APPLICATION OF THE MAP CORRELATION METHOD TO THE WESTERN BLACKSEA BASIN

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

BY

MEHMET KAYRA ERGEN

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

JUNE 2012

Approval of the thesis:

APPLICATION OF THE MAP CORRELATION METHOD TO

THE WESTERN BLACKSEA BASIN

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

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

Prof. Dr. Güney Özcebe
Head of Department, **Civil Engineering** _______________

Assoc. Prof. Dr. Elçin Kentel
Supervisor, **Civil Engineering Dept., METU** _______________

Examining Committee Members

Prof. Dr. Melih Yanmaz _______________
Civil Engineering Dept., METU

Assoc. Prof. Dr. Elçin Kentel
Civil Engineering Dept., METU _______________

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

Assist. Prof. Dr. Şahnaz Tiğrek
Civil Engineering Dept., METU _______________

Assoc. Prof. Dr. Ayşegül Aksoy
Environmental Engineering Dept., METU _______________

Date: 1 June 2012

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 Kayra, ERGEN

Signature :

ABSTRACT

APPLICATION OF THE MAP CORRELATION METHOD TO THE WESTERN BLACKSEA BASIN

Ergen, Mehmet Kayra

M.Sc., Department of Civil Engineering

Supervisor: Assoc. Prof. Dr. Elçin Kentel

June 2012, 99 Pages

Turkey is a developing country and its energy demand is increasing due to its growing population and industry. As a result, to fulfill this growing energy demand, Turkey is currently developing its unused hydropower potential, especially through small hydroelectric power plants (SHPPs). Estimation of annual electricity generation of a small hydropower plant strongly depends on streamflow data. In Turkey, there are a limited number of streamgaging stations so the estimation of streamflow at a potential SHPP location requires transferring streamflow time series from a reference streamgaging station to the ungaged basin. In order to determine daily streamflow time series for ungaged catchments, typically the nearest streamgaging station is chosen as the reference streamgaging station. However the distance between a reference streamgaging station and an ungaged catchment may not always be the most appropriate reference streamgaging station selection criterion. Archfield and Vogel (2010) proposed a new method called the Map Correlation Method (MCM) to select a reference streamgaging station to donate its observations to an ungaged catchment. MCM aims to identify the most correlated streamgaging station with the ungaged catchment. This new method is used at the Western Blacksea Basin in Turkey to select the best among possible reference streamgaging stations. The method proved to be promising; the most correlated streamgaging station for approximately one third of the study streamgaging stations are identified correctly by the MCM.

Keywords: The Map Correlation Method, Hydropower, The Reference Streamgaging Stations

ÖZ

HARİTA KORELASYON METODUNUN BATI KARADENİZ HAVZASI'NA UYGULANMASI

Ergen, Mehmet Kayra

Yüksek Lisans, İnşaat Mühendisliği Bölümü

Tez Yöneticisi: Doç. Dr. Elçin Kentel

June 2012, 99 Sayfa

Türkiye gelişmekte olan bir ülkedir ve enerji tüketimi, gün geçtikçe artan nüfusu ve endüstrisine bağlı olarak artış göstermektedir. Artan enerji ihtiyacını karşılamak için, hidroelektrik potansiyelini, özellikle küçük hidroelektrik santraller kurarak (HES) geliştirmektedir. Küçük HES'lerde elektrik üretiminin tahmin edilmesi akım verilerine bağlıdır. Türkiye'de yeterli miktarda akım gözlem istasyonu bulunmamaktadır; bu yüzden küçük HES'lerin bulunacağı kesitte akım verilerini hesaplamak için genellikle en yakın akım gözlem istasyonu referans olarak kullanılmaktadır. Ancak iki nokta arasındaki uzaklığın referans akım gözlem istasyonu seçilirken baz alınması her zaman doğru bir tercih olmayabilir. Archfield ve Vogel (2010) referans akım gözlem istasyonunu belirlemek için Harita Korelasyon Metodu (HKM) adını verdikleri yeni bir metod önermişlerdir. HKM, akım gözlem istasyonu bulunmayan bir havza için çevredeki akım gözlem istasyonları arasından bu istasyon ile en iyi korelasyona sahip olanı referans istasyon olarak seçmeyi hedeflemektedir. Bu metod referans akım gözlem istasyonlarını seçmek için Türkiye'de Batı Karadeniz Havzası'nda denenmiştir. HKM ile çalışma havzasındaki akım gözlem istasyonlarının yaklaşık üçte biri için en iyi korelasyonu veren istasyon tespit edilebilmiştir ve bu da metodun umut verici olduğunu göstermiştir.

Anahtar Kelimeler: Harita Korelasyon Metodu, Sugücü, Referans Akım Gözlem İstasyonu

To My Mother and Father

ACKNOWLEDGEMENTS

First and foremost, the author wishes to express his sincere graditude to his supervisor Assoc. Dr. Elçin Kentel for her valuable guidance, counsel, understanding and encouragement throughout this study.

The author is also grateful to Dr. Stacey A. Archfield for her great contribution and suggestions.

The author also would like to thank his friends Umut Egemen Çam, Ali Ersin Dinçer, Barış Hacıkerimoğlu, Adil Baran Çobanoğlu, Çağatay Dursun Hanedan, Miran Dzabic, Ufuk Çalışkan, Yılgün Gürcan, Nilay Doğulu, Ertuğrul Gören and Ece Boyacıoğlu for their motivation and supports.

Finally, the author would like to thank profoundly all his family members especially his mother Defne Ergen and his father Besim Ergen for their endless love and support.

TABLE OF CONTENT

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

a : Range

A_g : Drainage Area of Gaged Catchment

A_u : Drainage Area of Ungaged Catchment

$Cov\{XY\}$: Covariance of Different Random Variables, X and Y

$E\{X\}$: Mean of Random Variable X

EIEI: Elektrik İşleri Etüd İdaresi

h : Distance

HPP: Hydropower Plant

m_r : Average Error of a Set of k estimates

MCM: Map Correlation Method

NSE: Nash-Sutcliffe Efficiency

p_i : Probability of Random Variable x

p_j : Probability of Random Variable y

p_{ij} : Joint Probability of Random Variables, x and y

$\tilde{p}_X(h)$: Correlogram

Q_g : Flow at the Gaged Catchment

Q_u : Flow at the Ungaged Catchment

r : Pearson's Correlation Coefficient

r_i : Error of i^{th} estimate

USGS: United States Geological Survey

$Var\{X\}$: Variance of Random Variable X

w_j : Weight of Sample j

$\hat{x}$: Estimated Value by Using Weighted Linear Combination

x_j : Available Sample Data

Y^{mean} : Average Value of Observed Data

Y_i^{est} : Estimated Data Value

Y_i^{obs} : Observed Data Value

$\tilde{\sigma}^2$: Variance

σ_R^2 : Error Variance of a Set of k Estimates

$\tilde{\gamma}_X(h)$: Model Variogram

CHAPTER 1

INTRODUCTION

1.1 Statement of the Problem

In growing economies, like Turkey's, water resources play major role on sustainable economic development. For the efficient utilization of water resources, streamflow series must be analyzed and understood thoroughly.

Daily time series are commonly used in the design of the hydraulic structures such as small hydropower plants (SHPPs). At gaged catchments daily streamflow series can be obtained easily, however it is not possible to obtain daily streamflow series for an ungaged catchment. In order to do that, a gaged catchment is taken as a reference streamgaging station and its streamflow values are transferred to the ungaged catchment by using different methods. The most commonly used method is the drainage area ratio method (Korleski and Strickland, 2009):

$$Q_u = \frac{A_u}{A_g} \times Q_g \tag{1.1}$$

where Q_u is the flow at the ungaged catchment, Q_g is the flow at the gaged catchment, A_u is the drainage area of the ungaged catchment, and A_g is the drainage area of gaged catchment.

As can be seen from Equation 1.1 while transferring the streamflow series to an ungaged catchment, the flow rates of the reference streamgaging station are used. As a result, the flow characteristics at the ungaged catchment are directly affected by the reference streamgaging station. Therefore, selection of the reference streamgaging station is very important in estimating streamflow values at an ungaged location.

In practical applications, the reference streamgaging station is selected as the nearest streamgaging station to the ungaged location (Smakthin and Weregala 2005, Emerson et. al. 2005, Asquith et. al. 2006, Mohamoud and Parmar 2009, Esralew and Smith 2009, Patil and Stieglitz 2011). However distance between gaged and ungaged catchments may not be the only and most appropriate reference streamgaging station selection criterion. As an extension of this idea, Archfield and Vogel (2010) proposed a new method, called the Map Correlation Method (MCM), to select a reference streamgaging station to donate its observations to an ungaged catchment. MCM selects a reference streamgaging station by considering the correlations between gaged and ungaged catchments. They demonstrated the application of this method at a study area located across southern New England using 28 streamgaging stations. Archfield and Vogel (2010) concluded that the MCM generally provided improved estimates of daily streamflow time series over daily streamflow estimations from the nearest streamgaging station. In this study, the MCM is applied to the Western Blacksea Basin in Turkey for selecting reference streamgaging stations for ungaged catchments.

1.2 The Objective of the Study

The main objective of this study is to apply the MCM at the Western Blacksea Basin and evaluate the success of this method in selecting reference streamgaging stations in this region. In the study region, there are 13 streamgaging stations and their daily streamflow data is obtained from EIEI (Elektrik İşleri Etüd İdaresi) and used as input for the MCM.

In Turkey, the reference streamgaging station is traditionally selected as the nearest streamgaging station. In other words, the Euclidian distance is the main selection criteria for the selection of the reference streamgaging station. MCM utilizes the correlations between daily streamflows as the reference streamgaging station selection criteria. MCM methodology is summarized in Figure 1.1.

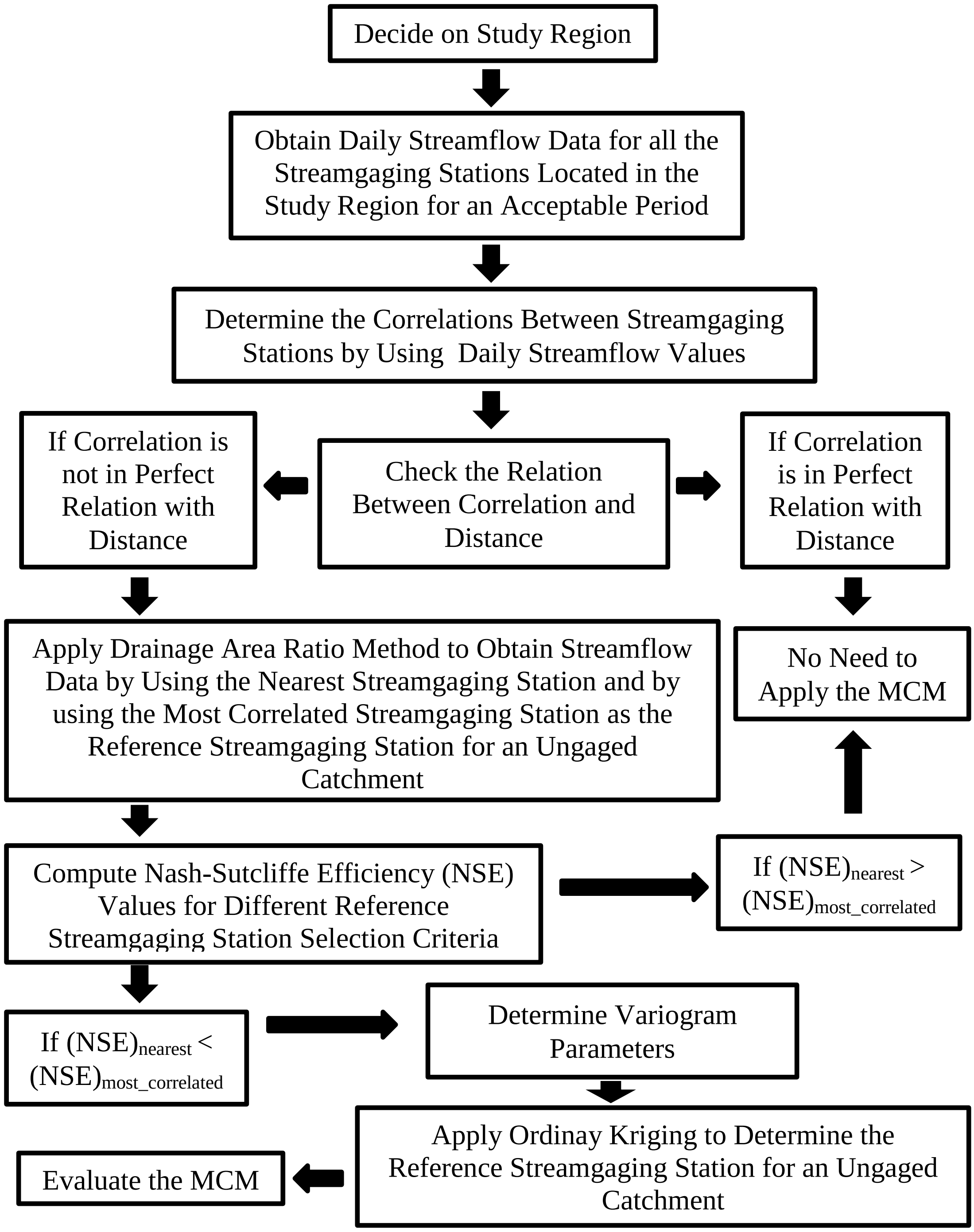

Figure 1.1 The MCM Methodology

1.3 Description of the Thesis

This thesis has five chapters. Chapter 1 is the Introduction and it highlights the current situation of hydropower development in Turkey and states the objectives of the study. Chapter 2 provides background information about selection criteria of the reference streamgaging station for an ungaged catchment and streamflow estimation techniques for ungaged catchments. In addition to that brief information about geostatistics is given in this chapter. The methodology of MCM is explained in Chapter 3. Chapter 4 covers the analyses and results of the application of MCM for the Western Blacksea Basin. Finally, the results are discussed in Chapter 5 and recommendations for future research are identified.

CHAPTER 2

LITERATURE REVIEW

2.1 Streamflow Estimation at Ungaged Sites

There are several methods to determine streamflow series at ungaged catchments. To apply these methods to an ungaged catchment, a reference streamgaging station needs to be determined and streamflow values of the gaged catchment need to be transferred. The accuracy of streamflow prediction at the ungaged catchment entirely depends on the selection of the reference streamgaging station. A number of example studies utilizing various streamflow estimation methods are given below.

For estimating streamflow for the Rappahannock watershed in Virginia, USA, a regionalized flow duration curve method was developed (Mohamoud and Parmar, 2009). In this method, streamflow time series were separated into magnitude and time sequence components. Then, the regionalized flow duration curve was used to estimate the magnitude component and the time sequence of the nearest gaged catchment was used to predict the time sequence component. Also in order to evaluate the performance of the regionalized flow duration curve method for the streamflow predictions at the ungaged catchments, estimated streamflows at the ungaged catchments and observed flows at nearby streamgaging stations were compared. In this study, while estimating streamflow series for an ungaged catchment, the nearest streamgaging station was used as the reference streamgaging station.

The regionalized flow duration curves were used in a different study conducted by Kocatepe (2011) to determine streamflow values in ungaged catchments. Three different approaches were applied to Oltu Basin in Turkey to develop regionalized flow duration curves in the study region and the performances of these methods are evaluated. The nearest or nearby streamgaging stations are used as reference streamgaging stations.

Another study was conducted for ungaged streams in Oklahoma, USA to estimate flow-duration and annual mean-flow characteristics (Esralew and Smith, 2009). The purpose of the study was to predict flow statistics for ungaged catchments by using regression equations. These flow statistics included annual, seasonal and monthly duration statistics. The preferred method to compute flow statistics at the ungaged catchment was the drainage area ratio method. The reference streamgaging station was taken as the nearest streamgaging station located on the same stream with the ungaged site. However, the method was considered as efficient if the drainage area ratio between the ungaged and gaged site was from 0.5 to 1.5.

In a different study, the application of drainage area ratio method in Texas, USA, was analyzed (Asquith W. H. et al., 2006). A statewide analysis of daily mean streamflow data is conducted to evaluate utilization of unity as the exponent of the area ratio and as a multiplicative bias correction. A total of 712 USGS streamgaging stations are used in this analysis. While using drainage area ratio method, one or more nearby streamflow gaging stations were used to estimate streamflow data.

According to Patil and Stieglitz (2011), while estimating streamflow series for an ungaged catchment, the hydrological parameters of gaged catchment which was used as the reference were transferred to the ungaged catchment. Generally, distance between the reference and the receiver is used as a reliable criterion to select the reference catchment for an ungaged site. As a result, the nearby gaged catchment is preferred to as the reference streamgaging station to predict streamflow series for an ungaged catchment. However, utilizing the nearest streamgaging station as the reference gage and transferring hydrological parameters to the ungaged catchment is investigated for different climatic and geographic regions. They investigated the reliability of the selection of the reference streamgaging station as the nearby streamgaging station by computing Nash-Sutcliffe efficiency values for 756 catchments across the United States. The results indicated that in humid run-off regions, streamflow estimation efficiency was higher when compared with dry-evapotranspiration dominated regions. Moreover, it was observed that distance between the ungaged and the gaged catchment was not an accurate selection criterion for the reference streamgaging station.

The drainage area ratio method is one of the most preferred streamflow estimation techniques due to its easy application. However, the results obtained from this method are not always reliable; so many researches have been trying to improve the performance of this method. One such study was conducted at Red River of North Basin, North Dakota and Minnesota where drainage area ratio method was applied to estimate streamflow by using nearby gages for 27 streamgaging stations (Emerson D. G. et al., 2005). The accuracy and reliability of drainage area ratio method was tested and development of an improved method to predict streamflow was initiated. Multiple regression technique was used to observe if drainage area, main slope, and precipitation are the most effective variables in predicting streamflows for three different seasons (winter, spring and summer). As a conclusion, drainage area was identified as the most significant variable in the drainage area ratio method.

In the same region, Red River of the North Basin, an investigation was conducted to analyze the effects of the operation of Garrison Diversion Unit in North Dakota on water quality and quantity of Sheyenne River and the Red River of the North (Guenthner R. S. et al., 1990). To investigate the impacts of the diversion unit, monthly observed and predicted flow data were collected and a complete monthly flow record for January 1931 through December 1984 for each of the 29 sites is tried to be obtained. For four sites, flow records were available but for the remaining 25 sites flow records were not complete. To form full flow records for these 25 sites, drainage area ratio method is used and the nearby streamgaging stations are used as the reference streamgaging stations. Another similar study for identifying the effects of Garrison Diversion unit was completed in this area at 2002 by Emerson and Dressler (2002). This time, flow records through 1931 to 1999 was estimated. There were a total of 35 sites; four of these sites had full flow records while ten of them only had partial flow records. For the remaining 21 sites, flow records were constructed by using drainage area ratio method and maintenance of variance extension methods.

In another study conducted by USGS, the flow duration and peak discharge frequency values were estimated for Kansas Stream Locations by using 4771 individual locations on streams (Perry A. C. et al., 2004). To predict the mean flow and flow-duration values of 90, 75, 50, 25 and 10 percent for ungaged locations, equations were developed by using least-squares multiple regression techniques. While applying regression analysis, climatic factors and basin characteristics including the drainage area, the mean annual precipitation, the mean basin permeability and the mean basin slope were taken into account. After the equations were developed, regression equations and interpolation methods were used to obtain the flow duration curve, the mean flow and estimates of peak-discharge frequency at ungaged locations. Streamflow statistics were obtained for ungaged locations by using streamflow data from weighted gaged locations which were selected according to the drainage area and bias between ungaged and gaged locations.

Another study was conducted by Smakhtin and Weregala (2005) at Walawe River Basin in southern region of Sri Lanka. The purpose of this research was to determine streamflow series for ungaged catchments and examine the effects of land-use changes and water resources development over the basin for a period of 40 years. Also an acceptable flow regime which can be used for future developments of water resources was constructed. To do that, streamflow series at ungaged catchments were estimated by using non-linear spatial interpolation method in which one or more nearby gaged catchments are used as reference gages.

GLM Engineering (2009) stated that minimum streamflow information is commonly used to determine the water available for extraction and to analyze instream environmental parameters. This requires determination of minimum streamflows at ungaged locations, either on a stream having gages at other locations, or on a stream without any gages. In order to estimate minimum streamflows at the ungaged sites, regional regression equations were used. Three multiple regression models (i.e. the ordinary least squares, the manual numeric search for least square error, and the generalized least squares) were evaluated. For predicting streamflow at ungaged sites, the drainage area ratio method was not used; instead the drainage area ratio method was modified by regression equations. The reason for this modification was large differences in the magnitudes of rainfalls over short distances over Puerto Rico. Although the drainage area method was modified, no modification is used in selecting the reference gage and the nearest streamgaging station is used at the reference streamgaging station.

Monthly and annual streamflow values at ungaged catchments are necessary for the planning and management of water resources. Hortness and Berenbock (2001) conducted a study to predict streamflow at the ungaged locations for determining flows occurring at 80, 50, and 20 percent of the time each month in Idaho. Estimating equations which were found to be reliable for estimating high monthly streamflow statics when compared with low streamflow statistics were developed. While determining streamflow for ungaged sites, regression equations were used by taking streamflow and basin characteristics as variables. So the streamflow data from more than one reference gage were used to estimate streamflow at the ungaged catchment.

Estimation of streamflow data at ungaged basins is an important issue in Turkey as well. Development of the unused hydropower potential of Turkey has accelerated in the recent years. Many small and large hydropower plants (HPPs) are planned and feasibility studies for these HPPs are being conducted. In the feasibility studies, the drainage area ratio method is used to estimate streamflow values at ungaged catchments. While using drainage area ratio method, nearby gaged catchments are taken as the reference streamgaging station (Ak M. 2011, Aydın B. E. 2010, Mutlu R. 2010). However the selection of the nearest streamgaging station or nearby streamgaging stations as the reference streamgaging station is questionable and reliability of the utilization of only distance as the selection criteria for ungaged catchment should be investigated.

In 2010, Archfield and Vogel developed a new method named as the Map Correlation Method (MCM) to select the reference streamgaging station for an ungaged catchment. In this study, it is stated that distance alone is not a good selection criterion for identifying the reference streamgaging station. Thus, the MCM is developed for selecting the most representative reference streamgaging station based on geostatistics. This new method utilizes the correlations between the streamflows and the relations between these correlations and distances as the guiding parameters of the reference streamgaging station identification process.

2.2 Geostatistics

The concept of geostatistics is first discussed by Kolmogorov and Wiener (1950's) and then improved by Matern (1960), Whittle (1963), Matheron (1965) and others (Journel A. G., 1989). Geostatistics was first used for predicting ore reserves in South African gold mines. Exact analysis of geological phenomena such as size of an oil reservoir requires too much effort and this analysis was not economical. Although the geostatistical analysis is easy to perform, the results obtained from this analysis are uncertain (Olea R.A., 2009).

In 1950's Daniel Krige, a South African engineer, developed a technique for estimating ore grades by using spatial correlation in data. Later, Matheron studied this technique, added mathematical content in it and introduced a new concept called kriging and geostatistics (Gomez-Hernandez 2005). Matheron (1971) comments that geostatistics is used for estimation of mineral deposits by applying the theory of regionalized variables. The usage of geostatistics extended to outside of earth sciences due to positive results obtained from its application. At the beginning, geostatistics aimed to estimate parameters at unsampled regions such as estimation of ore reserves in an area or total rainfall in a basin. Later the usage of geostatistics is expanded to stochastic simulation where the purpose is to create variable fields that demonstrate spatial variability patterns close to natural ones (Gomez-Hernandez 2005).

Journel (1989) states that geostatistics is mainly data analysis and spatial continuity modeling. Without understanding the origin of data, one cannot conduct such analysis and modeling. The aim to model spatial continuity is to investigate spatial uncertainty. Treatment of spatial uncertainty in natural processes is a major area of research.

Today geostatistical methods are used in hydrology, petroleum geology, mining operations, meteorology, and environmental engineering. One of the examples to geostatistical model for an ore reserve is given in Figure 2.1. According to Noppé (1994), one of the geological phenomena in which geostatistics can be applied is detecting the coal reserves. The sample data obtained by drilling can be used to create spatial models and these models provide information about reserve tonnage and quality on a variety of scales. In order to do that, semivariogram analysis can be conducted by using calorific value and seam thickness as data and geostatistical models are developed utilizing these semivariograms.

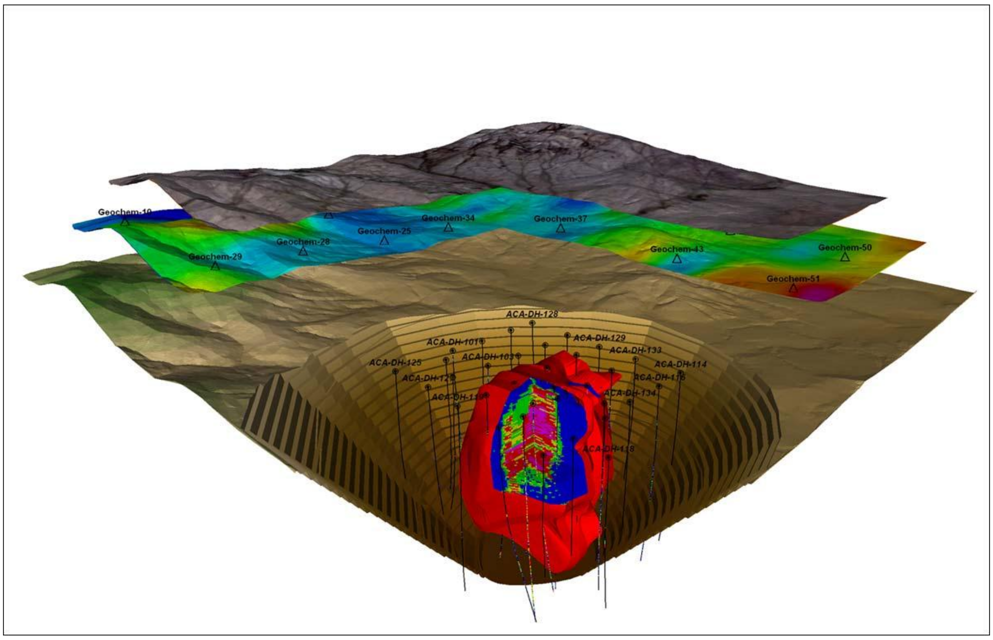

Figure 2.1 An Example of Geostatistical Model of an Ore Reserve (http://www.acahowe.co.uk, 2011)

The main purpose of geostatistics is to examine and analyze spatial systems which are incompletely known by using data from sampling locations. If there is a weak spatial correlation, there is no need to use geostatistics; however, this is usually not the case and there is significant spatial correlation between variables of natural processes. There are mainly two geostatistical tools to analyze spatial attributes which are probabilistic models and pattern recognition techniques (Olea R.A., 2009). MCM utilizes probability theory tools to identify spatial correlations between streamflow data.

Yaprak and Arslan (2008) indicates that geostatistics is a practical branch of statistics. It was firstly used to overcome problems encountered in geosciences. By using geostatistical analysis, unbiased and minimum variance results can be obtained from the phenomena in geosciences. To conduct geostatistical analysis, experimental variograms are designated and a model is fitted to designed variogram. Archfield and Vogel (2010) used this idea to develop variogram models to quantify the relation between the distances between pairs of possible reference streamgaging stations and the differences in the correlation coefficients of the streamgaging station pairs.

2.3 Variograms

Variogram which is a vector function is a simple tool in geostatistics to specify spatial continuity. “Structural” distance and relations of variables can be measured by variograms. If the variogram value is large, the correlation between variables is small (Gomez-Hernandez 2005).

According to Yaprak and Arslan (2008), a variogram shows the spatial relation of variables between each other and variogram analysis is conducted to investigate the change in the value of variables by considering positional dependence. To analyze the variogram, a model should be fitted to it. Detailed explanations about utilization of variograms in MCM are provided in Section 3.1.5.

CHAPTER 3

METHODOLOGY

3.1 Geostatistics

Phenomena of the environment can be measured at only a finite number of places. Especially for large scale applications, it is not economical to measure required properties everywhere. Instead, required properties can be estimated or predicted in a spatial sense. By using geostatistics, these predictions can be done without bias and with minimum error (Webster and Oliver 2007). These estimations require a model of how the phenomenon behaves at locations where it is not sampled. Without such a model, one cannot make estimations for an unknown location (Isaaks and Srivastava 1989).

There are two models which are used to make predictions at unsampled locations. First one is the deterministic model which can be applied when the context of data values is well understood. This means that the behavior of data can be explained by using laws of physics, economics, mathematics, etc. So, the change in the value of data can be determined by using appropriate governing equations. However, few earth science applications are understood in such a detail to allow the utilization of the deterministic approach for estimation of unknown variables. There is a lot of uncertainty about the nature and the behavior of processes at unsampled locations. As a result, in geostatistics, deterministic approach is not preferred and in order to make estimations at unknown locations, a probabilistic model is required.

In the probabilistic approach, the sample data are assumed to be produced as a result of some random process. For example, processes that create an ore deposit or a petroleum reservoir or decomposition and transportation of toxic compounds at a hazardous waste site are definitely complicated and our understanding of these processes may be so poor that the associated activities may appear as random behavior to us. In reality, these processes do not need to be random; it simply means that we are not capable of fully understanding their complexity and as a result we consider these processes as random. In practice, it is possible to specify a random process that generated any sample data set. To estimate values of a phenomenon at unknown locations we need to define a hypothetical random process which might have generated the observed values. However, the complete definition of the random process is usually not necessary, only certain parameters of the random process are sufficient for geostatistical estimation procedures (Isaaks and Srivastava 1989).

In the environment, locations that are close to each other tend to have similar properties, whereas properties at locations apart from each other differ more on the average. So in geostatistics, the distance is an important criterion in making predictions. Also, geostatistical predictions rely on spatial models whereas classical prediction methods do not. If the spatial correlation is high enough, the variable estimations are more accurate and precise with respect to their original values. Spatial correlations between data are important, because geostatistics treats a set of spatial data as a sample from the realization of a random process. To understand the nature of geostatistical predictions, variables and the outcome of random processes are needed to be analyzed (Webster and Oliver 2007). Definitions of a number of statistical concepts are provided below with related equations and examples. These concepts are used in MCM, thus a brief background on these issues is given for the sake of completeness.

3.1.1 Random Variables

A random variable is a variable whose values are randomly generated by some probabilistic mechanism (Isaaks and Srivastava 1989). Since the environmental variables are numerous and cannot be determined precisely and are in continuous interaction with each other in complex ways, outcomes of these variables can be regarded as random. By considering a stochastic view, at each point in space there is not just one value for a property but a whole set of possible values. Therefore, the observed value at a point can be judged as random due to the continuous change in the environmental properties at that point or points (Webster and Oliver 2007).

Throughout this thesis the following notation is used: upper case letters, such as X, indicate random variables and lower case letters, such as x, indicate the outcome of random variables. The possible outcome of a random variable X, is shown as $\{x_{(1)},...,x_{(n)}\}$ where n is the total number of possible outcomes and the outcomes that are actually observed, are denoted by $x_1, x_2, x_3, \ldots, x_n$.

For instance, assume outcome of flipping a coin is represented by a random variable D. The two possible outcomes of D, each having equal probability, are $d_{(1)} = T$, representing tail and $d_{(2)} = H$, representing head. The parentheses around the subscript remind that these symbols refer to a set of n possible values, for this example two, that the random variable can take. For example, a coin is flipped 8 times and the following sequence (Figure 3.1) is observed;

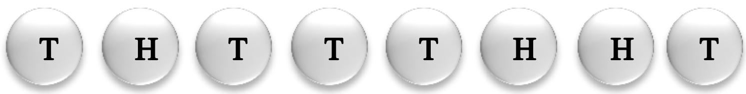

Figure 3.1 Resulting Sequence When a Coin Flipped 8 Times

The first observed outcome $d_1 = T$, the second is $d_2 = H$, the fourth and fifth one is $d_4 = T$ and $d_5 = T$, and the eight one $d_8 = T$. Without parentheses around the subscript, the lower case letters refer the outcomes that are actually observed through the repetition of the probabilistic mechanism.

Any random variable, X, can be defined by expressing its set of possible outcomes, $\{x_{(1)},...,x_{(n)}\}$, and the set of corresponding probabilities, $\{p_1,..., p_n\}$. The summation of probabilities $p_1,..., p_n$ is equal to 1.

3.1.2 Parameters of a Random Variable

The set of outcomes and their corresponding probabilities is sometimes represented as the *probability law* or *probability distribution* of a random variable. If these are well known, many parameters which are defining different features of random variables can be determined (Isaaks and Srivastava 1989). The parameters of a random variable cannot be calculated exactly from a few observations of the random variable but a set of sample statistics can be calculated from these observations. It is true that a different set of sample statistics will be produced from each different set of observations. However, as the size of the observations increases, the sample statistics tend to become more similar to the parameters of the conceptual model which is assumed to characterize the random variable.

The parameters of a conceptual model and sample statistics are different and they require different notation. A tilde (~) is used here to refer to a model parameter. For example, the mean of the model is represented by $\tilde{m}$ while m is used to represent the mean of the observed values.

The most commonly used random variable parameters in geostatistical estimation methods are the expected value or mean and the variance. The expected value (Equation 3.1) of a random variable is the mean or average outcome. It is the weighted average of the n possible outcomes:

$$E\{X\} = \tilde{m} = \sum_{i=1}^{n} p_i x_{(i)} \quad (3.1)$$

where, $\tilde{m}$ is the model mean, and n is the number of possible outcomes.

Variance measures the variability in the values of the random variable. It is defined as the squares of deviations of random variable values about the expected value $E\{X\}$ (Equation 3.2).

$$Var\{X\} = \tilde{\sigma}^2 = E\left\{\left[X - E\{X\}\right]^2\right\} \quad (3.2)$$

The variance of random variable is always positive. Larger value of variance indicates a greater spread in the distribution of random variable about the mean value (Hayter A. J., 2007).

3.1.3 Joint Random Variables

A probabilistic phenomenon can generate two random variables, X and Y. One of the most commonly used joint random variable parameter is the covariance (Isaaks and Srivastava 1989):

$$Cov\{XY\} = \tilde{C}_{XY} = E\{(X - E\{X\})(Y - E\{Y\})\}$$

$$Cov\{XY\} = \tilde{C}_{XY} = E\{XY\} - E\{X\}E\{Y\} \quad (3.3)$$

If the joint probability law is known, the covariance can be calculated by the following equation (Isaaks and Srivastava 1989):

$$Cov\{XY\} = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} x_{(i)} y_{(j)} - \sum_{i=1}^{n} p_i x_{(i)} \sum_{j=1}^{m} p_j y_{(j)} \quad (3.4)$$

where p_i and p_j are the corresponding probabilities of random variables x and y, and p_{ij} are the joint probability of x and y.

The other important parameter is the correlation coefficient. This parameter is an indicator of the strength of the linear relationship between two random variables. The correlation coefficient is defined as follows (Isaaks and Srivastava 1989):

$$\tilde{\rho}_{XY} = \frac{\tilde{C}_{XY}}{\sqrt{\tilde{\sigma}_X^2 \tilde{\sigma}_Y^2}} \quad (3.5)$$

where $\tilde{\sigma}_X^2$ and $\tilde{\sigma}_Y^2$ are the variances of X and Y. The value of the correlation coefficient varies between -1 and +1 depending on the relation between two different random variables. For instance, if the outcomes of two random variables are independent from each other, the value of the correlation coefficient becomes 0. If the observations of two random variables fall on a straight line having a positive slope then the correlation coefficient is +1.

3.1.4 Linear Combination of Random Variables

A probabilistic model may use a weighted linear combination of the available observations to estimate the unknown value at an unsampled location:

$$\hat{x} = \sum_{i=1}^{k} w_i x_i \tag{3.6}$$

where k is the number of observations, w_i and x_i are the weight and the observed value associated with observation i, respectively.

The expected value and the variance of a linear function are calculated as follows (Isaak and Srivastava 1989):

$$E\left\{\sum_{i=1}^{k} w_i X_i\right\} = \sum_{i=1}^{k} w_i E\left\{X_i\right\} \tag{3.7}$$

$$Var\left\{\sum_{i=1}^{k} w_i X_i\right\} = \sum_{i=1}^{k}\sum_{j=1}^{k} w_i w_j Cov\left\{X_i X_j\right\} \tag{3.8}$$

3.1.5. Random Functions

A random function is a set of random variables which have some spatial locations and their dependence to each other is specified with a probabilistic mechanism (Isaaks and Srivastava 1989). Similar to random variables, random functions may have several different outcomes as well.

For a stationary random function, the mean and the variance may be used to summarize the univariate behavior of the random function. To summarize the bivariate behavior of a stationary random function its covariance function, $\tilde{C}(h)$, its correlogram, $\tilde{\rho}(h)$, and its variogram, $\tilde{\gamma}(h)$ are used (Isaaks and Srivastava 1989).

The covariance function is the covariance between two random variables separated by a distance h and represented as follows:

$$\tilde{C}_X(h) = Cov\{X(i)X(i+h)\} \tag{3.9}$$

The correlogram is the correlation coefficient between two random variables that are separated by a separation distance h:

$$\tilde{p}_X(h) = \frac{Cov\{X(i)X(i+h)\}}{\sqrt{Var\{X(i)\}Var\{X(i+h)\}}} = \frac{\tilde{C}_X(h)}{\tilde{C}_X(0)} \tag{3.10}$$

Finally, the variogram is defined by the following equation:

$$\tilde{\gamma}_X(h) = \frac{1}{2}E\left\{\left[X(i) - X(i+h)\right]^2\right\} \tag{3.11}$$

The correlogram and the covariance has the same shape; with the difference being the maximum value of the correlogram is scaled to 1. Moreover the variogram function has the same shape as the covariance function except that it is upside down. The covariance starts from a maximum of $\tilde{\sigma}^2$ at $h = 0$ decreases to 0 whereareas variogram starts with 0 at $h = 0$ and increases to a maximum of $\tilde{\sigma}^2$ (Isaaks and Srivastava 1989). The two important parameters of the model variogram are the variance, $\tilde{\sigma}^2$ (Equation 3.2) and the range, a. The general equation of a spherical variogram model is given in Equation 3.12 and the spherical model is shown in Figure 3.2 (Kang L., 2012).

$$\gamma(h) = \begin{cases} \tilde{\sigma}^2\left(1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3\right) & 0 \le h < a \\ \tilde{\sigma}^2 & h \ge a \end{cases} \tag{3.12}$$

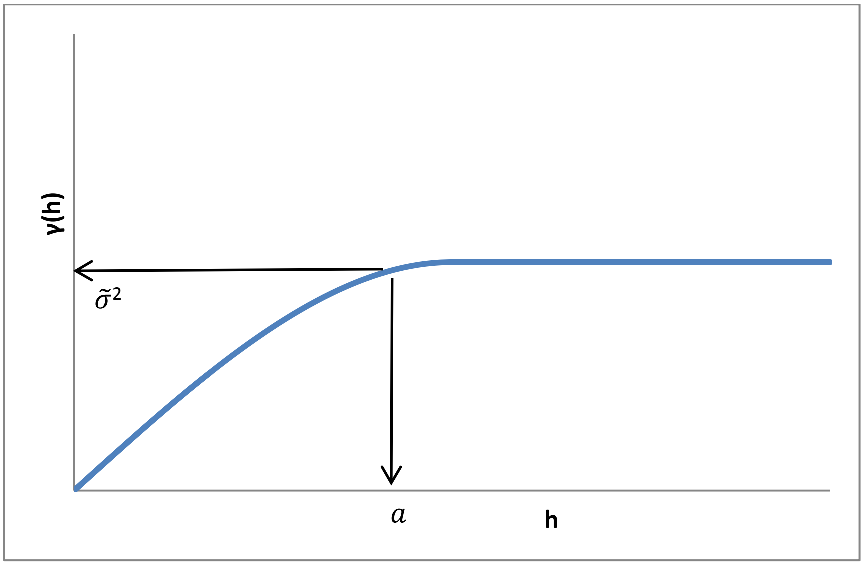

Figure 3.2 Spherical Variogram Model and Variogram Parameters

To clearly demonstrate the relation between a variogram and its corresponding correlogram and covariance, an example spherical model is considered:

$$\gamma(h) = \begin{cases} 0.0364\left(1.5\left(\dfrac{h}{85.86}\right) - 0.5\left(\dfrac{h}{85.86}\right)^3\right) & 0 \le h < 85.86 \\ 0.0364 & h \ge 85.86 \end{cases} \tag{3.13}$$

The variogram, the correlogram and the covariance of the spherical model given in Equation 3.13 are provided in Figures 3.3, 3.4, and 3.5, respectively.

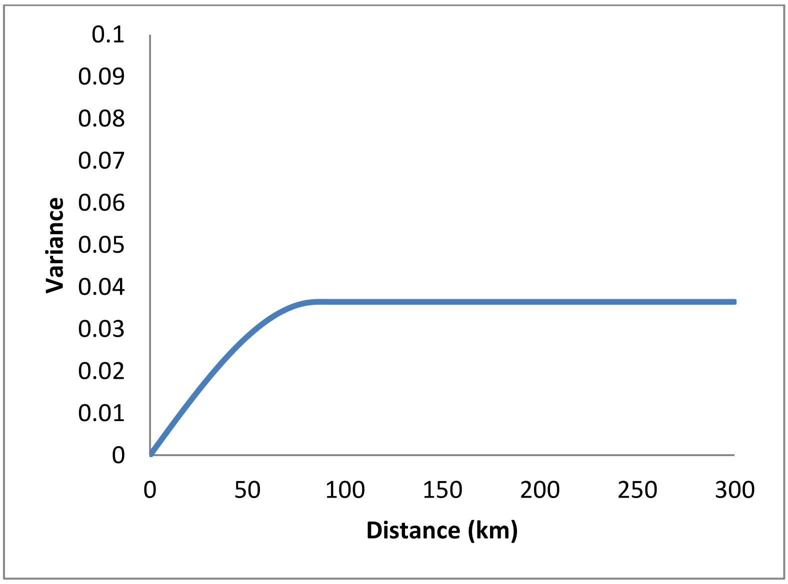

Figure 3.3 Variogram of the Random Function

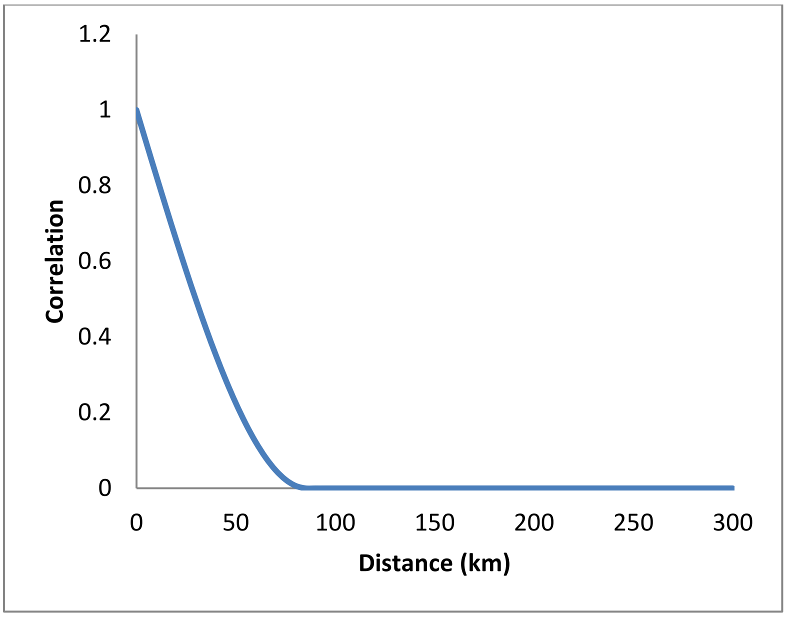

Figure 3.4 Correlogram of the Random Function

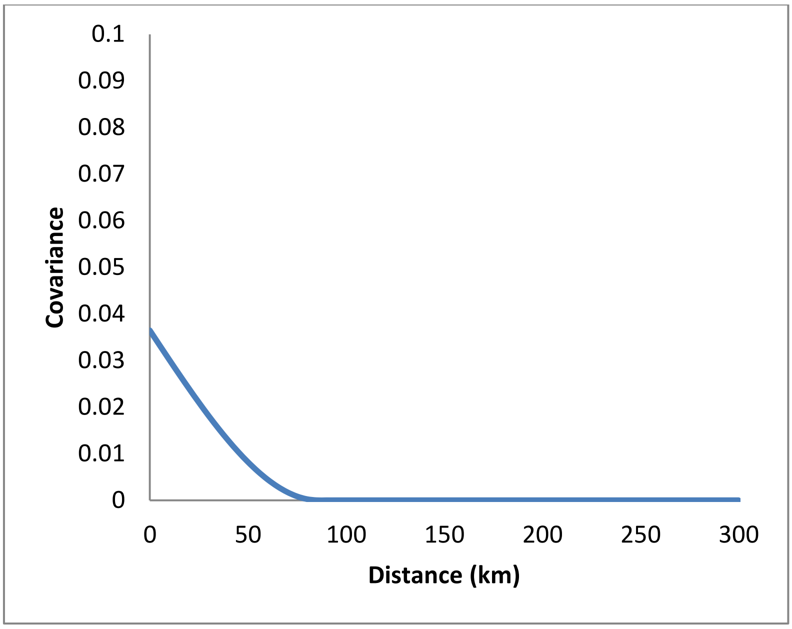

Figure 3.5 Covariance of Random the Function

3.2 The Map Correlation Method

Sub-catchments having similar hydrologic properties and climate characteristics are expected to produce similar responses, daily streamflow values in this case. Based on this fact, Archfield et al. (2009) hypothesized that correlation in daily streamflow between two catchments may be an improved metric of hydrologic similarity. Archfield and Vogel (2010) thinking that correlation between time series of daily streamflow integrates the complex physical processes that govern streamflow, proposed a geostatistical method, namely MCM, to identify the most correlated streamgaging station for an ungaged catchment. The most correlated streamgaging station can then be used as the reference streamgaging station to estimate streamflow values at the ungaged catchment.

While conducting the MCM, Pearson's r correlation coefficient and NSE values are used for testing the applicability of the method to the study region. NSE values are computed from the observed streamflow values and estimated streamflow values that are obtained by drainage area ratio method.

After testing the applicability of the MCM, ordinary kriging is applied to determine the most correlated streamgaging station for the ungaged catchment. In this case study, 12 correlation coefficients are estimated for an ungaged catchment by ordinary kriging and the streamgaging station, that has the highest correlation coefficient, is selected as the reference streamgaging station. The following sections give information about Pearson's r correlation coefficient, NSE and the geostatistical procedure of the MCM.

3.2.1 Pearson's r Correlation Coefficient

Pearson's correlation coefficient, r, is used to describe the degree of linear relation between observed and estimated data. In other words, it is the strength of the linear relation between two data sets. The range of r is from -1 to +1. When $r = 0$, it means there is no linear relationship between observed and estimated data values. If $r = 1$, the perfect positive linear relationship exists or if $r = -1$, the perfect negative linear relationship exists. Pearson’s correlation coefficient is defined as follows (Moriasi et al., 2007):

$$r_{yx} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - E\{X\})^2 \sum_{i=1}^{n}(Y_i - E\{Y\})^2}} \tag{3.14}$$

where X and Y are different variables, $E\{X\}$ and $E\{Y\}$ are the average of the data sets of X and Y, and n is the number of data points.

In practical applications, r is used extensively to evaluate models. According to Moriasi (2007), Pearson’s r correlation coefficient is highly sensitive to extreme values (outliers); thus, the logarithms of data values can be used in estimating the correlation coefficients.

3.2.2 Nash-Sutcliffe Efficiency

The Nash-Sutcliffe Efficiency (NSE) is a normalized statistical parameter and determines the relative magnitude of the unexplained variance compared to the observed data variance. NSE is defined as follows (Goswami et al. 2002):

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}\left(X_i^{obs} - X_i^{est}\right)^2}{\sum_{i=1}^{n}\left(X_i^{obs} - E\{X\}\right)^2} \right] \tag{3.15}$$

where X_i^{obs} is the i-th observed data value, X_i^{est} is the i-th estimated value and $E\{X\}$ is the average of the observed data.

The range of the value of NSE is from $-\infty$ to 1. While conducting MCM, values between 0 and 1 indicate acceptable levels of performance for the model. However the values that are smaller than 0, have unacceptable levels of performance (Moriasi et al., 2007). In that case, the average of observed values is accepted as a better predictor than the estimated value.

3.2.3 Ordinary Kriging

Kriging is a geostatistical method that is used for estimation of phenomenon at unsampled locations. An interpolation mechanism is used to combine weighed observed values around the unsampled location. The weights are determined according to the existing spatial covariance structure. Ordinary kriging is one of the kriging types that use weighted *linear* combinations of the available data for estimation (Bohling, 2005). According to Isaaks and Srivastava (1989), ordinary kriging is

1. "unbiased" because the mean residual or error, m_R equals 0 and

2. the "best" estimator because it tries to minimize the variance of the errors, σ_R^2.

These two conditions are unattainable since the values of m_R and σ_R^2 are unknown since the true value is not known. The value of m_R is not known so it is impossible to know if it is equal to zero or not. Similarly, the value of σ_R^2 is not known; so it cannot be minimized. In ordinary kriging, a probability model is built from the available data and the average error and the error variance of this model are calculated and used. Then, the weights of the linear probability model is selected such that the average error of the model, $\tilde{m}_R$ becomes 0 and the error variance, $\tilde{\sigma}_R^2$, is minimized (Isaaks and Srivastava 1989).

3.2.3.1 Unbiasednees Condition

At every location where there is no observation, the unknown true value is estimated by using a weighted linear combination of available samples:

$$\hat{x} = \sum_{j=1}^{n} w_j x_j \tag{3.16}$$

where, $\hat{x}$ is the estimated value by using weighted linear combination, w_j is the weight of the sample j, and x_j is the available sample data. The set of weights is free to change while making estimations at different locations (Isaaks and Srivastava 1989).

The error is defined as the difference between the estimated value and the true value at a given location by the following equation:

$$r_i = \hat{x}_i - x_i \tag{3.17}$$

where r_i is the error of i-th estimate.

If the error is defined like that, the average error of a set of k estimates given as follows:

$$m_r = \frac{1}{k}\sum_{i=1}^{k} r_i = \frac{1}{k}\sum_{i=1}^{k} \hat{x}_i - x_i \tag{3.18}$$

where m_r is the average error. However, m_r cannot be calculated since the true values, $x_1, x_2, \ldots, x_k$ are not known. Assuming the unknown values are outcomes of a random process, a conceptual model may be developed and used to ensure that the expected value of error at any particular location is zero (i.e. unbiasedness condition). This requirement is satisfied when the sum of the weights equals one:

$$\sum_{i=1}^{n} w_i = 1 \tag{3.19}$$

3.2.3.2 Minimization of Error Variance Condition

The error variance, σ_R^2, of a set of *k* estimates can be written as follows:

$$\sigma_R^2 = \frac{1}{k}\sum_{i=1}^{k}\left[r_i - m_r\right]^2 \tag{3.20}$$

As a can be seen from Equations 3.16, 3.17 and 3.20 the error variance is a function of the weights. To minimize the error variance, its derivative with respect to each weight has to be taken and equated to zero. As a result of this procedure the following system of linear equations is obtained:

$$\begin{pmatrix} \tilde{C}_{11} & \cdots & \tilde{C}_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{C}_{n1} & \cdots & \tilde{C}_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{pmatrix} = \begin{pmatrix} \tilde{C}_{10} \\ \vdots \\ \tilde{C}_{n0} \\ 1 \end{pmatrix} \tag{3.21}$$

where, $\tilde{C}_{ij}$ is the model covariance and μ is the Lagrange parameter. The indices i and j represent the streamgaging stations. The details of these derivations can be found in Isaaks and Srivastava (1989).

The relationship between model variogram and model covariance can be given as follows (Niu and Tarumi, 2009):

$$\gamma_{ij} = \tilde{\sigma}^2 - C_{ij} \tag{3.22}$$

where, $\tilde{\sigma}^2$ is the variance and γ_{ij} is the model variogram. Ungaged catchment is represented by $j = 0$.

3.2.4 Procedure of Geostatistical Analysis for MCM

In MCM, spatial prediction of variables is conducted mainly in four steps:

1. First it is assumed that one of the study streamgaging stations is an ungaged catchment and its correlation data is subtracted from the whole set of correlations. For example, among 13 study streamgaging stations located in the Western Blacksea Basin, one of the streamgaging stations is assumed to be an ungaged catchment. The available sample data (i.e. correlations between streamgaging stations which are determined by using daily streamflow data) of the remaining 12 streamgaging stations are used for the spatial prediction of correlations between ungaged catchment and the other streamgaging stations.
2. Calculating the sample variogram of streamgaging stations by using available sample data is the second step. If the example given in the first step is considered, sample variograms for the remaining 12 streamgaging stations are generated and plotted in this step.
3. Fitting the variogram models to the sample variograms is the third step. In MCM the preferred variogram model is the spherical model. For each of the 12 sample variograms, different spherical variogram models are fitted and variogram model parameters (range and variance) are determined from these fitted models.
4. Estimating correlations between streamflow values of the ungaged catchment and observed streamflows at each of the study streamgaging stations by using ordinary kriging is the fourth step. Thus, at the end of this step, 12 correlation coefficients are obtained. The streamgaging station having the highest correlation coefficient is identified as the reference streamgaging station.

CHAPTER 4

CASE STUDY

4.1 Selecting the Study Region

The main goal of this study is to determine applicability and usability of the map correlation method for catchments in Turkey. Thus, the first step is the selection of a study region. The most important step in selecting a study region is the identification of an area where human interference is minimal. Existence of hydraulic structures such as diversion weirs and dams disturb natural flow conditions in the sub-catchments. To identify an appropriate region for the application of the map correlation method, professional advice from İsmail Uçar, who is a chief engineer from EIEI, was taken (personal communication with Uçar, İ. on 14 May 2011). According to Uçar, it is difficult to find an undisturbed region in Turkey. However, northern parts of the country have relatively less interference compared to the western, eastern and southern parts. The reason for this is the recent development of water resources at the northern regions when compared to other regions. Development in the northern part of Turkey started later relative to the other regions, thus it is possible to obtain close to natural stream flow data for relatively long periods of time.

Second important criterion in selecting a study area is the availability of long time series of streamflow data. The MCM requires utilization of daily streamflow data from a large number of streamgaging stations located inside the study area. Moreover, observation periods of these streamgaging stations must be the same. In Turkey, many streamgaging stations have gaps in their streamflow data. These types of gaps in streamflow data cannot be tolerated while calculating correlations between streamflow values from two gaging stations. To summarize, in selecting a study region, the streamflow data availability of the streamgaging stations needs to be taken into account as well.

In this study, it is aimed to evaluate the performance of the MCM for a catchment in Turkey. Thus, a comparative evaluation of the results of this study with those presented in the original article (Archfield and Vogel, 2010) is seeked. The comparisons will be more realistic and meaningful if the areas of the study regions and the number of streamgaging stations within the study regions were similar in both studies. However, due to above stated restrictions it was not possible to select a study region that is fully compatible with the one used in the original study. Thus, comparison of the results and evaluation of the performance of MCM should be assessed carefully.

Considering all these factors, the Western Blacksea Basin (Figure 4.1) is selected as the study region. Streamflow recordings of all the streamgaging stations are determined (see Table 4.1) and a total of 13 streamgaging stations with 10 years of common streamflow data are identified. The streamflow data used for this study belongs to the period between 1995 and 2004. The elevation of the streamgaging stations above sea level can be found in Table 4.1 as well. The selected EIEI streamgaging stations are 1302, 1307, 1314, 1319, 1327, 1330, 1332, 1334, 1335, 1338, 1339, 1340, and 1343. Catchment locations, drainage areas, coordinates, and observation periods associated with each one of these 13 streamgaging stations are given in Table 4.1.

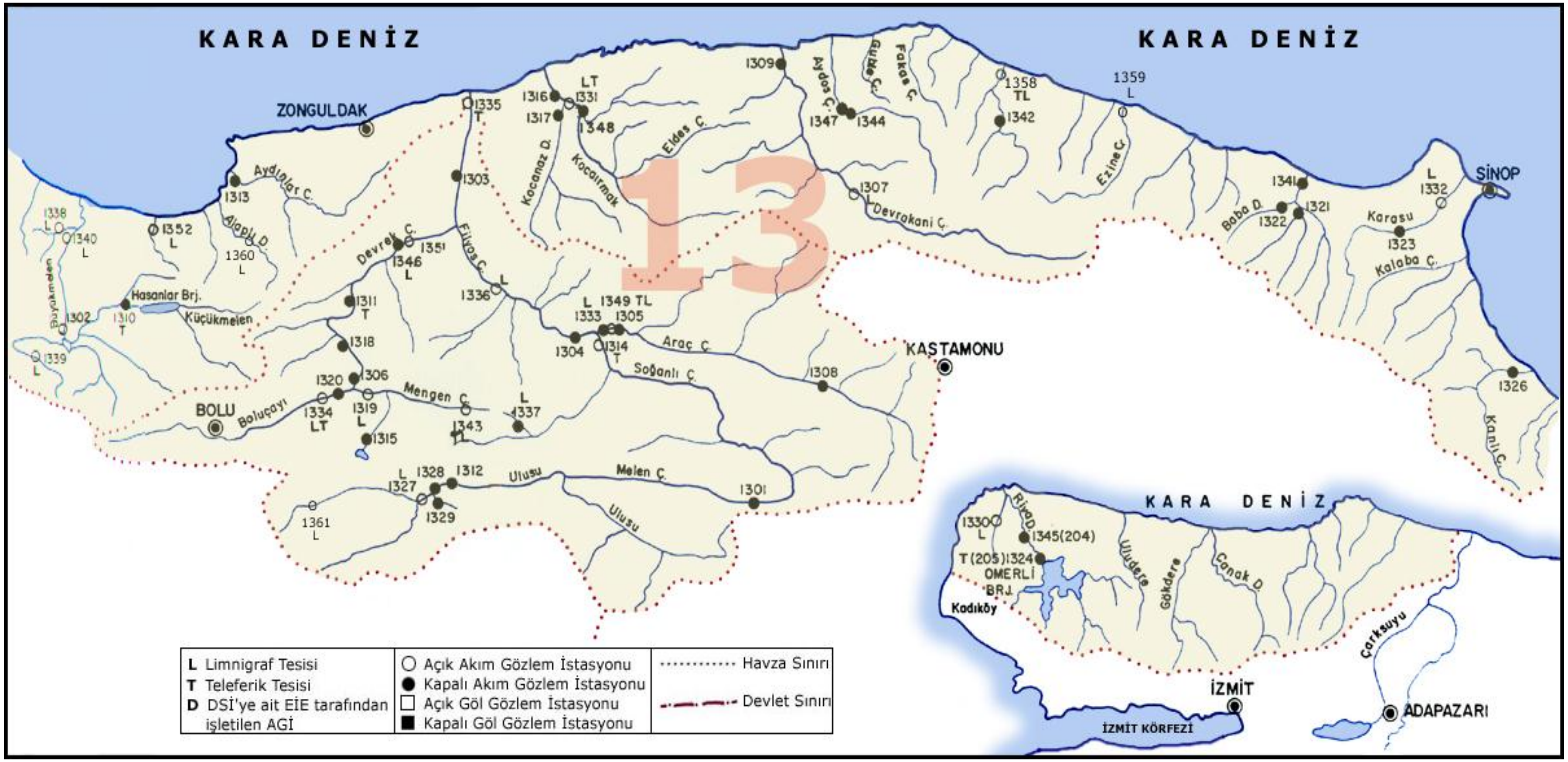

Figure 4.1 The Western Blacksea Basin
(www.eie.gov.tr)

Table 4.1 Brief Information About Study Streamgaging Stations in the Western Blacksea Basin

Streamgage Number	Catchment-Location (Elevation(m))	Drainage Area (km^2)	Coordinates	Observation Period (Years)
1302	Büyükmelen – Yakabaşı (115)	1988.00	$30^0$59'08" D- $40^0$51'22" K	1952-2012 (No record at 1963, 1971, 1972, 1992, 2008)
1307	Devrekani Çayı – Azdavay (815)	1097.60	$33^0$17'42" D- $41^0$38'19" K	1953-2012 (No record at 1955, 2005)
1314	Soğanlı Çayı – Karabük (271)	5086.80	$32^0$38'32" D- $41^0$10'11" K	1962-2012
1319	Mengen Çayı – Gökçesu (507)	766.40	$31^0$58'04" D- $40^0$53'44" K	1964-2012 (No record at 1981, 1998, 2008)
1327	Ulusu – Afatlar (1142)	953.60	$32^0$15'01" D- $40^0$44'30" K	1966-2012
1330	Yeniçiftlik Deresi - M. Şevket Paşa (39)	23.10	$29^0$11'25" D- $41^0$09'08" K	1966-2012 (No record at 1990, 1991)
1332	Karasu - Hacılar Köp. (20)	340.00	$35^0$01'57" D- $41^0$59'57" K	1968-2012
1334	Bolu Çayı – Beşdeğirmenler (541)	1102.80	$31^0$55'47" D- $40^0$53'11" K	1966-2012 (No record at 1994)
1335	Filyos Çayı – Derecikviran (2)	13300.40	$32^0$04'44" D- $41^0$32'49" K	1963-2012
1338	Lahana Deresi – Ortaköy (16)	104.80	$30^0$56'14" D- $40^0$59'58" K	1979-2012 (No record at 1980, 1985, 1986)
1339	Aksu Deresi – Çiftekese (634)	105.20	$30^0$55'13" D- $40^0$42'55" K	1980-2012
1340	Büyük Melen – Beyler (23)	2174.00	$30^0$57'20" D- $40^0$58'58" K	1980-2012 (No record at 2006)
1343	Korubaşı Deresi – Arak (780)	125.00	$32^0$15'38" D- $40^0$56'32" K	1991-2012

4.2 Outline of the Case Study

After selecting the study region, necessary calculations for the MCM is conducted. At the first step, Pearson's r correlation coefficients are computed between streamflow measurements of each pair of streamgaging station within the study region to observe the relation between distance and correlation. If the nearest streamgaging station to an assumed ungaged location is the most correlated streamgaging station for most of the study streamgaging stations, then the nearest streamgaging station can be used as the reference streamgaging station. But if distance and streamflow is not linearly related with each other, correlations instead of distances can be used as the reference streamgaging station selection criterion.

To test the feasibility of the reference streamgaging station selection criteria for the study region, Nash-Sutcliffe Efficiency (NSE) values are calculated as well. The drainage area ratio method is used to transfer streamflow data between catchments. The NSE values of streamflow at the nearest and at the most correlated streamgaging stations to the study streamgaging station are compared. If an NSE value of the most correlated streamgaging station is higher than that of the nearest streamgaging station, the correlation can be considered as a feasible reference streamgaging station selection criterion for an ungaged catchment.

The next step is the geostatistical computations. The variogram parameters are obtained from the model variogram and ordinary kriging is applied for determining the reference streamgaging stations for ungaged catchments. At the end, the NSE values are calculated using the nearest, the most correlated, and the streamgaging station identified by the MCM and compared. Procedures followed and the details of the computations are given in the following sections.

4.3 Pearson's r Correlation Coefficient Calculations

First, how the correlation between streamflow values at the streamgaging stations change with the Euclidian distance between the streamgaging stations is investigated. For this purpose, Pearson's r correlation coefficients are computed using the logarithms of the daily streamflow values between all possible pairs of 13 streamgaging stations. Pearson's r correlation coefficient measures the linear relationship between variables and it is highly sensitive to extreme changes (Moriasi, 2007). Thus, the logarithms of daily streamflow values are used in the calculation of correlation coefficients. Although the rest of the calculations are carried out with the logarithmic values of streamflow series, parallel calculations with arithmetic/real values of streamflow series are conducted as well to observe the differences that may occur due to linearization of the streamflow. Calculations carried out with real values of streamflow series are provided in Appendix A.

If there exists a perfect relation between distance and correlation then selection of the nearest streamgaging station as the reference streamgaging station for an ungaged catchment is reasonable. On the other hand, if the distance and correlation do not have a perfect relation then other approaches for identification of a reference streamgaging station might be implemented.

The Pearson's r correlation coefficients for logarithms of streamflow data from pairs of streamgaging stations are calculated. For each streamgaging station, 12 different correlation coefficients are obtained using 3650 data points; data is plotted as a scatter graph and a linear line is fitted to these points. Square of correlation coefficient is automatically calculated by the Excel.

A stock plot showing the minimum, maximum and the range of the remaining correlations is given in Figure 4.2. Correlation coefficients for all pairs are provided in a matrix format in Table 4.2.

As can be seen from Figure 4.2, for 1302 and 1340 the minimum and maximum values are very different than range of the rest of the values. As can be seen from Figure 4.1, 1302 and 1340 are located on the same reach and this results in a very high correlation between daily streamflow values of these two streamgaging stations. For 1335 the maximum correlation is 0.61 which is quite low compared to those of which indicates that streamflow values at 1335 are not strongly correlated with streamflow values of any other streamgaging stations. This is unexpected interms of its geographically position. It is the most connected streamgaging station in the selected study region. A diagram showing connectivity pattern of 1335 with other streamgaging stations is given in Figure 4.3.

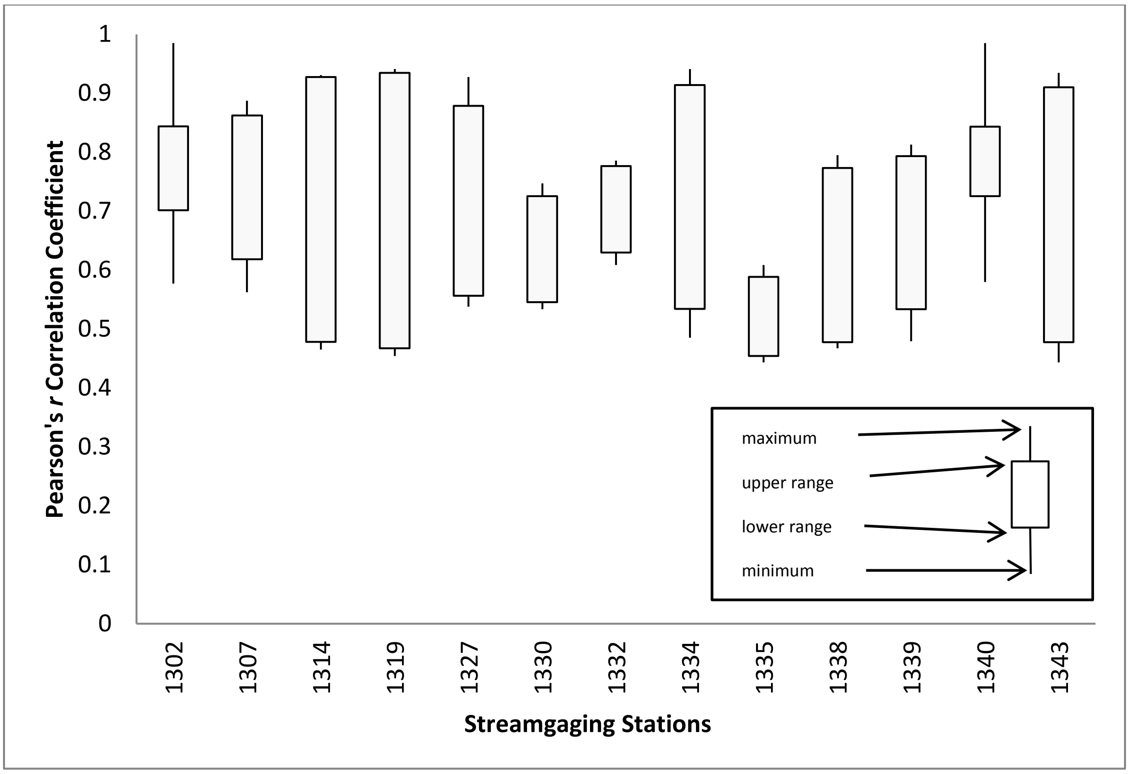

Figure 4.2 Stock Plot of Correlations for Streamgaging Stations at the Western Blacksea Basin

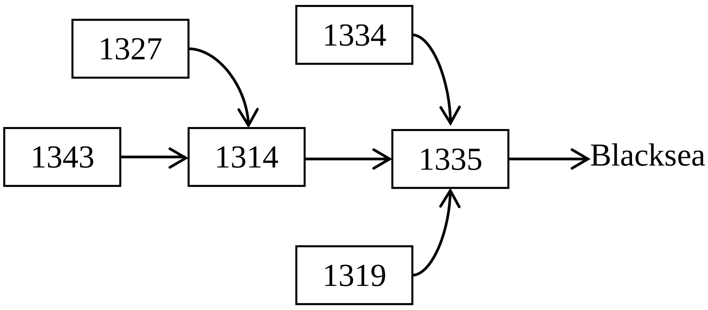

Figure 4.3 Connectivity Pattern of Streamgaging Station 1335

Table 4.2 Pearson's r Correlation Coefficients of 13 Streamgaging Stations Located in the Western Blacksea Basin

Streamgaging Stations	1302	1307	1314	1319	1327	1330	1332	1334	1335	1338	1339	1340	1343
1302	1.0000	0.8440	0.7876	0.8003	0.8155	0.7016	0.7764	0.8429	0.5769	0.7731	0.8130	0.9852	0.8004
1307	0.8440	1.0000	0.8520	0.8621	0.8529	0.6651	0.7660	0.8874	0.5621	0.6181	0.7321	0.8434	0.8535
1314	0.7876	0.8520	1.0000	0.9307	0.9276	0.5846	0.6481	0.9137	0.4649	0.4783	0.7571	0.7743	0.9058
1319	0.8003	0.8621	0.9307	1.0000	0.8784	0.5453	0.6298	0.9412	0.4540	0.4673	0.7331	0.7871	0.9345
1327	0.8155	0.8499	0.9276	0.8784	1.0000	0.6485	0.7100	0.8744	0.5381	0.5563	0.7877	0.8057	0.8687
1330	0.7016	0.6651	0.5846	0.5453	0.6485	1.0000	0.6981	0.5913	0.5883	0.7473	0.5335	0.7255	0.5485
1332	0.7764	0.7660	0.6481	0.6298	0.7100	0.6981	1.0000	0.6752	0.6087	0.6773	0.6570	0.7857	0.6382
1334	0.8429	0.8874	0.9137	0.9412	0.8744	0.5913	0.6752	1.0000	0.4854	0.5340	0.7671	0.8300	0.9101
1335	0.5769	0.5621	0.4649	0.4540	0.5381	0.5883	0.6087	0.4854	1.0000	0.5477	0.4791	0.5798	0.4436
1338	0.7731	0.6181	0.4783	0.4673	0.5563	0.7473	0.6773	0.5340	0.5477	1.0000	0.5699	0.7950	0.4778
1339	0.8130	0.7321	0.7571	0.7331	0.7877	0.5335	0.6570	0.7671	0.4791	0.5699	1.0000	0.7931	0.7124
1340	0.9852	0.8434	0.7743	0.7871	0.8057	0.7255	0.7857	0.8300	0.5798	0.7950	0.7931	1.0000	0.7825
1343	0.8004	0.8535	0.9058	0.9345	0.8687	0.5485	0.6382	0.9101	0.4436	0.4778	0.7124	0.7825	1.0000

As can be seen from Figure 4.3, catchments associated with streamgagin stations of 1327, 1343, 1334, and 1319 are sub-catchments of streamgaging station 1335. However, as can be seen from Table 4.1, the elevation of streamgaging station 1335 is 2 m while elevations of 1327, 1343, 1334, and 1319 are over 500 m. This factor may impact correlation coefficients for 1335.

4.4 Distance and Correlation Relations Between Streamgaging Stations

For determining the relation between distance and correlation; firstly, the Euclidian distances between the streamgaging stations need to be calculated. To do that, the coordinates of the streamgaging stations are obtained from EIEI and sketched on a map (Figure 4.4) in Google Earth software. The Euclidian distances between the streamgaging stations are computed approximately by using this software. The distance matrix is given in Table 4.3.

Figure 4.4 The Streamgaging Stations Plotted on Google Earth

A stock plot showing the minimum, maximum and the range of the distances is given in Figure 4.5. The average separation distance between streamgaging stations is more than 100 km, and it is high when compared with separation distances used in the original study.

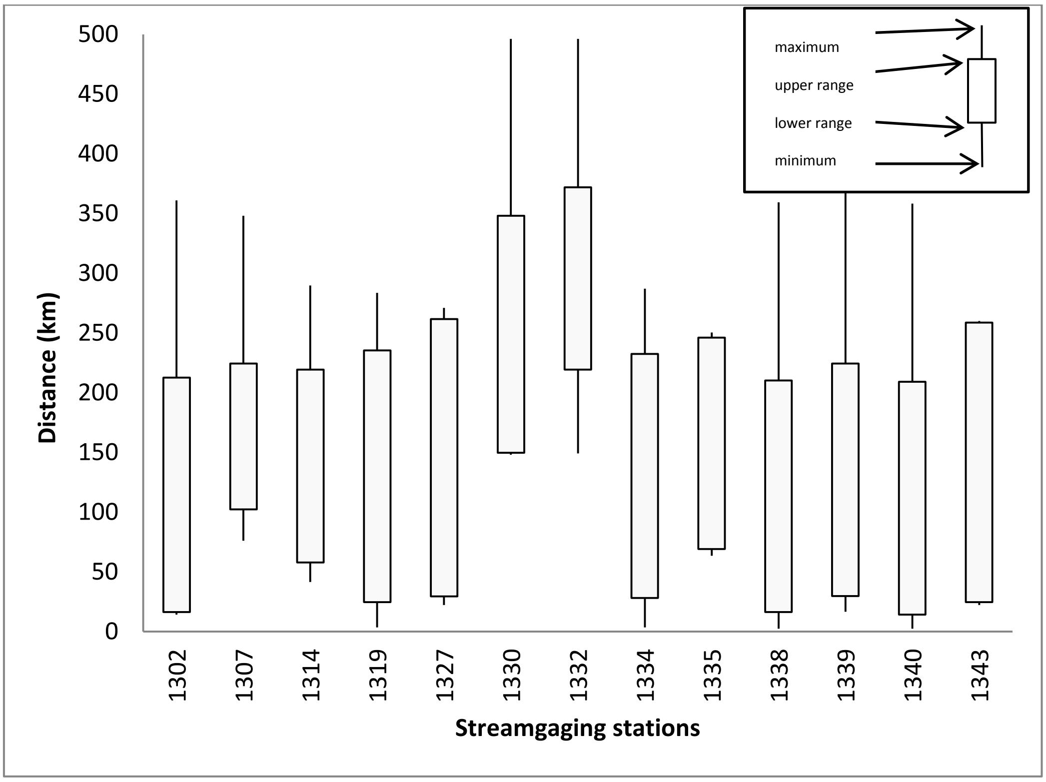

Figure 4.5 Stock Plot of Distances for Streamgaging Stations at the Western Blacksea Basin

Table 4.3 Distances (in km) between Streamgaging Stations

Streamgaging Stations	1302	1307	1314	1319	1327	1330	1332	1334	1335	1338	1339	1340	1343
1302	0.00	212.57	143.59	82.70	107.36	154.64	360.88	79.59	119.45	16.28	16.51	14.25	107.20
1307	212.57	0.00	75.87	138.80	133.05	348.07	149.06	141.99	102.26	210.25	224.32	209.17	117.08
1314	143.59	75.87	0.00	64.32	57.80	289.73	219.35	67.76	63.26	144.49	153.36	143.26	41.31
1319	82.70	138.80	64.32	0.00	29.37	235.37	283.56	3.45	73.14	87.67	90.39	85.51	24.68
1327	107.36	133.05	57.80	29.37	0.00	261.58	270.90	31.56	91.00	114.22	112.32	112.28	22.08
1330	154.64	348.07	289.73	235.37	261.58	0.00	496.13	232.36	246.07	147.97	153.57	149.62	258.70
1332	360.88	149.06	219.35	283.56	270.90	496.13	0.00	286.91	250.27	359.18	371.97	358.33	260.02
1334	79.59	141.99	67.76	3.45	31.56	232.36	286.91	0.00	74.43	84.39	87.19	82.69	28.00
1335	119.45	102.26	63.26	73.14	91.00	246.07	250.57	74.43	0.00	113.40	134.08	113.08	68.94
1338	16.28	210.25	144.49	87.67	114.22	147.97	359.18	84.39	113.40	0.00	31.35	2.26	110.92
1339	16.51	224.32	153.36	90.39	112.32	153.57	371.97	87.19	134.08	31.35	0.00	29.68	115.12
1340	14.25	209.17	143.26	85.51	112.28	149.62	358.33	82.69	113.08	2.26	29.68	0.00	109.43
1343	107.20	117.08	41.31	24.68	22.08	258.70	260.02	28.00	68.94	110.92	115.12	109.43	0.00

The relation between distance and correlation has significant importance because this relation is an indicator of the validity of the selection of the nearest streamgaging station as the reference streamgaging station. In order to observe the relation between distance and correlation, the correlations of a streamgaging station with the remaining streamgaging stations are plotted with respect to the distances between the pairs of streamgaging stations. As can be seen from Figure 4.6, approximately one third of the streamgaging stations show a strong relation between distance and r. For example, the most correlated streamgaging station is the nearest streamgaging station for 1319 and 1334, and correlation is decreasing with increasing distance. However, the majority of streamgaging stations have weak relation between distance and correlation. This can be due to big separation distances between streamgaging stations. On the top of it, streamgaging stations 1332 and 1335 have no considerable relation between correlation and distance. It can be concluded that the correlation between streamflow values is not strongly related with the separation distance in the study region. Thus, it is worth exploring appropriateness of other criteria for the identification of a reference streamgaging station, instead of the separation distance. The detailed graphical representations of the relations between correlation and distance of streamgaging stations at the Western Blacksea Basin are given in Appendix B.

The nearest and the most correlated streamgaging stations for each streamgaging station in the study region is identified and presented in Table 4.4. As can be seen from Table 4.4, for six of the streamgaging stations (i.e. 1302, 1319, 1330, 1334, 1338, and 1339) the nearest streamgaging station is the most correlated one. As can be seen from the comparison of Table 4.4 and Figure 4.1, most of these streamgaging stations are mostly correlated with streamgaging stations that are connected to them geographically. The bold marked streamgaging stations in Table 4.4 are the ones that are not the most correlated with the nearest streamgaging station but with some other streamgaging station. To sum up, for about half of the streamgaging stations in the study area, distance seems to be a good selection criterion as the reference streamgaging station; however for the remaining half distance alone cannot identify the best reference streamgaging station.

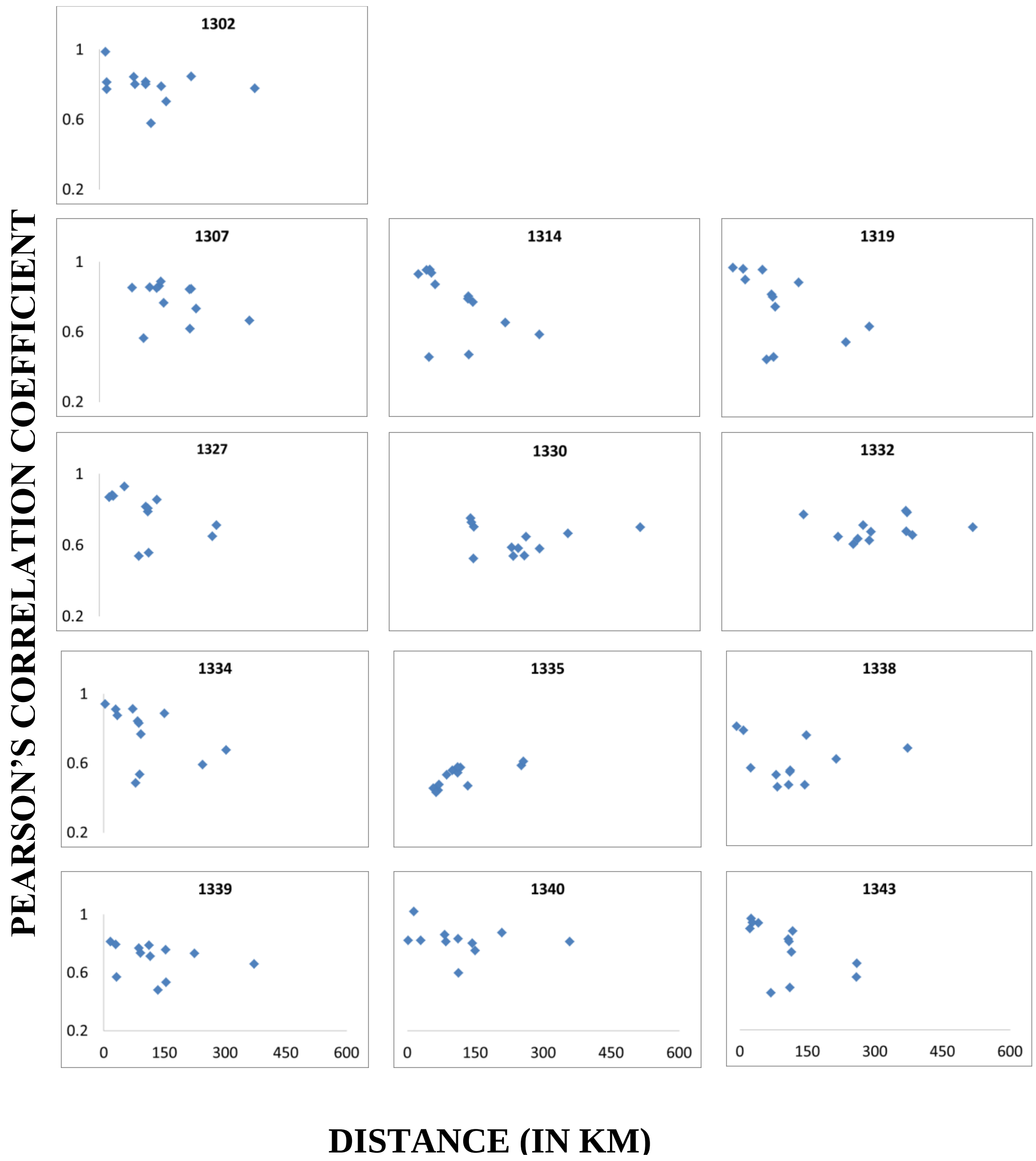

Figure 4.6 The Relation Between Pearson's r Correlation Coefficient and Distance for 13 Streamgaging Stations in the Western Blacksea Basin

Table 4.4 The Nearest and the Most Correlated Streamgaging Stations to 13 Streamgaging Station in the Western Blacksea Basin

Study SS*	The Nearest SS	Distance Between Study and the Nearest SS (km)	The Most Correlated SS	Correlation Coefficent Between Study SS and the Most Correlated SS
1302	1340	14.25	1340	0.99
1307	1314	75.87	**1334**	0.89
1314	1343	41.31	**1319**	0.93
1319	1334	3.45	1334	0.94
1327	1343	22.08	**1314**	0.93
1330	1338	147.97	1338	0.75
1332	1307	149.06	**1340**	0.79
1334	1319	3.45	1319	0.94
1335	1314	63.26	**1332**	0.61
1338	1340	2.26	1340	0.79
1339	1302	16.51	1302	0.81
1340	1338	2.26	**1302**	0.99
1343	1327	22.08	**1319**	0.93

* **Streamgaging Station**

As can be seen from Tables 4.3 and 4.4 for the seven streamgaging stations although the nearest streamgaging station was not identified as the most correlated streamgaging station, the most correlated streamgaging stations tend to be a streamgaging station that is close to the study streamgaging station. Thus, analysis of the relations between distances and correlation coefficients was not fully conclusive about the choice of the reference streamgaging station. To investigate the difference in the estimated daily streamflow values between utilization of the most correlated streamgaging station instead of the nearest streamgaging station as the reference streamgaging station, the following analysis is conducted. The drainage area ratio method is applied to each streamgaging station to estimate its daily streamflow values for two different cases: i) utilizing the nearest streamgaging station as the reference streamgaging station and ii) utilizing the most correlated streamgaging station as the reference streamgaging station. Results obtained from these two different cases are compared with the original observations to evaluate which one works with a better performance. The goodness of fit for two cases are evaluated by the Nash-Sutcliffe efficiency (NSE) value.

4.5 Nash-Sutcliffe Efficiency Calculations and Efficiency, Distance, Correlation Relations

To compare estimation performances of the nearest and the most correlated streamgaging stations as the reference streamgaging station the drainage area ratio method is used to estimate daily streamflow values for each of the 13 streamgaging stations. The procedure is as follows: The study streamgaging station is assumed to be an ungaged location, and its streamflow values are estimated by using the drainage area ratio method utilizing the nearest and the most correlated streamgaging station. This process is repeated for all 13 streamgaging stations located in the study region. Then the associated Nash-Sutcliff efficiency values are calculated. Calculated NSE values are given in Table 4.5.

As stated before, for six of the 13 streamgaging stations, the nearest and the most correlated streamgaging station were the same (see Table 4.4). Out of the remaining seven streamgaging stations, four of them (i.e. 1314, 1327, 1340, and 1343) generated higher NSE values with the most correlated streamgaging station as the reference streamgaging station. Three of the streamgaging stations (i.e. 1307, 1332, and 1335) generated higher NSE values with the nearest streamgaging station as the reference streamgaging station.

The average of NSE values computed from the most correlated streamgaging station is higher than the average of NSE values computed from the nearest streamgaging station (Table 4.5). The efficiency decreased in three of the streamgaging stations (i.e. 1307, 1332, 1335), however, it improved for four of the streamgaging stations (i.e. 1314, 1327, 1340, 1343) when the most correlated streamgaging station is taken as the reference streamgaging station. These results indicate that correlation might be a better selection criterion for the reference streamgaging station compared to the Euclidian distance between streamgaging stations. Although currently, distance is used as the selection criterion for the reference streamgaging station selection in Turkey, it might be possible to obtain better results by the utilization of the most correlated streamgaging station as the reference streamgaging station in estimating streamflow values at ungaged catchments. In the following sections validity of this idea is investigated.

Table 4.5 NSE Values That Are Calculated from the Nearest Streamgaging Station and the Most Correlated Streamgaging Station for 13 Streamgaging Stations in the Western Blacksea Basin

Streamgaging Station	The Nearest Streamgaging Station Taken as the Reference	Highest Correlated Streamgaging Station Taken as the Reference
1302	0.9502	0.9502
1307	**0.8091**	0.5790
1314	0.3190	**0.7121**
1319	0.8815	0.8815
1327	0.5813	**0.8715**
1330	0.5353	0.5353
1332	**0.6221**	0.2818
1334	0.8094	0.8094
1335	**0.5378**	-0.0630
1338	0.6577	0.6577
1339	0.3917	0.3917
1340	0.1981	**0.9511**
1343	-0.0650	**0.8002**
Mean of Nash-Sutcliffe Efficiency Values	0.5560	0.6430

Graphical representation of the comparison of NSE values based on the criteria (i.e. the distance and the correlation) used to identify the reference streamgaging station is shown in Figure 4.7. The relations between NSE and correlation (Figure 4.8) and NSE and distance (Figure 4.9) are investigated as well. As can be seen from Figure 4.8, as r increases the NSE increases, indicating a positive correlation between these two variables. Higher NSE values resulted when the reference streamgaging station had a correlation coefficient greater than 0.9. The relation between the distance and NSE is not as clear; however visual observation indicates there is a decreasing trend in NSE with increasing distance.

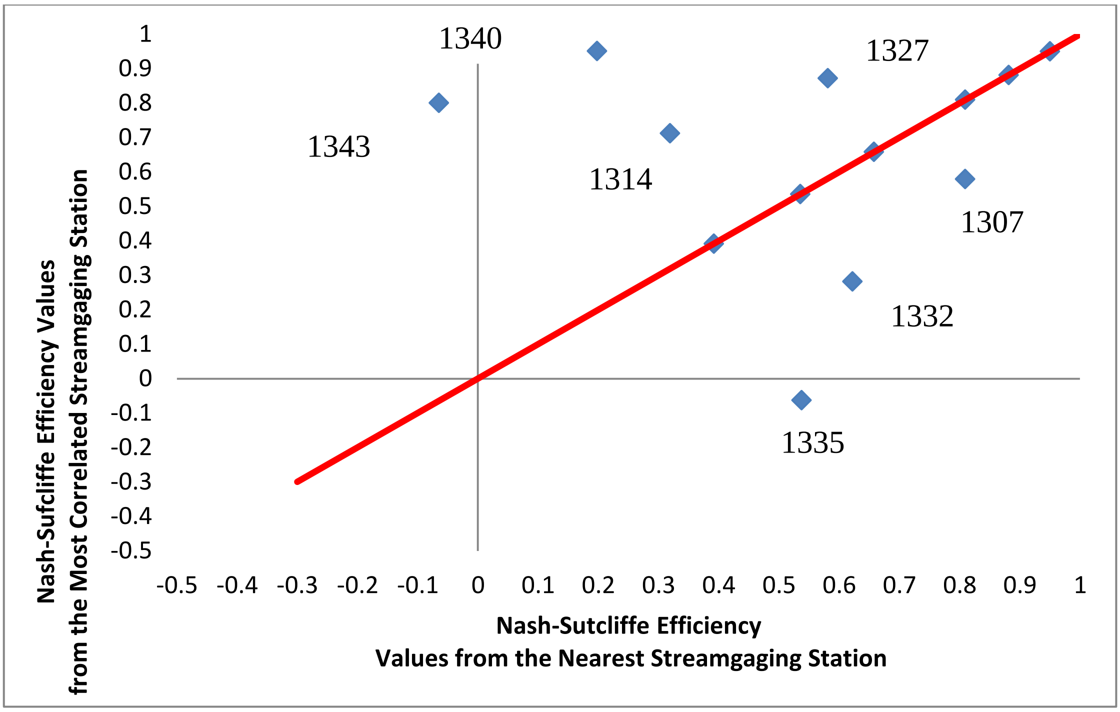

Figure 4.7 Comparison of NSE Values Based on the Reference Streamgaging Station Selection Criteria

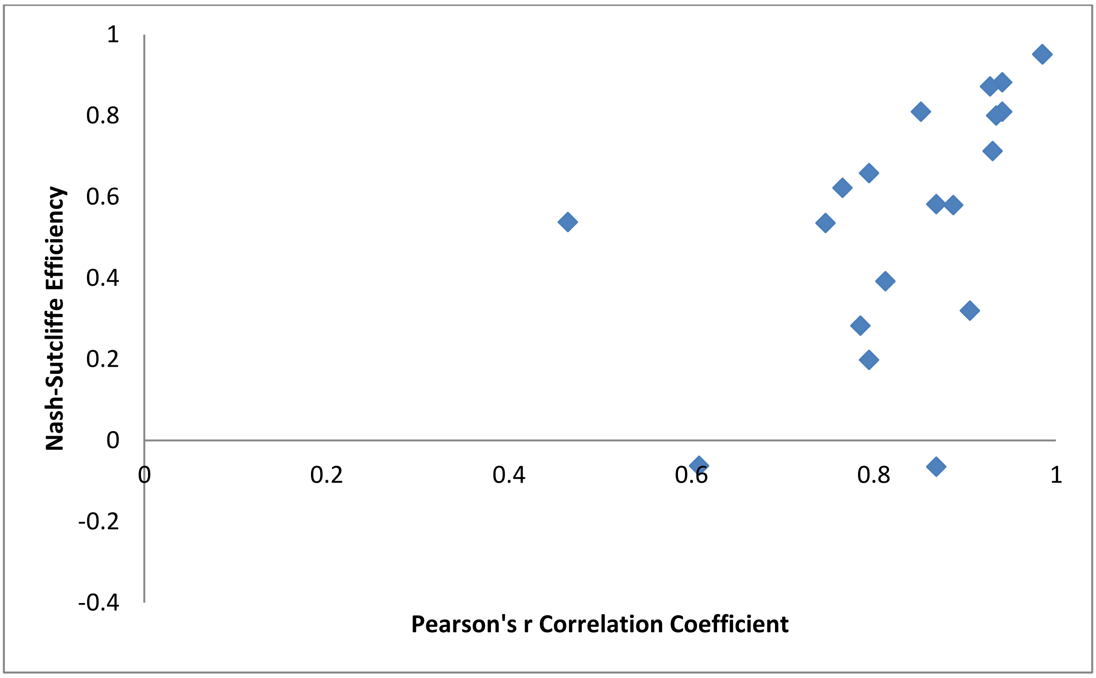

Figure 4.8 The Relation Between NSE and Pearson's *r* Correlation Coefficient

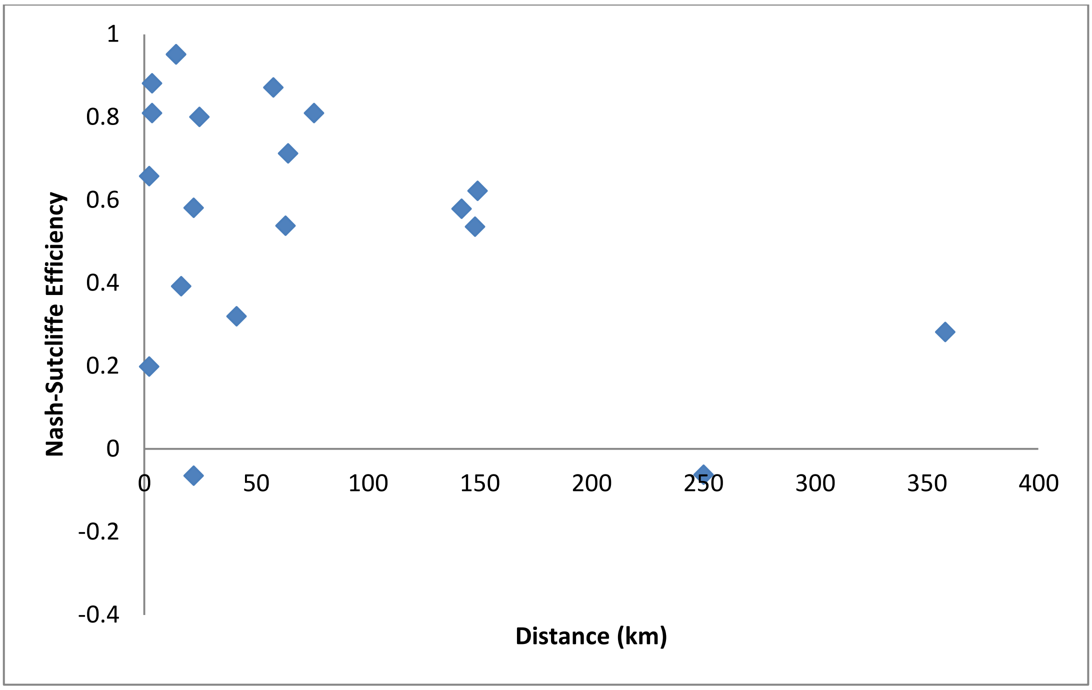

Figure 4.9 The Relation Between NSE and Separation Distance Between Streamgaging Stations

4.6 The Map Correlation Method

Results presented in the previous sections indicate that selecting the most correlated streamgaging station may be more appropriate than selecting the nearest streamgaging station as the reference streamgaging station for streamflow estimation. To test this hypothesis, the most correlated streamgaging station of an ungaged sub-catchment must be identified. However, no streamflow data is available at the ungaged catchment; thus correlation calculations are not possible. In the MCM, the most correlated streamgaging station is tried to be identified for an ungaged catchment by applying geostatistical methods. While applying the map correlation method, it is assumed that runoff exists in any location (Archfield and Vogel, 2010). This enables us to estimate cross-correlations between streamflows of a gaged and an ungaged catchment for any location in the study region. The MCM does not assume that the study area is limited by regions of homogeneity or other hydrologic boundaries (Archfield and Vogel, 2010), it assumes that hydrologic similarities are taken into account in terms of daily streamflow measurements and associated correlation coefficients. Another assumption of the MCM is that r values are isotropic across the study region.

4.6.1 Variograms and R-Software Application

To conduct the map correlation method, variograms need to be generated for each streamgaging station located in the study region. A variograms cloud can be plotted using the relation between semivariances and the corresponding separation distances between each pair of streamgaging stations. Semivariance is the squared difference between correlation coefficients of each pair of streamgaging station. The variogram cloud contains $\binom{n-1}{2}$ points, where n is the number of streamgaging stations in the study region. In this study, 13 streamgaging stations in the Western Blacksea Basin are used and this results in 66 points on each variogram cloud.

The points in the variogram cloud can be grouped into clusters with respect to the separation distance to obtain a sample variogram. This process is called binning. To determine the number of bins, a trial and error process based on visual observation of sample variograms is used. After the trial and error process, same bin length is used for all 13 streamgaging stations located in the Western Blacksea Basin. The semivariance value for each bin is calculated as the average of the semivariances that fall into each separation distance interval. After obtaining the sample variogram, a spherical variogram model is fitted to the sample variograms to obtain a continuous function that relates semivariance to the separation distance. In this study, a spherical variogram model is fitted to the sample variogram due to its simple formulation and to keep the application as similar as possible to the one presented in the original article. Moreover, spherical vaiogram model guarantees a unique solution to the system of linear equations given in Equation 3.21.

The R statistical software is used for binning the variogram clouds, generating sample variograms and fitting variogram models. The source code is obtained from Archfield, S. A. (personal communication with Archfiled, S. A. through 23 November 2011 to 4 May 2012) and modified for 13 streamgaging stations located in the Western Blacksea Basin. As a result of the trial and error process, the number of bins is selected as nine. As examples, the sample and model variograms generated by R statistical software for streamgaging stations 1307, 1334, and 1343 are given in Figures 4.10, 4.11 and 4.12, respectively. The model and sample variograms of the remaining stations are given in Appendix C.

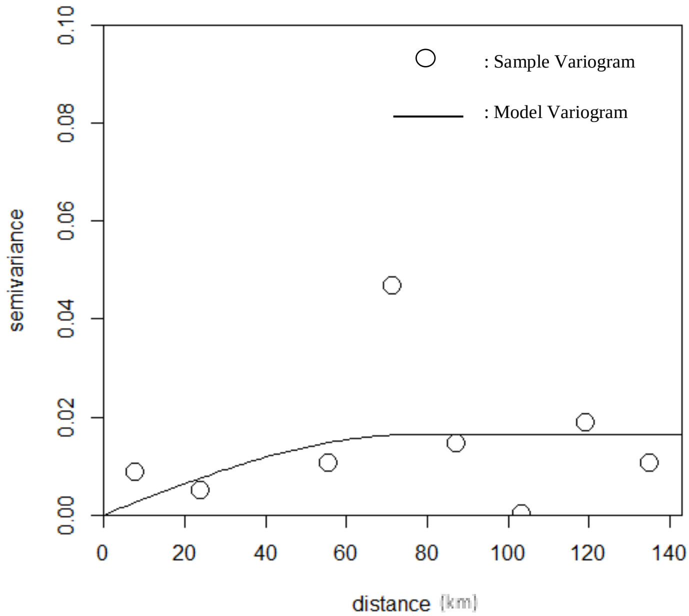

Figure 4.10 The Sample and The Model Variogram of Streamgaging Station 1307

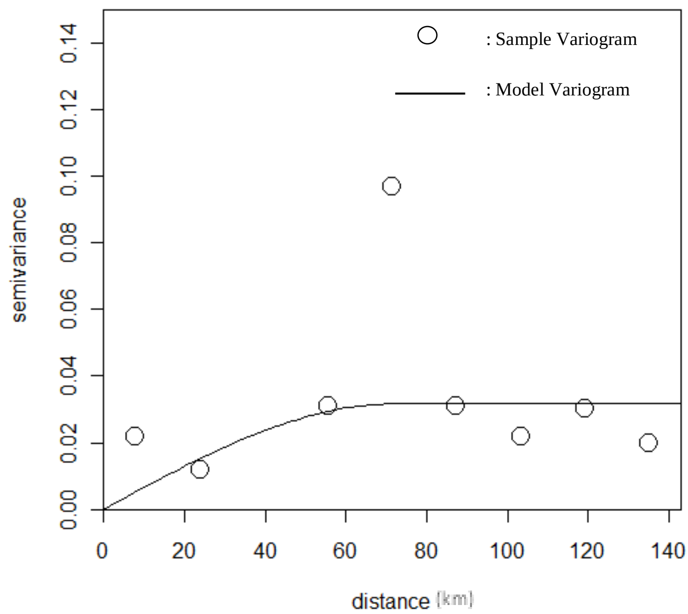

Figure 4.11 The Sample and The Model Variogram of Streamgaging Station 1334

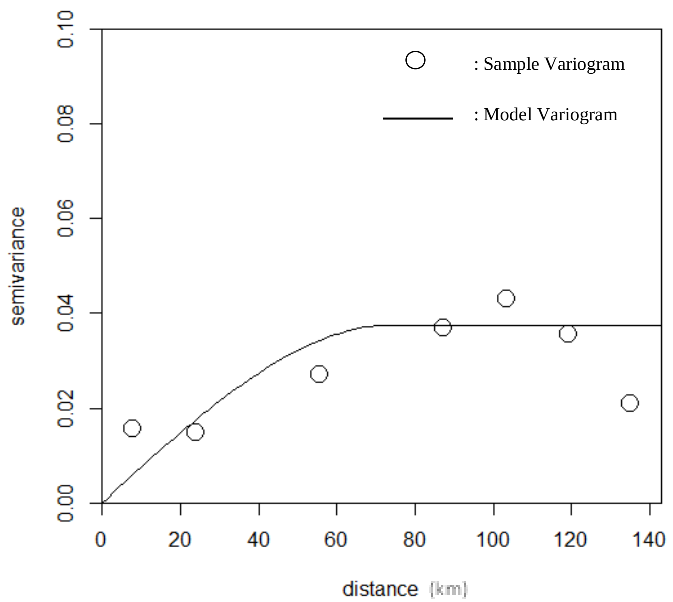

Figure 4.12 The Sample and The Model Variogram of Streamgaging Station 1343

Parameters of spherical variogram models, namely the range, a, and the variance, $\tilde{\sigma}^2$, for 13 streamgaging stations estimated by R software are given in Table 4.6.

Table 4.6 Spherical Variogram Model Parameters of 13 Streamgaging Stations Located in the Western Blacksea Basin

Streamgaging Stations	Variance ($\tilde{\sigma}^2$)	Range parameter (a)
1302	0.0127	82.9223
1307	0.0164	75.9359
1314	0.0390	75.3793
1319	0.0406	72.8733
1327	0.0220	60.6628
1330	0.0065	26.2683
1332	0.0054	128.3016
1334	0.0316	71.9597
1335	0.0037	48.5299
1338	0.0201	131.7337
1339	0.0138	73.0603
1340	0.0126	84.6343

4.6.2 Identification of the Correlation Coefficient of an Ungaged Stream by the MCM

In practical applications, the streamflow data at the ungaged catchment is unknown. So the Pearson's *r* correlation coefficients between the ungaged catchment and other catchments cannot be calculated. For selecting the most correlated streamgaging station for the ungaged catchment, ordinary kriging procedure is applied by using the variogram parameters determined from available daily streamflow data of study streamgaging stations. The flowchart of the MCM is given in Figure 4.13.

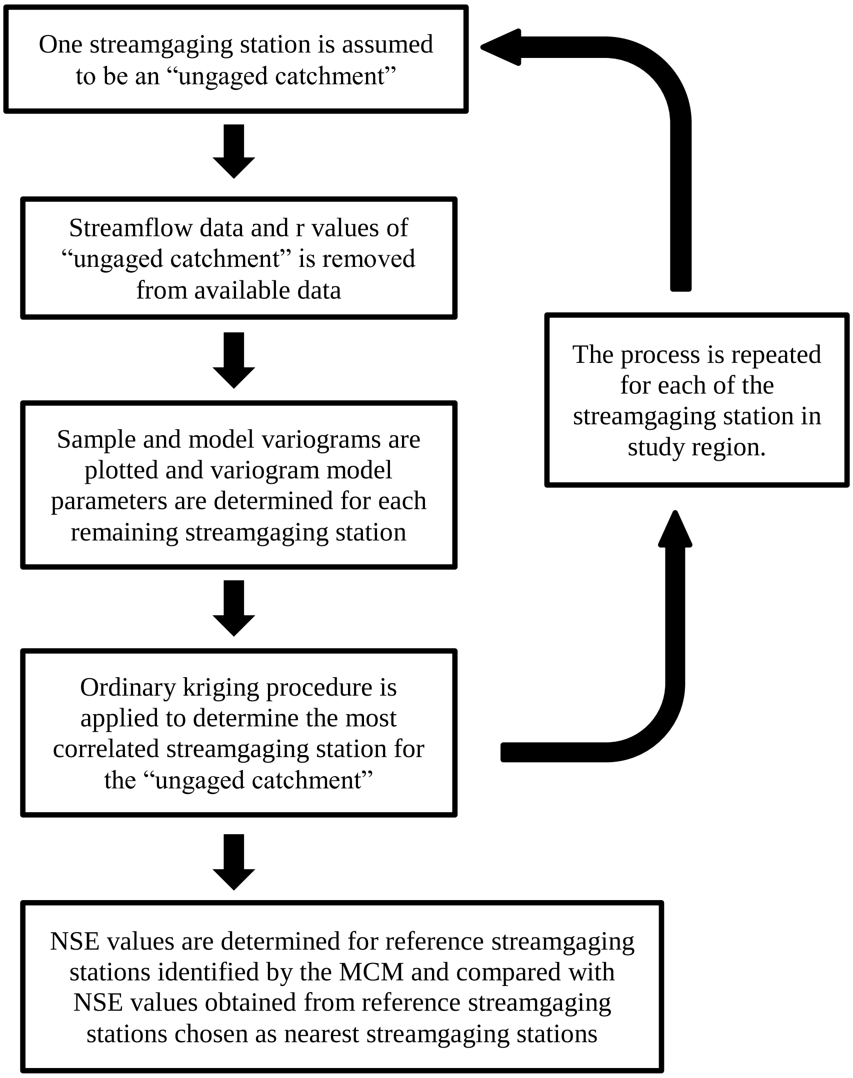

Figure 4.13 The Flowchart of the MCM

The procedure for the identification of the correlation coefficient between an ungaged catchment and one of the study streamgaging stations by using ordinary kriging was outlined in Chapter 3.2.4. Here details of calculations for the identification of the correlation coefficient for an assumed ungaged catchment with one of the study streamgaging stations is provided. As an example, streamgaging station 1302 is assumed to be the ungaged catchment and the correlation between streamgaging stations 1302 and 1307, $r_{1302-1307}$ is identified, through the following steps:

1. The correlation data of streamgaging station 1302 is removed from the set of correlations. The correlations of 1307 with other study streamgaging stations ($r_{1307-1314}, r_{1307-1319}, r_{1307-1327}, \cdots, r_{1307-1343}$) are utilized (See Figure 4.14).

$$\begin{pmatrix} r_{1307-1307} & \langle r_{1307-1314} & \cdots & r_{1307-1343} \rangle \\ r_{1314-1307} & r_{1314-1314} & \cdots & r_{1314-1343} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1343-1307} & r_{1343-1314} & \cdots & r_{1343-1343} \end{pmatrix}$$

Figure 4.14 Observed Correlation Matrix (1302 is Assumed as Ungaged)

where the values in $\langle \ \rangle$ are used to estimate $r_{1302-1307}$. Thus a total of 11 correlation coefficient values are used.

2. To estimate $r_{1302-1307}$, the sample variogram of 1307 need to be generated. Semivariances are calculated using 11 correlation coefficients of 1307 with other streamgaging stations ($r_{1307-1314}, r_{1307-1319}, r_{1307-1327}, \cdots, r_{1307-1343}$) present in the study region. A total of $\binom{11}{2} = 55$ different pairs of correlation coefficients are formed and one semivariance value is calculated for each of these 55 pairs. The corresponding separation

distances, h are identified for each of these 55 pairs and they are binned to generate the sample variogram for 1307.

3. The sample variogram is plotted and a spherical variogram model (See Equation 3.12) is fitted to the sample variogram by R Statistical Software. The sample variogram and the model variogram of 1307 is shown in Figure 4.15. The equation of the model variogram for 1307 is given as follows:

$$\gamma(h) = \begin{cases} 0.0185\left(1.5\left(\dfrac{h}{83.87}\right) - 0.5\left(\dfrac{h}{83.87}\right)^3\right) & 0 \le h < 83.87 \\ 0.0185 & h \ge 83.87 \end{cases} \quad (4.1)$$

As can be seen from Equation 4.1, the range a and the variance $\tilde{\sigma}^2$ are 83.87 and 0.0185, respectively.

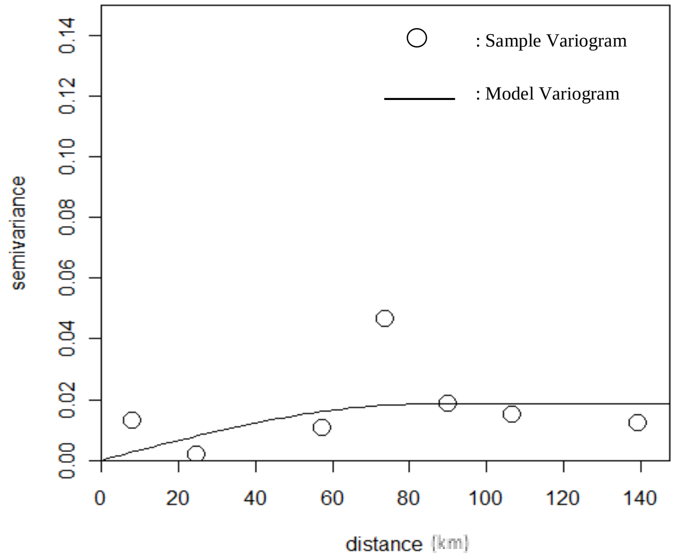

Figure 4.15 The Sample and The Model Variogram of 1307 (1302 is Assumed as Ungaged)

4. The next step is the identification of the covariance matrix. First, the distance matrix is used to determine the variogram matrix. The distance matrix which provides the distances between each pair of study streamgaging stations is given in Figure 4.16.

$$\begin{pmatrix} h_{1314-1314} & h_{1314-1319} & \cdots & h_{1314-1343} \\ h_{1319-1314} & h_{1314-1314} & \cdots & h_{1319-1343} \\ \vdots & \vdots & \ddots & \vdots \\ h_{1343-1314} & h_{1343-1319} & \cdots & h_{1343-1343} \end{pmatrix}$$

Figure 4.16 Distance Matrix

For example, $h_{1314-1314} = 0$ and $h_{1314-1319} = 64.32$ *km* (See Table 4.3).

Then the distance matrix together with Equation 4.1 is used to generate the variogram matrix (see Figure 4.17).

$$\begin{pmatrix} \gamma_{1314-1314} & \gamma_{1314-1319} & \cdots & \gamma_{1314-1343} \\ \gamma_{1319-1314} & \gamma_{1319-1319} & \cdots & \gamma_{1319-1343} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{1343-1314} & \gamma_{1343-1319} & \cdots & \gamma_{1343-1343} \end{pmatrix}$$

Figure 4.17 Variogram Matrix

For instance, to calculate $\gamma_{1314-1319}$, the distance $h_{1314-1319}$ is used in Equation 4.1:

$$\gamma_{1314-1319} = \begin{cases} 0.0185\left(1.5\left(\frac{h_{1314-1319}}{83.87}\right) - 0.5\left(\frac{h_{1314-1319}}{83.87}\right)^3\right) & 0 \le h_{1314-1319} < 83.87 \\ 0.0185 & h_{1314-1319} \ge 83.87 \end{cases} \quad (4.2)$$

where $h_{1314-1319} = 64.32$ *km*. The result of $\gamma_{1314-1319} = 0.0171$.

Then, the covariance matrix (see Figure 4.18) is calculated using the variogram matrix and Equation 3.22.

$$\begin{pmatrix} C_{1314-1314} & C_{1314-1319} & \cdots & C_{1314-1343} \\ C_{1319-1314} & C_{1319-1319} & \cdots & C_{1319-1343} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1343-1314} & C_{1343-1319} & \cdots & C_{1343-1343} \end{pmatrix}$$

Figure 4.18 Covariance Matrix

For example the covariance, $C_{1314-1319}$ can be determined by the following equation:

$$C_{1314-1319} = \tilde{\sigma}^2 - \gamma_{1314-1319} \tag{4.3}$$

where $\tilde{\sigma}^2 = 0.0185$ and $\gamma_{1314-1319} = 0.0174$, the value of $C_{1314-1319} = 0.0014$ is calculated.

5. The next step is the determination of the weights. To do this, the system of linear equations given in Equation 3.21 need to be formed and solved. The covariance matrix obtained in the previous step is modified (a column and a row of ones and a zero is added to it) to obtain the coefficient matrix (Figure 4.19) of Equation 3.21 and it is the values of modified covariance matrix given in Table 4.7. The right hand side matrix of Equation 3.21 is generated by calculating the covariance for $j = 0$ (for this example $j = 0$ corresponds to the streamgaging station 1302) and is given in Table 4.8.

$$\begin{pmatrix} C_{1314-1314} & C_{1314-1319} & \cdots & C_{1314-1343} & 1 \\ C_{1319-1314} & C_{1319-1319} & \cdots & C_{1319-1343} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{1343-1314} & C_{1343-1319} & \cdots & C_{1343-1343} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}$$

Figure 4.19 Modified Covariance Matrix

Table 4.7 Modified Covariance Matrix with Values

Streamgaging Stations	1314	1319	1327	1330	1332	1334	1335	1338	1339	1340	1343	-
1314	0.0185	0.0014	0.0024	0.0000	0.0000	0.0010	0.0015	0.0000	0.0000	0.0000	0.0059	1.0000
1319	0.0014	0.0185	0.0092	0.0000	0.0000	0.0174	0.0004	0.0000	0.0000	0.0000	0.0106	1.0000
1327	0.0024	0.0092	0.0185	0.0000	0.0000	0.0086	0.0000	0.0000	0.0000	0.0000	0.0114	1.0000
1330	0.0000	0.0000	0.0000	0.0185	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
1332	0.0000	0.0000	0.0000	0.0000	0.0185	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
1334	0.0010	0.0174	0.0086	0.0000	0.0000	0.0185	0.0003	0.0000	0.0000	0.0000	0.0096	1.0000
1335	0.0015	0.0004	0.0000	0.0000	0.0000	0.0003	0.0185	0.0000	0.0000	0.0000	0.0008	1.0000
1338	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0185	0.0086	0.0178	0.0000	1.0000
1339	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0086	0.0185	0.0091	0.0000	1.0000
1340	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0178	0.0091	0.0185	0.0000	1.0000
1343	0.0059	0.0106	0.0114	0.0000	0.0000	0.0096	0.0008	0.0000	0.0000	0.0000	0.0185	1.0000
-	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000

Table 4. 8 Covariance Matrix of the Ungaged Catchment (1302)

Streamgaging Stations	1302
1314	0.0000
1319	0.0000
1327	0.0000
1330	0.0000
1332	0.0000
1334	0.0001
1335	0.0000
1338	0.0132
1339	0.0131
1340	0.0138
1343	0.0000
-	1.0000

5. Then the system of linear equations is solved to obtain the weight matrix. The weights for each study streamgaging station are given in Table 4.9. The magnitude of the weights gives an idea about the contributions of the streamgaging stations to the correlation coefficient of the ungaged catchment. For instance, the value of $r_{1302-1307}$ is mainly influenced by the streamgaging stations 1339 and 1340.

Table 4.9 Weights Matrix

w_{1314}	0.0033
w_{1319}	-0.0269
w_{1327}	0.0017
w_{1330}	0.0035
w_{1332}	0.0035
w_{1334}	0.0311
w_{1335}	0.0033
w_{1338}	-0.0248
w_{1339}	0.4519
w_{1340}	0.5528
w_{1343}	0.0005
μ	-6E-05

6. Finally, the correlation between ungaged catchment (1302) and streamgaging station 1307, $r_{1302-1307}$ is determined by inserting the estimated weights into Equation 3.16. The modified version of Equation 3.16 for the estimation of $r_{1302-1307}$ is given as follows:

$$r_{1302-1307} = \sum_{j=1}^{11} w_j r_{1307-j} \tag{4.4}$$

where, j ranges from 1 to 11 and corresponds to the streamgaging stations 1314, 1319,...,1343 respectively. From Equation 4.4, $r_{1302-1307} = 0.7977$ is calculated.

This completes the full procedure for the estimation of the correlation coefficient between an ungaged streamgaging station and a selected study streamgaging station. The same calculation steps are followed for the rest of the streamgaging stations to estimate the correlations between each one of the study streamgaging stations and the ungaged catchment, 1302 (i.e. $r_{1302-1314}$, $r_{1302-1319}$, …, $r_{1302-1343}$). Table 4.10 shows the estimated correlations between each of the study streamgaging stations by 1302 using the MCM procedure. After estimating all the correlations, the streamgaging station having the highest correlation coefficient is selected as the reference streamgaging station. As can be seen from Table 4.10, 1334 is identified as the reference streamgaging station for 1302.

Table 4.10 Estimated Correlations for Ungaged Catchment 1302

Streamgaging Stations	Estimated Correlations
1307	0.7977
1314	0.7730
1319	0.7699
1327	0.8040
1330	0.6258
1332	0.7274
1334	**0.8078**
1335	0.5363
1338	0.6881
1339	0.7750
1340	0.7936
1343	0.7572

MCM procedure is applied for each streamgaging station in the study region and the most correlated streamgaging station (i.e. the reference streamgaging station) is identified. The reference streamgaging stations that are obtained for each streamgaging station at the Western Blacksea Basin are given in Table 4.11. Four out of thirteen most correlated streamgaging stations are identified correctly by the MCM. These are the streamgaging stations which are not the nearest streamgagin stations. The observed and estimated correlations are given in detail for each study streamgaging station located in the Western Blacksea Basin in Appendix D.

Table 4.11 Streamgaging Stations of the Western Blacksea Basin and Reference Streamgaging Stations Identified by the MCM

Study Streamgaging Station	The Nearest Streamgaging Station	The Most Correlated Streamgaging Station	The Reference Streamgaging Station Determined by MCM
1302	1340	1340	1334
1307	1314	1334	1302
1314	1343	1319	1302
1319	1334	1334	1314
1327	1343	**1314**	**1314**
1330	1338	1338	1302
1332	1307	**1340**	**1340**
1334	1319	1319	1343
1335	1314	1332	1319
1338	1340	1340	1302
1339	1302	1302	1340
1340	1338	**1302**	**1302**
1343	1327	**1319**	**1319**

After determining reference streamgaging stations for the Western Blacksea Basin by using the map correlation method, the NSE values of streamgaging stations are compared according to the reference streamgaging station selection criteria (Table 4.12). NSE values of the nearest and the most correlated streamgaging stations were calculated presented in Section 4.5 previously. Here, the NSE values of streamgaging stations that are determined by the MCM, are computed with the same methodology. As can be seen from Table 4.12, utilization of the nearest streamgaging station results in a little better average NSE value compared to that of the MCM streamgaging station. However, it should be recognized that the average NSE value even for the most correlated streamgaging station is only 0.64. This may be the outcome of the utilization of the drainage area ratio method and indicates that utilization of even the most correlated streamgaging station as the reference streamgaging station do not produce accurate daily streamflow estimates in the selected study area.

In four out of 13 streamgaging stations, the MCM successfully identified the most correlated streamgaging station. For three of these streamgaging stations (1327, 1340, 1343), the improvement in the NSE values compared to those obtained from the nearest streamgaging station is encouraging. On the other hand, for streamgaging station 1332 the MCM determines the most correlated streamgaging station successfully, but there is no improvement in NSE value. This is due to utilization of the drainage area ratio method and drainage areas of the most correlated streamgaging station (1340) and streamgaging station 1332 being very different than each other (see Table 4.1).

Table 4.12 NSE Values Determined for Different Reference Streamgaging Station Selection Criteria

Study Streamgaging Station	The Nearest Streamgaging Station	The Most Correlated Streamgaging Station	The MCM Streamgaging Station
1302	0.9502	0.9502	0.0435
1307	0.8091	0.5790	-0.3826
1314	0.3190	0.7121	-0.8592
1319	0.8815	0.8815	0.6372
1327	0.5813	0.8715	0.8715
1330	0.5353	0.5353	0.3324
1332	0.6221	0.2818	0.2818
1334	0.8094	0.8094	0.6578
1335	0.5378	-0.0630	0.8580
1338	0.6577	0.6577	0.6870
1339	0.3917	0.3917	0.5277
1340	0.1981	0.9511	0.9511
1343	-0.0650	0.8002	0.8002
Average of NSE Values	**0.5560**	**0.6430**	**0.4159**

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary and Conclusions

Water resources have significant impact on economic development of Turkey. In order to benefit effectively from water resources, streamflow series must be investigated carefully and understood entirely.

Planning and construction of SHPPs have accelerated in Turkey in the recent years. In the design of SHPPs, daily streamflow values are commonly used. Daily time series of streamflow data can be obtained easily at gaged catchments, however, at ungaged catchments daily streamflow values are estimated from data collected at a nearby or the nearest streamgaging station.

Although selecting the nearest streamgaging station as the reference streamgaging station is preferred in common practice, acceptance of the distance as the primary reference streamgaging station selection criterion may not be always correct. Archfield and Vogel (2010) introduced a new method called the MCM to select the reference streamgaging station by utilizing the correlations between gaged and ungaged catchments.

In this study, the MCM is analyzed and its applicability to the Western Blacksea Basin in Turkey is tested. The following results and conclusions are acquired:

1. Archfield and Vogel (2010) stated that the MCM was not tested for other geographic regions of larger and smaller sizes but it is reasonable to assume that it would be applicable to other study regions. Here, applicability of the MCM is tested for the Western Blacksea Basin which is larger than the study region used in Archfield and Vogel (2010). Results of this study are comparable to those of Archfield and Vogel's. Thus, it

can be concluded that the MCM is a potential tool for the identification of the reference streamgaging station.

2. Archfield and Vogel (2010) used 28 streamgaging stations and their streamflow values for a common 30 year period in their study. In the current study, 13 streamgaging stations with only 10 year common streamflow values are utilized. To lengthen the common time period of daily streamflow values with the motivation of obtaining more representative correlation coefficients, the number of streamgaging stations was limited to 13. Although the number of streamgaging stations in the Western Blacksea Basin is comparably less than that is used in the original study, the MCM determined the most correlated streamgaging station correctly in one third of the study streamgaging stations both in this and the original studies. However, the NSE values obtained by Archfield and Vogel (2010) were better compared to those obtained in this study; and one main reason for this is believed to be the difference in the total number of streamgaging stations used in the study region.

3. Approximately 60% of the study streamgaging stations of the Western Blacksea Basin were not mostly correlated with the nearest streamgaging station. This indicates that the separation distance might not be the best reference streamgaging station selection criterion all the time; and alternative methods should be evaluated. One such alternative method is the MCM and it produced promising results in its first application for a basin in Turkey.

4. The average NSE of the most correlated streamgaging stations is higher than those of the nearest streamgaging stations for estimating daily streamflow series for an ungaged catchment with the drainage area ratio method. Thus, if the most correlated streamgaging station to an ungaged catchment can be identified, it will produce better streamflow estimates at the ungaged location compared to those obtained by the nearest streamgaging station. The MCM method provides this opportunity; however with the current algorithm, the MCM succeeds in identifying the

most correlated streamgaging station only about one third of the time. Application of the MCM to the Western Blacksea Basin demonstrated the potential of this method in identifying the reference streamgaging station for an ungaged catchment in Turkey. However, the results of both Archfield and Vogel's (2010) work and this study showed that assumptions (i.e. correlation coefficients being isotropic, utilization of spherical variograms models, and minimum number of streamgaging stations required) used in the current algorithm of the MCM method need to be further evaluated and studied to improve the efficiency of the method.

The recommendations for the improvement of the MCM are discussed in the next section.

5.2 Recommendations for Future Research

The suggestions for the improvement of the MCM are listed as follows:

1. While performing the geostatistical analysis, the Euclidian distances are used. Instead of the Euclidian distances between the streamgaging stations, the distances along the river reaches between streamgaging stations can be used. This suggestion is not tested in this study because the number of streamgaging stations in most of the river reaches was limited to two or three in the Western Blacksea Basin. For basins where more streamgaging stations are located along the river reaches this recommendation might improve the efficiency of the MCM.

2. In the current study, the correlation coefficients were assumed to be isotropic within the study area similar to Archfield and Vogel's (2010) work. This resulted in utilization of omnidirectional variograms. However, correlation coefficients might be anisotropic and utilization of directional variograms might improve the efficiency of the MCM. Various grouping mechanism, for example with respect to the flow direction (i.e. upstream

and downstream streamgaging stations) or the bifurcation ratios, might be worth studying.

3. The simplest and most practical suggestion is to increase the number of study streamgaging stations by adding streamgaging stations of neighboring basins or using a shorter common period of daily streamflow series. As in all statistical applications, increasing the sample size is expected to increase accuracy of the results.

REFERENCES

Ak, M. (2011). *"Alternative Feasibility Studies for Altıparmak Dam and HEPP",* MSc Thesis. Middle East Technical University, Ankara.

Archfield, S. A., and Vogel, R. M. (2010). "Map Correlation Method: Selection of a reference streamgaging station to estimate daily streamflow at ungaged catchments". Water Resources Research, 46, W10513.

Archfield, S.A., Vogel, R.M., Wagener, T., and Singh, R. (2009). "Rainfall-runoff model calibration at an ungauged catchment using the map-correlation method". American Geophysical Union, Fall Meeting 2010, December 15, 2009.

Asquith, W. H., Roussel, M. C., and Vrabel, J. (2006). "Statewide Analysis of the Drainage-Area Ratio Method for 34 Streamflow Percentile Ranges in Texas". U. S. Geological Survey, Scientific Investigations Report, 2006-5286.

Aydın, B. E. (2010). *"Feasibility Study of Multiple Hydropower Projects",* MSc Thesis. Middle East Technical University, Ankara.

Bohling, G. (2005). "Kriging". Kansas Geological Survey, C&PE 940.

Emerson, D. G., and Dressler, V. M. (2002). "Historic and Unregulated Monthly Streamflow for Selected Sites in the Red River of the North Basin in North Dakota, Minnesota and South Dakota, 1931-99". U. S. Geological Survey, Water-Resources Investigations Report, 02-4095.

Emerson, D. G., Vecchia, A. V., and Dahl, A. L. (2005). "Evaluation of Drainage-Area Ratio Method Used to Estimate Streamflow for the Red River of the North Basin, North Dakota and Minnesota". U. S. Geological Survey, Scientific Investigations Report, 2005-5017.

Esralew, R. A., and Smith, S. J. (2009). “Methods for Estimating Flow-Duration and Annual Mean-Flow Statistics for Ungaged Streams in Oklahoma”. U. S. Geological Survey, Scientific Investigations Report, 2009-5267.

GLM Engineering (2009). “Minimum Instream Flow Estimation at Ungaged Stream Sites in Puerto Rico”. Retrieved January 24, 2012 from http://glmeng-coop.com/

Gomez-Hernandez, J. J. (2005). “Geostatistics”. Hydrogeophysics, 59-83, Netherlands.

Google Earth (2011, August 8), Google Inc.

Goswami, M., O’Connor, K. M., and Shamseldin, A. Y. (2002). “Structures and Performances of Five Rainfall-Runoff Models for Continuous River-Flow Simulation”. Department of Engineering Hydrology, National University of Ireland, Galway, Ireland.

Guenthner, R. S., Weigel, J. F., and Emerson, D. G. (1990). “Gaged and Estimated Monthly Streamflow During 1931-84 for Selected Sites in the Red River of the North Basin in North Dakota and Minnesota”. U. S. Geological Survey, Water-Resources Investigations Report, 90-4167.

Hayter, A. J. (2007). *Probability and Statistics for Engineers and Scientists,* 3^{rd} Edititon, Thomas Brooks/Cole, Canada.

Hortness, J. E., and Berenbock, J. (2001). “Estimating Monthly Annual Streamflow Statistics at Ungaged Sites in Idaho”. U. S. Geological Survey, Water-Resources Investigations Report, 01-4093.

Isaaks, E. H., and Srivastava, R. M. (1989). *Applied Geostatistics*, Oxford University Press, New York.

Journel, A. G. (1989). “Fundementals of Geostatistics In Five Lessons”. American Geophysical Union, Short Course in Geology, Volume 8.

Kang L. (2012). “Fitting Variogram Models by Weighted Least Squares”. Retrieved April 5, 2012, from www2.isye.gatech.edu

Kocatepe, Y. (2011). *“Developing Regionalized Flow Duration Curves and Evaluating the Performances in Ungauged Basins”,* MSc Thesis. Middle East Technical University, Ankara.

Korleski, C., and Strickland, T. (2009). “Biological and Water Quality Study of Grand River Basin”. State of Ohio Environmental Protection Agency, EPA Technical Report EAS/2009-6-5.

Matheron, G. (1971). “The Theory of Regionalized Variables and its Applications”. Ecole Nationale Supérieure des Mines de Paris, Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, No 5.

Mohamoud, Y., and Parmar, R. (2009). “A Regionalized Flow Duration Curve Method to Predict Streamflow for Ungaged Basins: A Case Study of the Rappahannock Watershed in Virginia, USA”. 8^{th} IAHS Scientific Assembly and 37^{th} IAH Congress, Hyderabad, India.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L. (2007). “Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations”. Soil & Water Division of ASABE, Vol. 50(3), 885-900.

Mutlu, R. (2010). *“Feasibility Study of a Hydropower Project: Case Study of Niksar HEPP, Turkey”,* MSc Thesis. Middle East Technical University, Ankara.

Niu, D., and Tarumi, T. (2009). “Estimation for the Parameters in Geostatistics”. *Cooperation in Classification and Data Analysis,* Akinori O., Tadashi I., and Hans-Hermann B., Springer-Verlag Berlin Heidelberg.

Noppé, M. A. (1994). “Practical Geostatistics for On-site Analysis-A Coal Example” Paper in Proceedings of the Mining Geostatistics Conference, Geostatistical Association of South Africa.

Olea, R. A. (2009). “A Practical Primer on Geostatistics”. U. S. Geological Survey, Open-File Report, 2009-1103.

Patil, S., and Stieglitz, M. (2011). “Controls on Hydrologic Similarity: Role of Nearby Gauged Catchments for Prediction at an Ungauged Catchment”. 8^{th} Hydrology and Earth System Sciences Discussions.

Perry, A. C., Wolock, D. M., and Artman, J. C. (2004). “Estimates of Flow Duration, Mean Flow, and Peak-Discharge Frequency Values for Kansas Stream Locations”. U. S. Geological Survey, Scientific Investigations Report, 2004-5033.

Smakhtin, V., and Weregala, N. (2005). “An Assessment of Hydrology and Environmental Flows in the Walawe River Basin, Sri Lanka”. International Water Management Institute, Working Paper 103.

Stat506 Notes (2012). “Fitting Theoretical Variogram Models: The R Package geoR and Custom Made Functions”. Retrieved February 12, 2012, from http://streaming.stat.iastate.edu

The R Development Core Team (2011). “R: A Language and Environment for Statistical Computing”. Reference Index, ISBN 3-900051-07-0.

The R Development Core Team (2011). “R Data Import/Export”. Reference Index, ISBN 3-900051-10-0.

Webster, R., and Oliver, M. A. (2007). *Geostatistics for Environmental Scientists,* John Wiley & Sons Ltd.

Yaprak S., and Arslan E. (2008). “Kriging Yöntemi ve Geoit Yüksekliklerinin Enterpolasyonu”. Jeodezi, Jeoinformasyon ve Arazi Yönetimi Dergisi, Sayı 98.

APPENDIX A

PEARSON'S *r* CORRELATION COEFFICIENT AND NSE CALCULATIONS (REAL/ARITHMETIC)

Pearson's r correlation coefficient and NSE values with real/arithmetic streamflow data is carried out in additon to log transformed streamflow calculations. Table A.1 gives Pearson's r correlation coefficients determined by using arithmetic daily streamflow data, respectively. Correlation versus distance relations for streamgaging stations located in the Western Blacksea Basin is given in Figures A.1 to A.13. Table A.3 provides the NSE values which are determined according to different reference streamgaging station selection criteria. Figure A.14 is the comparison of the NSE values that are obtained from the nearest streamgaging station and the most corralated streamgaging station. The relation between NSE and seperation distance is given at Figure A.15. The relation between NSE and Pearson's r correlation coefficient is provided at Fgure A.16.

Table A. 1 Pearson's *r* Correlation Coefficients of 13 Streamgaging Stations Located in the Western Blacksea Basin

Streamgaging Stations	1302	1307	1314	1319	1327	1330	1332	1334	1335	1338	1339	1340	1343
1302	1.0000	0.7082	0.7237	0.7548	0.7058	0.3106	0.6239	0.8035	0.3041	0.5027	0.6885	0.9850	0.7325
1307	0.7082	1.0000	0.7004	0.7475	0.7124	0.2766	0.5907	0.8079	0.2557	0.3012	0.5605	0.6956	0.7292
1314	0.7237	0.7004	1.0000	0.8594	0.8879	0.1257	0.3776	0.8403	0.1942	0.1949	0.6980	0.7046	0.8421
1319	0.7548	0.7475	0.8594	1.0000	0.7877	0.1466	0.4310	0.8874	0.1957	0.2530	0.6396	0.7340	0.9001
1327	0.7058	0.7124	0.8879	0.7877	1.0000	0.2000	0.4148	0.8420	0.2360	0.2159	0.7127	0.6719	0.8116
1330	0.3106	0.2766	0.1257	0.1466	0.2000	1.0000	0.2905	0.2460	0.1962	0.4521	0.2488	0.3288	0.1691
1332	0.6239	0.5907	0.3776	0.4310	0.4148	0.2905	1.0000	0.5510	0.2771	0.3788	0.4284	0.6155	0.4238
1334	0.8035	0.8079	0.8403	0.8874	0.8420	0.2460	0.5510	1.0000	0.2498	0.2680	0.7006	0.7760	0.8495
1335	0.3041	0.2557	0.1942	0.1957	0.2360	0.1962	0.2771	0.2498	1.0000	0.1382	0.2032	0.3005	0.2263
1338	0.5027	0.3012	0.1949	0.2530	0.2159	0.4521	0.3788	0.2680	0.1382	1.0000	0.4397	0.5341	0.2728
1339	0.6885	0.5605	0.6980	0.6396	0.7127	0.2488	0.4284	0.7006	0.2032	0.4397	1.0000	0.6677	0.6712
1340	0.9850	0.6956	0.7046	0.7271	0.6719	0.3288	0.6155	0.7760	0.3005	0.5341	0.6677	1.0000	0.7153
1343	0.7325	0.7292	0.8421	0.9001	0.8116	0.1691	0.4238	0.8495	0.2263	0.2728	0.6712	0.7153	1.0000

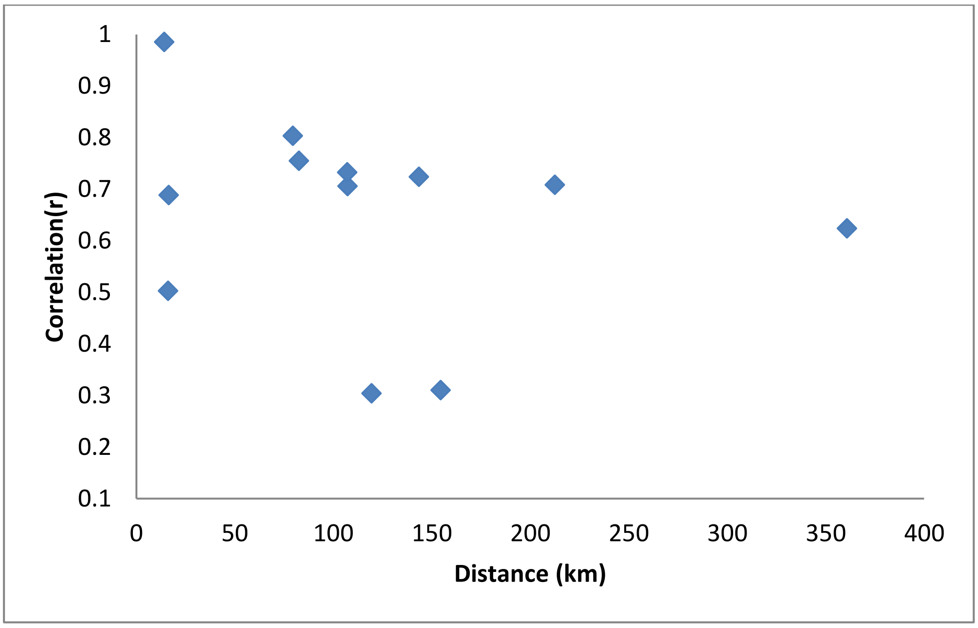

Figure A. 1 Distance versus Correlation Relation of Streamgaging Station 1302

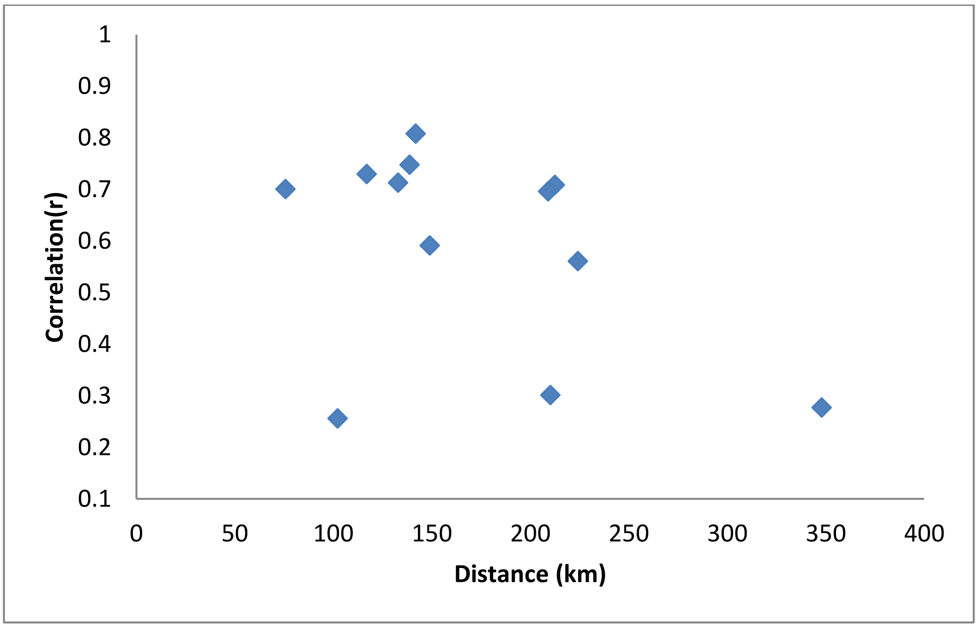

Figure A.2 Distance versus Correlation Relation of Streamgaging Station 1307

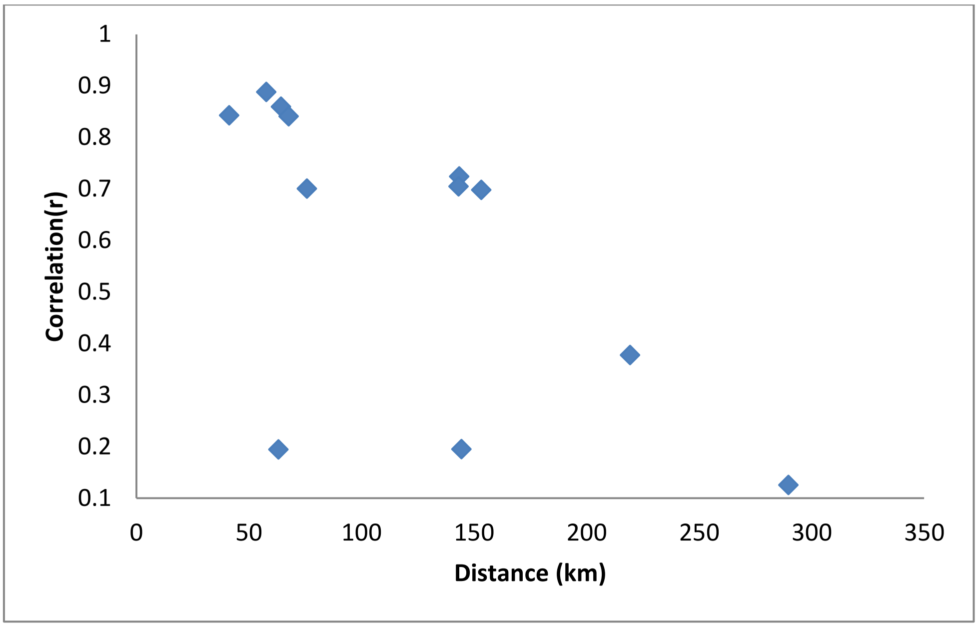

Figure A. 3 Distance versus Correlation Relation of Streamgaging Station 1314

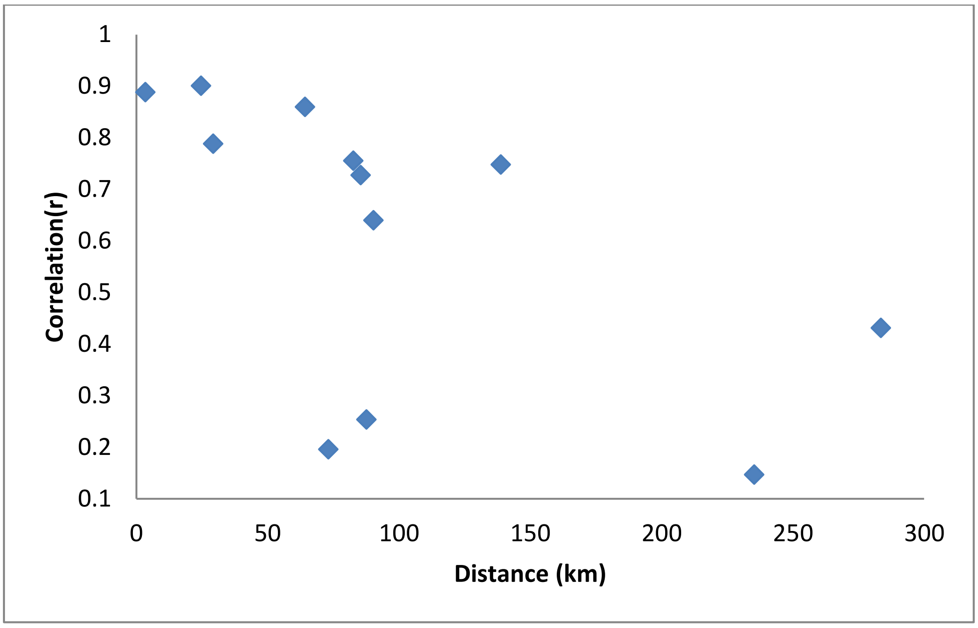

Figure A.4 Distance versus Correlation Relation of Streamgaging Station 1319

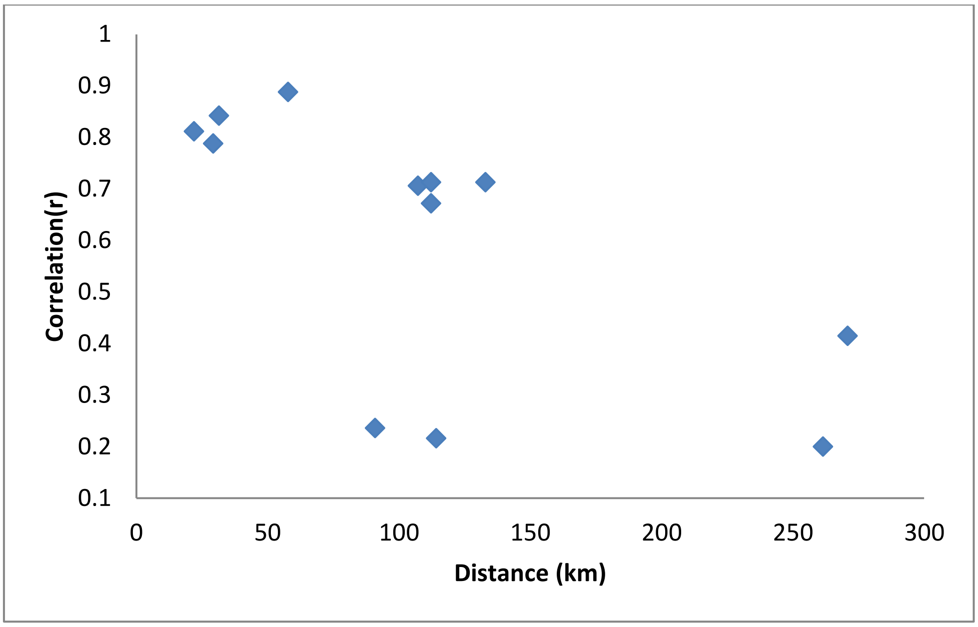

Figure A.5 Distance versus Correlation Relation of Streamgaging Station 1327

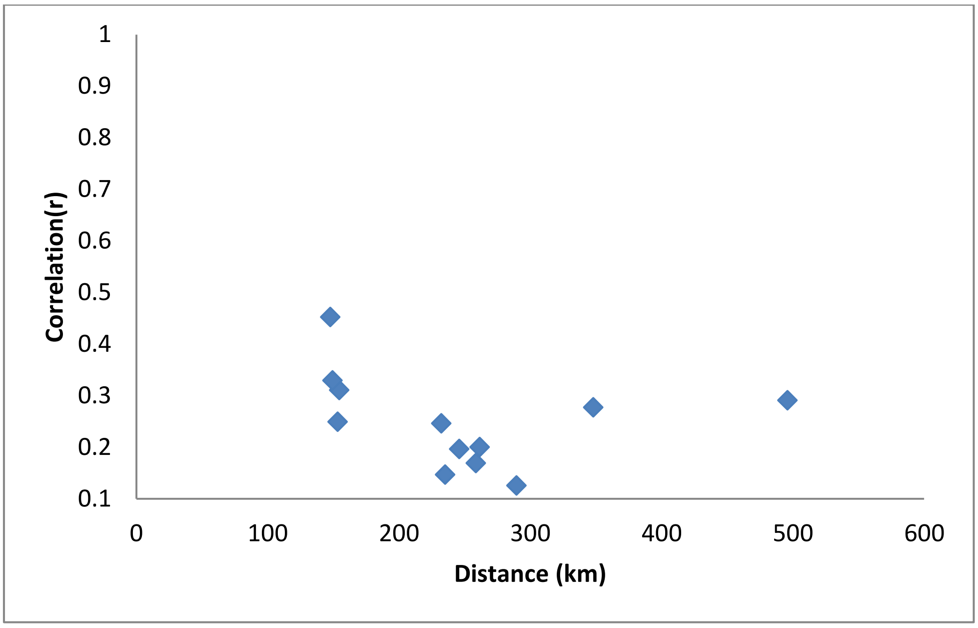

Figure A.6 Distance versus Correlation Relation of Streamgaging Station 1330

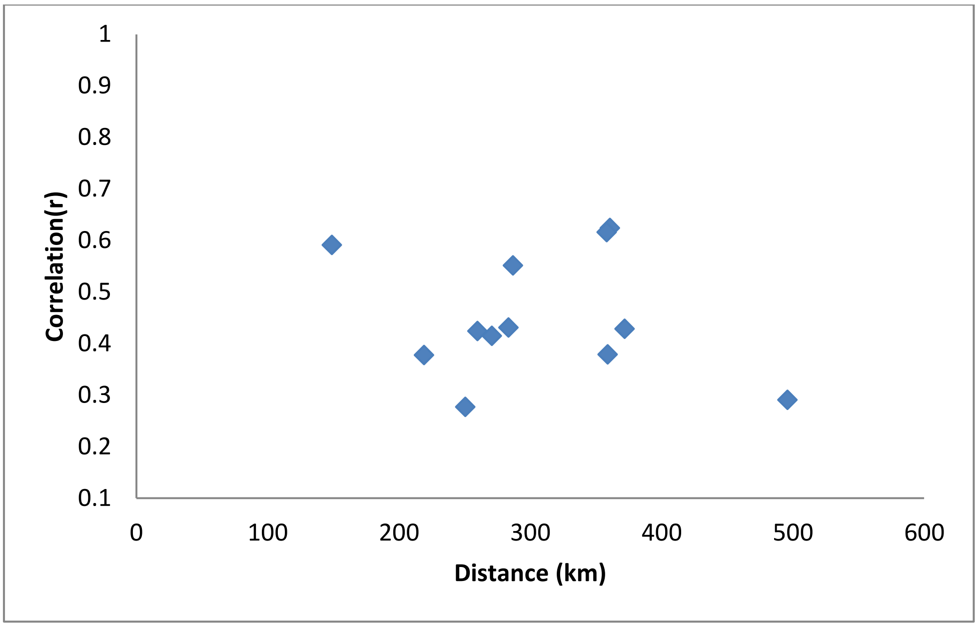

Figure A.7 Distance versus Correlation Relation of Streamgaging Station 1332

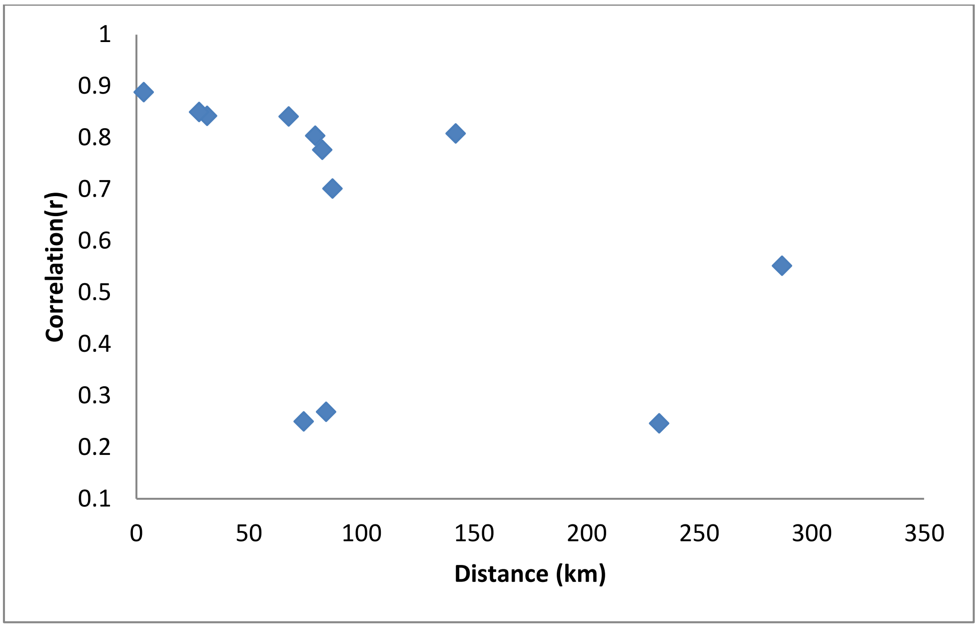

Figure A.8 Distance versus Correlation Relation of Streamgaging Station 1334

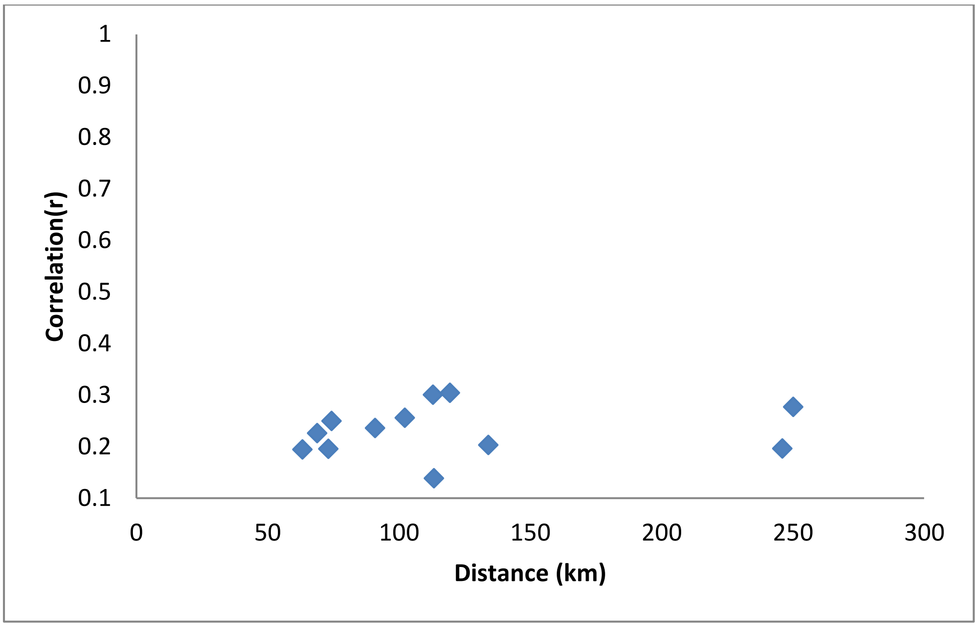

Figure A.9 Distance versus Correlation Relation of Streamgaging Station 1335

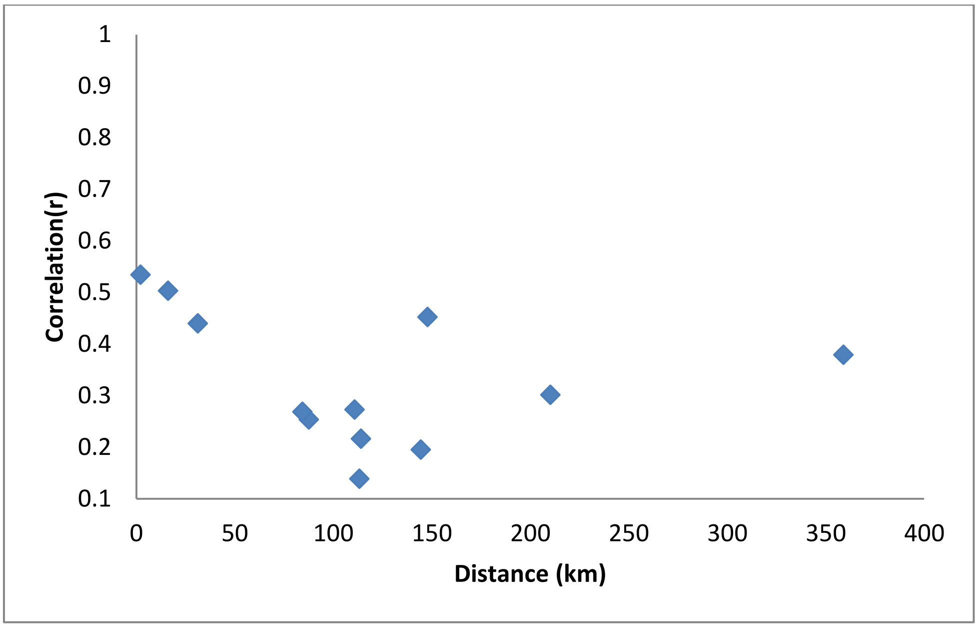

Figure A.10 Distance versus Correlation Relation of Streamgaging Station 1338

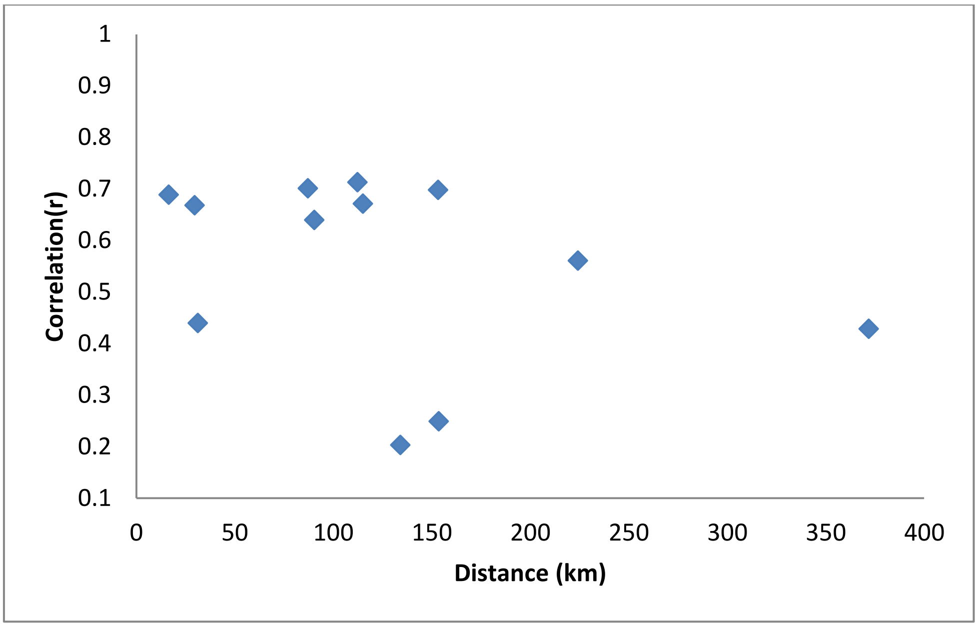

Figure A.11 Distance versus Correlation Relation of Streamgaging Station 1339

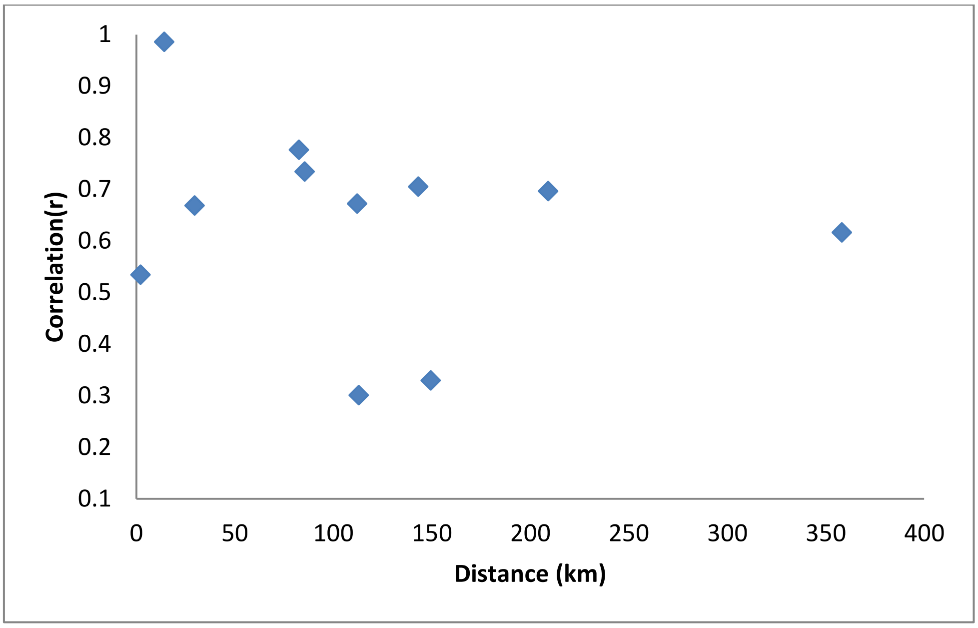

Figure A.12 Distance versus Correlation Relation of Streamgaging Station 1340

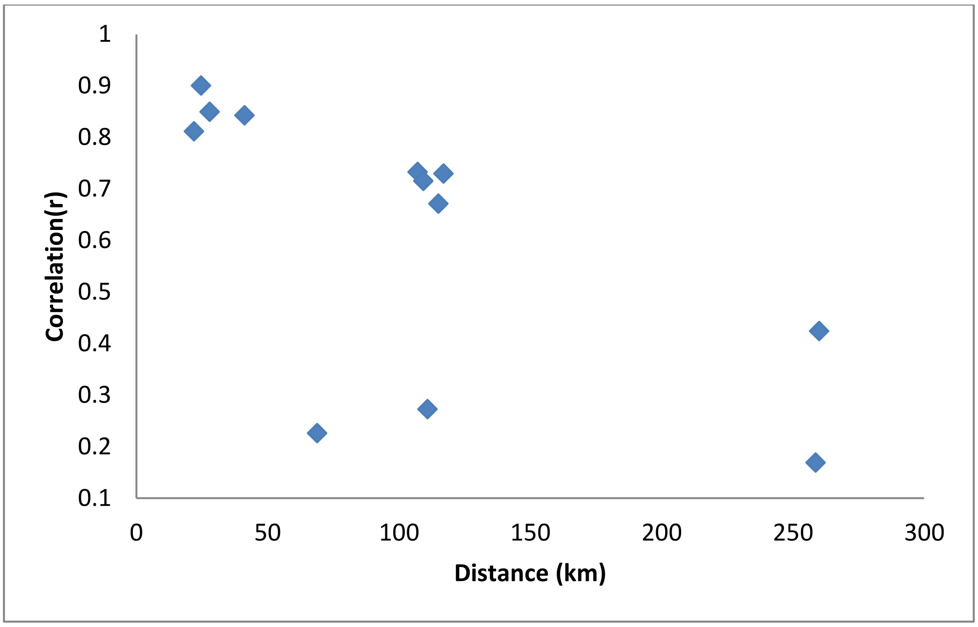

Figure A.13 Distance versus Correlation Relation of Streamgaging Station 1343

Table A.2 NSE Values That Are Calculated from the Nearest Streamgaging station and the Most Correlated Streamgaging Station for 13 Streamgaging Stations in the Western Blacksea Basin

Study Streamgaging Station	The Nearest Streamgaging Station taken as the reference	Highest Correlated Streamgaging Station taken as the reference
1302	0.8866	0.8866
1307	0.4761	**0.6143**
1314	-1.2471	**-0.1179**
1319	**0.7497**	0.5630
1327	0.4993	**0.6384**
1330	-1.4980	-1.4980
1332	**0.2146**	-0.2300
1334	0.4799	0.4799
1335	**0.7047**	-2.2136
1338	0.2853	0.2853
1339	**0.3017**	-0.0710
1340	-1.4768	**0.9222**
1343	0.5943	**0.7356**
Mean of NSE values	0.0746	0.0765

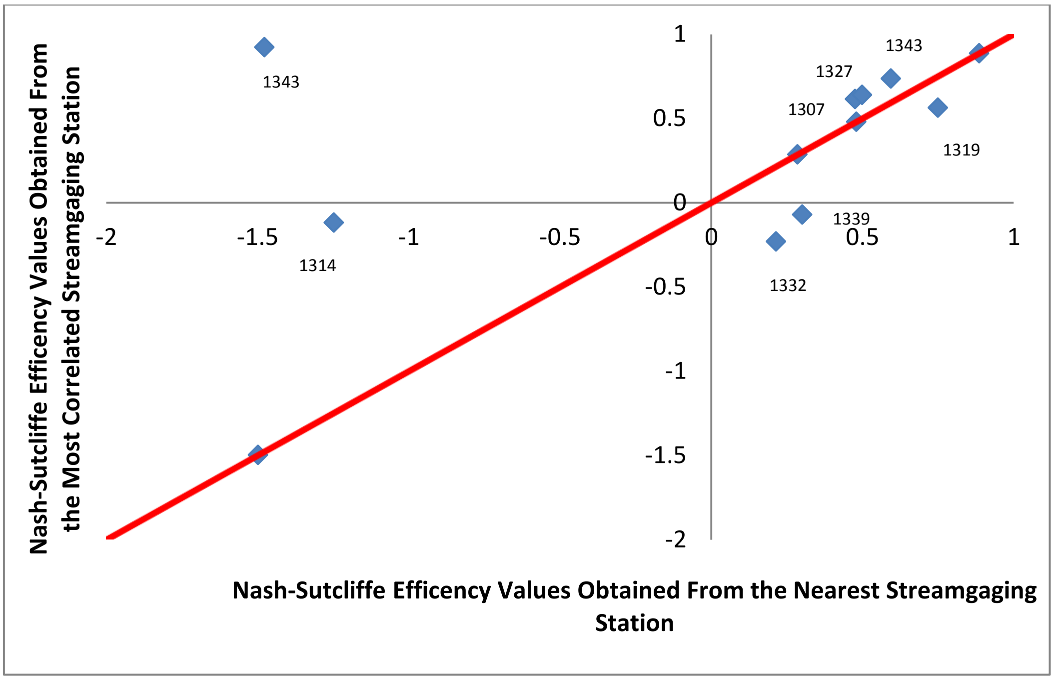

Figure A.14 Comparison of NSE Values Based on the Reference Streamgaging Station Selection Criteria

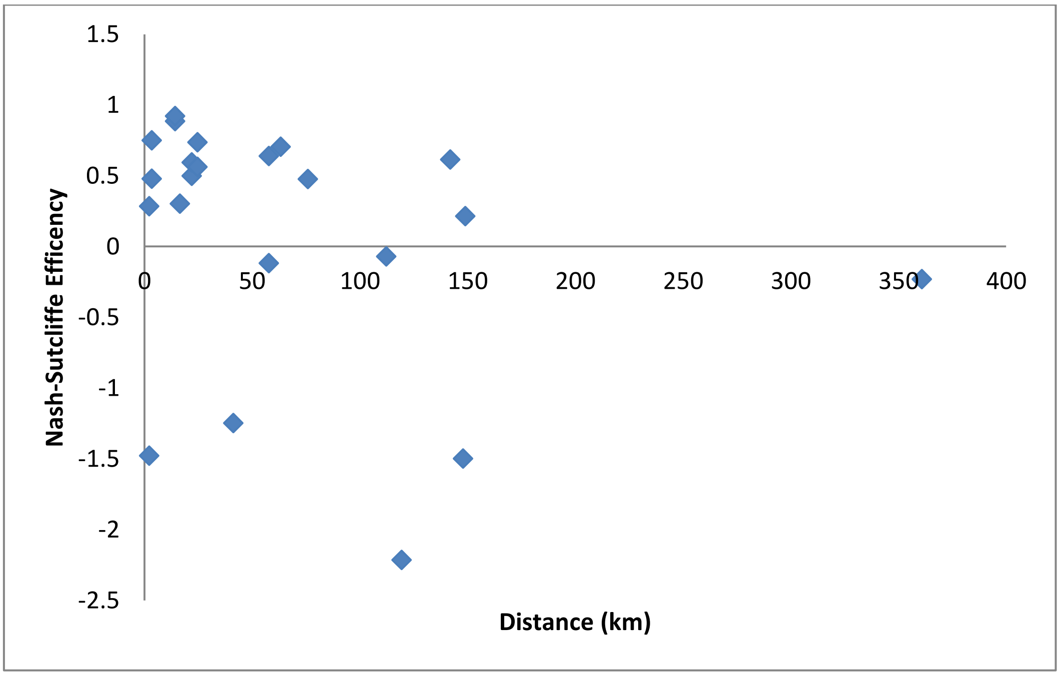

Figure A.15 The Relation Between NSE and Separation Distance Between Streamgaging Stations

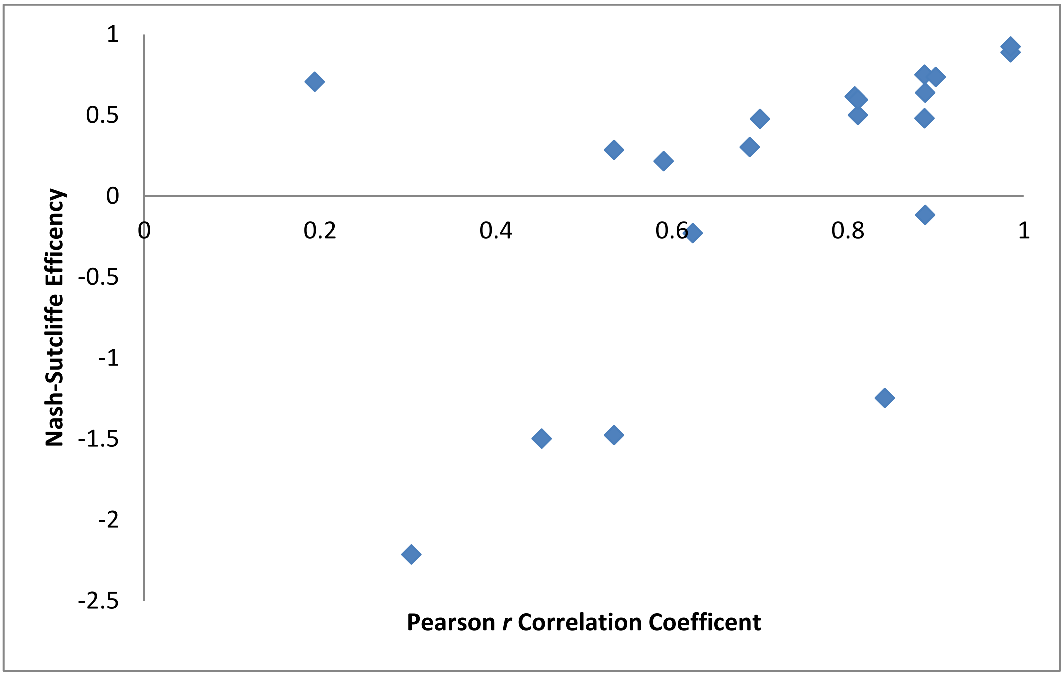

Figure A.16 The Relation Between NSE and Pearson's *r* Correlation Coefficient

APPENDIX B

PEARSON'S r CORRELATION COEFFICIENT AND DISTANCE RELATIONS

This section includes figures that give the correlation versus distance relations of streamgaging stations located in the Western Blacksea Basin (Figure B.1 to B.13).

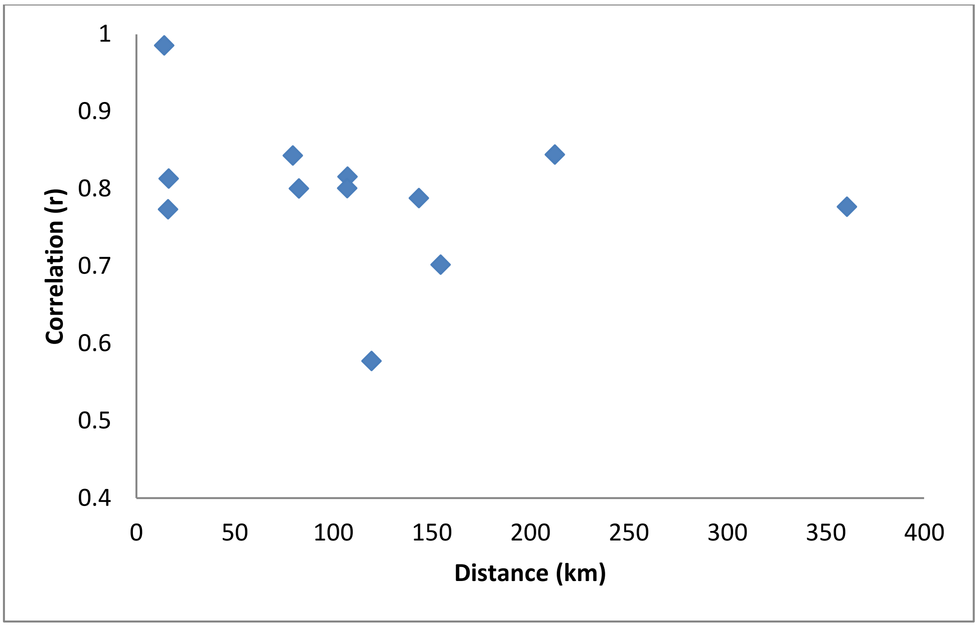

Figure B.1 Distance versus Correlation Relation of Streamgaging Station 1302

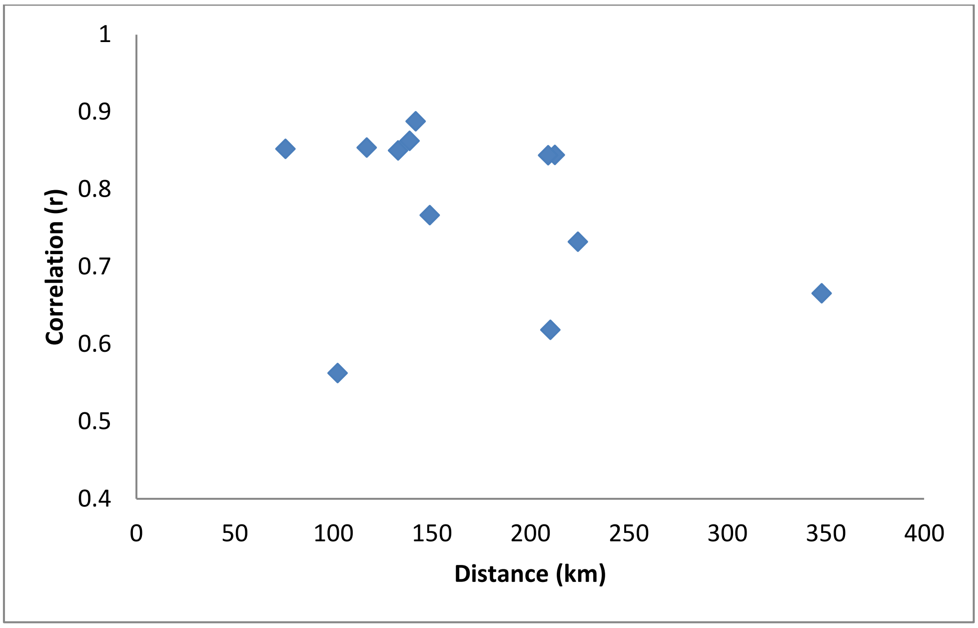

Figure B. 2 Distance versus Correlation Relation of Streamgaging Station 1307

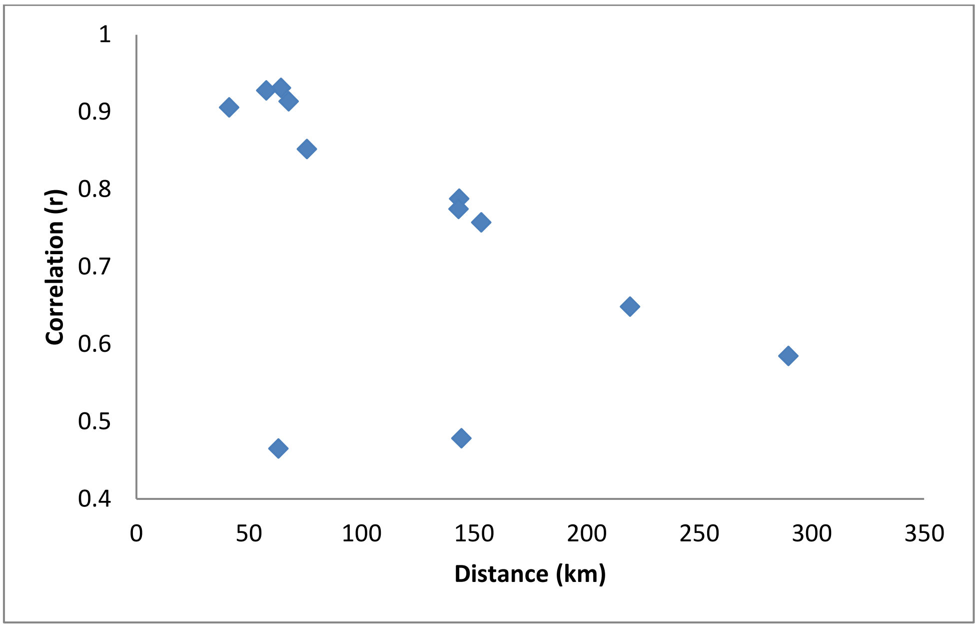

Figure B.3 Distance versus Correlation Relation of Streamgaging Station 1314

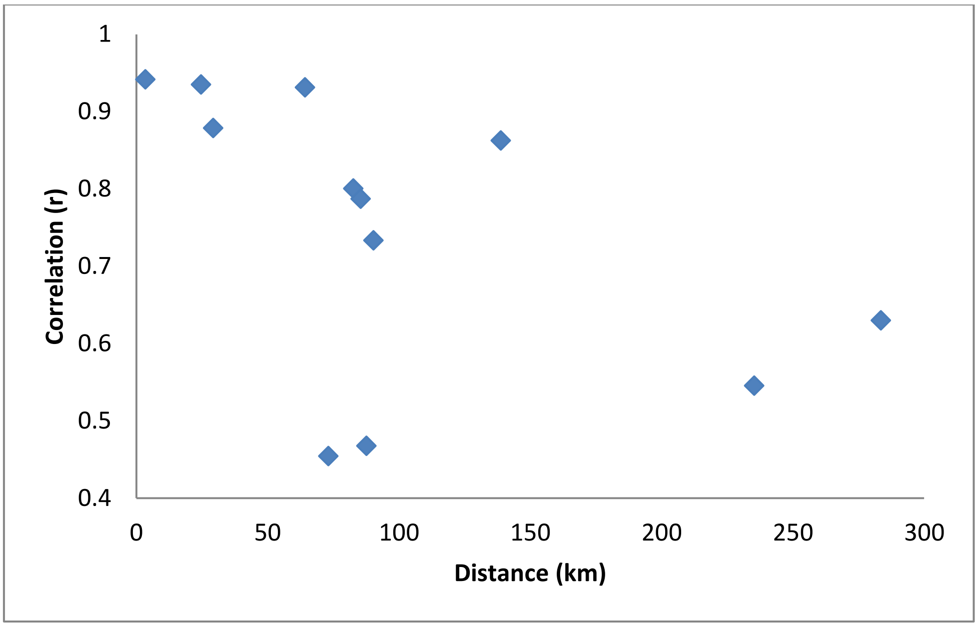

Figure B. 4 Distance versus Correlation Relation of Streamgaging Station 1319

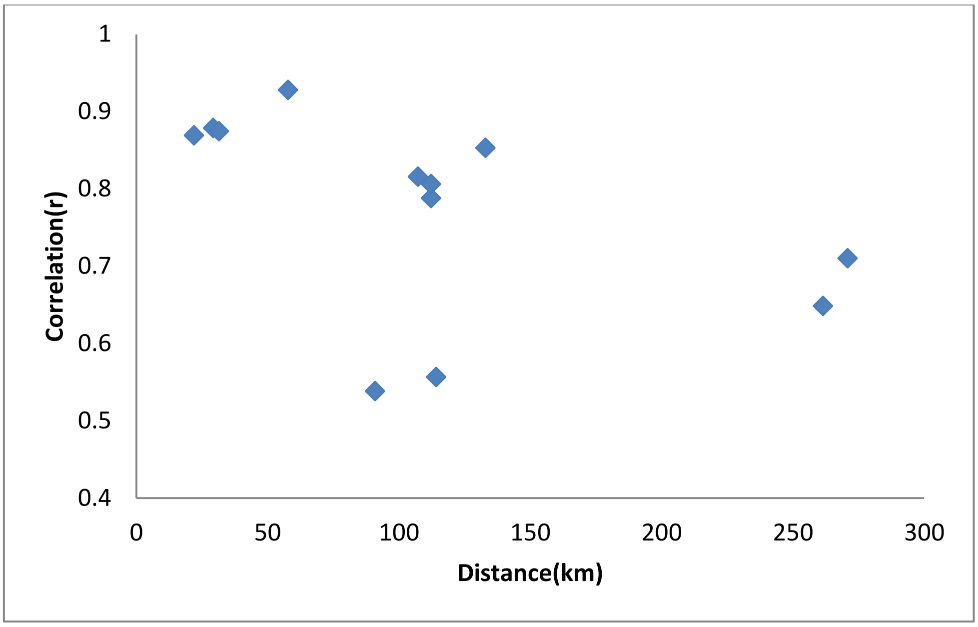

Figure B.5 Distance versus Correlation Relation of Streamgaging Station 1327

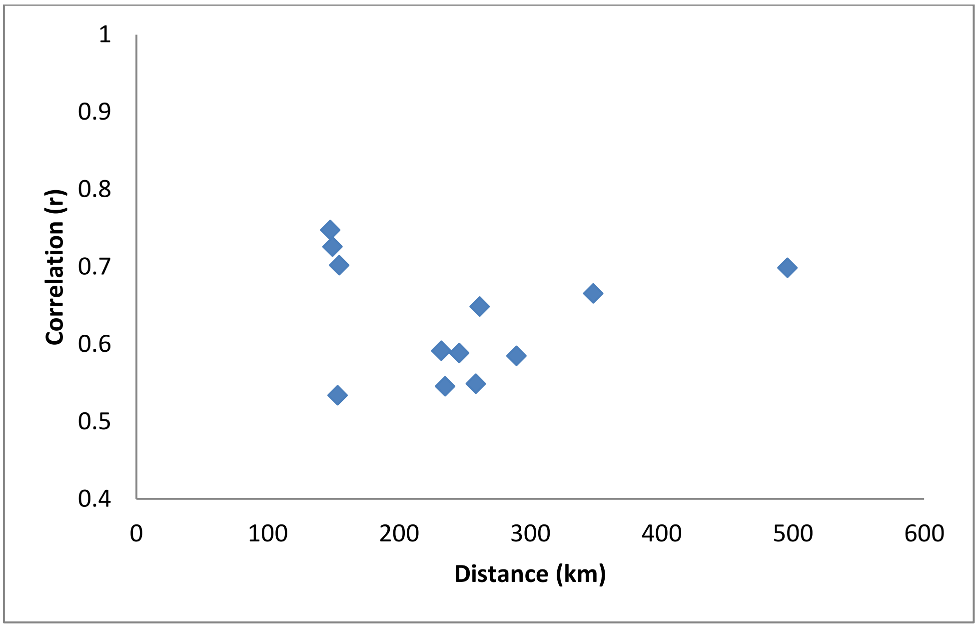

Figure B.6 Distance versus Correlation Relation of Streamgaging Station 1330

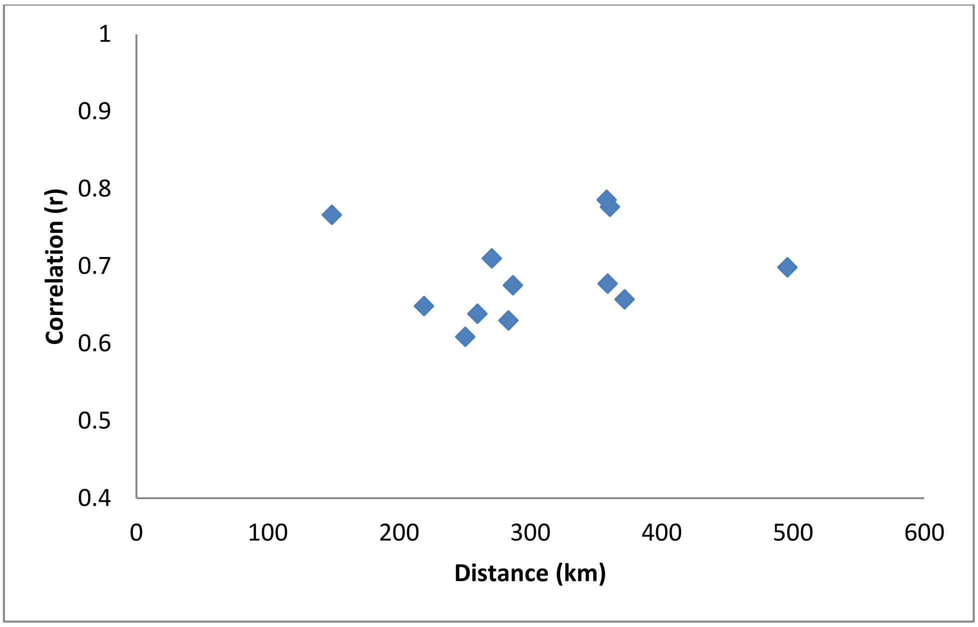

Figure B.7 Distance versus Correlation Relation of Streamgaging Station 1332

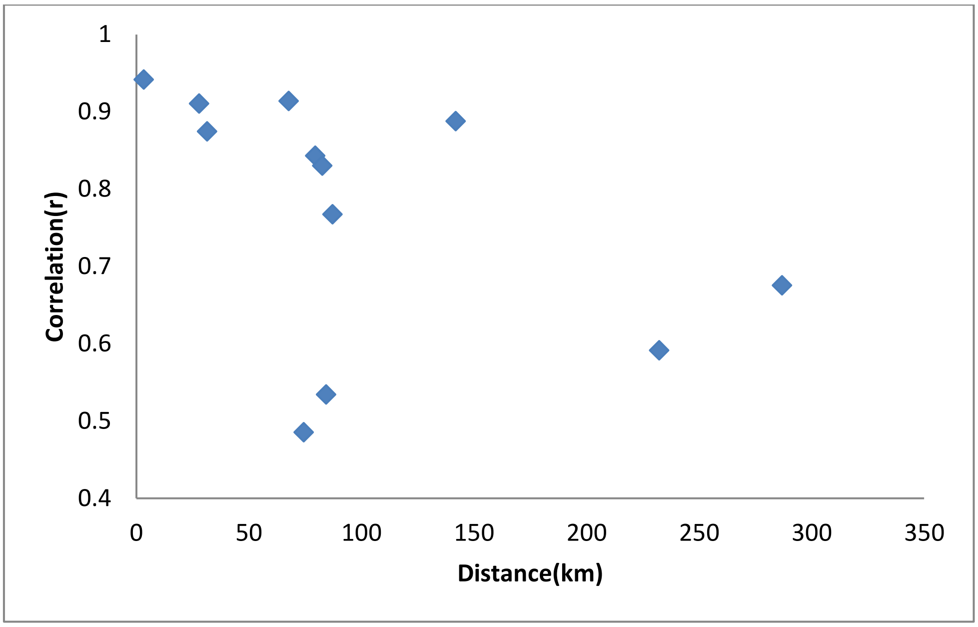

Figure B.8 Distance versus Correlation Relation of Streamgaging Station 1334

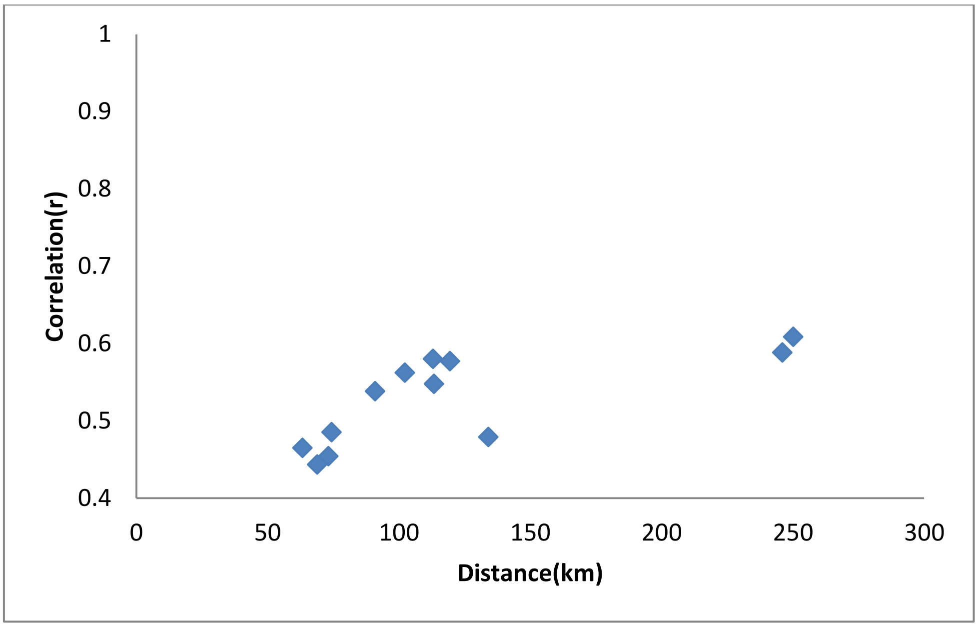

Figure B.9 Distance versus Correlation Relation of Streamgaging Station 1335

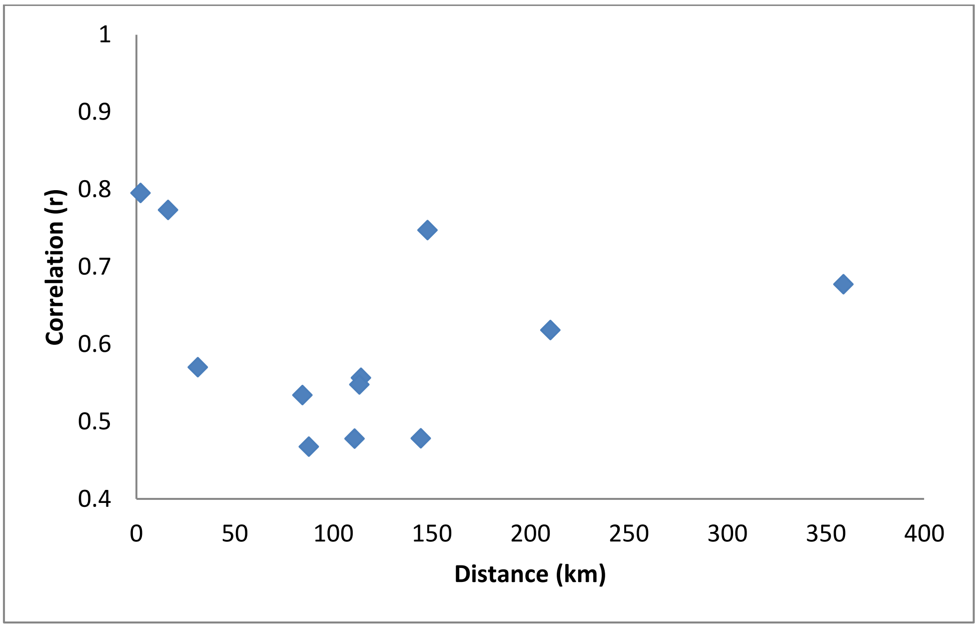

Figure B.10 Distance versus Correlation Relation of Streamgaging Station 1338

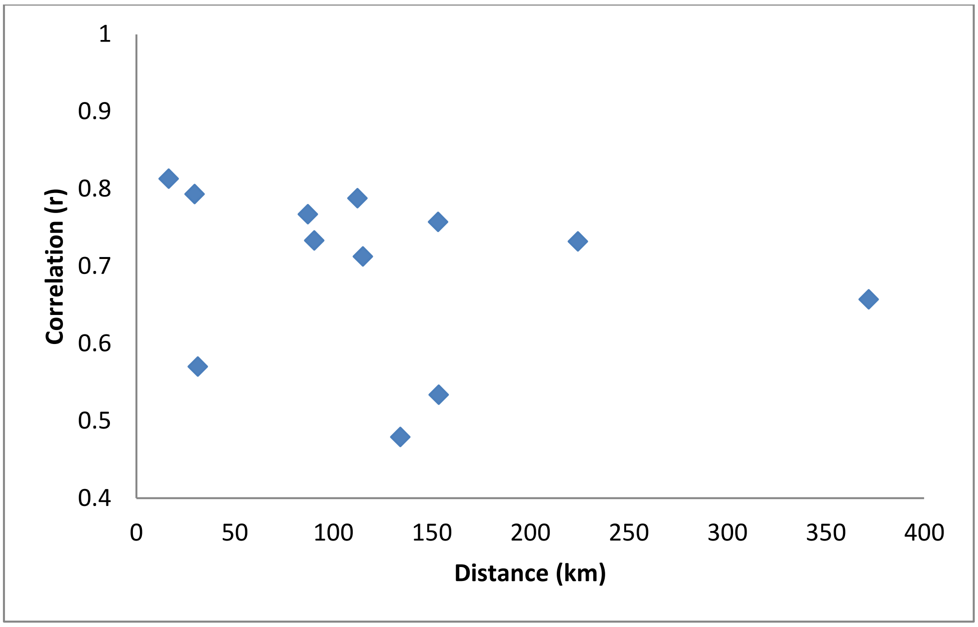

Figure B.11 Distance versus Correlation Relation of Streamgaging Station 1339

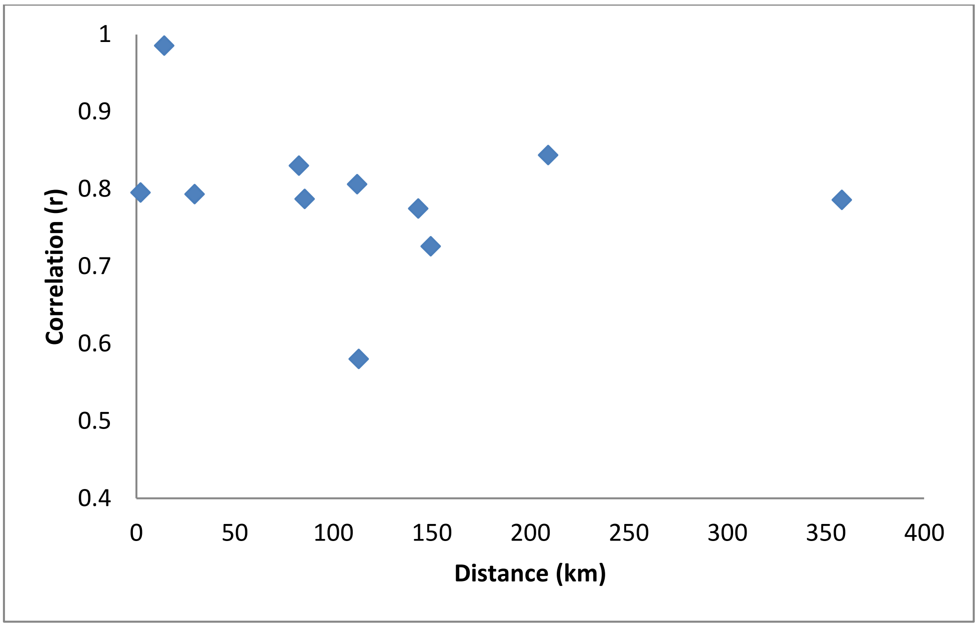

Figure B.12 Distance versus Correlation Relation of Streamgaging Station 1340

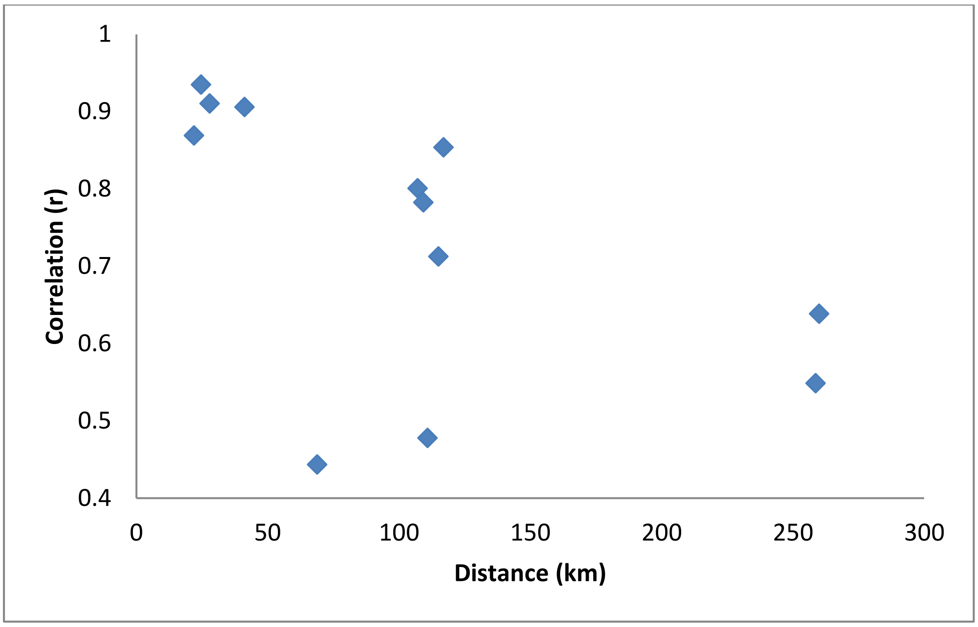

Figure B.13 Distance versus Correlation Relation of Streamgaging Station 1343

APPENDIX C

R SOFTWARE OUTPUTS

This section includes sample and model variograms that are obtained from R Statistical Software. Figures C.1 to C.10 provide the sample and the model variograms for remaining streamgaging stations in the Western Blacksea Basin.

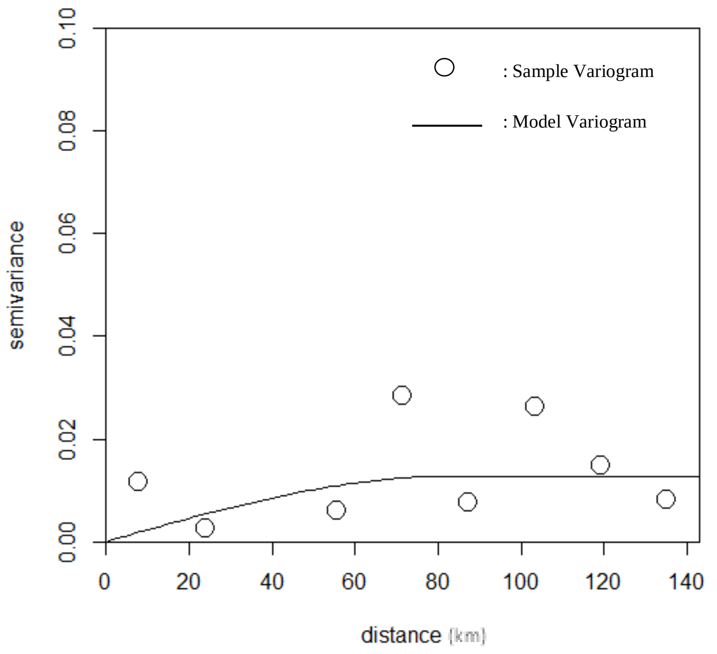

Figure C. 1 The Sample and The Model Variogram of Streamgaging Station 1302

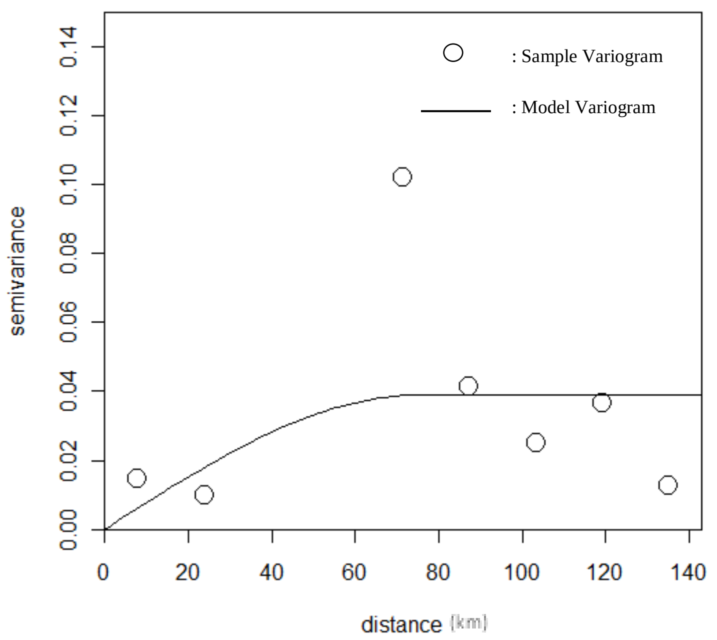

Figure C.2 The Sample and The Model Variogram of Streamgaging Station 1314

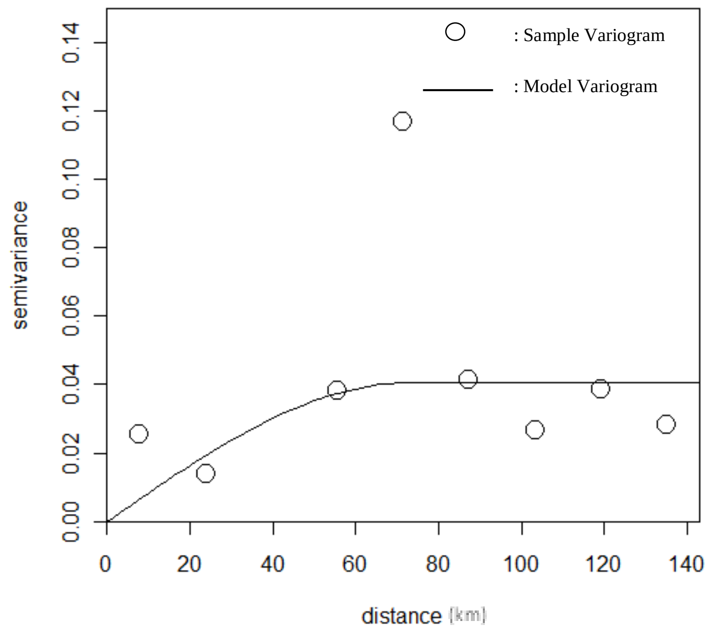

Figure C.3 The Sample and The Model Variogram of Streamgaging Station 1319

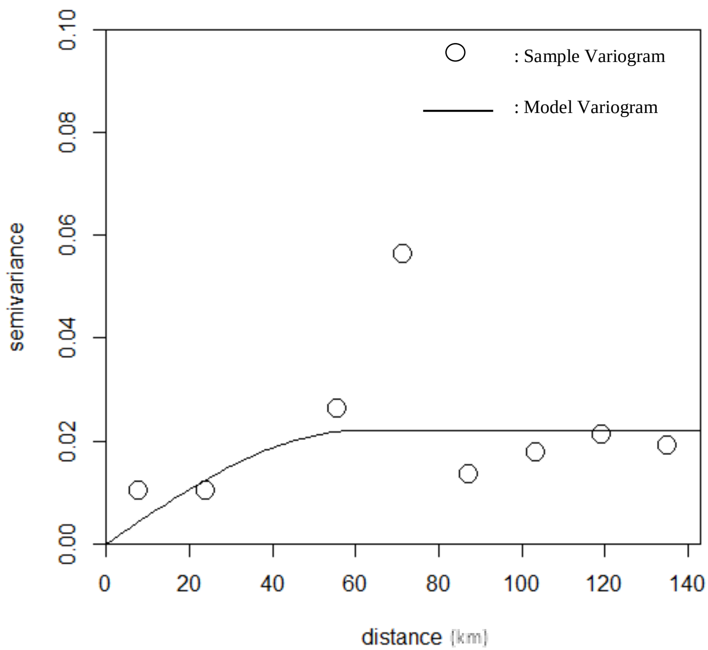

Figure C.4 The Sample and The Model Variogram of Streamgaging Station 1327

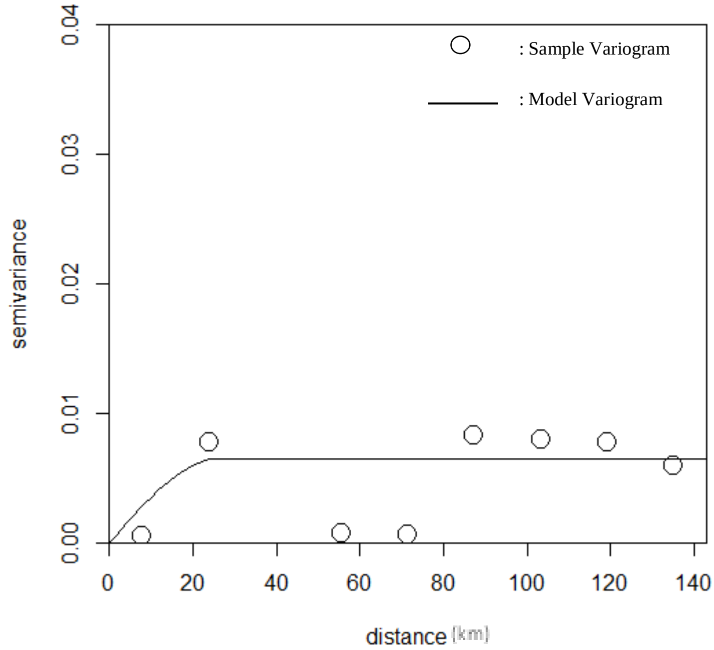

Figure C.5 The Sample and The Model Variogram of Streamgaging Station 1330

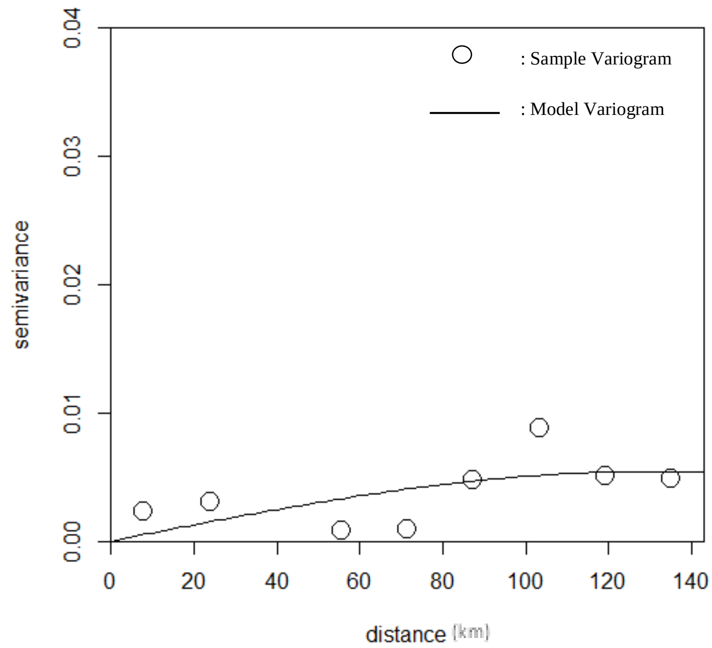

Figure C.6 The Sample and The Model Variogram of Streamgaging Station 1332

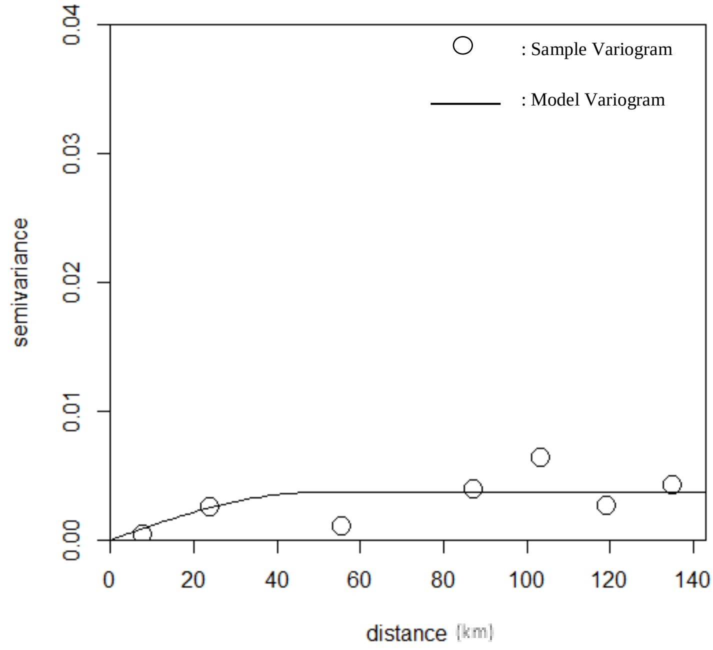

Figure C.7 The Sample and The Model Variogram of Streamgaging Station 1335

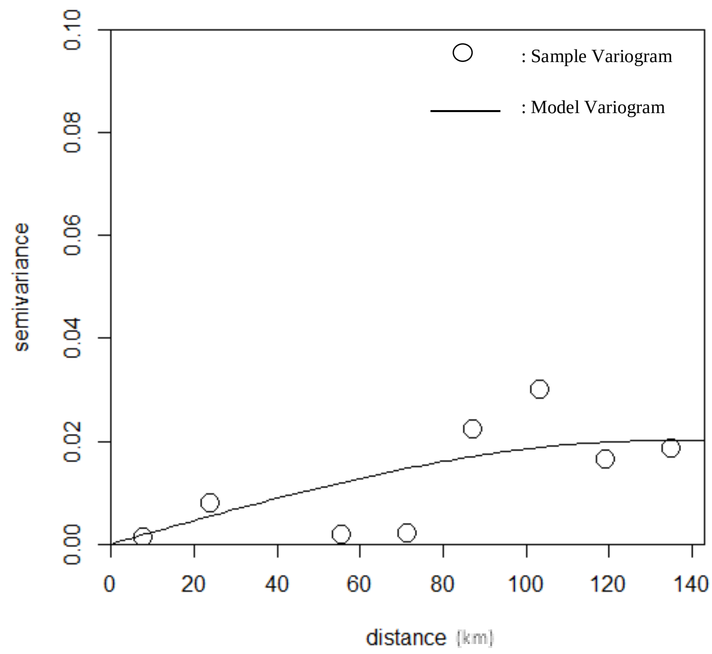

Figure C.8 The Sample and The Model Variogram of Streamgaging Station 1338

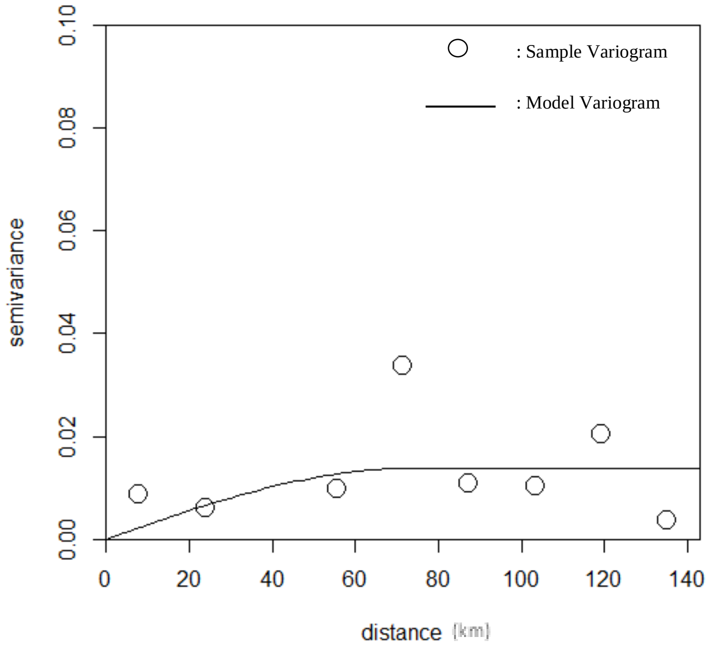

Figure C.9 The Sample and The Model Variogram of Streamgaging Station 1339

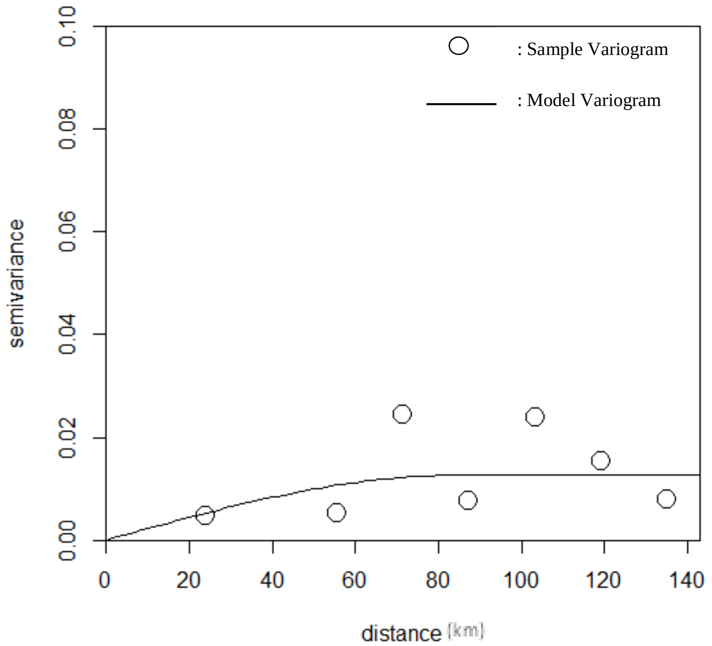

Figure C.10 The Sample and The Model Variogram of Streamgaging Station 1340

APPENDIX D

ORDINARY KRIGING CALCULATIONS

This section provides comparison of observed correlations and estimated correlations obtained from ordinary kriging analysis for each of the 13 streamgaging stations located in the study region (Tables D.1 to D.13).

Table D. 1 Comparison of Observed Correlations and Estimated Correlations Obtained by Ordinary Kriging for Streamgaging Station 1302

Streamgaging Stations	Observed Correlation	Estimated Correlation
1307	0.8440	0.7977
1314	0.7876	0.7730
1319	0.8003	0.7699
1327	0.8155	0.8040
1330	0.7016	0.6258
1332	0.7764	0.7274
1334	0.8429	0.8078
1335	0.5769	0.5363
1338	0.7731	0.6881
1339	0.8130	0.7750
1340	0.9852	0.7936
1343	0.8004	0.7572

Table D. 2 Comparison of Observed Correlations and Estimated Correlations Obtained by Ordinary Kriging for Streamgaging Station 1307

Streamgaging Stations	Observed Correlation	Estimated Correlation
1302	0.8440	0.7526
1314	0.8520	0.6751
1319	0.8621	0.6905
1327	0.8499	0.7326
1330	0.6651	0.6202
1332	0.7660	0.6448
1334	0.8874	0.7164
1335	0.5621	0.5348
1338	0.6181	0.5551
1339	0.7321	0.6571
1340	0.8434	0.7414
1343	0.8535	0.6823

Table D.3 Comparison of Observed Correlations and Estimated Correlations Obtained by Ordinary Kriging for Streamgaging Station 1314

Streamgaging Stations	Observed Correlation	Estimated Correlation
1302	0.7876	0.7581
1307	0.8520	0.7379
1319	0.9307	0.7264
1327	0.9276	0.7453
1330	0.5846	0.6295
1332	0.6481	0.6595
1334	0.9137	0.7407
1335	0.4649	0.5394
1338	0.4783	0.5613
1339	0.7571	0.6463
1340	0.7743	0.7538
1343	0.9058	0.6847

Table D. 4 Comparison of Observed Correlations and Estimated Correlations Obtained by Ordinary Kriging for Streamgaging Station 1319

Streamgaging Stations	Observed Correlation	Estimated Correlation
1302	0.8003	0.8372
1307	0.8621	0.8814
1314	0.9307	0.9132
1327	0.8784	0.8695
1330	0.5453	0.6287
1332	0.6298	0.6730
1334	0.9412	0.8195
1335	0.4540	0.4838
1338	0.4673	0.5281
1339	0.7331	0.7615
1340	0.7871	0.8229
1343	0.9345	0.9043

Table D.5 Comparison of Observed Correlations and Estimated Correlations Obtained by Ordinary Kriging for Streamgaging Station 1327

Streamgaging Stations	Observed Correlation	Estimated Correlation
1302	0.8155	0.7935
1307	0.8529	0.8118
1314	0.9276	0.8425
1319	0.8784	0.8368
1330	0.6485	0.6183
1332	0.7100	0.6336
1334	0.8744	0.8356
1335	0.5381	0.4635
1338	0.5563	0.4828
1339	0.7877	0.6849
1340	0.8057	0.7729
1343	0.8687	0.7287

Table D.6 Comparison of Observed Correlations and Estimated Correlations Obtained by Ordinary Kriging for Streamgaging Station 1330

Streamgaging Stations	Observed Correlation	Estimated Correlation
1302	0.7016	0.7731
1307	0.6651	0.7527
1314	0.5846	0.7200
1319	0.5453	0.7389
1327	0.6485	0.7634
1332	0.6981	0.6845
1334	0.5913	0.7607
1335	0.5883	0.5343
1338	0.7473	0.6199
1339	0.5335	0.6882
1340	0.7255	0.7709
1343	0.5485	0.7286

Table D. 7 Comparison of Observed Correlations and Estimated Correlations Obtained by Ordinary Kriging for Streamgaging Station 1332

Streamgaging Stations	Observed Correlation	Estimated Correlation
1302	0.7764	0.7593
1307	0.7660	0.7346
1314	0.6481	0.7093
1319	0.6298	0.7262
1327	0.7100	0.7540
1330	0.6981	0.6162
1334	0.6752	0.7481
1335	0.6087	0.5303
1338	0.6773	0.6361
1339	0.6570	0.6692
1340	0.7857	0.7595
1343	0.6382	0.7152

Table D.8 Comparison of Observed Correlations and Estimated Correlations Obtained by Ordinary Kriging for Streamgaging Station 1334

Streamgaging Stations	Observed Correlation	Estimated Correlation
1302	0.8429	0.7981
1307	0.8874	0.8528
1314	0.9137	0.9160
1319	0.9412	0.8078
1327	0.8744	0.8662
1330	0.5913	0.6189
1332	0.6752	0.6353
1335	0.4854	0.4619
1338	0.5340	0.4810
1339	0.7671	0.7280
1340	0.8300	0.7873
1343	0.9101	0.9163

Table D. 9 Comparison of Observed Correlations and Estimated Correlations Obtained by Ordinary Kriging for Streamgaging Station 1335

Streamgaging Stations	Observed Correlation	Estimated Correlation
1302	0.5769	0.7982
1307	0.5621	0.7979
1314	0.4649	0.7464
1319	0.4540	0.8464
1327	0.5381	0.8240
1330	0.5883	0.6321
1332	0.6087	0.6917
1334	0.4854	0.8461
1338	0.5477	0.6064
1339	0.4791	0.6910
1340	0.5798	0.7974
1343	0.4436	0.8231

Table D.10 Comparison of Observed Correlations and Estimated Correlations Obtained by Ordinary Kriging for Streamgaging Station 1338

Streamgaging Stations	Observed Correlation	Estimated Correlation
1302	0.7731	0.9775
1307	0.6181	0.8398
1314	0.4783	0.7717
1319	0.4673	0.7841
1327	0.5563	0.8036
1330	0.7473	0.6602
1332	0.6773	0.7820
1334	0.5340	0.8265
1335	0.5477	0.5786
1339	0.5699	0.7898
1340	0.7950	0.9369
1343	0.4778	0.7797

Table D. 11 Comparison of Observed Correlations and Estimated Correlations Obtained by Ordinary Kriging for Streamgaging Station 1339

Streamgaging stations	Observed Correlation	Estimated Correlation
1302	0.8130	0.8959
1307	0.7321	0.8183
1314	0.7571	0.7700
1319	0.7331	0.7822
1327	0.7877	0.8021
1330	0.5335	0.6881
1332	0.6570	0.7521
1334	0.7671	0.8201
1335	0.4791	0.5706
1338	0.5699	0.7453
1340	0.7931	0.9358
1343	0.7124	0.7819

Table D.12 Comparison of Observed Correlations and Estimated Correlations Obtained by Ordinary Kriging for Streamgaging Station 1340

Streamgaging Stations	Observed Correlation	Estimated Correlation
1302	0.9852	0.7760
1307	0.8434	0.6489
1314	0.7743	0.5193
1319	0.7871	0.5119
1327	0.8057	0.5907
1330	0.7255	0.6695
1332	0.7857	0.6910
1334	0.8300	0.5752
1335	0.5798	0.5496
1338	0.7950	0.7523
1339	0.7931	0.6021
1343	0.7825	0.5210

Table D.13 Comparison of Observed Correlations and Estimated Correlations Obtained by Ordinary Kriging for Streamgaging Station 1343

Streamgaging Stations	Observed Correlation	Estimated Correlation
1302	0.8004	0.7961
1307	0.8535	0.8470
1314	0.9058	0.8727
1319	0.9345	0.8979
1327	0.8687	0.8043
1330	0.5485	0.6355
1332	0.6382	0.6928
1334	0.9101	0.8946
1335	0.4436	0.4850
1338	0.4778	0.5063
1339	0.7124	0.7477
1340	0.7825	0.7848