DEVELOPMENT OF A GIS-BASED ESTIMATOR OF STREAMFLOW DATA AT UNGAGED CATCHMENTS

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

BY

DUYGU ÖCAL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
GEODETIC AND GEOGRAPHICAL INFORMATION TECHNOLOGIES

SEPTEMBER 2017

Approval of the thesis:

DEVELOPMENT OF A GIS-BASED ESTIMATOR OF STREAMFLOW DATA AT UNGAGED CATCHMENTS

submitted by **DUYGU ÖCAL** in partial fulfillment of the requirements for the degree of **Master of Science in Geodetic and Geographical Information Technologies Department, Middle East Technical University** by,

Prof. Dr. Gülbin DURAL ÜNVER
Dean, Graduate School of **Natural and Applied Sciences** ________________

Assoc. Prof. Dr. Uğur Murat LELOĞLU
Head of Department,
Geodetic and Geographical Information Technologies ________________

Prof. Dr. Elçin KENTEL ERDOĞAN
Supervisor, **Civil Engineering Dept.** ________________

Examining Committee Members:

Assoc. Prof. Dr. Ismail YUCEL
Civil Engineering Dept., METU ________________

Prof. Dr. Elçin KENTEL ERDOĞAN
Civil Engineering Dept., METU ________________

Prof. Dr. Ayşen DENER AKKAYA
Department of Statistics, METU ________________

Asst. Prof. Dr. M. Tuğrul YILMAZ
Civil Engineering Dept., METU ________________

Assoc. Prof. Dr. Aynur ŞENSOY SORMAN
Civil Engineering Dept., Anadolu University ________________

Date: 12.09.2017

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 materials and results that are not original to this work.

Name, Last name: Duygu ÖCAL
Signature:

ABSTRACT

DEVELOPMENT OF A GIS-BASED ESTIMATOR FOR STREAMFLOW AT UNGAGED CATCHMENTS

ÖCAL, Duygu
M.Sc., Department of Geodetic and Geographic Information Technologies
Supervisor: Prof. Dr. Elçin KENTEL ERDOĞAN

September 2017, 144 pages

Water resources management has been a critical component of sustainable resources planning. One of the most commonly used data in water resources management is streamflow measurements. Daily streamflow time series collected at a streamgage provide information on the temporal variation in water quantity at the gage location. However, streamflow information is often needed at ungaged catchments. One conventional approach to estimate streamflow at an ungaged catchment is to transfer streamflow measurements from the spatially closest streamgage. Recently, the correlation between daily streamflow time series is proposed as an alternative to distance for reference streamgage selection. The Map Correlation Method (MCM) enables development of a map that demonstrates the spatial distribution of correlation coefficients between daily streamflow time series at a selected streamgage and all other locations within a selected study area. Due to its geostatistical analysis procedure MCM is time-consuming and hard to implement for practical purposes such as installed capacity selection of run-of-river hydropower plants during their feasibility studies. In this study, an easy-to-use GIS-based tool, called MCM_GIS is developed to apply the MCM. MCM_GIS provides a user-friendly working environment and flexibility in choosing between two types of interpolation models, kriging and inverse distance weighting. The main motivation of this study is to increase practical application of the MCM by integrating it to the GIS environment. MCM_GIS can also carry out the leave-one-out cross-validation scheme to monitor the overall performance of the estimation. The tool is tested on two study area; Western Black Sea Region and Çoruh Basin, Turkey. ArcGIS for Desktop product along with a Python script is utilized. The outcomes of inverse distance weighting and ordinary kriging are compared, no significant difference between the two interpolation methods was observed. Results of GIS-based MCM are in good agreement with the observed hydrographs according to NSE values.

Keywords: Map Correlation Method, GIS, Water Resources, Kriging, Spatial Statistics

ÖZ

AKARSU AKIM VERİSİ OLMAYAN HAVZALARDA GIS TABANLI AKIM TAHMİNİ YAPAN BİR ARACIN GELİŞTİRİLMESİ

ÖCAL, Duygu
Yüksek Lisans, Jeodezi ve Coğrafi Bilgi Teknolojileri Bölümü
Tez Danışmanı: Prof. Dr. Elçin KENTEL ERDOĞAN
Eylül 2017, 144 sayfa

Su kaynakları yönetimi, sürdürülebilir kaynak planlamasının kritik bir bileşenidir. Su kaynakları yönetiminde en yaygın kullanılan verilerden biri de günlük akım ölçümleridir. Bir akım gözlem istasyonunda (AGİ) toplanan günlük akım zaman serisi, AGİ'nin bulunduğu konumdaki su miktarındaki zamansal değişime ilişkin bilgi sağlar. Yakın zamanda, günlük akım zaman serileri arasındaki korelasyon, referans AGİ seçimi için mesafeye alternatif olarak önerilmiştir. Harita Korelasyon Metodu (MCM), seçilen bir akım gözlem istasyonundaki günlük akım serisi ile seçilen çalışma alanı içindeki diğer noktalar arasındaki korelasyon katsayılarının mekansal dağılımını gösteren bir haritanın geliştirilmesini sağlar. Harita korelasyon metodunun kullanılabilirliği çeşitli çalışmalarda gösterilmiş olmasına rağmen, metodun uygulanması zaman alıcı ve zordur. Bu çalışmada, harita korelasyon metoduyla günlük akım serilerini tahmin etmede kullanmak üzere MCM_GIS olarak adlandırılan kullanımı kolay bir CBS tabanlı araç geliştirilmiştir. MCM_GIS, Kriging ve ters mesafe ağırlıklandırma (IDW) olmak üzere iki tür enterpolasyon modeli arasında seçim imkanı tanıyan kullanıcı dostu bir çalışma ortamı ve esneklik sağlar. Bu çalışmanın ana motivasyonu, harita korelasyon metodunun CBS ortamına entegre edilerek pratik problemlerde kullanımını arttırmaktır. MCM_GIS, tahminin genel performansını izlemek için, çapraz doğrulama işlemi de gerçekleştirebilmektedir. Araç, Batı Karadeniz Havzası ve Çoruh Havzasında yapılan uygulamar ile de test edilmiştir. ArcGIS for Desktop ürünü ve bir Python komut dosyası kullanılarak araçlar geliştirilmiştir. IDW ve Kriging ile elde edilen sonuçlar karşılaştırılmıştır. İki metod arasında önemli bir fark gözlemlenmemiştir. CBS tabanlı MCM'nin sonuçları gözlemlenmiş olan hidrograflarla, NSE değerleri bakımından uyum içindedir.

Anahtar Kelimeler: Harita Korelasyon Metodu, CBS, Su Kaynakları, Kriging, Mekansal İstatistik

To My Mother and Father,

ACKNOWLEDGEMENTS

First of all, I would like to express my graditude to my supervisor Prof. Dr. Elçin Kentel, for her constant support, guidance and patience throughout my thesis study. Our discussions during my graduate studies were my main motivation to pursue this thesis.

Secondly, I am very grateful to Ms. Fatma Sema Kandır for providing the data used in the application on case study. Testing of the GIS Tools were possible thanks to her. I hope people like Ms. Sema Kandır will contribute to many scientific research without obstacles. Also, I would like to thank Mr. Barış Uz for providing me ArcGIS for Desktop 10.5 license which was very valuable for my research.

During my graduate years, I had the opportunity to work in different companies related with GIS. Working and writing a thesis can be very challenging. Therefore, I want to thank my coworkers in NFB Mühendislik, HAVELSAN and ESRI Turkey for cheering me up and supporting me in difficult working hours.

I would also like to thank my friends Yeşim Can, Başak Atan, Selim Hikmet Şahin and Zeynep Özçam for always keeping me entertained but also focused in my studies.

I am grateful for all of Öcal family members for their love, care and jokes. Especially my sister Deniz Öcal, my mother Mihriye Öcal and my father Oktay Öcal. I am very lucky to be a part of this family and have their support.

Finally, I want to express my gratitude to Ekrem Bilgehan Uyar who is very dear to me. He has always been an inspiration to me and encouraged me during this study. His help on this thesis is always appreciated and I can’t thank him enough for the support he gave me.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

LIST OF TABLES

TABLES

LIST OF ABBREVIATIONS

DAR	: Drainage Area Ratio
DBMS	: Database Management System
GDB	: Geodatabase
GIS	: Geographic Information System
IDW	: Inverse Distance Weighting
MCM	: Map Correlation Method
OK	: Ordinary Kriging
SG	: Streamgage
SHPP	: Small Hydro-Power Plants

CHAPTER 1

INTRODUCTION

1.1. Problem Definition

The population and level of urbanization are growing at a high rate not only in Turkey but also in the world. These developments naturally result in greater demands for energy and water resources. Therefore, water resources projects should be handled carefully to effectively respond to existing and future needs. Due to its large potential, water resources are especially important for sustainable development of Turkey.

The design and implementation of projects like hydraulic structures require streamflow time series. When there is a streamgage at the point of interest, its data can be used directly. However, most of the hydraulic structures are planned at locations where no observation exists (i.e. ungaged basins). Thus, streamflow data need to be estimated at ungaged basins. The most widely used method for estimating daily streamflow time series is the drainage area-ratio (DAR) method where streamflow data of a reference streamgage is transferred to the ungaged basin using the drainage area ratios of these two basins. Consequently, the selection of a reference streamgage is very important.

In many recent studies, the reference streamgage is selected as the spatially closest 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, there are other approaches claiming the distance between gauged and ungaged locations is not be the best criteria for selecting the reference streamgage and may not provide the most accurate results. One such aproach, called the Map Correlation Method (MCM), is introduced by Archfield and Vogel (2010) and further studies were conducted using this framework (Ergen and Kentel 2016, Ergen 2012, Ocal and Kentel 2017, Monjardin et al 2017). MCM generates a correlation contour map for streamflow of the ungaged location with its surrounding, that allows selection of the reference streamgage based on the correlations between streamflows. Archfield and Vogel (2010) sucsessfully applied this method to a study area located around New England, USA with 28 streamgaging stations. Their findings showed that MCM generally provides improved estimates of daily streamflow data over those obtained from the closest streamgaging stations used as reference gages. Then, Ergen (2012) used the MCM in the Western Blacksea Basin and concluded that the method has potential to be used in the selected study area. However, one major drawback of the MCM in terms of its utilization for practical purposes is its time-consuming and relatively complex statistical procedure. As a remedy, a GIS tool is developed to carry out the MCM in this study.

1.2 The GIS Tool

The main aim of this study is to develop a GIS tool to facilitate use of MCM method for practical purposes. The integration of MCM with GIS eases application of time-consuming and complex statistical procedures of the MCM, provides a user-friendly tool for different types of users and allows easy implementation of various geospatial analysis. It helps decision makers in visualizing the outcomes and making more informed decisions. The newly develped GIS-based tool is tested in the Western Black Sea Region and Coruh Basin and its surrounding area.

The GIS tool developed in this study identifies the most correlated streamgage to be used as the reference gage in the drainage area-ratio method to estimate streamflow at ungaged catchments. Major contributions of the GIS tool are as follows:

1. The GIS tool fastens the process of streamflow data estimation with the MCM and has an easy to use interface both for GIS and non-GIS professionals. It is believed that the GIS tool will increase utilization of the MCM in streamflow

estimation and contribute to more accurate decision making related with planning of water resources. The tool is expected to be most beneficial in the feasibility analysis for water structures such as small hydropower plants in Turkey.

2. A leave-one-out cross-validation process in integrated into the GIS tool to assess the performance of the estimation in the selected study area. When the efficiency is poor, estimation can be repeated easily with larger number of streamgages and/or longer observation periods using the tool.

This thesis is composed of five chapters. This first chapter gives the problem definition, the motivation of the study and contributions of this study. The second chapter, Literature Review and Background Information, gives background information on geostatistics, hydrologic modelling and GIS usage in hydrology. The third chapter is, methodology where spatial estimation techniques used for streamflow data are explained in detail. Moreover, the GIS tool is introduced and implementation of the MCM with this tool is explained. Case studies are performed in chapter four and the results and discussions are provided. Finally, conclusions are presented in the last chapter.

CHAPTER 2

LITERATURE REVIEW AND BACKGROUND INFORMATION

In order to predict streamflow data at ungaged basins, hydrological models are widely used. However, input data requirements and model calibration are two main difficulties faced during development a hydrological model. As an alternative, Archfied and Vogel (2010) developed MCM to estimate daily streamflow at ungaged basins. MCM uses statistical methods to predict streamflow data at ungaged catchments based solely on streamflow measurements of existing streamgages in the vicinity of the ungaged location. In this method, cross-correlation maps are generated using ordinary kriging. Since MCM is based on spatial statistics, GIS appears to be a natural tool to implement this approach. In this thesis, a GIS tool, MCM_GIS, to implement MCM which is based on geostatistics is developed. The GIS tool is used to carry out the analysis, which combines database management systems along with a spatial analysis-working environment. Therefore, background information about hydrological modeling, spatial statistics and GIS is provided in this chapter to build the necessary background required to understand the development of the GIS tool.

2.1 Spatial Statistics

Spatial statistics, also referred to as geostatistics, were first discussed by Kolmogorov and Wiener in the 1940's (Kolmogorov, 1941; Wiener, 1949) and it continues to draw attention from different disciplines, especially earth sciences. Spatial statistics deals with different ways of interpreting variations of data in a spatial framework. Spatial data analysis first aims to understand an ongoing process in space and develop a model based on random functions that describe or explain the behavior of this process. Later on, having understood the spatial phenomena in statistical terms, spatial statistics tries to identify and formulate the possible relationship to other spatial processes of these phenomena. The main goal of such examination is to increase the understanding of a spatial process to predict values in areas where there are no observations available.

After the introduction of spatial statistics, it was improved by Matern (1963), Whittle (1963), Matheron (1965) and others (Journel, 1989). Their common contribution to the subject was to carry out statistical data analysis in spatial dimension and interpret results such as data variations.

In 1950's, a South African mining engineer and a statistician, D.G. Krige, developed a new technique that can be used for the estimation of ore grades by using spatial correlation in spatial data. This development created big interest in the field and led to the development of the concept by Huijbregts and Matheron (1971) who studied this technique and formalized it in the 1960's. This technique later on called as "kriging" named after D.G. Krige. Huijbregts and Matheron (1971) used geostatistics for estimation of mineral deposits by applying the theory of regionalized variables in Algeria. Studies in spatial statistics accelerated after the random functions are introduced and produced promising results. As a result, usage of the spatial statistics expanded from mining geology to other areas like hydrology. The research on geostatistics attracted other fields because of its quantitative nature some of which are, B.Matern (1960's) on forestry and Hardy and Göpfert (1975) in geodesy.

Hydrology was one of the early application fields of spatial statistics following mining engineering. One of the main study areas of hydrology, considering the contribution of geostatistics, is groundwater hydrology. Some of the first applications include Delhomme (1974). Afterwards, many other studies have been conducted such as, Delhomme (1978), Gambolati and Volpi (1979), Mizell et al. (1980), Darricau-Beucher (1981), Neuman and Jacobson (1984), Hoeksema and Kitanidis (1984), Thangarajan and Ahmed (1989), Ahmed and Murali (1992), Roth (1995), and Roth et al. (1996) to demonstrate the use of geostatistics in groundwater hydrology.

Geostatistical analysis are similar to regression analysis, but there are small differences between them. Regression analysis produces a statistical measure that determines level of relationship between one dependent variable and other independent variables. Correlation is the measure that quantifies this relationship. Geostatistics also uses correlation values, but the motivation is to determine the locations of the variables as well. For example, regression analysis can be used to determine the relationship between air pollution and lung cancer. Geostatistical correlation analysis can determine, which locations are most risky for lung cancer based on air pollution data.

Spatial statistics is a good working field for geoscientists. It helps researchers to model, monitor and test spatial phenomena. It is also possible to examine spatial distributions, patterns, connectivity and other spatial relationships of geographical features. Naturally, it is possible to determine to what extent a correlation exists between two observation points, spatially. Common area of geostatistics is data estimation. One alternative to spatial data estimation is done by producing interpolation maps. There are many different areal interpolation algorithms that are used in GIS programs. Some of the most used ones are explained below and detailed information on inverse distance weighting (IDW) and ordinary kriging (OK) are given in Section 3.5 Map Correlation Method Application for the Ungaged Basin).

2.2 Areal Interpolation

Interpolation is a mathematical term that describes a way to construct new data within a range of a set of known data. It is useful where there are available dataset around the unknown data point and the trend of the existing data is suitable for investigation. Interpolation applications coincides with study areas of mathematics, statistics, engineering and social sciences. The details of applications in geosciences are explained below.

Spatial interpolation is a widely used method for data estimation where a modeler is interested in a value at an unmeasured location which is surrounded with similar observations. It is useful for data visualization, spatial analyses especially in environmental sciences (Meng et al, 2013). Recently, large number of variations of spatial interpolation methods have been applied and developed. Existing remote sensing and GIS softwares provide ready to use algorithms that perform spatial interpolation. Some of these methods are kriging (e.g. universal kriging, ordinary kriging, regression kriging, cokriging), spline, radial basis and inverse distance weighting (IDW). Cokriging, universal kriging, ordinary kriging corresponds to stochastic interpolation methods, while IDW, spline and radial basis function are considered as deterministic methods. The GIS tool developed in this study allows the user to choose IDW or ordinary kriging (OK) as the spatial interpolation method. The details of these methods are provided in the Methodology Chapter.

Kriging is preferred by the data scientists for two main reasons. The first one is that it enables a more precise estimate because of the usage of surrounding data. Second, kriging is formulized to minimize the estimation variance. This aspect of kriging is one of the most remarkable difference from other estimations because it enables to monitor the uncertainty of the estimate.

Climatology studies benefited from geostatistical methods for estimating rainfall data. Performance of different kriging types are compared and as a result, kriging with an external drift is identified as the most advantageous for rainfall mean data (Pardo-Iguzquıza E., 1998, Goovaerts 1999b). Watershed management and hydrological modeling often require a preliminary spatial interpolation as part of the modeling process. Different spatial interpolation techniques were reviewed on rainfall data such as kriging and IDW (Ly et al, 2013). As a result, the performance of the interpolation methods depended on, temporal and spatial resolutions of the data and the parameters of the models, such as the semi-variogram in the case of kriging. Another comparison on the interpolation of hydrologic variables was between formal equivalence of multiquadratic surface fitting and kriging. Hourly rainfall maps of real storm events were used as the input data by Borga and Vizzaccaro (1997). Kriging is found to be very useful at low streamgage density while at high gage density the increase in accuracy of estimated fields is lower (Bacchi and Borga, 1995).

Geostatistics and areal interpolation provide smaller error variance than the error variance of other interpolation techniques which ignores all secondary information (Journel and Huijbregts 1978; Goovaerts 1997, 1999a, 1999b). Spatial interpolation is applied because it allows GIS analysts to take into account the spatial dependence between observations to estimate values at unsampled locations.

Areal interpolation algorithms are used to estimate different spatial variables. In the MCM, correlations between streamflow data are used as the input and areal interpolation algorithms are used to produce correlation maps where correlation values between streamflow measurements are represented as contours. By using these correlation maps, the correlation values at the ungaged locations are predicted.

2.3 Streamflow Estimation at Ungaged Basins

Hydrology is an important research area that studies earth's water and provides significant information on the utilization of water resources. It deals with the occurrence, circulation, distribution of the water in the environment.

Streamflow is one of the most important data that is used in hydrology. Whereas, it is difficult, actually most of the time, impossible to obtain this data at the location where it is needed because the streamgage network has limited number of streamgages. In practice, due to its simplicity, streamflow data is transferred from a spatially close reference streamgage to the ungaged location using the drainage area ratio method (DAR). This method transfers the daily streamflow time series from the gaged location to the ungaged location, by using the ratio of drainage areas of these two points. But, this method relies on the selection of an appropriate reference streamgage. Therefore, selection criteria become very important and affects the outcome significantly.

There are two main approaches that are used to estimate streamflow at ungaged basins. One is building a hydrological model using site-specific data. The other approach is using geostatistical methods with existing streamflow data that are collected at streamgages surrounding the unmeasured location.

2.3.1 Hydrological Modeling

According to Moradkhani and Sorooshian (2008), modelling is a clear and simple representation of the processes in the real world. According to this definition, a good model should predict a real world event with minimum variation and should not be complex. Hydrological model, same as any other model, tries to predict a system's behavior and its responses to various disturbances. A runoff model, which include streamflow data, is generally composed of a group of equations that estimate the runoff as a function of various parameters that are used for defining the watershed characteristics. The main inputs to this model are rainfall and drainage area data. Other inputs that can be considered include, soil properties, vegetation coverage, topography, ground water aquifer and soil moisture information (Devia et al., 2015).

Streamflow prediction is a benchmark issue for professionals in the fields of hydrology and water resources management. Hydrological models are commonly used to estimate streamflow data which is necessary for the design of hydraulic structures such as small hydropower plants (SSHP). Calibration of a hydrological model is as important as the construction phase and requires additional work (Pesti et al, 1996; Archfield, 2009; Ozdemir, 2016). Hydrological systems demonstrate a large scale of spatial diversity in their characteristics (Grayson and Bloeschl, 2000). Therefore, a regionalized perspective is one of the research areas that have been studied (Ajami, Newsha K., et al, 2004). Of course, with increasing attention to water quality, environmental and ecological concerns in water resources management, daily streamflow data started to be used instead of monthly streamflow data. This situation brought up several difficulties concerning data availability. For example, some of the available time series have gaps due to missing data usually because of unavailability of measurement (Hughes and Smakhtin, 1996). Time series from different locations within same basin may not coincide in time intervals and may show very different characteristics due to seasonal affects. These factors made calibration of hydrological models a challenging task. Several alternative methods to calibrate a hydrological model and create more accurate outputs such as, fuzzy sets are heuristic optimization algorithms have been introduced. Recently, rainfall forecasting, (Yu and Chen, 2000), groundwater simulation, (Dou et al., 1999: Schulz and Huwe, 1997), drought analysis (Pesti et al., 1996) and streamflow prediction (Chang and Chen, 2001; Kentel, 2009) made use of fuzzy sets and artificial neural networks.

Estimating streamflow using a hydrological model is a complicated process since it requires a well calibrated and validated hydrological model. The hydrological models require extensive spatial inputs such as meteorological parameters (i.e. rainfall, evaporation, wind speed, humidity, soil moisture content, etc.), soil types, slope, digital elevation models and vegetation and land use data. The hydrological model requires careful calibration and validation that are often time-consuming studies. Also, it is an important aspect to have an indicator about the uncertainty of the analysis so that they can be used to guide water management policies. Semi-distributed models are claimed to be easier to setup and require relatively shorter processing times compared to distributed hydrological models according to the studies of Phd Thesis of Ozdemir, 2016. However, even this type of modeling requires all of the components that are explained above. Field studies are needed in the case of absence of ready to use hydrological and geological data. Then, pre-processing of the data that will be an input to the modeling software in the proper format is required. Afterwards, building the model parameters, calibration and lastly the validation procedures are all mandatory steps to get a reliable output.

Development of hydrological models in Turkey is very hard due to unavailability of necessary input data. Streamgage and rain gage networks are not well designed and dense enough, so, most of the time do not provide necessary data for proper calibration. In addition, land-use and soil maps are not up to date and verified. Thus, application of hydrological models in Turkey is a challenging task (Ozcan et al., 2017; Keskin, 2007 (M.Sc. Thesis); Ozdemir, 2016 (Ph.D. Thesis)). Thus, alternative approaches to hydrological modeling are believed to be beneficial and useful especially for initial estimations of streamflow at ungaged locations. Such initial estimations will be necessary, for example, for evaluating the feasibility of small hydropower plants. One alternative is the MCM which is a geostatistical approach based on streamflow data of surrounding streamgages.

2.3.2 Geostatistical Modeling

Geospatial methods demonstrate promising results for streamflow estimation. Map correlation method (MCM) benefits from the geostatistical aspect of the streamflow time series (Archfield and Vogel, 2010). The application of MCM on different test sites produced promising results (Ergen and Kentel, 2016; Ocal and Kentel, 2017; Monjardin et al, 2017). In the MCM, daily streamflow time series are used to create a correlation matrix between the observed locations and a correlation value is estimated for the ungaged location. Therefore, daily streamflow time series and the coordinates of the observations are the two main inputs for this method. MCM procedure is naturally compatible with GIS since it is based on spatial statistics. Moreover, MCM will allow estimation of streamflow at ungaged locations with only streamflow measurements at streamgages found in the vicinity of the ungaged location instead of spatial meteorological data, soil and land use maps required by hydrological models. MCM is not suggested to replace hydrological models but is presented as an easy-to-implement alternative, especially for cases where first approximations for streamflow is required such as feasibility studies for various hydraulic structures such as small hydropower plants.

Archfield and Vogel (2010) proposed MCM to estimate daily stream flow data at ungaged basins using correlation coefficients between daily stream flow data collected at a number of streamgages within a study area. Later on, Ergen and Kentel (2016) investigated applicability of the MCM in the Western Black Sea Basin, Turkey and obtained satisfactory results. Moreover, GIS environment has been integrated with MCM and visualization of results have improved and application of the method became easier (Ocal and Kentel, 2017; Monjardin et al, 2017).

Main advantage of the MCM is its minimal data requirement. The inputs are locations of streamgages and the daily streamflow time series at the streamgages found around the ungaged location. These data can be obtained easily from the streamgage network databases. Even though, MCM simplifies the streamflow estimation process, manually calculating and adjusting the model still can be difficult. GIS can reduce the amount of time spent to apply the MCM and eases performance of the validation of the estimation. Provided user interface can enable analysts to perform streamflow estimation through GIS without the requirement of previous experience in GIS. Thus, it is expected that the GIS tool will increase implementation of the MCM for practical purposes.

2.4. GIS as a Spatial Analysis Tool

Geographical Information System (GIS) has many different definitions. In practice, it is an integrated system that includes database management systems (DBMS) and spatial features. This integration creates an environment which spatial and statistical analysis can be carried out together. The geographical data element is used as a reference to the attribute which is stored in the DBMS. For example, province boundaries, river networks are references for population counts and streamflows (Maguire, 1991).

Fast and easy application interfaces of different GIS softwares are suitable working environments for hydrologists. Different hydrological models can be created in the GIS environment. Processing and pre-processing of hydrological data such as soil type, rainfall, slope, and elevation are possible in the GIS environment as well. GIS enables linking spatial data with non-spatial data and performance of various spatial analysis. It is also very useful to visualize outputs of analysis. It can demonstrate complex problems in simpler forms both as a map and as a table which can be connected to a database.

There are several different areas of interest, when it comes to GIS usage in hydrology. First noticeable one is groundwater analysis. As mentioned before, groundwater hydrology was one of the application fields of geostatistics. Recently, GIS is widely used to create digital geographic databases to shape and prepare input data and to visualize model outputs. Hence, to understand the spatial variation of groundwater quality parameters GIS-based studies are applied to various locations (Khan et al, 2015). In another study, IDW was used as the interpolation technique inside GIS to create groundwater quality maps, with several parameters such as total hardness, iron and fluoride concentrations. Increasing water demands also creates a need to develop advanced site selection tools. GIS applications help to determine identification of suitable sites for recharge of groundwater as well (Indranil, 2009).

GIS has been integrated in flood risk analysis studies as well. Applications of remote sensing and GIS are used to create flood risk maps. Furthermore, vulnerability and damage analysis are carried out with the help of population and land use data inside the GIS. Detailed reports and output maps about flood risk are generated inside ArcGIS software (Thumerer et al, 2000; Samarasinghea, S. M. J. S., et al, 2010).

Climate change research have also increased in environmental studies. This phenomena, like other subjects mentioned in this section, has spatial consequences on settlements that need to be analyzed. Climate change has many consequences and is a very large study area. However, one of the most important studies involve coastal impact of global warming. GIS platform was used together with high resolution digital elevation model (DEM), wave and tidal observations on the site by various researchers (Thumerer et al, 2000; Baba et al, 2013). GIS and remote sensing technologies are used together to monitor the effects of climate change on vegetation and agricultural activities as well (Adams et al, 1998; Theurillat and Guisan, 2001).

Usage of GIS on different hydrological problems promote its application with MCM. The following chapter explains streamflow estimation at the ungaged basins using MCM and the newly developed MCM_GIS tool.

CHAPTER 3

METHODOLOGY

In this study, a GIS tool, the MCM_GIS, is developed to implement MCM. In this chapter, the methodology used in this study is explained in detail. The generalized flowchart of the methodology is given in Figure 1.

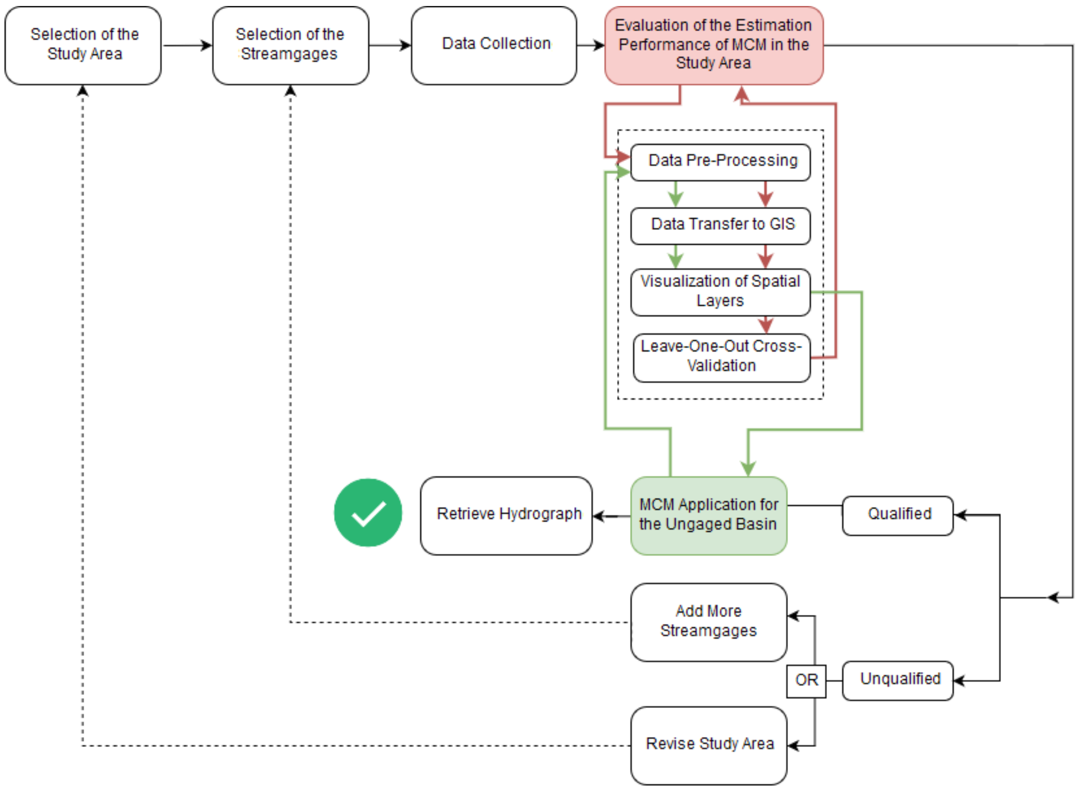


Figure 1: Methodology for streamflow estimation at an ungaged basin using MCM_GIS

Prediction of streamflow at ungaged location with MCM is a multi-step and rather time-consuming procedure. The procedure involves data visualization, generation of correlation maps and visualization of the outputs. All these steps require spatial analysis, which can be easily carried out using GIS tools. Thus, a GIS tool, MCM_GIS is developed to estimate daily streamflow at ungaged basins using MCM. The methodology given in Figure 1 is explained in detail in the following sections.

3.1. Selection of the Study Area

The main goal of this study is to predict streamflow at an ungaged location. Since MCM is based on the spatial analysis of existing streamflow data found around the ungaged location, the first step is the selection of the study area. All the streamgages with daily streamflow measurements around the ungaged location are candidate streamgages to be used in the analysis. This is an important step because the boundary of the study area affects the correlation maps that are composed of correlation contours. To be able to generate reasonable correlation contours around the ungaged location, sufficient number of streamgages that are as much as possible uniformly distributed in the study area shape need to be used in the analysis. Determination of the sufficient number of streamgages requires expertise and a trial-and-error procedure.

The study area can be determined by examining the existing streamgages and their relative locations on the map with respect to the ungaged location. A study area in which the observation points (i.e. streamgages in this study) are uniformly distributed is a good candidate. It has been experienced from the application trials that it becomes difficult for the interpolation algorithms to create contour lines in narrow rectangular study areas. Study areas that are close to a square-shaped bounding box, produces better contour lines. In addition, locating the ungaged basin near the center of the study area is important. This also helps creation of better correlation maps. For example, a study area like the one shown in Figure 2 (a) results in better correlation maps and subsequently better estimations; while a study area similar to the one shown in Figure 2 (b) is not an ideal choice.

After determining the study area, the next step is selection of the streamgages that will be used in generating the correlation maps among all candidate streamgages located within the study area.

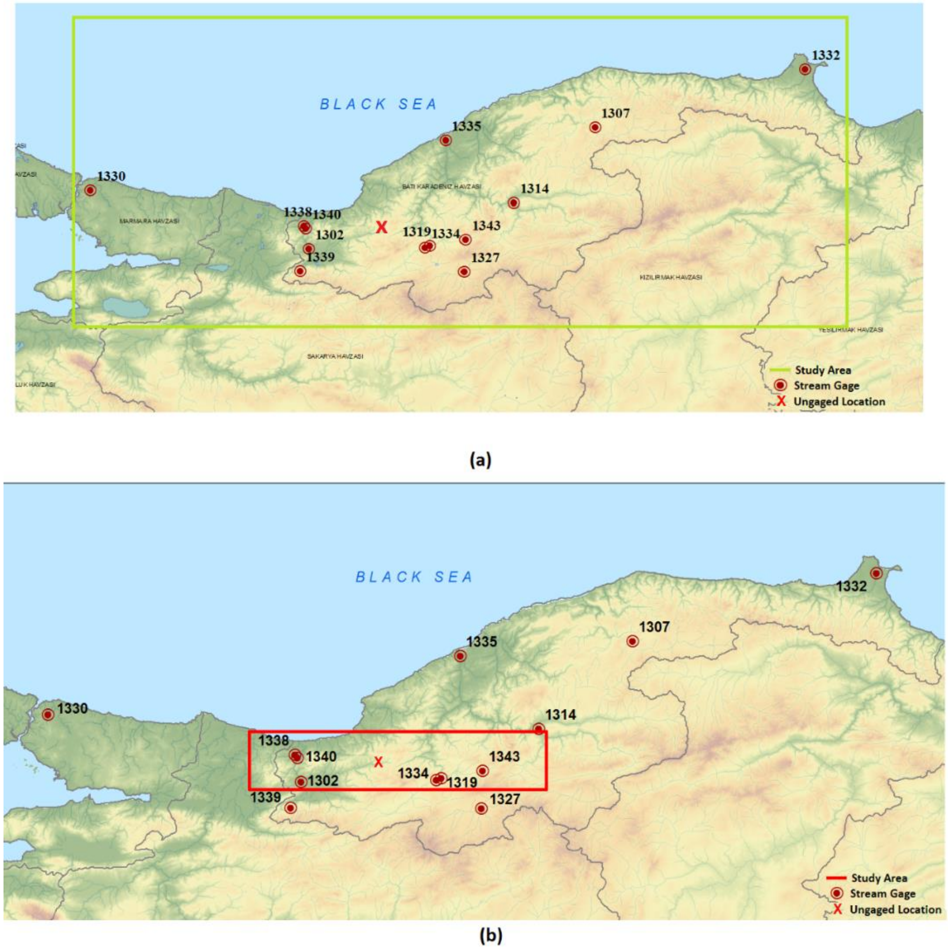


Figure 2: Examples of Study Areas, (a) green box represents suitable study area shape, (b) red box represents unsuitable study area shape

3.2 Selection of the Streamgages

First, all the streamgages inside the study area are identified. MCM requires utilization of streamflow data for the exact same period from all the streamgages. Thus, streamgages with daily streamflow data of common periods need to be identified. Choosing a shorter observation period may result in more streamgages. For example, larger number of streamgages with 10 year observation period, say 2000-2010, within the study area can be identified compared to those with 20 year observation period, say 1990-2010. In other words, in the case of very few number of streamgages with 20-year common observation period, the observation interval can be reduced and more streamgages can be included into the analysis. The trade-off between including more streamgages into the analysis versus choosing longer observation periods must be evaluated.

Once the study area and the streamgages are selected, a leave-one-out cross-validation experiment is applied first to evaluate the performance of MCM within the selected study area. If the performance is not satisfactory then either a new study area, a larger one, should be selected around the ungaged location or more streamgages with shorter observation periods should be included into the analysis (see Figure 1). When the performance is not found satisfactory (i.e. observed and predicted correlation values are not in good agreement) the whole procedure will start all over again, which is time-consuming and calculation intensive. MCM_GIS will ease application of the MCM procedure repeatedly when necessary.

3.3. Data Collection

Required inputs for the MCM_GIS are coordinates and catchment areas of the selected streamgages within the study area, daily streamflow measurements of these streamgages for the common observation period and the location and the catchment area of the point of interest on the streamflow network (i.e. ungaged basin for which streamflow estimations are required).

Streamflow data represents basin response and it reflects the hydrological character of the catchment. Many parameters are used in the hydrological models to predict this response, but streamflow observations convey the combined effect of all these parameters in them. Thus, utilization of streamflow observations directly in predicting the basin response is a practical and realistic approach. Of course, this approach is only useful for streamflow prediction at ungaged locations for the common observation period in which the surrounding operational streamgages collected streamflow data.

For future streamflow predictions, this approach will not work and one should rely on hydrological models. To summarize, MCM is based on the idea that streamflow measurements within a selected area are best indicators of the responses of the sub-basins of the study area, and they can be used to estimate streamflow at ungaged locations within this study area.

The streamflow data for Turkey are collected and distributed by The General Directorate of State Hydraulic Works (DSI). The location of the stations and other general attributes belonging to these stations are publicly available at http://rasatlar.dsi.gov.tr/ while daily streamflow time series is sold upon request by DSI.

In addition to the above stated data, some spatial layers including digital elevation model, dam locations, elevations of streamgages and stream network are beneficial for interpreting and visualizing the results. For example, when the performance of the estimation is poor, the selected reference streamgage can be comprehensively examined by evaluating these data and the reasons behind poor performance can be identified (e.g. reference streamgage being located at an isolated basin or at the downstream of a dam or at a very different elevation). In some situations, such data may help the analyst in data preparation as well. For example, if a streamgage at the downstream of a dam is selected for the analysis, its streamflow measurements need to be corrected first. Effects of regulation should be removed and streamflow data should be naturalized. Stream network is used to see the connections between streamgages since streamgages located in the same stream are expected to have higher correlation values and the elevation is helpful while interpreting the effects of high altitude differences.

3.4. Evaluation of the Estimation Performance of MCM in the Study Area

Before applying MCM to estimate daily streamflow at the ungaged location, the estimation performance of MCM in the selected study area is tested. A leave-one-out cross-validation experiment is carried out for this purpose. One of the streamgages in the study area is assumed to be ungaged. Cross-correlation maps for this streamgage location is prepared using the MCM and the correlation values are determined from these map for the streamgage that is assumed to be ungaged. Then, the observed correlation values for that location are calculated and compared to the estimated values. This procedure is carried out for all the streamgages found in the study area one by one. This is called the leave-one-out cross-validation experiment.

The following steps are used in the leave-one-out cross-validation experiment in this study:

1. Data pre-processing
2. Data transfer to GIS
3. Visualization of spatial layers
4. Leave-one-out cross-validation
    i. Estimation of Pearson's r correlation values using MCM
    ii. Comparison of observed and estimated Pearson's r correlation values to evaluate the performance of MCM in the study area.

Each of these steps is explained in detail in the following paragraphs. The study area and streamgages given in Figure 2 (a) is used as an example to explain the procedure.

Step 1. Data pre-processing

The daily streamflow time series and x- and y-coordinates of the streamgages are obtained from DSI. In MCM, Pearson's r correlation coefficient is used as an indicator of the linear correlation between two data series. For this reason, the logarithms of the daily streamflow values are taken to linearize the relation between two measurements (Archfield and Vogel, 2010). A correlation matrix is calculated using the logarithms of the daily streamflow time series as shown in Equation (1).

$$\begin{bmatrix} r_{1,1} & \cdots & r_{1,N} \\ \vdots & \ddots & \vdots \\ r_{N,1} & \cdots & r_{N,N} \end{bmatrix} \tag{1}$$

Where $r_{i,j}$ is the Pearson's r correlation coefficient between logarithms of the daily streamflow values measured at streamgages i and j, and N is the total number of streamgages located in the study area. The correlation values need to be saved in a special format before being transferred to the GIS environment to carry out leave-one-out cross-correlation experiment.

In the leave-one-out cross-correlation experiment, each streamgage is assumed to be ungaged and one folder is created for it. As can be seen in Figure 2(a), there are 13 streamgages in the study area and one folder is generated for each of these streamgages (see first column of Figure 3). Correlation values of all remaining streamgages (SG) with each other are stored in text files within this folder (see second column of Figure 3). As can be seen in Figure 3, for each streamgage that is assumed to be ungaged there is one folder. To generalize, when SG1 is assumed to be ungaged, the name of the folder is SG1_tables. In each folder, there is a text file for each of the remaining streamgages (which have streamflow data). Therefore, there are 12 text files in 1302_tables folder. To generalize, if there are a total of N streamgages within the study area (including the one assumed as ungaged), in SG1_tables folder, there are $N-1$ text files named as agi_SG1_out_SGj.txt where, $j = 2,3,\ldots,N$. Each text file has a table in it where the columns are separated by commas (see the third column of Figure 3). The first row provides explanation of each column: row number, Pearson's r correlation coefficient between SGj and streamgage listed in the next column, streamgage (i.e. each row will have SGk where $k =$ all gages other than ungaged one), x-coordinate of SGk and y-coordinate of SGk. The following rows provide the related values as shown in Figure 3. For example, in Figure 3, the third row provides the Pearson's r correlation coefficient between 1307 and 1319 in its second column, x- and y-coordinates of 1319 in its fourth and fifth columns, respectively.

The text files must be in certain format, in order to avoid errors while importing these files into the GIS environment. Some key issues are as follows (i) column names should not include space or foreign characters, (ii) x- and y-coordinates should be in the same reference coordinate system for all streamgages. If the data points are obtained from different data sources, it should be checked if they are in the same reference coordinate system, (iii) each value should be separated with a comma. Input files (i.e. txt files shown in Figure 3) of GIS should be prepared following the stated format. Special attention should be paid to the reference coordinate system because this is the basis for the following step, transfer to GIS environment, as it can be seen from the flowchart of this transfer (see Figure 4).

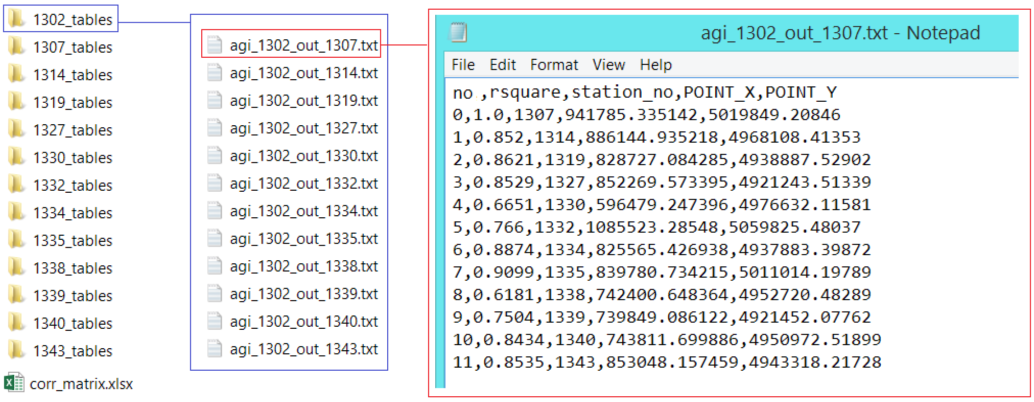


Figure 3: Example file structure of correlation data (the first column shows one folder for each of the assumed ungaged location, the second column shows the contents of the first folder and the third column shows the contents of the first text file in Notepad)

Step 2. Data transfer to GIS

ArcGIS for Desktop Advanced 10.5 program is used in this study. To work in the GIS environment, firstly the input data need to be imported. The text files containing correlation values are converted into feature classes (which are spatial layers) as point layers with a model created in the model builder (see Figure 4). These point layers, corresponding to each text file are represented according to their POINT_X and POINT_Y coordinates and correlation values are stored as their attributes.

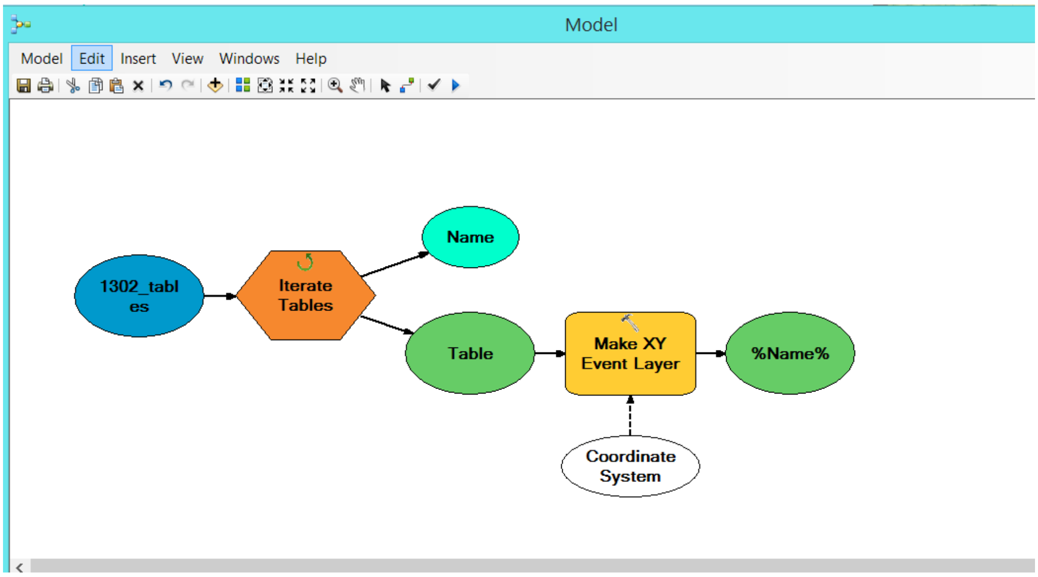


Figure 4: Converting files into GIS layers

Step 3. Visualization of spatial layers

Transferred data can be visualized inside the GIS environment along with other spatial layers such as the river network and the digital elevation model as given in Section 3.3. Locations of the streamgages on the stream network or on the digital elevation map will help the user to understand the system and interpret the results better.

Step 4. Leave-one-out cross-validation

i) Estimation of Pearson's r correlation values using MCM

After correlation between daily streamflow data is imported to the GIS environment, MCM is used to calculate Pearson's r correlation values at the assumed ungaged location. For the sake of simplicity, these values are called estimated Pearson's r correlation values from here after. Application of MCM at the ungaged basin is given in detail in Section 3.5 Map Correlation Method Application for the Ungaged Basin. In the leave-one-out cross-validation experiment, for each streamgage that is assumed to be ungaged, this procedure is applied once. For example, if there are N streamgages within the study area, MCM procedure is applied N times in the leave-one-out cross-validation experiment.

ii) Comparison of the observed and estimated Pearson's r correlation values to evaluate the performance of MCM in the study area

Estimated and observed Pearson's r correlation values are plotted against each other. High positive relationship between these values is an indicator of successful prediction performance of MCM within the selected study area. If the performance of the MCM is found to be sufficient then the method can be applied to find daily streamflow at the point of interest (i.e. the real ungaged location) within the study area. However, if the estimated correlation values are not acceptably close enough to the observed correlation values then the estimation is classified as unqualified (See Figure 1). To overcome this problem, there are two options. One is including more streamgages in the analysis. This can be done by shortening the common observation period, which generally increases the number of streamgages with the same observation period. The other option is to revise the study area. This may also result in including additional streamgages. Nonetheless, a new study area can improve the performance of correlation maps.

Validating the estimation is a very important step in any data estimation study. In this case, unqualified results cause the workflow to repeat from the start. In the feasibility studies of water resources projects like small hydropower plants, data need to be generated quickly. (This is the main motivation for current practice in Turkey: transfer of daily streamflow observations from the closest streamgage to the small hydropower plant location.) MCM searches for the most correlated streamgage around the point of interest and allows transfer of its daily streamflow observations. However, application of MCM by hand is time consuming. Thus, MCM_GIS tool is designed to fasten the process and ease practical application of MCM.

3.5 Map Correlation Method Application for the Ungaged Basin

The MCM assumes that the correlation between their daily stream flow measurements of two catchments is an indicator of hydrologic similarities. The spatial nature of MCM represent an opportunity for its easy implementation in the GIS environment. The procedure for the MCM is given in Figure 5.

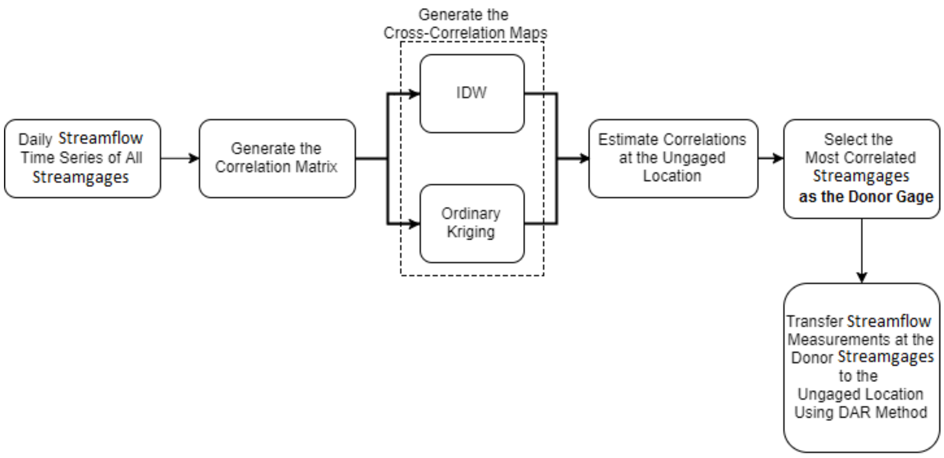


Figure 5: Workflow of MCM (Modified from Ocal and Kentel, 2017)

Four main steps of the MCM to estimate the stream flow time series at an ungaged location are provided below:

1. Calculate the correlation matrix
2. Create the Cross-Correlation Maps (Ordinary Kriging or IDW)
3. Select the reference streamgage
4. Use the DAR method to estimate daily streamflow at the ungaged basin.

Although Step 1 was explained previously, for the sake of completeness it is included in MCM procedure as well. All the step of MCM is explained in detail in the following paragraphs.

Step 1. Calculate the correlation matrix

As explained in Step 1 of Section 3.4. Evaluation of the Estimation Performance, the correlation matrix is prepared using daily streamflow data of all the streamgages in the study area as shown inEquation (1). Then, input text files for GIS are generated similar to those shown in Figure 3: Example file structure of correlation data. This time, since there is a single ungaged location where streamflow will be estimated, a single folder for this ungaged location is necessary. Under this single folder there has to be one text file for each of the selected streamgages within the study area. These text files are transferred into the GIS environment as explained in Step 2 of Section 3.4.

Step 2. Create the Cross-Correlation Maps using Ordinary Kriging or Inverse Distance Weighting

In the second step, cross-correlation maps are generated. This step is one of the main motivations for the GIS integration to the MCM method. The correlation between each gaged location and the ungaged point is estimated using spatial interpolation algorithms of GIS (i.e. in this study OK or IDW are used).

Application of the method is explained on the sample river network given in Figure 2 (a). Assuming that all 13 streamgages given in Figure 2 (a) have common daily streamflow observation periods, these 13 streamgages are selected to be used in MCM. In order to estimate the daily streamflow at the ungaged location, shown with “X” in Figure 2(a), the most correlated streamgage within the study area with the ungaged location need to be identified. Since ungaged location does not have streamflow measurements, this cannot be done directly. Instead, spatial analysis is used to generate a correlation map with each existing streamgage and the ungaged basin to estimate its correlation with the ungaged location. For the sample river network given in Figure

2(a), 13 different correlation maps need to be generated. Group of these maps are referred to as cross-correlation maps in the following paragraphs. Each correlation map is created using one streamgage's correlation values with those of remaining streamgages located within the study area. As a result, these maps created through IDW or OK contain estimated correlation values at every point within the study area. Details of this procedure is given in detail in Ergen and Kentel (2016). The correlation values used in these maps are estimated using inverse distance weighting (IDW) and ordinary kriging (OK) algorithms.

i) Inverse Distance Weighting

Inverse distance weighting (IDW) interpolation assumes that things that are spatially close to each another are more alike than those that are farther apart. To predict a value for any unmeasured location, IDW uses the measured values surrounding the prediction location. The measured values closest to the prediction location have more influence on the predicted value than those farther away. IDW assumes that each measured point has a local influence that diminishes with distance. It gives greater weights to points closest to the prediction location, and the weights lessen as a function of distance. The general formula for IDW is formed as a weighted sum of the data:

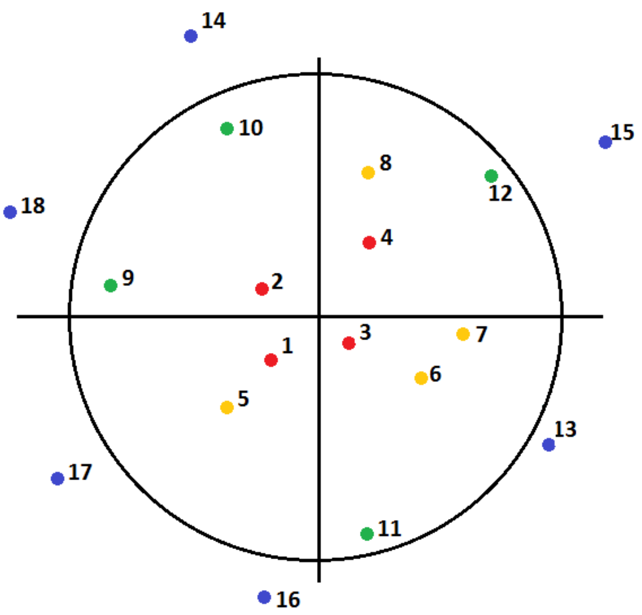
$$\hat{Z}(s_o) = \sum_{i=1}^{N} \lambda_i Z(s_i) \tag{2}$$

where $Z(s_i)$ is the measured value at the location s_i, λ_i are the weights assigned to each measured point at the ith location, s_o is the prediction location, $\hat{Z}(s_o)$ is the value that is being predicted and N is the number of measured values surrounding the prediction location that will be used in the prediction.

Equation (3) is used to determine the weights with the assumption that the sum of the weights assigned to each measured point is equal to one.

$$\lambda_i = d_{i0}^{-p} \Big/ \sum_{i=1}^{N} d_{i0}^{-p} \qquad and \qquad \sum_{i=1}^{N} \lambda_i = 1 \tag{3}$$

As the distance becomes larger, the weight is reduced by a factor of p. The quantity d_{i0} is the distance between the prediction location, s_0, and each of the measured locations, s_i. A sample neighborhood illustration is given in Figure 6.



Data Points	Weights
1	0.2020
2	0.1802
3	0.1212
4	0.1011
5	0.0700
6	0.0672
7	0.0587
8	0.0542
9	0.0413
10	0.0410
11	0.0329
12	0.0297
13	Not Included
14	Not Included
15	Not Included
16	Not Included
17	Not Included
18	Not Included

Figure 6: Neighborhood illustration

As it can be seen from Figure 6, the legend box contains the list of weights assigned to each data point (red, orange, green) that is used to generate a predicted value at the center of crossed circle and weights gets smaller as the distance from the prediction location increases. The sum of the weights assigned to 12 data points is equal to one. In this illustration, it is assumed that the maximum number of neighbors is limited to 12. Therefore, data points after 12 are not included in the analysis. The main reason for this limitation is to reduce the computation time. The details of neighborhood parameters are given in the following paragraphs.

As mentioned above, weights are proportional to the inverse of the distance (between the data point and the prediction location) raised to the power p. As the distance increases, the weight decreases rapidly. The rate at which the weights decrease dependent on the value of p. If p = 0, there is no decrease with distance, and because each weight λ_i is the same, the prediction will be the mean of all the data values in the search neighborhood. As p increases, the weights for distant points decrease rapidly. If the p value is very high, only the immediate surrounding points will influence the prediction. The optimal p value is determined by minimizing the root-mean-square prediction error (RMSPE) (Johnston et al. 2001). The effect of p value on relative height with distance is illustrated in Figure 7.

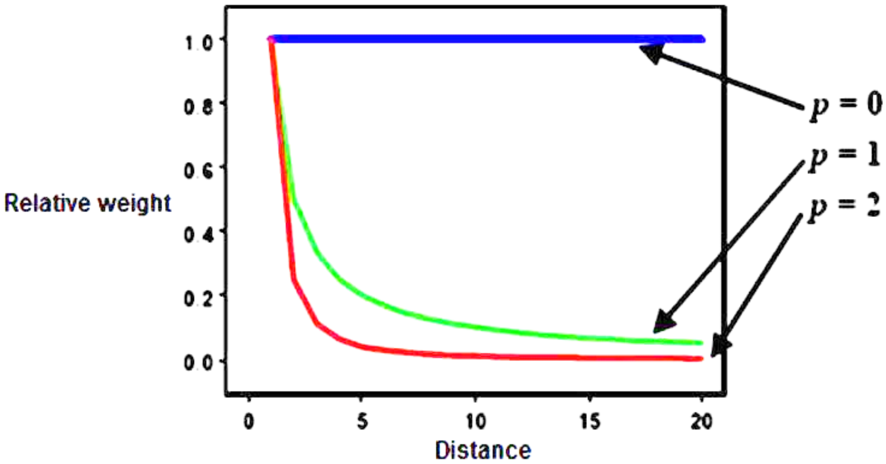


Figure 7: Decrease of weight with distance (ESRI Development Team, 2010)

The tool of IDW in ArcGIS software uses power values greater than or equal to one. When p = 2, the method is known as the inverse distance squared weighted interpolation. The default value of p is two, although there is no theoretical justification to prefer this value over others, and the effect of changing p can be investigated through Geostatistical Wizard tool by previewing the output and examining the cross-validation statistics (Johnston et al. 2001). Inside MCM_GIS tool,

p value is taken as two as default, because it was experienced that, this increases the performance of the estimation. However, the tool allows the user to adjust the p value.

In addition to parameter p, the search neighborhood can be adjusted in the IDW method. The search neighborhood has four major components: search radius, shape, minimum/maximum number of neighbors and sector. The search radius is used to determine how far the interpolator searches the area around the unknown location. The shape of the search neighborhood is influenced by the data itself. If the direction of data is not important, then the search neighborhood should be a circle so that each data point is considered equally in all directions. If direction influences the data, the shape of the search neighborhood can be adjusted. In order to minimize the computation time, a search neighborhood can be used. Distant points that will have little influence on the prediction can be excluded with this search limitation. As a result, it is common practice to limit the number of measured values by specifying a search neighborhood. The shape of the neighborhood restricts how far and where to look for the measured values to be used in the prediction. Sector type selection is provided by Geostatistical Analyst toolbox of ArcGIS. This parameter creates divisions on the neighborhood shape. The specified minimum and maximum number of neighbors is applied to each sector. MCM_GIS tool contains default values of these parameters which were determined by trial and error experiment on several study areas. However, they are adjustable and the users is encouraged to adjust the values for the neighborhood parameters for the specific problem in hand.

Application of IDW through Geostatistical Analyst toolbox in a selected study area is explained in detail in the following steps:

1) To begin with, under the Geostatistical Wizard tool, IDW method is selected from ArcGIS software. Input data set for the application is the text files that are transferred into the GIS environment (Figure 3). Input data field is the Pearson's r correlation coefficient value which is named as, "r_square", in the example shown in Figure 3. The input data set and input data field is selected

for the application. The correlation values are prepared as explained in Step 1 of Section 3.4; then they are transferred into the GIS environment (see Step 2 in Section 3.4. Evaluation of the Estimation Performance for details).

2) As the second step, parameters of IDW are selected. Best values for the neighborhood properties, type of neighborhood, maximum and minimum number of neighborhood and the sector type, has to be determined through a trial-and-error procedure. Effect of each parameter must be investigated and values that produce the minimum error should be determined. In doing this, cross-validation results of Geostatistical Wizard for IDW, which is explained in the next step can be used. An example IDW map is given in Figure 8.

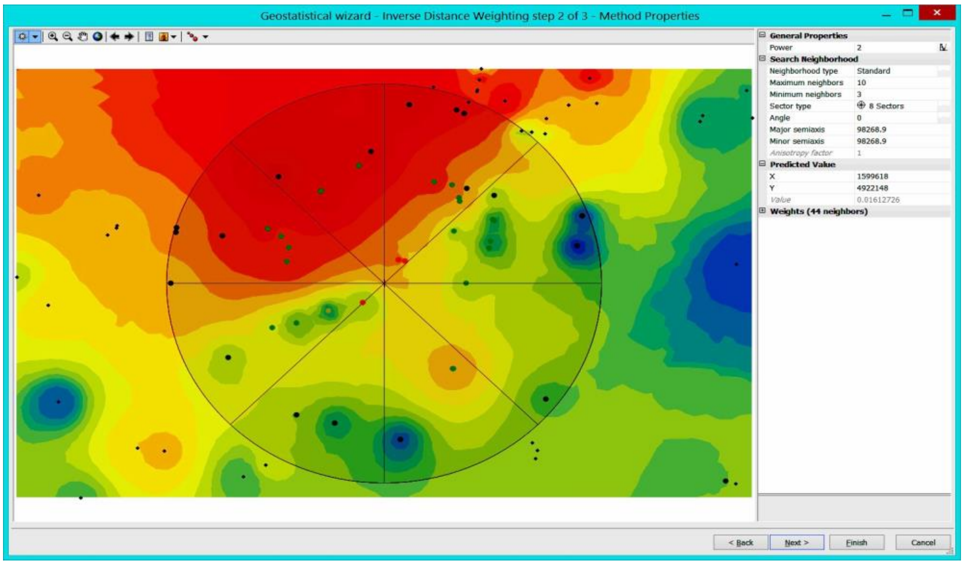


Figure 8: Parameters of IDW

3) Last step of Geostatistical Wizard for IDW is cross-validation. An example cross validation window can be seen in Figure 9. In this step, the observed and predicted values are compared and the performance of the prediction is measured. The plot in Figure 9 shows measured against predicted values. Smaller the angle between two lines, better the performance of the estimator.

The aim of this step is to find and adjust the parameters that minimizes the error. Before finishing the process, certain parameters can be adjusted by going back to previous steps (i.e. minimum/maximum number of neighbors, sector type etc.).

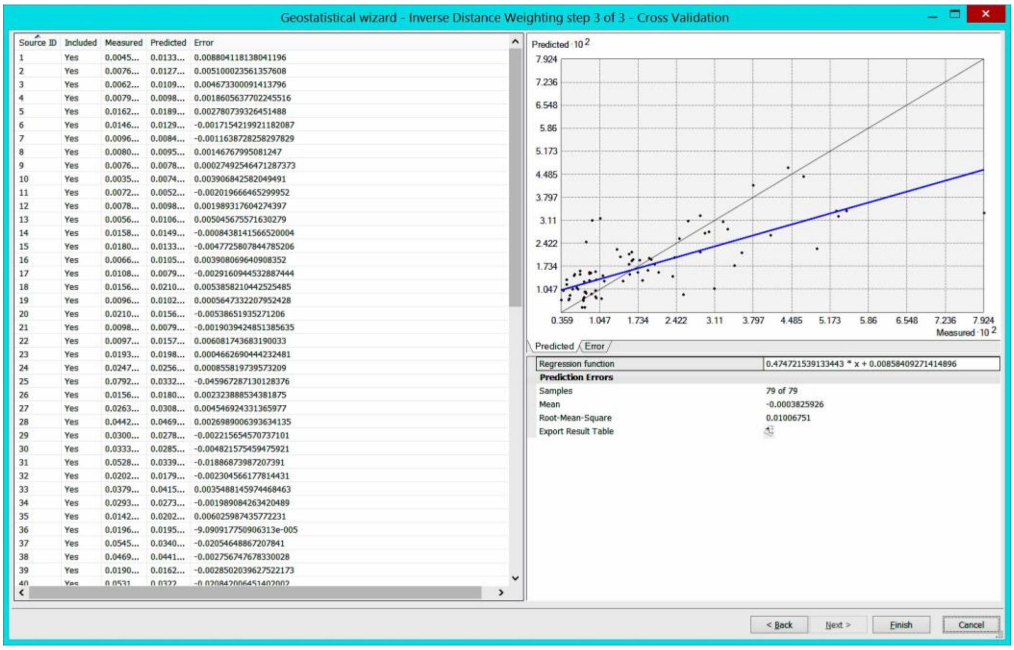


Figure 9: Step 3 - Cross-validation

4) The output of the IDW process is a raster surface map where at each pixel Pearson's r correlation coefficient can be read.

ii) Ordinary Kriging

There are two main groups of models used for predicting spatially continuous data like streamflow data. These are deterministic and stochastic models. Generally, stochastic models are also called geostatistical methods. One of the deterministic models is IDW, which is explained in the previous section. The geostatistical models are a group of different types of kriging. Also, these methods produce not only prediction surfaces, but also error or uncertainty surfaces which provide indications of how good the predictions are. Here, firstly the principle and formulation of kriging are going to be explained. Later, basic steps of ordinary kriging are given in detail.

Kriging assumes that the distance or direction between sample points reflects a spatial correlation that can be used to explain variation in the surface. The kriging tool in GIS fits a mathematical function to a specified number of points, or all points within a specified radius (i.e. neighborhood), to determine the output value for each location.

For all kriging types, the value to be estimated is calculated using Equation (4).

$$Z(s) = \mu(s) + \varepsilon(s) \tag{4}$$

Where $Z(s)$ is the value to be estimated, $\mu(s)$ is the deterministic trend and $\varepsilon(s)$ is the errors due to spatial dependence. This formula varies and constitutes the basis for all types of kriging. The ordinary kriging assumes that $\mu(s)$ is constant and unknown.

As mentioned at the beginning of this section, there are several types of kriging. Most general and widely used one is ordinary kriging and it is selected as the stochastic model in this study.

Geostatistical methods for spatially continuous data include

- Simple Kriging
- Ordinary Kriging
- Universal Kriging
- Block Kriging
- Co-Kriging

Kriging is similar to IDW in that it weights the surrounding measured values to derive a prediction for an unmeasured location. The general formula for both interpolators is formed as a weighted sum of the data as given in Equation (2).

In IDW, the weight, λ_i, depends only on the distance to the prediction location. However, in kriging, the weights are based not only on the distance between the measured points and the prediction location but also on the overall spatial arrangement of the measured points. To use the spatial arrangement in the weights, the spatial autocorrelation are quantified.

In ordinary kriging the weight, λ_i, depens on several parameters which are;

- The variogram model,
- The distance to the prediction location,
- The spatial relationships among the measured values around prediction location.

The sum of weight, λ_i, must be equal to one to ensure that the prediction is unbiased. Using this constraint, the difference between the true value $Z(s_o)$, and the predictor (see Equation (2)), must be as small as possible. In other words, the expectation of Equation (5) must be minimized.

$$\left[Z(s_o) - \sum_{i=1}^{N} \lambda_i Z(s_i) \right]^2 \tag{5}$$

Where N is the number of observations near prediction location. The solution of the minimization, constrained by unbiasedness gives us the kriging equations as shown in Equation (6).

$$\Gamma \lambda = g \tag{6}$$

Where Γ matrix contains the modelled variogram values and g matrix contains modelled variogram between measured and predicted locations, and λ is the weight matrix. Open form of Equation (6) is given in Equation (7). Ones and zeros in the bottom row and the right-last column are used to represent the unbiasedness constraint.

$$\begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1N} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{N1} & \cdots & \gamma_{NN} & \vdots \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \vdots \\ \gamma_{N0} \\ 1 \end{bmatrix} \tag{7}$$

$$\Gamma \qquad \lambda = g$$

The basic steps of kriging include the following;

1. Calculating the empirical variogram
2. Fitting a model
3. Creating matrices
4. Making a prediction

Each of these steps are explained in detail below.

1. Calculating the empirical variogram

The empirical variogram explores the spatial correlation between data. Pairs that are spatially close in distance should have smaller measurement difference than those farther away from one another. The variogram helps examining this assumption and to what extent the data is spatially correlated.

The variogram, also called semi-variance or semi-variogram, is estimated over all directions for a given distance separation h as in Equation (8).

$$\hat{\gamma}(h) = \frac{1}{2} \cdot \frac{1}{n(h)} \sum_{i=1}^{n(h)} \left(Z(y_i + h) - Z(y_i) \right)^2 \tag{8}$$

Where $Z(y_i)$ is the value at a particular location, $\hat{\gamma}(h)$ is the estimated semi-variogram, $n(h)$ is the number of pairs separated by the distance h. The summation is over all pairs of observed data point with a vector separation of h. Figure 10 represents the pairing of one point (the red point in the middle) with all other measured locations. This process is applied for each of the measured points.

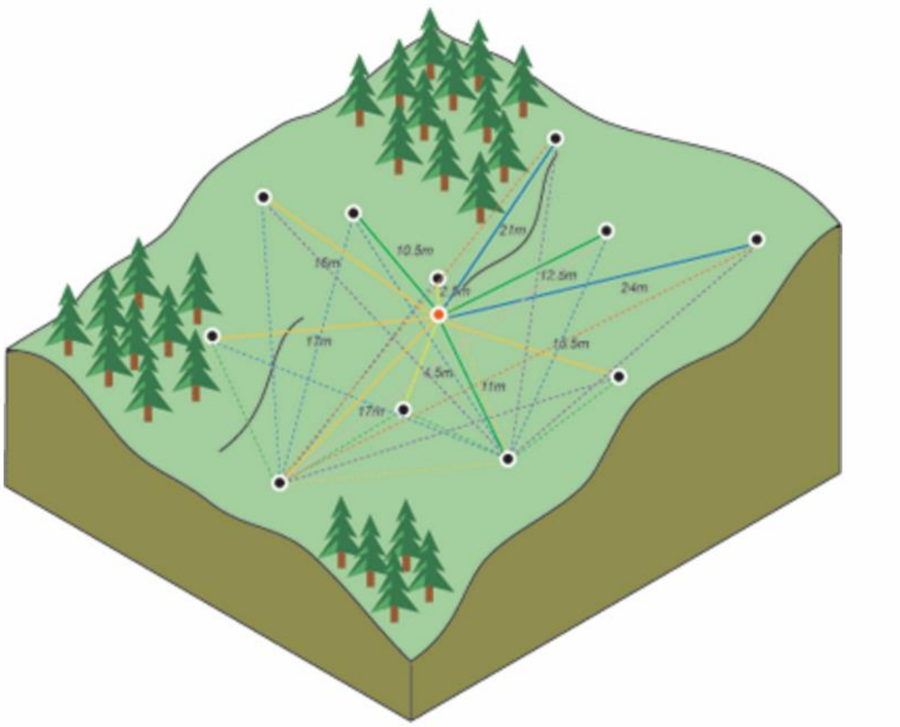


Figure 10: Variogram calculation on sample data. (ESRI Development Team, 2010)

Plotting each pair separately is time-consuming. To ease this, the pairs are grouped into bins. The empirical semi-variogram is a graph of the averaged semi-variogram values on the y-axis and the distance (i.e. lag) on the x-axis. A sample semi-variogram is given in Figure 11 along with its characteristics.

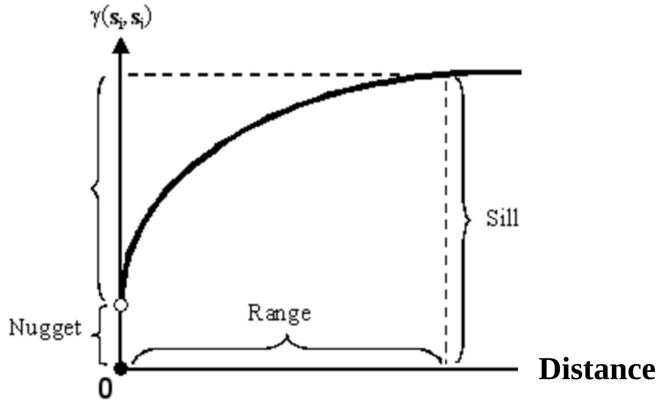


Figure 11: Example of a semi-variogram

As seen in Figure 11, at a certain distance the model levels out. This distance value is known as the range. Sample locations separated by the distances closer than range are spatially correlated, whereas locations farther apart than the range are not correlated. Sill is the value that the variogram attains at the range. Theoretically, $\hat{\gamma}(0)$ should be zero. If the variogram value is not zero at the origin, it is called the nugget (i.e. nugget effect). The nugget effect is often a result of measurement error.

2. Fitting a model

Second step of ordinary kriging is fitting a model. This process is done by defining a line that provides the best fit through the points in the empirical variogram plot. The weighted squared difference between each point and the line should be as small as possible. This line is considered as a model that quantifies the spatial autocorrelation in the data. This step is similar to regression analysis, in which a continuous line or curve is fitted to the data points. There are several semi-variogram models such as; circular, spherical, exponential, Gaussian, and linear. The selected model influences the prediction of the unknown values, especially when the shape of the curve near the origin differs significantly. The steeper the curve near the origin, the more influence the closest neighbors will have on the prediction. As a result, the output kriging surface will appear less smooth. Examples of spherical and exponential models are given in Figure 12 (A) and (B), respectively.

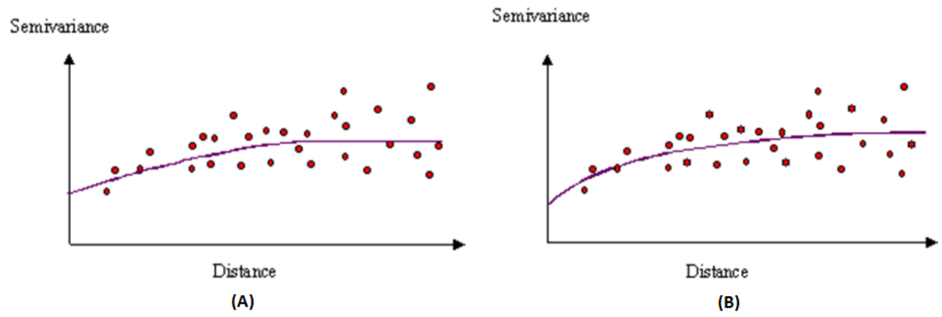


Figure 12: Different types of variogram models. (A) shows a spherical model while (B) shows an exponential model (ESRI Development Team, 2010)

3. Creating the matrices

The equations for ordinary kriging are stored in matrices that depend on the spatial correlation between the measured and prediction locations. The correlation values are obtained from the fitted semi-variogram and put into Γ and g matrices as shown in Equation (7).

4. Making a prediction

As the last step, after Γ and g matrices are obtained, the weight matrix λ given in Equation (7) can be calculated. The matrices that are shown in Equation (7) determine the kriging weights that are assigned to each measured point. Finally, the value at the unknown location is predicted using Equation (2).

Application of ordinary kriging in the Geostatistical Analyst toolbox of GIS is explained in detail in the following steps:

1. To begin with, under Geostatistical Wizard tool, Kriging/Cokriging method is selected in ArcMap software. The input data set and input data field is selected for the application similar to IDW. Input data field is the Pearson's r correlation coefficient value which is named as, "r_square", in the example

given in Figure 3. The correlation values are prepared as explained in Step 1 of Section 3.4. Evaluation of the Estimation Performance; then they are transferred into the GIS environment (see Step 2 in Section 3.4 for details).

2. As the second step of kriging, ordinary for type and prediction surface are selected and no transformation is applied to input data.

3. At this step, the semi-variogram is created. An example semi-variogram is given in Figure 13 where the red points are binned values, blue crosses represent averaged values and the blue line is the variogram model. At this step, the range, sill and nugget effect is calculated automatically. When the bin size and the number of bins change, the fitted model changes. The selection of the number of lag/bins is very important for the variogram. When the lag size is too large, the effect of the correlation between closer points on the estimation can be ignored. Similarly, when the lag size is too small, the number of data included in one-bin increases too much. Bin size and number of bins can be adjusted based on cross validation results. MCM_GIS tool has default values for these parameters. However, adjustment of these parameters by the user for the specific problem in hand is encouraged.

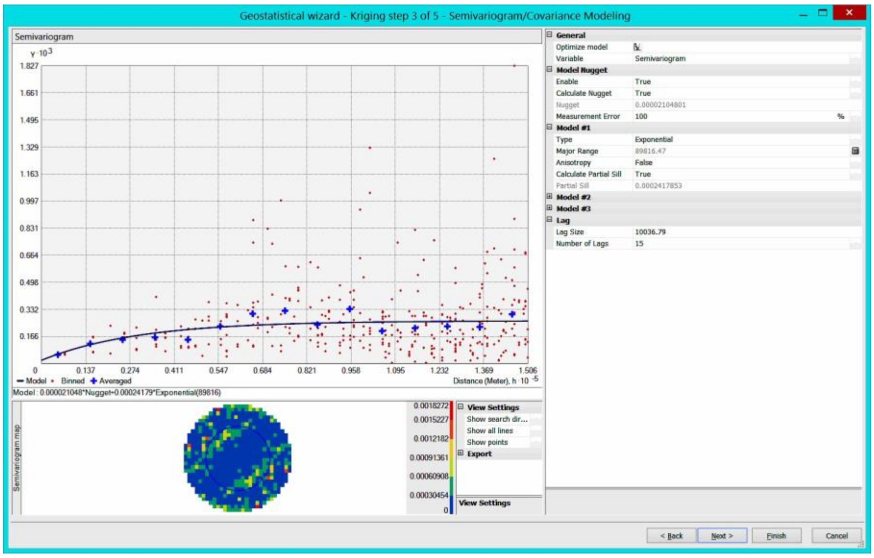


Figure 13: Step 3 - Semivariogram modeling

4. At the forth step best values for the neighborhood properties are identified. Type of neighborhood, maximum and minimum number of neighborhood and the sector type has to be determined through a trial-and-error procedure. Effect of each parameter must be investigated and values that produce the minimum error should be determined. This can be done using the cross validation procedure explained in the following step.

5. Cross validation step compares observed and estimated data. Unlike IDW result tables, standardized error is also calculated for OK. Calculated error helps the user to evaluate the accuracy of the prediction model.

6. The output of the OK process is a raster surface map where at each pixel Pearson's r correlation coefficient can be read.

Output map of both OK and IDW represent a prediction raster surface. In these raster files, each pixel corresponds to an estimated Pearson's r correlation coefficient value. These raster files containing correlation values that can be represented as correlation maps. The correlation value at the point of interest where no observation exists (i.e. the ungaged basin) is determined from these maps.

Step 3. Select the reference streamgage

Using each correlation map that belongs to selected streamgages in the study area, the Pearson's r correlation coefficient value is read at the ungaged location. Then, these Pearson's r correlation coefficient values are evaluated and the streamgage with the largest correlation with the ungaged location is selected as the reference/donor streamgage.

Step 4. Use the DAR method to estimate daily streamflow at the ungaged basin

Finally, the last step is the estimation of the daily streamflow at the ungaged. Daily streamflow at the point of interest (i.e. the ungaged basin) can be estimated using DAR method which is the most common and widely used method to estimate daily streamflow at an ungaged catchment. DAR method, uses a reference streamgage and its streamflow observations on a given day t are transferred to the ungaged location by Equation (9).

$$Qu_t = \frac{A_u}{A_g} Qg_t \tag{9}$$

Qu_t is the streamflow on day t at the ungaged site, Qg_t is the streamflow on day t at the reference streamgage, A_u is the drainage area of the ungaged catchment, and A_g is the drainage area to the reference streamgage.

Reference streamgage identified in the previous step will be used to estimate streamflow values at the ungaged location. The reference streamgage identified through MCM is the one that is expected to have the largest correlation with the ungaged location. This is the main difference between MCM and traditional approach where the spatially closest streamgage is used as the reference streamgage. Benefit of using the most correlated streamgage as the reference streamgage has been demonstrated in a number of recent studies (Archfield and Vogel, 2010; Ergen and Kentel, 2016; Ocal and Kentel, 2017; Monjardin et al, 2017).

3.6 Description of the GIS Tool, MCM_GIS

MCM_GIS is the GIS tool developed in this study to carry out MCM procedure. In MCM_GIS, Python is used as the scripting language since it is available in many GIS softwares and has many applications in open source GIS studies (Zambelli et al, 2013). The tool is ready to be used in ArcGIS products with the help of Arcpy library, however, the tool can be shared on different GIS platforms that supports Python with minor changes as well. MCM_GIS is a collection of tools. All of these tools aim to automatize the application of MCM. General workflow of the tool follows MCM steps, which are explained in the previous sections. Utilization of MCM_GIS is demonstrated on the example study area given in Figure 2(a), in the following paragraphs.

The general structure of MCM_GIS is shown in Figure 14. MCM_GIS can perform two different sets of operations, evaluation of the estimation performance of MCM in the study area and MCM application for the ungaged basin as shown in red and green boxes in Figure 14, respectively. As explained in Section 3.4, evaluation of the estimation performance of MCM in the study area is nothing but application of the MCM for an ungaged basin multiple times. Thus, MCM application for the ungaged basin using MCM_GIS is explained first.

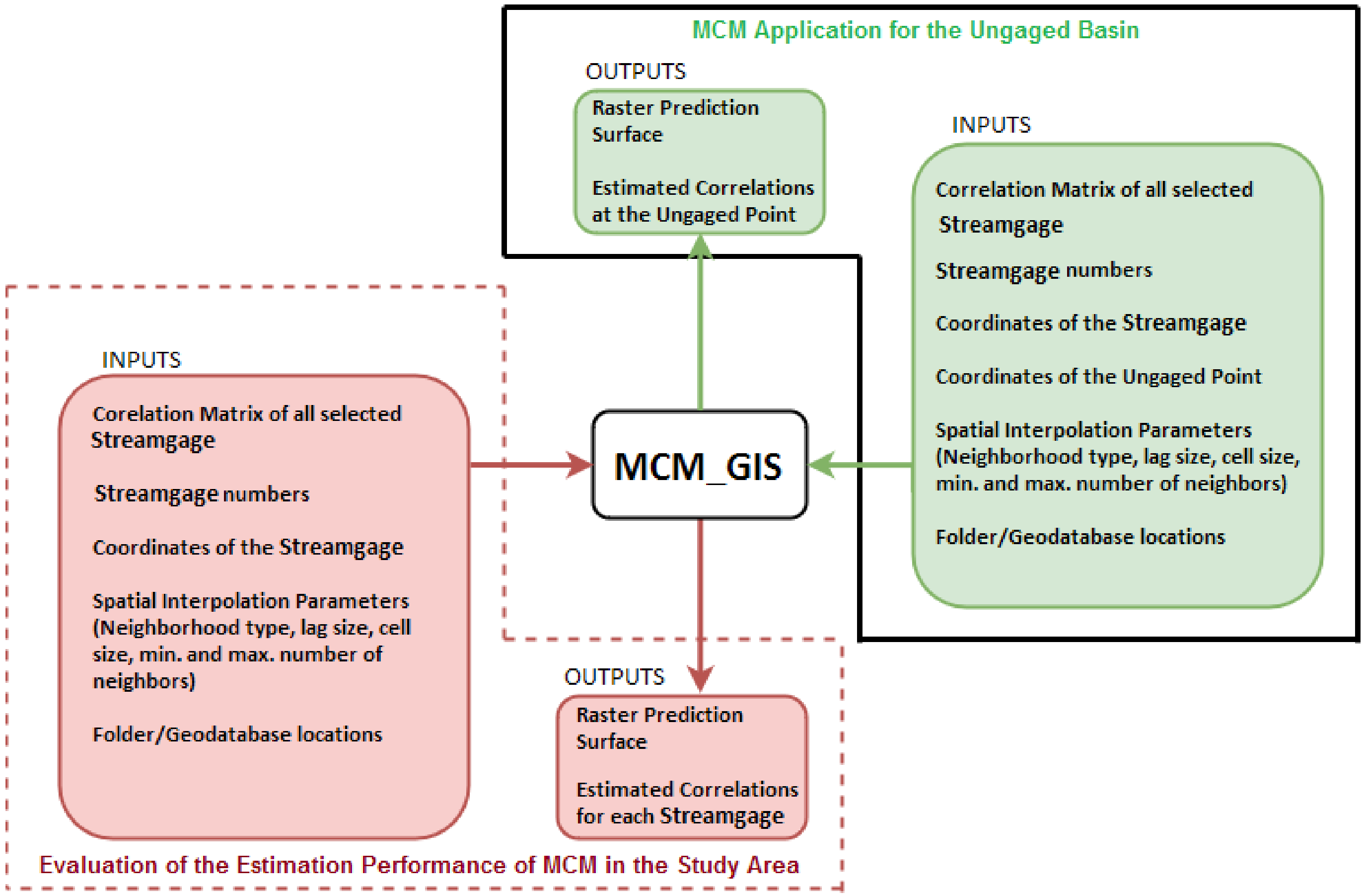


Figure 14: The general structure of MCM_GIS

3.6.1 MCM application for the ungaged basin using MCM_GIS

Application of MCM for the ungaged basin using MCM_GIS is explained using the example study area given in Figure 2(a). The folder and file structure that is required by MCM_GIS as input, is given in Figure 15. In the first level, there is a folder for the point of interest (i.e. where streamflow values are missing and will be calculated through MCM). The name of the folder should have the following format: **SG**_tables. Here the bold part is the name given to the point of interest and the regular part is fixed (i.e. has to be "tables" all the time). For the study area given in Figure 2(a), the point of interest is called ungaged. Thus, the name of the folder is ungaged_tables for this example (see Figure 15). Inside the folder, there is one text file for each streamgage located within the study area. The name of the text files should have the following format: agi_**SG**_out_***SGi***.txt. Here the bold part is the name given to the point of interest that has to be the same with the folder name. The bold and italic part is the name of the streamgage whose correlations are provided with all other streamgages within the text file. For this example, since there are 13 streamgages within the study area, there are 13 text files inside ungaged_tables folder (see Figure 15) for each of these streamgages. These text files are named as agi_ungaged_out_*SGi*.txt, *i*=1302, 1307, …., 1343 as shown in the second level of Figure 15. In the case of ungaged folder, the text files are named as: ungaged_ ***SGi***.txt. Finally, each text file has the correlation of *SGi* with all other streamgages within the study area and x- and y-coordinates of each of these streamgages. For example, in the third level of Figure 15, the text file for ungaged_1302.txt is given. This file has the correlation of 1302 with all other streamgages in the study area, corresponding streamgage number to the correlation value and x- and y-coordinates of each of these streamgages.

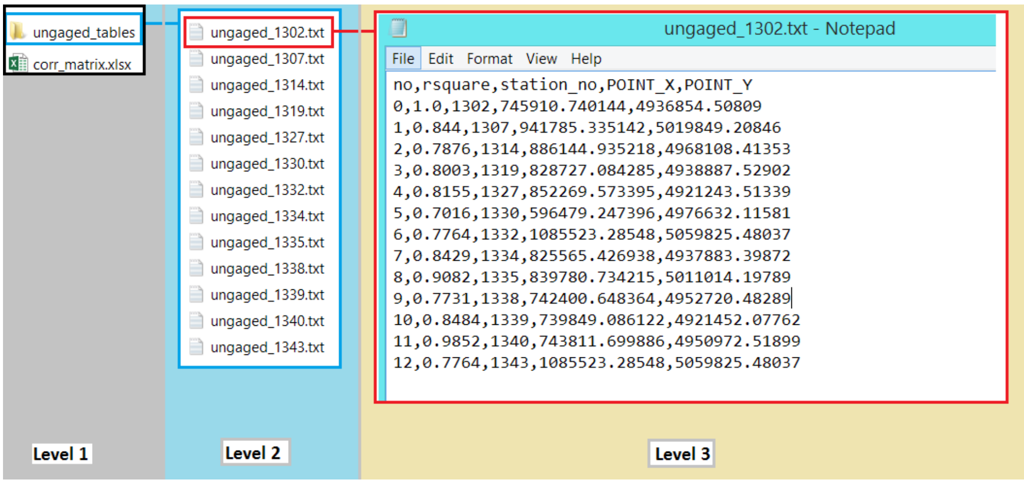


Figure 15: File Structure for the Ungaged Location

Since it will be time consuming to prepare this folder and file structure by hand, a MATLAB script called File_Generator (see Figure 16) is written to convert a single Excel file which has the correlation matrix, the name of the streamgages and the coordinates of the streamgages into separate Excel files that contain correlation values for each streamgage with all others, the name of the streamgage and the coordinates of the streamgages. The File_Generator is an M file and should be executed in MATLAB. The original Excel file which will be called by the MATLAB script should have a specific format which is given in Figure 17. When MATLAB File_Generator is executed, the original Excel file is converted to a set of Excel files (i.e., one for each streamgage) as shown in Figure 17, as well.

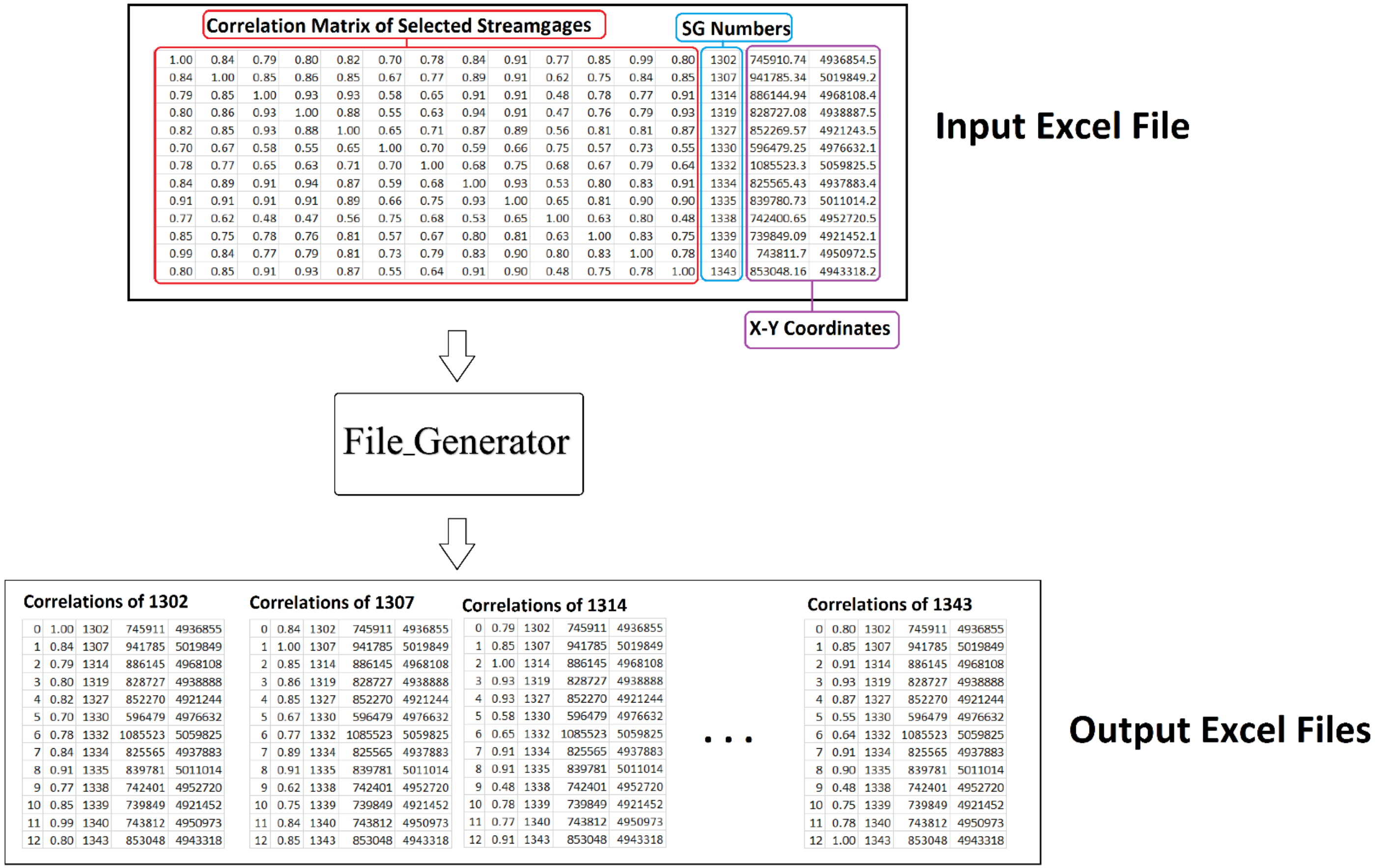
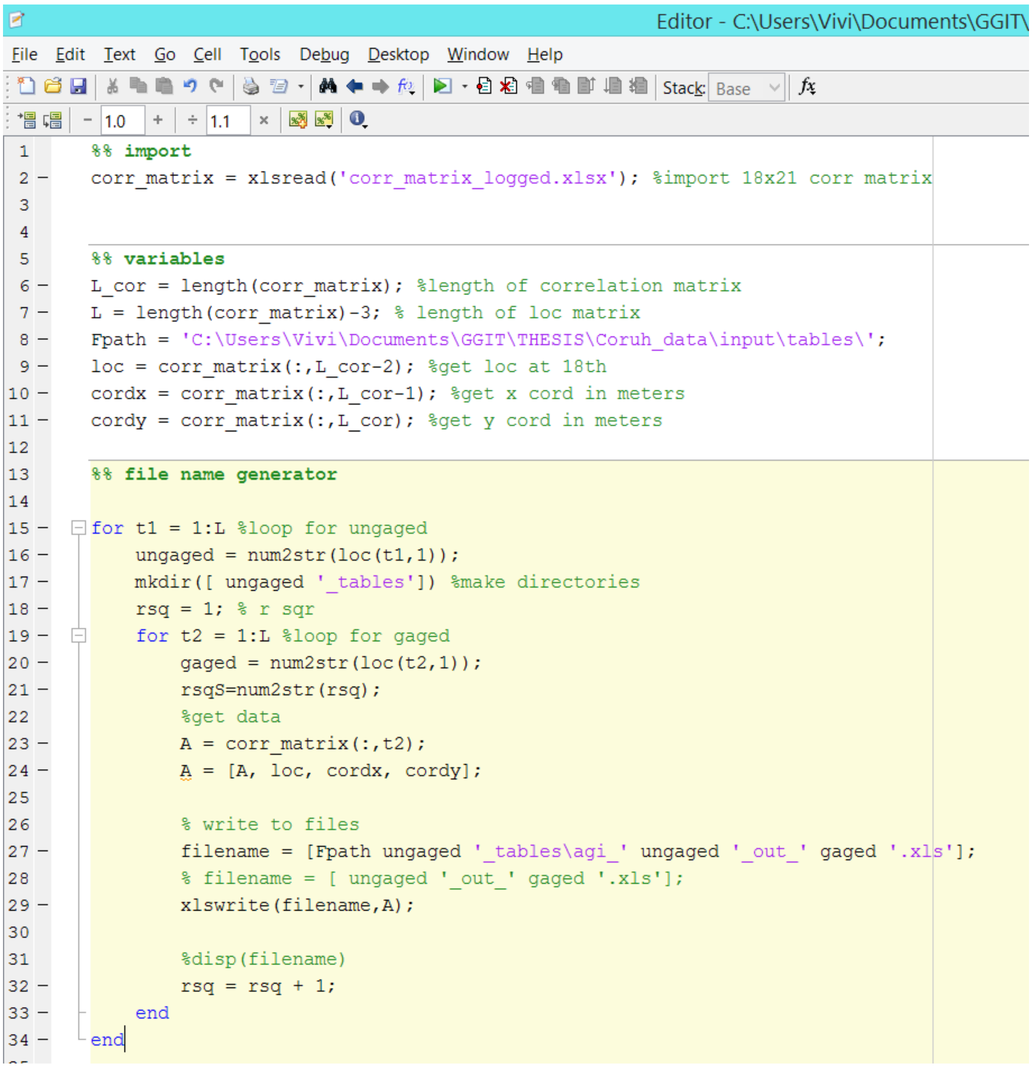


Figure 17: Input Excel file and output Excel files for File_Generator




```matlab
%% import
corr_matrix = xlsread('corr_matrix_logged.xlsx'); %import 18x21 corr matrix


%% variables
L_cor = length(corr_matrix); %length of correlation matrix
L = length(corr_matrix)-3; % length of loc matrix
Fpath = 'C:\Users\Vivi\Documents\GGIT\THESIS\Coruh_data\input\tables\';
loc = corr_matrix(:,L_cor-2); %get loc at 18th
cordx = corr_matrix(:,L_cor-1); %get x cord in meters
cordy = corr_matrix(:,L_cor); %get y cord in meters

%% file name generator

for t1 = 1:L %loop for ungaged
    ungaged = num2str(loc(t1,1));
    mkdir([ ungaged '_tables']) %make directories
    rsq = 1; % r sqr
    for t2 = 1:L %loop for gaged
        gaged = num2str(loc(t2,1));
        rsqS=num2str(rsq);
        %get data
        A = corr_matrix(:,t2);
        A = [A, loc, cordx, cordy];

        % write to files
        filename = [Fpath ungaged '_tables\agi_' ungaged '_out_' gaged '.xls'];
        % filename = [ ungaged '_out_' gaged '.xls'];
        xlswrite(filename,A);

        %disp(filename)
        rsq = rsq + 1;
    end
end
```

Figure 16: MATLAB script: File Generator

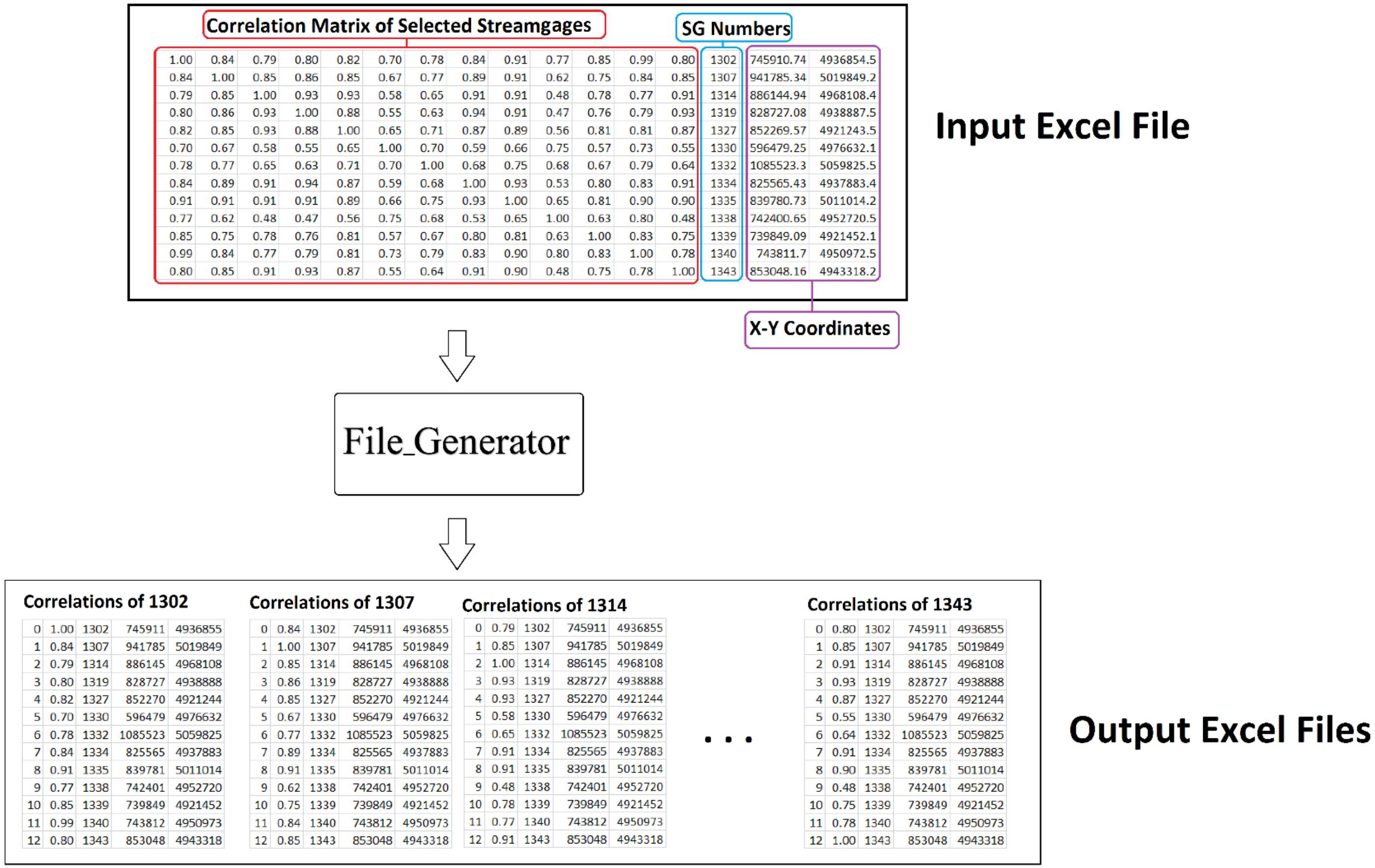


Figure 17: Input Excel file and output Excel files for File_Generator

Finally, the Excel files for each streamgage should be converted to text files that will be imported by MCM_GIS. The rest of the MCM procedure is carried out in the GIS environment.

First, ArcMap is run to start the analysis inside the GIS environment. To convert Excel files into text files an ArcMap tool named xls2txt is written in Python scripting language. This tool is stored in a toolbox (see the right pane of the window shown in Figure 18) together with a number of other tools that are written to carry out the rest of the MCM procedure. The user can either load the Python script into the Python window of ArcMap program and run or directly from the user interface (UI) shown in Figure 19. The UI can be activated by double clicking on xls2txt. This conversion tool requires the location of the Excel files as the input.

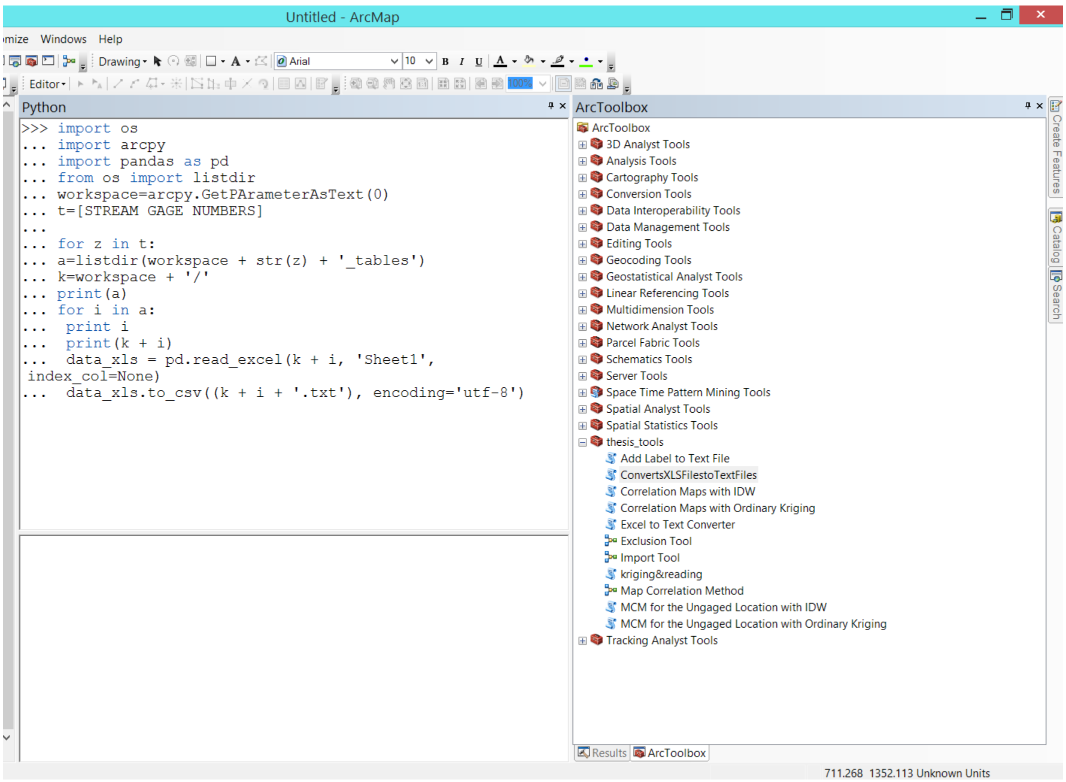


Figure 18: ArcGIS tool: xls2txt

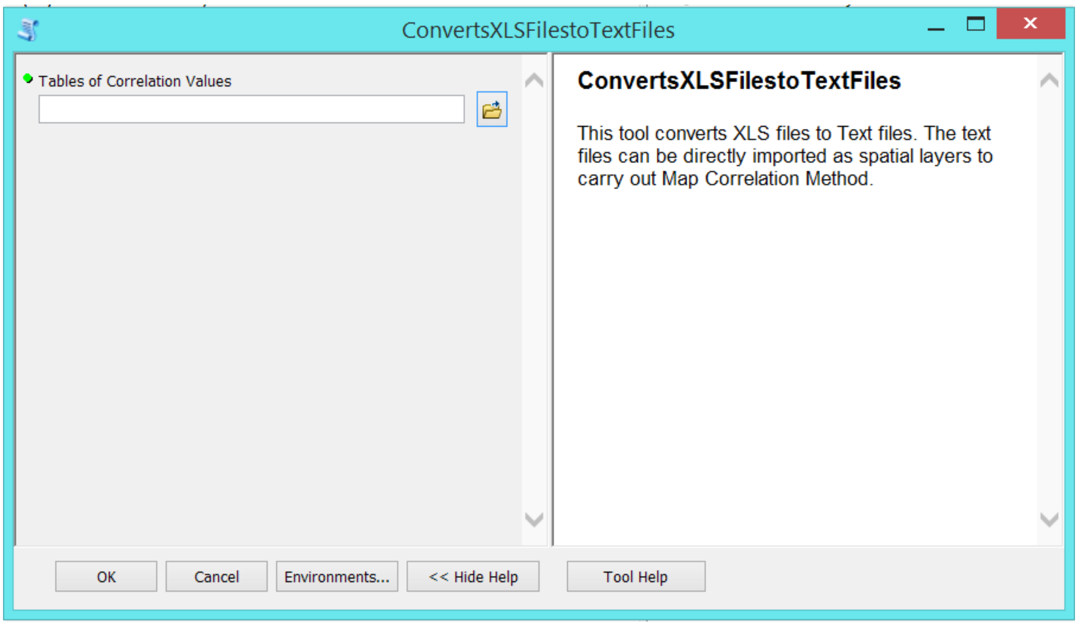


Figure 19: The UI for xls2txt

The last step of input files preparation is adding labels (i.e. the first row of the text files given in Level 3 of Figure 15) to the text files. A labeling tool is written in ArcMap to add a labeling line to the text files prepared using File_Generator. The UI of the labeling tool, Add Label to Text File is given in Figure 20. Now, all the text files given in Level 3 of Figure 15 are ready to be used in the MCM procedure.

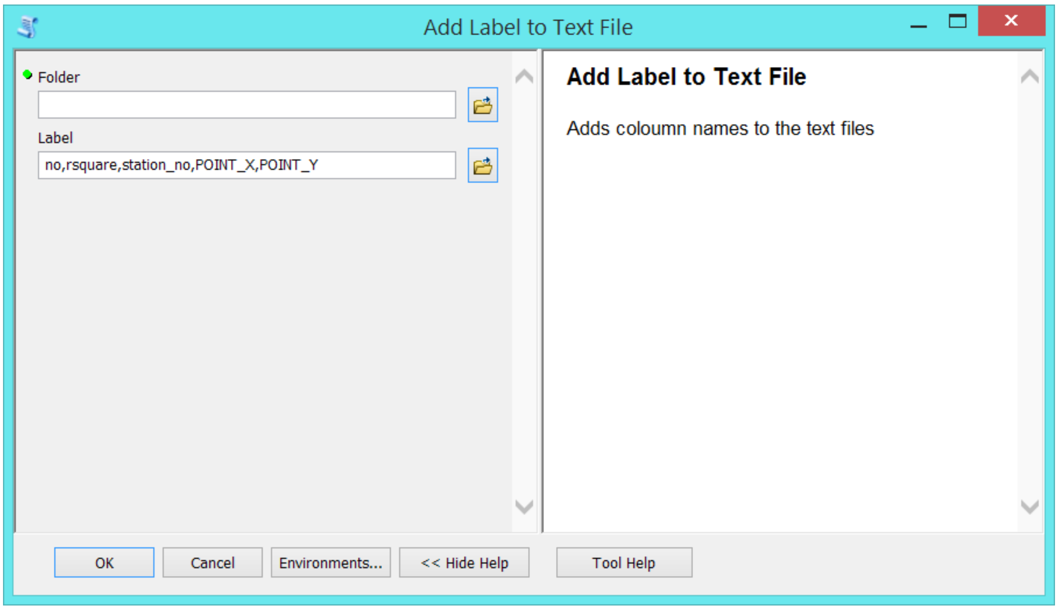


Figure 20: The UI for Add Label to Text File

The next step is the generation of geodatabases. For each folder, one geodatabase is created. The geodatabase can be manually created using ArcCatalog application of ArcMap. However, when there are too many streamgages in the analysis, this step may become time consuming so a small script called Create_Geodatabase is written in Python as shown in Figure 21. This script can be used to convert all folders into geodatabases in a single execution. Each geodatabase is named in accordance with the name of the folder. For the example study area given in Figure 2(a), ungaged_tables folder is converted into ungaged.gdb.

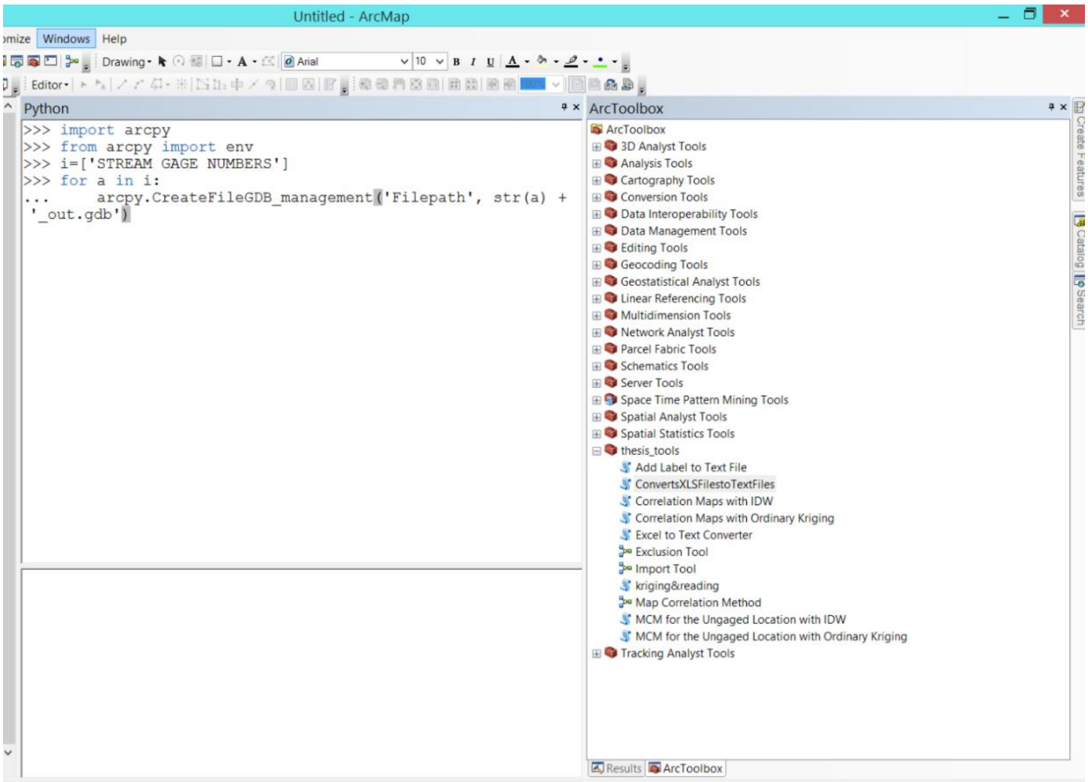


Figure 21: Python script: Create_Geodatabase

The text files inside the folder are transferred into its corresponding geodatabase. This process' requirements and details are given in Step 2 of Section 3.4. Transfer of text files into the geodatabase is carried out with the Import Tool (see Figure 22) that is prepared with the Model Builder application of ArcMap. The Import Tool takes the text files as main inputs and generates feature classes as outputs (see Figure 23). Other inputs to this tool are reference coordinate system and the ungaged.gdb location. After this transfer, the example geodatabase structure is similar to Figure 24.

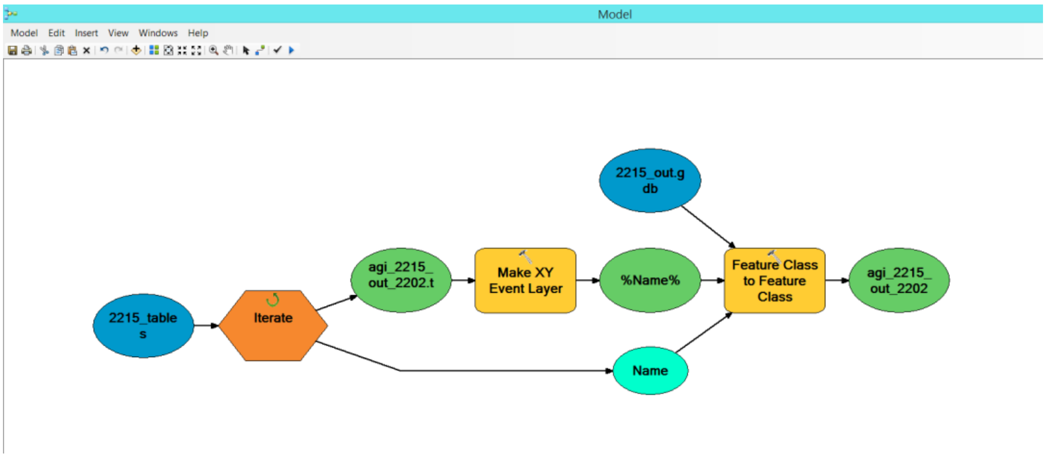


Figure 22: Model Builder: Import tool

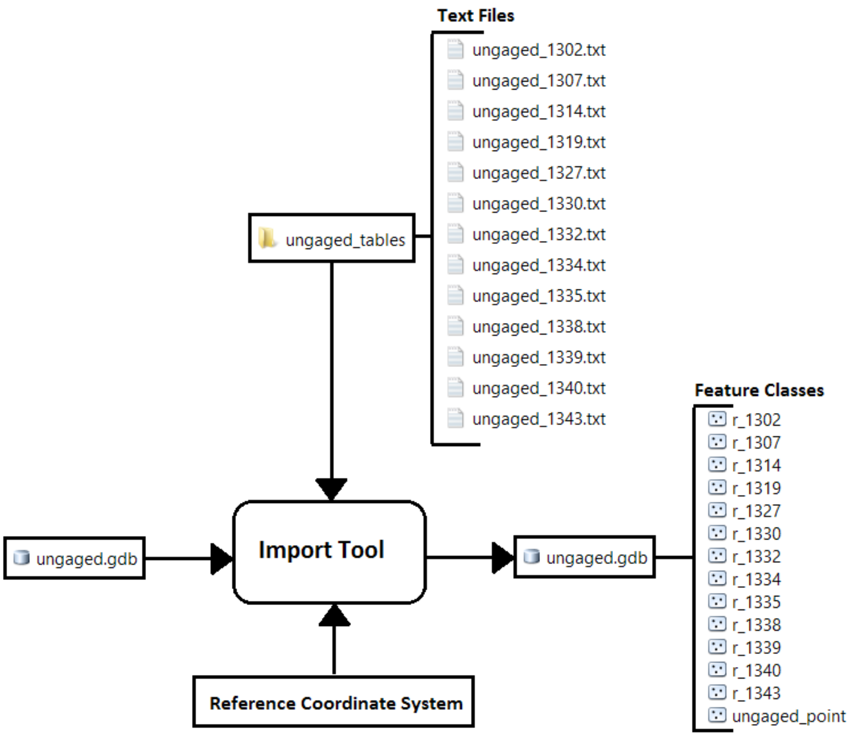


Figure 23: Import tool inputs and outputs

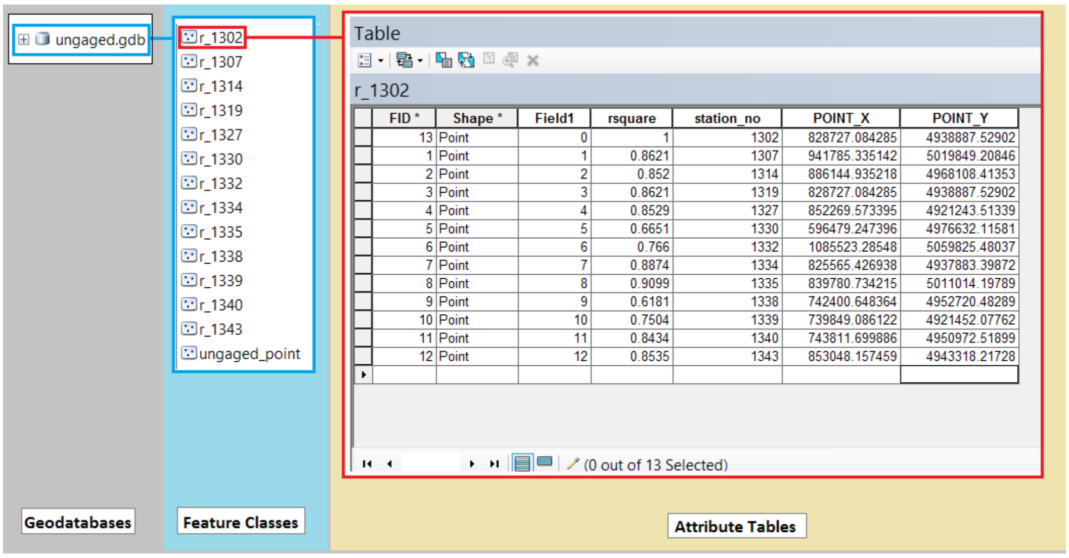


Figure 24: Geodatabase Structure for the Ungaged.gdb

Before proceeding to correlation map generation, it will be beneficial to visualize the river network, dams and locations of streamgages within the study area. This will help the user to identify potential errors such as utilization of wrong coordinate system which will result in wrong locations of streamgages on the river network or utilization of streamflow measurements of streamgages that are located downstream of a dam without naturalizing the data.

Finally, correlation maps are generated using either IDW or OK, which is selected by the user. To apply IDW, a Python script called MCM for the Ungaged Basin with IDW is written (see Figure 25) and this script is run through the designed UI. The UI for MCM for the Ungaged Basin with IDW is given in Figure 26. This UI requires geodatabase file, streamgage numbers, neighborhood parameters, correlation field name (r_square in this case), output raster location as inputs. Best parameters for IDW has to be identified using the Geostatistical Wizard of ArcMap through a trial-and-error procedure as explained in 3.5. In this study, MCM_GIS is applied at two different study areas and based on the experience gained from these studies, the values for sector and minimum neighbor can be selected as 4 and 10, respectively to start with the trial-and-error procedure. Moreover, Standard worked fine as the neighborhood type in both case studies.

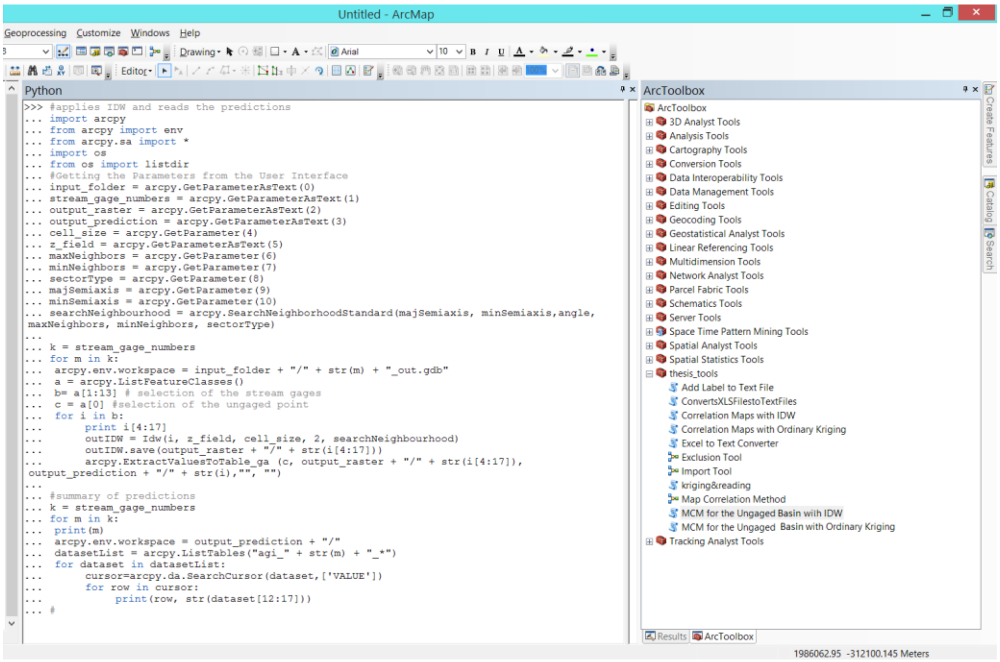


Figure 25: Python script: MCM for the Ungaged Location with IDW

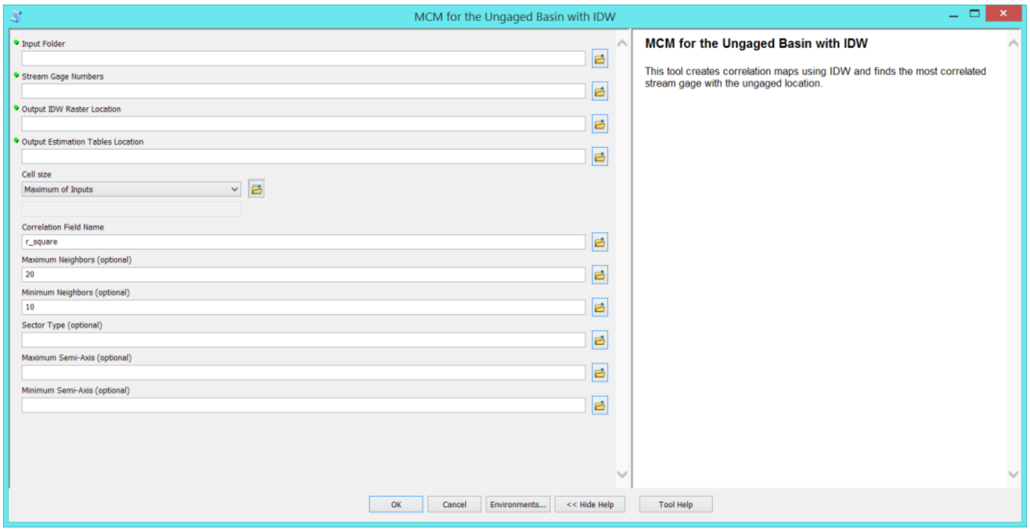


Figure 26: MCM for the Ungaged Basin with IDW UI

After all the necessary input is provided for MCM for the Ungaged Location with IDW, when clicked OK, the script runs and cross-correlation maps for the ungaged basin with each of the streamgages are generated. To display the correlation map, from Table of Contents window of ArcMap, the corresponding layer of the map is selected. For the study area given in Figure 2(a), the cross-correlation maps for 12 of the streamgages are generated and as an example, the map for 1314 is displayed in Figure 27. Estimated correlation values are stored as tables. Example of estimated correlation values for streamgage 1314 is given in Table 1. The corresponding table can be added from the *Output Estimation Tables Location* provided inside the MCM for the Ungaged Basin with IDW UI to Arcmap. The table can be opened by right-clicking to the attribute table of the corresponding table. The estimated correlation value is stored as an attribute of the output table.

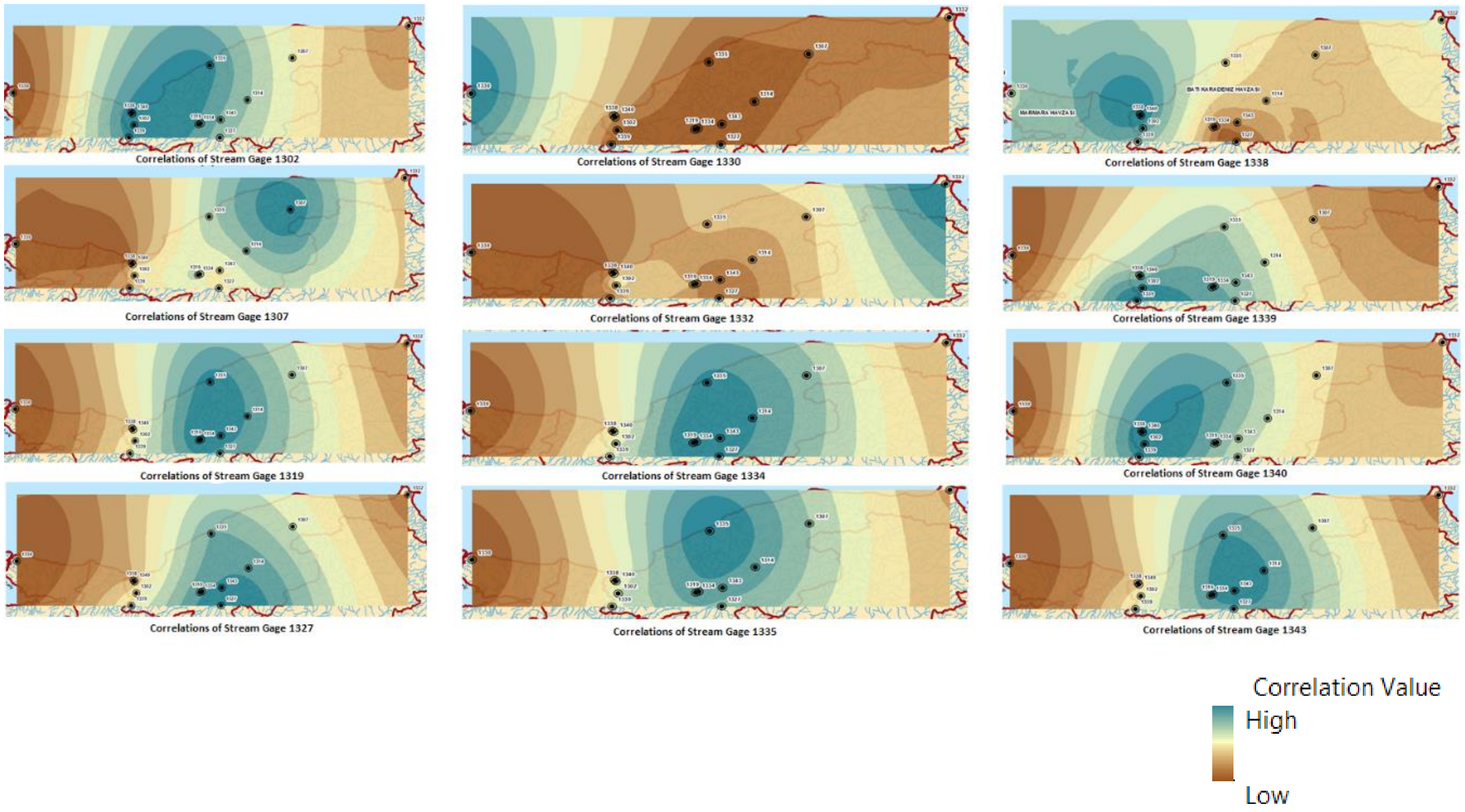


Figure 27: Correlation maps for Western Blacksea Basin where streamgage number 1314 is assumed to be ungaged

Table 1: Comparison of correlation coefficients estimated by the MCM and calculated from measurements when 1314 is assumed to be ungaged.

Correlations with streamgages	Correlations calculated from the measurements	Correlations estimated by the MCM_GIS
1302	0.79	0.84
1307	0.85	0.87
1319	0.93	0.90
1327	0.93	0.88
1330	0.58	0.61
1332	0.65	0.69
1334	0.91	0.90
1335	0.91	0.91
1338	0.48	0.56
1339	0.78	0.78
1340	0.77	0.82
1343	0.91	0.91

Cross-correlation maps can be generated by OK as well, following a similar procedure. This time, the Python script called MCM for the Ungaged Basin with Ordinary Kriging is run using the designed UI. The UI for MCM for the Ungaged Basin with Ordinary Kriging is given in Figure 28. This UI requires the same inputs as the MCM for the Ungaged Basin with IDW tool except for lag size. Again, lag size has to be determined through a trial-and-error procedure.

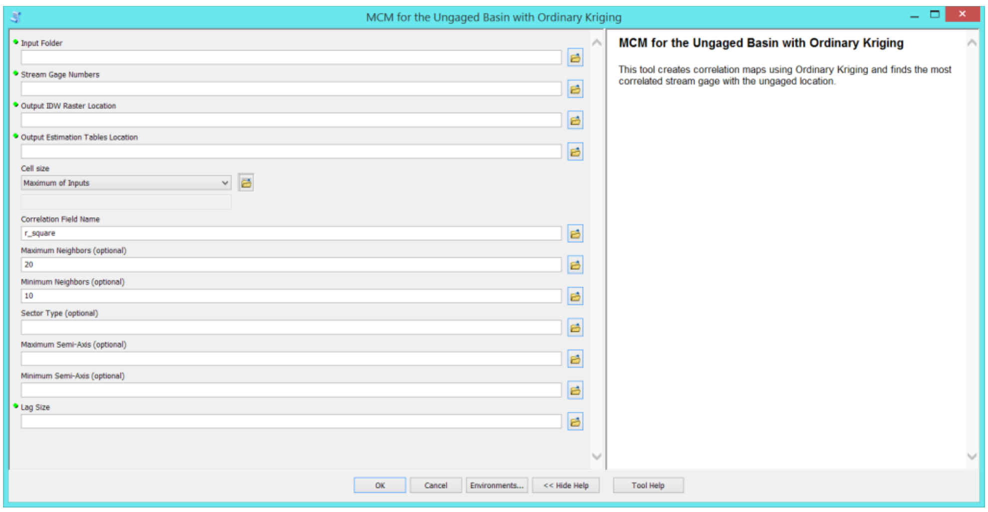


Figure 28: MCM for the Ungaged Basin with Ordinary Kriging UI

The output of MCM_GIS is a vector composed of the correlation values of the ungaged basin and each of the streamgages. Following the execution of MCM for the Ungaged Basin with Ordinary Kriging or MCM for the Ungaged Basin with IDW, estimated correlation values of streamgages with the ungaged location appears in the execution window. The most correlated streamgage can be identified from the vector of correlation coefficients and streamflow values at the ungaged basin can be estimated using DAR with the most correlated streamgage as the reference streamgage. This step has to be carried out outside the GIS environment (i.e. MCM_GIS does not carry out DAR calculations).

3.6.2 Evaluation of the Estimation Performance of MCM in the Study Area

Evaluation of the estimation performance of MCM in the study area procedure is very similar to that of MCM application for the ungaged basin using MCM_GIS. The differences are highlighted in the following paragraphs. This step is the most time-consuming step of MCM since it performs the leave-one-out cross-validation experiment.

The folder and file structure that is required by MCM_GIS as input for the evaluation of the estimation performance of MCM in the study area is given in Figure 29. In the performance evaluation of MCM in the study area, each streamgage is assumed as an ungaged location and correlations are estimated using MCM. Then these estimated correlations are compared with observed correlations to evaluate the performance of MCM in the study area. Thus, in the first level of Figure 29, there is one folder for each streamgages found in the study area. These folders correspond to the assumed ungaged locations. For the study area given in Figure 2(a), instead of one folder (i.e. ungaged_tables), there will be 11 folders named as *SGi*_tables, *i*=1302, 1307, … , 1343. Inside each folder, there is one text file for each streamgage located in the study area (see Figure 29). File_Generator script, xls2txt tool, and Add Label to Text File tool are used to prepare required text files for MCM_GIS tool using the same procedure explained above.

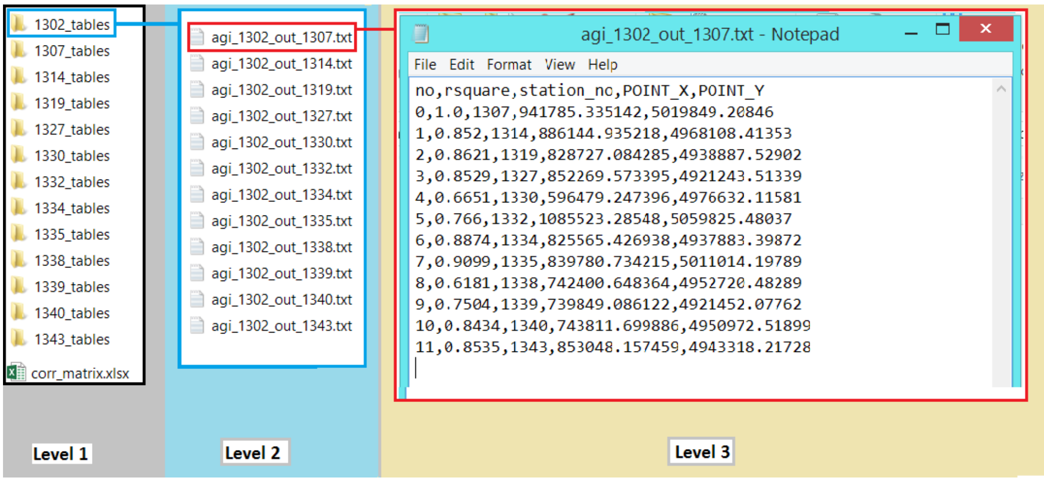


Figure 29: Folder Structure for Evaluation of Estiomation Performance of MCM

In the GIS environment, first geodatabases for each folder (i.e. streamgage assumed to be ungaged) are created using Create_Geodatabase script. Then Import Tool is used to generate feature classes. The geodatabase structure for the example given in Figure 2(a) for the evaluation of the performance of MCM in the study area is provided in Figure 30.

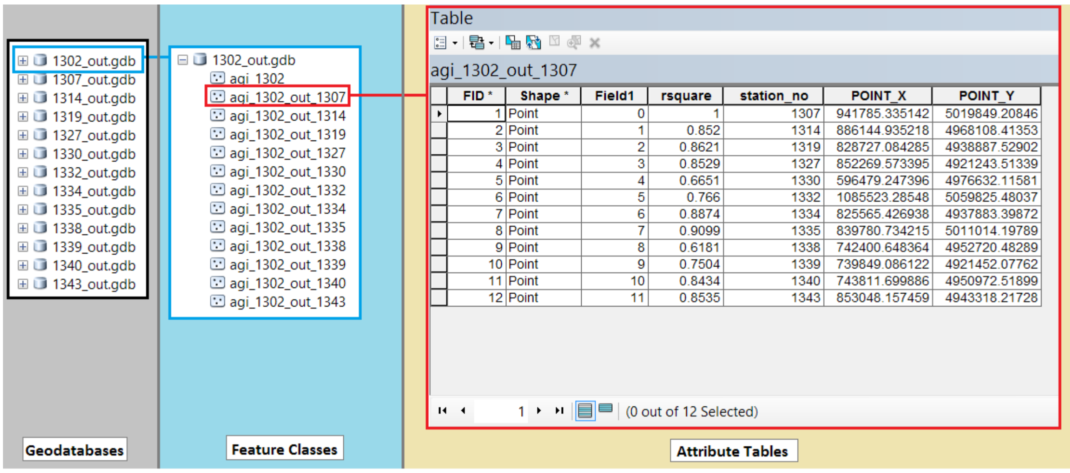


Figure 30: Geodatabase structure for the Evaluation of the Estimation performance

The final step is generation of the cross-correlation maps using either IDW or OK. To apply IDW inside the leave-one-out cross-validation procedure a Python script called Evaluation of the Estimation Performance with IDW is written (See Figure 31). This script is run through the designed UI given in Figure 32. The UI requires parameters of IDW which should be identified through a trial-and-error procedure that can be carried out using the Geostatistical Wizard of ArcMap. When the best parameters are identified, OK button is clicked to carry out leave-one-our cross-validation experiment for all the streamgages found in the study area using IDW to generate cross-correlation maps. The output of Evaluation of the Estimation Performance with IDW are the prediction surface rasters and tables containing observed and estimated correlation values for each streamgage. The same procedure can be carried out using Ordinary Kriging as well using Evaluation of the Estimation Performance with OK which is the Python script written to carry out leave-one-our corss-validation experiment with OK. The UI of Evaluation of the Estimation Performance with OK is given in Figure 33.

These tables are examined by the user to evaluate the performance of the MCM in the study area. High correlation between observed and estimated correlation values is an indicator of successful estimation. On the other hand, if there is no significant relation between the observed and the estimated correlation values then, the estimation is unqualified. In this case, the user exits MCM_GIS and revises the study area or selects additional streamgages and starts the whole procedure all over again (See Figure 1). When acceptable correlations are obtained between estimated and observed correlation values MCM_GIS can be used to identify the most correlated streamgage in the study area with any ungaged basin.

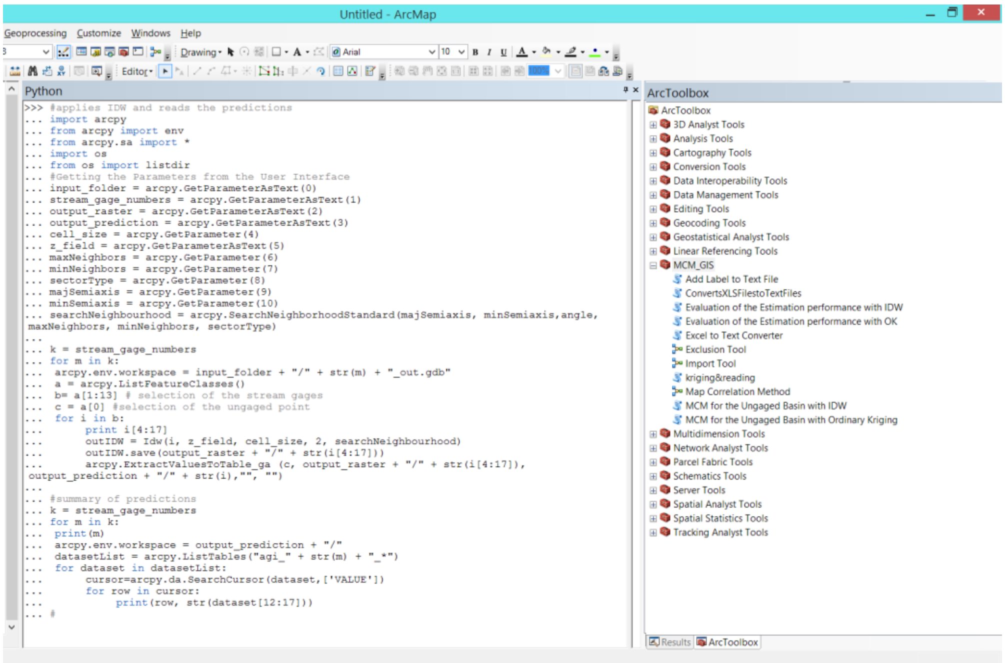


Figure 31: Python Script: Evaluation of the Estimation performance with IDW

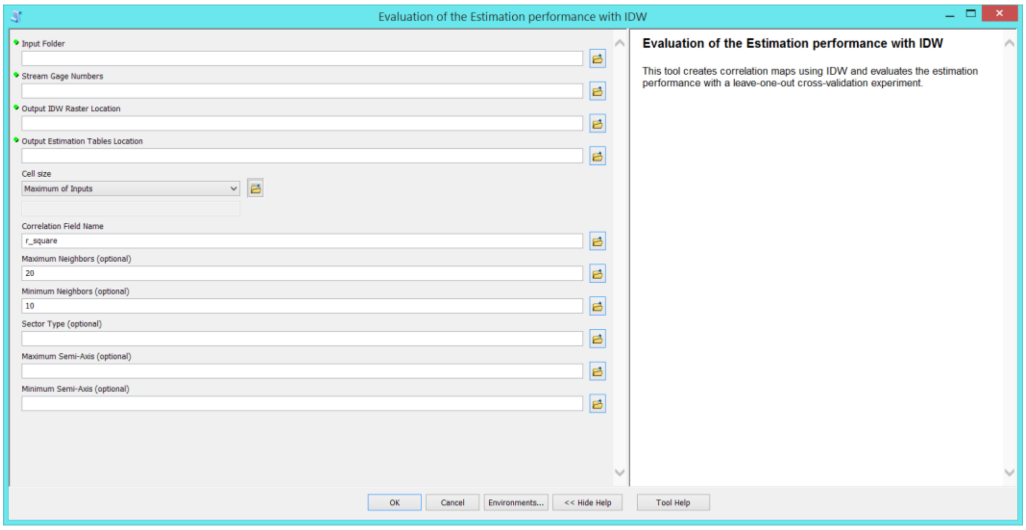


Figure 32: Evaluation of the Estimation performance with IDW UI

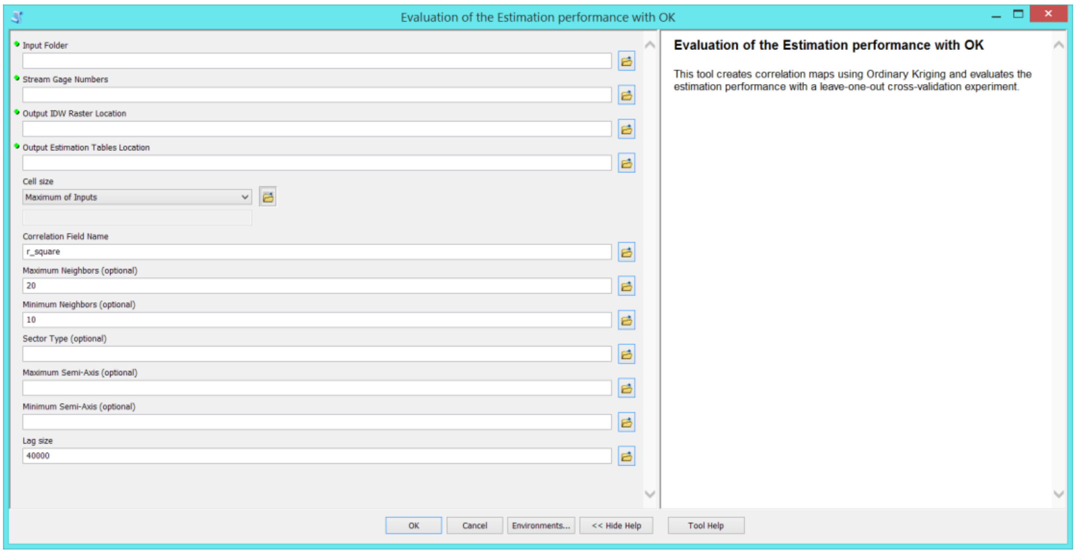


Figure 33: Evaluation of the Estimation performance with OK UI

CHAPTER 4

THE CASE STUDIES: WESTERN BLACK SEA AND CORUH REGIONS

A GIS tool, MCM_GIS is developed in this study to estimate daily streamflow at ungaged locations. Application of MCM_GIS tool is demonstrated on two case studies, at Western Black Sea and Coruh regions, and the details of these case studies are provided in this chapter.

4.1 Application of MCM_GIS at Western Black Sea Region

4.1.1 Description of the Study Area

Application of MCM_GIS is first demonstrated at a study area located in the Western Black Sea Region (Figure 34). A total of 13 stream gauges are used with 10 years (1995-2004) of common streamflow data. While selecting the streamgages, the procedure explained in Section 3.3 is followed. MCM requires utilization of daily streamflows from a large number of stream gauges. Streamgages that are numbered as 1302, 1307, 1314, 1319, 1327, 1330, 1332, 1334, 1335, 1338, 1339, 1340, and 1343 are included for this case (see Figure 34). Catchment locations, drainage areas, coordinates, and observation periods associated with each one of these 13 streamgages are given in Table 2.

Table 2: Summary of Western Black Sea Basin Streamgages

No	Catchment Location	Drainage area (km²)	Elevation (m)	Observation period (years)
1302	Buyukmelen	1988	115	1952-2004 (excluding 1963, 1971, 1972, 1992, 2008)
1307	Devrekani Cayi	1097.6	815	1953-2004 (excluding 1955, 2005)
1314	Soganli Cayi	5086.8	271	1962-2004
1319	Mengen Cayi	766.4	507	1964-2004 (excluding 1981, 1998, 2008)
1327	Ulusu	953.6	1142	1966-2004
1330	Yeniciftlik D.	23.1	39	1966-2004 (excluding 1990,1991)
1332	Karasu	340	20	1968-2004
1334	Bolu Cayi	1095.3	541	1966-2004 (excluding 1994)
1335	Filyos Cayi	13300.4	2	1963-2004
1338	Lahana Deresi	104.8	16	1979-2004 (excluding 1980, 1985, 1986)
1339	Aksu Deresi	105.2	634	1980-2004
1340	Buyukmelen	2174	23	1980-2004 (excluding 2006)
1343	Korubasi Deresi	125	780	1991-2004

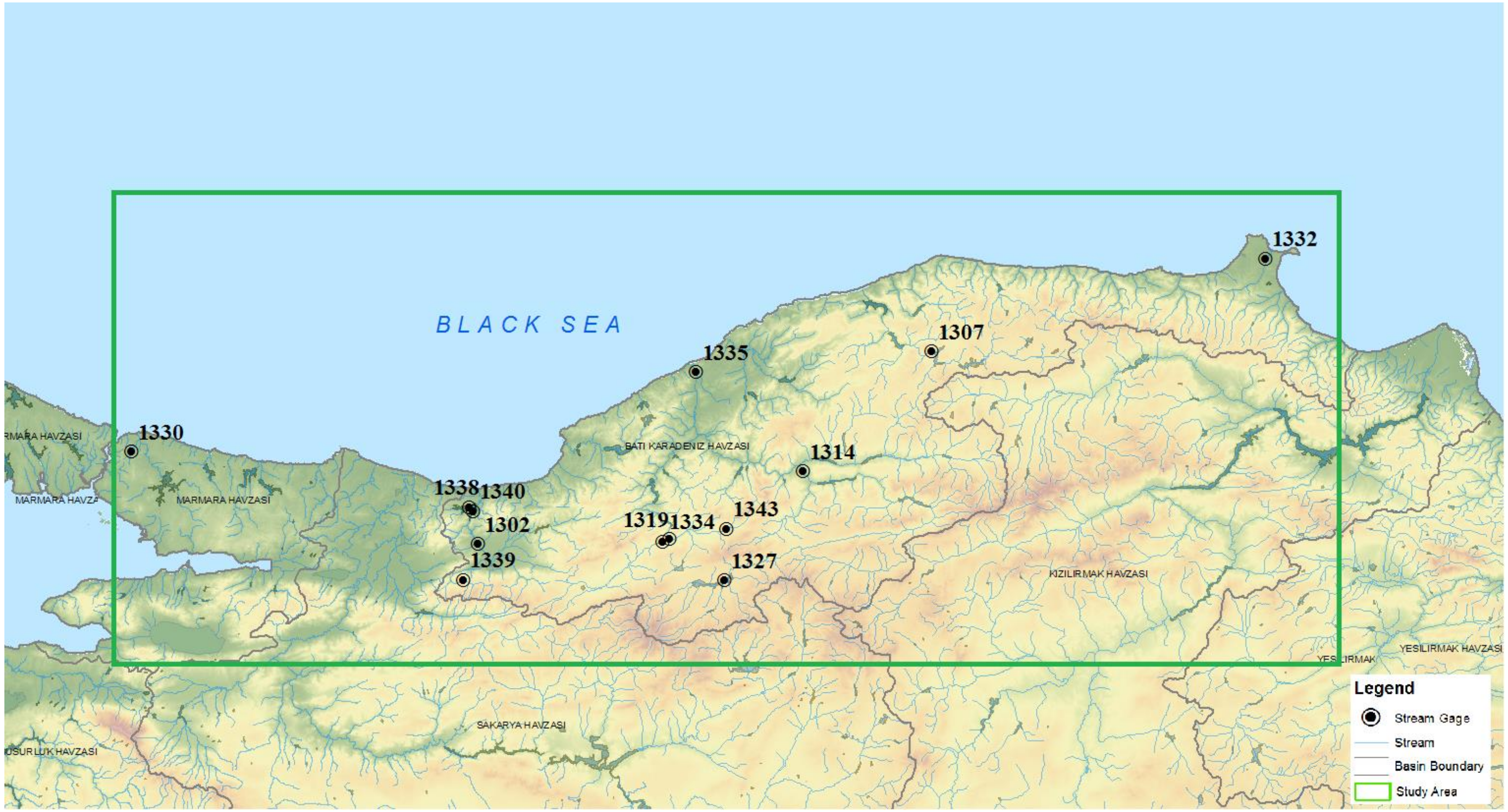


Figure 34: Study Area in Western Black Sea Region

4.1.2 Application of MCM by the GIS Tool

The methodology that is shown in the flowchart (Figure 1) is followed in the application. First, xls2txt Tool and Add Label to Text File Tool are used to obtain the file structure shown in Figure 35. The xls2txt Tool and Add Label to Text File Tool work inside the GIS environment. They are stored inside the MCM_GIS toolbox and has an easy to use interface as explained in Section 3.6.1. The correlation matrix for the streamgages found in the Western Black Sea Region are calculated and given in right hand side of Figure 36. The File_Generator script is used to convert the correlation matrix Excel file into a set of Excel files. The *corr_matrix* variable that is used inside the File Generator script, is the Excel file shown in the Figure 36.

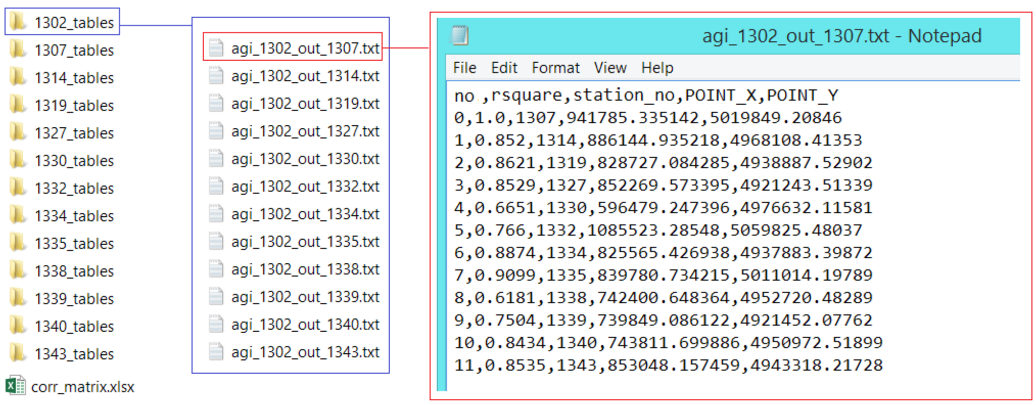


Figure 35: File Structure for Streamgages in Western Black Sea Region

Next, Import tool is used to convert the text files into feature classes (see Figure 37). At this stage, the correlation values for each streamgage are stored as spatial layers (i.e. point layers). Before moving to the next step, these point layers and other spatial layers can be viewed in the GIS environment (see Figure 38). In this map, there are streams, basin boundaries, dam locations and a digital elevation model.

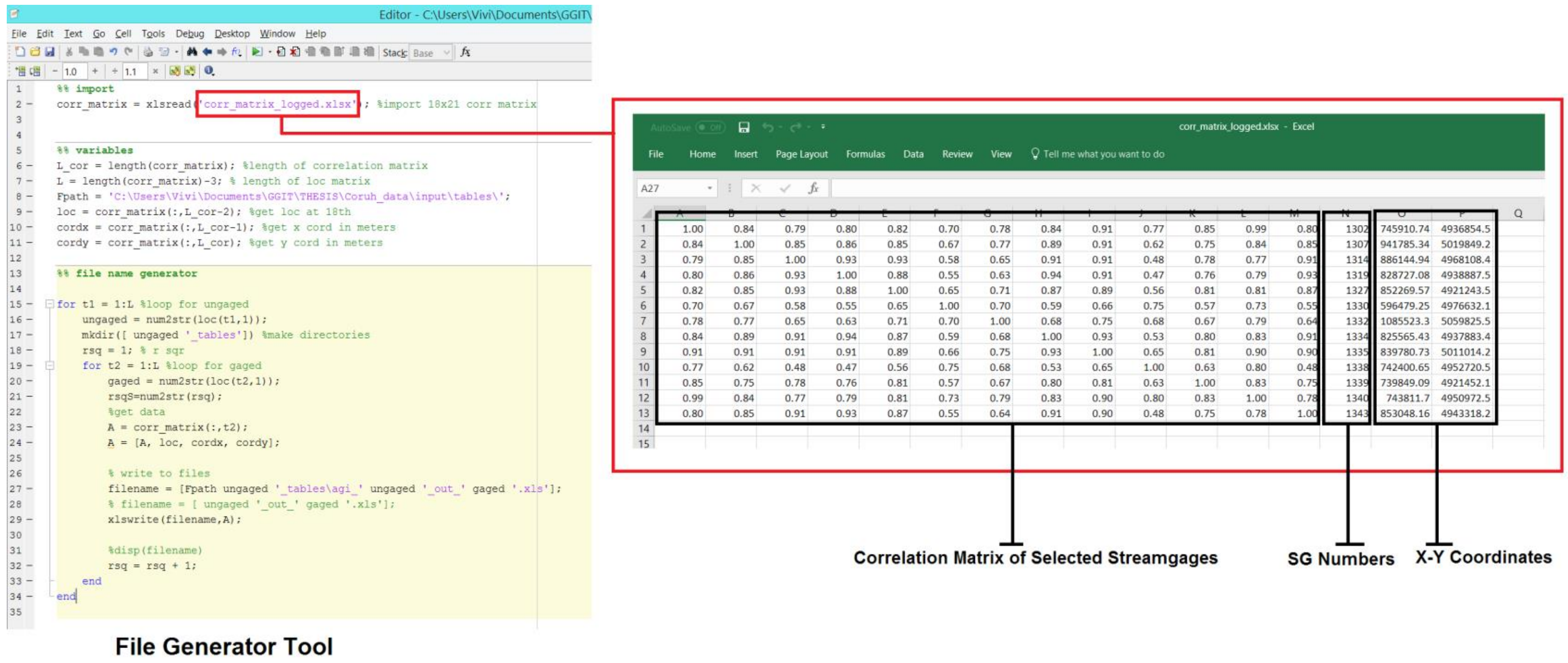


Figure 36: The correlation matrix Excel file for the streamgages found in the Western Black Sea Region

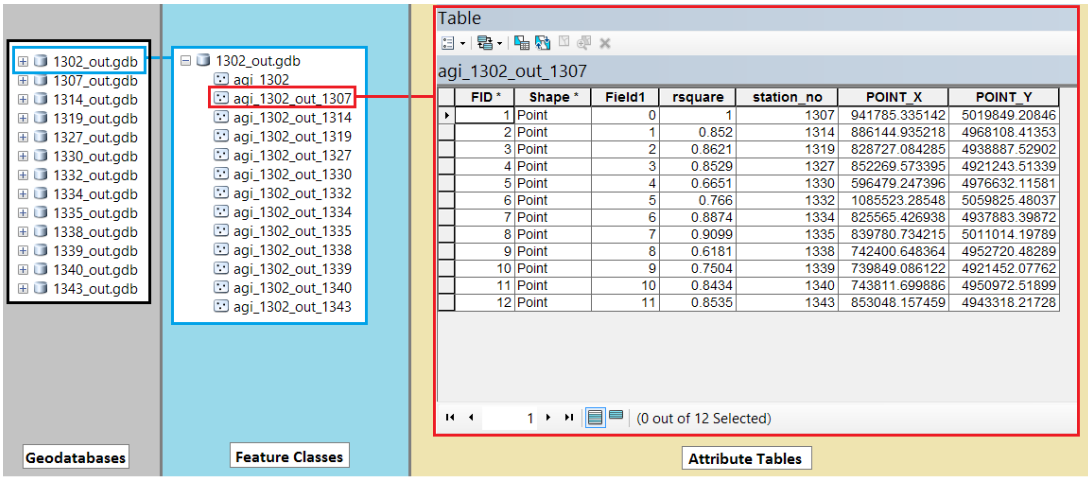


Figure 37: File Structure in GIS for Western Black Sea Basin

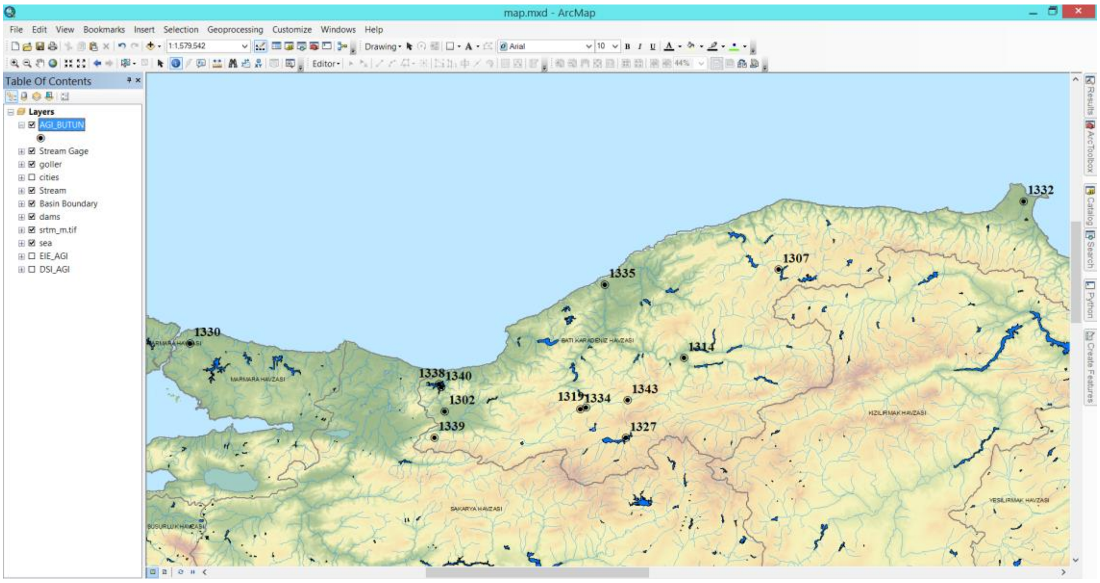


Figure 38: Digital elevation model, streamgages, dams at Western Black Sea Region

As can be seen in Table 3, out of 13 streamgages, results are obtained for 11 of them since IDW and OK were not able to carry out spatial interpolation for remaining two streamgages (i.e. 1330 and 1332). The reason for this is that, these particular streamgages are located near the boundary of the study area. As explained in Section 3.1, this problem can be fixed by keeping the point of interest (i.e. the ungaged location) close to the center while determining the study area boundaries. 1330 and 1332 are shown in Figure 39 within red squares.

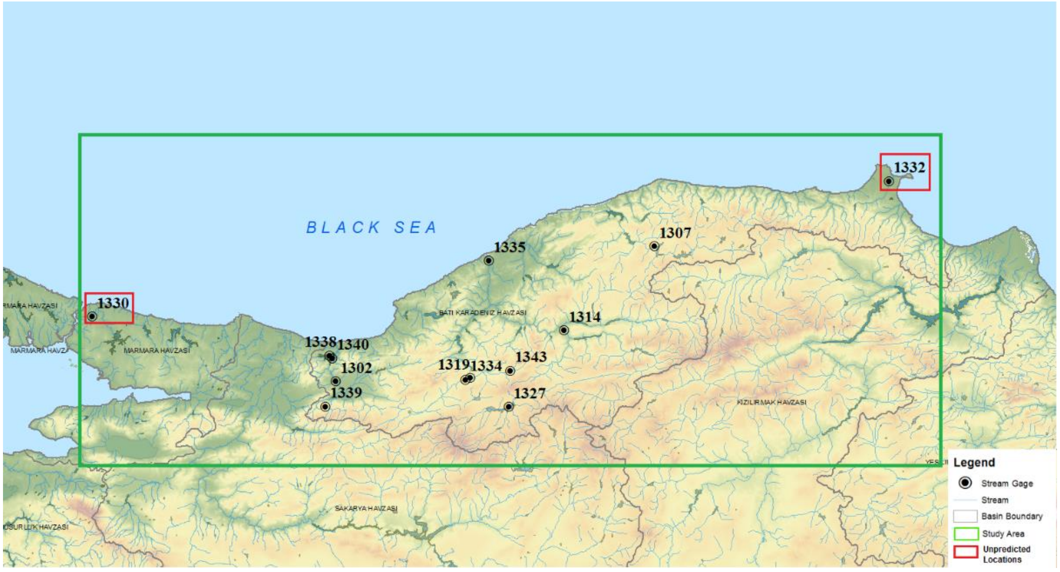


Figure 39: Unpredicted Locations in Western Black Sea Region

Leave-one-out cross-validation is carried out in this study to evaluate the performance of MCM in the study area. Estimation of r values is done by performing spatial interpolation. Cross-correlation maps created using OK, where a selected streamgage is assumed to be ungaged are given in the Figure 40, Figure 41, Figure 42, Figure 43, Figure 44, Figure 45, Figure 46, Figure 47, Figure 48, Figure 49 and Figure 50 for streamgages 1302, 1307, 1319, 1327, 1334, 1335, 1338, 1339, 1340 and 1343, respectively.

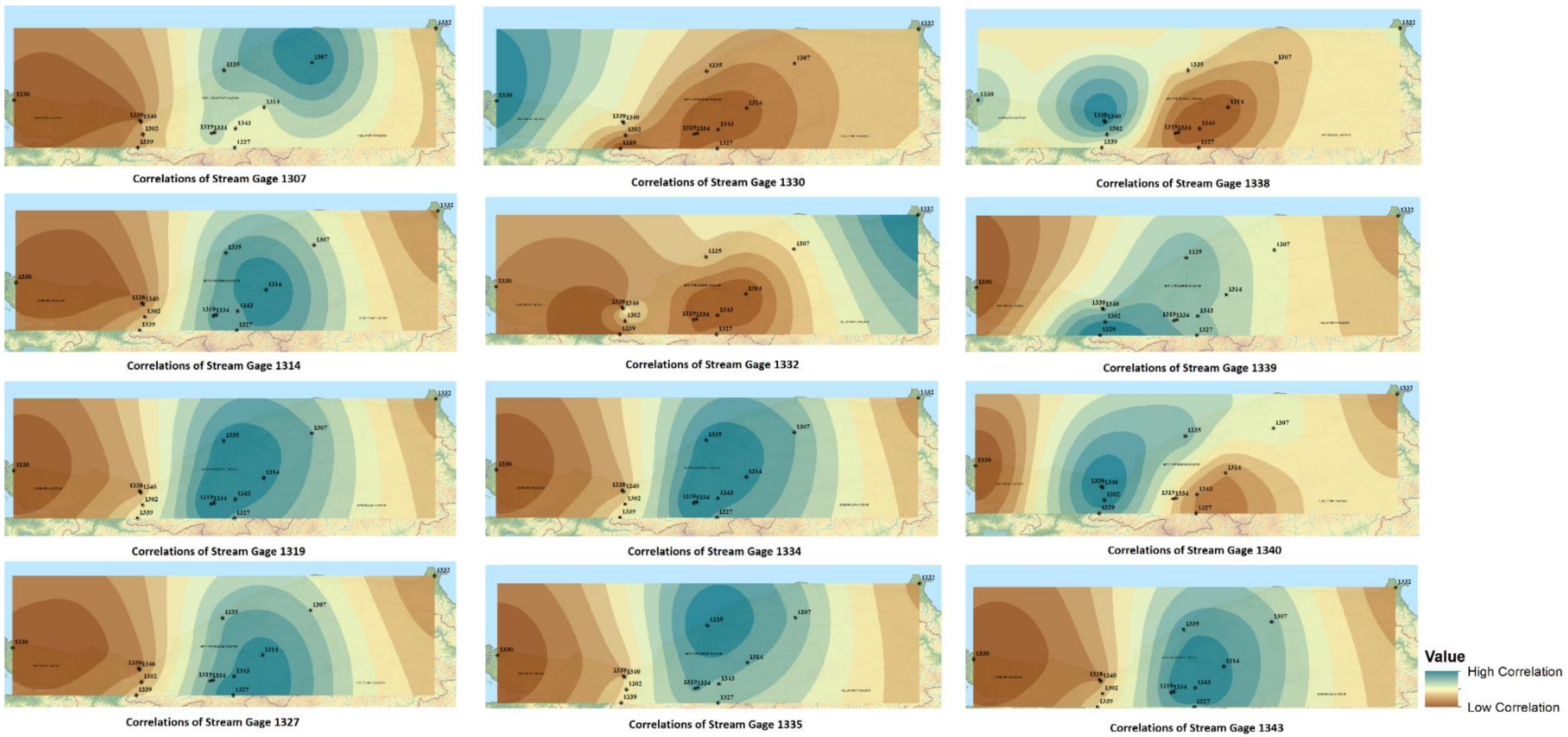


Figure 40: Correlation maps for Western Black Sea Basin where streamgage number 1302 is assumed to be ungaged (Created with OK).

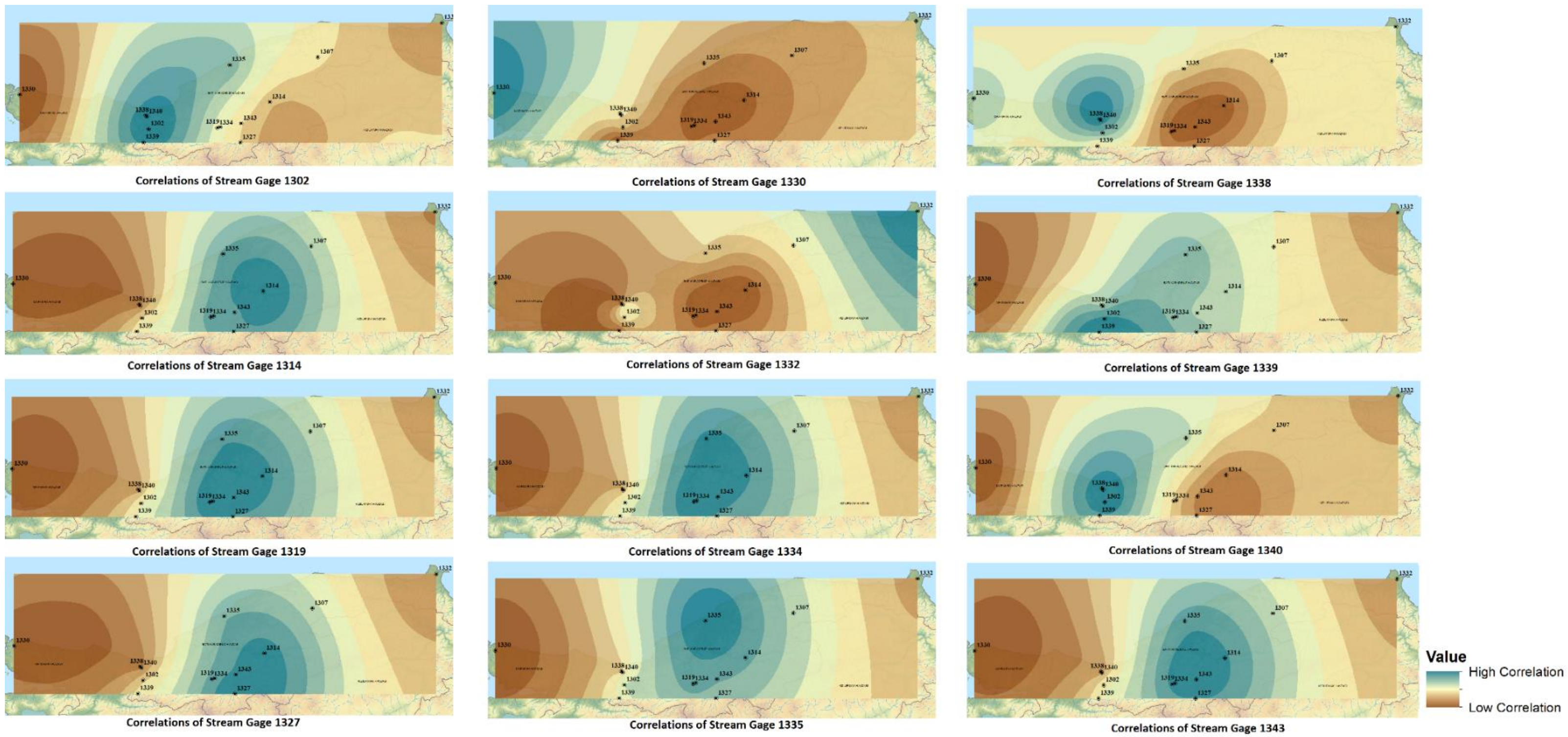


Figure 41: Correlation maps for Western Black Sea Basin where streamgage number 1307 is assumed to be ungaged (Created with OK).

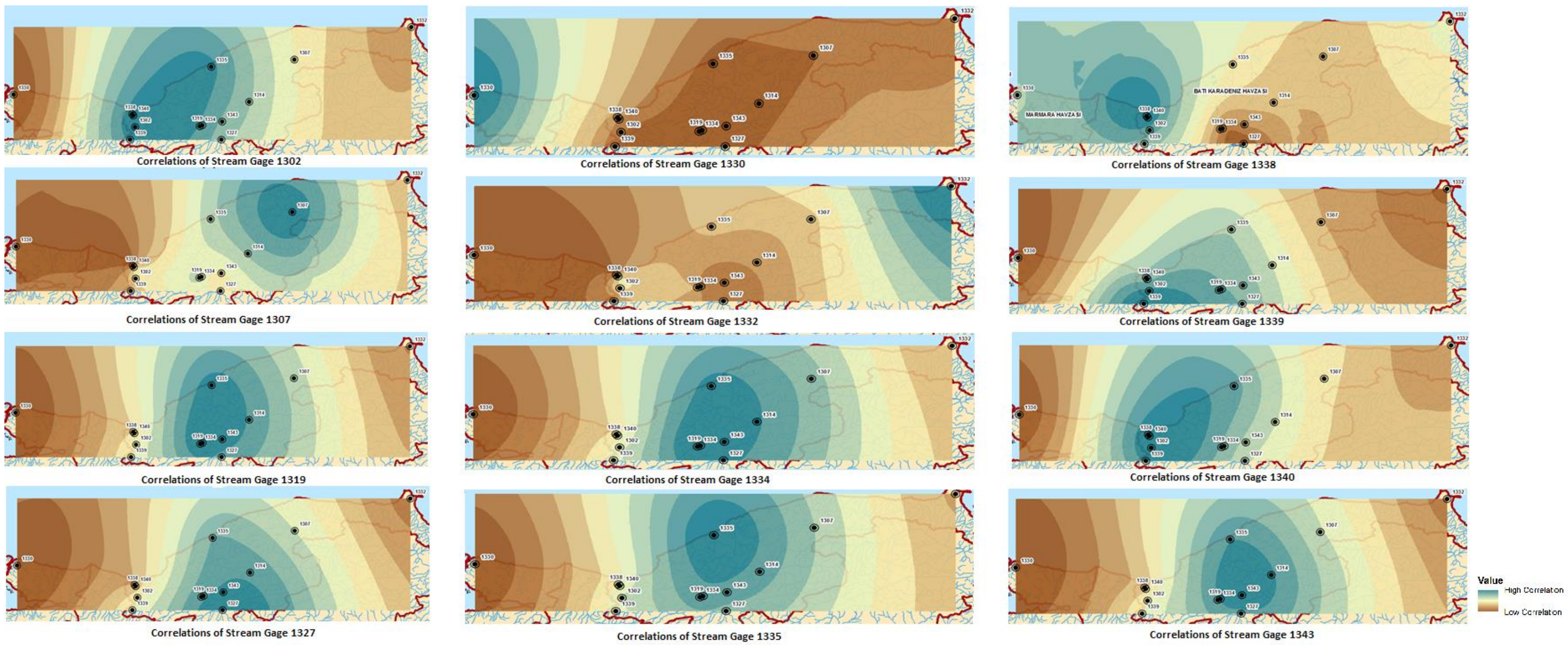


Figure 42: Correlation maps for Western Black Sea Basin where streamgage number 1314 is assumed to be ungaged (Created with OK).

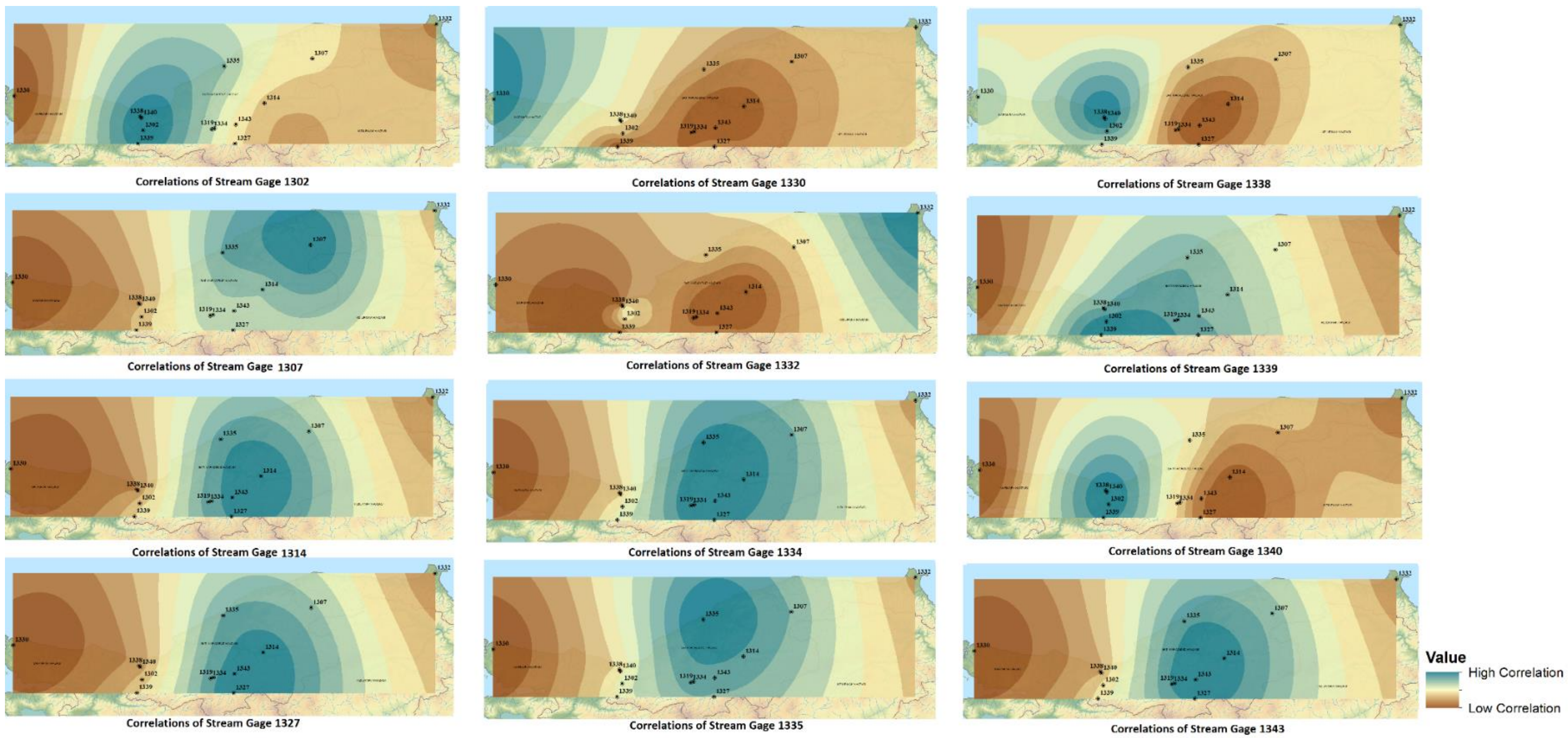


Figure 43: Correlation maps for Western Black Sea Basin where streamgage number 1319 is assumed to be ungaged (Created with OK).

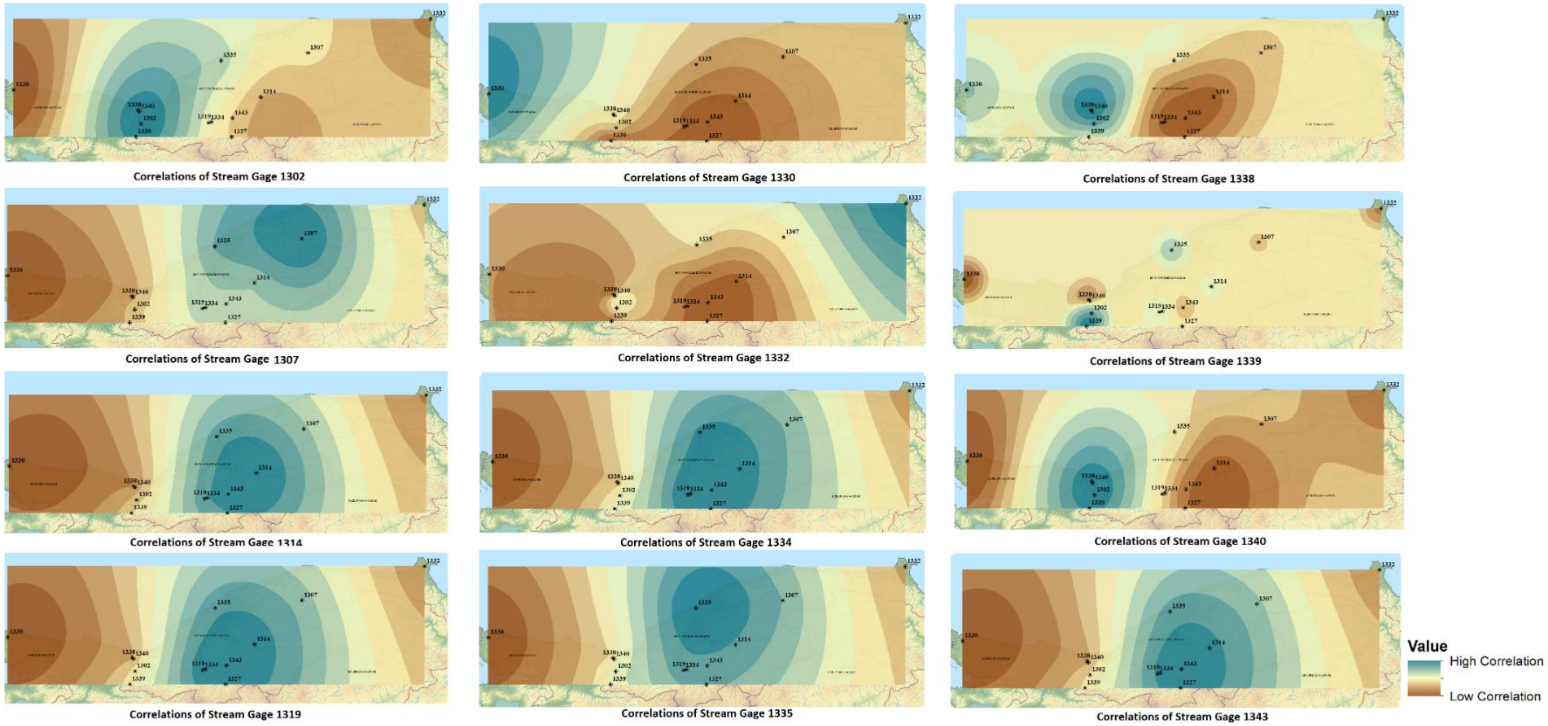


Figure 44: Correlation maps for Western Black Sea Basin where streamgage number 1327 is assumed to be ungaged (Created with OK).

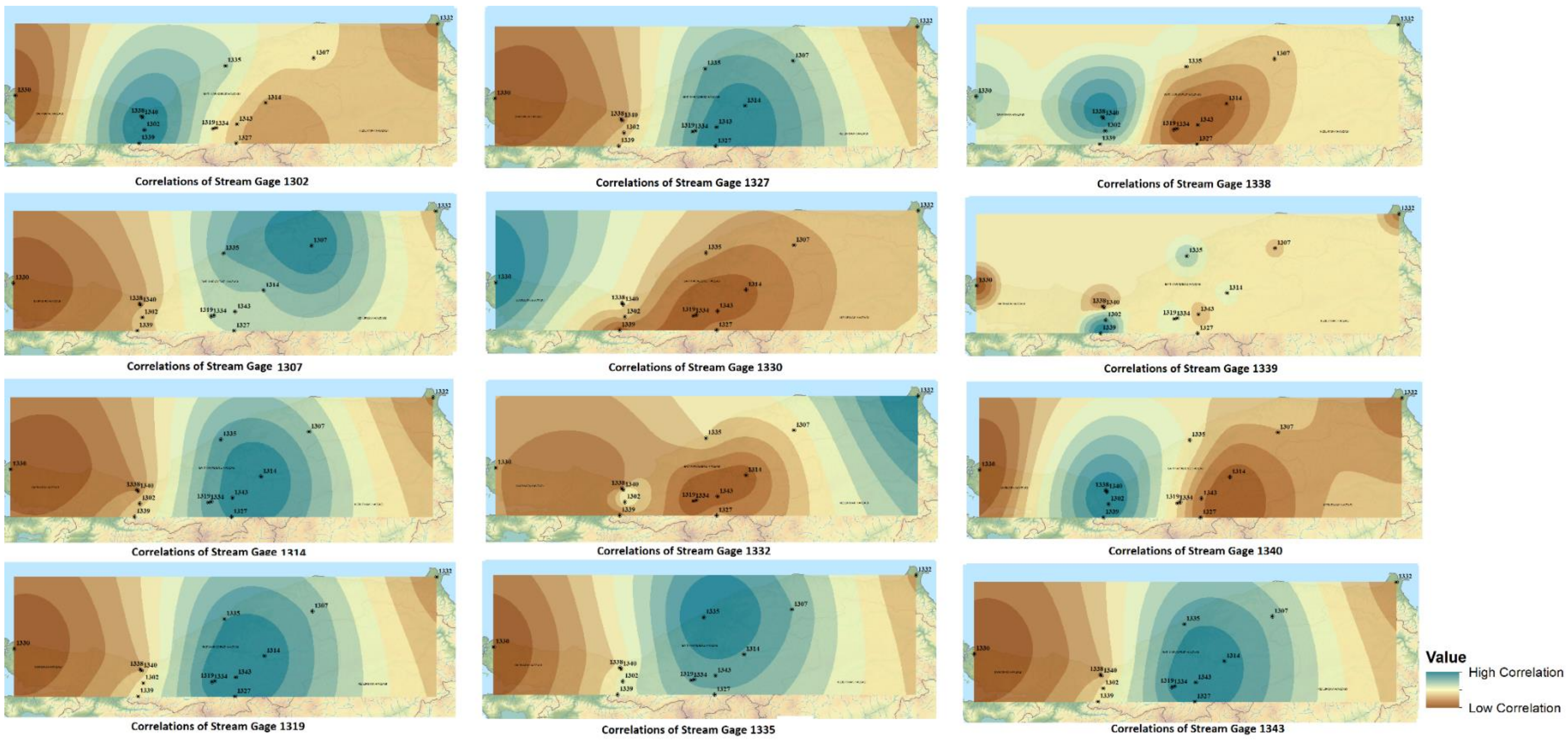


Figure 45: Correlation maps for Western Black Sea Basin where streamgage number 1334 is assumed to be ungaged (Created with OK).

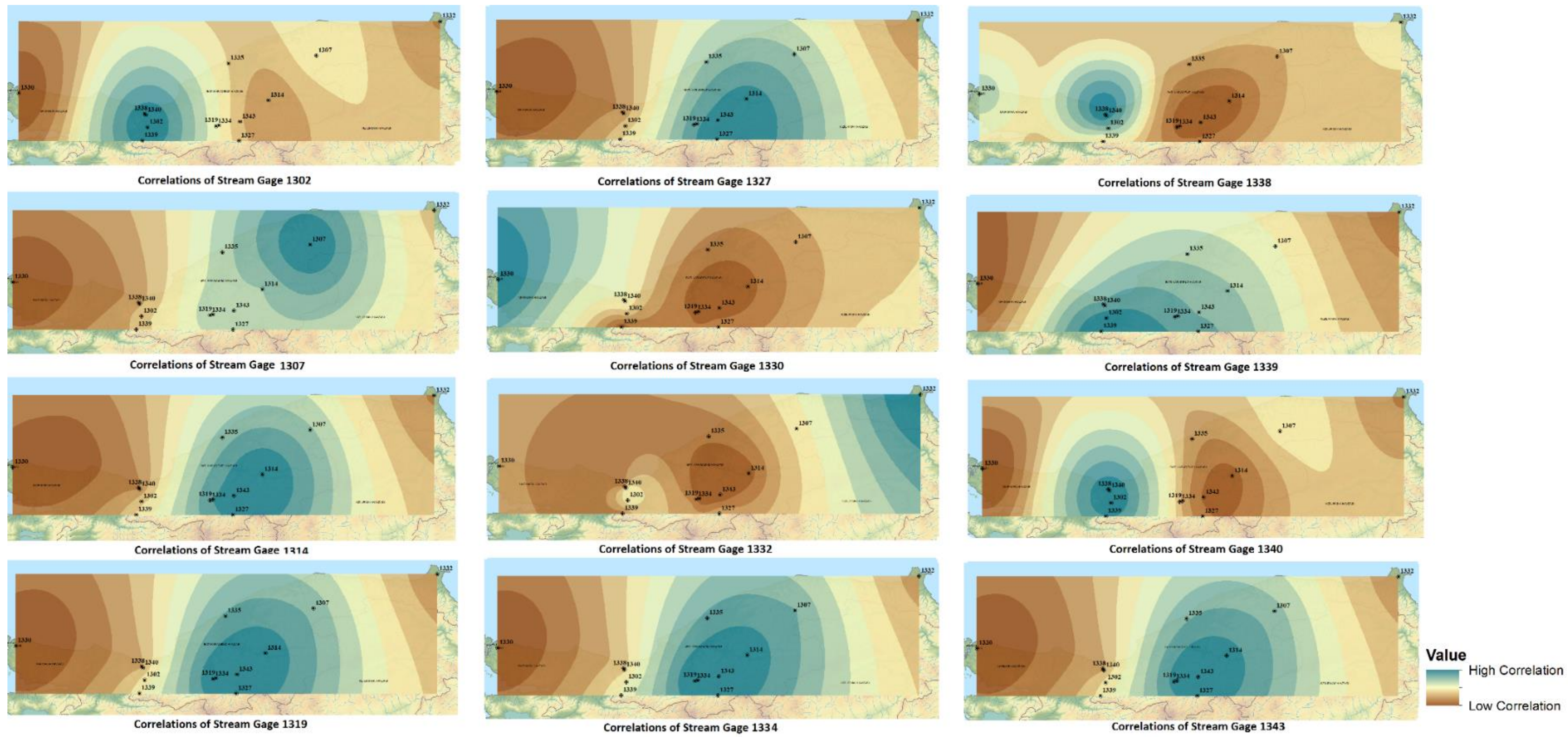


Figure 46: Correlation maps for Western Black Sea Basin where streamgage number 1335 is assumed to be ungaged (Created with OK).

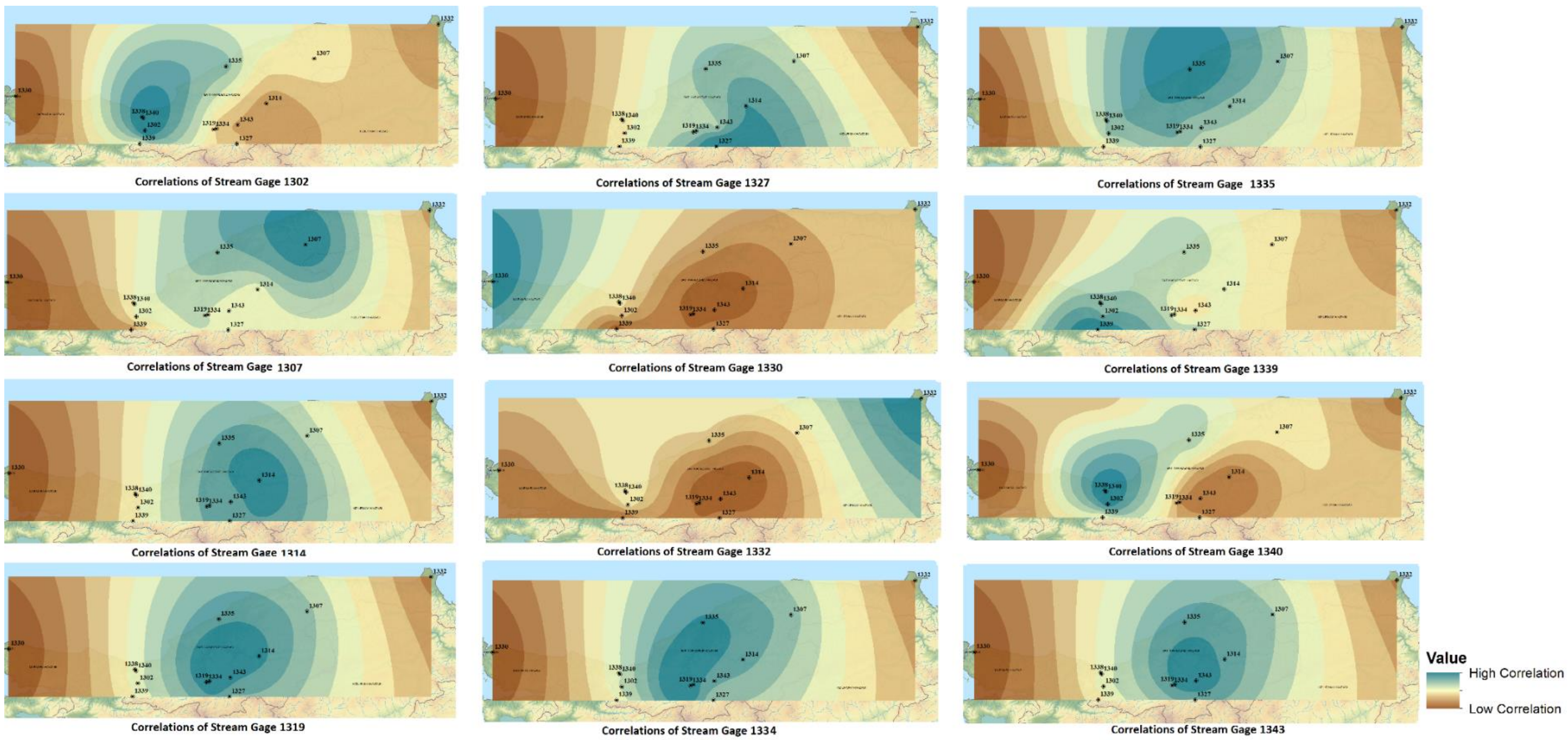


Figure 47: Correlation maps for Western Black Sea Basin where streamgage number 1338 is assumed to be ungaged (Created with OK).

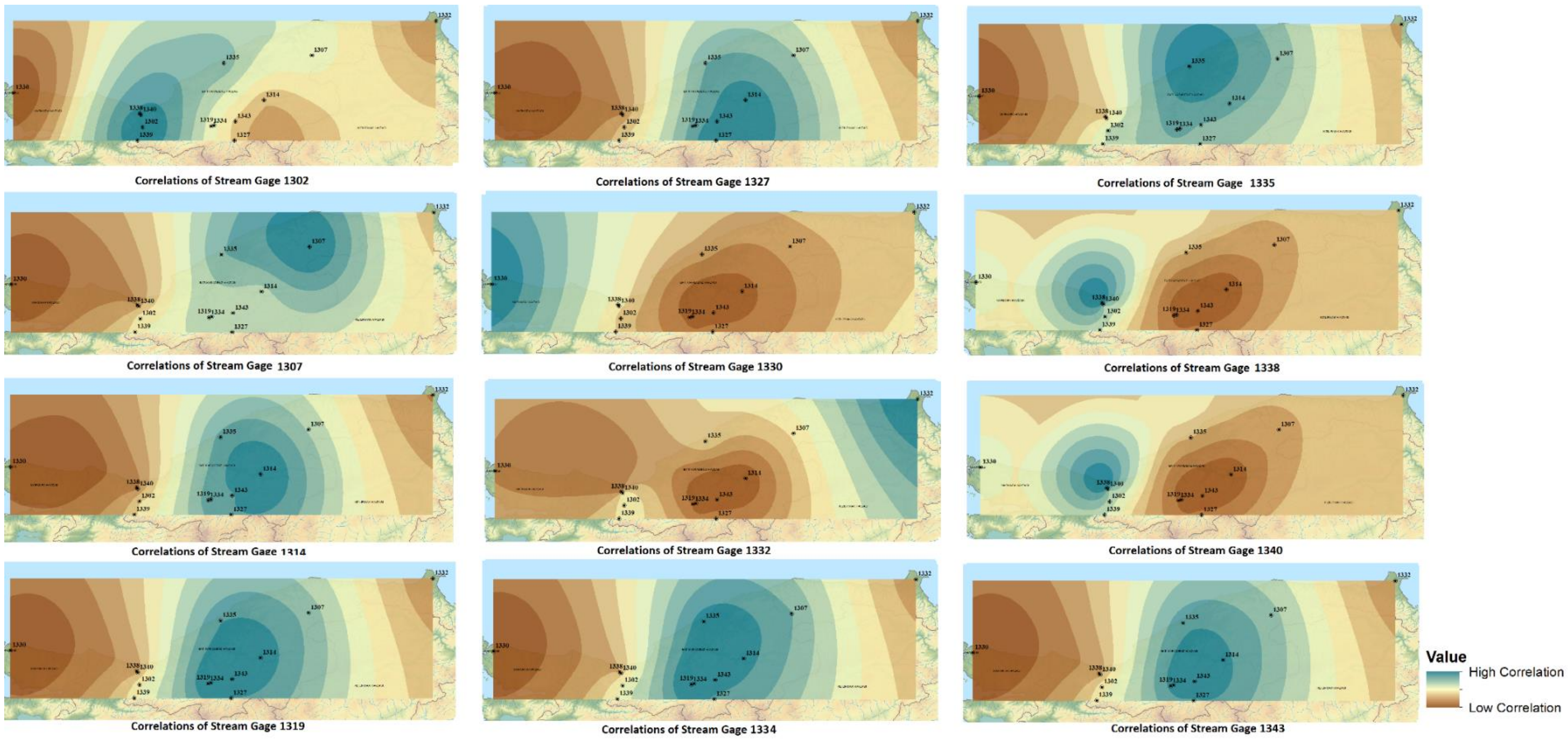


Figure 48: Correlation maps for Western Black Sea Basin where streamgage number 1339 is assumed to be ungaged (Created with OK).

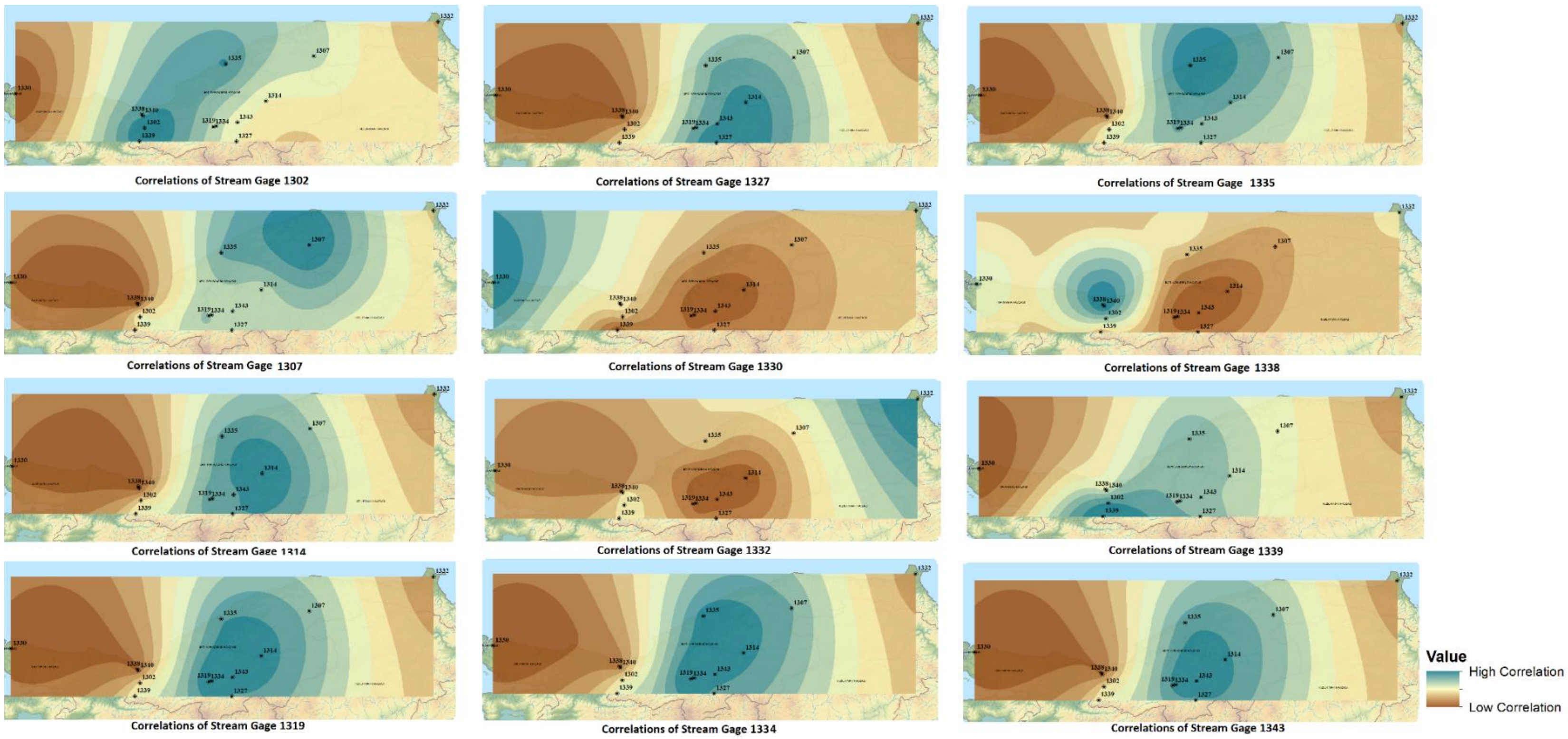


Figure 49: Correlation maps for Western Black Sea Basin where streamgage number 1340 is assumed to be ungaged (Created with OK).

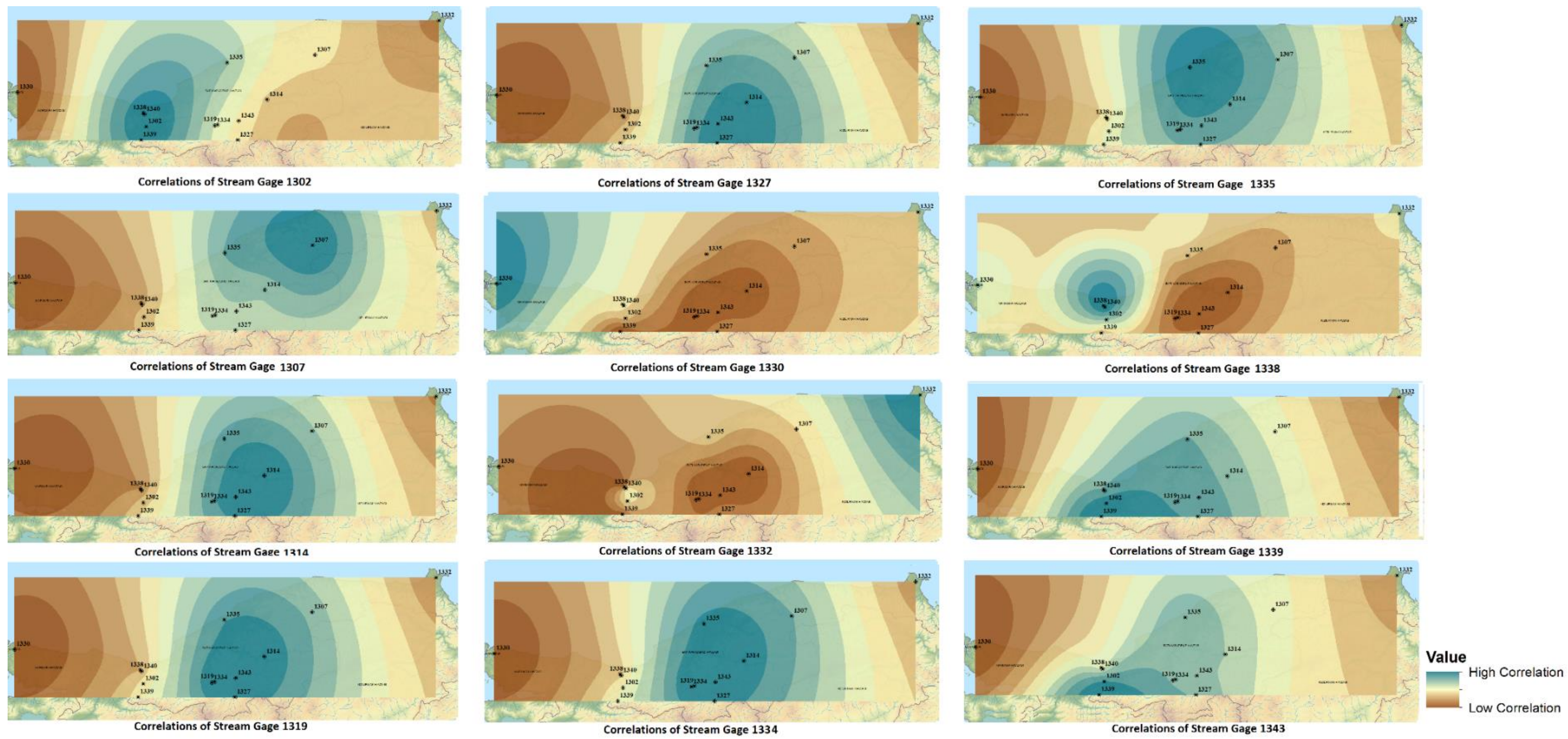


Figure 50: Correlation maps for Western Black Sea Basin where streamgage number 1343 is assumed to be ungaged (Created with OK).

4.1.3 Evaluation of the Results for Western Black Sea Region

Pearson's r correlation values are read from the cross-correlation maps. Then these estimated values are compared with the observed Pearson's r correlation values. Outcome of OK for Western Black Sea Region is given in Figure 51. Generally, the estimated r values are in good agreement for both study areas, except for a few of the streamgages.

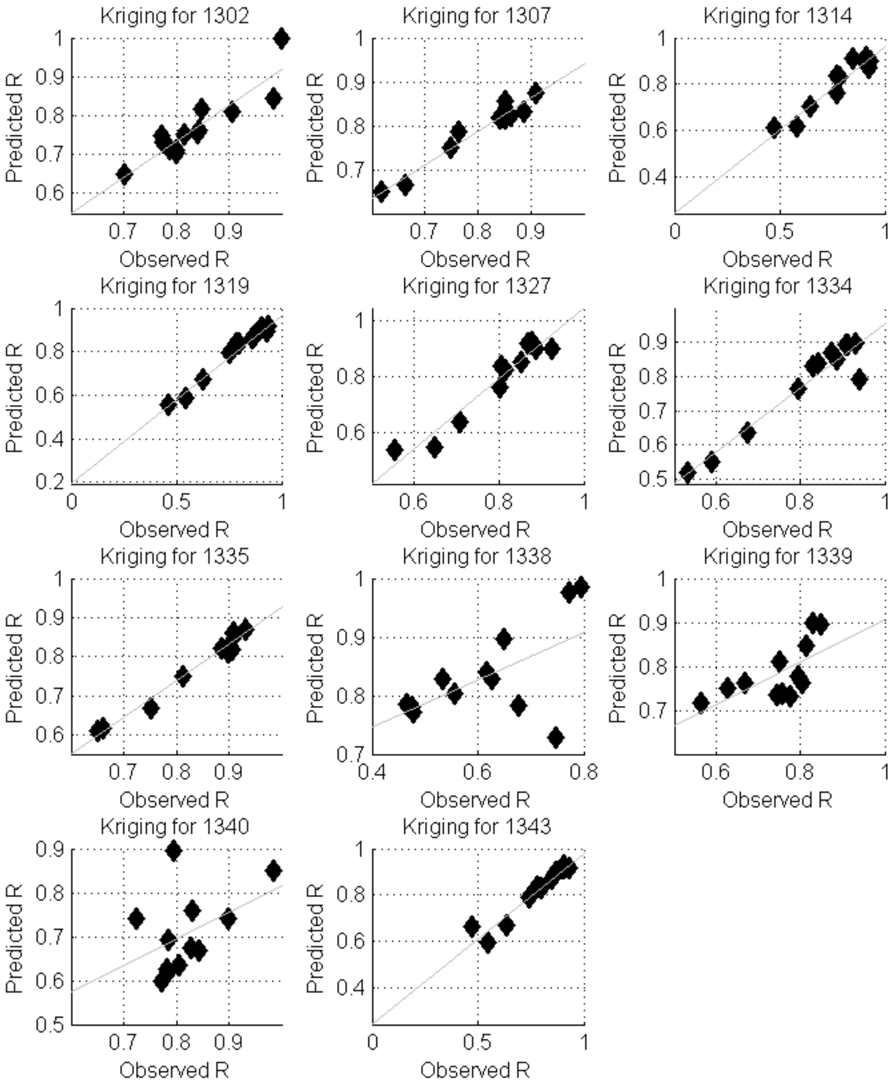


Figure 51: Evaluation of Correlation Estimations with Ordinary Kriging for Western Black Sea Region

Comparison of the observed r and estimated r values provides an important indicator about the performance of the estimation. As it can be seen from the Figure 51, for streamgages 1338 and 1340 estimated r values are not close to observed ones. Streamgage 1338 is located in a very small and isolated basin, this might be the reason for low correlation values with the rest of the streamgages. For streamgage 1340, as can be seen from Table 2, all methods (i.e. ones calculated using the nearest or the most correlated streamgages) result in low estimations. One reason might be unknown or unrecorded regulations around this streamgage or errors in recording.

There was no significant performance difference between OK with circular model and IDW methods. However OK with spherical model was able to identify the most correlated streamgage for two of the streamgages where OK with circular model identified second or third most correlated streamgages. The output of correlation estimations are shown in Figure 52, Figure 53 and Figure 54 for IDW, OK with circular model and OK with spherical model respectively in detail. Most correlated streamgages for the measurements and the estimations are shown with bold characters.

Ungaged	1302		1307		1314		1319		1327	
	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation
1302	1.00	1.00	0.84	0.83	0.79	0.84	0.80	0.84	0.82	0.82
1307	0.84	0.75	1.00	1.00	0.85	0.87	0.86	0.89	0.85	0.86
1314	0.79	0.70	0.85	0.89	1.00	1.00	0.93	0.91	**0.93**	0.91
1319	0.80	0.69	0.86	0.88	**0.93**	0.90	1.00	1.00	0.88	0.93
1327	0.82	0.74	0.85	0.87	0.93	0.88	0.88	0.88	1.00	1.00
1330	0.70	0.68	0.67	0.62	0.58	0.61	0.55	0.59	0.65	0.57
1332	0.78	0.72	0.77	0.71	0.65	0.69	0.63	0.68	0.71	0.65
1334	0.84	0.74	0.89	0.88	0.91	0.90	**0.94**	**1.00**	0.87	0.93
1335	0.91	0.80	**0.91**	**0.90**	0.91	**0.91**	0.91	0.93	0.89	0.91
1338	0.77	0.79	0.62	0.57	0.48	0.56	0.47	0.53	0.56	0.51
1339	0.85	0.82	0.75	0.77	0.78	0.78	0.76	0.80	0.81	0.77
1340	**0.99**	**0.88**	0.84	0.82	0.77	0.82	0.79	0.83	0.81	0.80
1343	0.80	0.69	0.85	0.87	0.91	0.91	0.93	0.91	0.87	**0.94**

Ungaged	1334		1335		1338		1339		1340		1343	
	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation
1302	0.84	0.80	0.91	0.83	0.77	0.98	**0.85**	**0.94**	**0.99**	0.78	0.80	0.82
1307	0.89	0.86	0.91	0.85	0.62	0.84	0.75	0.81	0.84	0.63	0.85	0.86
1314	0.91	0.93	0.91	0.88	0.48	0.77	0.78	0.75	0.77	0.49	0.91	0.92
1319	**0.94**	**1.00**	0.91	0.88	0.47	0.79	0.76	0.75	0.79	0.48	**0.93**	**0.92**
1327	0.87	0.88	0.89	0.87	0.56	0.81	0.81	0.78	0.81	0.57	0.87	0.92
1330	0.59	0.55	0.66	0.61	0.75	0.72	0.57	0.71	0.73	0.74	0.55	0.61
1332	0.68	0.63	0.75	0.68	0.68	0.78	0.67	0.75	0.79	0.68	0.64	0.68
1334	1.00	1.00	**0.93**	**0.89**	0.53	0.83	0.80	0.80	0.83	0.55	0.91	0.92
1335	0.93	0.91	1.00	1.00	0.65	0.90	0.81	0.86	0.90	0.66	0.90	0.91
1338	0.53	0.47	0.65	0.57	1.00	1.00	0.63	0.79	0.80	**0.99**	0.48	0.54
1339	0.80	0.76	0.81	0.78	0.63	0.83	1.00	1.00	0.83	0.64	0.75	0.79
1340	0.83	0.79	0.90	0.82	**0.80**	**1.00**	0.83	0.94	1.00	1.00	0.78	0.81
1343	0.91	0.93	0.90	0.87	0.48	0.78	0.75	0.75	0.78	0.49	1.00	1.00

Figure 52: Output correlation estimation table for MCM_GIS application with IDW

Ungaged	1302		1307		1314		1319		1327	
	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation
1302	1.00	1.00	0.84	0.82	0.79	0.83	0.80	0.84	0.82	0.83
1307	0.84	0.76	1.00	1.00	0.85	0.91	0.86	0.87	0.85	0.85
1314	0.79	0.71	0.85	0.86	1.00	1.00	0.93	0.89	**0.93**	0.90
1319	0.80	0.71	0.86	0.82	**0.93**	0.90	1.00	1.00	0.88	**0.92**
1327	0.82	0.75	0.85	0.83	0.93	0.87	0.88	0.87	1.00	1.00
1330	0.70	0.65	0.67	0.66	0.58	0.62	0.55	0.59	0.65	0.55
1332	0.78	0.73	0.77	0.79	0.65	0.70	0.63	0.67	0.71	0.64
1334	0.84	0.75	0.89	0.83	0.91	0.90	**0.94**	**0.92**	0.87	0.92
1335	0.91	0.81	**0.91**	**0.87**	0.91	**0.92**	0.91	0.91	0.89	0.90
1338	0.77	0.75	0.62	0.65	0.48	0.61	0.47	0.56	0.56	0.54
1339	0.85	0.82	0.75	0.75	0.78	0.76	0.76	0.79	0.81	0.76
1340	**0.99**	**0.84**	0.84	0.82	0.77	0.83	0.79	0.84	0.81	0.84
1343	0.80	0.70	0.85	0.82	0.91	0.90	0.93	0.89	0.87	0.92

Ungaged	1334		1335		1338		1339		1340		1343	
	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation
1302	0.84	0.84	0.91	0.82	0.77	0.98	**0.85**	0.89	**0.99**	0.85	0.80	0.83
1307	0.89	0.85	0.91	0.86	0.62	0.84	0.75	0.81	0.84	0.67	0.85	0.87
1314	0.91	0.89	0.91	0.85	0.48	0.77	0.78	0.73	0.77	0.60	0.91	**0.93**
1319	**0.94**	0.79	0.91	0.86	0.47	0.78	0.76	0.74	0.79	0.62	**0.93**	0.92
1327	0.87	0.87	0.89	0.82	0.56	0.80	0.81	0.76	0.81	0.63	0.87	0.90
1330	0.59	0.55	0.66	0.62	0.75	0.73	0.57	0.72	0.73	0.74	0.55	0.59
1332	0.68	0.64	0.75	0.67	0.68	0.78	0.67	0.76	0.79	0.69	0.64	0.67
1334	1.00	1.00	**0.93**	**0.87**	0.53	0.83	0.80	0.78	0.83	0.68	0.91	0.92
1335	0.93	**0.90**	1.00	1.00	0.65	0.90	0.81	0.85	0.90	0.74	0.90	0.92
1338	0.53	0.52	0.65	0.61	1.00	1.00	0.63	0.75	0.80	**0.90**	0.48	0.66
1339	0.80	0.77	0.81	0.75	0.63	0.83	1.00	1.00	0.83	0.76	0.75	0.79
1340	0.83	0.83	0.90	0.81	**0.80**	**0.98**	0.83	**0.90**	1.00	1.00	0.78	0.84
1343	0.91	0.89	0.90	0.83	0.48	0.78	0.75	0.73	0.78	0.63	1.00	1.00

Figure 53: Output correlation estimation table for MCM_GIS application with OK, circular model

Ungaged	1302		1307		1314		1319		1327	
	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation
1302	1.00	1.00	0.84	0.82	0.79	0.83	0.80	0.84	0.82	0.83
1307	0.84	0.77	1.00	1.00	0.85	0.91	0.86	0.87	0.85	0.85
1314	0.79	0.72	0.85	0.86	1.00	1.00	0.93	0.89	**0.93**	0.90
1319	0.80	0.80	0.86	0.82	**0.93**	0.90	1.00	1.00	0.88	**0.92**
1327	0.82	0.75	0.85	0.83	0.93	0.87	0.88	0.88	1.00	1.00
1330	0.70	0.65	0.67	0.66	0.58	0.61	0.55	0.59	0.65	0.55
1332	0.78	0.73	0.77	0.79	0.65	0.70	0.63	0.67	0.71	0.64
1334	0.84	0.75	0.89	0.83	0.91	0.90	**0.94**	**0.92**	0.87	0.92
1335	0.91	0.81	**0.91**	**0.87**	0.91	**0.92**	0.91	0.91	0.89	0.90
1338	0.77	0.75	0.62	0.65	0.48	0.62	0.47	0.56	0.56	0.54
1339	0.85	0.82	0.75	0.74	0.78	0.76	0.76	0.79	0.81	0.76
1340	**0.99**	**0.84**	0.84	0.82	0.77	0.83	0.79	0.84	0.81	0.84
1343	0.80	0.70	0.85	0.81	0.91	0.90	0.93	0.89	0.87	0.92

Ungaged	1334		1335		1338		1339		1340		1343	
	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation
1302	0.84	0.83	0.91	0.82	0.77	0.98	**0.85**	**0.89**	**0.99**	0.85	0.80	0.83
1307	0.89	0.85	0.91	0.86	0.62	0.84	0.75	0.81	0.84	0.66	0.85	0.87
1314	0.91	0.90	0.91	0.86	0.48	0.77	0.78	0.74	0.77	0.59	0.91	**0.93**
1319	**0.94**	**0.92**	0.91	0.86	0.47	0.78	0.76	0.74	0.79	0.60	**0.93**	0.92
1327	0.87	0.87	0.89	0.84	0.56	0.80	0.81	0.76	0.81	0.68	0.87	0.90
1330	0.59	0.55	0.66	0.62	0.75	0.73	0.57	0.72	0.73	0.74	0.55	0.60
1332	0.68	0.64	0.75	0.67	0.68	0.78	0.67	0.76	0.79	0.69	0.64	0.67
1334	1.00	1.00	**0.93**	**0.87**	0.53	0.83	0.80	0.78	0.83	0.67	0.91	0.92
1335	0.93	0.90	1.00	1.00	0.65	0.90	0.81	0.85	0.90	0.73	0.90	0.91
1338	0.53	0.52	0.65	0.61	1.00	1.00	0.63	0.75	0.80	**0.95**	0.48	0.66
1339	0.80	0.77	0.81	0.76	0.63	0.83	1.00	1.00	0.83	0.76	0.75	0.79
1340	0.83	0.83	0.90	0.81	**0.80**	**0.99**	0.83	0.89	1.00	1.00	0.78	0.84
1343	0.91	0.90	0.90	0.85	0.48	0.78	0.75	0.74	0.78	0.61	1.00	1.00

Figure 54: Output correlation estimation table for MCM_GIS application with OK, spherical model

The Nash-Sutcliffe Efficiency, *NSE* values are determined in order to evaluate the goodness of fit of the estimation which is performed by MCM_GIS. *NSE* is a commonly used performance indicator in hydrological estimations (Archfield and Vogel, 2009; Johnston et al., 2001; Shu and Ouarda, 2012; Zambelli et al., 2013). *NSE* values are calculated using the following equation (Nash and Sutcliffe, 1970) ;

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}(X_{\mathrm{i}}^{obs} - X_{i}^{est})^2}{\sum_{i=1}^{n}(X_{\mathrm{i}}^{obs} - \overline{X^{obs}})^2}\right] \tag{10}$$

where X_i^{obs} is the *i*-th observed value, X_i^{est} is the *i*-th estimated data and $\overline{X^{obs}}$ is the averaged value of all observed data. *NSE* values that are lower than zero indicates that the average of the observed data is a better estimator than the model applied.

Donor/reference streamgages are identified using both IDW, OK with circular model type and OK with spherical model type, and DAR is used to estimate streamflow at ungaged locations. The results are compared with those obtained using the closest streamgage as the reference streamgage. Performances of these four approaches are compared in Table 3. The performance of the estimations improved for most of the streamgages when the donor is selected using the MCM instead of the nearest streamgage. The performance decreased only for 1343 when OK is used for the estimation. There are certain studies stating kriging is a better interpolator than IDW, while estimating geographical data (Johnston et al., 2001; Luo et al., 2008; Mbilinyi et al., 2007). Even though the overall performance of two different interpolation methods are very similar in our study, for one of the streamgages (i.e. 1343) IDW performed better.

Table 3: Nash-Sutcliffe Efficiency values identified using different reference gage selection criteria

Ungaged	Nearest Streamgage	The Most Correlated Streamgage using MCM with OK - CIRCULAR	The Most Correlated Streamgage using MCM with IDW	The Most Correlated Streamgage using MCM with OK - SPHERICAL
1302	0.89	0.89	0.89	0.89
1307	0.47	0.60	0.60	0.60
1314	-1.23	0.52	0.52	0.52
1319	0.74	0.82	0.74	0.82
1327	0.50	0.61	0.50	0.61
1334	0.44	0.71	0.44	0.44
1335	0.70	0.80	0.80	0.80
1338	0.29	0.29	0.29	0.29
1339	0.31	0.39	0.31	0.31
1340	-1.44	-1.44	-1.44	-1.44
1343	0.59	0.42	0.75	0.42

As an example, comparison of hydrographs for the streamgage 1307 and 1343 for the year 2003 and 2000 are given in Figure 55 and Figure 56, respectively. It can be seen that the streamflow estimation obtained using the most correlated streamgage is in better agreement than those obtained from the nearest streamgage in Figure 55. However, it should be recognized that MCM does not always guarantee better estimations compared to those obtained from the nearest streamgage used as the reference streamgage (see Figure 56). As can be seen from Table 2, MCM gives improved estimates for most of the streamgages in the study area. Generation of the hydrograph for any year within the observation period at any of the streamgages when it is assumed to be ungaged is relatively easy using the outputs of MCM_GIS.

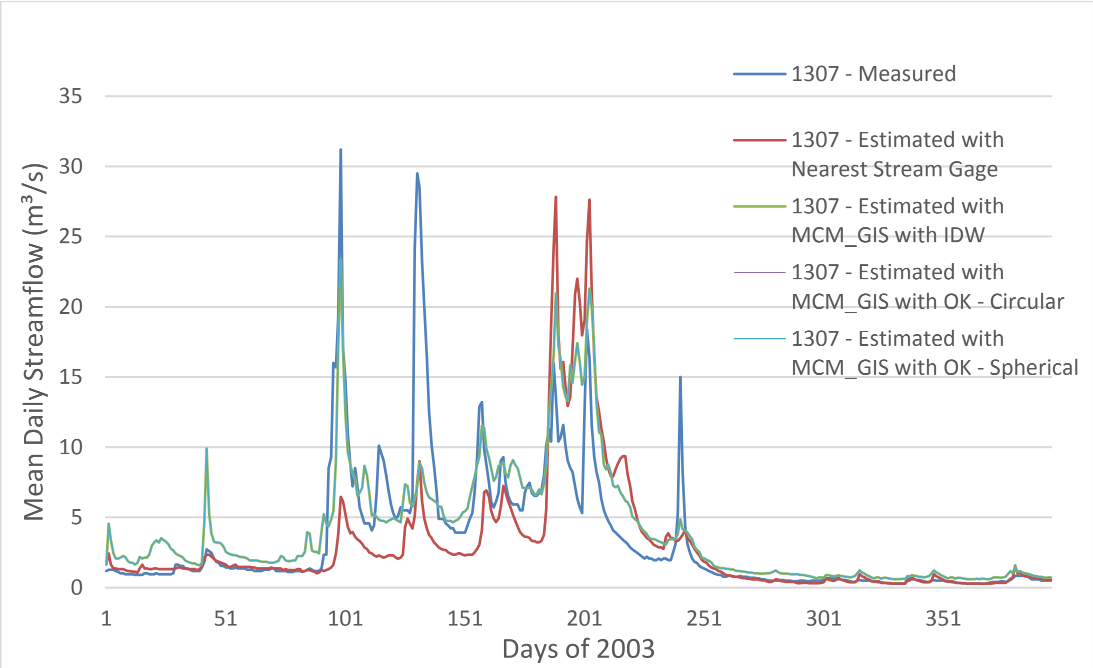


Figure 55: Hydrograph of Streamgage 1307 obtained by MCM_GIS

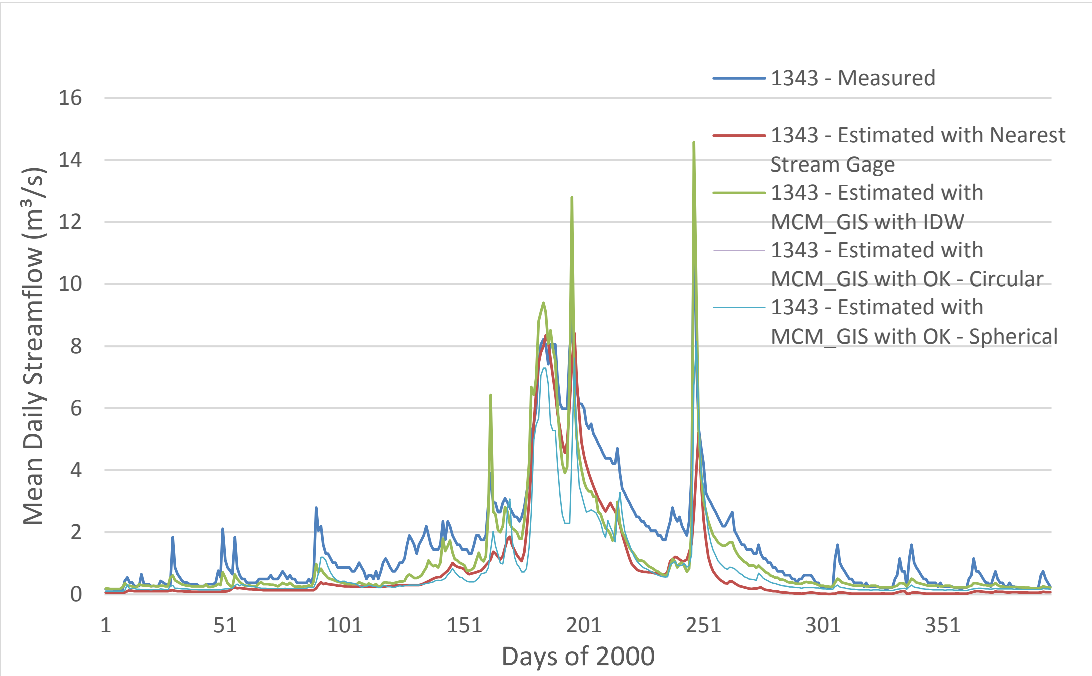


Figure 56: Hydrograph of Streamgage 1343 obtained by MCM_GIS

4.2. Application of MCM_GIS at Coruh Region

4.2.1 Description of the Study Area

A study area that is located in Coruh and Eastern Black Sea Basins (for the sake of simplicity this is called Coruh Region) is selected (Figure 57) as the second case study area. Initially a total of 18 streamgages with 10 years (1993-2003) of common streamflow data are identified and used. Streamgages that are numbered as 2202, 2215, 2218, 2232, 2233, 2304, 2305, 2316, 2320, 2321, 2323, 2325, 2328, 2329, 2330, 2337, 2340 and 2342 are included in the analysis (see Figure 57). The details of these streamgages are given in Table 4.

Table 4: Summary of Coruh Region Streamgages

No	Catchment Location	Drainage area (km^2)	Elevation (m)	Observation period (years)
2202	Kara Dere	635.7	78	1967-2003
2215	Çamlık Dere	445.2	942	1965-2003 (excluding 1991)
2218	İyidere	834.9	308	1954-2009
2232	Fırtına Deresi	763.2	237	1964-2003
2233	Tozköy Deresi	223.1	1296	1964-2003
2304	Çoruh Ana Kol	1734	1545	1962-2003
2305	Çoruh Ana Kol	7272	654	1963-2003
2316	Çoruh Ana Kol	5505.2	1170	1965-2003
2320	Çoruh Ana Kol	4759.2	1365	1971-2003 (excluding 1990, 1991, 1992)
2321	Parhal Deresi	586	705	1972-2003
2323	Oltu Suyu	6854	572	1963-2003
2325	Oltu Suyu	1762	1129	1974-2003 (excluding 1990)
2328	Ardanuç Deresi	546.8	365	1982-2003
2329	Oltu Suyu	3518.5	1004	1982-2003
2330	Çamlıkaya Deresi	113.6	995	1982-2003
2337	Çoruh Ana Kol	6634.2	892	1990-2003
2340	Öğdem Deresi	202	682	1992-2003
2342	Parhal Deresi	318.4	112	1993-2003

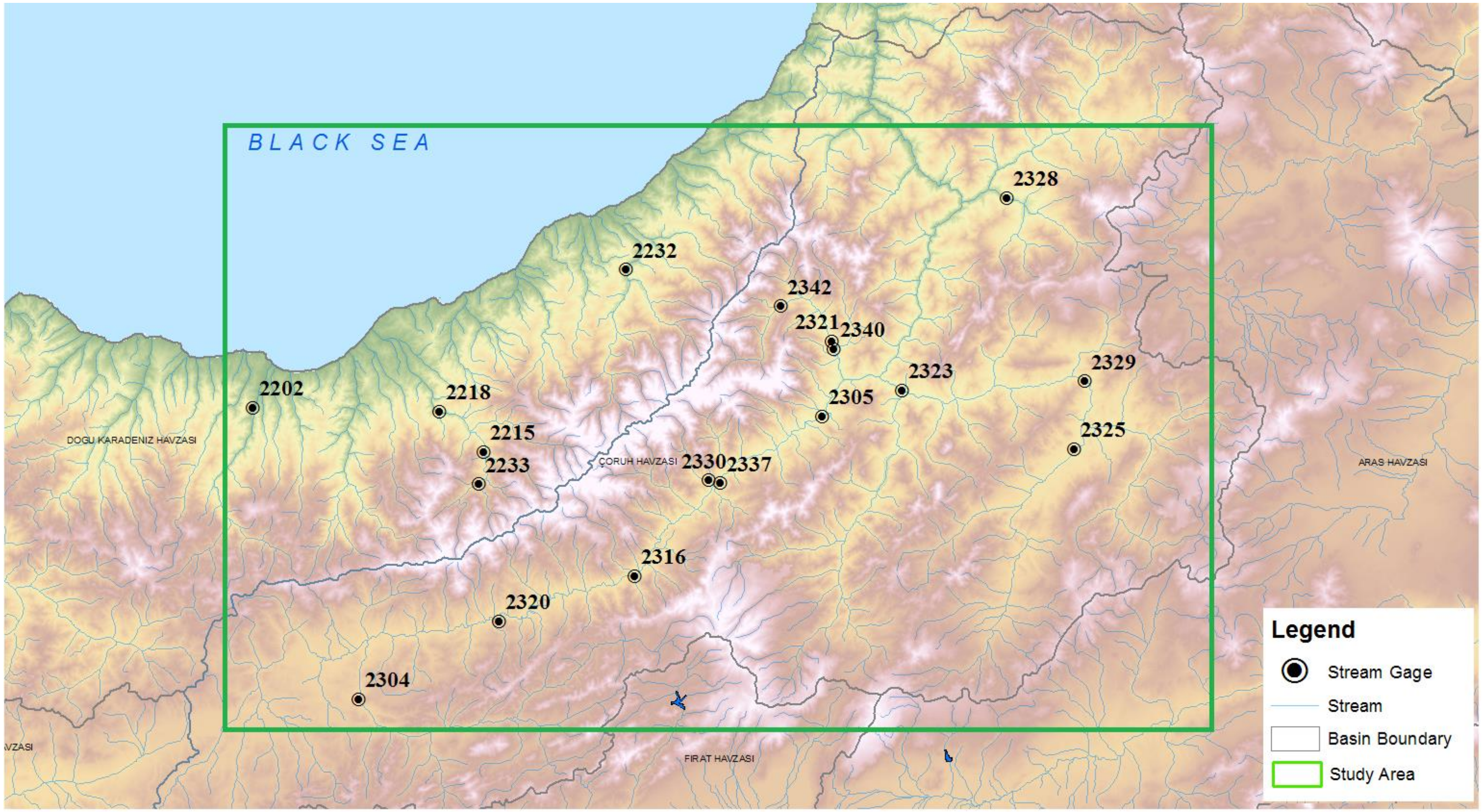


Figure 57: Study Area in Coruh Basin and Eastern Black Sea Basin

4.2.2 Application of MCM by the GIS Tool

First, the correlation matrix for the streamgages found in the Coruh Region are calculated and given in Figure 60. The file structure obtained after using File Generator, xls2txt Tool and Add Label to Text File Tool and is given in Figure 58. Newly created txt files are provided as the inputs to the Import Tool and file structure inside the GIS environment is created (See Figure 59). Similar to the application to the Western Black Sea Region, an overlook to the spatial layers is provided inside the GIS environment (see Figure 61). In this map, the digital elevation model, basin boundaries, dam locations can be seen before moving into the next step of the application.

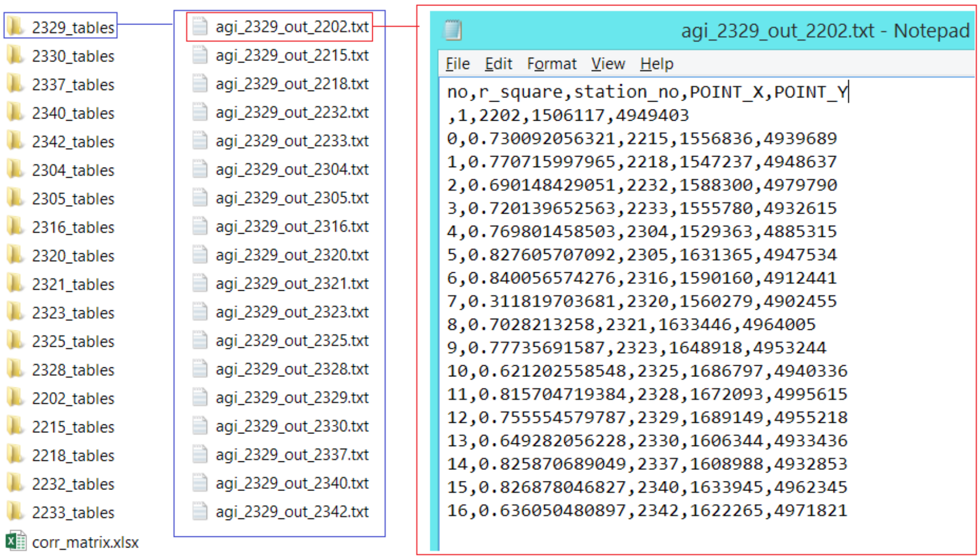


Figure 58: File Structure for Streamgages in Coruh Region

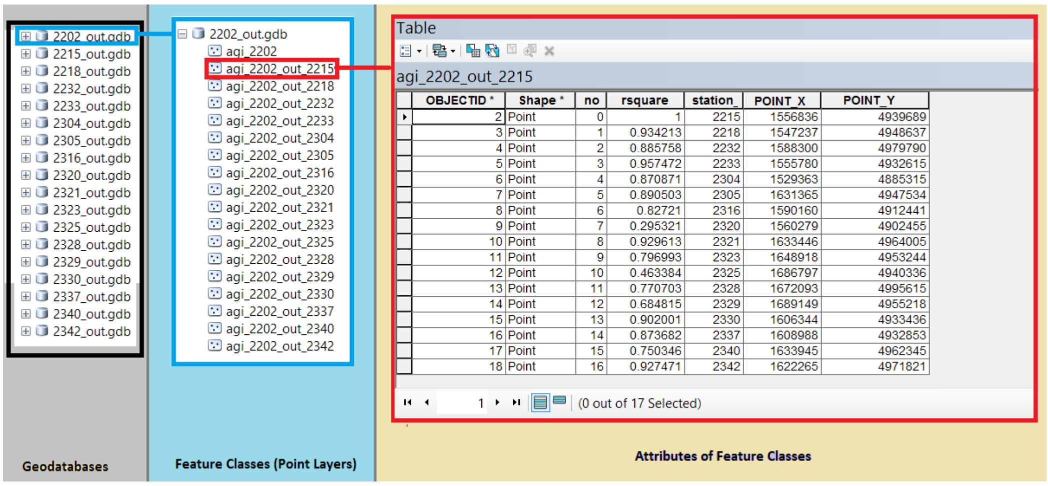


Figure 59: File Structure in GIS for Western Black Sea Basin

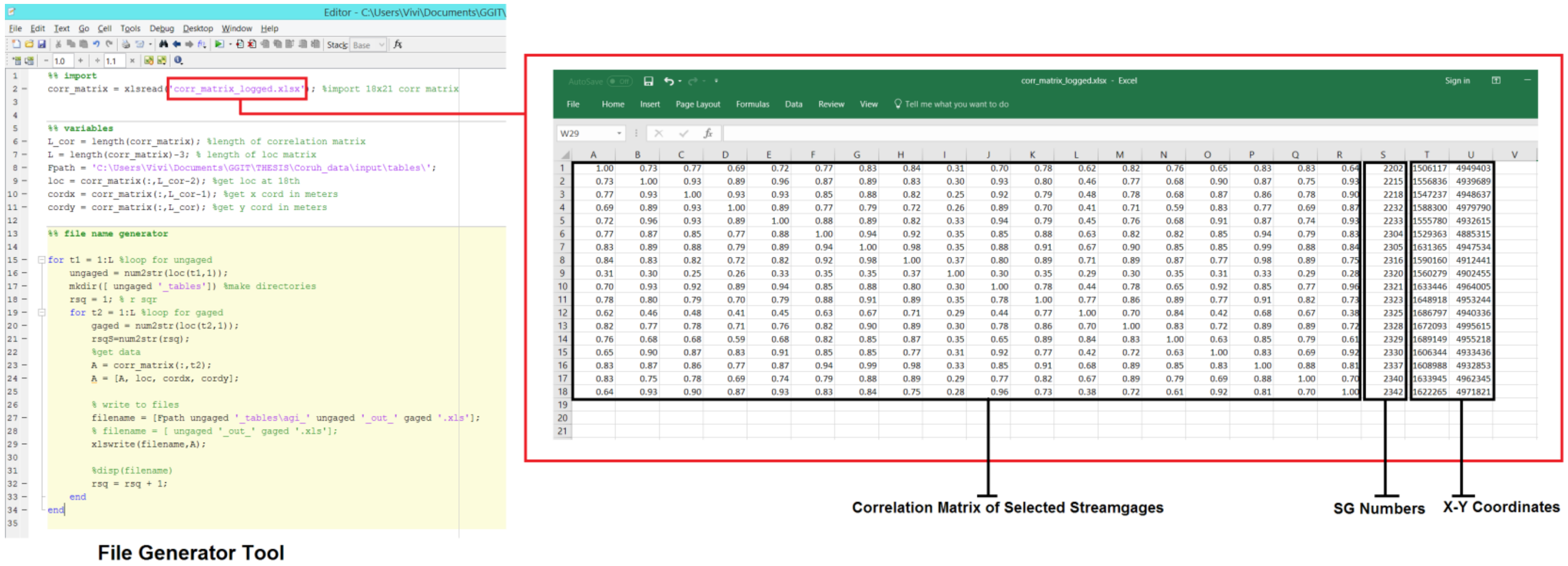


Figure 60: The correlation matrix Excel file for the streamgages found in the Coruh Region

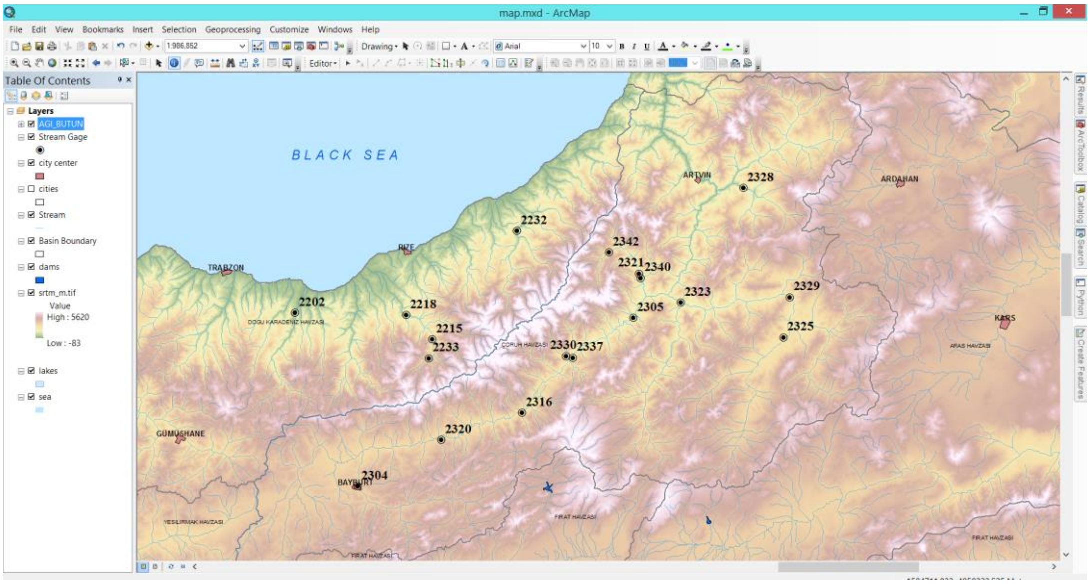


Figure 61: Digital elevation model, streamgages, dams at Coruh Region

Leave-one-out cross-validation is carried out for this study area to evaluate the performance of MCM in the Coruh Basin as well. Estimation of *r* values is done by performing spatial interpolation. Cross-correlation maps created using IDW, where a selected streamgage is assumed to be ungaged are given in Figure 62, Figure 63, Figure 64, Figure 65, Figure 66, Figure 67, Figure 68, Figure 69, Figure 70, Figure 71, Figure 72, Figure 73, Figure 74 and Figure 75 for streamgages 2215, 2218, 2232, 2233, 2305, 2316, 2320, 2321, 2323, 2325, 2330, 2337, 2340 and 2342, respectively.

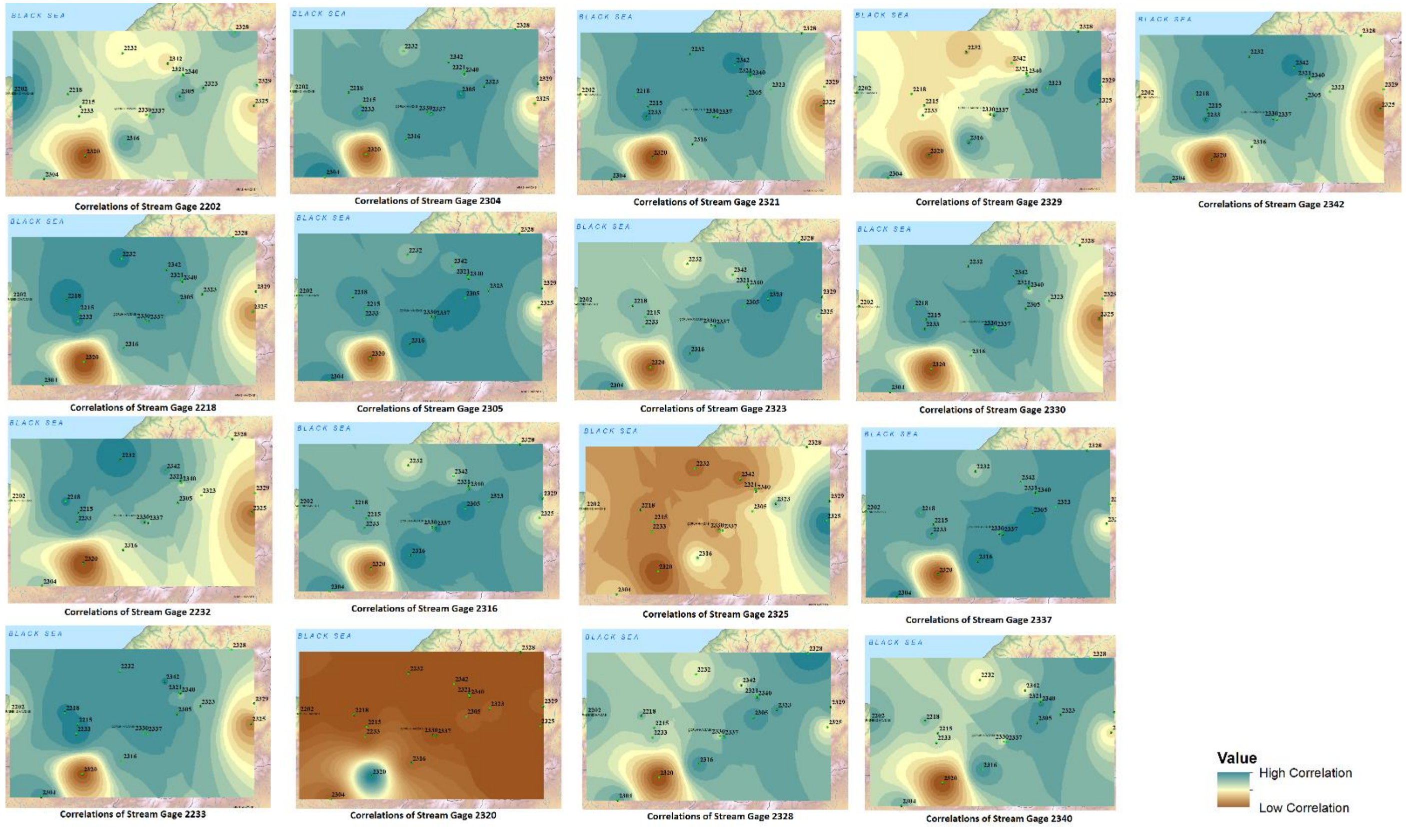


Figure 62: Correlation maps for Coruh Basin where streamgage number 2215 is assumed to be ungaged (Created with IDW)

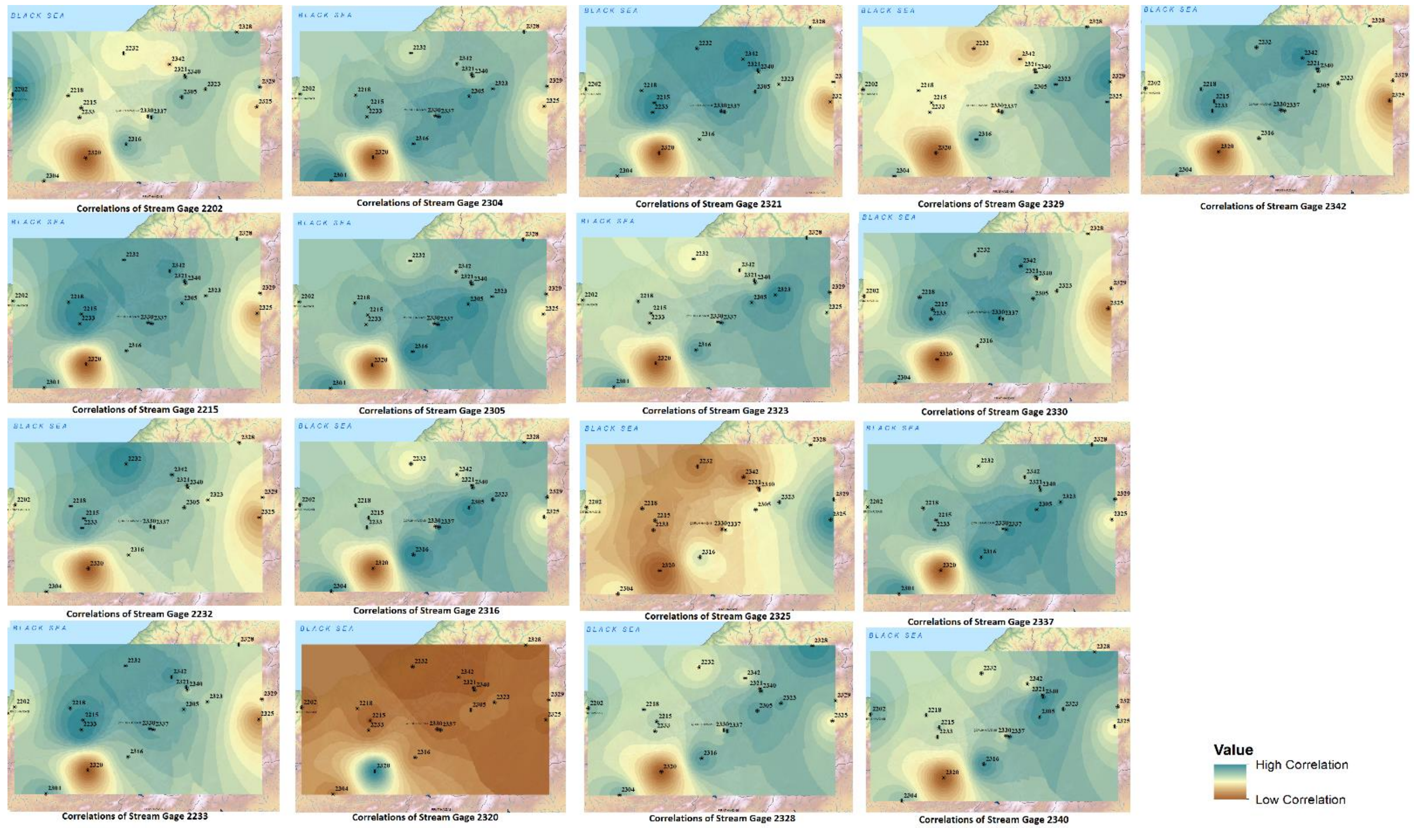


Figure 63: Correlation maps for Coruh Basin where streamgage number 2218 is assumed to be ungaged (Created with IDW)

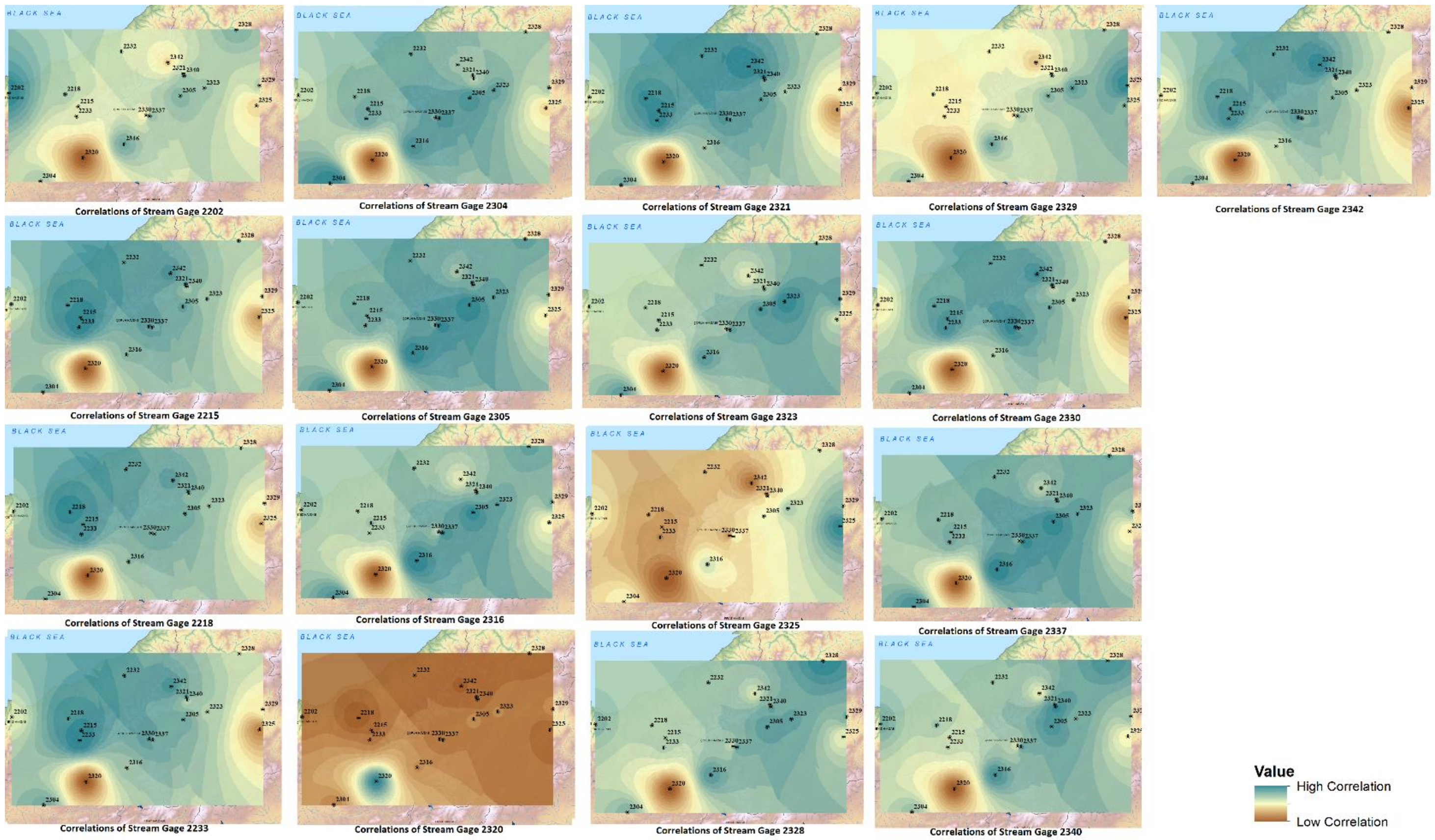


Figure 64: Correlation maps for Coruh Basin where streamgage number 2232 is assumed to be ungaged (Created with IDW)

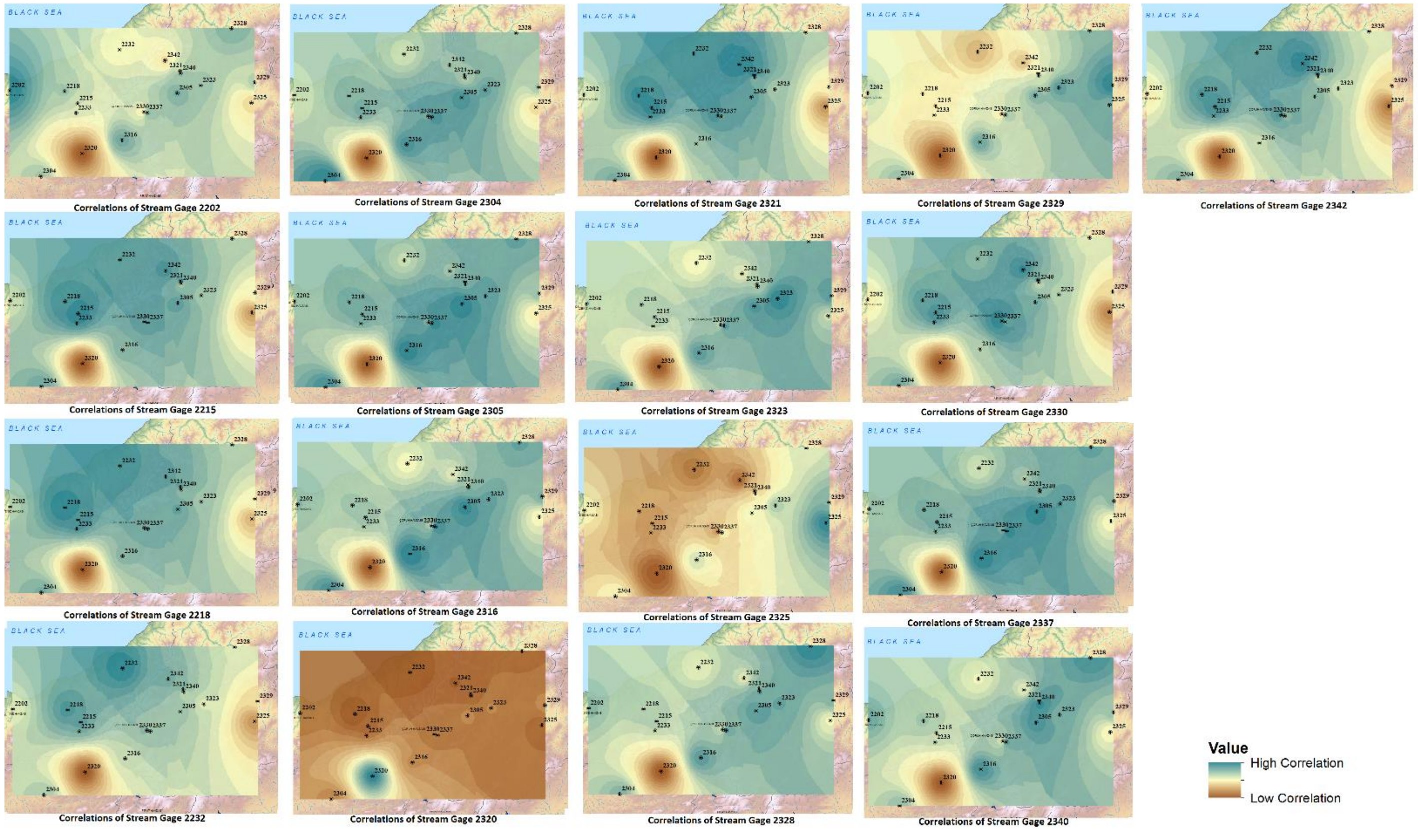


Figure 65: Correlation maps for Coruh Basin where streamgage number 2233 is assumed to be ungaged (Created with IDW)

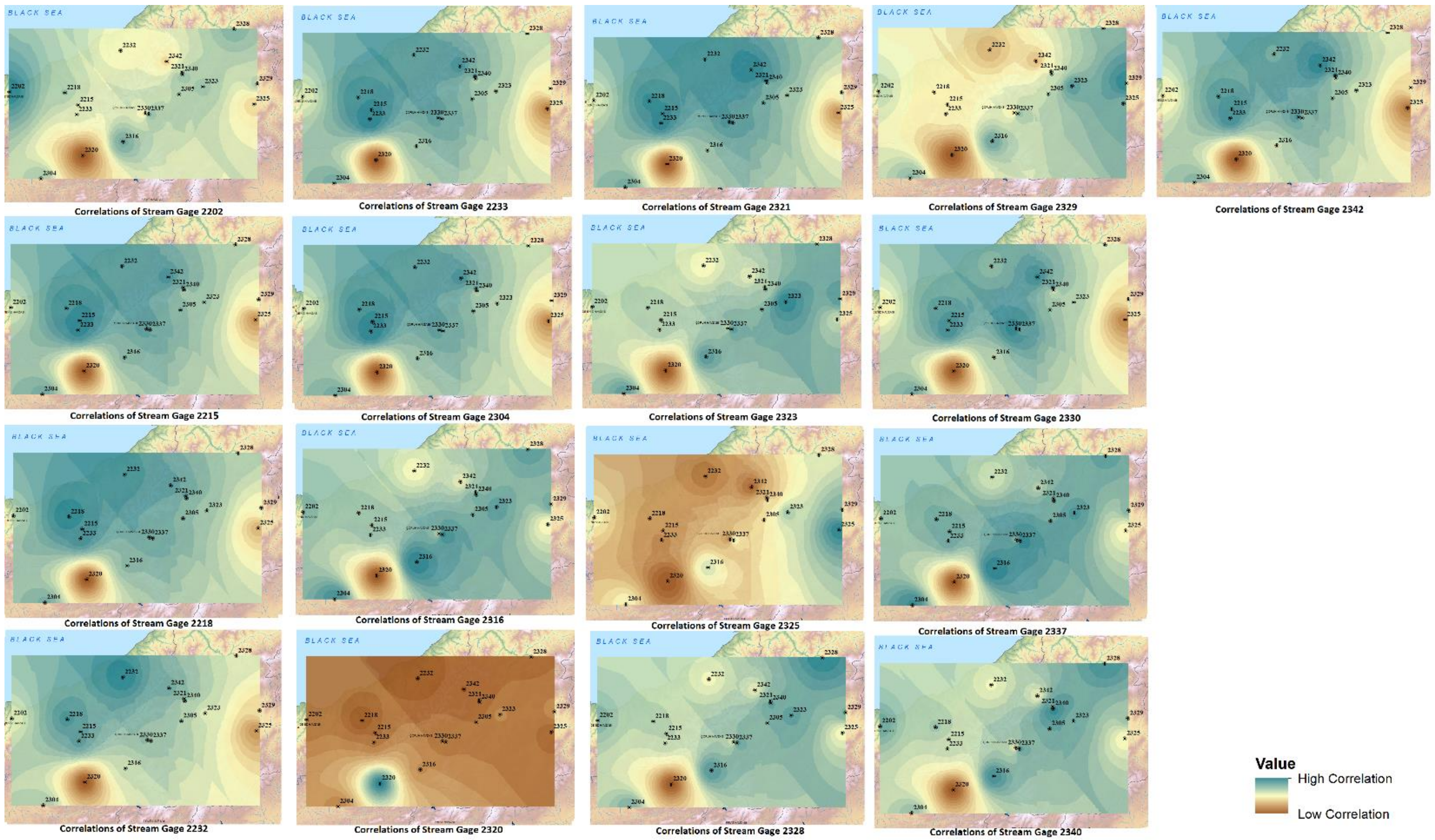


Figure 66: Correlation maps for Coruh Basin where streamgage number 2305 is assumed to be ungaged (Created with IDW)

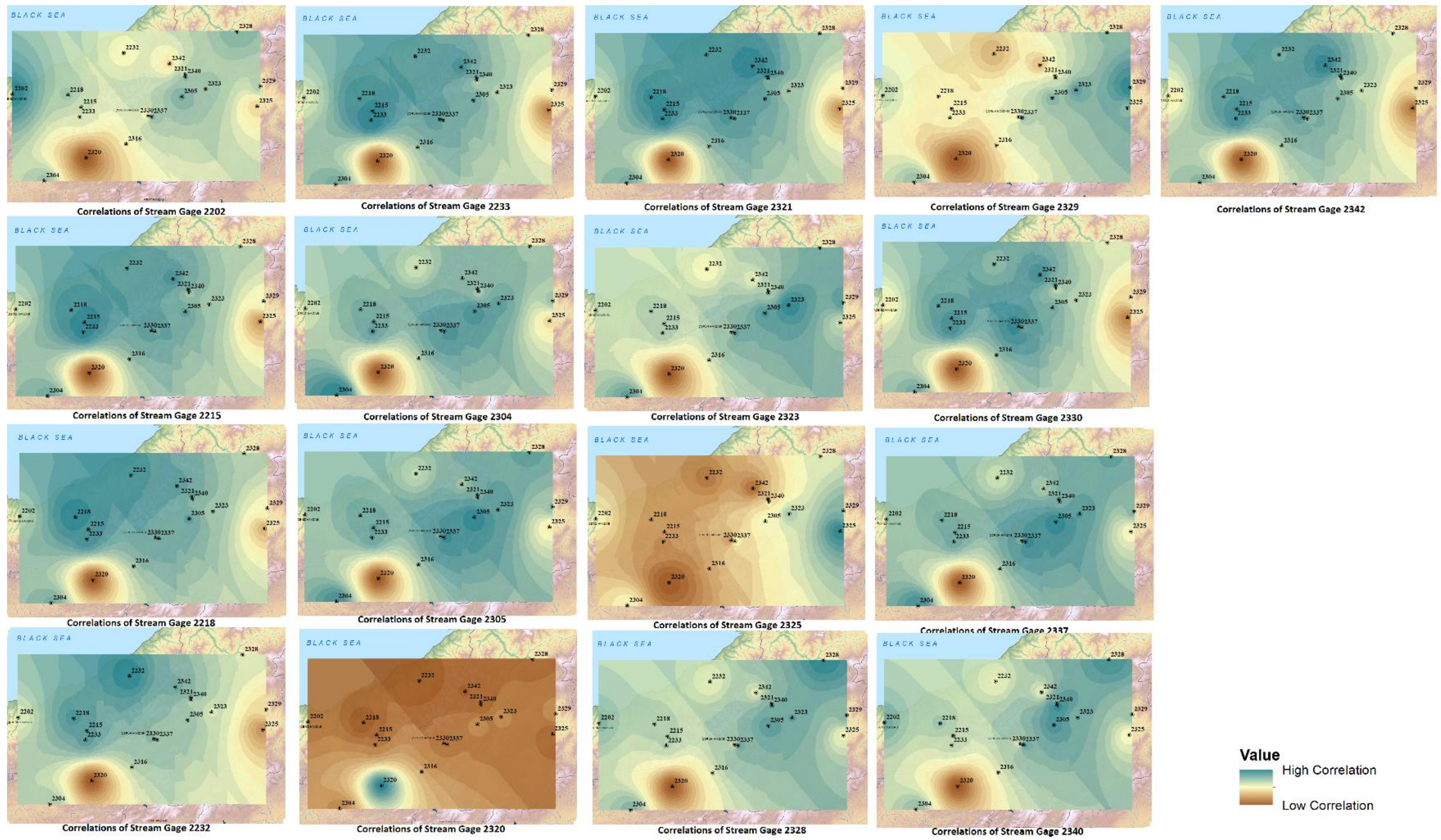


Figure 67: Correlation maps for Coruh Basin where streamgage number 2316 is assumed to be ungaged (Created with IDW)

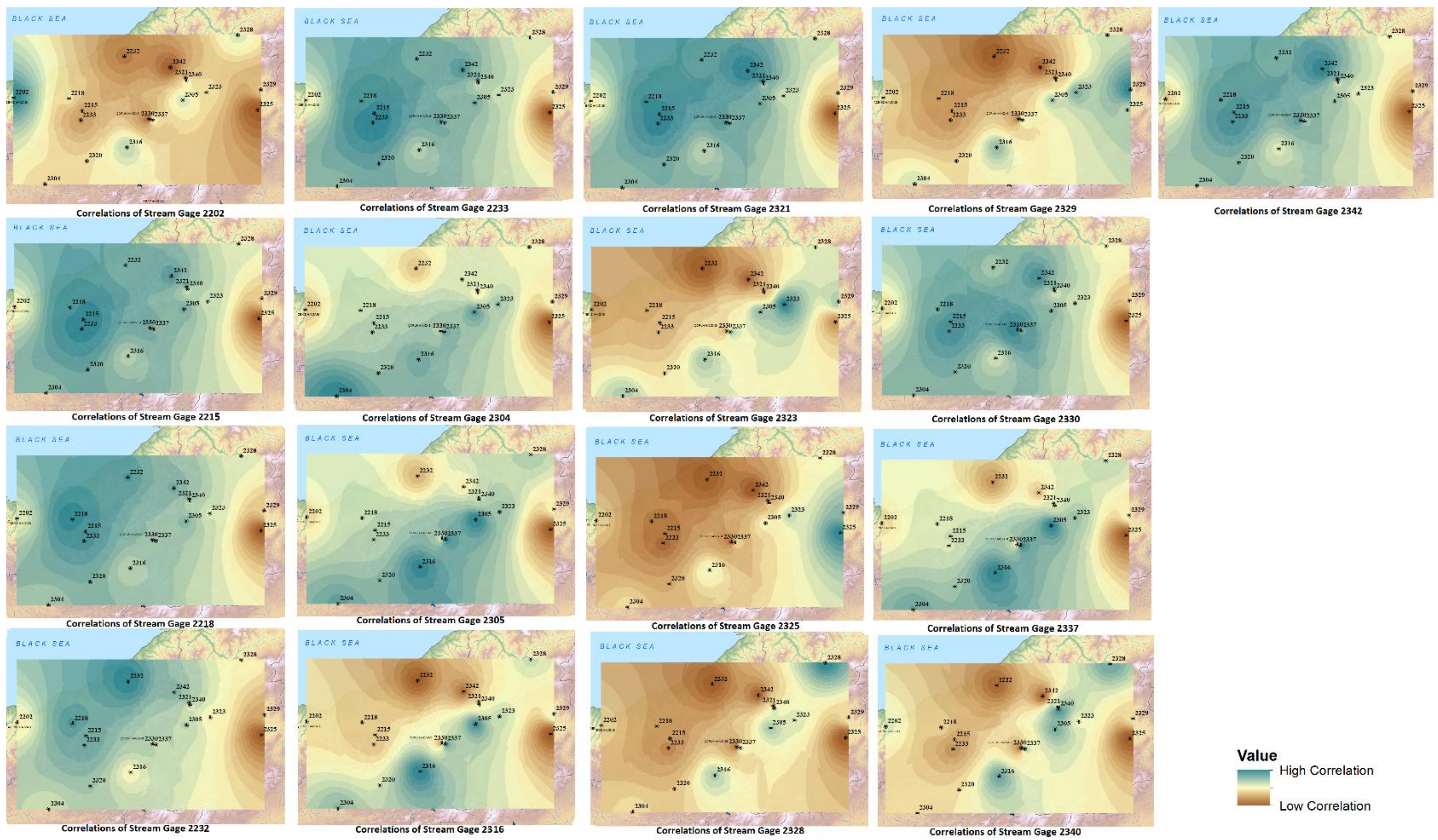


Figure 68: Correlation maps for Coruh Basin where streamgage number 2320 is assumed to be ungaged (Created with IDW)

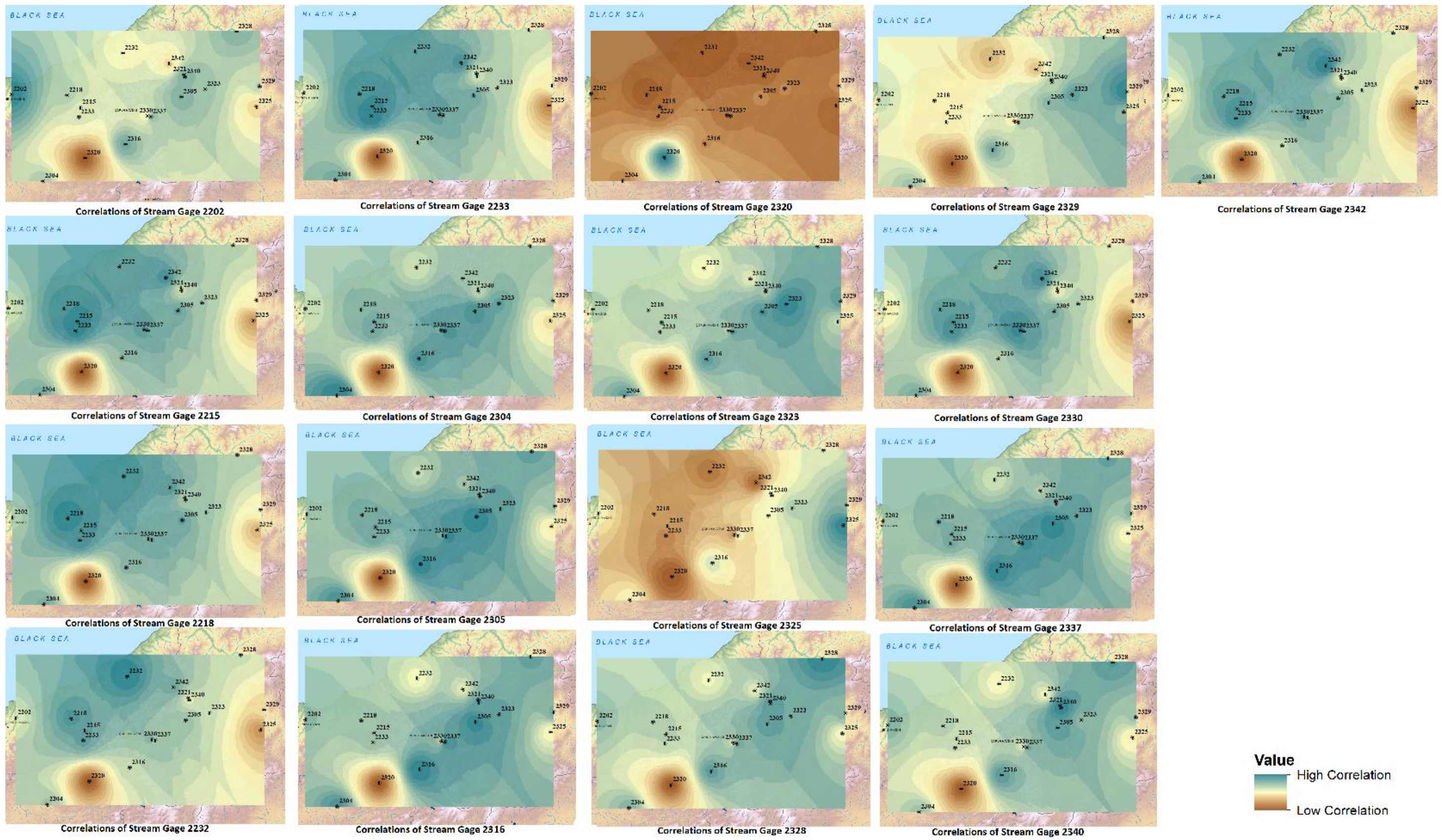


Figure 69: Correlation maps for Coruh Basin where streamgage number 2321 is assumed to be ungaged (Created with IDW)

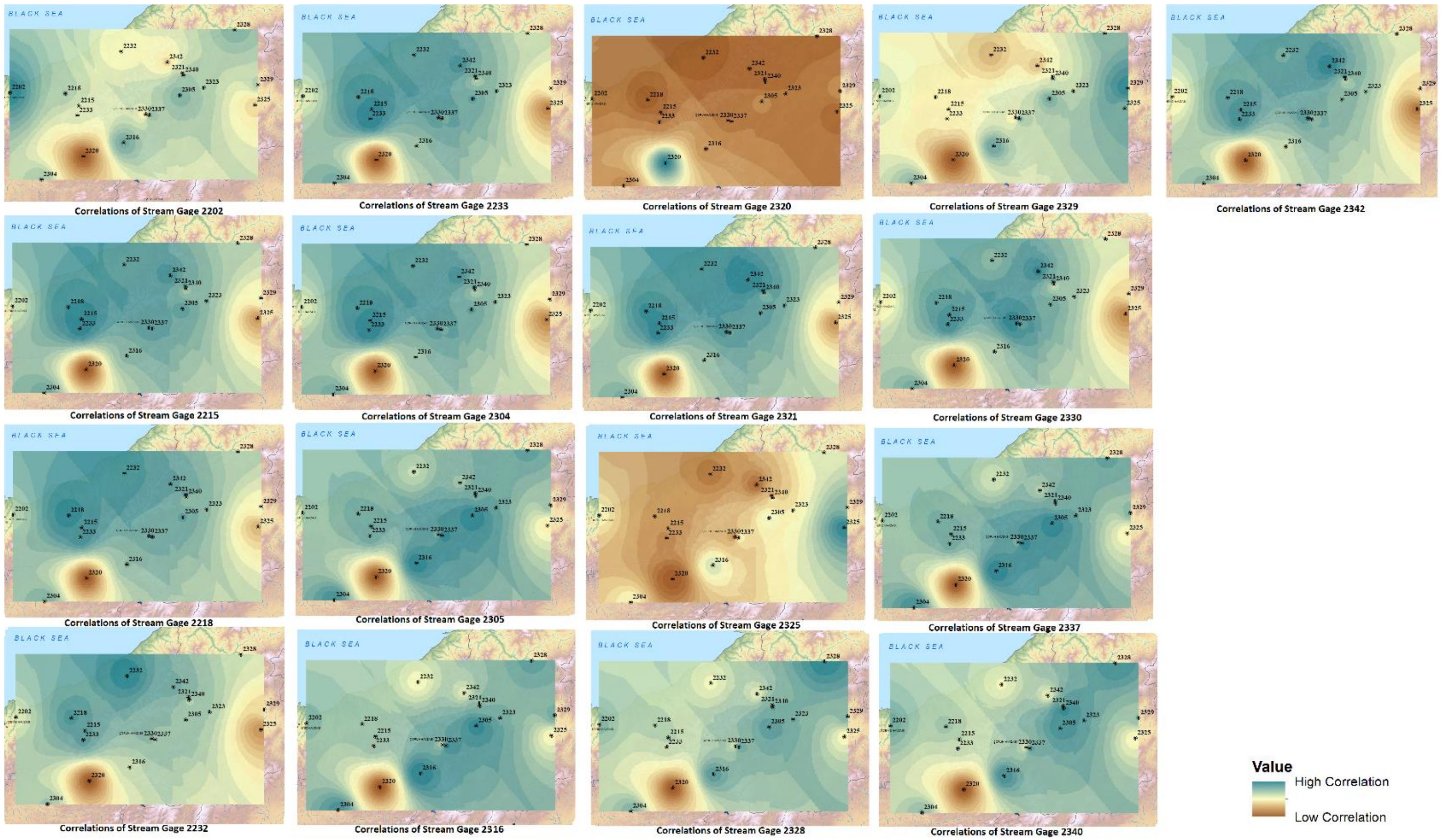


Figure 70: Correlation maps for Coruh Basin where streamgage number 2323 is assumed to be ungaged (Created with IDW)

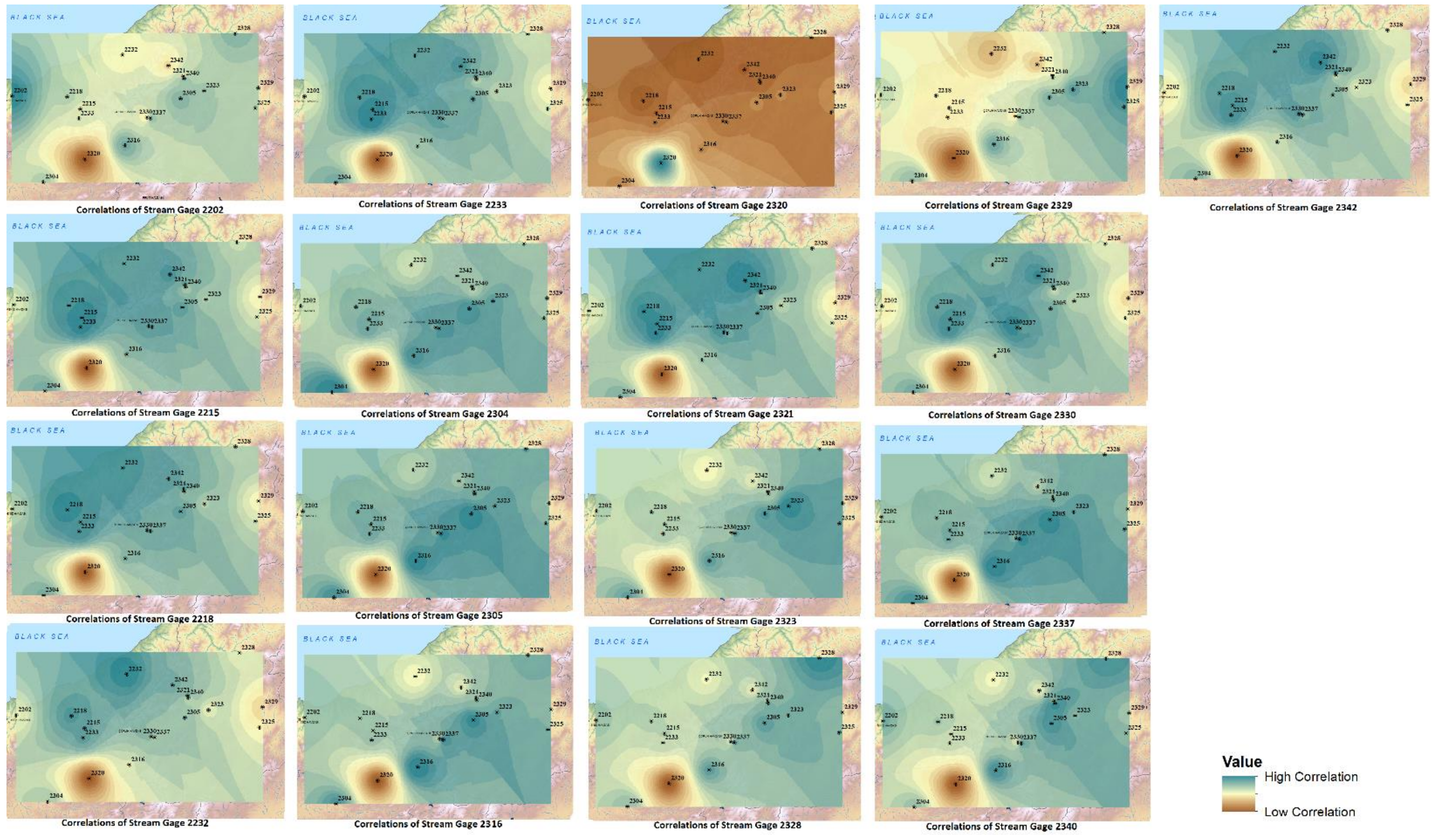


Figure 71: Correlation maps for Coruh Basin where streamgage number 2325 is assumed to be ungaged (Created with IDW)

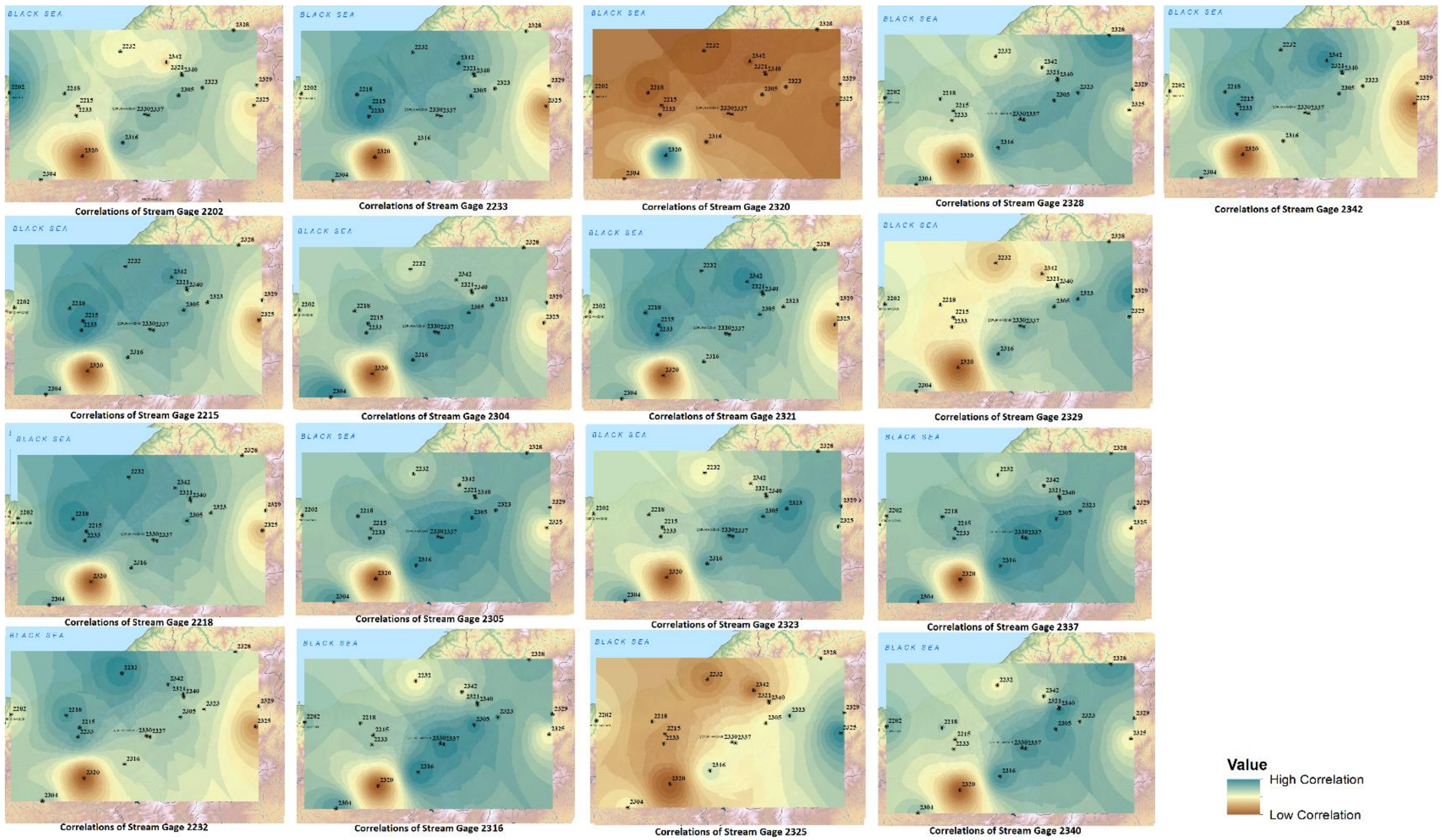


Figure 72: Correlation maps for Coruh Basin where streamgage number 2330 is assumed to be ungaged (Created with IDW)

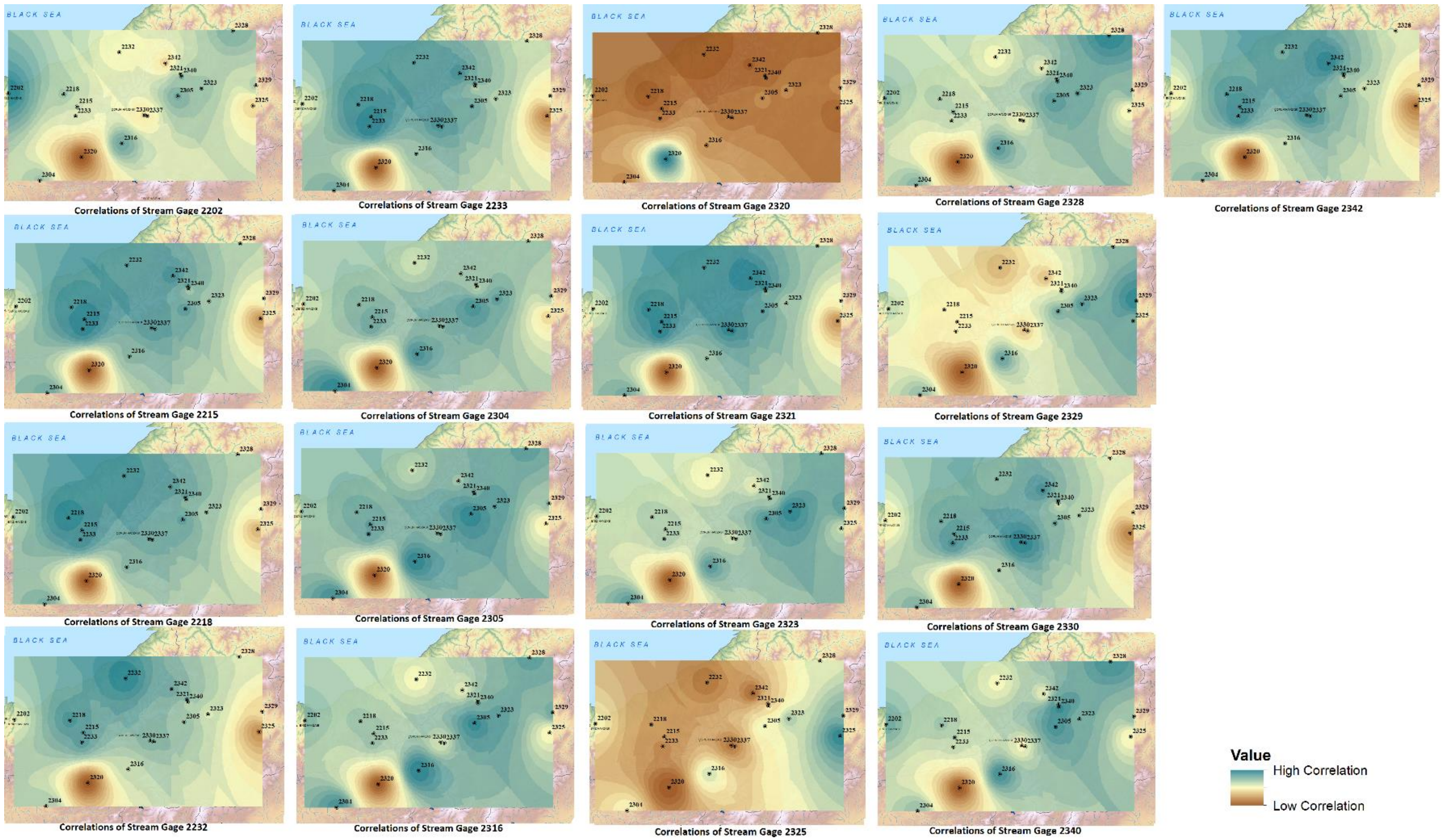


Figure 73: Correlation maps for Coruh Basin where streamgage number 2337 is assumed to be ungaged (Created with IDW).

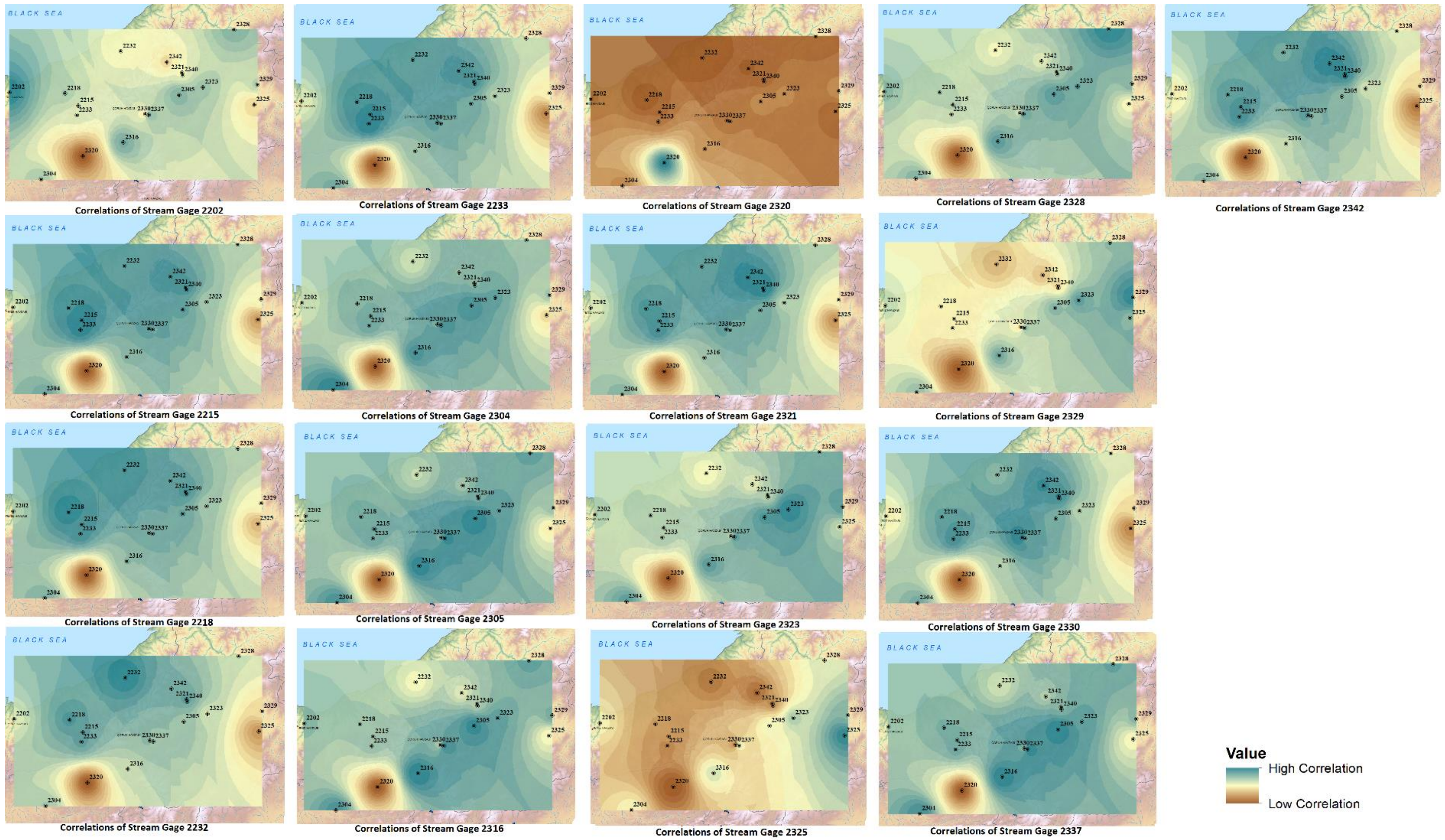


Figure 74 Correlation maps for Coruh Basin where streamgage number 2340 is assumed to be ungaged (Created with IDW).

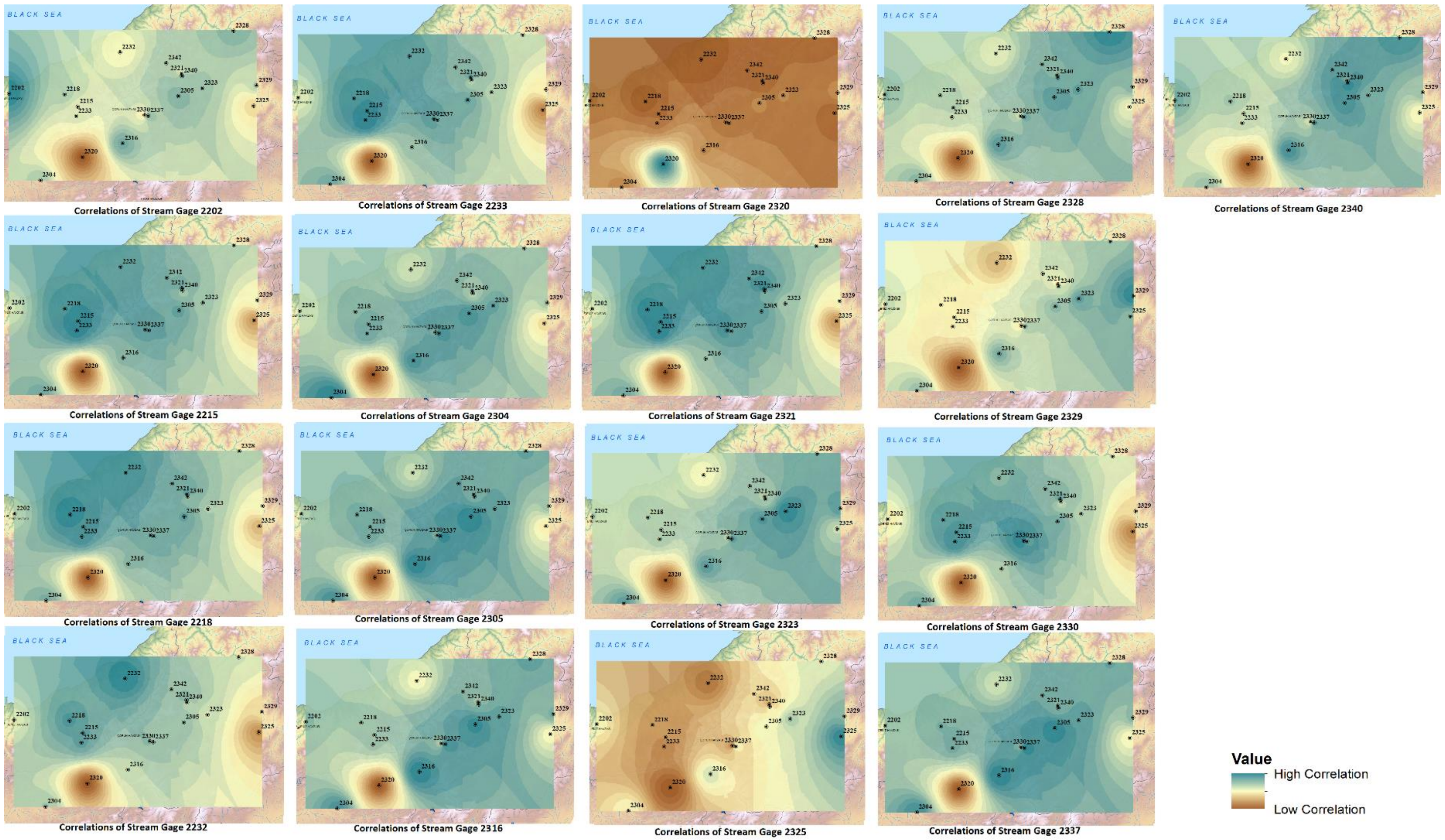


Figure 75: Correlation maps for Coruh Basin where streamgage number 2342 is assumed to be ungaged (Created with IDW)

4.2.3 Evaluation of the Results for Coruh Region

Pearson's r correlation values are read from the cross-correlation maps. Then these estimated values are compared with the observed Pearson's r correlation values. Outcome of the applications of IDW and Ordinary Kriging at Coruh Region are given in Figure 76 and Figure 77, respectively. Generally, the estimated r values are in good agreement for both study areas, except for a few of the streamgages. For example, for streamgage 2320 there is not a good corelation between the observed and predicted correlation values. Low correlation values with the rest of the streamgages as can be seen from Figure 60 as well. One possible reason for this may be the proximity of streamgage 2320 to the study area boundary. As it was mentioned before, locating the ungaged basin near the center of the study area produce more accurate estimations. Other reason might be unknown or unrecorded regulations in the vicinity of this streamgage. NSE values obtained with different types of interpolation methods are given in Table 5. For most of the streamgages, correlation maps created using IDW produced higher NSE values. Three of the streamgages (i.e. 2305, 2316 and 2323) produced extremely low NSE values when the nearest streamgage is selected as the reference. The reason for this is the assumed to be ungaged basin has relatively large drainage area and is located in the Coruh River while the reference streamgage is located in a very small different basin. As can be seen in Table 5, out of 18 streamgages, results are obtained for 14 of them since IDW and OK were not able to carry out spatial interpolation for remaining four streamgages (i.e. 2202, 2304, 2328 and 2329) due to these particular streamgages being located near the boundary of the study area.

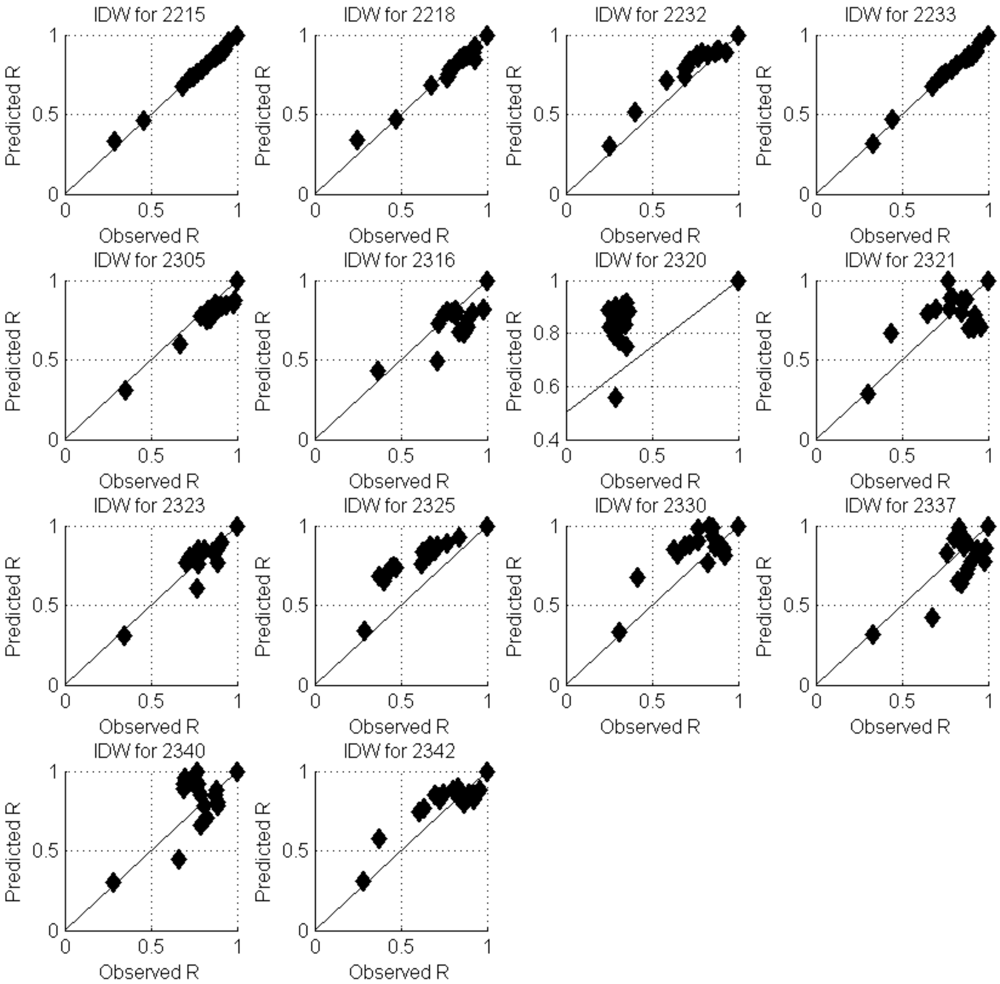


Figure 76: Evaluation of Correlation Estimations with IDW for Coruh Region

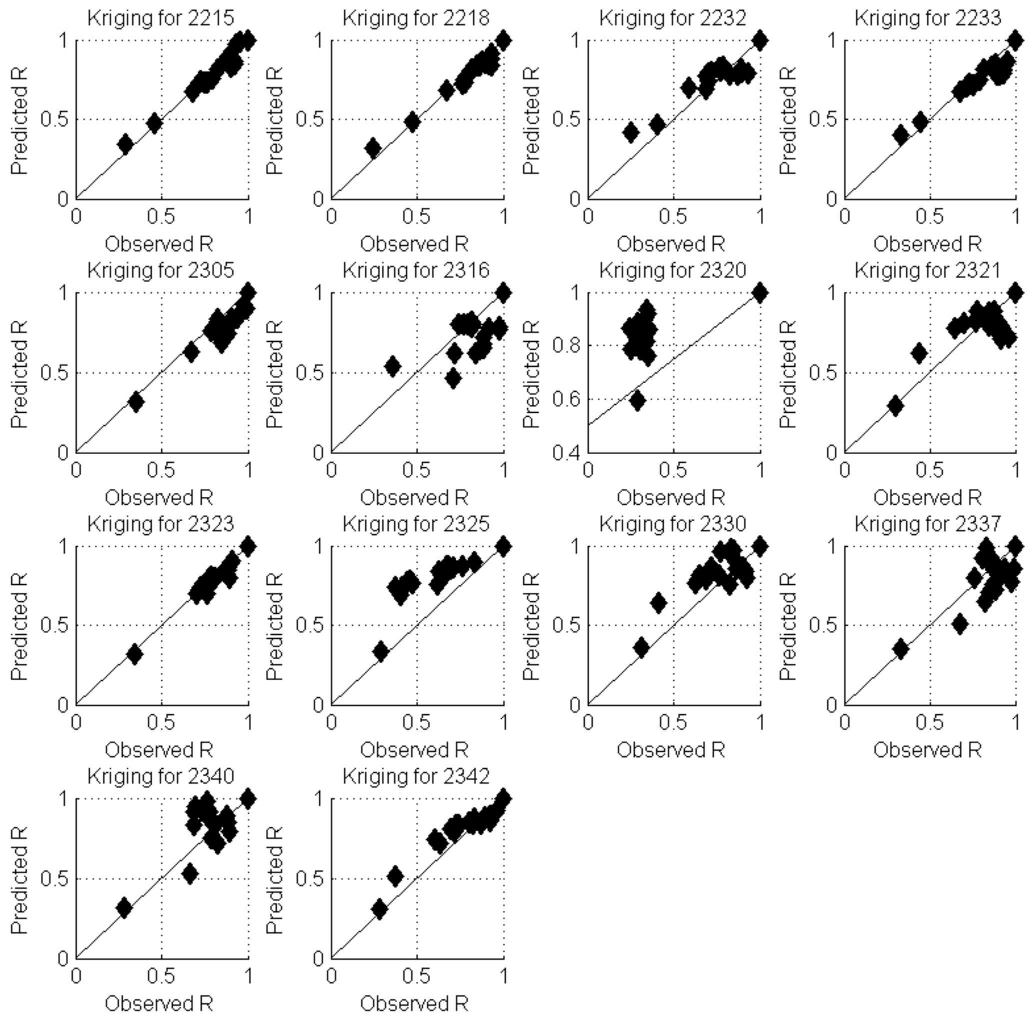


Figure 77: Evaluation of Correlation Estimations with Ordinary Ordinary Kriging for Coruh Region

IDW performed better for the application on Coruh Region. One possible reason for this might be that Ordinary Kriging correlation maps were influenced by the outliers in the data. The output of correlation estimations are shown in Figure 78 and Figure 79 for OK and IDW, respectively. Most correlated streamgages are shown with bold characters.

Ungaged	2215		2218		2232		2233		2305		2316		2320	
	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation
2202	0,73	0,74	0,77	0,73	0,69	0,69	0,72	0,70	0,83	0,83	0,84	0,62	0,31	0,78
2215	1,00	1,00	0,93	**0,91**	0,89	0,80	**0,96**	**0,87**	0,89	0,80	0,83	0,80	0,30	0,88
2218	0,93	0,97	1,00	1,00	**0,93**	0,79	0,93	0,82	0,88	0,77	0,82	0,78	0,25	0,86
2232	0,89	0,87	0,93	0,84	1,00	1,00	0,89	0,78	0,79	0,76	0,72	0,62	0,26	0,79
2233	**0,96**	**0,99**	**0,93**	0,88	0,89	0,81	1,00	1,00	0,89	0,74	0,82	**0,81**	0,33	0,90
2304	0,87	0,87	0,85	0,84	0,77	0,80	0,88	0,83	0,94	0,84	0,92	0,78	0,35	**0,93**
2305	0,89	0,89	0,88	0,86	0,79	**0,83**	0,89	0,85	1,00	1,00	0,98	0,78	0,35	0,92
2316	0,83	0,82	0,82	0,81	0,72	0,80	0,82	0,81	0,98	**0,92**	1,00	1,00	**0,37**	0,86
2320	0,30	0,34	0,25	0,32	0,26	0,42	0,33	0,40	0,35	0,31	0,37	0,54	1,00	1,00
2321	0,93	0,86	0,92	0,85	0,89	0,80	0,94	0,81	0,88	0,74	0,80	0,80	0,30	0,86
2323	0,80	0,76	0,79	0,76	0,70	0,79	0,79	0,74	0,91	0,83	0,89	0,72	0,35	0,81
2325	0,46	0,47	0,48	0,48	0,41	0,47	0,45	0,48	0,67	0,63	0,71	0,46	0,29	0,59
2328	0,77	0,73	0,78	0,75	0,71	0,74	0,76	0,71	0,90	0,81	0,89	0,65	0,30	0,81
2329	0,68	0,67	0,68	0,68	0,59	0,70	0,68	0,67	0,85	0,77	0,87	0,65	0,35	0,76
2330	0,90	0,83	0,87	0,84	0,83	0,78	0,91	0,78	0,85	0,68	0,77	0,80	0,31	0,85
2337	0,87	0,87	0,86	0,85	0,77	0,82	0,87	0,83	**0,99**	0,90	**0,98**	0,77	0,33	0,90
2340	0,75	0,73	0,78	0,73	0,69	0,77	0,74	0,72	0,88	0,79	0,89	0,68	0,29	0,80
2342	0,93	0,85	0,90	0,85	0,87	0,78	0,93	0,79	0,84	0,74	0,75	0,80	0,28	0,83

Ungaged	2321		2323		2325		2330		2337		2340		2342	
	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation
2202	0,70	0,81	0,78	0,75	0,62	0,75	0,65	0,82	0,83	0,65	0,83	0,71	0,64	0,71
2215	0,93	0,76	0,80	0,80	0,46	0,78	0,90	0,86	0,87	0,90	0,75	0,92	0,93	0,89
2218	0,92	0,79	0,79	0,80	0,48	0,77	0,87	0,85	0,86	0,87	0,78	0,91	0,90	0,88
2232	0,89	0,75	0,70	0,70	0,41	0,69	0,83	0,76	0,77	0,80	0,69	0,83	0,87	0,84
2233	0,94	0,75	0,79	0,78	0,45	0,78	0,91	0,86	0,87	0,91	0,74	0,93	0,93	0,89
2304	0,85	0,79	0,88	0,84	0,63	0,84	0,85	0,92	0,94	0,85	0,79	0,86	0,83	0,84
2305	0,88	**0,88**	**0,91**	**0,90**	0,67	0,87	0,85	0,97	**0,99**	0,85	0,88	0,89	0,84	0,87
2316	0,80	0,86	0,89	0,85	0,71	0,86	0,77	0,96	0,98	0,78	0,89	0,85	0,75	0,83
2320	0,30	0,29	0,35	0,32	0,29	0,34	0,31	0,36	0,33	0,35	0,29	0,31	0,28	0,31
2321	1,00	1,00	0,78	0,79	0,44	0,76	0,92	0,84	0,85	0,92	0,77	**0,98**	**0,96**	**0,92**
2323	0,78	0,84	1,00	1,00	0,77	0,87	0,77	0,82	0,91	0,79	0,82	0,83	0,73	0,83
2325	0,44	0,62	0,77	0,70	1,00	1,00	0,42	0,64	0,68	0,51	0,67	0,53	0,38	0,51
2328	0,78	0,88	0,86	0,84	0,70	0,85	0,72	0,87	0,89	0,72	**0,89**	0,79	0,72	0,79
2329	0,65	0,78	0,89	0,80	**0,84**	**0,90**	0,63	0,77	0,85	0,71	0,79	0,75	0,61	0,74
2330	0,92	0,71	0,77	0,75	0,42	0,74	1,00	1,00	0,83	**0,99**	0,69	0,91	0,92	0,86
2337	0,85	0,87	0,91	0,89	0,68	0,87	0,83	**0,98**	1,00	1,00	0,88	0,87	0,81	0,85
2340	0,77	0,81	0,82	0,80	0,67	0,82	0,69	0,79	0,88	0,76	1,00	1,00	0,70	0,81
2342	**0,96**	0,72	0,73	0,74	0,38	0,74	**0,92**	0,80	0,81	0,92	0,70	0,94	1,00	1,00

Figure 78: Output correlation estimation table for MCM_GIS application with OK, circular model

Ungaged	2215		2218		2232		2233		2305		2316		2320	
	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation
2202	0.73	0.73	0.77	0.73	0.69	0.73	0.72	0.73	0.83	0.75	0.84	0.67	0.31	0.77
2215	1.00	1.00	0.93	**0.93**	0.89	0.90	**0.96**	**0.95**	0.89	0.83	0.83	0.81	0.30	0.90
2218	0.93	0.92	1.00	1.00	**0.93**	0.89	0.93	0.91	0.88	0.83	0.82	0.78	0.25	0.89
2232	0.89	0.88	0.93	0.85	1.00	1.00	0.89	0.86	0.79	0.77	0.72	0.73	0.26	0.82
2233	**0.96**	**0.96**	**0.93**	0.92	0.89	0.90	1.00	1.00	0.89	0.83	0.82	0.82	0.33	0.90
2304	0.87	0.86	0.85	0.85	0.77	0.86	0.88	0.85	0.94	0.84	0.92	0.80	0.35	0.90
2305	0.89	0.87	0.88	0.87	0.79	0.89	0.89	0.87	1.00	1.00	0.98	**0.82**	0.35	**0.92**
2316	0.83	0.81	0.82	0.81	0.72	0.84	0.82	0.81	0.98	0.85	1.00	1.00	**0.37**	0.88
2320	0.30	0.33	0.25	0.34	0.26	0.30	0.33	0.32	0.35	0.31	0.37	0.43	1.00	1.00
2321	0.93	0.91	0.92	0.89	0.89	**0.91**	0.94	0.90	0.88	0.85	0.80	0.80	0.30	0.88
2323	0.80	0.78	0.79	0.78	0.70	0.81	0.79	0.78	0.91	0.84	0.89	0.75	0.35	0.83
2325	0.46	0.46	0.48	0.47	0.41	0.52	0.45	0.47	0.67	0.60	0.71	0.49	0.29	0.56
2328	0.77	0.76	0.78	0.75	0.71	0.80	0.76	0.76	0.90	0.83	0.89	0.72	0.30	0.81
2329	0.68	0.68	0.68	0.68	0.59	0.71	0.68	0.68	0.85	0.76	0.87	0.67	0.35	0.75
2330	0.90	0.88	0.87	0.86	0.83	0.88	0.91	0.87	0.85	0.81	0.77	0.80	0.31	0.86
2337	0.87	0.86	0.86	0.85	0.77	0.88	0.87	0.85	**0.99**	**0.87**	**0.98**	0.81	0.33	0.91
2340	0.75	0.74	0.78	0.74	0.69	0.79	0.74	0.74	0.88	0.84	0.89	0.71	0.29	0.79
2342	0.93	0.90	0.90	0.88	0.87	0.89	0.93	0.89	0.84	0.81	0.75	0.78	0.28	0.85

Ungaged	2321		2323		2325		2330		2337		2340		2342	
	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation	Measured Correlation	Estimated Correlation
2202	0.70	0.82	0.78	0.76	0.62	0.76	0.65	0.82	0.83	0.66	0.83	0.70	0.64	0.76
2215	0.93	0.75	0.80	0.83	0.46	0.74	0.90	0.87	0.87	0.90	0.75	0.93	0.93	0.85
2218	0.92	0.78	0.79	0.83	0.48	0.73	0.87	0.86	0.86	0.87	0.78	0.92	0.90	0.86
2232	0.89	0.70	0.70	0.77	0.41	0.65	0.83	0.77	0.77	0.83	0.69	0.89	0.87	0.80
2233	0.94	0.74	0.79	0.83	0.45	0.74	0.91	0.87	0.87	0.91	0.74	0.93	0.93	0.85
2304	0.85	0.79	0.88	0.84	0.63	0.84	0.85	0.94	0.94	0.85	0.79	0.85	0.83	0.84
2305	0.88	0.89	**0.91**	**0.90**	0.67	0.87	0.85	0.99	**0.99**	0.86	0.88	0.88	0.84	**0.89**
2316	0.80	0.89	0.89	0.87	0.71	0.87	0.77	0.98	0.98	0.78	0.89	0.80	0.75	0.86
2320	0.30	0.29	0.35	0.31	0.29	0.34	0.31	0.33	0.33	0.32	0.29	0.30	0.28	0.31
2321	1.00	1.00	0.78	0.85	0.44	0.72	0.92	0.85	0.85	0.92	0.77	**0.99**	**0.96**	0.88
2323	0.78	0.82	1.00	1.00	0.77	0.89	0.77	0.90	0.91	0.77	0.82	0.78	0.73	0.82
2325	0.44	0.66	0.77	0.61	1.00	1.00	0.42	0.68	0.68	0.43	0.67	0.45	0.38	0.57
2328	0.78	0.89	0.86	0.84	0.70	0.84	0.72	0.88	0.89	0.73	**0.89**	0.78	0.72	0.84
2329	0.65	0.79	0.89	0.77	**0.84**	**0.93**	0.63	0.85	0.85	0.64	0.79	0.66	0.61	0.74
2330	0.92	0.70	0.77	0.80	0.42	0.70	1.00	1.00	0.83	**0.99**	0.69	0.92	0.92	0.82
2337	0.85	0.88	0.91	0.88	0.68	0.87	0.83	**1.00**	1.00	1.00	0.88	0.85	0.81	0.88
2340	0.77	**0.99**	0.82	0.85	0.67	0.81	0.69	0.87	0.88	0.70	1.00	1.00	0.70	0.85
2342	**0.96**	0.70	0.73	0.81	0.38	0.68	**0.92**	0.81	0.81	0.92	0.70	0.96	1.00	1.00

Figure 79: Output correlation estimation table for MCM_GIS application with IDW

Table 5: NSE values for Coruh Region with different types of interpolation methods

Ungaged	Nearest Streamgage	The Most Correlated Streamgage	The Most Correlated Streamgage using MCM with IDW	The Most Correlated Streamgage using MCM with OK - CIRCULAR
2215	0.85	0.85	0.85	0.85
2218	0.86	0.70	0.86	0.86
2232	0.64	0.84	0.60	-0.51
2233	0.90	0.90	0.90	0.90
2305	-0.28	0.98	0.98	0.85
2316	-9.87	0.88	0.76	0.68
2320	0.96	0.96	0.59	0.59
2321	0.33	0.82	0.33	0.17
2323	-10.80	-1.31	-1.31	-1.31
2325	0.76	0.76	0.76	0.76

Table 5	continued.			
2330	0.10	0.65	0.10	0.10
2337	-5.52	0.97	-5.52	-5.52
2340	-0.20	0.72	-0.20	-0.20
2342	0.88	0.88	0.03	0.88

As an example, comparison of hydrographs for the streamgage 2316 and 2305 for the year 1998 and 2000 are given in Figure 80 and Figure 81 respectively. It can be seen that the streamflow estimation obtained using the most correlated streamgage is in better agreement than those obtained from the nearest streamgage. Hydrograph for any year within the observation period at any streamgage can be generated easily using the results of MCM_GIS.

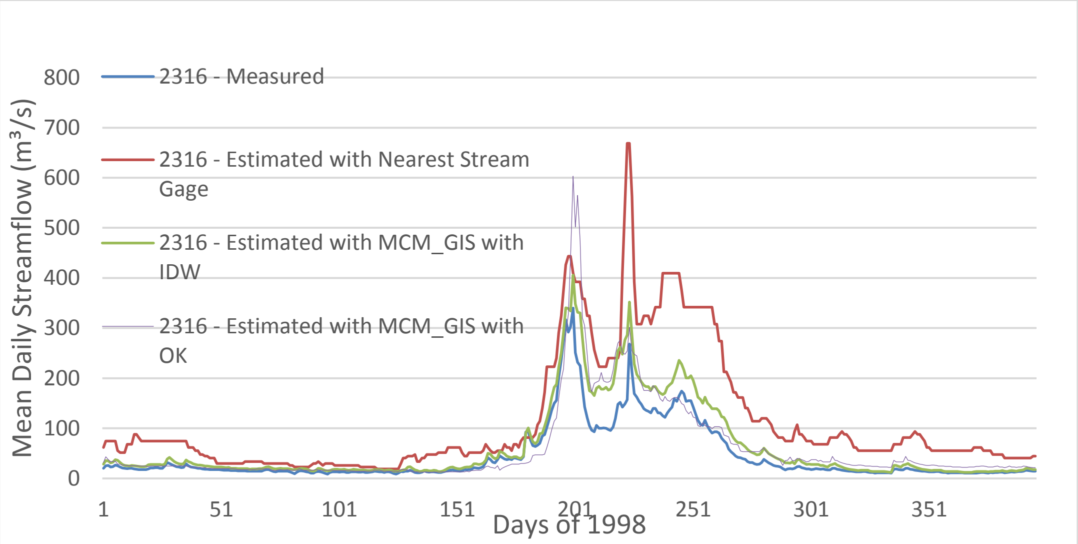


Figure 80: Hydrograph of Streamgage 2316 obtained by MCM_GIS

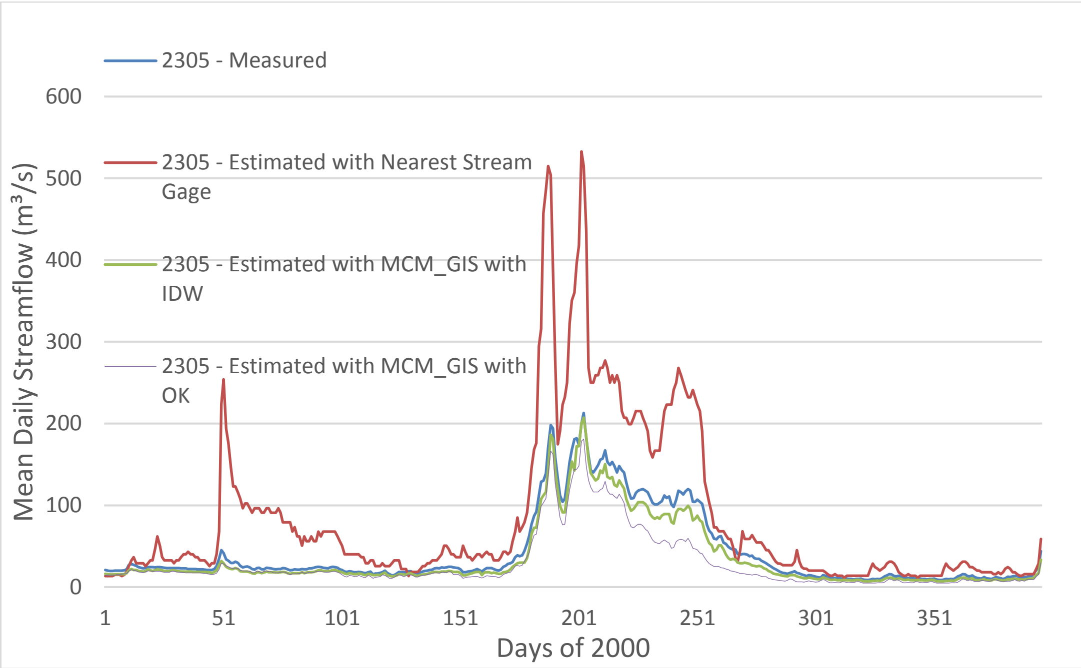


Figure 81:Hydrograph of Streamgage 2305 obtained by MCM_GIS

CHAPTER 5

CONCLUSION

Estimating streamflow at ungaged locations is still a challenge in Turkey. One of the most commonly used approaches is transferring streamflow observations of the spatially closest streamgage to the ungaged location. However, response of the basin corresponding to the spatially closest streamgage may not be representative of the ungaged basin. Utilization of the basin with the most similar hydrological response (i.e. similar streamflow values) as the reference streamgage will produce better streamflow estimations at the ungaged location. MCM allows identification of the most correlated streamgage with the ungaged location. Although the utility of MCM has been demonstrated in a number of recent studies, its application in practice has been limited. One of the major drawbacks of the method is its multi-step and rather time-consuming procedure. The goal of this study is to ease utilization of MCM through GIS integration. A new tool, called MCM_GIS that is composed of easy-to-use interfaces is developed in this study. MCM_GIS is tested on two different study areas, Western Black Sea and Coruh Regions. Major contributions of this study are:

1. Faster application of MCM is made possible through the newly developed GIS tool. The tool is partially-automated and is able to perform time consuming and recursive steps relatively quickly. Main output of the tool is the most correlated streamgage which is used as the reference streamgage to estimate streamflow time series at the ungaged basin. It is believed that daily streamflow values will be useful in the feasibility studies of water structures such Small Hydropower Plants.

2. MCM_GIS tool is developed with the ArcGIS platform which is commonly used in private and public institutions. The toolset is composed of several modules written in Python scripting language. User interfaces are designed to guide the user through the MCM application steps. Due to these characteristics, it is believed that MCM_GIS will increase utilization of MCM for practical applications.
3. Originally, MCM is proposed to work with OK. MCM_GIS enables comparison of different spatial interpolation algorithms that are embedded in ArcGIS. In this study, results obtained using IDW and OK are compared at two case study sites. For these sites, no significant difference in *NSE* efficiency values is observed.
4. Based on case study trials, it is observed that location of the ungaged basin close to the center of the study area improves the efficiency of estimations.

Future research on the application of MCM and the developed GIS tool, MCM_GIS is recommended to focus on:

a. Effect of other interpolation algorithms and associated model parameters such as neighbourhood type, lag size, minimum and maximum number of neighbours can be further investigated. Initial suggestions for model parameters may be generated to ease the model parameter optimization procedure.
b. Currently, MCM_GIS tool is partially-automated. One module works in MATLAB and the rest of the modules work in ArcMap and written in Python. Fully automation may contribute to widespread use of the tool by large number of users. The number of modules can be reduced and they may be further combined into a single fully automated tool.

REFERENCES

Adams, R. M., Hurd, B. H., Lenhart, S., & Leary, N. (1998). Effects of global climate change on agriculture: an interpretative review. Climate Research, 11(1), 19-30.

Ahmed, S., & Murali, G. (1992). Regionalization of fluoride content in an aquifer. *J Environmental Hydrology,*, *1*(1).

Ajami, N. K., Gupta, H., Wagener, T., & Sorooshian, S. (2004). Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. *Journal of Hydrology*, *298*(1), 112-135.

Archfield, S. A., & Vogel, R. M. (2010). Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungaged catchments. Water Resources Research, 46(10).

Archfield, S. A., Vogel, R. M., & Wagener, T. (2009, December). Rainfall-runoff model calibration at an ungaged catchment using the map-correlation method. In AGU Fall Meeting Abstracts.

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.

Baba, M., Unnikrishnan, A. S., Rajawat, A. S., Bhattacharya, S., Ratheesh, R., Kurian, N. P. & Sundar, D. (2013). Demarcation of coastal vulnerability line along the Indian coast.

Bacchi, B., & Borga, M. (1993). Spatial correlation patterns and rainfall fields analysis. *Excerpta of the Italian Contributions to the Field of Hydraulic Engineering*, *7*, 7-40.

Borga, M., & Vizzaccaro, A. (1997). On the interpolation of hydrologic variables: formal equivalence of multiquadratic surface fitting and kriging. *Journal of Hydrology*, *195*(1-4), 160-171.

Chang, F. J., & Chen, Y. C. (2001). A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction. Journal of hydrology, 245(1), 153-164.

Darricau-Beucher, H. (1981). *Approche géostatistique du passage des données de terrain aux paramètres des modèles en hydrogéologie* (Doctoral dissertation).

Delhomme, J. P. (1974). La cartographie d'une grandeur physique a partir de donnees de differentes qualites. In *Proceedings of IAH Congress, Montpelier, France* (Vol. 10, pp. 185-194).

Devia, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). A review on hydrological models. *Aquatic Procedia*, *4*, 1001-1007.

Dou, C., Woldt, W., & Bogardi, I. (1999). Fuzzy rule-based approach to describe solute transport in the unsaturated zone. Journal of Hydrology, 220(1), 74-85.

Emerson, D. G., Vecchia, A. V., and Dahl, A. L. (2005). "Evaluation of DrainageArea 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.

Ergen, K. (2012). Application Of The Map Correlation Method To The Western Blacksea Basin (M.Sc. Thesis, Middle East Technical University).

Ergen, K., & Kentel, E. (2016). An integrated map correlation method and multiple-source sites drainage-area ratio method for estimating streamflows at ungaged catchments: A case study of the Western Black Sea Region, Turkey. Journal of environmental management, 166, 309-320.

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.

ESRI Development Team (2010). Geostatistical Analyst Tutorial

Gambolati, G., & Volpi, G. (1979). A conceptual deterministic analysis of the kriging technique in hydrology. *Water Resources Research*, *15*(3), 625-629.

Goovaerts, P. (1997). Geostatistics for natural resources evaluation. Oxford University Press on Demand.

Goovaerts, P. (1999) a. Geostatistics in soil science: state-of-the-art and perspectives. Geoderma, 89(1), 1-45.

Goovaerts, P. (1999) b. Using elevation to aid the geostatistical mapping of rainfall erosivity. Catena, 34(3), 227-242.

Grayson, R., & Blöschl, G. (Eds.). (2001). *Spatial patterns in catchment hydrology: observations and modelling*. CUP Archive.

Hardy, R. L., & Göpfert, W. M. (1975). Least squares prediction of gravity anomalies, geoidal undulations, and deflections of the vertical with multiquadric harmonic functions. *Geophysical Research Letters*, *2*(10), 423-426.

Hoeksema, R. J., & Kitanidis, P. K. (1984). An application of the geostatistical approach to the inverse problem in two- dimensional groundwater modeling. *Water Resources Research*, *20*(7), 1003-1020.

Hughes, D. A., & Smakhtin, V. (1996). Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves. Hydrological Sciences Journal, 41(6), 851-871.

Huijbregts, C., & Matheron, G. (1971). Universal kriging (an optimal method for estimating and contouring in trend surface analysis). In *Proceedings of Ninth International Symposium on Techniques for Decision-making in the Mineral Industry*.

Johnston, K., Ver Hoef, J. M., Krivoruchko, K., & Lucas, N. (2001). Using ArcGIS geostatistical analyst (Vol. 380). Redlands: Esri.

Journel, A.G. and C.J. Huijbregts. 1978. Mining Geostatistics. New York: Academic Press.

Journel, A. G., & Journel, A. G. (1989). *Fundamentals of geostatistics in five lessons* (Vol. 8). Washington, DC: American Geophysical Union.

Kentel, E. (2009). Estimation of river flow by artificial neural networks and identification of input vectors susceptible to producing unreliable flow estimates. Journal of hydrology, 375(3), 481-488.

Keskin, F. (2007). Hydrological Model Study In Yuvacık Dam Basin By Using GIS Analysis, (M. Sc. Thesis, 2007)

Khan, S.U., Noor, A. and Farooqi, I. H. (2015). GIS Application for Groundwater Management and Quality Mapping in Rural Areas of District Agra, India, International Journal of Water Resources and Arid Environments 4(1): 89-96, 2015 (http://www.psipw.org/attachments/article/365/89-96.pdf)

Kolmogorov, A. N. (1941, February). The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. In *Dokl. Akad. Nauk SSSR* (Vol. 30, No. 4, pp. 299-303).

Luo, W., Taylor, M. C., & Parker, S. R. (2008). A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. International journal of climatology, 28(7), 947-959.

Ly, S., Charles, C., & Degré, A. (2013). Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. A review. *Biotechnologie, Agronomie, Société et Environnement*, *17*(2), 392.

Maguire, D. J. (1991). An overview and definition of GIS. Geographical information systems: Principles and applications, 1, 9-20.

Matern, B. (1963). Spatial Variations. 2nd. *Lecture notes in Statistics*, *36*.

Matheron, G. (1965). Les variables régionalisées et leur estimation.

Meng, Q., Liu, Z., & Borders, B. E. (2013). Assessment of regression kriging for spatial interpolation–comparisons of seven GIS interpolation methods. *Cartography and Geographic Information Science, 40*(1), 28-39.

Mizell, S. A., Gelhar, L. W., & Gutjahr, A. L. (1980). *Stochastic analysis of spatial variability in two-dimensional groundwater flow with implications for observation-well-network design*. Hydrology Research Program, Geophysical Research Center, Research and Development Division, New Mexico Institute of Mining & Technology.

Mbilinyi, B. P., Tumbo, S. D., Mahoo, H., & Mkiramwinyi, F. O. (2007). GIS-based decision support system for identifying potential sites for rainwater harvesting. Physics and Chemistry of the Earth, Parts A/B/C, 32(15), 1074-1081.

Meng, Q., Liu, Z., & Borders, B. E. (2013). Assessment of regression kriging for spatial interpolation–comparisons of seven GIS interpolation methods. Cartography and Geographic Information Science, 40(1), 28-39.

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”. 8th IAHS Scientific Assembly and 37th IAH Congress, Hyderabad, India.

Moradkhani, H. and Sorooshian, S., 2008. General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. Hydrological modeling and the water cycle.Springer. 291 p. ISBN 978-3-540-77842-4.

Monjardin, C. E. F., Uy, F. A. A., & Tan, F. J. (2017, June). Estimation of River Discharge at Ungauged Catchment using GIS Map Correlation Method as Applied in Sta. Lucia River in Mauban, Quezon, Philippines. In *Materials Science and Engineering Conference Series* (Vol. 216, No. 1, p. 012045).

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

Neuman, S. P., & Jacobson, E. A. (1984). Analysis of nonintrinsic spatial variability by residual kriging with application to regional groundwater levels. *Mathematical Geology*, *16*(5), 499-521.

Ocal, D., & Kentel, E. (2017, July). A GIS Tool to Estimate Flow at Ungaged Basins Using the Map Correlation Method. In International Conference on Computational Science and Its Applications (pp. 377-391). Springer, Cham.

Özcan, Z., Başkan, O., Düzgün, H. Ş., Kentel, E., & Alp, E. (2017). A pollution fate and transport model application in a semi-arid region: Is some number better than no number?. Science of The Total Environment, 595, 425-440.

Özdemir, A. (2016). Hierarchical Approach To Semi-Distributed Hydrological Model Calibration (Doctoral Dissertation, Middle East Technical University).

Pardo-Igúzquiza, E. (1998). Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. *Journal of Hydrology*, *210*(1), 206-220.

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

Pesti, G., Shrestha, B. P., Duckstein, L., & Bogárdi, I. (1996). A fuzzy rule- based approach to drought assessment. Water Resources Research, 32(6), 1741-1747.

Roth, K. (1995). Soil physics lecture notes. *Institute of Environmental Physics, Heidelberg University*, *2*.

Roth, N. E., Allan, J. D., & Erickson, D. L. (1996). Landscape influences on stream biotic integrity assessed at multiple spatial scales. *Landscape ecology*, *11*(3), 141-156.

Samarasinghea, S. M. J. S., Nandalalb, H. K., Weliwitiyac, D. P., Fowzed, J. S. M., Hazarikad, M. K., & Samarakoond, L. (2010). Application of remote sensing and GIS for flood risk analysis: a case study at Kalu-Ganga river, Sri Lanka. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science, 38(Pt 8), 110-115.

Schulz, K., & Huwe, B. (1997). Water flow modeling in the unsaturated zone with imprecise parameters using a fuzzy approach. Journal of Hydrology, 201(1-4), 211-229.

Shu, C., & Ouarda, T. B. (2012). Improved methods for daily streamflow estimates at ungauged sites. *Water Resources Research*, *48*(2).

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.

Thangarajan, M., & Ahmed, S. (1989). Kriged estimates of water levels from the sparse measurements in a hard rock aquifer. In *Proc. of Internat. Groundwater Workshop (IGW-89), Hyderabad, India, Feb* (Vol. 28, pp. 287-302).

Theurillat, J. P., & Guisan, A. (2001). Potential impact of climate change on vegetation in the European Alps: a review. Climatic change, 50(1-2), 77-109.

Thumerer, T., Jones, A. P., & Brown, D. (2000). A GIS based coastal management system for climate change associated flood risk assessment on the east coast of England. International Journal of Geographical Information Science, 14(3), 265-281.

Whittle, P., 1963, Prediction and regulation by linear least-square methods: van Nostrand, Princeton, 141 pp.

Wiener, N. (1949). *Extrapolation, interpolation, and smoothing of stationary time series* (Vol. 7). Cambridge, MA: MIT press.

Yu, P. S., Chen, C. J., & Chen, S. J. (2000). Application of gray and fuzzy methods for rainfall forecasting. Journal of Hydrologic Engineering, 5(4), 339-345.

Zambelli, P., Gebbert, S., & Ciolli, M. (2013). Pygrass: An object oriented python application programming interface (API) for geographic resources analysis support system (GRASS) geographic information system (GIS). ISPRS International Journal of Geo-Information, 2(1), 201-219.

APPENDIX A

HYDROGRAPHS

This section includes hydrographs of all streamgages that are selected in both of the study areas. The hydrographs represent the measured and estimated hydrographs with different reference streamgage selection criteria. Figure 82 to Figure 92 show streamgage estimations in Western Blacksea Region and Figure 93 to Figure 106 show estimations performed in Coruh Region.

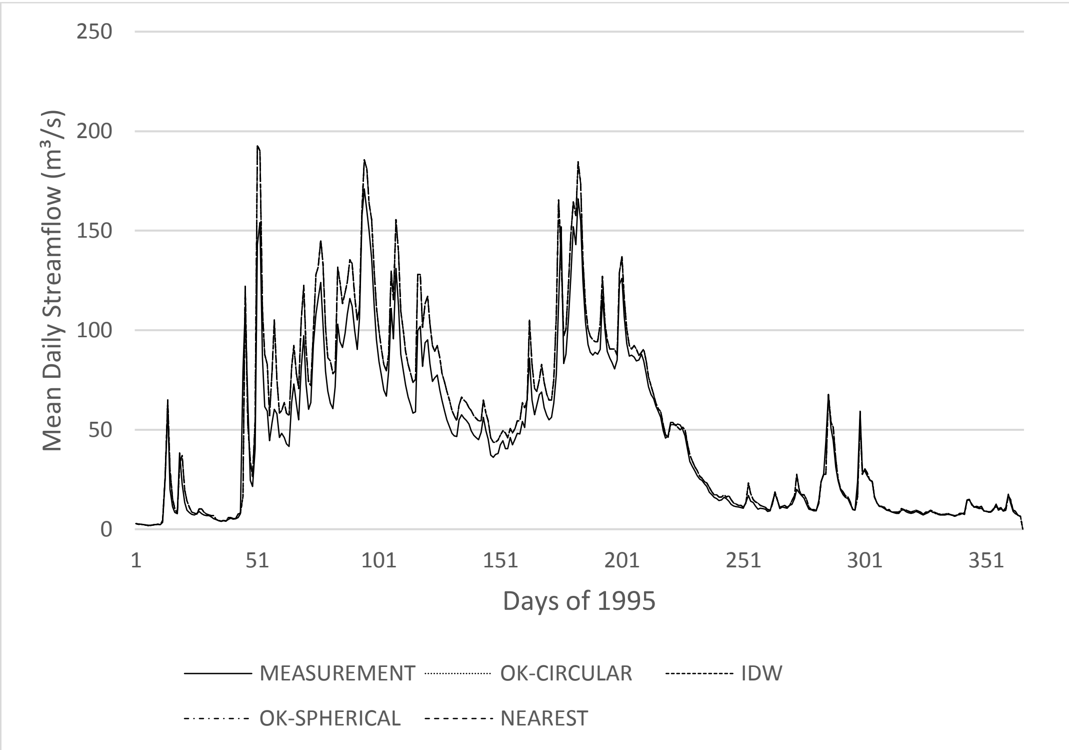


Figure 82: Hydrograph of streamgage 1302

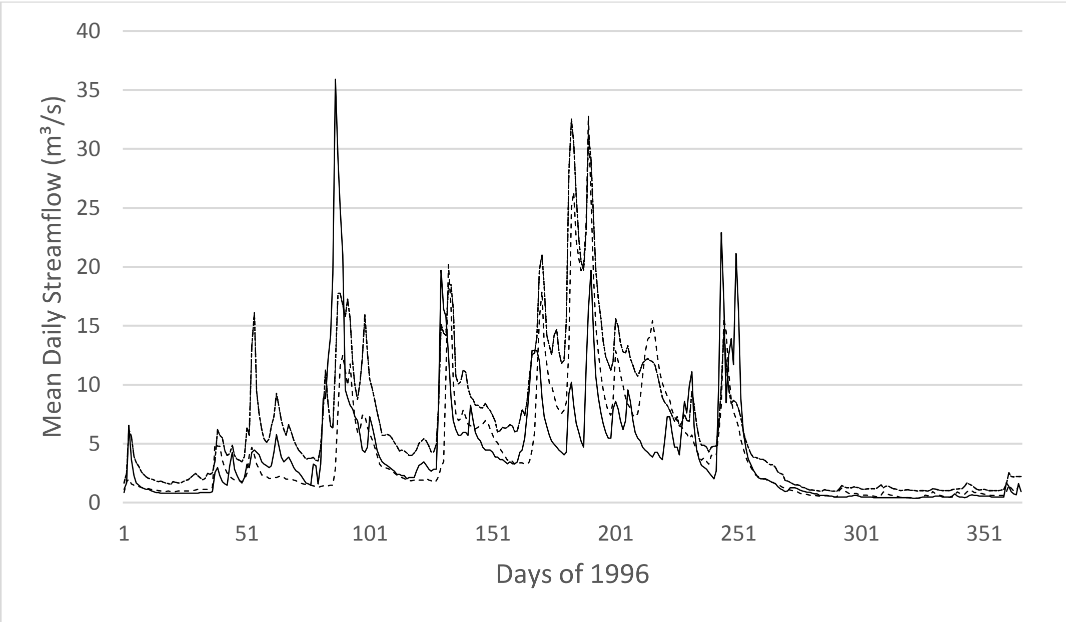


Figure 83: Hydrograph of streamgage 1307

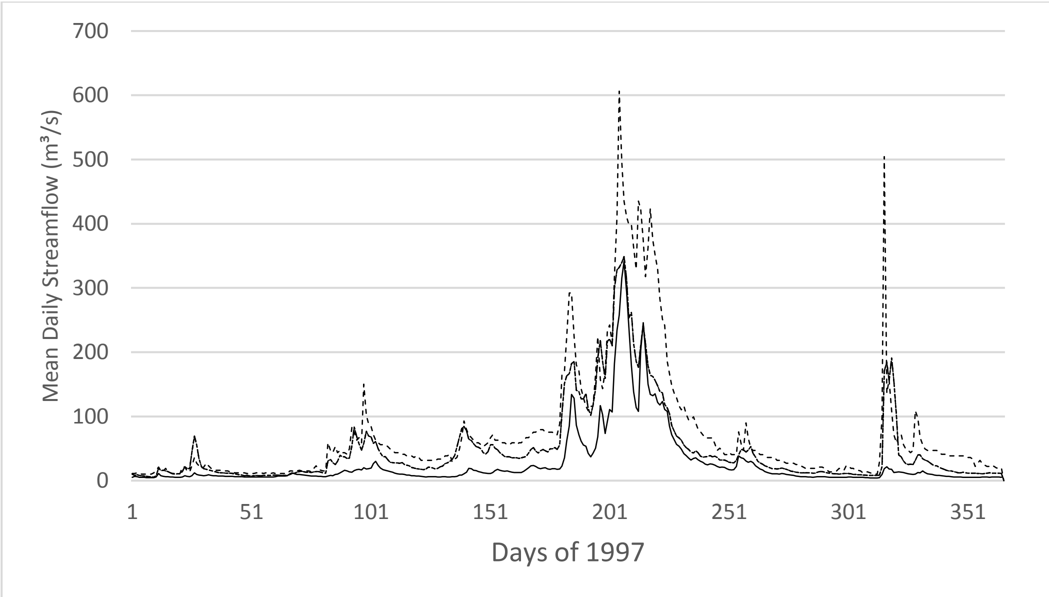


Figure 84: Hydrograph of streamgage 1314

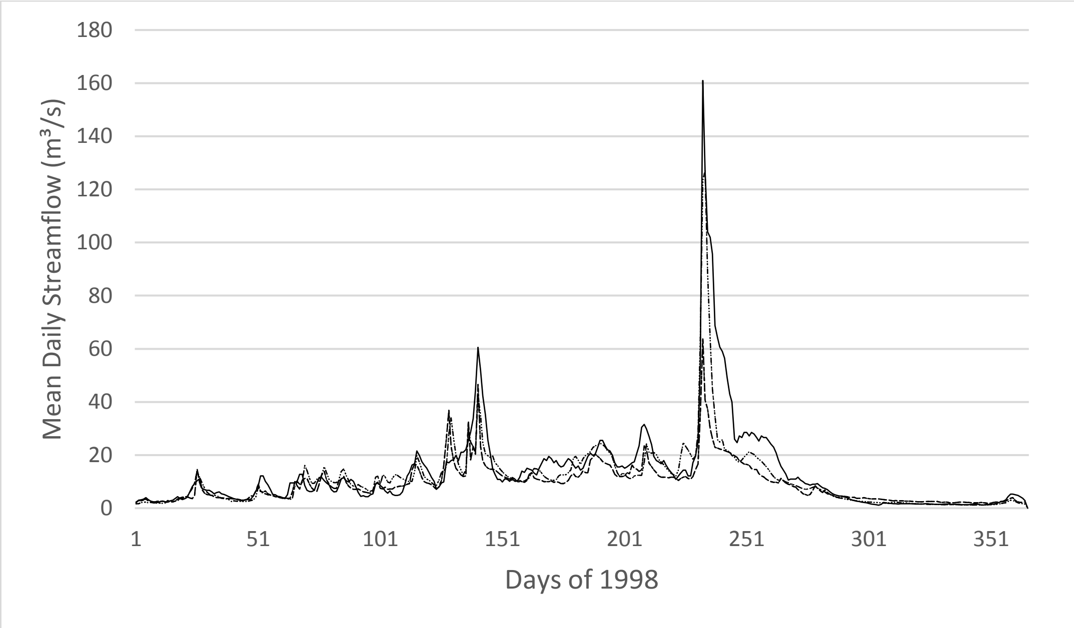


Figure 85: Hydrograph of streamgage 1319

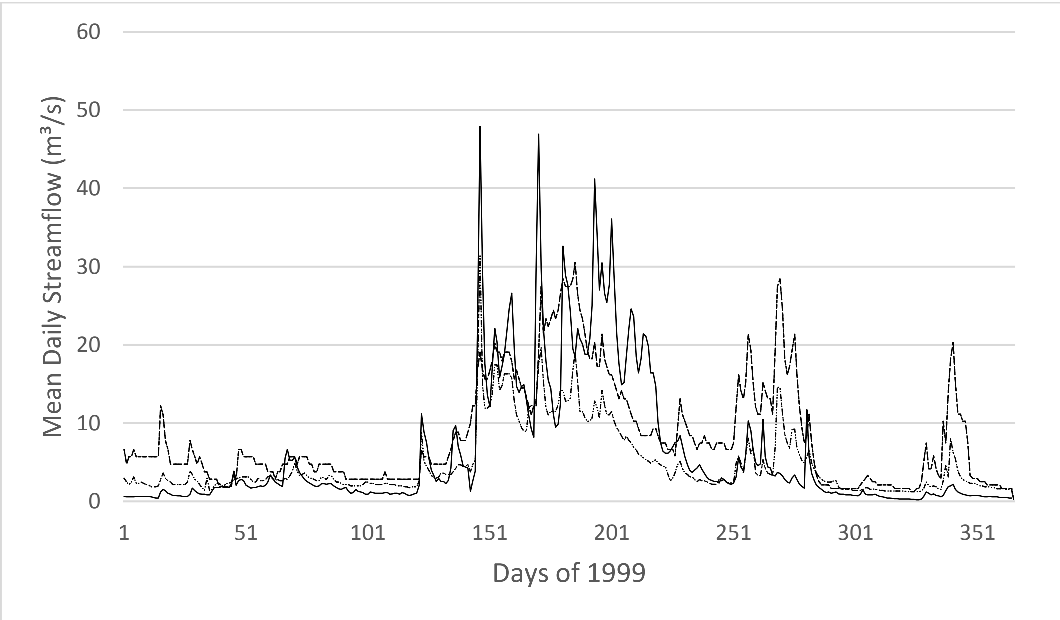


Figure 86: Hydrograph of streamgage 1327

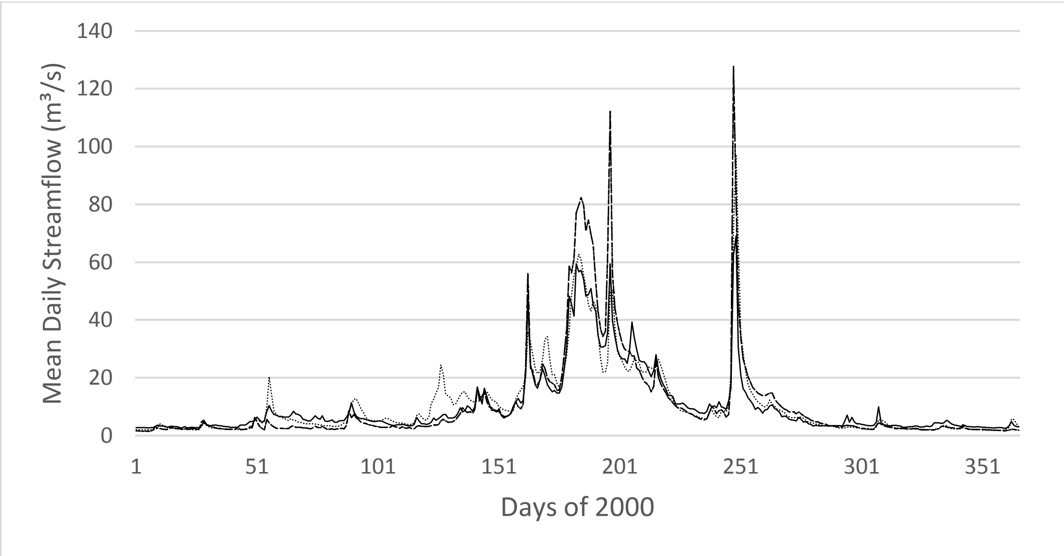


Figure 87: Hydrograph of streamgage 1334

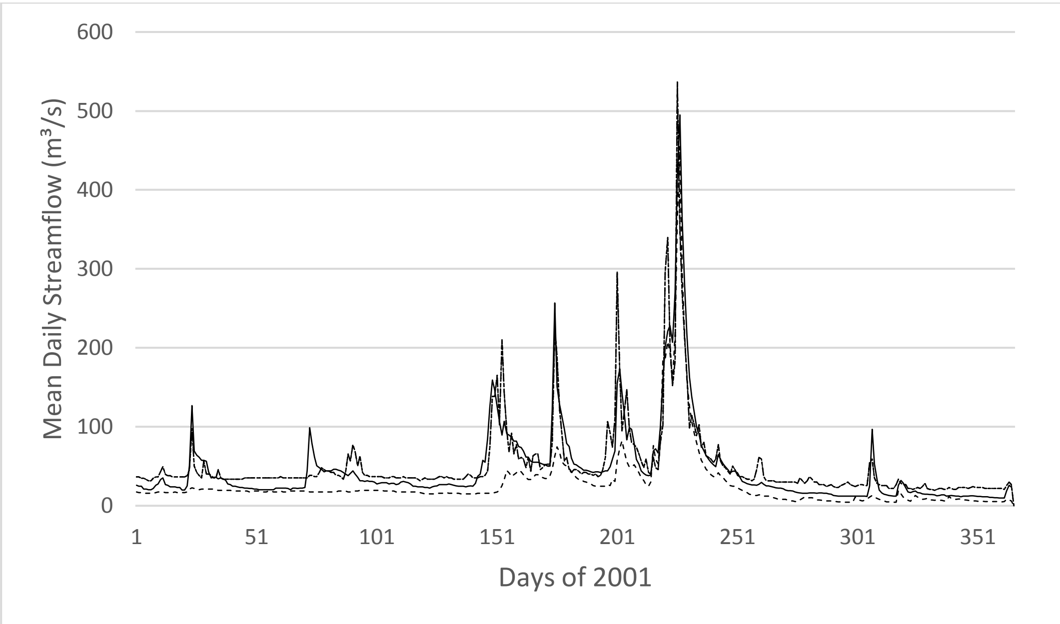


Figure 88: Hydrograph of streamgage 1335

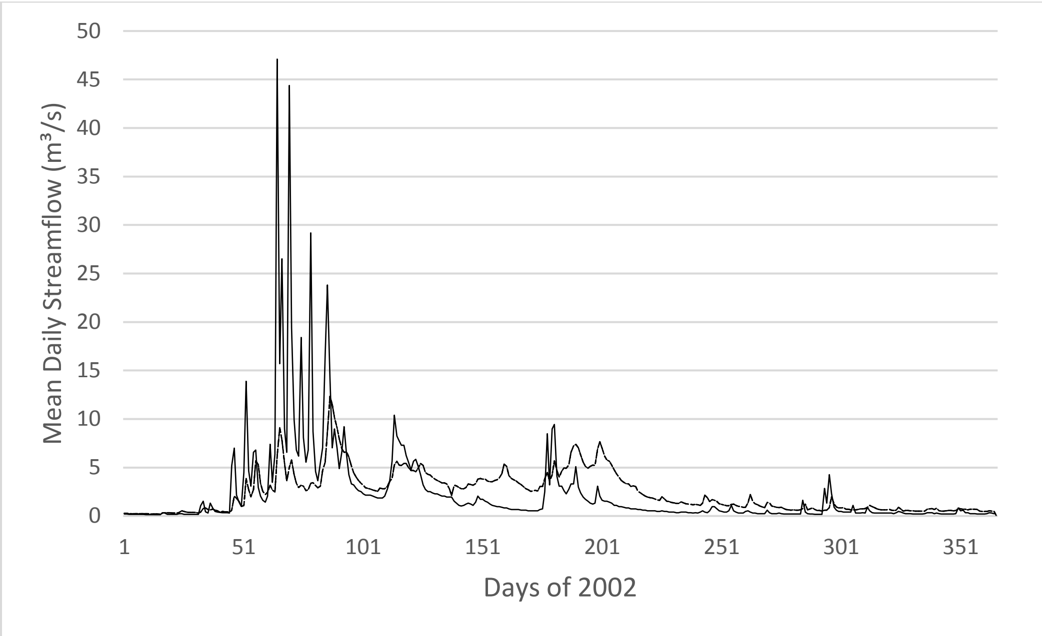


Figure 89: Hydrograph of streamgage 1338

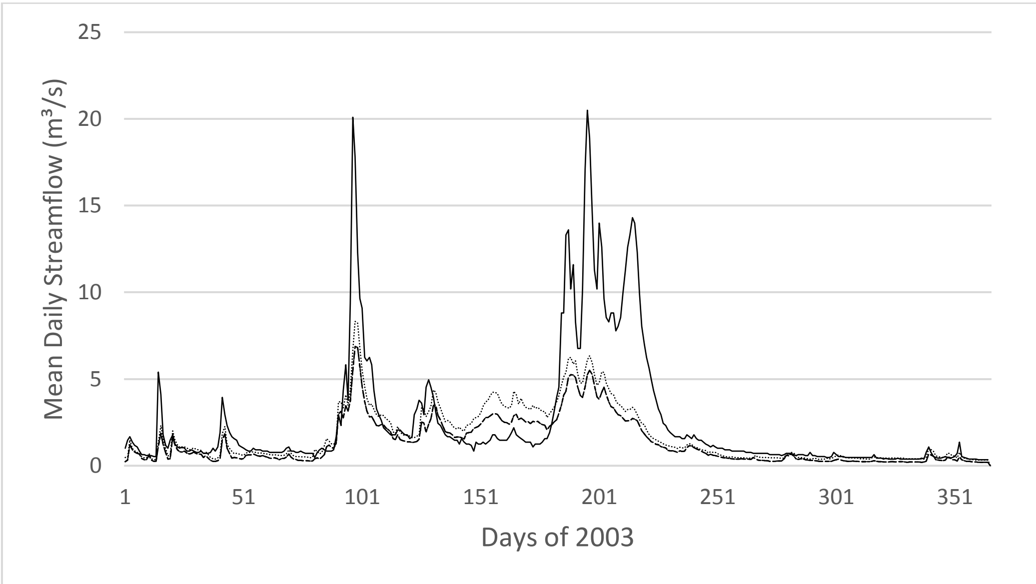


Figure 90: Hydrograph of streamgage 1339

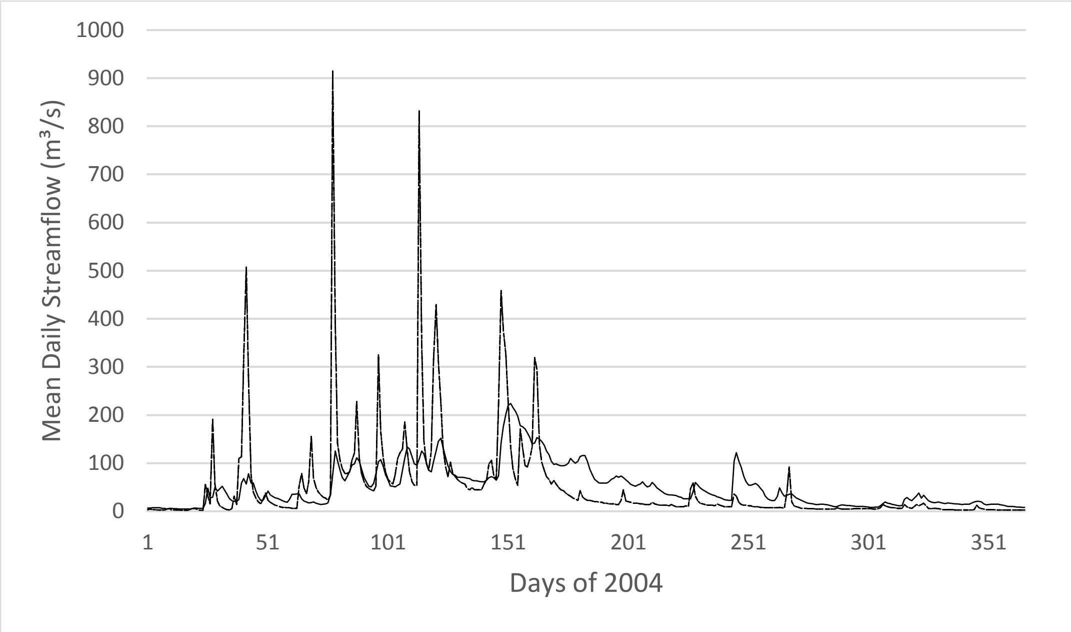


Figure 91: Hydrograph of streamgage 1340

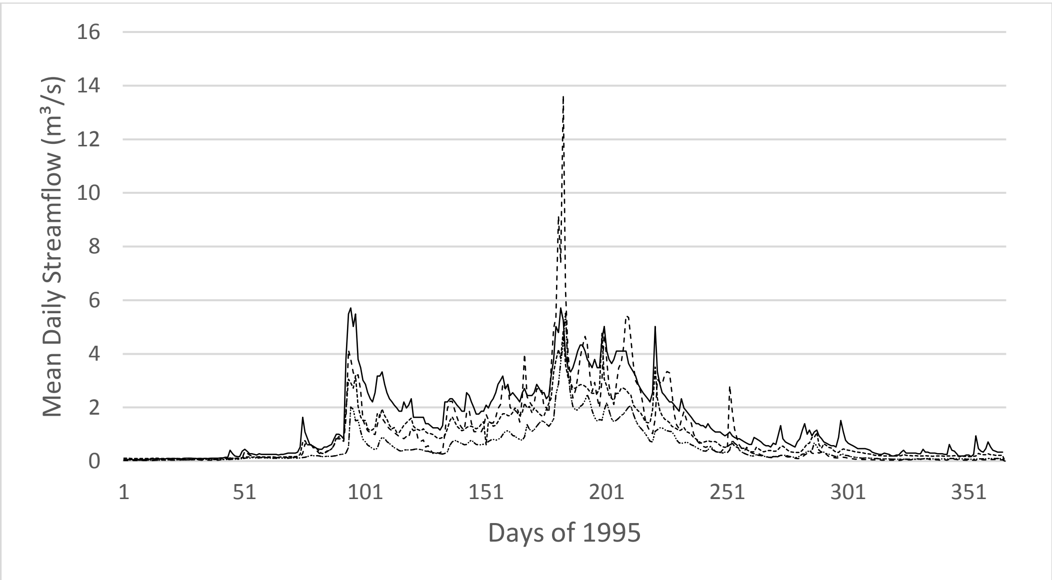


Figure 92: Hydrograph of streamgage 1343

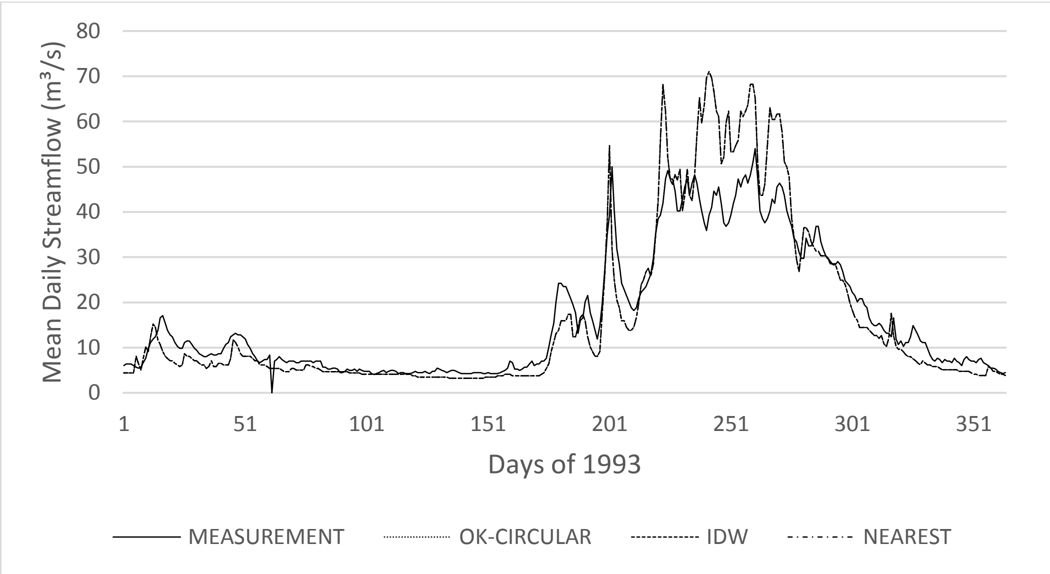


Figure 93: Hydrograph of streamgage 2215

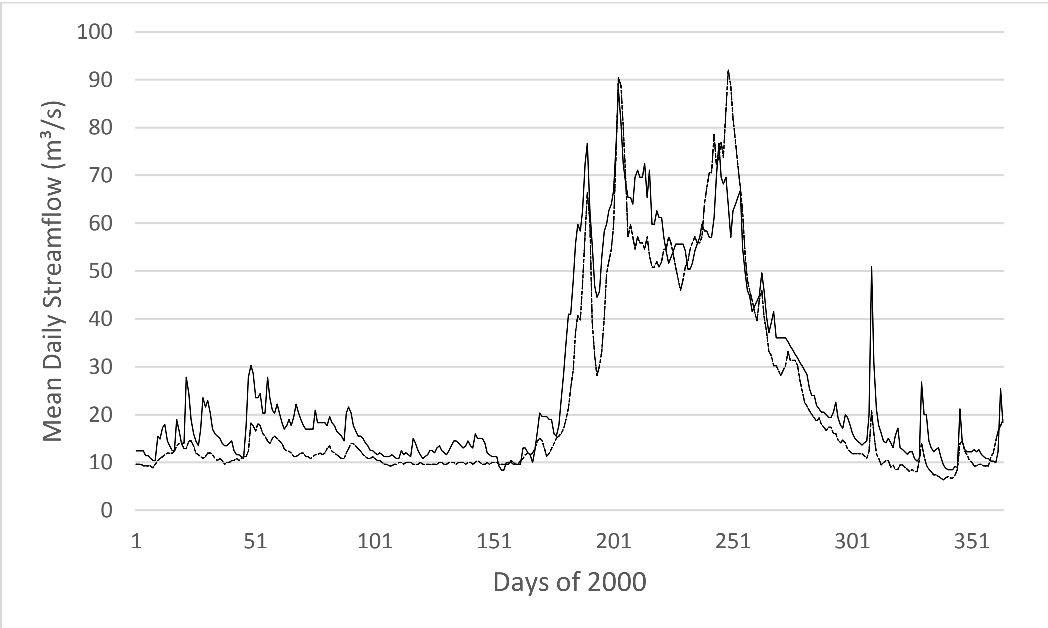


Figure 94: Hydrograph of streamgage 2218

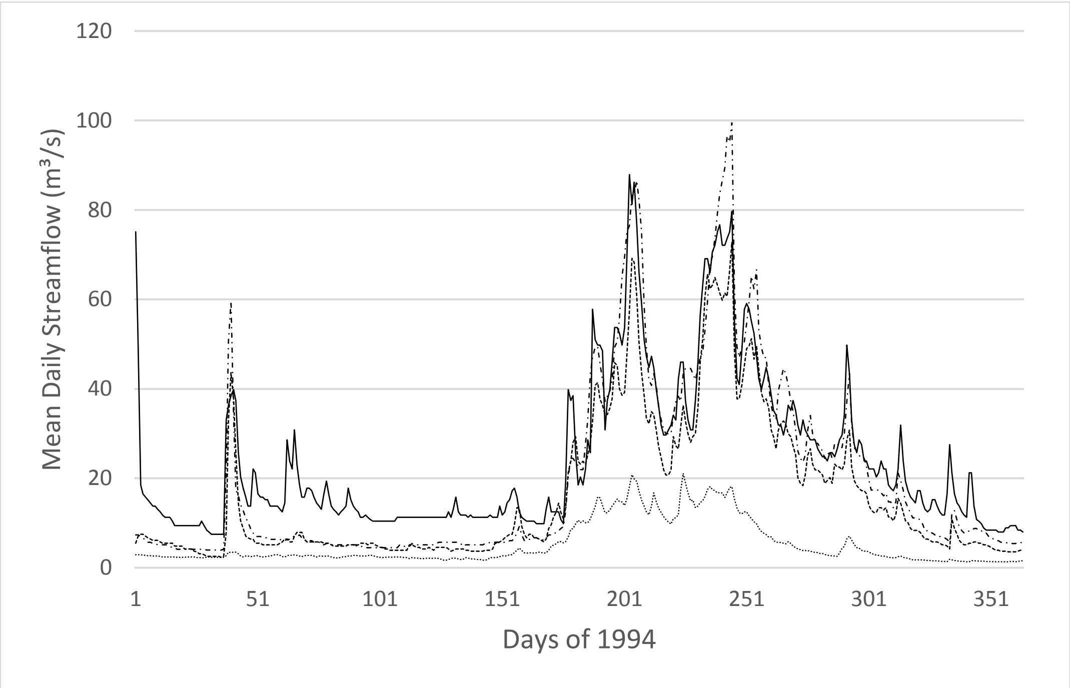


Figure 95: Hydrograph of streamgage 2232

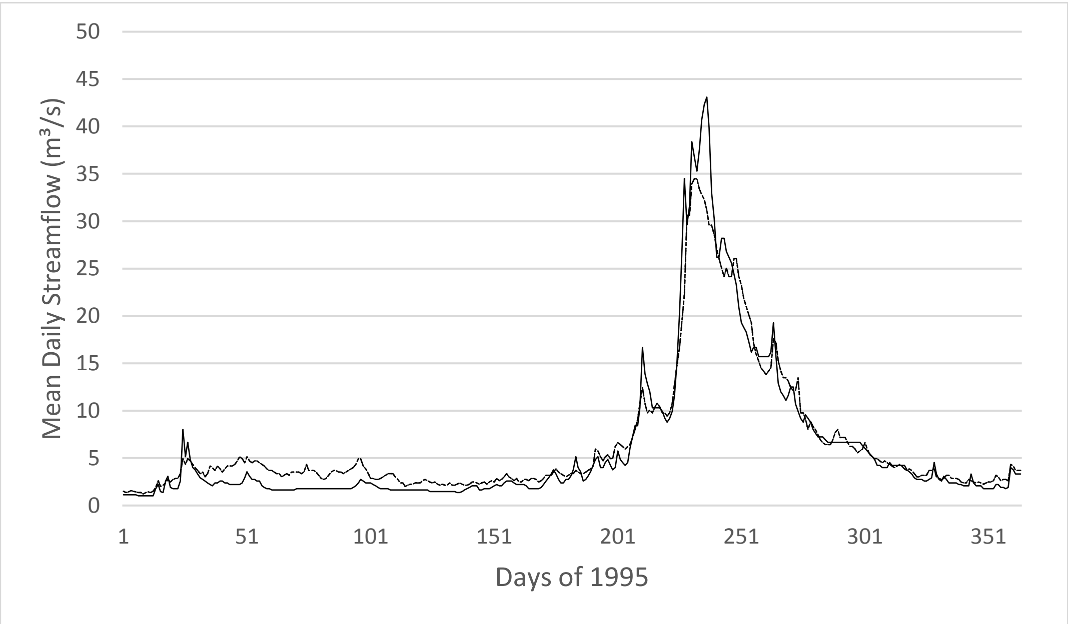


Figure 96: Hydrograph of streamgage 2233

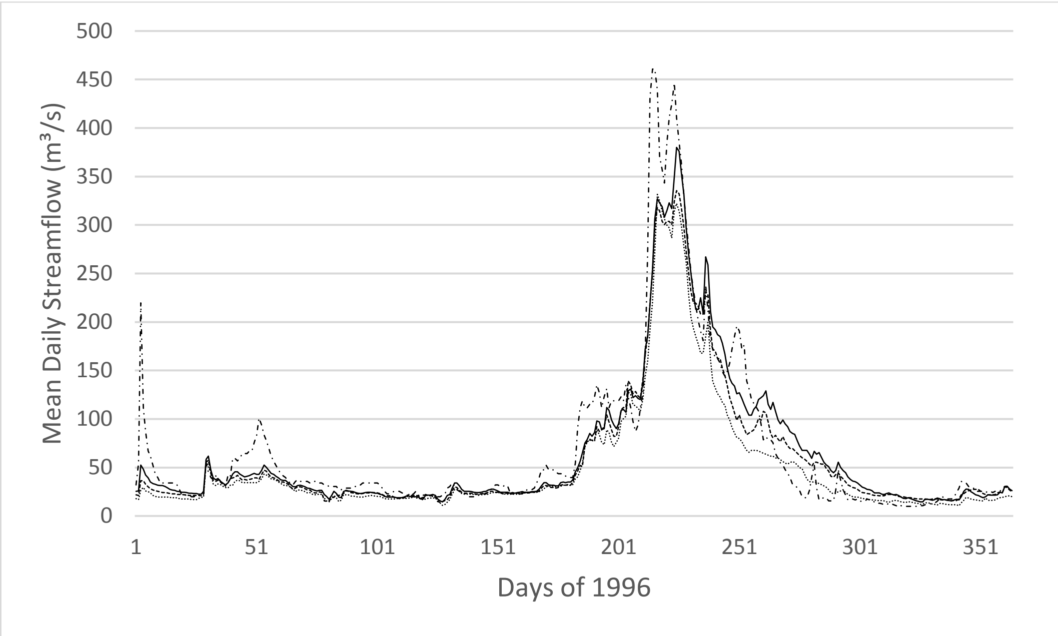


Figure 97: Hydrograph of streamgage 2305

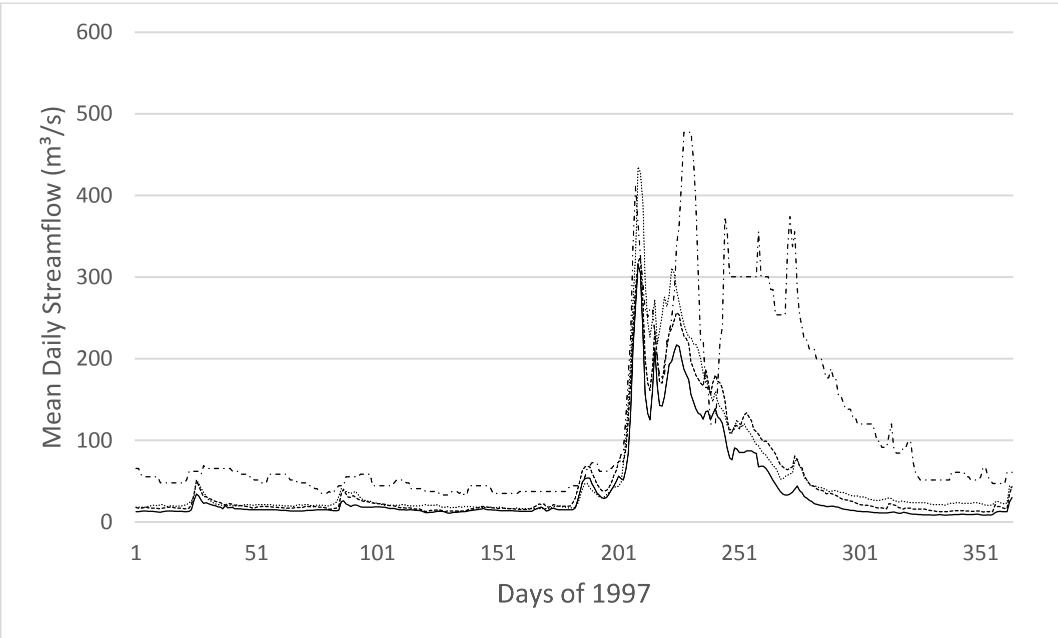


Figure 98: Hydrograph of streamgage 2316

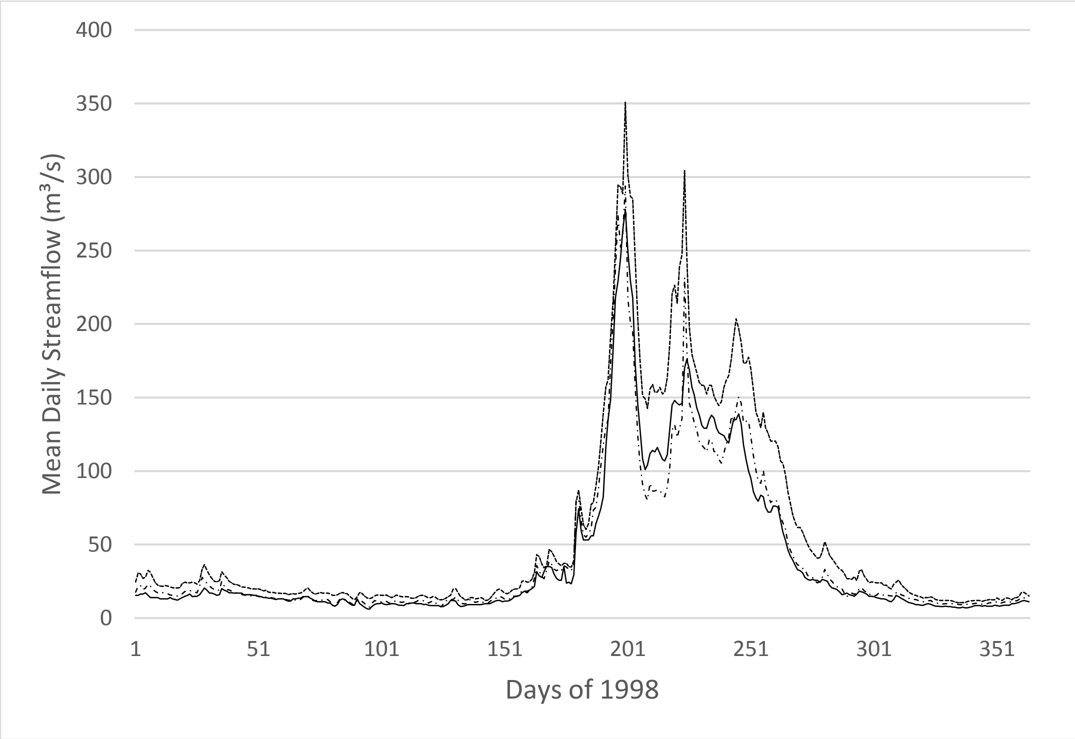


Figure 99: Hydrograph of streamgage 2320

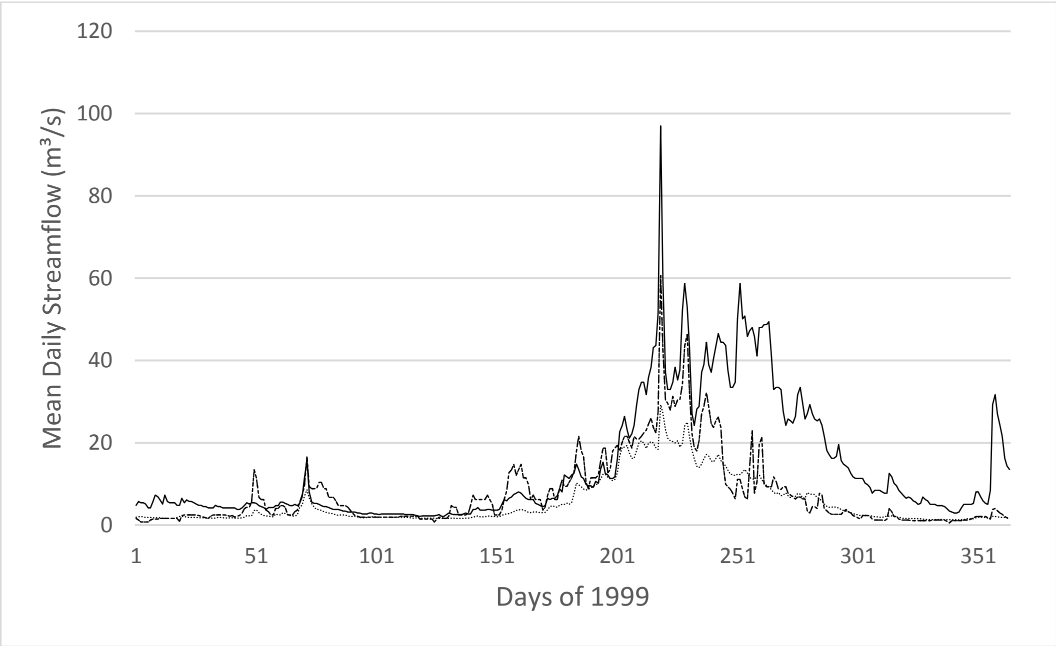


Figure 100: Hydrograph of streamgage 2321

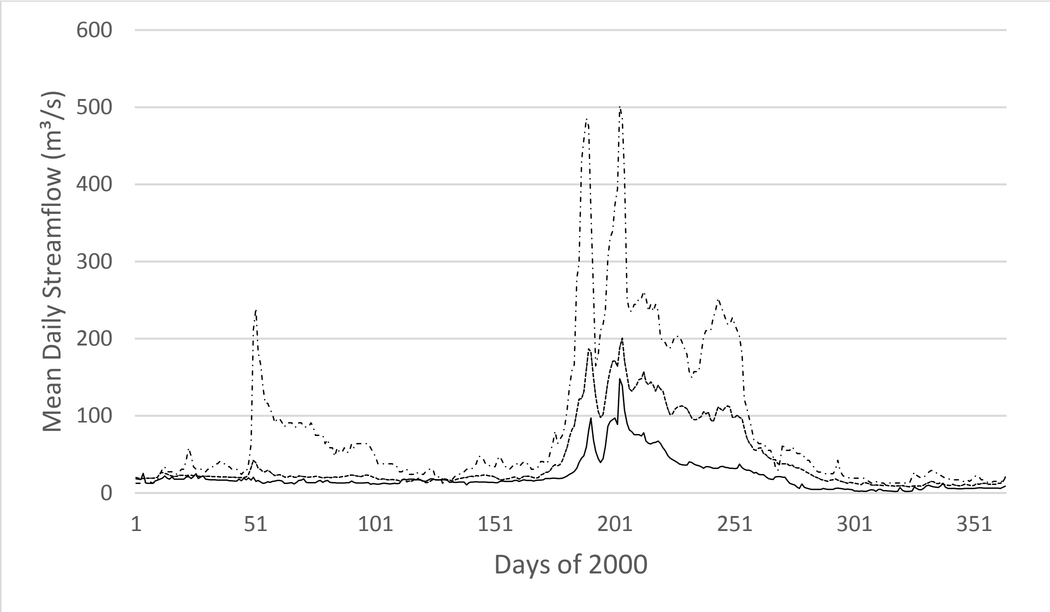


Figure 101: Hydrograph of streamgage 2323

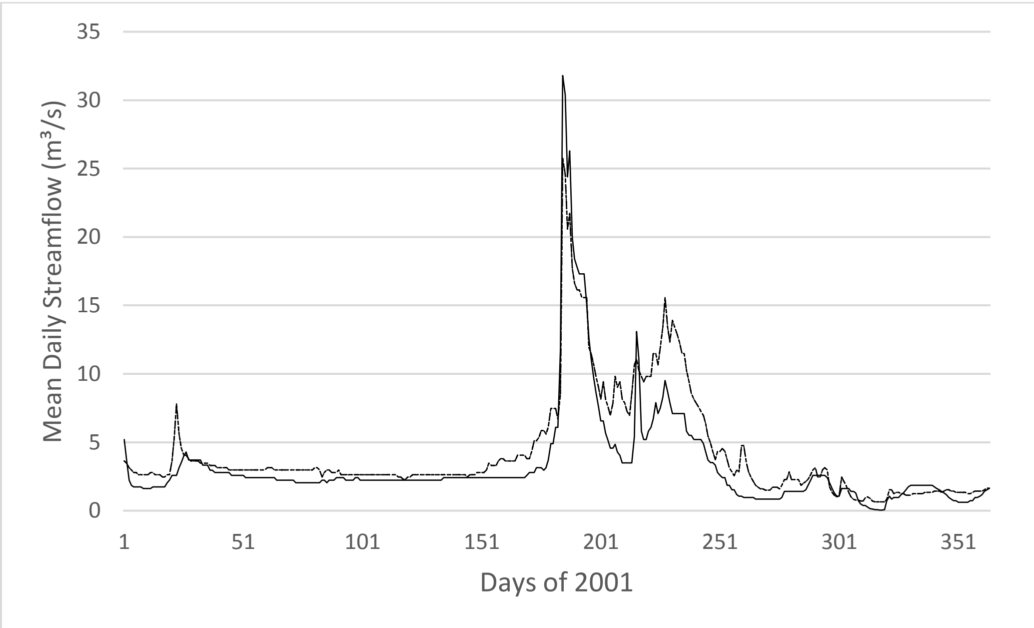


Figure 102: Hydrograph of streamgage 2325

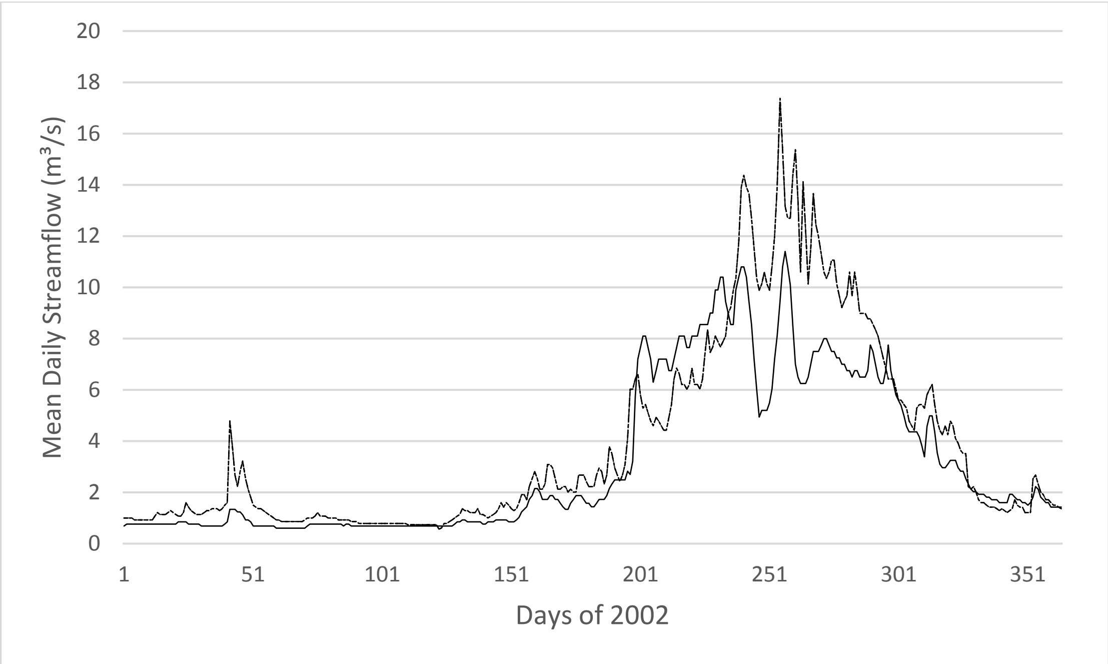


Figure 103: Hydrograph of streamgage 2330

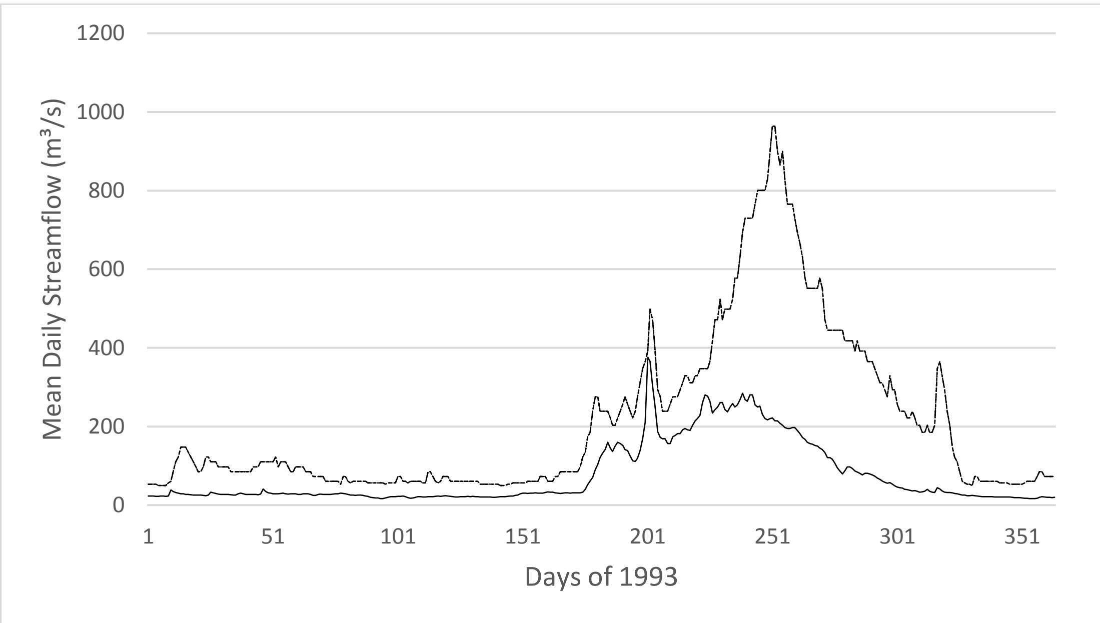


Figure 104: Hydrograph of streamgage 2337

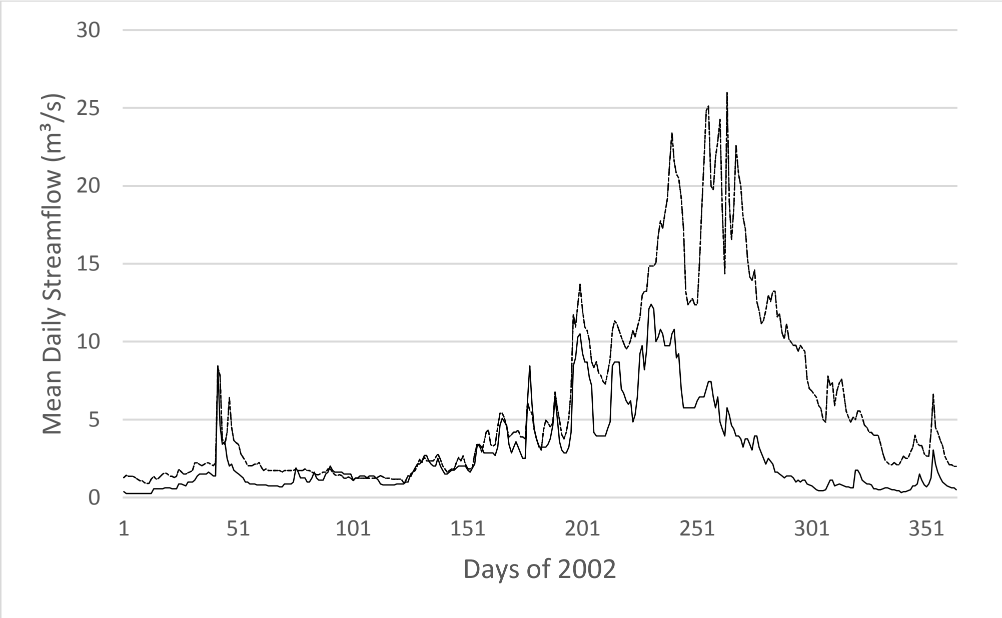


Figure 105: Hydrograph of streamgage 2340

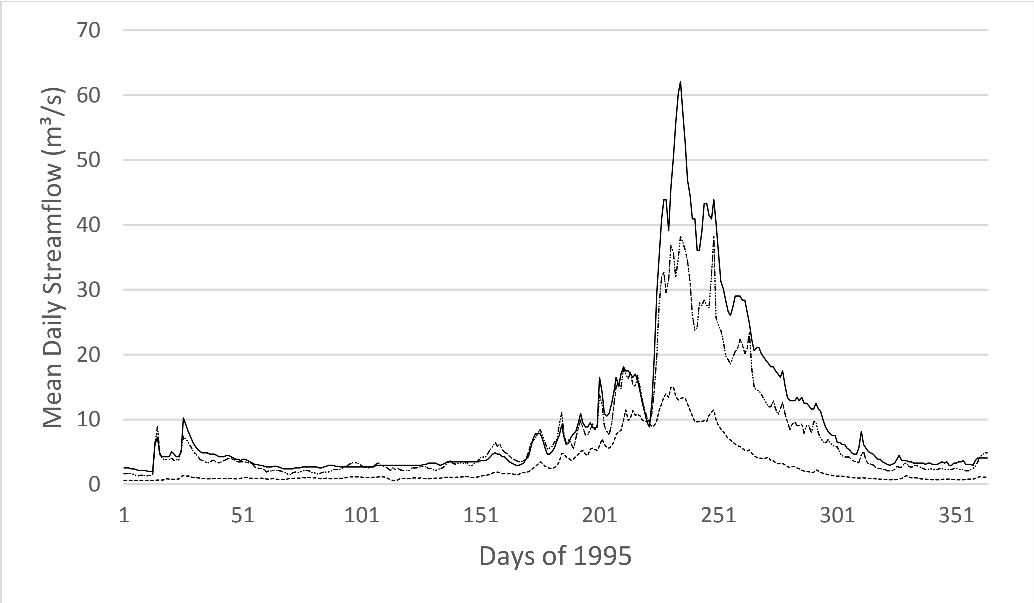


Figure 106: Hydrograph of streamgage 2342

APPENDIX B

ERROR STATISTICS

This section includes Root Mean Square Error (RMSE) of all streamgages that are selected in both of the study areas. The RMSE represent the root of average squared error between measured and estimated daily streamflow time series in m^3/s with different reference streamgage selection criteria. Table 6 show RMSE in Western Blacksea Region and Table 7 show RMSE values in Coruh Region.

Table 6: RMSE Values for different streamgage selection criteria in Western Blacksea Region

Ungaged	Nearest Stream Gage	Most correlated measured	The Most Correlated Stream Gage using MCM with OK - CIRCULAR	The Most Correlated Stream Gage using MCM with IDW	The Most Correlated Stream Gage using MCM with OK - SPHERICAL
1302	13.33	13.33	13.33	13.33	13.33
1307	7.60	6.68	6.68	6.68	6.68
1314	57.38	36.50	26.50	26.50	26.50
1319	4.57	4.57	3.83	4.57	3.83
1327	8.93	7.59	7.87	8.93	7.87
1334	6.53	6.53	4.70	6.53	6.53
1335	69.28	57.09	57.09	57.09	57.09
1338	3.96	3.96	3.96	3.96	3.96
1339	3.95	3.95	3.71	3.95	3.95
1340	82.23	14.57	82.23	82.23	82.23
1343	1.17	0.93	1.41	0.93	1.41

Table 7: RMSE Values for different streamgage selection criteria in Coruh Region

Ungaged	Nearest Stream Gage	Most correlated measured	The Most Correlated Stream Gage using MCM with IDW	The Most Correlated Stream Gage using MCM with OK
2215	4.85	4.85	4.85	4.85
2218	8.18	11.84	8.18	8.18
2232	13.45	8.96	14.21	27.70
2233	2.43	2.43	2.43	2.43
2305	83.50	11.56	11.56	28.14
2316	142.97	14.91	21.30	24.50
2320	7.31	7.31	22.36	22.36
2321	11.65	6.10	11.65	13.02
2323	115.64	0.99	0.99	0.99
2325	4.34	4.34	4.34	4.34
2330	2.72	1.69	2.72	2.72
2337	7.06	10.55	7.06	7.06
2340	2.84	1.94	2.84	2.84
2342	5.91	5.91	9.35	5.91