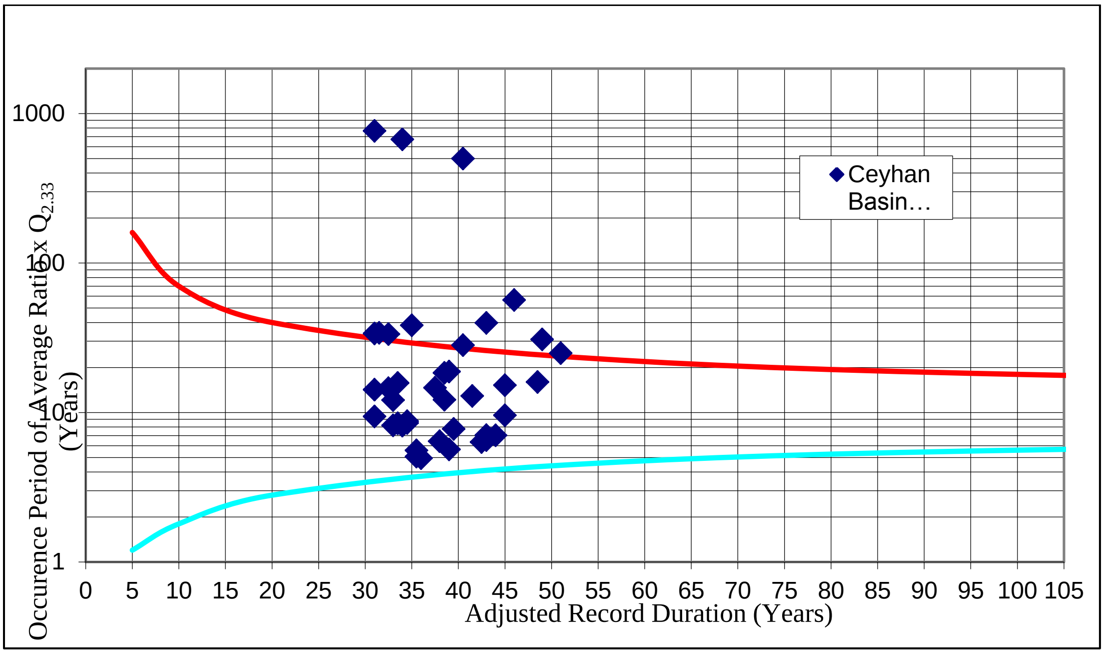


Figure 4.7 Homogeneity Test Graphic For RFFA with Different Distribution Parameters.

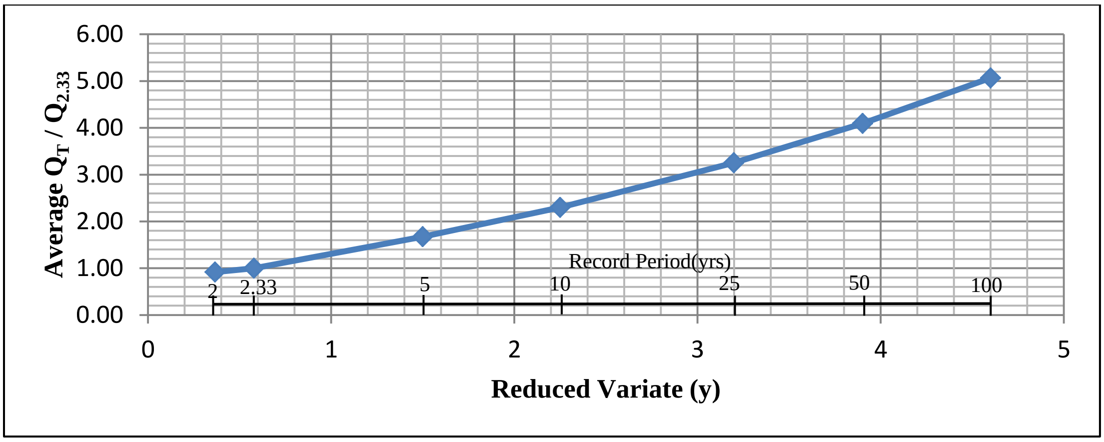


Figure 4.8 RFFA Curve For Ceyhan Basin with Different Distribution Parameters.

Table 4.10 Flood Frequency Ratios Over $Q_{2.33}$ Values in RFFA with Different Distribution Parameters.

| Station | Drainage Area ($km^2$) | $Q_2$ | $Q_{2.33}$ | $Q_5$ | $Q_{10}$ | $Q_{25}$ | $Q_{50}$ | $Q_{100}$ |
|---|---|---|---|---|---|---|---|---|
| **EIE/2008** | 444.00 | 0.89 | 1.00 | 1.50 | 1.96 | 2.60 | 3.11 | 3.65 |
| **EIE/2009** | 1387.20 | 0.90 | 1.00 | 1.46 | 1.86 | 2.41 | 2.85 | 3.30 |
| **EIE/2015** | 915.20 | 0.88 | 1.00 | 1.58 | 2.13 | 2.94 | 3.62 | 4.36 |
| **EIE/2016** | 846.80 | 0.89 | 1.00 | 1.51 | 1.99 | 2.66 | 3.22 | 3.81 |
| **EIE/2022** | 428.00 | 0.94 | 1.00 | 1.70 | 2.39 | 3.52 | 4.57 | 5.84 |
| **DSI/20-02** | 197.10 | 0.89 | 1.00 | 1.51 | 1.99 | 2.66 | 3.21 | 3.80 |
| **DSI/20-04** | 178.00 | 0.92 | 1.00 | 1.97 | 2.81 | 3.99 | 4.91 | 5.86 |
| **DSI/20-05** | 94.00 | 0.90 | 1.00 | 1.80 | 2.49 | 3.40 | 4.09 | 4.77 |
| **DSI/20-06** | 174.90 | 0.96 | 1.00 | 1.45 | 1.80 | 2.28 | 2.66 | 3.06 |
| **DSI/20-07** | 2084.00 | 0.93 | 1.00 | 1.78 | 2.52 | 3.71 | 4.78 | 6.03 |
| **DSI/20-08** | 131.10 | 0.93 | 1.00 | 1.82 | 2.67 | 4.09 | 5.46 | 7.13 |
| **DSI/20-13** | 105.10 | 0.95 | 1.00 | 1.54 | 1.92 | 2.38 | 2.70 | 3.00 |
| **DSI/20-14** | 310.50 | 0.95 | 1.00 | 1.57 | 1.96 | 2.40 | 2.69 | 2.94 |
| **DSI/20-15** | 189.70 | 0.92 | 1.00 | 1.95 | 3.06 | 5.14 | 7.35 | 10.32 |
| **DSI/20-16** | 291.00 | 0.93 | 1.00 | 1.83 | 2.77 | 4.55 | 6.46 | 9.01 |
| **DSI/20-17** | 1740.00 | 0.91 | 1.00 | 2.10 | 3.13 | 4.65 | 5.92 | 7.27 |
| **DSI/20-36** | 174.20 | 0.88 | 1.00 | 1.51 | 1.92 | 2.44 | 2.83 | 3.22 |
| **DSI/20-40** | 79.00 | 0.94 | 1.00 | 1.67 | 2.26 | 3.10 | 3.80 | 4.56 |
| **DSI/20-43** | 163.00 | 0.93 | 1.00 | 1.84 | 2.69 | 4.10 | 5.44 | 7.06 |
| **DSI/20-44** | 35.00 | 0.91 | 1.00 | 2.07 | 3.18 | 5.03 | 6.78 | 8.87 |
| **DSI/20-45** | 170.00 | 0.94 | 1.00 | 1.64 | 2.23 | 3.16 | 3.98 | 4.94 |
| **DSI/20-46** | 477.00 | 0.90 | 1.00 | 1.44 | 1.81 | 2.29 | 2.66 | 3.04 |
| **DSI/20-51** | 131.40 | 0.94 | 1.00 | 1.65 | 2.33 | 3.47 | 4.59 | 5.99 |
| **DSI/20-52** | 23.00 | 0.87 | 1.00 | 1.66 | 2.33 | 3.33 | 4.20 | 5.18 |
| **DSI/20-54** | 207.50 | 0.93 | 1.00 | 1.53 | 1.93 | 2.43 | 2.78 | 3.13 |
| **DSI/20-58** | 24.38 | 0.89 | 1.00 | 1.49 | 1.89 | 2.39 | 2.77 | 3.14 |
| **DSI/20-59** | 171.50 | 0.94 | 1.00 | 1.70 | 2.25 | 2.98 | 3.54 | 4.09 |
| **DSI/20-63** | 57.50 | 0.88 | 1.00 | 1.58 | 2.15 | 2.97 | 3.67 | 4.43 |
| | **AVERAGE** | **0.92** | **1.00** | **1.67** | **2.30** | **3.25** | **4.09** | **5.06** |

First equation is determinated by drawing an envelope line which is shown in Figure 4.8:

$$\mathbf{Q_{2.33} = 1.4266 * AREA^{0.8123}} \quad \text{(Equation 4.12)}$$

equation is determinated by using SPSS program by trying to fit different equations. The results of the fitted equations are given in Table 4.11. Cubic equation seems more efficient than the others In Table 4.11, but this equation has a decreasing part after some region. Therefore power equation is selected as second best equation:

$$\mathbf{Q_{2.33} = 2.509 * AREA^{0.504}} \quad \text{(Equation 4.13)}$$

The Flood results which are found by these two equations are compared with the results of model which is built with basin characteristics. However, 12 stations flood values are not calculated by using these two equations because of these are not in homogeneous region.

Table 4.11 Best Equation fits over RFFA results Different Distribution Parameters.

| Equation | Model Summary | | | | | Parameter Estimates | | | |
|---|---|---|---|---|---|---|---|---|---|
| | R Square | F | df1 | df2 | Sig. | Constant | b1 | b2 | b3 |
| Linear | ,072 | 2,029 | 1 | 26 | ,166 | 43,472 | ,023 | | |
| Logarithmic | ,249 | 8,601 | 1 | 26 | ,007 | -51,594 | 19,538 | | |
| Inverse | ,198 | 6,421 | 1 | 26 | ,018 | 68,879 | -1828,214 | | |
| Quadratic | ,315 | 5,750 | 2 | 25 | ,009 | 17,490 | ,176 | -8,190E-005 | |
| Cubic | ,408 | 5,507 | 3 | 24 | ,005 | -3,144 | ,359 | ,000 | 9,096E-008 |
| Compound | ,140 | 4,220 | 1 | 26 | ,050 | 28,525 | 1,001 | | |
| Power | ,405 | 17,664 | 1 | 26 | ,000 | 2,508 | ,504 | | |
| S | ,382 | 16,072 | 1 | 26 | ,000 | 4,064 | -51,345 | | |
| Growth | ,140 | 4,220 | 1 | 26 | ,050 | 3,351 | ,001 | | |
| Exponential | ,140 | 4,220 | 1 | 26 | ,050 | 28,525 | ,001 | | |
| Logistic | ,140 | 4,220 | 1 | 26 | ,050 | ,035 | ,999 | | |

After the finishing of these calculations, drainage area versus $Q_{2.33}$ graphic should be drawn to get a relationship between the drainage area and flood frequencies in Figure 4.9.

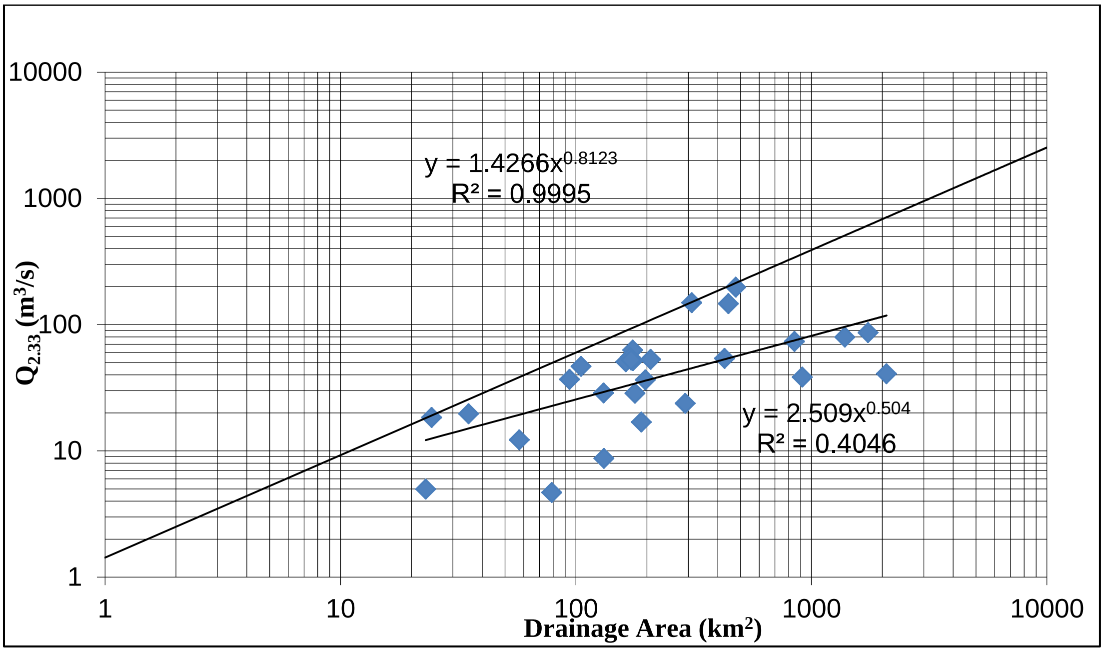


Figure 4.9 Drainage Area versus $Q_{2.33}$ Graph for RFFA with Different Distribution Parameters.

### 4.3 Point Flood Analysis (At-Site Flood Analysis)

#### 4.3.1 Methodology of Point Flood Analysis

Point Flood Frequency Analysis is commonly used in ungauged basins which has a gauged station in the upstream or downstream on the same river branch. Moreover, ungauged and gauged basins are so close and has the same basin characteristics, one can also use the method for ungauged basins. Any desired flood period value for ungauged station can be calculated by using this formula:

$$Qt(ungauged) = Qt(gauged)\left( {A(ungauged)}/{A(gauged)} \right)^{(n)} \qquad \text{(Equation 4.14)}$$

t : Desired period to calculate
n : Constant for calculation.
A(gauged) : Drainage are of gauged station
A(ungauged) : Drainage area of ungauged station.
Qt(gauged) : Flood value for related t.
Qt(ungauged): Result of ungauged basin Flood Frequency.

Firstly, two gauged stations annual time series are taken into account and the best distribution fits over this station is determined. This procedure is explained in details in section 4.1. n value is the only unknown when considering two gauged stations. Therefore, n value is calculated for this two gauged stations. After, one can obtain the flood result of an ungauged station by using this n value and the value for the gauged stations. However, the two gauged stations should be on same river branch to make this analysis. There are some exceptions which this procedure can be applied.

In this research, Point Flood Analysis is used for comparing the Regional Flood Frequency Analysis results, so flood discharge for gauged station is used to calculate the flood discharge for another gauged station for comparison. Therefore, n value could not be calculated for any two gauged station, for this reason n = 2/3 is used in the calculations.

#### 4.3.2 Point Flood Analysis Calculations.

In Ceyhan Basin, all stations do not have upstream/downstream relationship. Therefore, this method can be applied for limited number of stations shown in Table 4.12. Two stations which are on the same river branch are used for this analysis. However, there are some exceptional stations that are not on the same river branch but stay so close to each other. Therefore, this kind of stations also taken into consideration because of having the same basin parameters.

All the Flood values are calculated in Chapter 4.1 for all stations. However, idea behind this research is relied on to check the methods dependency. Therefore one should consider the stations, which have upstream-downstream relationship, on same river branch. First, calculate the upstream station by using downstream station Flood discharge and Area values. Then, calculate the downstream station flood values by using upstream flood values. For example, EIE-2006 and EIE-2009 are the stations which are on the same river branch are shown in Figure 4.10. First, EIE-2006 Point Flood Frequency values are obtained by EIE-2009 station values. Then, EIE-2009 values are obtained by using EIE-2006 station values which are calculated by the method explained in previous part. By doing this, the result of analysis can be checked by the real calculated values.

By applying this method for all of the selected stations which are given in Table 4.12, Gumbel Distribution parameter values and different distribution parameter values are taken into account in Point Flood Frequency Analysis. Table 4.12 means that if two stations are on same row, Point Flood Frequency Analysis is applied between these two stations.

Results of the stations for both Gumbel and different distributions model are given in Table 4.13 and Table 4.14, respectively.

Table 4.12 Stations which can be used together for Point Flood Frequency Analysis.

| # | Upstream Station | Drainage Area (km$^2$) | Downstream Station | Drainage Area (km$^2$) |
|---|---|---|---|---|
| **1** | **EIE/2005** | 4219.1 | **EIE/2001** | 8484.0 |
| **2** | **EIE/2012** | 19727.2 | **EIE/2004** | 20466.0 |
| **3** | **EIE/2006** | 733.2 | **EIE/2009** | 1387.2 |
| **4** | **EIE/2015** | 915.2 | **DSI/20-07** | 2084.0 |
| **5** | **EIE/2022** | 428.0 | **EIE/2016** | 846.8 |
| **6** | **DSI/20-43** | 163.0 | **EIE/2007** | 623.0 |
| **7** | **DSI/20-14** | 310.5 | **DSI/20-46** | 477.0 |
| **8*** | **DSI/20-53** | 178.5 | **DSI/20-54** | 207.5 |
| **9*** | **DSI/20-58** | 24.4 | **DSI/20-59** | 171.5 |

* These stations are not on the same branch, but so close to each other.

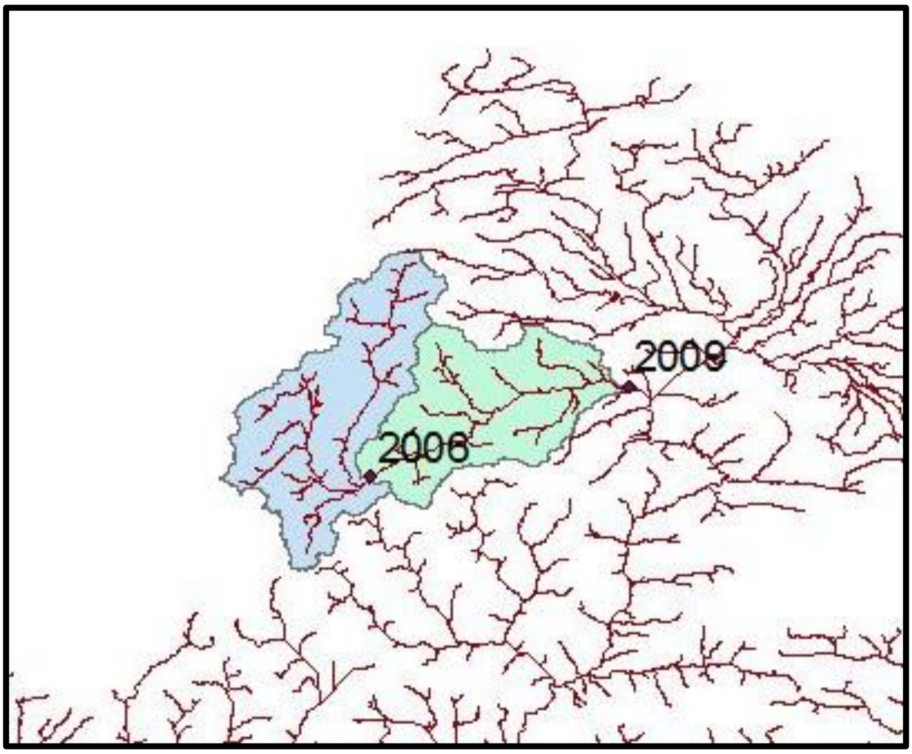


Figure 4.10 Relation between the stations EIE-2006 and EIE-2009.

Table 4.13 Point Flood Frequency Analysis with Gumbel Distribution.

| Group Number | Station | Drainage Area (km$^2$) | $Q_{2.33}$ (m$^3$/s) | Results of Point Flood Frequency Analysis $Q_{2.33}$ (m$^3$/s) |
|---|---|---|---|---|
| 1 | EIE/2001 | 8484 | 552.95 | 177.51 |
| | EIE/2005 | 4219.08 | 111.42 | 347.08 |
| 2 | EIE/2004 | 20466 | 1027.82 | 1121.89 |
| | EIE/2012 | 19727.2 | 1094.73 | 1002.93 |
| 3 | EIE/2006 | 733.2 | 53.43 | 55.61 |
| | EIE/2009 | 1387.2 | 85.06 | 81.73 |
| 4 | EIE/2015 | 915.2 | 43.95 | 31.51 |
| | DSI/20-07 | 2084 | 54.54 | 76.07 |
| 5 | EIE/2016 | 846.8 | 83.56 | 108.63 |
| | EIE/2022 | 428 | 68.93 | 53.02 |
| 6 | DSI/20-53 | 178.5 | 38.27 | 53.25 |
| | DSI/20-54 | 207.5 | 58.87 | 42.31 |
| 7 | DSI/20-58 | 24.38 | 18.38 | 17.02 |
| | DSI/20-59 | 171.5 | 62.49 | 67.48 |
| 8 | EIE/2007 | 623 | 59.86 | 166.62 |
| | DSI/20-43 | 163 | 68.16 | 24.49 |
| 9 | DSI/20-14 | 310.5 | 163.28 | 155.99 |
| | DSI/20-46 | 477 | 207.68 | 217.38 |

Table 4.14 Point Flood Frequency Analysis Results with Different Distributions Results.

| Group Number | Station | Drainage Area (km$^2$) | $Q_{2.33}$ (m$^3$/s) | Results of Point Flood Frequency Analysis (m$^3$/s) |
|---|---|---|---|---|
| 1 | EIE/2001 | 8484 | 536.54 | 172.72 |
| | EIE/2005 | 4219.08 | 108.41 | 336.78 |
| 2 | EIE/2004 | 20466 | 1089.20 | 1116.72 |
| | EIE/2012 | 19727.2 | 1089.68 | 1062.83 |
| 3 | EIE/2006 | 733.2 | 53.55 | 52.01 |
| | EIE/2009 | 1387.2 | 79.56 | 81.91 |
| 4 | EIE/2015 | 915.2 | 38.34 | 23.61 |
| | DSI/20-07 | 2084 | 40.86 | 66.36 |
| 5 | EIE/2016 | 846.8 | 73.65 | 84.99 |
| | EIE/2022 | 428 | 53.93 | 46.73 |
| 6 | DSI/20-53 | 178.5 | 36.37 | 47.96 |
| | DSI/20-54 | 207.5 | 53.02 | 40.21 |
| 7 | DSI/20-58 | 24.38 | 18.38 | 14.28 |
| | DSI/20-59 | 171.5 | 52.43 | 67.48 |
| 8 | EIE/2007 | 623 | 58.86 | 124.16 |
| | DSI/20-43 | 163 | 50.79 | 24.08 |
| 9 | DSI/20-14 | 310.5 | 149.17 | 148.79 |
| | DSI/20-46 | 477 | 198.10 | 198.60 |

## 4.4 Cluster Analysis

### 4.4.1 Methodology of Cluster Analysis

Cluster analysis (CA) groups sites on the basis of a statistical distance measure reflecting the similarity (or dissimilarity) among a set of attributes (similarity measures) selected to represent each gauging station. Several clustering techniques are available in the statistical literature, including hierarchical approaches such as single linkage, complete linkage, average linkage and Ward’s method, as well as non-hierarchical approaches such as the k-means method (Johnson and Wichern 2007). These methods have been widely used in the delineation of hydrologically homogeneous regions (Burn 1989, 1990, 1997, 2000; Hosking and Wallis 1997; Chiang et al. 2002a; Castellarin et al. 2001; Rao and Srinivas 2006).

In this research, both Ward’s Cluster and k-means cluster methods are used to delinate homogeneous regions by using SPSS program as the name with Twostep Clustring Analysis which is shown Figure 4.11.

The Twostep Cluster Analysis procedure is an exploratory tool designed to reveal natural groupings (or clusters) within a dataset that would otherwise not be apparent. The algorithm employed by this procedure has several desirable features that differentiate it from traditional clustering techniques. Here is the reasons why this method is selected to find different clusters:

• Handling of categorical and continuous variables. By assuming variables to be independent, a joint multinomial-normal distribution can be placed on categorical and continuous variables.

• Automatic selection of number of clusters. By comparing the values of a model-choice criterion across different clustering solutions, the procedure can automatically determine the optimal number of clusters.

• Scalability. By constructing a cluster features (CF) tree that summarizes the records, the TwoStep algorithm allows you to analyze large data files.

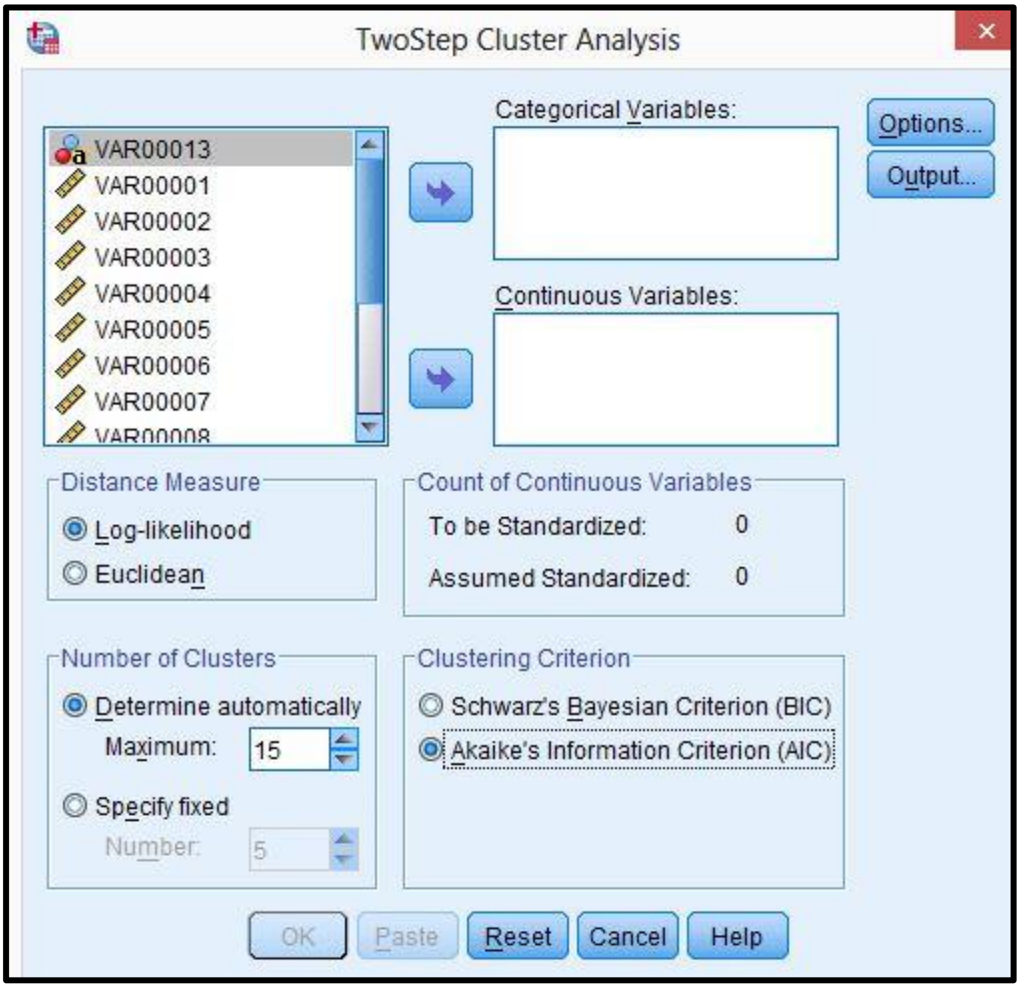


Figure 4.11 TwoStep Cluster Analysis Main Menu in SPSS.

### 4.4.2 Cluster Analysis Calculations

Regional Flood Frequency Analysis is done by using generally accepted method in Turkey. However, this method only uses area and flood relationship, but there are a lot of basin characteristics which should affect the Flood Frequency Analysis. Therefore, Cluster Analysis is done for dividing basin into homogeneous regions due to basin characteristics. By doing this method one can compare the effect of clustering and non-clustering.

To make this analysis, independent basin characteristics which are given in Table 3.3 are used for the analysis. $Q_{2.33}$ flood values which is dependent value of basin equation are not used in the analysis because of the research aim is to find these values.

As $Q_{2.33}$ values not used in Cluster Analysis, two models which are developed by Gumbel Distribution and Different Distribution Parameters could use the same clusters. The summary of the Clustering Analysis Result is given in Figure 4.12.

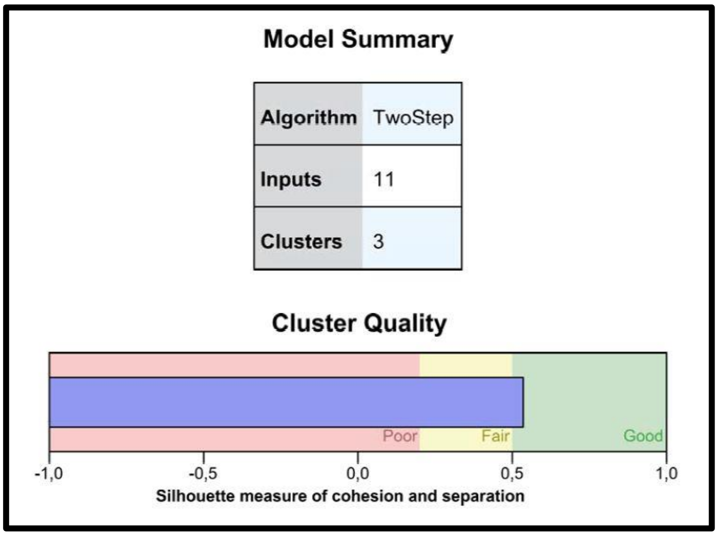


Figure 4.12 Summary of Twostep Cluster Analysis which has done by SPSS.

There are only 3 clusters selected by the TwoStep Clustering Analysis. The best cluster number is obtained by this method. The most effective variables in the selection criteria is given in Figure 4.13

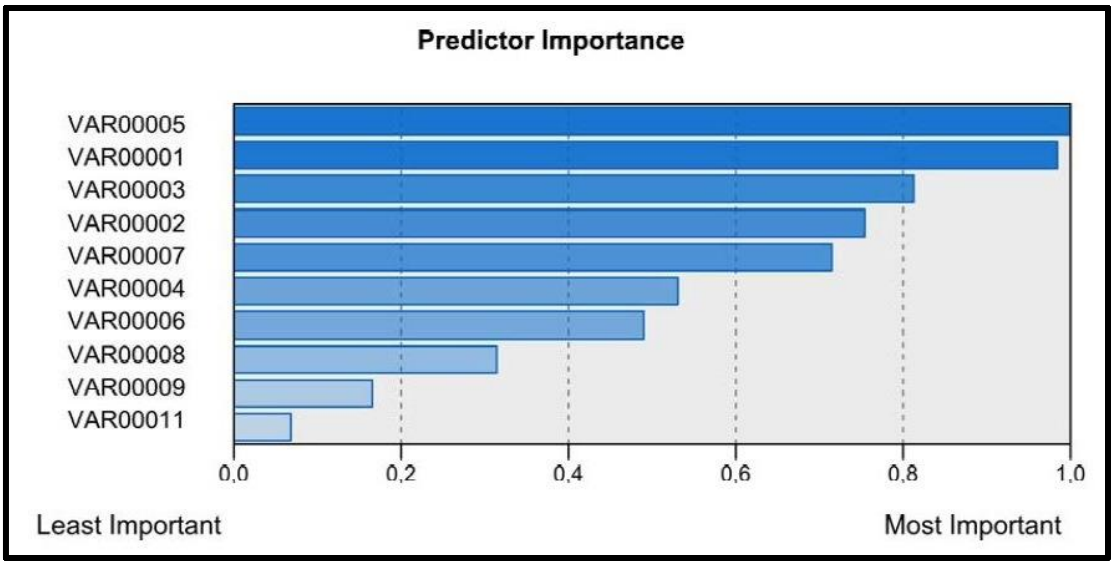


Figure 4.13 Result of Variable Effects in Clustering.

As one can understand that $H_{min}$ and Drainage Area are the two parameters which are more effective in Twostep clustering technique. The clusters with respect to Area and $H_{min}$ are given in Table 4.15.

Table 4.15 Clustered Regions and Important Clustering Variables.

| Cluster Region | Station | AREA ($km^2$) | $H_{min}$ (m) |
|---|---|---|---|
| # 1 | EIE/2001 | 8484 | 420 |
| | EIE/2004 | 20466 | 2 |
| | EIE/2012 | 19727.2 | 2 |
| # 2 | EIE/2007 | 623 | 7 |
| | EIE/2008 | 444 | 56 |
| | EIE/2010 | 3498.8 | 401 |
| | DSI/20-04 | 178 | 214 |
| | DSI/20-05 | 94 | 196 |
| | DSI/20-06 | 174.9 | 340 |
| | DSI/20-08 | 131.1 | 211 |
| | DSI/20-10 | 217.3 | 486 |
| | DSI/20-13 | 105.1 | 183 |
| | DSI/20-14 | 310.5 | 417 |
| | DSI/20-17 | 1740 | 478 |
| | DSI/20-40 | 79 | 617 |
| | DSI/20-43 | 163 | 213 |
| | DSI/20-44 | 35 | 300 |
| | DSI/20-45 | 170 | 274 |
| | DSI/20-46 | 477 | 171 |
| | DSI/20-56 | 238.4 | 488 |
| | DSI/20-63 | 57.5 | 530 |
| | DSI/20-65 | 161 | 71 |
| # 3 | EIE/2005 | 4219.08 | 1059 |
| | EIE/2006 | 733.2 | 1313 |
| | EIE/2009 | 1387.2 | 1102 |
| | EIE/2015 | 915.2 | 1209 |
| | EIE/2016 | 846.8 | 1153 |
| | EIE/2022 | 428 | 1335 |
| | DSI/20-02 | 197.1 | 889 |
| | DSI/20-07 | 2084 | 1095 |
| | DSI/20-15 | 189.7 | 1350 |
| | DSI/20-16 | 291 | 1363 |
| | DSI/20-36 | 174.2 | 1232 |
| | DSI/20-51 | 131.4 | 1099 |
| | DSI/20-52 | 23 | 1275 |
| | DSI/20-53 | 178.5 | 1060 |
| | DSI/20-54 | 207.5 | 758 |
| | DSI/20-55 | 111.6 | 830 |
| | DSI/20-58 | 24.38 | 970 |
| | DSI/20-59 | 171.5 | 1018 |

Cluster-1 has only three stations, this is because these stations have very huge drainage areas. Therefore, cluster analysis is produced a different cluster for this kind of huge drainage areas. In addition, huge drainage areas are mostly have gauged stations, so Regional Flood Frequency Analysis is generally not used in these basins.

Cluster-2 and Cluster-3 are clustered due to the elevation of stations. This seems logical because of drainage basins which have the same elevation have more chance to have same basin characteristics. Cluster-2 and Cluster-3 stations are given in Figure 4.14.

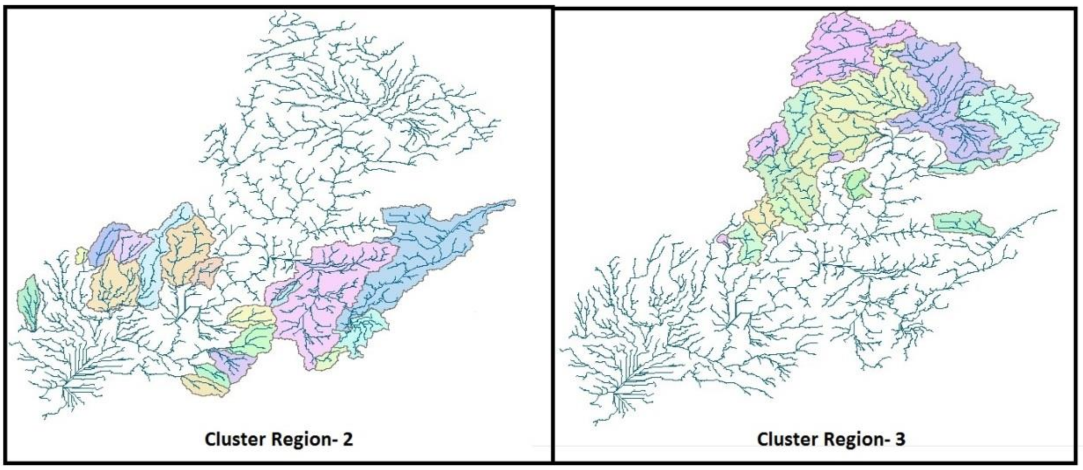


Figure 4.14 Physical Comparison of Cluster-2 and Cluster-3.

### 4.5 Stepwise Regression Analysis

#### 4.5.1 Methodology of Stepwise Regression Analysis

Stepwise regression is an automatic regression algorithm that enters X variables into the regression model, one X variable at a time. The X variables are entered based on statistical criteria, usually partial F ratios and their corresponding p values (Schmee and Openlander, 2010).

The objective of stepwise regression is to develop a prediction equation relating a criterion (dependent) variable to *p* predictor variables Although it is a type of multiple regression analysis, it differs from the commonly used multiple regression technique in that stepwise regression, in addition to calibrating a prediction equation, introduces predictor variables sequentially based on a partial-F statistic; thus, stepwise regression analysis yields *p* prediction equations from which one must be selected as the “best” model. The multiple regression technique includes all available predictor variables in the equation but is often plagued by irrational regression coefficients because of the high intercorelation between the predictor variables. Stepwise regression usually avoids the irrational coefficients because the final model can be selected so that only predictor variables with low intercorelation are included.

The Forward Stepwise Regression Analysis is applied by using SPSS program. The algorithm and steps of this analysis is given as follows.

If there are independent variables currently entered in the model, choose $X_k$ such that F−to−remove$_k$ is minimum. $X_k$ is removed if F−to−remove$_k$ < $F_{out}$ (default = 2.71) or, if probability criteria are used, P(F−to−remove$_k$) > $P_{out}$ (default = 0.1). If the inequality does not hold, no variable is removed from the model.

If there are no independent variables currently entered in the model or if no entered variable is to be removed, choose $X_k$ such that F−to−enter$_k$ is maximum. $X_k$ is entered if F−to−enter$_k$ > $F_{in}$ (default = 3.84) or, P(F−to−enter$_k$) < $P_{in}$ (default = 0.05). If the inequality does not hold, no variable is entered.

At each step, all eligible variables are considered for removal and entry.

Stepwise (STEP) logistic regression algorithms:

1. If STEP is the first method requested, estimate the parameter and likelihood function for the initial model. Otherwise, the final model from the previous method is the initial model for STEP. Obtain the necessary information: Maximum Likelihood Estimates(MLEs) of the parameters for the current model, predicted probability, likelihood function for the current model, and so on.

2. Based on the MLEs of the current model, calculate the score statistic for every variable eligible for inclusion and find its significance.

3. Choose the variable with the smallest significance. If that significance is less than the probability for a variable to enter, then go to step 4; otherwise, stop STEP.

4. Update the current model by adding a new variable. If this results in a model which has already been evaluated, stop STEP.

5. Calculate LR or Wald statistic or conditional statistic for each variable in the current model. Then calculate its corresponding significance.

6. Choose the variable with the largest significance. If that significance is less than the probability for variable removal, then go back to step 2; otherwise, if the current model with the variable deleted is the same as a previous model, stop STEP; otherwise, go to the next step.

7. Modify the current model by removing the variable with the largest significance from the previous model. Estimate the parameters for the modified model and go back to step 5.

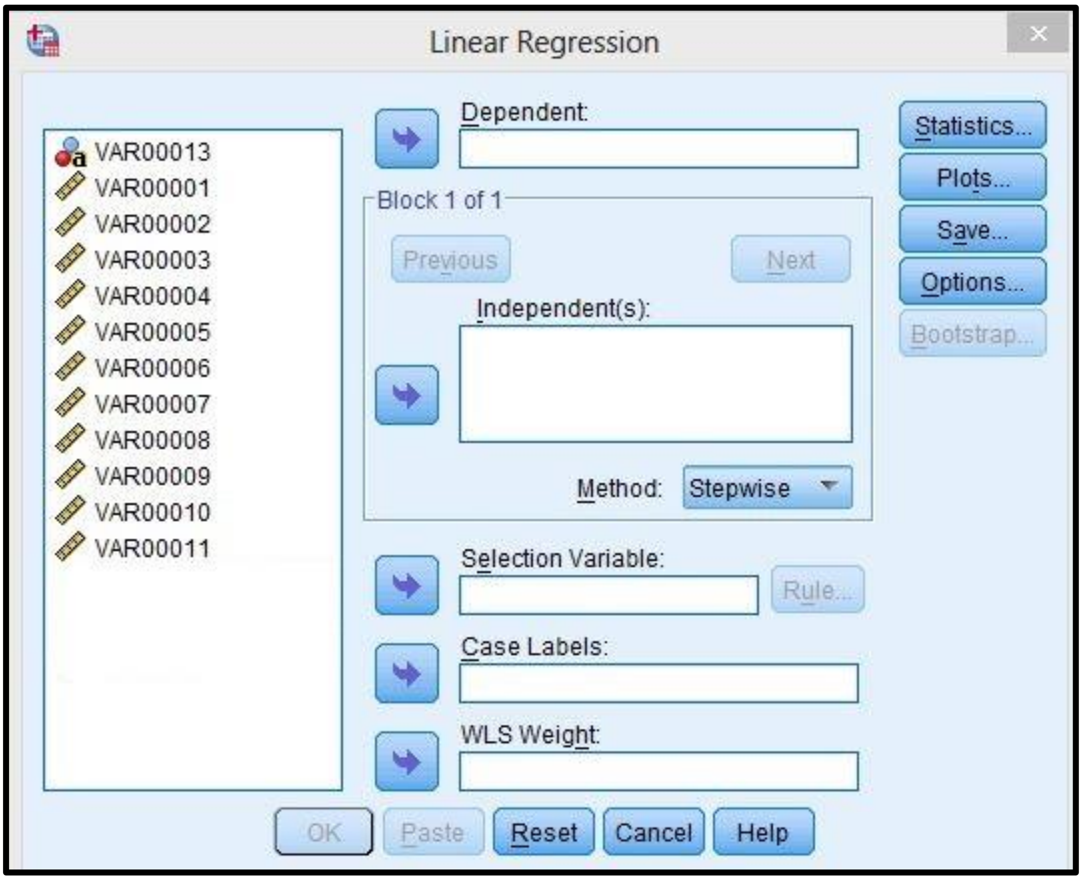


Figure 4.15 Stepwise Regression Analysis Main Menu in SPSS.

### 4.5.2 Stepwise Regression Analysis Calculations

After the cluster regions are obtained by Twostep Clustering Analysis, the important basin characteristics and representative model equations should be obtained by Stepwise Regression Analysis which is explained in part 4.5. However, $Q_{2.33}$ flood values are obtained with Gumbel Distribution and different distribution parameters. Therefore, two models should be obtained to compare the results. First model is developed by the $Q_{2.33}$ flood values which are obtained by Gumbel Distribution, and the second model is developed by the $Q_{2.33}$ flood values which are obtained by different distribution parameters.

In addition, to see the effect of clustering, these analysis are done with whole region stations and also are applied for the clustered region in their own right. Therefore, one equation is obtained for 40 stations which are the selected station number in the region and three equations are obtained for the cluster regions respectively. At the end of the Stepwise Regression Analysis, four equations where one of them is for whole region and others for the three cluster regions are obtained for Gumbel distribution and for different distributions.

#### 4.5.2.1 Stepwise Regression Analysis with Gumbel Distribution $Q_{2.33}$ values

For this analysis, the basin characteristics and $Q_{2.33}$ values of stations which are given in Table 3.3 are used for whole region analysis. Cluster region basin characteristics and $Q_{2.33}$ values which is obtained by Gumbel Distribution are given in Table 4.16.

Table 4.16 Cluster Region Basin Characteristics and $Q_{2.33}$ Flood Values with Gumbel Distribution.

| Cluster Region | Station | AREA ($km^2$) | Perimeter (km) | Length of Main River (km) | Annual Mean Rainfall (mm) | $H_{min}$ (m) | $H_{max}$ (m) | $H_{mean}$ (m) | $H_{\Delta}$ (m) | Basin Relief (m) | Relative Relief | Circularity | $Q_{2.33}$ ($m^3/s$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | Y |
| # 1 | EIE/2001 | 8484 | 742 | 209.83 | 540.72 | 420 | 3073 | 1554.32 | 2653 | 1134.32 | 0.1529 | 0.1936 | 552.95 |
| | EIE/2004 | 20466 | 1466 | 385.76 | 665.98 | 2 | 3073 | 1027.29 | 3071 | 1025.29 | 0.0699 | 0.1197 | 1027.82 |
| | EIE/2012 | 19727.2 | 1417 | 371.98 | 662.04 | 2 | 3073 | 1048.80 | 3071 | 1046.80 | 0.0739 | 0.1235 | 1094.73 |
| # 2 | EIE/2007 | 623 | 196 | 59.58 | 830.54 | 7 | 1959 | 441.71 | 1952 | 434.71 | 0.2218 | 0.2038 | 59.86 |
| | EIE/2008 | 444 | 209 | 64.68 | 835.72 | 56 | 2244 | 908.44 | 2188 | 852.44 | 0.4079 | 0.1277 | 160.70 |
| | EIE/2010 | 3498.8 | 546 | 147.34 | 644.19 | 401 | 2436 | 871.73 | 2035 | 470.73 | 0.0862 | 0.1475 | 287.69 |
| | DSI/20-04 | 178 | 88 | 35.19 | 830.57 | 214 | 1953 | 903.84 | 1739 | 689.84 | 0.7839 | 0.2888 | 38.11 |
| | DSI/20-05 | 94 | 61 | 17.66 | 772.37 | 196 | 2070 | 1178.83 | 1874 | 982.83 | 1.6112 | 0.3175 | 45.57 |
| | DSI/20-06 | 174.9 | 80 | 21.32 | 800.25 | 340 | 2215 | 1184.68 | 1875 | 844.68 | 1.0558 | 0.3434 | 57.01 |
| | DSI/20-08 | 131.1 | 72 | 21.03 | 774.20 | 211 | 2140 | 1283.14 | 1929 | 1072.14 | 1.4891 | 0.3178 | 39.33 |
| | DSI/20-10 | 217.3 | 95 | 25.77 | 790.71 | 486 | 1782 | 1045.02 | 1296 | 559.02 | 0.5884 | 0.3026 | 62.53 |
| | DSI/20-13 | 105.1 | 75 | 18.52 | 862.97 | 183 | 1602 | 840.45 | 1419 | 657.45 | 0.8766 | 0.2348 | 50.53 |
| | DSI/20-14 | 310.5 | 126 | 27.93 | 725.30 | 417 | 2244 | 1269.95 | 1827 | 852.95 | 0.6769 | 0.2458 | 163.28 |
| | DSI/20-17 | 1740 | 479 | 103.63 | 594.89 | 478 | 2436 | 992.20 | 1958 | 514.20 | 0.1073 | 0.0953 | 121.32 |
| | DSI/20-40 | 79 | 63 | 14.51 | 716.34 | 617 | 1419 | 971.30 | 802 | 354.30 | 0.5624 | 0.2501 | 5.66 |
| | DSI/20-43 | 163 | 87 | 19.45 | 775.15 | 213 | 1959 | 1012.86 | 1746 | 799.86 | 0.9194 | 0.2706 | 68.16 |
| | DSI/20-44 | 35 | 33 | 5.94 | 735.88 | 300 | 1353 | 750.89 | 1053 | 450.89 | 1.3663 | 0.4039 | 28.96 |
| | DSI/20-45 | 170 | 90 | 25.75 | 750.63 | 274 | 1938 | 719.15 | 1664 | 445.15 | 0.4946 | 0.2637 | 67.52 |
| | DSI/20-46 | 477 | 146 | 40.65 | 758.38 | 171 | 2244 | 1086.26 | 2073 | 915.26 | 0.6269 | 0.2812 | 207.68 |
| | DSI/20-56 | 238.4 | 145 | 25.47 | 645.65 | 488 | 1466 | 804.98 | 978 | 316.98 | 0.2186 | 0.1425 | 21.53 |
| | DSI/20-63 | 57.5 | 60 | 14.62 | 694.28 | 530 | 1522 | 992.00 | 992 | 462.00 | 0.7700 | 0.2007 | 14.61 |
| | DSI/20-65 | 161 | 89 | 26.34 | 737.03 | 71 | 698 | 257.43 | 627 | 186.43 | 0.2095 | 0.2554 | 41.45 |

Table 4.16 Cluster Region Basin Characteristics and $Q_{2.33}$ Flood Values with Gumbel Distribution. (Continued)

| Cluster Region | Station | AREA ($km^2$) | Perimeter (km) | Length of Main River (km) | Annual Mean Rainfall (mm) | $H_{min}$ (m) | $H_{max}$ (m) | $H_{mean}$ (m) | $H_{\Delta}$ (m) | Basin Relief (m) | Relative Relief | Circularity | $Q_{2.33}$ ($m^3/s$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | Y |
| # 3 | EIE/2005 | 4219.08 | 516 | 110.55 | 473.06 | 1059 | 3073 | 1603.12 | 2014 | 544.12 | 0.1054 | 0.1991 | 111.42 |
| | EIE/2006 | 733.2 | 232 | 49.305 | 557.89 | 1313 | 2941 | 1828.41 | 1628 | 515.41 | 0.2222 | 0.1712 | 53.43 |
| | EIE/2009 | 1387.2 | 300 | 101.57 | 560.21 | 1102 | 2941 | 1672.16 | 1839 | 570.16 | 0.1901 | 0.1937 | 85.06 |
| | EIE/2015 | 915.2 | 277 | 62.62 | 471.80 | 1209 | 2851 | 1826.80 | 1642 | 617.80 | 0.2230 | 0.1499 | 43.95 |
| | EIE/2016 | 846.8 | 242 | 66.44 | 481.87 | 1153 | 3073 | 1640.36 | 1920 | 487.36 | 0.2014 | 0.1817 | 83.56 |
| | EIE/2022 | 428 | 149 | 37.98 | 517.23 | 1335 | 3073 | 1808.58 | 1738 | 473.58 | 0.3178 | 0.2423 | 68.93 |
| | DSI/20-02 | 197.1 | 99 | 27.93 | 555.61 | 889 | 2436 | 1473.67 | 1547 | 584.67 | 0.5906 | 0.2527 | 40.83 |
| | DSI/20-07 | 2084 | 340 | 92.92 | 474.08 | 1095 | 2887 | 1661.00 | 1792 | 566.00 | 0.1665 | 0.2265 | 54.54 |
| | DSI/20-15 | 189.7 | 93 | 22.99 | 558.76 | 1350 | 2737 | 1785.93 | 1387 | 435.93 | 0.4687 | 0.2756 | 26.15 |
| | DSI/20-16 | 291 | 124 | 38.54 | 537.76 | 1363 | 2941 | 2092.48 | 1578 | 729.48 | 0.5883 | 0.2378 | 35.84 |
| | DSI/20-36 | 174.2 | 102 | 19.52 | 626.09 | 1232 | 2340 | 1634.99 | 1108 | 402.99 | 0.3951 | 0.2104 | 62.79 |
| | DSI/20-51 | 131.4 | 96 | 31.09 | 546.21 | 1099 | 3069 | 1824.14 | 1970 | 725.14 | 0.7554 | 0.1792 | 10.14 |
| | DSI/20-52 | 23 | 27 | 3.25 | 605.52 | 1275 | 2411 | 1629.32 | 1136 | 354.32 | 1.3123 | 0.3965 | 6.12 |
| | DSI/20-53 | 178.5 | 103 | 20.95 | 598.75 | 1060 | 2379 | 1655.54 | 1319 | 595.54 | 0.5782 | 0.2114 | 38.27 |
| | DSI/20-54 | 207.5 | 88 | 24.54 | 651.58 | 758 | 2470 | 1571.25 | 1712 | 813.25 | 0.9241 | 0.3367 | 58.87 |
| | DSI/20-55 | 111.6 | 74 | 17.11 | 601.50 | 830 | 3010 | 1906.45 | 2180 | 1076.45 | 1.4547 | 0.2561 | 21.06 |
| | DSI/20-58 | 24.38 | 29 | 5.3 | 673.66 | 970 | 2171 | 1556.12 | 1201 | 586.12 | 2.0211 | 0.3643 | 18.38 |
| | DSI/20-59 | 171.5 | 97 | 20.12 | 697.23 | 1018 | 2244 | 1362.08 | 1226 | 344.08 | 0.3547 | 0.2291 | 62.49 |

**i) Stepwise Regression Analysis with Gumbel Distribution Q2.33 values for whole region**

In this analysis, basin characteristics for 40 stations given in Table 3.3 are used in Stepwise Regression Analysis. The standard procedure of Forward Stepwise Regression Analysis procedure which is explained in part 4.5 is applied to the data. The summary of selected variables are given in Table 4.17

Table 4.17 Summary of Stepwise Regression Analysis for Whole Region

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | ,977[a] | ,954 | ,953 | 51,44044 | ,954 | 790,032 | 1 | 38 | ,000 |
| 2 | ,981[b] | ,963 | ,961 | 46,64894 | ,009 | 9,207 | 1 | 37 | ,004 |

a. Predictors: (Constant), VAR00001

b. Predictors: (Constant), VAR00001, VAR00005

c. Dependent Variable: VAR00012

As one can see from the Table 4.17, two main parameters which are Area and $H_{min}$ are selected as the main basin characteristics. These two basin characteristics equation constants are given in Table 4.18

Table 4.18 Stepwise Regression Analysis Results Equation Constants.

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Zero-order | Partial | Part | Tolerance | VIF |
| 1 | (Constant) | 38,213 | 8,731 | | 4,377 | ,000 | | | | | |
| | VAR00001 | ,051 | ,002 | ,977 | 28,108 | ,000 | ,977 | ,977 | ,977 | 1,000 | 1,000 |
| 2 | (Constant) | 74,966 | 14,471 | | 5,180 | ,000 | | | | | |
| | VAR00001 | ,049 | ,002 | ,945 | 28,514 | ,000 | ,977 | ,978 | ,899 | ,903 | 1,107 |
| | VAR00005 | -,052 | ,017 | -,101 | -3,034 | ,004 | -,395 | -,446 | -,096 | ,903 | 1,107 |

a. Dependent Variable: VAR00012

The model equation which contain two parameters is given below:

$\mathbf{Q_{2.33} = 74.966 + 0.049 * AREA - 0.052* H_{min}}$ (Equation 4.15)

**ii) Stepwise Regression Analysis with Gumbel Distribution Q2.33 values for cluster regions.**

This procedure is applied for the three cluster regions separately. First, 3 station's basin characteristics is used to make Stepwise Regression Analysis to get an equation for this cluster. Then, 19 station's basin characteristics, which are named as Cluster-2, are used to make analysis to get an equation for this cluster. Lastly, 18 station's basin characteristics, which are named as Cluster-3, are used to make analysis to get an equation for this cluster. At the end of the analysis, three different equations for each cluster are obtainedAll stations basin characteristics and cluster regions are given in Table 4.16.

- **Stepwise Regression Analysis for Cluster-1**

This cluster have 3 stations. However, Stepwise Regression Analysis is not applied for this cluster, because of less number of stations. Therefore, there is not any Stepwise Regression Equation for cluster-1.

- **Stepwise Regression Analysis for Cluster-2**

This cluster region has 19 stations which are shown in Table 4.16. These stations generally located at bottom parts of the basin. The same Stepwise Regression Analysis procedure is applied to these stations. The summary and selected variables are shown in Table 4.19 and Table 4.20, respectively.

Table 4.19 Summary of Stepwise Regression Analysis for Cluster-2.

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | ,768[a] | ,590 | ,565 | 48,86147 | ,590 | 24,416 | 1 | 17 | ,000 |
| 2 | ,835[b] | ,697 | ,659 | 43,28614 | ,107 | 5,661 | 1 | 16 | ,030 |

a. Predictors: (Constant), VAR0003

b. Predictors: (Constant), VAR0003, VAR0009

c. Dependent Variable: VAR00012

Table 4.20 Selected Variables and Resultant Equation Constants for Cluster-2.

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Zero-order | Partial | Part | Tolerance | VIF |
| 1 | (Constant) | 19,811 | 16,723 | | 1,185 | ,252 | | | | | |
| | VAR0003 | 1,629 | ,330 | ,768 | 4,941 | ,000 | ,768 | ,768 | ,768 | 1,000 | 1,000 |
| 2 | (Constant) | -44,309 | 30,752 | | -1,441 | ,169 | | | | | |
| | VAR0003 | 1,711 | ,294 | ,807 | 5,819 | ,000 | ,768 | ,824 | ,801 | ,986 | 1,014 |
| | VAR0009 | ,098 | ,041 | ,330 | 2,379 | ,030 | ,235 | ,511 | ,328 | ,986 | 1,014 |

a. Dependent Variable: VAR00012

After applying this procedure to the Cluster-2, the resultant equation become as follows:

$\mathbf{Q_{2.33}}$ **= -44.309 + 1.711 * (Length of Main River) + 0.098 * Basin Relief** (Equation 4.16)

In Table 5.1, this equation which is derived from using Stepwise Regression Analysis is used for the calculation of flood values which are in the second cluster region.

- **Stepwise Regression Analysis for Cluster-3**

This cluster region has 18 stations which are shown in Table 4.16. These stations generally located at upper parts of the basin. The same Stepwise Regression Analysis procedure is applied to these stations. The summary and selected variables are shown in Table 4.21 and Table 4.22, respectively.

Table 4.21 Summary of Stepwise Regression Analysis for Cluster-3.

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | ,760[a] | ,578 | ,551 | 18,71691 | ,578 | 21,902 | 1 | 16 | ,000 |

a. Predictors: (Constant), VAR00002

b. Dependent Variable: VAR00012

Table 4.22 Selected Variables and Resultant Equation Constants for Cluster-3.

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Zero-order | Partial | Part | Tolerance | VIF |
| 1 | (Constant) | 21,284 | 7,383 | | 2,883 | ,011 | | | | | |
| | VAR00002 | ,167 | ,036 | ,760 | 4,680 | ,000 | ,760 | ,760 | ,760 | 1,000 | 1,000 |

a. Dependent Variable: VAR00012

After applying this procedure to the Cluster-3, the resultant equation becomes:

$\mathbf{Q_{2.33}}$ **= 21.284 + 0.167 * Perimeter** (Equation 4.17)

#### 4.5.2.2 Stepwise Regression Analysis with Different Distribution Parameter $Q_{2.33}$ values

For this analysis, the basin characteristics and $Q_{2.33}$ values of stations which are given in Table 3.3 are used for whole region analysis. In this way, different distributions for $Q_{2.33}$ values which are given in Table 4.3 are used rather than Gumbel Distribution. However, the whole dependent and independent variables are given in Table 4.23 for the comprehensibility of the analysis.

Table 4.23 Cluster Region Basin Characteristics and $Q_{2.33}$ Flood Values with Different Distribution Parameters.

| Cluster Region | Station | AREA ($km^2$) | Perimeter (km) | Length of Main River (km) | Annual Mean Rainfall (mm) | $H_{min}$ (m) | $H_{max}$ (m) | $H_{mean}$ (m) | $H_{\Delta}$ (m) | Basin Relief (m) | Relative Relief | Circularity | $Q_{2.33}$ ($m^3/s$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | Y |
| # 1 | EIE/2001 | 8484 | 742 | 209.83 | 540.72 | 420 | 3073 | 1554.32 | 2653 | 1134.32 | 0.1529 | 0.1936 | 536.54 |
| | EIE/2004 | 20466 | 1466 | 385.76 | 665.98 | 2 | 3073 | 1027.29 | 3071 | 1025.29 | 0.0699 | 0.1197 | 1089.20 |
| | EIE/2012 | 19727.2 | 1417 | 371.98 | 662.04 | 2 | 3073 | 1048.80 | 3071 | 1046.80 | 0.0739 | 0.1235 | 1089.68 |
| # 2 | EIE/2007 | 623 | 196 | 59.58 | 830.54 | 7 | 1959 | 441.71 | 1952 | 434.71 | 0.2218 | 0.2038 | 58.86 |
| | EIE/2008 | 444 | 209 | 64.68 | 835.72 | 56 | 2244 | 908.44 | 2188 | 852.44 | 0.4079 | 0.1277 | 146.34 |
| | EIE/2010 | 3498.8 | 546 | 147.34 | 644.19 | 401 | 2436 | 871.73 | 2035 | 470.73 | 0.0862 | 0.1475 | 285.82 |
| | DSI/20-04 | 178 | 88 | 35.19 | 830.57 | 214 | 1953 | 903.84 | 1739 | 689.84 | 0.7839 | 0.2888 | 28.54 |
| | DSI/20-05 | 94 | 61 | 17.66 | 772.37 | 196 | 2070 | 1178.83 | 1874 | 982.83 | 1.6112 | 0.3175 | 36.82 |
| | DSI/20-06 | 174.9 | 80 | 21.32 | 800.25 | 340 | 2215 | 1184.68 | 1875 | 844.68 | 1.0558 | 0.3434 | 51.84 |
| | DSI/20-08 | 131.1 | 72 | 21.03 | 774.20 | 211 | 2140 | 1283.14 | 1929 | 1072.14 | 1.4891 | 0.3178 | 28.72 |
| | DSI/20-10 | 217.3 | 95 | 25.77 | 790.71 | 486 | 1782 | 1045.02 | 1296 | 559.02 | 0.5884 | 0.3026 | 62.53 |
| | DSI/20-13 | 105.1 | 75 | 18.52 | 862.97 | 183 | 1602 | 840.45 | 1419 | 657.45 | 0.8766 | 0.2348 | 46.61 |
| | DSI/20-14 | 310.5 | 126 | 27.93 | 725.30 | 417 | 2244 | 1269.95 | 1827 | 852.95 | 0.6769 | 0.2458 | 149.17 |
| | DSI/20-17 | 1740 | 479 | 103.63 | 594.89 | 478 | 2436 | 992.20 | 1958 | 514.20 | 0.1073 | 0.0953 | 86.40 |
| | DSI/20-40 | 79 | 63 | 14.51 | 716.34 | 617 | 1419 | 971.30 | 802 | 354.30 | 0.5624 | 0.2501 | 4.68 |
| | DSI/20-43 | 163 | 87 | 19.45 | 775.15 | 213 | 1959 | 1012.86 | 1746 | 799.86 | 0.9194 | 0.2706 | 50.79 |
| | DSI/20-44 | 35 | 33 | 5.94 | 735.88 | 300 | 1353 | 750.89 | 1053 | 450.89 | 1.3663 | 0.4039 | 19.68 |
| | DSI/20-45 | 170 | 90 | 25.75 | 750.63 | 274 | 1938 | 719.15 | 1664 | 445.15 | 0.4946 | 0.2637 | 55.11 |
| | DSI/20-46 | 477 | 146 | 40.65 | 758.38 | 171 | 2244 | 1086.26 | 2073 | 915.26 | 0.6269 | 0.2812 | 198.10 |
| | DSI/20-56 | 238.4 | 145 | 25.47 | 645.65 | 488 | 1466 | 804.98 | 978 | 316.98 | 0.2186 | 0.1425 | 22.30 |
| | DSI/20-63 | 57.5 | 60 | 14.62 | 694.28 | 530 | 1522 | 992.00 | 992 | 462.00 | 0.7700 | 0.2007 | 12.22 |
| | DSI/20-65 | 161 | 89 | 26.34 | 737.03 | 71 | 698 | 257.43 | 627 | 186.43 | 0.2095 | 0.2554 | 39.47 |

Table 4.23 Cluster Region Basin Characteristics and $Q_{2.33}$ Flood Values with Different Distribution Parameters. (Continued)

| Cluster Region | Station | AREA ($km^2$) | Perimeter (km) | Length of Main River (km) | Annual Mean Rainfall (mm) | $H_{min}$ (m) | $H_{max}$ (m) | $H_{mean}$ (m) | $H_{\Delta}$ (m) | Basin Relief (m) | Relative Relief | Circularity | $Q_{2.33}$ ($m^3/s$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | Y |
| # 3 | EIE/2005 | 4219.08 | 516 | 110.55 | 473.06 | 1059 | 3073 | 1603.12 | 2014 | 544.12 | 0.1054 | 0.1991 | 108.41 |
| | EIE/2006 | 733.2 | 232 | 49.305 | 557.89 | 1313 | 2941 | 1828.41 | 1628 | 515.41 | 0.2222 | 0.1712 | 53.55 |
| | EIE/2009 | 1387.2 | 300 | 101.57 | 560.21 | 1102 | 2941 | 1672.16 | 1839 | 570.16 | 0.1901 | 0.1937 | 79.56 |
| | EIE/2015 | 915.2 | 277 | 62.62 | 471.80 | 1209 | 2851 | 1826.80 | 1642 | 617.80 | 0.2230 | 0.1499 | 38.34 |
| | EIE/2016 | 846.8 | 242 | 66.44 | 481.87 | 1153 | 3073 | 1640.36 | 1920 | 487.36 | 0.2014 | 0.1817 | 73.65 |
| | EIE/2022 | 428 | 149 | 37.98 | 517.23 | 1335 | 3073 | 1808.58 | 1738 | 473.58 | 0.3178 | 0.2423 | 53.93 |
| | DSI/20-02 | 197.1 | 99 | 27.93 | 555.61 | 889 | 2436 | 1473.67 | 1547 | 584.67 | 0.5906 | 0.2527 | 36.54 |
| | DSI/20-07 | 2084 | 340 | 92.92 | 474.08 | 1095 | 2887 | 1661.00 | 1792 | 566.00 | 0.1665 | 0.2265 | 40.86 |
| | DSI/20-15 | 189.7 | 93 | 22.99 | 558.76 | 1350 | 2737 | 1785.93 | 1387 | 435.93 | 0.4687 | 0.2756 | 16.90 |
| | DSI/20-16 | 291 | 124 | 38.54 | 537.76 | 1363 | 2941 | 2092.48 | 1578 | 729.48 | 0.5883 | 0.2378 | 23.80 |
| | DSI/20-36 | 174.2 | 102 | 19.52 | 626.09 | 1232 | 2340 | 1634.99 | 1108 | 402.99 | 0.3951 | 0.2104 | 62.79 |
| | DSI/20-51 | 131.4 | 96 | 31.09 | 546.21 | 1099 | 3069 | 1824.14 | 1970 | 725.14 | 0.7554 | 0.1792 | 8.71 |
| | DSI/20-52 | 23 | 27 | 3.25 | 605.52 | 1275 | 2411 | 1629.32 | 1136 | 354.32 | 1.3123 | 0.3965 | 4.97 |
| | DSI/20-53 | 178.5 | 103 | 20.95 | 598.75 | 1060 | 2379 | 1655.54 | 1319 | 595.54 | 0.5782 | 0.2114 | 36.37 |
| | DSI/20-54 | 207.5 | 88 | 24.54 | 651.58 | 758 | 2470 | 1571.25 | 1712 | 813.25 | 0.9241 | 0.3367 | 53.02 |
| | DSI/20-55 | 111.6 | 74 | 17.11 | 601.50 | 830 | 3010 | 1906.45 | 2180 | 1076.45 | 1.4547 | 0.2561 | 21.06 |
| | DSI/20-58 | 24.38 | 29 | 5.3 | 673.66 | 970 | 2171 | 1556.12 | 1201 | 586.12 | 2.0211 | 0.3643 | 18.38 |
| | DSI/20-59 | 171.5 | 97 | 20.12 | 697.23 | 1018 | 2244 | 1362.08 | 1226 | 344.08 | 0.3547 | 0.2291 | 52.43 |

**i) Stepwise Regression Analysis with Different Distribution Parameter $Q_{2.33}$ values for whole region**

In this analysis, 40 station's basin characteristics which are given in Table 4.23 are used in Stepwise Regression Analysis. The standard procedure of Forward Stepwise Regression Analysis procedure which is explained in previous part is applied to the data. In this analysis, cluster regions are not taken into consideration.

The summary of selected variables and equation constants are given in Table 4.24 and Table 4.25, respectively.

Table 4.24 Summary of Stepwise Regression Analysis for Whole Region

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | ,980[a] | ,961 | ,960 | 48,71020 | ,961 | 934,512 | 1 | 38 | ,000 |
| 2 | ,985[b] | ,970 | ,968 | 43,29528 | ,009 | 11,100 | 1 | 37 | ,002 |

a. Predictors: (Constant), VAR00001

b. Predictors: (Constant), VAR00001, VAR00004

c. Dependent Variable: VAR00012

Table 4.25 Selected Variables and Resultant Equation Constants for Whole Region.

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Zero-order | Partial | Part | Tolerance | VIF |
| 1 | (Constant) | 30,143 | 8,268 | | 3,646 | ,001 | | | | | |
| | VAR00001 | ,053 | ,002 | ,980 | 30,570 | ,000 | ,980 | ,980 | ,980 | 1,000 | 1,000 |
| 2 | (Constant) | -105,808 | 41,463 | | -2,552 | ,015 | | | | | |
| | VAR00001 | ,053 | ,002 | ,991 | 34,547 | ,000 | ,980 | ,985 | ,985 | ,986 | 1,014 |
| | VAR00004 | ,205 | ,062 | ,096 | 3,332 | ,002 | -,021 | ,480 | ,095 | ,986 | 1,014 |

a. Dependent Variable: VAR00012

The model equation which contains two parameters is given below:

$$Q_{2.33} = -105.808 + 0.053 * AREA + 0.205 * MAR \qquad \text{(Equation 4.18)}$$

This analysis is for whole region, therefore drainage area has an important role in the Flood Analysis.

**ii) Stepwise Regression Analysis with Different Distribution Parameter $Q_{2.33}$ values for cluster regions.**

This procedure is applied for the three cluster regions separately. First, 3 station's basin characteristics are used to make Stepwise Regression Analysis to get an equation for this cluster. Then, 19 station's basin characteristics, which are named as Cluster-2, are used to make analysis to get an equation for this cluster. Lastly, 18 station's basin characteristics, which are named as Cluster-3, are used to make analysis to get an equation for this cluster. At the end of the analysis, three equations are obtained from the results for different cluster regions. All station's basin characteristics and cluster regions are given in Table 4.23.

- **Stepwise Regression Analysis for Cluster-1**

This cluster has 3 stations. Stepwise Regression Analysis is not applied for this cluster, because of less number of station's value. Therefore, there is not any Stepwise Regression Equation for cluster-1.

- **Stepwise Regression Analysis for Cluster-2**

This cluster region has 19 stations which are shown in Table 4.23. These stations generally located at bottom parts of the basin. The same Stepwise Regression Analysis procedure is applied to these stations. The summary and selected variables are shown in Table 4.26 and Table 4.27, respectively.

Table 4.26 Summary of Stepwise Regression Analysis for Cluster-2.

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | ,755[a] | ,571 | ,545 | 48,90594 | ,571 | 22,600 | 1 | 17 | ,000 |
| 2 | ,822[b] | ,676 | ,635 | 43,81851 | ,105 | 5,177 | 1 | 16 | ,037 |
| 3 | ,873[c] | ,763 | ,715 | 38,71513 | ,087 | 5,496 | 1 | 15 | ,033 |
| 4 | ,920[d] | ,847 | ,803 | 32,16112 | ,084 | 7,737 | 1 | 14 | ,015 |

a. Predictors: (Constant), VAR00001

b. Predictors: (Constant), VAR00001, VAR00009

c. Predictors: (Constant), VAR00001, VAR00009, VAR00010

d. Predictors: (Constant), VAR00001, VAR00009, VAR00010, VAR00002

e. Dependent Variable: VAR00012

Table 4.27 Selected Variables and Result Equation Constants for Cluster-2.

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Zero-order | Partial | Part | Tolerance | VIF |
| 1 | (Constant) | 41,807 | 12,981 | | 3,221 | ,005 | | | | | |
| | VAR00001 | ,066 | ,014 | ,755 | 4,754 | ,000 | ,755 | ,755 | ,755 | 1,000 | 1,000 |
| 2 | (Constant) | -19,703 | 29,430 | | -,669 | ,513 | | | | | |
| | VAR00001 | ,071 | ,013 | ,807 | 5,598 | ,000 | ,755 | ,814 | ,797 | ,975 | 1,025 |
| | VAR00009 | ,095 | ,042 | ,328 | 2,275 | ,037 | ,201 | ,494 | ,324 | ,975 | 1,025 |
| 3 | (Constant) | -8,091 | 26,470 | | -,306 | ,764 | | | | | |
| | VAR00001 | ,054 | ,013 | ,617 | 4,083 | ,001 | ,755 | ,726 | ,514 | ,694 | 1,442 |
| | VAR00009 | ,168 | ,048 | ,580 | 3,480 | ,003 | ,201 | ,668 | ,438 | ,570 | 1,756 |
| | VAR00010 | -71,820 | 30,635 | -,449 | -2,344 | ,033 | -,401 | -,518 | -,295 | ,431 | 2,320 |
| 4 | (Constant) | 45,655 | 29,273 | | 1,560 | ,141 | | | | | |
| | VAR00001 | ,137 | ,032 | 1,566 | 4,307 | ,001 | ,755 | ,755 | ,450 | ,083 | 12,096 |
| | VAR00009 | ,223 | ,045 | ,767 | 4,984 | ,000 | ,201 | ,800 | ,521 | ,461 | 2,171 |
| | VAR00010 | -128,032 | 32,497 | -,801 | -3,940 | ,001 | -,401 | -,725 | -,412 | ,264 | 3,784 |
| | VAR00002 | -,610 | ,219 | -1,160 | -2,781 | ,015 | ,699 | -,597 | -,291 | ,063 | 15,935 |

a. Dependent Variable: VAR00012

After applying this procedure to the Cluster-2, the resultant equation becomes:

$$Q_{2.33} = 45.655 + 0.137 * \text{Area} - 0.610 * \text{Perimeter} + 0.223 * \text{Basin Relief} - 128.032 * \text{Relative Relief} \quad \text{(Equation 4.19)}$$

- **Stepwise Regression Analysis for Cluster-3**

This cluster region has 18 stations which are shown in Table 4.23. These stations generally located at upper parts of the basin. The same Stepwise Regression Analysis procedure is applied to these stations. The summary and selected variables are shown in Table 4.28 and Table 4.29, respectively.

Table 4.28 Summary of Stepwise Regression Analysis for Cluster-3.

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | ,753[a] | ,566 | ,539 | 18,20751 | ,566 | 20,900 | 1 | 16 | ,000 |
| 2 | ,833[b] | ,693 | ,653 | 15,81182 | ,127 | 6,216 | 1 | 15 | ,025 |

a. Predictors: (Constant), VAR00002

b. Predictors: (Constant), VAR00002, VAR00004

c. Dependent Variable: VAR00012

Table 4.29 Selected Variables and Resultant Equation Constants for Cluster-3.

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Zero-order | Partial | Part | Tolerance | VIF |
| 1 | (Constant) | 17,186 | 7,182 | | 2,393 | ,029 | | | | | |
| | VAR00002 | ,159 | ,035 | ,753 | 4,572 | ,000 | ,753 | ,753 | ,753 | 1,000 | 1,000 |
| 2 | (Constant) | -109,065 | 51,022 | | -2,138 | ,049 | | | | | |
| | VAR00002 | ,237 | ,043 | 1,122 | 5,448 | ,000 | ,753 | ,815 | ,779 | ,482 | 2,076 |
| | VAR00004 | ,200 | ,080 | ,514 | 2,493 | ,025 | -,294 | ,541 | ,356 | ,482 | 2,076 |

a. Dependent Variable: VAR00012

After applying this procedure to the Cluster-3, the resultant equation becomes:

**$Q_{2.33}$ = - 109.065 + 0.237 * Perimeter + 0.200 *MAR** (Equation 4.20)

### 4.6 Nonlinear Regression Analysis

#### 4.6.1 Methodology of Nonlinear Regression Analysis

Parametric regression methods have been widely used for obtaining regional flood estimates. By using these methods, the relationship between the flood quantile $Q_T$ and the catchment characteristics are assumed to be the powerform function (Thomas and Benson, 1970) which has the following form (Shu, 2008).:

$$Q_T = aX_1^{\Theta_1}X_2^{\Theta_2}X_3^{\Theta_3} \ldots X_n^{\Theta_n}$$ (Equation 4.21)

where $X_i$ is the $i^{th}$ model parameter, a is the multiplicative error term and n is the number of catchment characteristics.

Using linear regression techniques generally requires linearizing the power-form model by a logarithmic transformation to the form. However, the estimation of the linearized model is theoretically unbiased in the logarithmic domain, but will be biased in the real flow domain (McCuen et al., 1990). Using nonlinear regression (NLR) methods, model parameters can be directly estimated by minimizing the estimation error in the actual flow domain. Nonlinear regression, with a properly selected objective function, can generally provide more accurate estimates than linear regression (Nguyen and Pandey, 1999; Grover et al., 2002).

Nonlinear regression is a method of finding a nonlinear model of the relationship between the dependent variable and a set of independent variables. Unlike traditional linear regression, which is restricted to estimating linear models, nonlinear regression can estimate models with arbitrary relationships between independent and dependent variables. This is accomplished using iterative estimation algorithms.

Nonlinear regression is appropriate when the relationship between the dependent and independent variables is not intrinsically linear.

Regression models, whether linear or nonlinear, assume that the form of the model is Y=F(X,B)+error, where Y is the dependent variable, X represents the predictors, and F is a function of X. In linear models, F is of the form

$$F(X,B) = \sum_{j=1}^{p} b_j x_j \qquad \text{(Equation 4.22)}$$

where $X_j$ is the $j^{th}$ predictor, and $b_j$ is the $j^{th}$ regression coefficient. Note that for a model to be considered linear, F must be a linear function of the parameters, not necessarily the predictors. Thus, $y=bx^2+error$ is a linear model. Additionally, some models in which the error is multiplicative, such as $y=e^{bx}error$, are linear models under the log-transformation: $\ln(y)=bx+\ln_{(error).}$ These models are known as intrinsically linear. Nonlinear models are all other forms of F.

Therefore, Nonlinear Linear Regression Analysis is done by after the Stepwise Regression Analysis. It is because of that Stepwise Regression Analysis selected the best basin characteristics which shows more dependency for whole region and different cluster regions. Than these selected basin characteristics are used in the Nonlinear Regression Analysis to find a nonlinear equation.

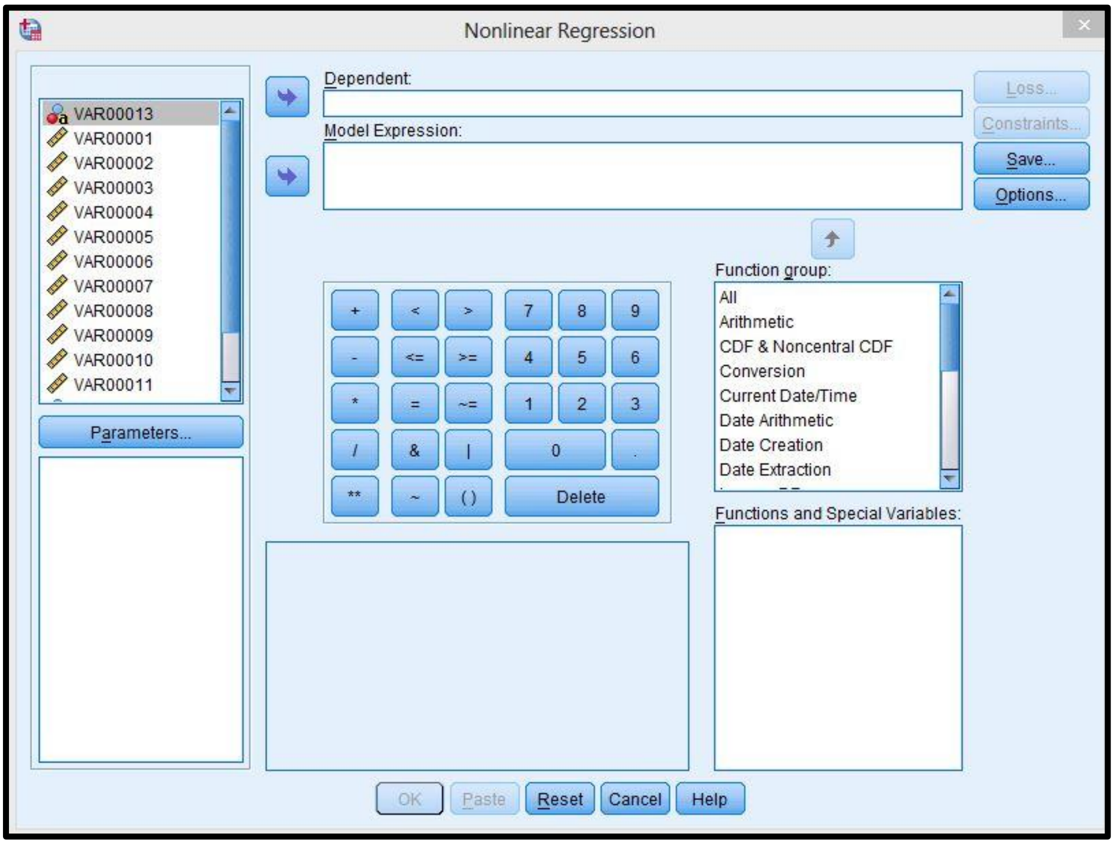


Figure 4.16 Nonlinear Regression Analysis Main Menu in SPSS.

### 4.6.2 Nonlinear Regression Analysis Calculations

Stepwise Regression Analysis has linear equation. However, in basin characteristics model, there are a lot of basin characteristics. Therefore, these characteristics should have nonlinear relationship. Nonlinear Regression Analysis is applied for the characteristics which are found by Stepwise Regression Analysis. This analysis is also done for the Gumbel Distribution and Different Distribution parameters.

#### 4.6.2.1 Nonlinear Regression Analysis for $Q_{2.33}$ values with Gumbel Distributions

The basin characteristics data and $Q_{2.33}$ values are shown in Table 4.16. First this analysis is done for whole region, than this is applied for all cluster regions.

**i) Nonlinear Regression Analysis for $Q_{2.33}$ values with Gumbel Distributions for whole region**

In the Stepwise Regression Part, two main parameters which are Area and $H_{min}$ are selected as the main basin characteristics in part 4.5.2. Therefore, these two parameters are used as nonlinear equation input data.

For the equation which is entered the SPSS program is:

$Q_{2.33} = b0 * Area^{b1} * H_{min}{}^{b2}$

After the running of the SPSS Nonlinear Analysis tool the summary of results is shown in Table 4.30.

Table 4.30 Summary of Nonlinear Regression Analysis Estimated Parameters

| Parameter | Estimate | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|
| | | | Lower Bound | Upper Bound |
| b0 | 2,610 | 1,715 | -,864 | 6,085 |
| b1 | ,612 | ,065 | ,481 | ,743 |
| b2 | -,076 | ,026 | -,130 | -,022 |

The Nonlinear Regression Analysis Equation becomes:

$$\mathbf{Q_{2.33} = 2.610 * AREA^{0.612} * H_{min}{}^{(-0,076)}} \qquad \text{(Equation 4.23)}$$

In Table 5.31, this equation which is derived from using Nonlinear Regression Analysis is used for the calculation of flood values which are distributed by Gumbel Distribution for whole region.

**ii) Nonlinear Regression Analysis for $Q_{2.33}$ values with Gumbel Distributions for Cluster Regions**

The same procedure which is applied for whole region stations is used for the different clusters. However, Stepwise Regression Analysis procedure could not be used for the Cluster Region-1. Therefore, this cluster has no selected basin characters. The Nonlinear Analysis procedure is applied for Cluster-2 and Cluster-3 regions.

- **Nonlinear Regression Analysis for Cluster-1**

This cluster have 3 stations. Stepwise Regression Analysis is not applied for this cluster, because of less number of station’s value. Therefore, there is not any Nonlinear Equation for cluster-1

- **Nonlinear Regression Analysis for Cluster-2**

This cluster has 19 stations. Therefore, Stepwise Regression Analysis is applied for this cluster and the Length of Main River and Basin Relief are selected main variables to obtain $Q_{2.33}$ in part 4.5.2.1.2.2. Therefore, these two parameters are used as nonlinear equation input data.

For the equation which is entered the SPSS program is:

$Q_{2.33} = b0 * \text{Length of Main River}^{b1} * \text{Basin Relief}^{b2}$

After the running of the SPSS Nonlinear Analysis tool the summary of results is shown in Table 4.31.

Table 4.31 Summary of Nonlinear Regression Analysis Estimated Parameters for Cluster Region-2

| Parameter | Estimate | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|
| | | | Lower Bound | Upper Bound |
| b0 | ,0026 | ,009 | -,016 | ,021 |
| b1 | ,991 | ,188 | ,592 | 1,389 |
| b2 | 1,054 | ,419 | ,166 | 1,941 |

The Nonlinear Regression Analysis Equation becomes:

$\mathbf{Q_{2.33} = 0.0026 * Length\ of\ Main\ River^{(0.991)} * Basin\ Relief^{(1.054)}}$ (Equation 4.24)

In Table 5.31, this equation which is derived from using Nonlinear Regression Analysis is used for the calculation of flood values which are distributed by Gumbel Distribution for the cluster region-2.

- **Nonlinear Regression Analysis for Cluster-3**

This cluster has 18 stations. Therefore, Stepwise Regression Analysis is applied for this cluster and Perimeter is selected main variables to obtain $Q_{2.33}$ in part 4.5.2. Therefore, this parameter is used as nonlinear equation input data.

For the equation which is entered the SPSS program is:

$Q_{2.33} = b0 * Perimeter^{b1}$

After the running of the SPSS Nonlinear Analysis tool the summary of results is shown in Table 4.32.

Table 4.32 Summary of Nonlinear Regression Analysis Estimated Parameters for Cluster Region-3

| Parameter | Estimate | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|
| | | | Lower Bound | Upper Bound |
| b0 | 2,896 | 1,955 | -1,248 | 7,040 |
| b1 | ,566 | ,124 | ,304 | ,829 |

The Nonlinear Regression Analysis Equation becomes:

$\mathbf{Q_{2.33} = 2.896 * Perimeter^{(0.566)}}$ (Equation 4.25)

In Table 5.31, this equation which is derived from using Nonlinear Regression Analysis is used for the calculation of flood values which are distributed by Gumbel Distribution for the cluster region-3.

#### 4.6.2.2 Nonlinear Regression Analysis for $Q_{2.33}$ values with Different Distribution Parameters

The basin characteristics data and $Q_{2.33}$ values are shown in Table 4.23. First this analysis is done for whole region, than this is applied for all cluster regions.

**i) Nonlinear Regression Analysis for $Q_{2.33}$ values with Different Distribution Parameters for whole region**

In the Stepwise Regression Part, two main parameters which are Area and Mean Annual Rainfall are selected as the main basin characteristics in part 4.5.2. Therefore, these two parameters are used as nonlinear equation input data.

For the equation which is entered the SPSS program is:

$Q_{2.33}$ = b0 * $Area^{b1}$ * Mean Annual $Rainfall^{b2}$

After the running of the SPSS Nonlinear Analysis tool the summary of results is shown in Table 4.33.

Table 4.33 Summary of Nonlinear Regression Analysis Estimated Parameters for Whole Region

| Parameter | Estimate | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|
| | | | Lower Bound | Upper Bound |
| B0 | 6,011E-006 | ,000 | -2,064E-005 | 3,266E-005 |
| B1 | ,724 | ,031 | ,661 | ,787 |
| B2 | 1,824 | ,343 | 1,129 | 2,519 |

The Nonlinear Regression Analysis Equation becomes:

$\mathbf{Q_{2.33} = 6.011*10^{-6} * Area^{(0.724)} * MAR^{(1.824)}}$ (Equation 4.26)

In Table 5.31, this equation which is derived from using Nonlinear Regression Analysis is used for the calculation of flood values which are distributed by Different Distribution Parameters for whole region.

**ii) Nonlinear Regression Analysis for $Q_{2.33}$ values with Different Distribution Parameters for Cluster Regions**

The same procedure which is applied for whole region stations is used for the different clusters. However, Stepwise Regression Analysis procedure could not use for the Cluster Region-1. Because of, this cluster has no selected basin characters in Stepwise Analysis. The Nonlinear Analysis procedure is applied for Cluster-2 and Cluster-3 regions.

- **Nonlinear Regression Analysis for Cluster-1**

This cluster have 3 stations. However, Stepwise Regression Analysis is not applied for this cluster, because of less station value. Therefore, there is not any Nonlinear Equation for cluster-1

- **Nonlinear Regression Analysis for Cluster-2**

This cluster has 19 stations. Therefore, Stepwise Regression Analysis is applied for this cluster and the Drainage Area and Basin Relief are selected main variables to obtain $Q_{2.33}$ in part 4.5.2. Therefore, these two parameters are used as nonlinear equation input data.

For the equation which is entered the SPSS program is:

$Q_{2.33}$ = b0 * $Area^{b1}$ * $Perimeter^{b2}$ * Basin $Relief^{b3}$ * Relative $Relief^{b4}$

After the running of the SPSS Nonlinear Analysis tool the summary of results is shown in Table 4.34.

Table 4.34 Summary of Nonlinear Regression Analysis Estimated Parameters for Cluster Region-2

| Parameter | Estimate | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|
| | | | Lower Bound | Upper Bound |
| b0 | ,201 | 371,901 | -797,447 | 797,849 |
| b1 | 1,527 | ,420 | ,625 | 2,428 |
| b2 | 2,191 | 825,249 | -1767,793 | 1772,175 |
| b3 | -1,743 | 824,998 | -1771,188 | 1767,702 |
| b4 | 3,385 | 824,881 | -1765,809 | 1772,580 |

The Nonlinear Regression Analysis Equation becomes:

$\mathbf{Q_{2.33} = 0.201 * Area^{(1.527)} * Perimeter^{(2.191)} * Basin\ R.^{(-1.743)} * Relative\ Relief^{(3.385)}}$ (Equation 4.27)

- **Nonlinear Regression Analysis for Cluster-3**

This cluster has 18 stations. Therefore, Stepwise Regression Analysis is applied for this cluster and the Perimeter and Mean Annual Rainfall are selected main variables to obtain $Q_{2.33}$ in Chapter 4.5.2. Therefore, these two parameters are used as nonlinear equation input data.

For the equation which is entered the SPSS program is:

$Q_{2.33} = b0 * Perimeter^{b1} * Mean\ Annual\ Rainfall^{b2}$

After the running of the SPSS Nonlinear Analysis tool the summary of results is shown in Table 4.35.

Table 4.35 Summary of Nonlinear Regression Analysis Estimated Parameters for Cluster Region-3

| Parameter | Estimate | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|
| | | | Lower Bound | Upper Bound |
| b0 | 6,454E-009 | ,000 | -9,037E-008 | 1,033E-007 |
| b1 | ,973 | ,188 | ,573 | 1,373 |
| b2 | 2,810 | ,987 | ,707 | 4,913 |

The Nonlinear Regression Analysis Equation becomes:

$\mathbf{Q_{2.33} = 6.454 * 10^{-9} * Perimeter^{(0.973)} * MAR^{(2.810)}}$ (Equation 4.28)

### 4.7 Statistical Measures in Comparison of the Models

After the calculation of Flood Frequency Analysis with Point Flood Analysis, Regional Flood Frequency Analysis, Stepwise Regression Analysis and Nonlinear Regression Analysis methods, results are compared by using Root Mean Squared Errors, Nash-Sutcliffe Efficiency Index and % difference of calculated results.

1) Root Mean Squared Error:

RMSE has the same units with the data and an unbiased estimator. The smaller the Mean Squared Error, the closer the fit is to the data.

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N_D}(q_{s,j}-\hat{q}_{s,j})^2}{N_D}} \qquad \text{(Equation 4.29)}$$

2) Nash-Sutcliffe efficiency criterion, ( Castiglioni et al., 2009)(Nash and Sutcliffe, 1970);

Takes values between 1 and -∞. 1 means perfect fit, closer values to 1 have better results.(Nash and Sutcliffe, 1970).

$$E = 1 - \frac{\sum_{j=1}^{N_D}(q_{s,j}-\hat{q}_{s,j})^2}{\sum_{j=1}^{N_D}(q_{s,j}-\bar{q}_{s,j})^2} \qquad \text{(Equation 4.30)}$$

3) In addition, percentage error is calculated by:

% Difference $= \frac{(Q_{calculated}-Q_{observed})}{Q_{observed}} * 100$

Test results and comparison of models are given in Chapter 5.

# CHAPTER 5

# DISCUSSION OF RESULTS

Equations for the $Q_{2.33}$ flood value calculation are derived in the previous chapters and all the analysis are completed. In this part, accuracy of the equations are tested with several statistical measures like Root Mean Squared Errors, Nash-Sutcliffe efficiency criterion and % differences. In addition, the Ceyhan basin $Q_{2.33}$ flood equation developed by Topaloglu(2005) is used for the comparison. All of the equations which are used in the calculations are given in Table 5.1.

There are a lot of equations in Table 5.1. Therefore, for the sake of simplicity, one should have analyze these data in steps. First, Stepwise Regression Analysis and Nonlinear Regression Analysis results are compared due to the Root Mean Squared Errors, Nash-Sutcliffe efficiency criterion and % differences. This procedure is calculated for $Q_{2.33}$ flood values with Gumbel Distribution and $Q_{2.33}$ flood values with Different Distributions and shown in Table 5.2 and Table 5.3, respectively.

According to RMSE and Nash-Sutcliffe statistical measures Nonlinear Analysis with three cluster region gives the best result. This result is valid for both Gumbel Distribution and Different Distribution results. In addition, Nonlinear Regression Analysis relationship is more effective than Stepwise Regression Analysis in Ceyhan Basin.

There is big difference between the results which have been calculated for whole region and cluster regions. This shows, clustering is very effective for both Stepwise Regression Analysis and Nonlinear Regression Analysis Methods. However, Stepwise Regression Analysis could not be done for cluster region 1, so this cluster do not have any offered equation. Because of huge drainage areas and gauged stations, cluster region 1 is not so important for this study.

Moreover, Stepwise Regression Analysis gave one negative flood result in Different Distributions with three clusters analysis. Therefore, this could not be selected as a distribution parameter method.

Nonlinear Regression Analysis with three cluster region is selected the as the best method and this is also compared with Dalyrmple Method, Topaloglu (2005), Point Flood Analysis and the Regional Flood Frequency Method which is used in Turkey. These comparisons are made differently for Gumbel Distribution Parameter and Different Distribution Parameters.

Table 5.1 Computed Model Equations with Different Methods for $Q_{2.33}$ flood values.

| Calculation Methodology | Equation Number | Using Gumbel Distribution | Related Chapter | Using Different Distributions |
|---|---|---|---|---|
| **Dalyrmple Method-Envelope Equation** | Equation 4.10 | $Q_{2.33} = 1.7284 * AREA^{0.8076}$ | Equation 4.12 | $Q_{2.33} = 1.4266 * AREA^{0.8123}$ |
| **Dalyrmple Method-Best Fit** | Equation 4.11 | $Q_{2.33} = 0.0492 * AREA + 39.713$ | Equation 4.13 | $Q_{2.33} = 2.509 * AREA^{0.504}$ |
| **Point Flood Analysis** | Equation 4.14 | Table 4.13 | Equation 4.14 | Table 4.14 |
| **Stepwise Analysis-Whole Region** | Equation 4.15 | $Q_{2.33} = 74.966 + 0.049 * AREA - 0.052 * H_{min}$ | Equation 4.18 | $Q_{2.33} = -105.808 + 0.053 * AREA + 0.205 * MAR$ |
| **Stepwise Analysis-Cluster-1** | - | Could not computed | - | Could not computed |
| **Stepwise Analysis-Cluster-2** | Equation 4.16 | $Q_{2.33} = -44.309 + 1.711 * (\text{Length of Main River}) + 0.098 * BR$ | Equation 4.19 | $Q_{2.33} = 45.655 + 0.137 * Area - 0.610 * P + 0.223 * BR – 128.032 * RR$ |
| **Stepwise Analysis-Cluster-3** | Equation 4.17 | $Q_{2.33} = 21.284 + 0.167 * Perimeter$ | Equation 4.20 | $Q_{2.33} = - 109.065 + 0.237 * Perimeter + 0.200 * MAR$ |
| **Nonlinear Analysis-Whole Region** | Equation 4.23 | $Q_{2.33} = 2.610 * AREA^{0.612} * H_{min}^{(-0.076)}$ | Equation 4.26 | $Q_{2.33} = 6.011 * 10^{-6} * Area^{(0.724)} * MAR^{(1.824)}$ |
| **Nonlinear Analysis-Cluster-1** | | Could not be computed | | Could not be computed |
| **Nonlinear Analysis-Cluster-2** | Equation 4.24 | $Q_{2.33} = 0.0026 * \text{Length of Main River}^{(0.991)} * BR^{(1.054)}$ | Equation 4.27 | $Q_{2.33} = 0.201 * Area^{(1.527)} * Perimeter^{(2.191)} * BR^{(-1.743)} * RR^{(3.385)}$ |
| **Nonlinear Analysis-Cluster-3** | Equation 4.25 | $Q_{2.33} = 2.896 * Perimeter^{(0.566)}$ | Equation 4.28 | $Q_{2.33} = 6.454 * 10^{-9} * Perimeter^{(0.973)} * MAR^{(2.810)}$ |
| **RFFA by Topaloglu(2004)** | - | $Q_{2.33} = 0.585 * Area^{0.727}$ | - | There is not any derived equation |

Table 5.2 Comparison of Stepwise and Nonlinear Regression Analysis for $Q_{2.33}$ Flood Values with Gumbel Distribution

| Stream Gauging Station # | | AREA ($km^2$) | $Q_{2.33}$ Values with Gumbel Distribution ($m^3/s$) | Stepwise Analysis for Whole Region ($m^3/s$) | Stepwise Analysis for Three Cluster Regions ($m^3/s$) | Nonlinear Analysis for Whole Region ($m^3/s$) | Nonlinear Analysis For Three Cluster Regions ($m^3/s$) | Stepwise Analysis for Whole Region % Difference | Stepwise Analysis for Three Cluster Regions % Difference | Nonlinear Analysis for Whole Region % Difference | Nonlinear Analysis For Three Cluster Regions % Difference |
|---|---|---|---|---|---|---|---|---|---|---|---|
| # 1 | EIE/2001 | 8484 | 552.95 | 596.71 | | 418.39 | | 7.91% | | -24.34% | |
| | EIE/2004 | 20466 | 1027.82 | 1236.36 | | 1076.75 | | 20.29% | | 4.76% | |
| | EIE/2012 | 19727.2 | 1094.73 | 1199.42 | | 1052.79 | | 9.56% | | -3.83% | |
| # 2 | EIE/2007 | 623 | 59.86 | 193.59 | 100.23 | 115.52 | 90.11 | 223.41% | 67.45% | 92.98% | 50.53% |
| | EIE/2008 | 444 | 160.70 | 192.71 | 149.90 | 80.16 | 198.78 | 19.92% | -6.72% | -50.12% | 23.69% |
| | EIE/2010 | 3498.8 | 287.69 | 320.63 | 253.92 | 244.16 | 240.37 | 11.45% | -11.74% | -15.13% | -16.45% |
| | DSI/20-04 | 178 | 38.11 | 150.99 | 83.51 | 41.38 | 87.00 | 296.20% | 119.12% | 8.58% | 128.28% |
| | DSI/20-05 | 94 | 45.57 | 153.80 | 82.22 | 28.18 | 63.80 | 237.51% | 80.44% | -38.16% | 40.00% |
| | DSI/20-06 | 174.9 | 57.01 | 150.41 | 74.95 | 39.52 | 65.54 | 163.82% | 31.46% | -30.68% | 14.97% |
| | DSI/20-08 | 131.1 | 39.33 | 157.35 | 96.74 | 34.35 | 83.14 | 300.09% | 145.98% | -12.65% | 111.38% |
| | DSI/20-10 | 217.3 | 62.53 | 118.88 | 54.57 | 43.93 | 51.19 | 90.12% | -12.73% | -29.75% | -18.14% |
| | DSI/20-13 | 105.1 | 50.53 | 134.56 | 51.81 | 30.33 | 43.78 | 166.30% | 2.53% | -39.97% | -13.37% |
| | DSI/20-14 | 310.5 | 163.28 | 151.02 | 87.07 | 55.29 | 86.54 | -7.51% | -46.68% | -66.14% | -47.00% |
| | DSI/20-17 | 1740 | 121.32 | 225.22 | 183.39 | 157.12 | 186.15 | 85.64% | 51.17% | 29.51% | 53.43% |
| | DSI/20-40 | 79 | 5.66 | 82.92 | 15.24 | 23.22 | 17.92 | 1365.05% | 169.25% | 310.31% | 216.53% |
| | DSI/20-43 | 163 | 68.16 | 150.61 | 67.36 | 39.22 | 56.50 | 120.97% | -1.18% | -42.45% | -17.10% |
| | DSI/20-44 | 35 | 28.96 | 108.50 | 10.04 | 14.91 | 9.53 | 274.66% | -65.33% | -48.53% | -67.09% |
| | DSI/20-45 | 170 | 67.52 | 144.10 | 43.37 | 39.48 | 40.23 | 113.42% | -35.76% | -41.52% | -40.41% |
| | DSI/20-46 | 477 | 207.68 | 183.21 | 114.94 | 76.95 | 135.21 | -11.78% | -44.66% | -62.95% | -34.89% |
| | DSI/20-56 | 238.4 | 21.53 | 105.52 | 30.33 | 46.48 | 27.82 | 390.11% | 40.89% | 115.87% | 29.24% |
| | DSI/20-63 | 57.5 | 14.61 | 94.92 | 25.98 | 19.34 | 23.88 | 549.70% | 77.84% | 32.39% | 63.43% |
| | DSI/20-65 | 161 | 41.45 | 107.54 | 19.03 | 42.32 | 16.44 | 159.44% | -54.09% | 2.09% | -60.33% |

Table 5.2 Comparison of Stepwise and Nonlinear Regression Analysis for $Q_{2.33}$ Flood Values with Gumbel Distribution (Continued)

| Stream Gauging Station # | | AREA (km$^2$) | $Q_{2.33}$ Values with Gumbel Distribution (m$^3$/s) | Stepwise Analysis for Whole Region (m$^3$/s) | Stepwise Analysis for Three Cluster Regions (m$^3$/s) | Nonlinear Analysis for Whole Region (m$^3$/s) | Nonlinear Analysis For Three Cluster Regions (m$^3$/s) | Stepwise Analysis for Whole Region % Difference | Stepwise Analysis for Three Cluster Regions % Difference | Nonlinear Analysis for Whole Region % Difference | Nonlinear Analysis For Three Cluster Regions % Difference |
|---|---|---|---|---|---|---|---|---|---|---|---|
| # 3 | EIE/2005 | 4219.08 | 111.42 | 321.48 | 107.46 | 254.32 | 99.35 | 188.53% | -3.56% | 128.25% | -10.83% |
| | EIE/2006 | 733.2 | 53.43 | 54.60 | 60.03 | 85.74 | 63.19 | 2.19% | 12.35% | 60.47% | 18.27% |
| | EIE/2009 | 1387.2 | 85.06 | 107.77 | 71.38 | 128.36 | 73.09 | 26.70% | -16.08% | 50.90% | -14.07% |
| | EIE/2015 | 915.2 | 43.95 | 69.74 | 67.54 | 98.82 | 69.86 | 58.68% | 53.68% | 124.84% | 58.96% |
| | EIE/2016 | 846.8 | 83.56 | 81.74 | 61.70 | 94.57 | 64.72 | -2.18% | -26.16% | 13.17% | -22.55% |
| | EIE/2022 | 428 | 68.93 | 43.15 | 46.17 | 61.60 | 49.18 | -37.41% | -33.02% | -10.64% | -28.65% |
| | DSI/20-02 | 197.1 | 40.83 | 46.20 | 37.82 | 39.53 | 39.02 | 13.15% | -7.38% | -3.20% | -4.42% |
| | DSI/20-07 | 2084 | 54.54 | 140.86 | 78.06 | 164.74 | 78.45 | 158.26% | 43.13% | 202.06% | 43.85% |
| | DSI/20-15 | 189.7 | 26.15 | 14.66 | 36.82 | 37.40 | 37.67 | -43.95% | 40.78% | 43.03% | 44.04% |
| | DSI/20-16 | 291 | 35.84 | 27.64 | 41.99 | 48.56 | 44.33 | -22.88% | 17.17% | 35.50% | 23.68% |
| | DSI/20-36 | 174.2 | 62.79 | 7.46 | 38.32 | 35.75 | 39.69 | -88.12% | -38.97% | -43.06% | -36.79% |
| | DSI/20-51 | 131.4 | 10.14 | 51.03 | 37.32 | 30.35 | 38.35 | 403.23% | 268.01% | 199.27% | 278.20% |
| | DSI/20-52 | 23 | 6.12 | -1.07 | 25.79 | 10.33 | 18.70 | -117.55% | 321.45% | 68.75% | 205.63% |
| | DSI/20-53 | 178.5 | 38.27 | 26.12 | 38.49 | 36.70 | 39.91 | -31.76% | 0.56% | -4.09% | 4.28% |
| | DSI/20-54 | 207.5 | 58.87 | 60.96 | 35.98 | 41.29 | 36.51 | 3.54% | -38.88% | -29.87% | -37.99% |
| | DSI/20-55 | 111.6 | 21.06 | 73.48 | 33.64 | 28.05 | 33.10 | 248.89% | 59.74% | 33.20% | 57.15% |
| | DSI/20-58 | 24.38 | 18.38 | 17.78 | 26.13 | 10.93 | 19.48 | -3.26% | 42.15% | -40.55% | 5.97% |
| | DSI/20-59 | 171.5 | 62.49 | 23.77 | 37.48 | 35.93 | 38.58 | -61.97% | -40.02% | -42.51% | -38.27% |
| | | | | | | | | | | | |
| | | | | | | Root Mean Square Errors | | 79.20 | 31.01 | 52.25 | 29.54 |
| | | | | | | Nash-Sutcliffen Efficiency | | 0.013 | 0.708 | 0.286 | 0.735 |

Table 5.3 Comparison of Stepwise and Nonlinear Regression Analysis for $Q_{2.33}$ Flood Values with Different Distributions

| Stream Gauging Station # | | AREA ($km^2$) | $Q_{2.33}$ Values with Different Distribution Parameters ($m^3/s$) | Stepwise Analysis for Whole Region ($m^3/s$) | Stepwise Analysis for Three Cluster Regions ($m^3/s$) | Nonlinear Analysis for Whole Region ($m^3/s$) | Nonlinear Analysis For Three Cluster Regions ($m^3/s$) | Stepwise Analysis for Whole Region % Difference | Stepwise Analysis for Three Cluster Regions % Difference | Nonlinear Analysis for Whole Region % Difference | Nonlinear Analysis For Three Cluster Regions % Difference |
|---|---|---|---|---|---|---|---|---|---|---|---|
| # 1 | **EIE/2001** | 8484 | 536.54 | 454.69 | | 405.69 | | -15.26% | | -24.39% | |
| | **EIE/2004** | 20466 | 1089.20 | 1115.42 | | 1122.34 | | 2.41% | | 3.04% | |
| | **EIE/2012** | 19727.2 | 1089.68 | 1075.45 | | 1081.11 | | -1.31% | | -0.79% | |
| # 2 | **EIE/2007** | 623 | 58.86 | 97.47 | 79.99 | 133.99 | 59.53 | 65.61% | 35.91% | 127.66% | 1.15% |
| | **EIE/2008** | 444 | 146.34 | 89.05 | 116.87 | 106.05 | 99.39 | -39.15% | -20.14% | -27.53% | -32.08% |
| | **EIE/2010** | 3498.8 | 285.82 | 211.69 | 285.86 | 294.02 | 277.43 | -25.94% | 0.02% | 2.87% | -2.94% |
| | **DSI/20-04** | 178 | 28.54 | 73.89 | 69.83 | 54.10 | 48.93 | 158.89% | 144.66% | 89.54% | 71.44% |
| | **DSI/20-05** | 94 | 36.82 | 57.51 | 34.21 | 29.85 | 51.19 | 56.20% | -7.08% | -18.93% | 39.04% |
| | **DSI/20-06** | 174.9 | 51.84 | 67.51 | 74.00 | 49.91 | 74.44 | 30.23% | 42.73% | -3.72% | 43.58% |
| | **DSI/20-08** | 131.1 | 28.72 | 59.85 | 68.13 | 38.14 | 80.46 | 108.40% | 137.23% | 32.79% | 180.14% |
| | **DSI/20-10** | 217.3 | 62.53 | 67.80 | 66.80 | 57.14 | 42.86 | 8.44% | 6.83% | -8.61% | -31.45% |
| | **DSI/20-13** | 105.1 | 46.61 | 76.67 | 48.68 | 39.61 | 24.50 | 64.48% | 4.44% | -15.02% | -47.43% |
| | **DSI/20-14** | 310.5 | 149.17 | 59.33 | 114.87 | 63.21 | 105.52 | -60.22% | -22.99% | -57.63% | -29.26% |
| | **DSI/20-17** | 1740 | 86.40 | 108.36 | 92.77 | 153.35 | 129.24 | 25.42% | 7.37% | 77.48% | 49.58% |
| | **DSI/20-40** | 79 | 4.68 | 45.23 | 25.05 | 22.94 | 7.08 | 865.75% | 434.98% | 389.80% | 51.08% |
| | **DSI/20-43** | 163 | 50.79 | 61.74 | 75.57 | 44.75 | 55.30 | 21.55% | 48.79% | -11.89% | 8.88% |
| | **DSI/20-44** | 35 | 19.68 | 46.90 | -44.07 | 13.36 | 6.57 | 138.36% | -323.95% | -32.09% | -66.59% |
| | **DSI/20-45** | 170 | 55.11 | 57.08 | 49.99 | 43.51 | 21.63 | 3.57% | -9.30% | -21.05% | -60.75% |
| | **DSI/20-46** | 477 | 198.10 | 74.94 | 145.79 | 93.56 | 191.23 | -62.17% | -26.41% | -52.77% | -3.47% |
| | **DSI/20-56** | 238.4 | 22.30 | 39.18 | 32.56 | 42.22 | 11.74 | 75.73% | 46.03% | 89.35% | -47.36% |
| | **DSI/20-63** | 57.5 | 12.22 | 39.57 | 21.37 | 17.21 | 7.14 | 223.68% | 74.85% | 40.83% | -41.56% |
| | **DSI/20-65** | 161 | 39.47 | 53.82 | 28.18 | 40.46 | 4.83 | 36.33% | -28.62% | 2.49% | -87.76% |

Table 5.3 Comparison of Stepwise and Nonlinear Regression Analysis for $Q_{2.33}$ Flood Values with Different Distribution (Continued)

| Stream Gauging Station # | | AREA ($km^2$) | $Q_{2.33}$ Values with Different Distribution Parameters ($m^3/s$) | Stepwise Analysis for Whole Region ($m^3/s$) | Stepwise Analysis for Three Cluster Regions ($m^3/s$) | Nonlinear Analysis for Whole Region ($m^3/s$) | Nonlinear Analysis For Three Cluster Regions ($m^3/s$) | Stepwise Analysis for Whole Region % Difference | Stepwise Analysis for Three Cluster Regions % Difference | Nonlinear Analysis for Whole Region % Difference | Nonlinear Analysis For Three Cluster Regions % Difference |
|---|---|---|---|---|---|---|---|---|---|---|---|
| # 3 | **EIE/2005** | 4219.08 | 108.41 | 214.78 | 107.84 | 191.71 | 92.42 | 98.11% | -0.53% | 76.83% | -14.76% |
| | **EIE/2006** | 733.2 | 53.55 | 47.42 | 57.50 | 72.96 | 67.49 | -11.45% | 7.38% | 36.25% | 26.04% |
| | **EIE/2009** | 1387.2 | 79.56 | 82.56 | 74.08 | 116.64 | 87.69 | 3.76% | -6.90% | 46.60% | 10.21% |
| | **EIE/2015** | 915.2 | 38.34 | 39.42 | 50.94 | 63.10 | 50.08 | 2.80% | 32.87% | 64.57% | 30.60% |
| | **EIE/2016** | 846.8 | 73.65 | 37.86 | 44.66 | 61.99 | 46.59 | -48.60% | -39.36% | -15.83% | -36.74% |
| | **EIE/2022** | 428 | 53.93 | 22.91 | 29.69 | 43.04 | 35.46 | -57.52% | -44.94% | -20.19% | -34.24% |
| | **DSI/20-02** | 197.1 | 36.54 | 18.54 | 25.52 | 27.97 | 29.13 | -49.26% | -30.15% | -23.44% | -20.27% |
| | **DSI/20-07** | 2084 | 40.86 | 101.83 | 66.33 | 115.50 | 61.96 | 149.21% | 62.33% | 182.67% | 51.63% |
| | **DSI/20-15** | 189.7 | 16.90 | 18.79 | 24.73 | 27.49 | 27.85 | 11.20% | 46.33% | 62.68% | 64.82% |
| | **DSI/20-16** | 291 | 23.80 | 19.86 | 27.88 | 34.95 | 33.09 | -16.58% | 17.10% | 46.81% | 39.01% |
| | **DSI/20-36** | 174.2 | 62.79 | 31.77 | 40.33 | 31.81 | 41.95 | -49.40% | -35.77% | -49.34% | -33.19% |
| | **DSI/20-51** | 131.4 | 8.71 | 13.13 | 22.93 | 20.22 | 26.95 | 50.70% | 163.18% | 132.07% | 209.34% |
| | **DSI/20-52** | 23 | 4.97 | 19.54 | 18.44 | 6.91 | 10.48 | 293.16% | 270.93% | 39.00% | 110.82% |
| | **DSI/20-53** | 178.5 | 36.37 | 26.40 | 35.10 | 29.84 | 37.36 | -27.41% | -3.49% | -17.94% | 2.73% |
| | **DSI/20-54** | 207.5 | 53.02 | 38.76 | 42.11 | 38.83 | 40.65 | -26.89% | -20.58% | -26.77% | -23.33% |
| | **DSI/20-55** | 111.6 | 21.06 | 23.41 | 28.77 | 21.42 | 27.43 | 11.17% | 36.62% | 1.70% | 30.25% |
| | **DSI/20-58** | 24.38 | 18.38 | 33.58 | 32.54 | 8.75 | 15.16 | 82.72% | 77.04% | -52.37% | -17.53% |
| | **DSI/20-59** | 171.5 | 52.43 | 46.21 | 53.37 | 38.27 | 54.06 | -11.85% | 1.80% | -27.00% | 3.11% |
| | | | | | | **Root Mean Square Errors** | | **41.65** | **22.20** | **41.03** | **21.33** |
| | | | | | | **Nash-Sutcliffen Efficiency** | | **0.462** | **0.841** | **0.575** | **0.854** |

### 5.7.1 Comparison of results which are derived by $Q_{2.33}$ flood values with Gumbel Distribution

$Q_{2.33}$ flood values which are calculated by using Gumbel Distribution, are compared with values obtained from the equations which are given in Table 5.1. However, Nonlinear Regression Analysis with three cluster region is selected as the final model for Ceyhan Basin. Therefore, other Stepwise Regression Analysis for whole basin, Stepwise Regression Analysis for three clusters and Nonlinear Analysis for whole basin are not used in this comparison.

In Table 5.4, the results of $Q_{2.33}$ flood values are given due to related calculation equations. Moreover, % difference from the real calculated values, Root Mean Square Errors and Nash-Sutcliffe Efficiency values are also given in this table.

There are some missing results in the Table 5.4. Nonlinear Regression Analysis for cluster-1 could not be calculated due to the explained reasons before. In Dalyrmple Method, there are only two station results which are EIE/2012 and EIE/2007 could not calculated because of these two stations are located in non-homogeneous region in the analysis. In Point Flood Analysis, there are only 18 station results, other 22 stations could not be calculated because of there is no relationship between any of them. These missing results are considered while calculating the Root Mean Square Errors. % differences and Nash-Sutcliffe Efficiency values.

Considering the Root Mean Square Errors result, there is big difference between the Nonlinear Regression Analysis and the other methods. This shows that grouping and making analysis for basin characteristics gives more dependable results with less errors. And also, Nonlinear Regression Analysis % differences are the dependable lowest values. There are only some stations that have unexpected result. For example DSI/20-40 station has %216.53 difference.

Comparing the % differences of results shows that Topaloglu (2005) equation, gives more low values than the real values. This could be explained with the number of stations which are used in Topaloglu(2005) have less record periods than the record period used in this study. In addition, Topaloglu(2005) used only 15 stations to derive the related equation.

In Point Flood Analysis, % differences are not so different, but there are some significant differences in some stations. For example, results of EIE/2007 and EIE/2005 have so significant change. In addition, this method has a significant problem that the method could not applied to all ungauged basins easily.

In Dalyrmple Method which is made by envelope line equation, % differences are more higher than the real calculated values. This is because of envelope line is drawn to include all points including in the graph. Therefore, the results of this equation gives higher results than the real results.

In Dalyrmple Method which is made by best fit equation, % differences are the most closest to the Nonlinear Regression Analysis with three clusters.

The relationships between the obtained results and the real station values are shown in Figure 5.1, Figure 5.2, Figure 5.3, Figure 5.4 and Figure 5.5

Moreover, results obtained from Gumbel Distribution equations are compared with the flood peak values presented in Dagdelen HEPP report, which has approved by DSI. Regional Flood Frequency Analysis $Q_{100}$ result is given as 1030.50 m3/s in the report. Gumbel Distribution equations $Q_{100}$ results for this basin are obtained as follows:

- Dalyrmple Method-Envelope Equation : 6586.8 m3/s
- Dalyrmple Method-Best Fit Equation : 1130.40 m3/s
- Nonlinear for Whole Region Equation : 1075.24 m3/s
- Nonlinear for Cluster-3 Equation : 396.78 m3/s

At first view, one can see that Nonlinear with three cluster equation does not provide good result. However, when we consider station EIE/2005 which is near Dagdelen HEPP and having the same drainage area, the best accurate method is Nonlinear Analysis with three clusters for this station. Therefore, one can say that Regional Flood Frequency Analysis results were overestimated in Dagdelen HEPP report.

Table 5.4 Comparison of Results which are derived for $Q_{2.33}$ flood values for Gumbel Distribution.

| Stream Gauging Station # | | AREA ($km^2$) | $Q_{2.33}$ Values with Gumbel Distribution ($m^3/s$) | Nonlinear Analysis For Three Cluster Regions ($m^3/s$) | Dalyrmple Method - Best Fit Equation ($m^3/s$) | Dalyrmple Method - Envelope Line Equation ($m^3/s$) | Point Flood Analysis ($m^3/s$) | Equation of Ceyhan Basin in Topaloglu (2004) ($m^3/s$) | Nonlinear Analysis For Three Cluster Regions % Difference | Dalyrmple Method - Best Fit Equation % Difference | Dalyrmple Method - Envelope Line Equation % Difference | Point Flood Analysis % Difference | Equation of Ceyhan Basin in Topaloglu (2004) % Difference |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # 1 | EIE/2001 | 8484 | 552.95 | | 457.13 | 2572.66 | 177.51 | 420.00 | | -17.33% | 365.26% | -67.90% | -24.04% |
| | EIE/2004 | 20466 | 1027.82 | | 1046.64 | 5238.83 | 1121.89 | 796.67 | | 1.83% | 409.70% | 9.15% | -22.49% |
| | EIE/2012 | 19727.2 | 1094.73 | | | | 1002.93 | 775.65 | | | | -8.39% | -29.15% |
| # 2 | EIE/2007 | 623 | 59.86 | 90.11 | | | 166.62 | 62.91 | 50.53% | | | 178.35% | 5.10% |
| | EIE/2008 | 444 | 160.70 | 198.78 | 61.56 | 237.50 | | 49.18 | 23.69% | -61.69% | 47.79% | | -69.40% |
| | EIE/2010 | 3498.8 | 287.69 | 240.37 | 211.85 | 1258.09 | | 220.59 | -16.45% | -26.36% | 337.31% | | -23.32% |
| | DSI/20-04 | 178 | 38.11 | 87.00 | 48.47 | 113.52 | | 25.31 | 128.28% | 27.19% | 197.88% | | -33.60% |
| | DSI/20-05 | 94 | 45.57 | 63.80 | 44.34 | 67.79 | | 15.91 | 40.00% | -2.70% | 48.75% | | -65.09% |
| | DSI/20-06 | 174.9 | 57.01 | 65.54 | 48.32 | 111.92 | | 24.98 | 14.97% | -15.25% | 96.32% | | -56.18% |
| | DSI/20-08 | 131.1 | 39.33 | 83.14 | 46.16 | 88.68 | | 20.26 | 111.38% | 17.37% | 125.47% | | -48.49% |
| | DSI/20-10 | 217.3 | 62.53 | 51.19 | 50.40 | 133.37 | | 29.25 | -18.14% | -19.39% | 113.29% | | -53.21% |
| | DSI/20-13 | 105.1 | 50.53 | 43.78 | 44.88 | 74.18 | | 17.25 | -13.37% | -11.17% | 46.81% | | -65.86% |
| | DSI/20-14 | 310.5 | 163.28 | 86.54 | 54.99 | 177.92 | 155.99 | 37.92 | -47.00% | -66.32% | 8.97% | -4.47% | -76.78% |
| | DSI/20-17 | 1740 | 121.32 | 186.15 | 125.32 | 715.67 | | 132.75 | 53.43% | 3.30% | 489.90% | | 9.42% |
| | DSI/20-40 | 79 | 5.66 | 17.92 | 43.60 | 58.91 | | 14.02 | 216.53% | 670.31% | 940.76% | | 147.69% |
| | DSI/20-43 | 163 | 68.16 | 56.50 | 47.73 | 105.73 | 24.49 | 23.74 | -17.10% | -29.97% | 55.12% | -64.07% | -65.18% |
| | DSI/20-44 | 35 | 28.96 | 9.53 | 41.44 | 30.52 | | 7.76 | -67.09% | 43.08% | 5.40% | | -73.21% |
| | DSI/20-45 | 170 | 67.52 | 40.23 | 48.08 | 109.38 | | 24.47 | -40.41% | -28.80% | 62.00% | | -63.75% |
| | DSI/20-46 | 477 | 207.68 | 135.21 | 63.18 | 251.66 | 217.39 | 51.81 | -34.89% | -69.58% | 21.18% | 4.67% | -75.05% |
| | DSI/20-56 | 238.4 | 21.53 | 27.82 | 51.44 | 143.73 | | 31.29 | 29.24% | 138.93% | 567.59% | | 45.35% |
| | DSI/20-63 | 57.5 | 14.61 | 23.88 | 42.54 | 45.58 | | 11.13 | 63.43% | 191.18% | 211.96% | | -23.83% |
| | DSI/20-65 | 161 | 41.45 | 16.44 | 47.63 | 104.68 | | 23.52 | -60.33% | 14.92% | 152.55% | | -43.25% |

Table 5.4 Comparison of Results which are derived for $Q_{2.33}$ flood values for Gumbel Distribution. (Continued)

| Stream Gauging Station # | | AREA ($km^2$) | $Q_{2.33}$ Values with Gumbel Distribution ($m^3/s$) | Nonlinear Analysis For Three Cluster Regions ($m^3/s$) | Dalyrmple Method - Best Fit Equation ($m^3/s$) | Dalyrmple Method - Envelope Line Equation ($m^3/s$) | Point Flood Analysis ($m^3/s$) | Equation of Ceyhan Basin in Topaloglu (2004) ($m^3/s$) | Nonlinear Analysis For Three Cluster Regions % Difference | Dalyrmple Method - Best Fit Equation % Difference | Dalyrmple Method - Envelope Line Equation % Difference | Point Flood Analysis % Difference | Equation of Ceyhan Basin in Topaloglu (2004) % Difference |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # 3 | **EIE/2005** | 4219.08 | 111.42 | 99.35 | 247.29 | 1463.42 | 347.08 | 252.75 | -10.83% | 121.95% | 1213.43% | 211.51% | 126.84% |
| | **EIE/2006** | 733.2 | 53.43 | 63.19 | 75.79 | 356.12 | 55.61 | 70.82 | 18.27% | 41.84% | 566.52% | 4.07% | 32.55% |
| | **EIE/2009** | 1387.2 | 85.06 | 73.09 | 107.96 | 595.99 | 81.73 | 112.59 | -14.07% | 26.93% | 600.66% | -3.91% | 32.36% |
| | **EIE/2015** | 915.2 | 43.95 | 69.86 | 84.74 | 425.96 | 31.51 | 83.21 | 58.96% | 92.81% | 869.18% | -28.30% | 89.33% |
| | **EIE/2016** | 846.8 | 83.56 | 64.72 | 81.38 | 400.06 | 108.63 | 78.64 | -22.55% | -2.61% | 378.76% | 30.01% | -5.89% |
| | **EIE/2022** | 428 | 68.93 | 49.18 | 60.77 | 230.57 | 53.02 | 47.89 | -28.65% | -11.84% | 234.50% | -23.08% | -30.53% |
| | **DSI/20-02** | 197.1 | 40.83 | 39.02 | 49.41 | 123.26 | | 27.25 | -4.42% | 21.01% | 201.89% | | -33.26% |
| | **DSI/20-07** | 2084 | 54.54 | 78.45 | 142.25 | 827.92 | 76.07 | 151.35 | 43.85% | 160.81% | 1418.00% | 39.48% | 177.51% |
| | **DSI/20-15** | 189.7 | 26.15 | 37.67 | 49.05 | 119.51 | | 26.50 | 44.04% | 87.56% | 357.02% | | 1.35% |
| | **DSI/20-16** | 291 | 35.84 | 44.33 | 54.03 | 168.84 | | 36.17 | 23.68% | 50.75% | 371.10% | | 0.93% |
| | **DSI/20-36** | 174.2 | 62.79 | 39.69 | 48.28 | 111.56 | | 24.91 | -36.79% | -23.10% | 77.67% | | -60.33% |
| | **DSI/20-51** | 131.4 | 10.14 | 38.35 | 46.18 | 88.84 | | 20.29 | 278.20% | 355.40% | 776.16% | | 100.14% |
| | **DSI/20-52** | 23 | 6.12 | 18.70 | 40.84 | 21.75 | | 5.72 | 205.63% | 567.40% | 255.32% | | -6.59% |
| | **DSI/20-53** | 178.5 | 38.27 | 39.91 | 48.50 | 113.78 | 53.25 | 25.36 | 4.28% | 26.72% | 197.31% | 39.14% | -33.74% |
| | **DSI/20-54** | 207.5 | 58.87 | 36.51 | 49.92 | 128.49 | 42.31 | 28.29 | -37.99% | -15.20% | 118.26% | -28.13% | -51.95% |
| | **DSI/20-55** | 111.6 | 21.06 | 33.10 | 45.20 | 77.86 | | 18.02 | 57.15% | 114.64% | 269.73% | | -14.42% |
| | **DSI/20-58** | 24.38 | 18.38 | 19.48 | 40.91 | 22.79 | 17.02 | 5.96 | 5.97% | 122.59% | 24.01% | -7.39% | -67.55% |
| | **DSI/20-59** | 171.5 | 62.49 | 38.58 | 48.15 | 110.16 | 67.48 | 24.63 | -38.27% | -22.95% | 76.29% | 7.98% | -60.59% |
| | | | | | | | | | | | | | |
| | | | | | | | **Root Mean Square Errors** | | **29.54** | **50.34** | **831.41** | **112.86** | **83.11** |
| | | | | | | | **Nash-Sutcliffen Efficiency** | | **0.735** | **0.288** | **-2.124** | **-0.361** | **-0.482** |

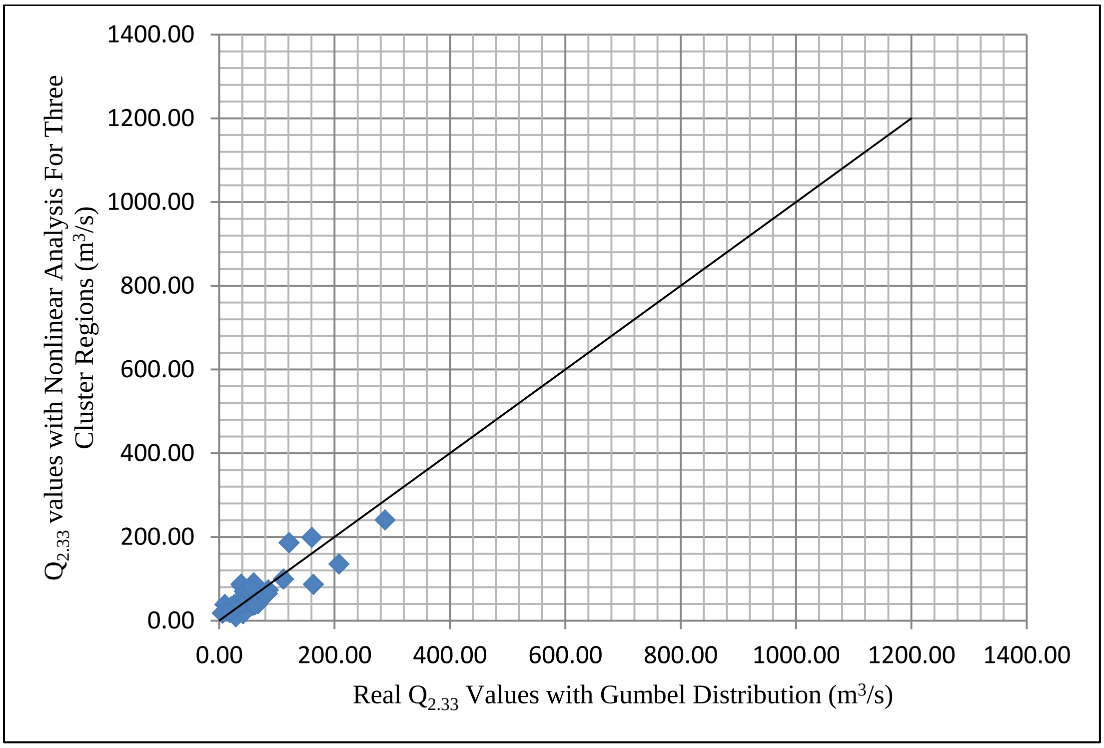


Figure 5.1 Comparison between the Real and Nonlinear Analysis for Three Cluster Region $Q_{2.33}$ values

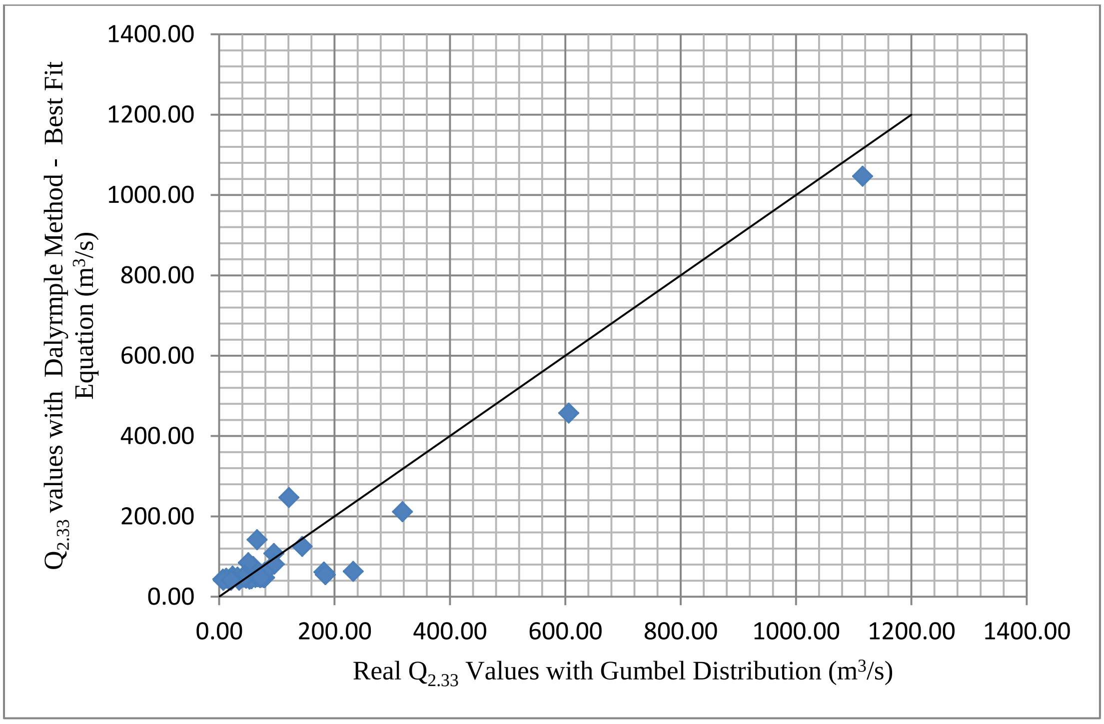


Figure 5.2 Comparison between the Real and Dalyrmple Method for Best Fit Equation $Q_{2.33}$ values

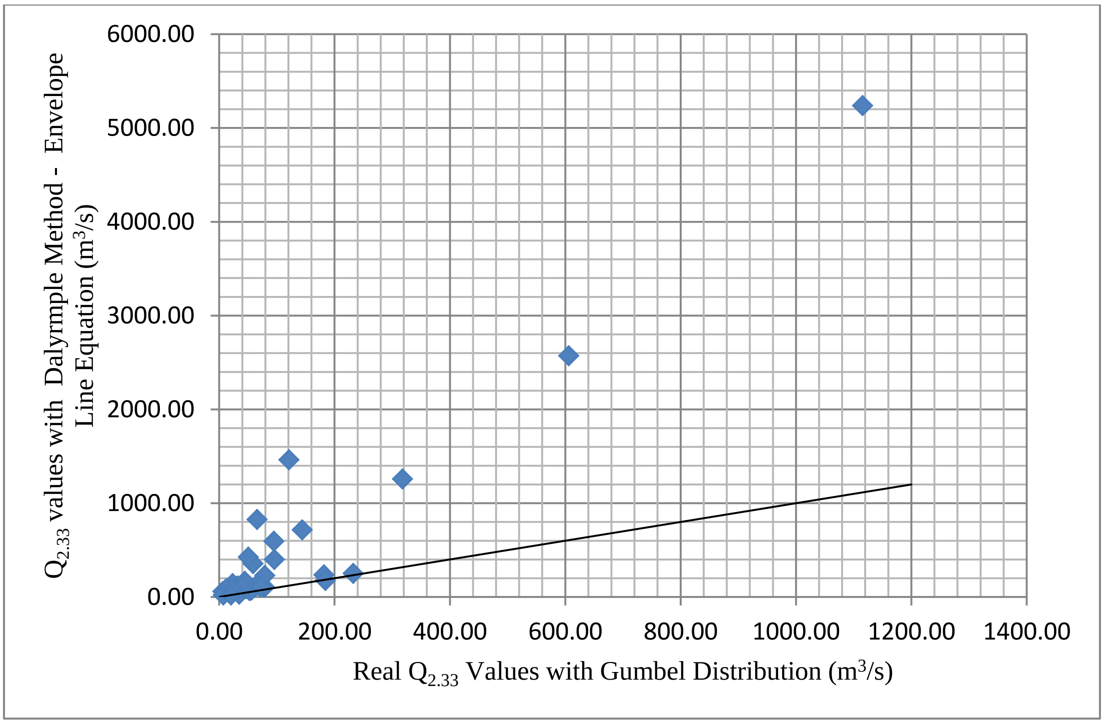


Figure 5.3 Comparison between the Real and Dalyrmple Method for Envelope Line Equation $Q_{2.33}$ values

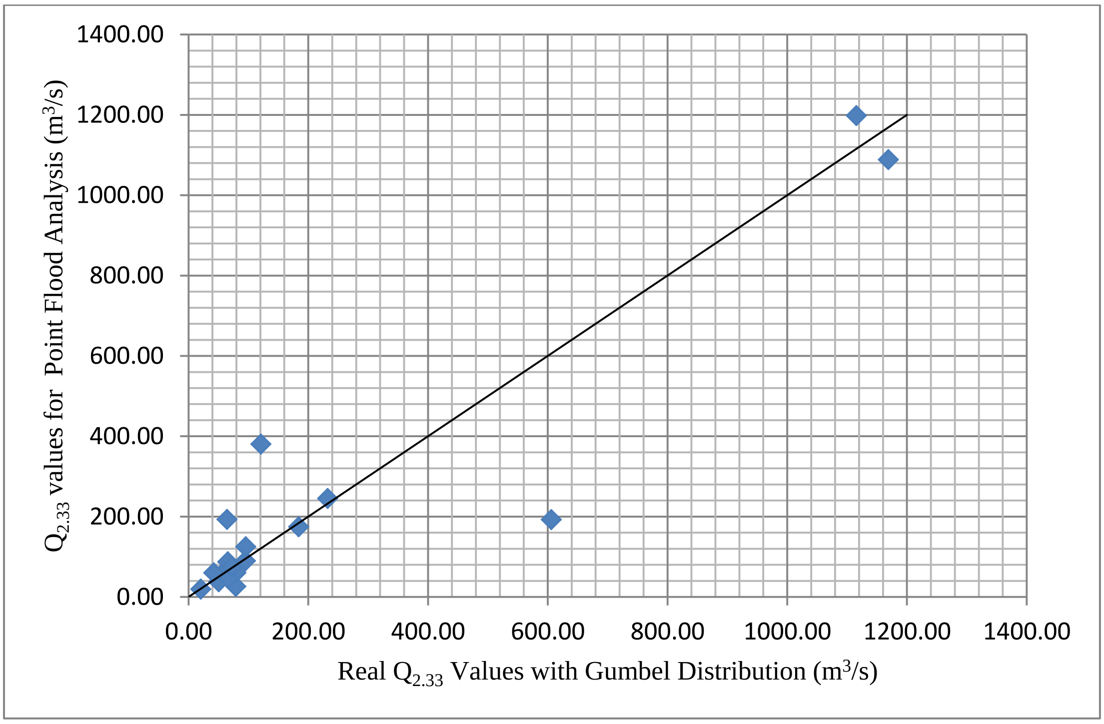


Figure 5.4 Comparison between the Real and Point Flood Analysis $Q_{2.33}$ values

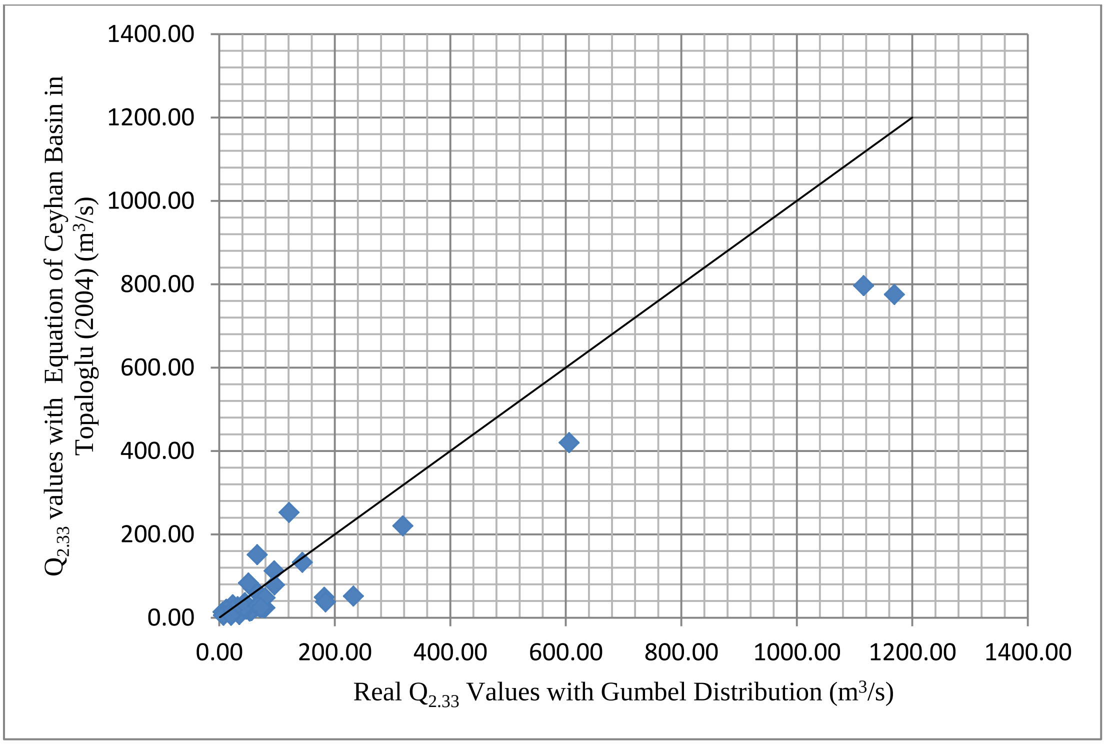


Figure 5.5 Comparison between the Real and Equation of Topaloglu(2005) for Ceyhan Basin $Q_{2.33}$ values

As the calculation results show Nonlinear Regression Analysis with three clusters is the best method to calculate $Q_{2.33}$ flood values. $Q_{100}$, $Q_{50}$ ,. flood values can be calculated by using $Q_{2.33}$ flood values. $Q_{2.33}$ flood values are multiplied by related number to obtain any Flood Frequency for cluster region. Therefore, Table 5.5 is used to obtain multiplication factors for any Flood Frequency in cluster region. However, there is not any multiplication factors for cluster region-1 because Nonlinear Regression Analysis could not be applied for this cluster.

In addition, averages of multiplication is shown with reduced variate values in Figure 5.6

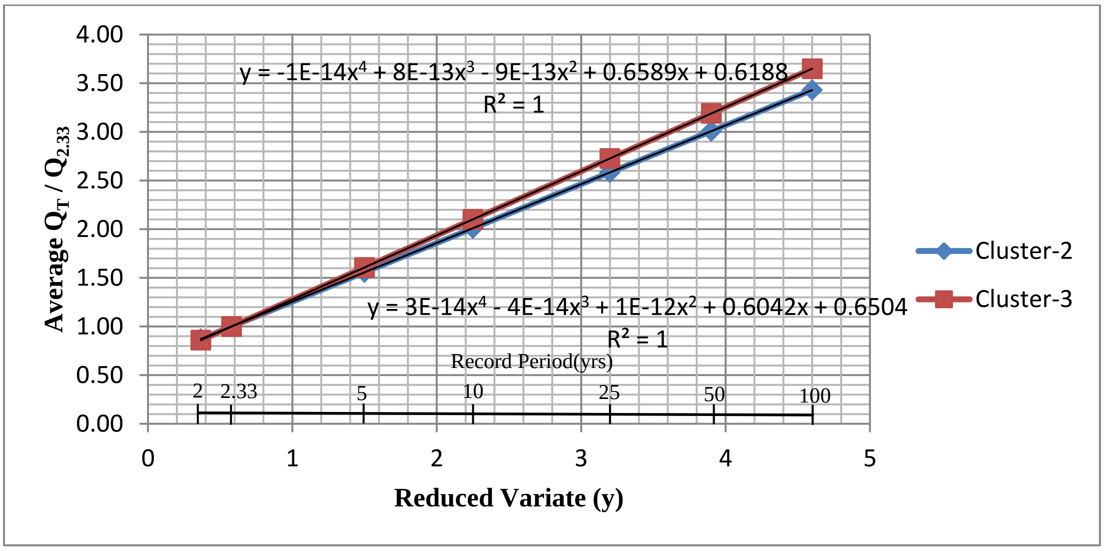


Figure 5.6 RFFA Curve for Different Clusters with Gumbel Distribution

Table 5.5 Average Ratios of Multiplication Factors for Different Clusters with Gumbel Distribution

| Cluster Region | Station | Drainage Area ($km^2$) | $Q_2$ | $Q_{2.33}$ | $Q_5$ | $Q_{10}$ | $Q_{25}$ | $Q_{50}$ | $Q_{100}$ |
|---|---|---|---|---|---|---|---|---|---|
| # 2 | **EIE/2007** | 623.00 | 0.93 | 1.00 | 1.32 | 1.57 | 1.90 | 2.14 | 2.38 |
| | **EIE/2008** | 444.00 | 0.88 | 1.00 | 1.54 | 1.98 | 2.54 | 2.95 | 3.36 |
| | **EIE/2010** | 3498.80 | 0.90 | 1.00 | 1.43 | 1.79 | 2.23 | 2.57 | 2.89 |
| | **DSI/20-04** | 178.00 | 0.82 | 1.00 | 1.79 | 2.44 | 3.25 | 3.86 | 4.46 |
| | **DSI/20-05** | 94.00 | 0.85 | 1.00 | 1.65 | 2.18 | 2.84 | 3.34 | 3.83 |
| | **DSI/20-06** | 174.90 | 0.90 | 1.00 | 1.41 | 1.75 | 2.18 | 2.49 | 2.81 |
| | **DSI/20-08** | 131.10 | 0.82 | 1.00 | 1.76 | 2.38 | 3.17 | 3.75 | 4.33 |
| | **DSI/20-10** | 217.30 | 0.92 | 1.00 | 1.33 | 1.60 | 1.95 | 2.20 | 2.45 |
| | **DSI/20-13** | 105.10 | 0.90 | 1.00 | 1.46 | 1.83 | 2.30 | 2.64 | 2.99 |
| | **DSI/20-14** | 310.50 | 0.88 | 1.00 | 1.53 | 1.96 | 2.50 | 2.90 | 3.30 |
| | **DSI/20-17** | 1740.00 | 0.82 | 1.00 | 1.77 | 2.39 | 3.19 | 3.77 | 4.36 |
| | **DSI/20-40** | 79.00 | 0.86 | 1.00 | 1.60 | 2.08 | 2.70 | 3.15 | 3.60 |
| | **DSI/20-43** | 163.00 | 0.85 | 1.00 | 1.65 | 2.18 | 2.85 | 3.35 | 3.84 |
| | **DSI/20-44** | 35.00 | 0.81 | 1.00 | 1.81 | 2.47 | 3.31 | 3.93 | 4.54 |
| | **DSI/20-45** | 170.00 | 0.87 | 1.00 | 1.56 | 2.01 | 2.58 | 3.00 | 3.43 |
| | **DSI/20-46** | 477.00 | 0.89 | 1.00 | 1.49 | 1.89 | 2.40 | 2.77 | 3.14 |
| | **DSI/20-56** | 238.40 | 0.92 | 1.00 | 1.35 | 1.63 | 1.99 | 2.26 | 2.52 |
| | **DSI/20-63** | 57.50 | 0.84 | 1.00 | 1.70 | 2.27 | 3.00 | 3.53 | 4.07 |
| | **DSI/20-65** | 161.00 | 0.90 | 1.00 | 1.43 | 1.78 | 2.22 | 2.54 | 2.87 |
| | | **AVERAGE** | **0.87** | **1.00** | **1.56** | **2.01** | **2.58** | **3.01** | **3.43** |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| # 3 | **EIE/2005** | 4219.08 | 0.92 | 1.00 | 1.35 | 1.64 | 2.00 | 2.27 | 2.53 |
| | **EIE/2006** | 733.20 | 0.90 | 1.00 | 1.41 | 1.75 | 2.18 | 2.49 | 2.81 |
| | **EIE/2009** | 1387.20 | 0.89 | 1.00 | 1.48 | 1.88 | 2.37 | 2.74 | 3.11 |
| | **EIE/2015** | 915.20 | 0.86 | 1.00 | 1.61 | 2.11 | 2.74 | 3.21 | 3.67 |
| | **EIE/2016** | 846.80 | 0.86 | 1.00 | 1.60 | 2.08 | 2.69 | 3.15 | 3.60 |
| | **EIE/2022** | 428.00 | 0.86 | 1.00 | 1.61 | 2.12 | 2.75 | 3.22 | 3.68 |
| | **DSI/20-02** | 197.10 | 0.87 | 1.00 | 1.56 | 2.02 | 2.59 | 3.02 | 3.45 |
| | **DSI/20-07** | 2084.00 | 0.81 | 1.00 | 1.85 | 2.53 | 3.40 | 4.05 | 4.69 |
| | **DSI/20-15** | 189.70 | 0.79 | 1.00 | 1.90 | 2.64 | 3.56 | 4.25 | 4.94 |
| | **DSI/20-16** | 291.00 | 0.78 | 1.00 | 1.97 | 2.76 | 3.76 | 4.50 | 5.24 |
| | **DSI/20-36** | 174.20 | 0.88 | 1.00 | 1.51 | 1.92 | 2.44 | 2.83 | 3.22 |
| | **DSI/20-51** | 131.40 | 0.80 | 1.00 | 1.89 | 2.61 | 3.53 | 4.21 | 4.88 |
| | **DSI/20-52** | 23.00 | 0.83 | 1.00 | 1.75 | 2.36 | 3.14 | 3.71 | 4.28 |
| | **DSI/20-53** | 178.50 | 0.91 | 1.00 | 1.41 | 1.75 | 2.17 | 2.49 | 2.80 |
| | **DSI/20-54** | 207.50 | 0.88 | 1.00 | 1.53 | 1.96 | 2.51 | 2.92 | 3.32 |
| | **DSI/20-55** | 111.60 | 0.92 | 1.00 | 1.35 | 1.63 | 1.99 | 2.26 | 2.52 |
| | **DSI/20-58** | 24.38 | 0.89 | 1.00 | 1.49 | 1.89 | 2.39 | 2.77 | 3.14 |
| | **DSI/20-59** | 171.50 | 0.85 | 1.00 | 1.65 | 2.17 | 2.84 | 3.33 | 3.82 |
| | | **AVERAGE** | **0.86** | **1.00** | **1.61** | **2.10** | **2.73** | **3.19** | **3.65** |

The equations of flood frequency curve vs $Q_T/Q_{2.33}$ values are shown without reduced variate:

Cluster-2: $Q_T/Q_{2.33} = 3*10^{-14}y^4 - 4*10^{-14}y^3 + 10^{-12}y^2 + 0.6042y + 0.6504$ (Equation 5.1)

Cluster-3: $Q_T/Q_{2.33} = -10^{-14}y^4 + 8*10^{-13}y^3 - 9*10^{-13}y^2 + 0.6589y + 0.6188$ (Equation 5.2)

$y = -ln\left(-ln\left(1 - \frac{1}{T}\right)\right)$ (Equation 5.3)

Reduced variate values for Cluster regions and whole region are compared in Figure 5.7. In whole basin reduced variate values are calculated by using stations which are in homogeneous regions. As one could see there is not big difference between whole region and cluster regions. However, cluster-3 has more reduced value for higher time periods. In addition, cluster-2 has generally lower values rather than whole basin values.

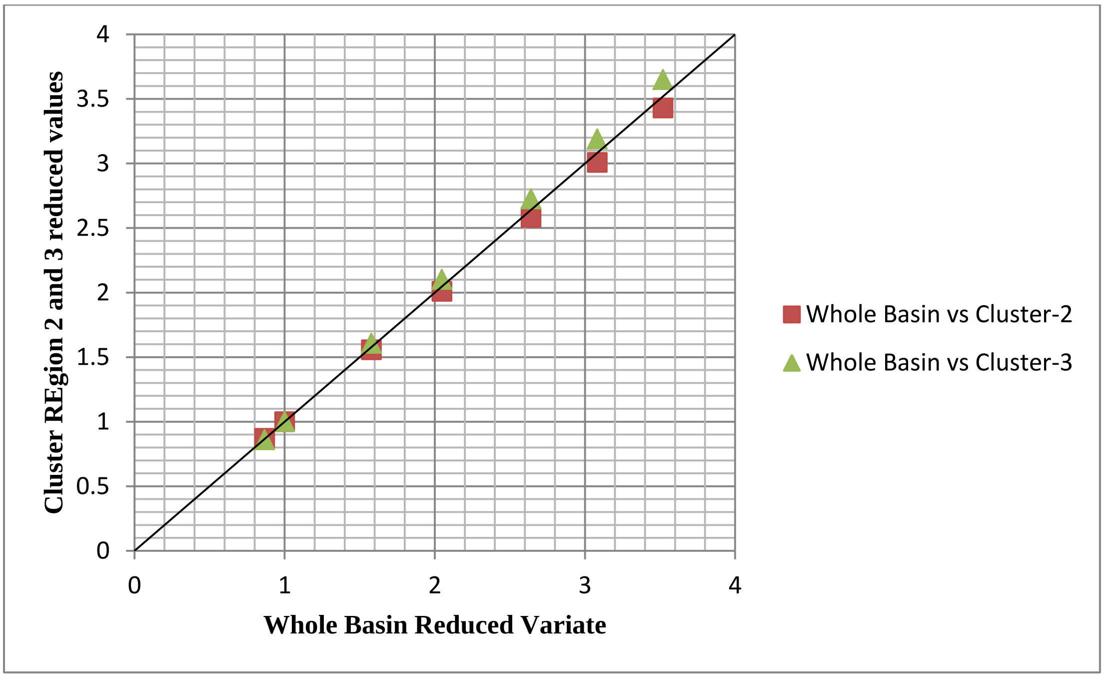


Figure 5.7 Comparison of Whole Basin and Cluster Regions Reduced Variates with Gumbel Distribution.

### 5.7.2 Comparison of equation results which are derived by $Q_{2.33}$ flood values with Different Distribution Parameters

Second, $Q_{2.33}$ flood values which are calculated by using Different Distribution Parameters, are also compared due to the equations which are given in Table 5.1. However, only Nonlinear Regression Analysis with three cluster regions are taken into comparison because of explained reasons and results of Table 5.1. In addition, previous study which is done by Topaloglu(2005) could be not used in this comparison because of previous study includes only Gumbel Distribution Parameters.

In Table 5.6, the results of $Q_{2.33}$ flood values are given due to related calculation equations. Moreover, % difference from the real calculated values, Root Mean Square Errors and Nash-Sutcliffe Efficiency values are also given in this table.

There are some missing results in the Table 5.6. Nonlinear Regression Analysis for cluster-1 could not be calculated due to the explained reasons before. In Dalyrmple Method, there are 12 station results which could not calculated because of these 12 stations are located in non-homogeneous region in the analysis. In Point Flood Analysis, there are only 18 station results, other 22 stations could not be calculated because of there is no relationship between any of them. These missing results are considered while calculating the Root Mean Square Errors.

Considering the Root Mean Square Errors result, there is big difference between the Nonlinear Regression Analysis and the other methods. This shows that grouping and making analysis for basin characteristics gives more accurate results with less errors. And also, Nonlinear Regression Analysis % differences are the most lower values. There are only some stations that have unexpected result. For example DSI/20-51 station has %218.15 difference.

In Point Flood Analysis, % differences are not close to zero and differs much, this shows that Point Analysis could not give reliable results for Ceyhan Basin with Different Distribution Parameters. Therefore, Nonlinear Regression Analysis with three clusters give much more reliable results rather than Point Analysis.

In Dalyrmple Method which is made by envelope line equation, % differences are more higher than the real calculated values. This is because of envelope line is drawn to include all points including in the graph. Therefore, the results of this equation gives higher results than the real results. Moreover, Dalyrmple offer this method for only Gumbel Distributions, but in this part of the study Different Distribution Parameters are used.

In Dalyrmple Method which is made by best fit equation, % differences are the most closest to the Nonlinear Regression Analysis with three clusters. Moreover, in some stations the results of this analysis is better than Nonlinear Regression Analysis.

The relationships between the obtained results and the real station values are shown in Figure 5.8, Figure 5.9, Figure 5.10 and Figure 5.11.

Table 5.6 Comparison of Results which are derived for $Q_{2.33}$ flood values for Different Distributions.

| Stream Gauging Station # | | AREA ($km^2$) | $Q_{2.33}$ Values with Different Distribution Parameters ($m^3/s$) | Nonlinear Analysis For Three Cluster Regions ($m^3/s$) | Dalyrmple Method - Best Fit Equation ($m^3/s$) | Dalyrmple Method - Envelope Line Equation ($m^3/s$) | Point Flood Analysis ($m^3/s$) | Nonlinear Analysis For Three Cluster Regions % Difference | Dalyrmple Method - Best Fit Equation % Difference | Dalyrmple Method - Envelope Line Equation % Difference | Point Flood Analysis % Difference |
|---|---|---|---|---|---|---|---|---|---|---|---|
| # 1 | EIE/2001 | 8484 | 536.54 | | | | 172.72 | | | | -67.81% |
| | EIE/2004 | 20466 | 1089.20 | | | | 1116.72 | | | | 2.53% |
| | EIE/2012 | 19727.2 | 1089.68 | | | | 1062.83 | | | | -2.46% |
| # 2 | EIE/2007 | 623 | 58.86 | 59.53 | | | 124.16 | 1.15% | | | 110.97% |
| | EIE/2008 | 444 | 146.34 | 99.39 | 54.17 | 201.73 | | -32.08% | -62.98% | 37.85% | |
| | EIE/2010 | 3498.8 | 285.82 | 277.43 | | | | -2.94% | | | |
| | DSI/20-04 | 178 | 28.54 | 48.93 | 34.18 | 96.01 | | 71.44% | 19.74% | 236.38% | |
| | DSI/20-05 | 94 | 36.82 | 51.19 | 24.77 | 57.16 | | 39.04% | -32.72% | 55.25% | |
| | DSI/20-06 | 174.9 | 51.84 | 74.44 | 33.87 | 94.65 | | 43.58% | -34.66% | 82.57% | |
| | DSI/20-08 | 131.1 | 28.72 | 80.46 | 29.29 | 74.89 | | 180.14% | 2.00% | 160.76% | |
| | DSI/20-10 | 217.3 | 62.53 | 42.86 | | | | -31.45% | | | |
| | DSI/20-13 | 105.1 | 46.61 | 24.50 | 26.21 | 62.58 | | -47.43% | | | |
| | DSI/20-14 | 310.5 | 149.17 | 105.52 | 45.24 | 150.87 | 148.79 | -29.26% | | | -0.25% |
| | DSI/20-17 | 1740 | 86.40 | 129.24 | 107.83 | 611.79 | | 49.58% | 24.80% | 608.07% | |
| | DSI/20-40 | 79 | 4.68 | 7.08 | 22.69 | 49.63 | | 51.08% | 384.57% | 959.74% | |
| | DSI/20-43 | 163 | 50.79 | 55.30 | 32.69 | 89.38 | 24.08 | 8.88% | -35.64% | 75.98% | -52.60% |
| | DSI/20-44 | 35 | 19.68 | 6.57 | 15.06 | 25.62 | | -66.59% | -23.48% | 30.19% | |
| | DSI/20-45 | 170 | 55.11 | 21.63 | 33.39 | 92.49 | | -60.75% | -39.41% | 67.82% | |
| | DSI/20-46 | 477 | 198.10 | 191.23 | 56.17 | 213.83 | 198.60 | -3.47% | -71.65% | 7.94% | 0.25% |
| | DSI/20-56 | 238.4 | 22.30 | 11.74 | | | | -47.36% | | | |
| | DSI/20-63 | 57.5 | 12.22 | 7.14 | 19.34 | 38.34 | | -41.56% | 58.18% | 213.66% | |
| | DSI/20-65 | 161 | 39.47 | 4.83 | | | | -87.76% | | | |

Table 5.6 Comparison of Results which are derived for $Q_{2.33}$ flood values for Different Distributions. (Continued)

| Stream Gauging Station # | | AREA ($km^2$) | $Q_{2.33}$ Values with Different Distribution Parameters ($m^3/s$) | Nonlinear Analysis For Three Cluster Regions ($m^3/s$) | Dalyrmple Method - Best Fit Equation ($m^3/s$) | Dalyrmple Method - Envelope Line Equation ($m^3/s$) | Point Flood Analysis ($m^3/s$) | Nonlinear Analysis For Three Cluster Regions % Difference | Dalyrmple Method - Best Fit Equation % Difference | Dalyrmple Method - Envelope Line Equation % Difference | Point Flood Analysis % Difference |
|---|---|---|---|---|---|---|---|---|---|---|---|
| # 3 | **EIE/2005** | 4219.08 | 108.41 | 92.42 | | | 336.78 | -14.76% | | | 210.65% |
| | **EIE/2006** | 733.2 | 53.55 | 67.49 | | | 52.01 | 26.04% | | | -2.87% |
| | **EIE/2009** | 1387.2 | 79.56 | 87.69 | 96.19 | 508.93 | 81.91 | 10.21% | 20.90% | 539.66% | 2.95% |
| | **EIE/2015** | 915.2 | 38.34 | 50.08 | 78.00 | 363.03 | 23.61 | 30.60% | 103.44% | 846.81% | -38.43% |
| | **EIE/2016** | 846.8 | 73.65 | 46.59 | 75.01 | 340.83 | 84.99 | -36.74% | 1.84% | 362.78% | 15.40% |
| | **EIE/2022** | 428 | 53.93 | 35.46 | 53.18 | 195.81 | 46.73 | -34.24% | -1.39% | 263.09% | -13.35% |
| | **DSI/20-02** | 197.1 | 36.54 | 29.13 | 35.98 | 104.30 | | -20.27% | -1.54% | 185.45% | |
| | **DSI/20-07** | 2084 | 40.86 | 61.96 | 118.09 | 708.34 | 66.36 | 51.63% | 189.01% | 1633.52% | 62.41% |
| | **DSI/20-15** | 189.7 | 16.90 | 27.85 | 35.29 | 101.11 | | 64.82% | 108.83% | 498.30% | |
| | **DSI/20-16** | 291 | 23.80 | 33.09 | 43.78 | 143.13 | | 39.01% | 83.93% | 501.29% | |
| | **DSI/20-36** | 174.2 | 62.79 | 41.95 | 33.81 | 94.34 | | -33.19% | -46.16% | 50.25% | |
| | **DSI/20-51** | 131.4 | 8.71 | 26.95 | 29.33 | 75.03 | | 209.34% | 236.63% | 761.22% | |
| | **DSI/20-52** | 23 | 4.97 | 10.48 | 12.18 | 18.22 | | 110.82% | 145.13% | 266.45% | |
| | **DSI/20-53** | 178.5 | 36.37 | 37.36 | | | 47.96 | 2.73% | | | 31.87% |
| | **DSI/20-54** | 207.5 | 53.02 | 40.65 | 36.92 | 108.75 | 40.21 | -23.33% | -30.36% | 105.11% | -24.17% |
| | **DSI/20-55** | 111.6 | 21.06 | 27.43 | | | | 30.25% | | | |
| | **DSI/20-58** | 24.38 | 18.38 | 15.16 | 12.55 | 19.10 | 14.28 | -17.53% | -31.73% | 3.91% | -22.31% |
| | **DSI/20-59** | 171.5 | 52.43 | 54.06 | 33.54 | 93.15 | 67.48 | 3.11% | -36.02% | 77.68% | 28.71% |
| | | | | | | | | | | | |
| | | | | | | **Root Mean Square Errors** | | **21.33** | **43.32** | **203.30** | **103.43** |
| | | | | | | **Nash-Sutcliffen Efficiency** | | **0.854** | **0.101** | **-1.664** | **-0.578** |

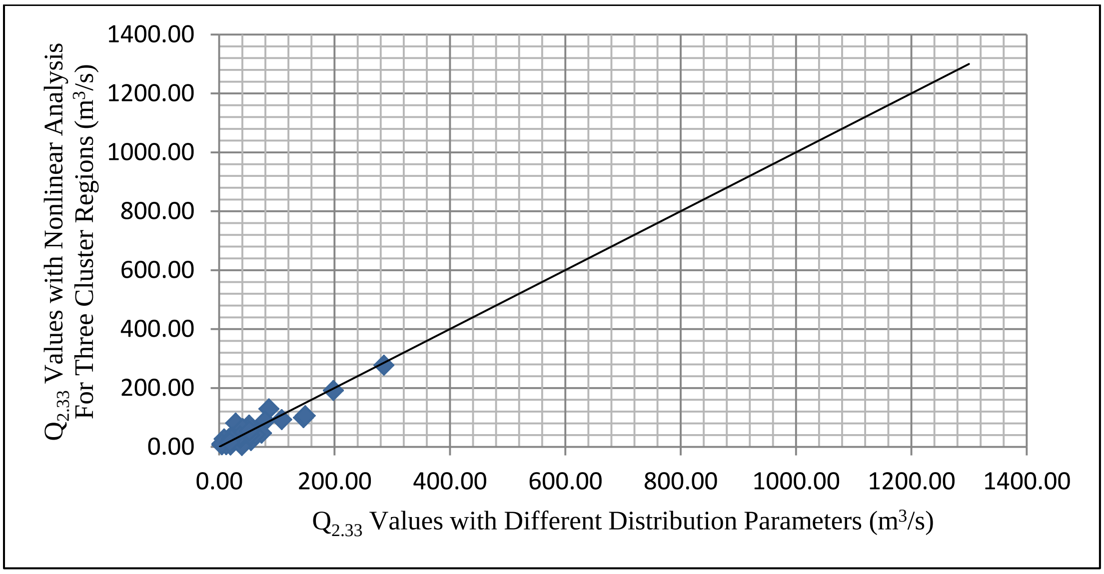


Figure 5.8 Comparison between the Real and Nonlinear Analysis for Three Cluster Region $Q_{2.33}$ Values.

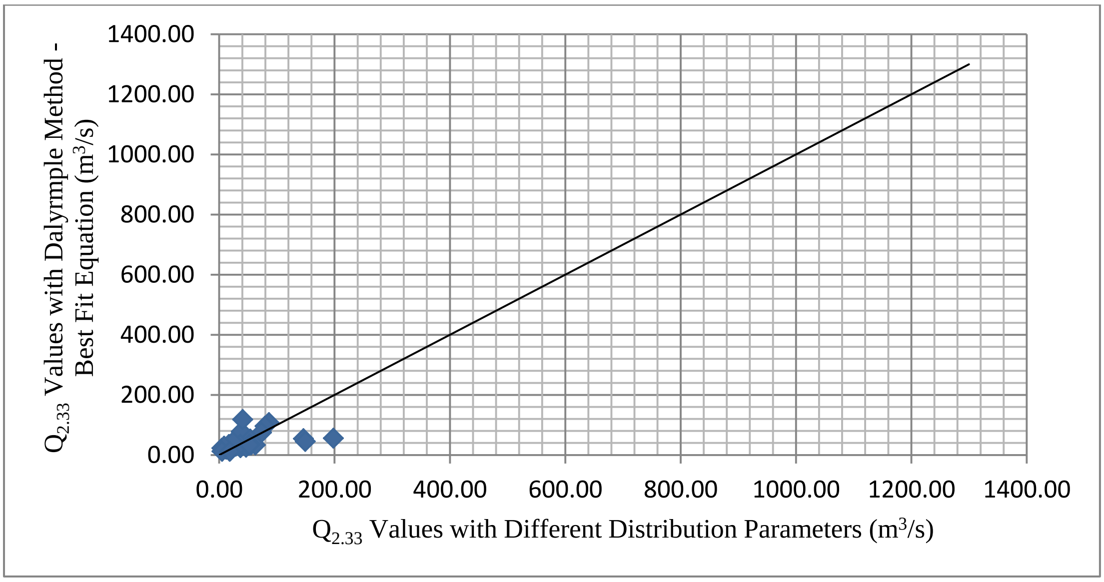


Figure 5.9 Comparison between the Real and Dalyrmple Method - Best Fit Equation $Q_{2.33}$ Values.

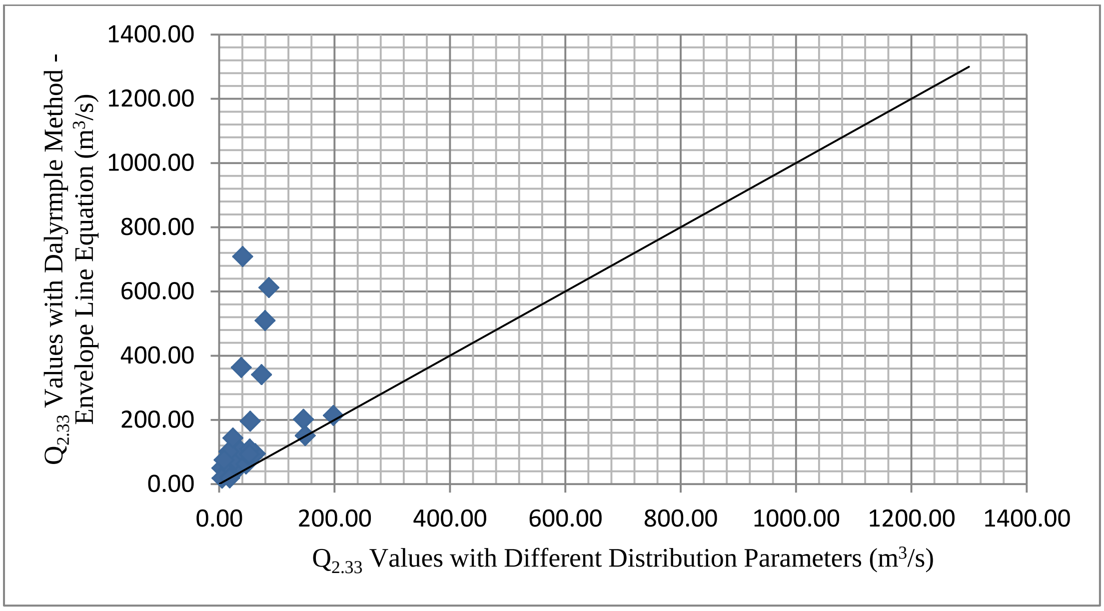


Figure 5.10 Comparison between the Real and Dalyrmple Method - Envelope Line Equation $Q_{2.33}$ Values.

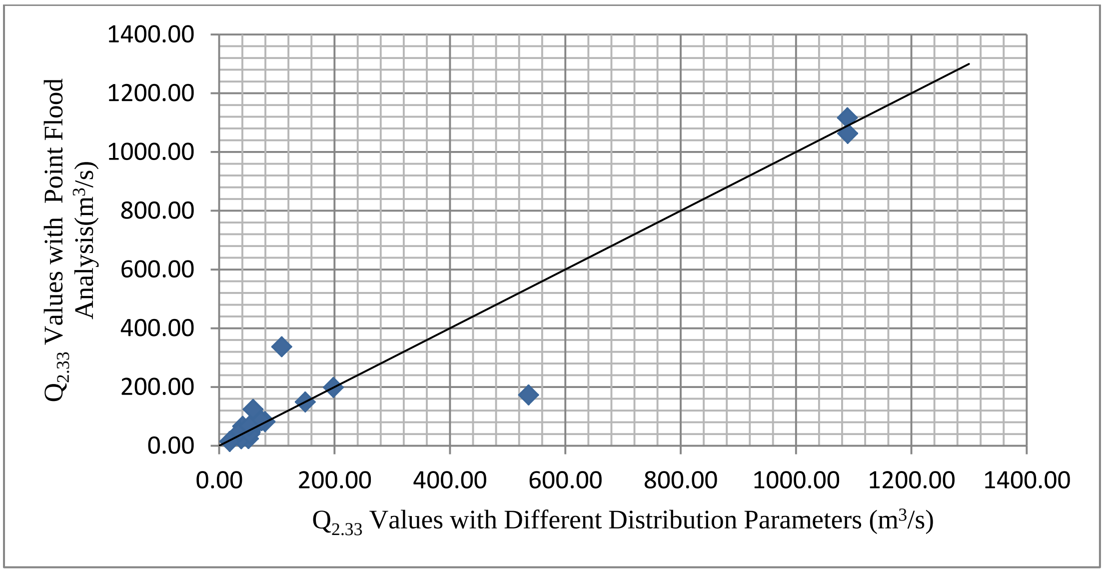


Figure 5.11 Comparison between the Real and Point Flood Analysis $Q_{2.33}$ Values.

As the calculation results show Nonlinear Regression Analysis with three clusters is the best method to calculate $Q_{2.33}$ flood values. $Q_{100}$, $Q_{50}$ ,. flood values can be calculated by using $Q_{2.33}$ flood values. $Q_{2.33}$ flood values are multiplied by related number to obtain any Flood Frequency for cluster region. Therefore, Table 5.7 is used to obtain multiplication factors for any Flood Frequency in cluster region. However, there is not any multiplication factor for cluster region-1 because Nonlinear Regression Analysis could not be performed for this cluster.

In addition, averages of multiplication is shown with reduced variate values in Figure 5.12

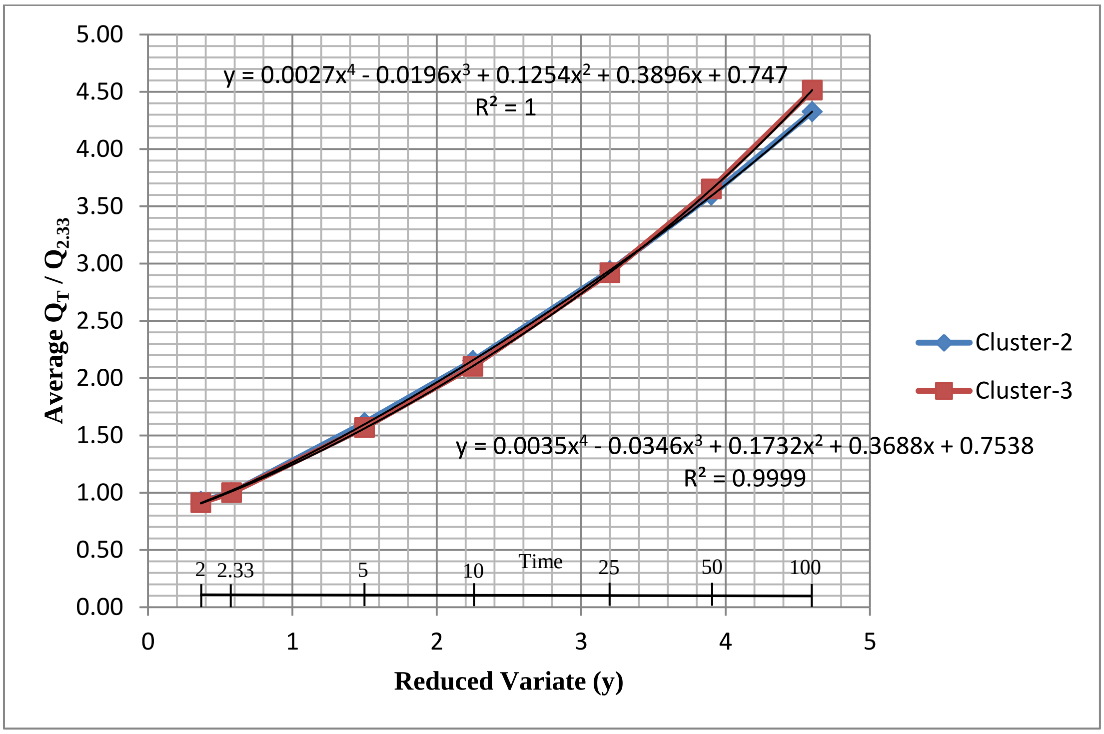


Figure 5.12 RFFA Curve for Different Clusters with Different Distribution Parameters.

The equations of flood frequency curve vs $Q_T/Q_{2.33}$ values are shown without reduced variate:

Cluster-2: $Q_T/Q_{2.33} = 0.0035y^4 - 0.0346y^3 + 0.1732y^2 + 0.3688y + 0.7538$ (Equation 5.4)

Cluster-3: $Q_T/Q_{2.33} = 0.0027y^4 - 0.0196y^3 + 0.1254y^2 + 0.3896y + 0.747$ (Equation 5.5)

$$y = -ln\left(-ln\left(1 - \frac{1}{T}\right)\right)$$ (Equation 5.6)

Reduced variate values for Cluster regions and whole region are compared in the Figure 5.13. In whole basin reduced variate values are calculated by using stations which are in homogeneous region. There are 28 stations in homogeneous region. Therefore there is significant difference between the results of reduced variates. Both cluster-2 and cluster-3 values are lower than the whole region results.

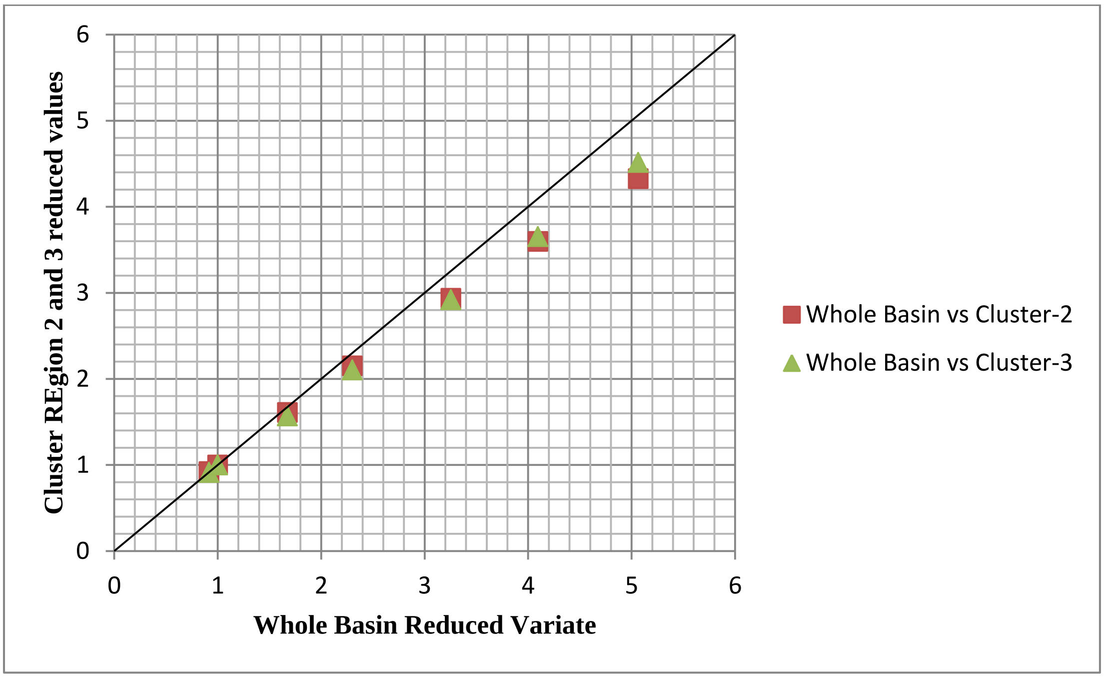


Figure 5.13 Comparison of Whole Basin and Cluster Regions Reduced Variates With Different Distribution Parameters.

Table 5.7 Average Ratios of Multiplication Factors for Different Clusters.

| Cluster Region | Station | Drainage Area ($km^2$) | $Q_2$ | $Q_{2.33}$ | $Q_5$ | $Q_{10}$ | $Q_{25}$ | $Q_{50}$ | $Q_{100}$ |
|---|---|---|---|---|---|---|---|---|---|
| # 2 | **EIE/2007** | 623.00 | 0.93 | 1.00 | 1.29 | 1.53 | 1.83 | 2.05 | 2.28 |
| | **EIE/2008** | 444.00 | 0.89 | 1.00 | 1.50 | 1.96 | 2.60 | 3.11 | 3.65 |
| | **EIE/2010** | 3498.80 | 0.91 | 1.00 | 1.38 | 1.68 | 2.05 | 2.31 | 2.58 |
| | **DSI/20-04** | 178.00 | 0.92 | 1.00 | 1.97 | 2.81 | 3.99 | 4.91 | 5.86 |
| | **DSI/20-05** | 94.00 | 0.90 | 1.00 | 1.80 | 2.49 | 3.40 | 4.09 | 4.77 |
| | **DSI/20-06** | 174.90 | 0.96 | 1.00 | 1.45 | 1.80 | 2.28 | 2.66 | 3.06 |
| | **DSI/20-08** | 131.10 | 0.93 | 1.00 | 1.82 | 2.67 | 4.09 | 5.46 | 7.13 |
| | **DSI/20-10** | 217.30 | 0.92 | 1.00 | 1.33 | 1.60 | 1.95 | 2.20 | 2.45 |
| | **DSI/20-13** | 105.10 | 0.95 | 1.00 | 1.54 | 1.92 | 2.38 | 2.70 | 3.00 |
| | **DSI/20-14** | 310.50 | 0.95 | 1.00 | 1.57 | 1.96 | 2.40 | 2.69 | 2.94 |
| | **DSI/20-17** | 1740.00 | 0.91 | 1.00 | 2.10 | 3.13 | 4.65 | 5.92 | 7.27 |
| | **DSI/20-40** | 79.00 | 0.94 | 1.00 | 1.67 | 2.26 | 3.10 | 3.80 | 4.56 |
| | **DSI/20-43** | 163.00 | 0.93 | 1.00 | 1.84 | 2.69 | 4.10 | 5.44 | 7.06 |
| | **DSI/20-44** | 35.00 | 0.91 | 1.00 | 2.07 | 3.18 | 5.03 | 6.78 | 8.87 |
| | **DSI/20-45** | 170.00 | 0.94 | 1.00 | 1.64 | 2.23 | 3.16 | 3.98 | 4.94 |
| | **DSI/20-46** | 477.00 | 0.90 | 1.00 | 1.44 | 1.81 | 2.29 | 2.66 | 3.04 |
| | **DSI/20-56** | 238.40 | 0.94 | 1.00 | 1.23 | 1.39 | 1.56 | 1.66 | 1.76 |
| | **DSI/20-63** | 57.50 | 0.88 | 1.00 | 1.58 | 2.15 | 2.97 | 3.67 | 4.43 |
| | **DSI/20-65** | 161.00 | 0.92 | 1.00 | 1.34 | 1.62 | 1.99 | 2.26 | 2.54 |
| | | **AVERAGE** | **0.92** | **1.00** | **1.61** | **2.15** | **2.94** | **3.60** | **4.33** |
| # 3 | **EIE/2005** | 4219.08 | 0.93 | 1.00 | 1.32 | 1.59 | 1.93 | 2.19 | 2.45 |
| | **EIE/2006** | 733.20 | 0.91 | 1.00 | 1.37 | 1.67 | 2.02 | 2.28 | 2.53 |
| | **EIE/2009** | 1387.20 | 0.90 | 1.00 | 1.46 | 1.86 | 2.41 | 2.85 | 3.30 |
| | **EIE/2015** | 915.20 | 0.88 | 1.00 | 1.58 | 2.13 | 2.94 | 3.62 | 4.36 |
| | **EIE/2016** | 846.80 | 0.89 | 1.00 | 1.51 | 1.99 | 2.66 | 3.22 | 3.81 |
| | **EIE/2022** | 428.00 | 0.94 | 1.00 | 1.70 | 2.39 | 3.52 | 4.57 | 5.84 |
| | **DSI/20-02** | 197.10 | 0.89 | 1.00 | 1.51 | 1.99 | 2.66 | 3.21 | 3.80 |
| | **DSI/20-07** | 2084.00 | 0.93 | 1.00 | 1.78 | 2.52 | 3.71 | 4.78 | 6.03 |
| | **DSI/20-15** | 189.70 | 0.92 | 1.00 | 1.95 | 3.06 | 5.14 | 7.35 | 10.32 |
| | **DSI/20-16** | 291.00 | 0.93 | 1.00 | 1.83 | 2.77 | 4.55 | 6.46 | 9.01 |
| | **DSI/20-36** | 174.20 | 0.88 | 1.00 | 1.51 | 1.92 | 2.44 | 2.83 | 3.22 |
| | **DSI/20-51** | 131.40 | 0.94 | 1.00 | 1.65 | 2.33 | 3.47 | 4.59 | 5.99 |
| | **DSI/20-52** | 23.00 | 0.87 | 1.00 | 1.66 | 2.33 | 3.33 | 4.20 | 5.18 |
| | **DSI/20-53** | 178.50 | 0.93 | 1.00 | 1.33 | 1.61 | 1.98 | 2.26 | 2.54 |
| | **DSI/20-54** | 207.50 | 0.93 | 1.00 | 1.53 | 1.93 | 2.43 | 2.78 | 3.13 |
| | **DSI/20-55** | 111.60 | 0.92 | 1.00 | 1.35 | 1.63 | 1.99 | 2.26 | 2.52 |
| | **DSI/20-58** | 24.38 | 0.89 | 1.00 | 1.49 | 1.89 | 2.39 | 2.77 | 3.14 |
| | **DSI/20-59** | 171.50 | 0.94 | 1.00 | 1.70 | 2.25 | 2.98 | 3.54 | 4.09 |
| | | **AVERAGE** | **0.91** | **1.00** | **1.57** | **2.10** | **2.92** | **3.65** | **4.51** |

### 5.7.3 Comparison between Gumbel Distribution and Different Distributions $Q_T/Q_{2.33}$ values

As all the results are obtained with Gumbel Distribution and Different Distributions, one could see that there is big difference between the reduced variate values. The Gumbel Distribution reduced variate results are given in Table 5.8 and Different Distributions reduced variate results are given in Table 5.9.

Table 5.8 Gumbel Distribution Reduced Variate Results.

| Gumbel Distribution | | | | |
|---|---|---|---|---|
| **T** | **Whole Region** | **Cluster-1** | **Cluster-2** | **Cluster-3** |
| **2** | 0.87 | - | 0.87 | 0.86 |
| **2.33** | 1 | - | 1 | 1 |
| **5** | 1.58 | - | 1.56 | 1.61 |
| **10** | 2.05 | - | 2.01 | 2.1 |
| **25** | 2.64 | - | 2.58 | 2.73 |
| **50** | 3.08 | - | 3.01 | 3.19 |
| **100** | 3.52 | - | 3.43 | 3.65 |

Table 5.9 Different Distributions Reduced Variate Results.

| Different Distributions | | | | |
|---|---|---|---|---|
| **T** | **Whole Region** | **Cluster-1** | **Cluster-2** | **Cluster-3** |
| **2** | 0.92 | - | 0.92 | 0.91 |
| **2.33** | 1 | - | 1 | 1 |
| **5** | 1.67 | - | 1.61 | 1.57 |
| **10** | 2.3 | - | 2.15 | 2.1 |
| **25** | 3.25 | - | 2.94 | 2.92 |
| **50** | 4.09 | - | 3.6 | 3.65 |
| **100** | 5.06 | - | 4.33 | 4.51 |

$Q_{100}/Q_{2.33}$ value for whole region with Gumbel Distribution value is 3.52 and $Q_{100}/Q_{2.33}$ value for whole region with Different Distributions value is 5.06. There is big difference between these factors so the Different Distributions should give higher $Q_{100}$ flood values than Gumbel Distribution values.

For cluster regions, Different Distributions $Q_{100}/Q_{2.33}$ values are also higher than Gumbel Distribution values.

$Q_{100}/Q_{2.33}$ value for, whole region value is 3.52, cluster-2 value is 3.43 and cluster-3 value is 3.65 with Gumbel Distribution. This shows that clustering have effect on the calculation of $Q_{100}$ flood results.

$Q_{100}/Q_{2.33}$ value for, whole region value is 5.06, cluster-2 value is 4.33and cluster-3 value is 4.51 with Different Distributions. This also shows that clustering have effect on the calculation of $Q_{100}$ flood results. There is drastic decrease on the clustered reduced variate values, this is because of whole region reduced variate values are calculated by using 28 homogeneous stations. This shows that homogeneity test eliminated the smaller reduced variates for Different Distributions so mean reduced varied values is higher.

# CHAPTER 6

# CONCLUSIONS AND RECOMMENDATIONS

## 6.1 Conclusions

For Ceyhan Basin, Dalyrmple (1960) Regional Flood Frequency Analysis Method is compared with Stepwise Regression Analysis, Nonlinear Regression Analysis and Point Flood Analysis in this research.

Dalyrmple (1960) offers to use Gumbel Distribution in the Dalyrmple Method. Several references in the literature followed closely Dalrymple's (1960) method by taking the 2.33-year event as the index flood. Several well-known references in the literature (Linsley et al. 1982; Viessman and Lewis 1996) followed closely Dalrymple's method of using the 2.33-year event as the index flood. However, this method is not used with only Gumbel Distribution in DSI projects. Three Parameter Lognormal, Pearson Type III, Log-pearson Type III, Two Parameter Lognormal, Normal are the parameters which are used as another option to Gumbel distribution. Therefore, this procedure is also compared with Stepwise Regression Analysis, Nonlinear Regression Analysis and Point Flood Analysis.

To make Nonlinear Regression Analysis and Stepwise Regression Analysis, different basin characteristics which are Mean Annual Rainfall, Perimeter, Drainage Area, $H_{min}$, $H_{max}$, $H_{mean}$, $H_{\Delta}$, Basin Relief, Relative Relief and Circularity are selected to use in the model as independent variables.

As there are 40 stations in the analysis, clustering technique is applied to find more homogeneous groups in the region. At the end of the Analysis, three cluster regions are obtained. These regions are divided into groups according to the station's Drainage Areas and $H_{min}$ values. In addition, clustered regions and non-clustered whole basin results are also compared with each other.

The most effective basin characteristics are selected by Stepwise Regression Analysis and the selected basin characteristics are used in Nonlinear Regression Analysis for comparison of results.

For only Gumbel Distribution parameters, the most effective method is Nonlinear Regression Analysis with clusters. This method could not be used for cluster region-1 but this region stations have very high drainage areas. The aim of this study is about to calculate the flood values in ungauged areas so this is not necessary to present an equation for that kind of huge drainage areas. The obtained equation for Cluster-2 which are developed with the bottom stations of Ceyhan Basin is:

$$\mathbf{Q_{2.33} = 0.0026 * Length\ of\ Main\ River^{(0.991)} * Basin\ Relief^{(1.054)}}$$

The equation for Cluster-3 which is developed with upper stations of Ceyhan Basin is:

$$\mathbf{Q_{2.33} = 2.896 * Perimeter^{(0.566)}}$$

For Different Distribution Parameters, there are 12 stations which are stated nonhomogeneous region in Dalyrmple Method. This number is only 2 with Gumbel Distribution. Therefore, using different kind of distribution parameters could not give accurate results with Dalyrmple(1960) Method.

After calculation of $Q_{2.33}$ flood values, one could obtain the related flood values easily by using related Reduced Variate vs. $Q_T/Q_{2.33}$ figures. These figures are not only for $Q_{100}$ flood values, this could give the result of whatever period is wanted.

Topaloglu (2005) used 15 stations, this research includes 40 stations. Comparing the results of Topaloglu (2005) and Dalyrmple (1960) results, increasing the number of results give more accurate results in Regional Flood Frequency Analysis.

There is big difference between the results which have been calculated for whole region and cluster regions. This shows, clustering is very effective for both Stepwise Regression Analysis and Nonlinear Regression Analysis Methods.

Clustered regions are divided into groups mostly for according to the station's Drainage Areas and $H_{min}$ values. Therefore these parameters are not so effective in the analysis and not selected as an important basin parameter in the equations.

In Dalyrmple(1960) Method, only the drainage area of basin is used to calculate flood values. In Stepwise and Nonlinear Regression Analysis methods, different basin parameters are used. The results show that additional basin characteristics decreased the final results.

$Q_{100}/Q_{2.33}$ value for, whole region value is 3.52, cluster-2 value is 3.43 and cluster-3 value is 3.65 with Gumbel Distribution. $Q_{100}/Q_{2.33}$ value for, whole region value is 5.06, cluster-2 value is 4.33 and cluster-3 value is 4.51 with Different Distributions. Therefore this shows that there is big difference between the values obtained from Different Distributions and Gumbel Distribution. Using Different Distributions give higher $Q_{100}$ results in Ceyhan Basin. In addition, clustered and non-clustered results has also diversity.

To simplify the application procedure of Nonlinear Regression Analysis with three clusters:

- One should decide that Gumbel Distriution or Different Distributions will be used in the analysis.
- Decide the cluster regions of the basin.
- Select related equation in Table 5.1
- Evaluate basin characteristics those are used in the selected equation
- Calculate of $Q_{2.33}$ flood value by using the selected equation
- Multiply $Q_{2.33}$ flood value with reduced variate for any desired record period

**6.2 Recommendations**

Here are the recommendations:

• The basin characteristics are selected due to the accessibility of the data and to provide an easy methodology depending less data for practical use. There should be additional basin characteristics for geology, land cover, snow, temperature etc. to get more accurate results.

• The proposed models can be tested for different basins.

• Other clustering techniques and Regression Analysis Method can be also tried for this basin.

# REFERENCES

Abbott M.B., Bathurst J.C., Cunge J.A., Oconnell P.E., Rasmussen J. (1986). An Introduction to the European Hydrological System - Systeme Hydrologique Europeen, She .1. History and Philosophy of a Physically-Based, Distributed Modeling System. Journal of Hydrology. 87(1-2):45-59.

Benson, M. A. (1962). Plotting positions and economics of engineering planning. Proc. Amer. Soc. Civ. Eng. Hydraul. Div.,88 (HY6), 57–71.

Bertoldi L, Massironi M, Visona D, Carosi R, Montomoli C, Gubert F, Naletto G, Pelizzo MG. (2011). Mapping the Buraburi granite in the Himalaya of Western Nepal: Remote sensing analysis in a collisional belt with vegetation cover and extreme variation of topography. Remote Sensing of Environment. 115(5):1129-1144.

Beven K.J., Kirkby M.J., Schofield N., Tagg A.F. (1984). Testing a Physically-Based Flood Forecasting-Model (Topmodel) for 3 UK Catchments. Journal of Hydrology. 69(1-4):119-143.

Bobee, B. (1975). The log pearson type 3 distribution and its applications in hydrology, Water Resour. Res., 11(3), 365–369.

Bobee, B., and R. Robitaille. (1977). The use of the Pearson type 3 and log Pearson type 3 distributions revisited, Water Resour. Res., 13(2), 427–443.

Bobee, B.; Ashkar, F. (1991): The gamma family and derived distributions applied in hydrology. Water Resources Publications, Yevjevich Edt., Littleton, Co. 217.

Boyle D.P., Gupta H.V., Sorooshian S., Koren V., Zhang Z.Y., Smith M. (2001). Toward improved streamflow forecasts: Value of semidistributed modeling. Water Resources Research. 37(11):2749-2759.

Brink A.B., Eva H.D. (2009). Monitoring 25 years of land cover change dynamics in Africa: A sample based remote sensing approach. Applied Geography. 29(4):501-512.

Burn, D.H., (1989). Cluster analysis as applied to regional flood frequency. J. Water Resour. Plann. Mgmt. 115 (5), 567-582.

Burn, D.H., (1990). Evaluation of regional flood frequency analysis with a region of influence approach. Water Resour. Res. 26, 2257-2265.

Burn D.H. (1997). Catchment similarity for regional flood frequency analysis using seasonality measures. Journal of Hydrology, 202(1-4):212-230.

Burn, D.H., Goel, N.K., (2000). The formation of groups for regional flood frequency estimation. Hydrol. Sci. J. 45 (1), 97e112.

Castellarin A, Burn D.H., Brath A. (2001). Assessing the effectiveness of hydrological similarity measures for flood frequency analysis. Journal of Hydrology. 241(3- 4):270-285.

Castiglioni, S., Castellarin, A., and Montanari, A. (2009). Prediction of low-flow indices in ungauged basins through physiographical space-based interpolation. Journal of Hydrology, 378(3-4), 272-280.

Chiang S.M., Tsay T.K., Nix S.J. (2002a). Hydrologic regionalization of watersheds. I: Methodology development. Journal of Water Resources Planning and Management. 128(1):3-11.

Chiang SM, Tsay T.K., Nix S.J. (2002b). Hydrologic regionalization of watersheds. II: Applications. Journal of Water Resources Planning and Management. 128(1):12- 20.

Condie, R. (1977). The log-pearson type 3 distribution: The t-year event and its asymptotic standard error by maximum likelihood, Water Resour. Res., 13(6), 987–991.

Corbane C, Raclot D, Jacob F, Albergel J, Andrieux P. (2008). Remote sensing of soil surface characteristics from a multiscale classification approach. Catena. 75(3):308-318.

Cunnane, C. (1988). Methods and merits of regional flood frequency analysis. Journal of Hydrology, 100, 269-290.

Dalrymple, T. (1960). Flood-Frequency Analyses. Water-Supply Paper: 1543-A, U.S. Washington, 80p. Geological Survey, Washington, D.C.

De Michele C, Rosso R. (2001). Uncertainty assessment of regionalized flood frequency estimates. Journal of Hydrologic Engineering. 6(6):453-459.

DSİ, (2007). DSİ Akım Gözlem İstasyonları Yılda Anlık Maksimum Akımları 1954 – 2000. General Directorate of State Hydraulic Works, Ankara.

Duffy C.J. (2004). Semi-discrete dynamical model for mountain-front recharge and water balance estimation: Rio Grande of southern Colorado and New Mexico, in Groundwater Recharge in a Desert Environment: The Southwestern United States. Water Sci. Appl. Ser., vol. 9, JF Hogan, F Phillips, and B Scanlon (eds.), 255 – 271.

Eng K, Tasker G.D., Milly P.C.D. (2005). An analysis of region-of-influence methods for flood regionalization in the gulf-atlantic rolling plains. Journal of the American Water Resources Association. 41(1):135-143.

Eng K, Stedinger J.R., Gruber A.M. (2007a). Regionalization of streamflow characteristics for the Gulf-Atlantic rolling plains using leverage-guided region-of-influence regression. World Environmental & Water Resources Conference - May 15-18.

Eng K, Milly P.C.D., Tasker G.D. (2007b). Flood regionalization: A hybrid geographic and predictor-variable region-of-influence regression method. Journal of Hydrologic Engineering. 12(6), 585-591.

Fill H.D., Stedinger J.R. (1998). Using regional regression within index flood procedures and an empirical Bayesian estimator. Journal of Hydrology. 210(1-4):128-145.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. Wat. Resour. Res., 15, 1049-1054

Greis, N.P. & Wood, E.F. (1981) Regional flood frequency estimation and network design. Wat. Resour. Res. 17(4), 1167-1177.

Griffis V.W., Stedinger J.R. (2007a). The use of GLS regression in regional hydrologic analyses. Journal of Hydrology. 344(1-2):82-95.

Griffis V.W., Stedinger J.R. (2007b). Evolution of flood frequency analysis with Bulletin 17. Journal of Hydrologic Engineering. 12(3):283-297.

Grover, P. L., D. H. Burn, and J. M. Cunderlik (2002), A comparison of index flood estimation procedures for ungauged catchments, Can. J. Civ. Eng., 29, 734–741.

Gruber A.M., Stedinger J.R. (2008). Models of LP3 regional skew, data selection, and Bayesian GLS regression. In proc., World Environmental & Water Resources Conf., R Babcock and RWalton, eds., ASCE, Reston, Va., Paper No. 596.

Inbar M, Gilichinsky M, Melekestsev I, Melnikov D, Zaretskaya N. (2011). Morphometric and morphological development of Holocene cinder cones: A field and remote sensing study in the Tolbachik volcanic field, Kamchatka. Journal of Volcanology and Geothermal Research. 201(1-4):301-311.

Interagency Committee on Water Data (IACWD). (1982). Guidelines for Determining Flood Flow Frequency, Bulletin 17B, 28.

Viessman, W., and G. L. Lewis, (1996) Introduction to Hydrology, 4th ed.,HarperCollins, New York.

Jeong D, Stedinger J.R., Kim Y, Sung J.H. (2007). Bayesian GLS for Regionalization of Flood Characteristics in Korea. World Environmental and Water Resources Congress.

Johnson R.A., Wichern D.W. (2007). Applied Multivariate Statistical Analysis. Pearson Education, Upper Saddle River, 773.

Haan, C. T. (1977). Statistical Methods in Hydrology. Iowa State University Press, Ames, Iowa, 378.

Hosking J.R.M., Wallis J.R. (1988). The Effect of Intersite Dependence on Regional Flood Frequency-Analysis. Water Resources Research. 24(4):588-600.

Hosking J.R.M, Wallis J.R. (1997). Regional Frequency Analysis: An Approach Based on Lmoments. Cambridge University Press, New York, 224.

Karahan, H. (2010). Şiddet-Süre-Frekans Bağıntısının Armoni Araştırma Tekniği ile Belirlenmesi ve Ege Bölgesi İstasyonları İçin Uygulama, 6. Ulusal Hidroloji Kongresi Bildiriler Kitabı, pp. 210-228.

Kite, G.W., (1977). Frequency and Risk Analysis in Hydrology, Fort Collins. Water Resources. Publications, 224.

Kjeldsen T.R., Rosbjerg D. (2002). Comparison of regional index flood estimation procedures based on the extreme value type I distribution. Stochastic Environmental Research and Risk Assessment. 16(5):358-373.

Kolmogorov A. (1933). "Sulla determinazione empirica di una legge di distribuzione". G. Inst. Ital. Attuari 4: 83.

Linsley, R. K., M. A. Kohler, and J. L. H. Paulhaus, (1982). Hydrology for Engineers, 3rd ed., McGraw-Hill, New York.

Lowery, M.D., Nash, J.E. (1970) A comparison of methods of fitting the double exponential distribution. J. Hydrol. 10, 259-275.

MacKinnon D, Tetzlaff D. (2009). Conceptualising Scale in Regional Studies and Catchment Science – Towards an Integrated Characterisation of Spatial Units. Geography Compass. 3(3): 976-996.

Madsen H, Rosbjerg D. (1997). Generalized least squares and empirical Bayes estimation in regional partial duration series index-flood modeling. Water Resources Research. 33(4):771-781.

Malekinezhad H, Nachtnebel HP, Klik A. (2011). Comparing the index-flood and multiple regression methods using L-moments. Physics and Chemistry of the Earth. 36(1-4):54-60.

Mamun, A., Hashim, A., and Amir, Z. (2012). ”Regional Statistical Models for the Estimation of Flood Peak Values at Ungauged Catchments: Peninsular Malaysia.” J. Hydrol. Eng., 17(4), 547–553.

Matalas, N. C., J. R. Wallis. (1973). Eureka! It fits a Pearson type 3 distribution, Water Resour. Res., 9(2), 281–289.

McCuen R.H, Leahy R.B., and Johnson P.A. (1990). Problems with logarithmic transformations in regression. J. Hydraulic Engineering, 116.

McDonnell J.J, Sivapalan M, Vache K, Dunn S, Grant G, Haggerty R, Hinz C, Hooper R, Kirchner J, Roderick ML et al. (2007). Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. Water Resources Research. 43(7).

Mishra A.K., Coulibably P. (2009). Development in hydrometric networks design: A review. Rev. Geophys. 47, RG2001.

Nash, J. E., Shaw, B. L. (1965). “Flood frequency as a function of catchment characteristics.” Proc., River Flood Hydrology Symposium, Institute of Civil Engineers, London.

Nash, J. E., Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology, 10(3), 282-290

NERC (Natural Environment Research Council). (1975). Flood Studies Report, in Five Volumes, Vol. 1. Hydrological Studies, London, 550p.

NERC (Natural Environment Research Council). (1988). Estimating Probabilities of Extreme Floods. National Academy Press, Washington, D.C., 141.

Neykov N.M., Neytchev P.N., Van Gelder PHAJM, Todorov V.K.. (2007). Robust detection of discordant sites in regional frequency analysis. Water Resources Research.43(6).

Nguyen, V.T.V., Pandey, G.R. (1999). New statistical approaches to regional estimation of floods. Proceedings of the International Workshop on Floods and Droughts, 102-113.

P.A. Bostan, G.B.M. Heuvelink, S.Z. Akyurek, (2012). Comparison of regression and kriging techniques for mapping the average annual precipitation of Turkey. International Journal of Applied Earth Observation and Geoinformation 19,115–126.

Panday S, Huyakorn PS. (2004). A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow. Advances in Water Resources. 27(4):361-382.

Rao AR, Srinivas VV. (2006). Regionalization of watersheds by hybrid-cluster analysis. Journal of Hydrology. 318(1-4):37-56.

Reed, S., Schaake, J., Zhang, Z. (2007) A Distributed Hydrologic Model and Threshold Frequency-based Method for Flash Flood Forecasting at Ungauged Locations. In: Journal of Hydrology 337, 402–420

Rhee J, Im J, Carbone GJ, Jensen JR. (2008). Delineation of climate regions using in-situ and remotely-sensed data for the Carolinas. Remote Sensing of Environment. 112(6):3099-3111.

Robson, A. J. & Reed, D. W. (1999) Flood Estimation Handbook, vol. 3: Statistical Procedures for Flood Frequency Estimation. Institute of Hydrology.

Saf B. (2009). Regional Flood Frequency Analysis Using L-Moments for the West Mediterranean Region of Turkey. Water Resources Management. 23(3):531-551.

Schmee, J., Openlander, J., (2010) JMP Means Business: Statistical Models for Management, SAS Publishing.

Shaw, E.M. (1983). Hydrology in Practice. Van Nostrand Reinhold, Berkshire, UK.

Shu C, Ouarda TBMJ. (2008). Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. Journal of Hydrology. 349(1-2):31-43.

Srinivas VV, Tripathi S, Rao AR, Govindaraju RS. (2008). Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. Journal of Hydrology, 348(1-2):148-166.

Stedinger JR, Lu LH. (1995). Appraisal of Regional and Index Flood Quantile Estimators. Stochastic Hydrology and Hydraulics. 9(1):49-75.

Stedinger, J. (1983). "Confidence Intervals for Design Events." J. Hydraul. Eng., 109(1), 13–27.

Stedinger JR, Vogel RM, Foufoula-Georgiou E. (1993). "Frequency Analysis of Extreme Events", in Handbook of Hydrology, chap. 18, pp. 18.1-18.66, McGraw-Hill Book Co., NY.

Tasker GD, Stedinger JR. (1989). An Operational GLS Model for Hydrologic Regression. Journal of Hydrology. 111(1-4):361-375.

Tasker GD, Hodge SA, Barks CS. (1996). Region of influence regression for estimating the 50-year flood at ungaged sites. Water Resources Bulletin. 32(1):163-170.

Tetzlaff D., McDonnell J.J., Uhlenbrook S., McGuire K.J., Bogaart P.W., Naef F. (2008). Conceptualizing catchment processes: simply too complex? Hydrological Processes 22 (11), 1727-1730.

Thomas, D. M., Benson, M. A. (1970) Generalization of streamflow characteristics from drainage-basin characteristics. USGS Water Supply Paper.

Todorovic, P., Zelenhasic, V. (1970) A stochastic model for flood analysis. Wat. Resour. Res. 6(6), 1641-1648.

Topaloğlu F. (2005). Regional Flood Frequency Analysis of the Basins of the East Mediterranean Region. Turk. J. Agric. For., 29, 287-295.

U.S. Army Corps of Engineers (USACE). (1994). "Engineering and Design – Hydrologic Engineering Analysis Concepts for Cost-Shared Flood Damage Reduction Studies," Engineer Pamphlets EP 1110-2-10.

U.S. Water Resources Council, (1976), Guidelines For Determining Flood Flow Frequency: Bulletin 17 of the Hydrology Subcommittee of the U.S. Water Resources Council, Washington, DC.

Ward, J. H., Jr. (1963), Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association, 48, 236–244.

Wallis, J. R., E. F. Wood. (1985), Relative accuracy of log Pearson III procedures, J. Hydraul. Eng., 111(7), 1043–1056.

Wallis, J. R. (1980) Risk and uncertainties in the evaluation of flood events for the design of hydraulic structures. In Piene e Siccitd (eds E. Guggino, G. Rossi and E. Todini), 3-36.

Wallis, J.R. (1989). Regional frequency studies using L-Moments, Res. Report 14597, IBM Res. Div. T.J. Watson Res. Cent. Yorktown Heights, NY. 1–17.

Zhang J.Y., Hall M.J. (2004) Regional flood frequency analysis for the Gan-Ming River basin in China. J Hydrol 296:98–117.

Zrinji, Z., D.H. Burn. (1994). Flood frequency analysis for ungauged sites using a region of influence approach. Journal of Hydrology, 153: 1-21.