A black background with white text

Description automatically generated with low confidence

MSc Data Science Project

7PAM2002-0509-2023

Department of Physics, Astronomy and Mathematics

Data Science FINAL PROJECT

REPORT

Project Title:

Predicting Water Quality Parameters in Groundwater Using Ordinary Kriging

Student Name: Muhammad Rashid Munir

SRN: 21072699

Supervisor: Will Cooper

Date Submitted: August 29, 2024

Word Count: 7800

DECLARATION STATEMENT

This report is submitted in partial fulfilment of the requirement for the degree of Master of Science in Data Science at the University of Hertfordshire.

I have read the guidance to students on academic integrity, misconduct and plagiarism information at [Assessment Offences and Academic Misconduct](https://www.herts.ac.uk/__data/assets/pdf_file/0007/237625/AS14-Apx3-Academic-Misconduct-v17.0.pdf) and understand the University process of dealing with suspected cases of academic misconduct and the possible penalties, which could include failing the project module or course.

I certify that the work submitted is my own and that any material derived or quoted from published or unpublished work of other persons has been duly acknowledged. (Ref. UPR AS/C/6.1, section 7 and UPR AS/C/5, section 3.6). I have not used chatGPT, or any other generative AI tool, to write the reportor code (other than where declared or referenced).

I did not use human participants or undertake a survey in my MSc Project.

I hereby give permission for the report to be made available on module websites provided the source is acknowledged.

Student Name printed: Muhammad Rashid Munir

Student Name signature:

Student SRN number: 21072699

UNIVERSITY OF HERTFORDSHIRE

SCHOOL OF PHYSICS, ENGINEERING AND COMPUTER SCIENCE

DEDICATION

I dedicate this work to my beloved late parents, whom I love dearly

ACKNOWLEDGMENTS

Special thanks to Pakistan Council of Research in Water Resources (PCRWR), Lahore, Pakistan for allowing to use the spatial data on WQPs in this project report.

ABSTRACT

In this project report, the research question is to evaluate how multivariate statistical and geostatistical techniques can jointly explore and predict the water quality parameters (WQPs) and how ordinary kriging outperform the Inverse Distance Weighting (IDW) method in recognizing the contaminated areas and evaluating the groundwater quality in different spatial locations. To explore the research question, primary objective is to examine the WQPs and to predict their concentration levels at unsampled locations. The secondary objective to compare the predictive performance of ordinary kriging with IDW prediction method. At the first stage of statistical analysis, mean values of each WQP are computed and compared with the permissible limits of the World Health Organization (WHO). Beside this, coefficients of variation (CV) of each WQP are calculated as a relative measure of dispersion. The graphical display of correlation coefficients highlighted the pairwise associations among WQPs. The dendrogram of hierarchical cluster analysis and Factor analysis is considered sequentially which suggested four most significant and vital WQP including TDS, sulfate, sodium and chloride. As a second part of the analysis, these four WQPs are initially transformed using Box-Cox transformation to deal the asymmetrical WQPs. To capture the spatial correlation of the parameters, variogram model is estimated and the parameters (sill range and nugget) are plugged into the Ordinary Kriging model. The performance of the ordinary kriging technique is compared with the Inverse Distance Weighting (IDW) methods where Ordinary kriging outperforms the IDW. The prediction values resulted from the Ordinary kriging are than processed for contour plots which highlighted the contaminated areas. This research may be helpful for the management of the groundwater agencies and other policy makers for ensuring the pure and safe ground water to all residents of the underlying area.

Contents

[CHAPTER 1: INTRODUCTION 9](#_Toc175849001)

[1.1 Introduction 9](#_Toc175849002)

[1.2 Background of the Study 9](#_Toc175849003)

[1.2.1 Importance of Water Quality 9](#_Toc175849004)

[1.2.2 Overview of Groundwater Resources 9](#_Toc175849005)

[1.3 Significance of Groundwater Quality Monitoring 9](#_Toc175849006)

[1.3.1 Impact on Public Health 9](#_Toc175849007)

[1.3.2 Environmental and Agricultural Relevance 10](#_Toc175849008)

[1.4 Water Quality Parameters Overview 10](#_Toc175849009)

[1.5 Ordinary Kriging: An Overview 10](#_Toc175849010)

[1.6 Research Gap 10](#_Toc175849011)

[1.6.1 Current Methods of Predicting Water Quality 10](#_Toc175849012)

[1.6.2 Limitations of Existing Models 11](#_Toc175849013)

[1.7 Objectives of the Study 11](#_Toc175849014)

[1.8 Research Question 11](#_Toc175849015)

[1.9 Scope and Limitations 11](#_Toc175849016)

[1.10 Conclusion 12](#_Toc175849017)

[CHAPTER 2: REVIEW OF LITERATURE 13](#_Toc175849018)

[2.1 Introduction 13](#_Toc175849019)

[2.2 Overview of Groundwater Quality Parameters 13](#_Toc175849020)

[2.2 Geostatistical Methods in Environmental Science 14](#_Toc175849021)

[2.3 Conclusion 15](#_Toc175849022)

[CHAPTER 3: METHODOLOGY 15](#_Toc175849023)

[3.1 Introduction 15](#_Toc175849024)

[3.2 Study Area and Data Collection 16](#_Toc175849025)

[3.2.1 Description of the study area 16](#_Toc175849026)

[3.2.2 Groundwater Sampling Locations 16](#_Toc175849027)

[3.2.3 Data sources and types of WQPs 17](#_Toc175849028)

[3.3 Exploratory Data Analysis 17](#_Toc175849029)

[3.4 Dendrogram and Factor Analysis 18](#_Toc175849030)

[3.5 Kriging Theory and Concepts 19](#_Toc175849031)

[3.5.1 Data transformation and normalization 19](#_Toc175849032)

[3.6 Spatial Prediction and Mapping 20](#_Toc175849033)

[3.7 Conclusion 20](#_Toc175849034)

[CHAPTER 4: ANALYSIS AND DISCUSSION 20](#_Toc175849035)

[4.1. Introduction 20](#_Toc175849036)

[4.2. Exploratory Data Analysis 21](#_Toc175849037)

[4.3. Multivariate Analysis of WQPs 22](#_Toc175849038)

[4.4 Spatial Analysis of WQPs 27](#_Toc175849039)

[4.5 Conclusion 34](#_Toc175849040)

[CHAPTER 5: CONCLUSION 35](#_Toc175849041)

[REFERENCES 37](#_Toc175849042)

# CHAPTER 1: INTRODUCTION

## 1.1 Introduction

Groundwater is a basic source of fresh and pure water which meets the industrial, agricultural and domestic (Triki et al, 2014). Millions of people in the world use groundwater to fulfill their need. Therefore, ensuring the quality of groundwater is very imperative for the betterment of human health’s and environmental sustainability (Ahmad et al, 2017). The quality of the groundwater is characterized by its chemical parameters. Many techniques have been developed to analyze the WQPs so that groundwater quality could to judged. In this chapter includes the importance of the WQPs, significance of the monitoring of WQPs, research gap, research question, objectives of the study and the scope and limitations of the study have been discussed in detail.

## 1.2 Background of the Study

### 1.2.1 Importance of Water Quality

To evaluate the quality of the groundwater is very imperative for the betterment of the public health and ecological balance. Pure groundwater is also boost up the environmental health. The contaminated and impure groundwater may cause chronic health diseases (WHO, 2011) and disturb the ecosystem and life cycles of many species (Chapman, 1996). Due to these reasons, effective monitoring of the groundwater quality parameters is vital to ensure good human healths and environmental betterment (Smith and Desvousges 1986).

### 1.2.2 Overview of Groundwater Resources

Groundwater have a significant part in universal freshwater supply, providing drinking water for nearly half of the world’s population and supporting agricultural and industrial activities (Foster and Chilton, 2003). The quality of groundwater is influenced by natural processes such as mineral dissolution and human activities including agriculture, industrial discharge, and urbanization (Zektser and Everett, 2004). As groundwater is a crucial resource, understanding its quality and predicting changes through reliable methods like ordinary kriging is essential for sustainable water resource management (Ahmad et al 2015).

## 1.3 Significance of Groundwater Quality Monitoring

### 1.3.1 Impact on Public Health

Monitoring the groundwater quality has great impact on public health because a lot of global population is dependent on groundwater. Contaminants in groundwater such as heavy metals and nitrates may cause serious public health problems. This contamination may can cause serious chronic diseases in children and adults (Smith, Lingas, and Rahman, 2000). Thus, ensuring safe drinking water is very important to prevent waterborne chronic diseases and preserve the public health (WHO, 2017).

### 1.3.2 Environmental and Agricultural Relevance

Monitoring the quality of groundwater has great impact on environmental and agricultural productivity. The impure and contaminated groundwater can damage ecosystems as it changes the chemical balance of rivers, lakes, and other groundwater sources (Morris et al, 2003). In agriculture sector, it is used for irrigation; however, the contaminated groundwater damages the soil health which reduces the crop yields (Vörösmarty et al, 2010). Thus, monitoring the groundwater quality is very important to maintain the ecosystems. Regular evaluation assists in handling the groundwater resources effectively.

## 1.4 Water Quality Parameters Overview

In this section, we have provided a brief introduction of the key WQPs considered in this project report. The first variable is TDS measures the concentration of dissolved substances in water, including salts, minerals, and organic matter. High TDS levels can affect water taste, and suitability for drinking, and agricultural irrigation (Ahmad et al, 2015). The WQPs Calcium and Magnesium both are important for determining water hardness. High concentrations can lead to scaling in pipes and appliances, while low levels might affect the nutritional value of water. **Hardness** represents the combined concentration of calcium and magnesium. Hard water can cause scaling, while soft water may be more corrosive, affecting plumbing and appliances. The WQP **Chloride shows that its h**igh levels can impart a salty taste to water and contribute to corrosion in pipes, affecting the quality of drinking water and the longevity of infrastructure. The WQPs **Potassium and Sodium** are essential electrolytes but, in high concentrations, can lead to health issues such as hypertension. Sulfate is also an important WQPs whose high concentration in drinking water may result many chronic diseases like diarrhea, stomach disorder, laxative problems and food poisoning in in infants and adults (Ahmad et al, 2016). **Fluoride is** beneficial in small amounts for dental health, excessive fluoride can lead to dental destruction, a significant health concern in areas with naturally high fluoride levels (Din et al, 2024; Ahmad et al, 2023).

## 1.5 Ordinary Kriging: An Overview

Ordinary Kriging is a geostatistical prediction technique which consider the spatial correlations of the data to predict the value response variable at unsampled location. This technique was developed by the South African mining engineer Danie Krige in the 1950s and further improved by Georges Matheron in the 1960s (Webster and Oliver 2007). In spatial statistics, Ordinary Kriging has become a basic tool. Ordinary kriging technique assumes of Stationarity which means that the spatial response variable is alike through the study area (Ahmad et al, 2016). Ordinary kriging has the superiority that it consider the spatial correlation using the suitable variogram model (Triki et al. 2014).

## 1.6 Research Gap

### 1.6.1 Current Methods of Predicting Water Quality

To deal the spatial data of groundwater quality parameters; there are many prediction models. First type of models are the deterministic models which are mathematical in nature and use the physical principles to predict the response variable. Second type of models are the Statistical methods like regression modeling and IDW which are unable to consider the spatial correlation while prediction. Third type of models are the Machine learning techniques which are very popular in handling the complex and large data sets. Despite the availability of many prediction techniques, there is always gap of using the accurate prediction technique of environmental variables.

### 1.6.2 Limitations of Existing Models

The available models for predicting the WQPs often face many limitations. Deterministic models need extensive data; therefore, they can be computationally rigorous which make these models less practical. Statistical models such as regression model and IDW interpolations methods are relatively simple but these are unable to consider the spatial autocorrelation that’s why are less accurate in spatial prediction. Thus, Ordinary Kriging have superiority as it incorporates the spatial correlation. This technique has also limitation of its basic assumption of stationarity and based on the accurate selection of variogram model for valid prediction. These research gaps identify the requirement of more robust and adaptable prediction techniques which can improve the accuracy and reliability prediction of WQPs in diverse settings.

## 1.7 Objectives of the Study

The primary objective of this research is to explore the key WQPs and to predict the spatial distribution of these parameters at unsampled locations. The secondary objective of this research is to compare the ordinary kriging with inverse distance weighting technique.

## 1.8 Research Question

The research question of this project report is to evaluate that how can multivariate statistical and geostatistical techniques jointly explore and predict the WQPs in groundwater, and how does ordinary kriging outperform as compared to Inverse Distance Weighting (IDW) method in recognizing the contaminated areas and evaluating the groundwater quality in different spatial locations?

## 1.9 Scope and Limitations

As far as the scope of this research is concerned, it focuses on the spatial interpolation of the WQPs at unsampled locations using the ordinary kriging. Mainly following WQPs are included in the study: TDS, Calcium, Magnesium, Hardness, Chloride, Potassium, Sodium, Sulfate, and Fluoride. The scope includes:

1. **Geographical Scope:** This research is conducted for a specific area due to the availability of the ground water data.
2. **Temporal Scope:** The groundwater data is for a specific period of time; therefore, the temporal changes have not been considered.
3. **Methodological Scope:** Ordinary Kriging have been considered for the spatial prediction of the unsampled locations. The performance of Ordinary kriging has been compared with IDW method. Thus, this research also has methodological scope.
4. **Practical Application:** The findings of this study may contribute to better groundwater management and may be helpful for water management agencies and policy makers for ensuring safe and sustainable water resources.

The limitations include:

1. **Data Availability:** The unevenly distributed spatial data may the reliability of the prediction of unsampled locations using Ordinary kriging because quality of the data has great impact in spatial prediction.
2. **Model Assumptions:** The geostatistical Ordinary Kriging technique have stationarity assumption which means that the spatial data do not change across the study region. If this assumption is violated, the predictions of the Ordinary kriging may become less accurate.
3. **Variogram Model:** The prediction efficiency of Ordinary Kriging depends upon the choice of the variogram model as it symbolizes the spatial autocorrelation of the data. An incorrect fitted variogram can lead to unreliable predictions.
4. **Environmental Variability:** This research study does not include the temporal changes in water quality data as it only depends on the spatial locations. Variations in environmental conditions

## 1.10 Conclusion

This chapter has laid the basis for predicting the WQPs using Ordinary Kriging through highlighting the importance of key WQPs, groundwater and the application of the adopted technique in environmental science. The objectives and research questions along with research gap has been narrated in detail. The scope of the research includes a detailed analysis within a specific geographic area. The limitations related to the availability of data, assumptions of ordinary kriging and environmental variability are also discussed. Through these aspects, the study aims to provide valuable insights for groundwater management and policy-making, contributing to more effective and informed decision-making regarding water quality.

# CHAPTER 2: REVIEW OF LITERATURE

## 2.1 Introduction

This chapter includes the literature on prediction techniques on groundwater quality parameters and spatial interpolation techniques which has significant importance regarding the environmental sustainability and public health. The literature review inspects few studies that have used Ordinary Kriging to predict WQPs at unsampled locations. The review also explores advancements in geostatistical modeling and the comparative effectiveness of different interpolation methods. By reviewing the past research studies, this chapter identifies gaps in latest techniques and highlights the potential for improved accuracy and applicability in predicting groundwater quality across diverse spatial locations.

## 2.2 Overview of Groundwater Quality Parameters

In this section, we have overviewed literature on WQPs. It is known fact that millions of people in the world use groundwater to fulfill their daily needs. It is a basic source of fresh and pure water which meets the industrial, agricultural and domestic (Triki et al, 2014). Therefore, ensuring the quality of groundwater is very imperative for the betterment of human health’s and environmental sustainability (Ahmad et al, 2017). The quality of the groundwater is characterized by its chemical parameters. Many techniques have been developed to analyze the WQPs so that groundwater quality could to judged. Exploring the groundwater quality is a very important aspect of environmental health which influence both human well-being as well as ecological balance (Idrees et al, 2020; Chapman, 1996; Smith and Desvousges 1986).

The quality of groundwater is influenced by natural processes such as mineral dissolution and human activities including agriculture, industrial discharge, and urbanization (Zektser and Everett, 2004). As groundwater is a crucial resource, understanding its quality and predicting changes through reliable methods like ordinary kriging is essential for sustainable water resource management (Ahmad et al 2015). Monitoring the groundwater quality has great impact on public health because a lot of global population is dependent on groundwater. Contaminants in groundwater such as heavy metals and nitrates may cause serious public health problems. This contamination may can cause serious chronic diseases in children and adults (Smith, Lingas, and Rahman, 2000). Thus, ensuring safe drinking water is very important to prevent waterborne chronic diseases and preserve the public health (World Health Organization, 2017). Monitoring the quality of groundwater has great impact on environmental and agricultural productivity. The impure and contaminated groundwater can damage ecosystems as it changes the chemical balance of rivers, lakes, and other groundwater sources (Morris et al., 2003). In agriculture sector, it is used for irrigation; however, the contaminated groundwater damages the soil health which reduces the crop yields (Vörösmarty et al., 2010). Thus, monitoring the groundwater quality is very important to maintain the ecosystems. Regular evaluation assists in handling the groundwater resources effectively. In hydrological researches, water samples are collected and analyzed for different WQPs and compared with the permissible limits. Generally, much more water samples are collected as compared to the samples that are statistically analyzed (Li et al 2019). The reason behind this gap is the challenges faced during the analysis. Moreover, the analysis of groundwater quality has gained much importance in order to assess the factors making groundwater contaminated.

## 2.2 Geostatistical Methods in Environmental Science

Following studies collectively provide a robust foundation for understanding and application of groundwater geostatistical prediction techniques. They cover theoretical aspects, practical applications, and comparative analyses, offering a well-rounded perspective on the strengths and limitations of this geostatistical method. Predicting water quality in groundwater systems is a complex task due to the numerous factors that influence it, including natural geological processes and human activities. Many techniques have been developed to predict the water quality data, yet there exist challenges to deal the spatial variability (Idrees et al 2020).

Goovaerts (1997) stated in his book that ordinary kriging is very effective in providing reliable spatial estimates and highlight its advantages over other prediction methods of spatial response variables. This comparison is essential for understanding the strengths and limitations of ordinary kriging in groundwater studies. McGrath et al. (2004) considered geostatistical methods, including ordinary kriging, to predict the soil contamination. Journal and Huijbregts' (1989) provided detailed introduction of kriging prediction methods along with its applications in mining. Rouhani (1996) narrated the concept of Kriging in the context of geostatistical Estimation for its application in environmental and geotechnical data. He argued that Kriging is widely used to model applied to predict the spatial data as it also deal the variations and spatial correlations. The paper provides an overview of these techniques and their application in analyzing environmental variables.

Ahmad et al. (2016) used univariate data of sulfate concentration in the groundwater of Jampur Pakistan to predict its spatial distribution at unsampled locations. In this research ordinary kriging has been compared with Bayesian kriging and both techniques are compared on the basis of validation statistics. Parameters of the variogram (sill, range and nugget) were estimated using OLS and WLS techniques and plugged into kriging model. Prediction results were mapped and results showed the superiority of the Bayesian kriging due to its ability of considering the prior knowledge. Idrees et al (2020) conducted a hydrological research in district Okara-Pakistan. They applied geostatistical approaches on the multivariate data of 14 WQPs. They used cross-variography which is a distance-based correlation matrix of the variables. Moreover, they also applied dimension reduction techniques along with geostatistical techniques and mapped the prediction results to predict the unsampled locations.

Ahmad et al, (2017) considered the multivariate hydrological data of Faisalabad city of Pakistan and used both multivariate statistical and geostatistical techniques to analyze this data. Zhou et al, (2007) also applied the multivariate statistical methods to deal the WQPs data of Hong Kong. Nas B (2009) applied the ordinary kriging techniques to assess the distribution of chemical WQPs. He also mapped the results of the prediction to evaluate the most contaminated areas. Triki et al, (2014) compared the prediction performance of multivariate cokriging with ordinary kriging and IDW methods for the groundwater quality data. Results highlighted that cokriging outperformed as compared to other prediction techniques.

Subyani and Al-Dakheel (2009) used the multivariate geostatistical techniques to predict the data of mean annual and seasonal precipitation in Saudi Arabia. The study focused on analyzing rainfall patterns using geostatistical techniques, providing valuable information about the spatial and temporal distribution of rainfall in the region.

While using the geostatistical approaches, the choice of variogram model is very important as best variogram model capture the correlation of the spatial data in a best way (Ribeiro and Diggle 2006). Pebesma and Graler (2015) introduced the variography techniques for the Spatio-Temporal data focusing on the analysis of spatial and temporal variations in geospatial datasets. Triki et al. (2014) suggested to use cross-variography as it also captuire the distance based association among the spatial response variables. Most commonly used variogram models for the univariate spatial response variable are spherical, Matern and Gaussian

## 2.3 Conclusion

This chapter has provided a detailed literature review on the importance of groundwater with respect to domestic requirement, industrial need and agricultural application. It also provided literature review on the importance of WQPs along with their prediction techniques. Moreover, many studies have been discussed where geostatistical techniques have been used in hydrological context. Some studied focused on the jointly use of multivariate statistical techniques and spatial prediction methods.

# CHAPTER 3: METHODOLOGY

## 3.1 Introduction

This research investigates the prediction of groundwater quality parameters using Ordinary Kriging, a geostatistical interpolation technique. The methodology begins with a comprehensive description of the study area, including groundwater sampling locations and the water quality parameters analyzed. Data collection involves field sampling and laboratory analysis, ensuring the reliability and accuracy of the data. Following data collection, preprocessing steps address data cleaning, validation, and normalization to prepare the dataset for analysis. Exploratory data analysis is conducted to understand the statistical characteristics and spatial distribution of the parameters. This includes correlation analysis to identify relationships between variables. The core of the methodology focuses on applying Ordinary Kriging. This involves calculating and modeling the variogram to understand spatial variability and selecting an appropriate variogram model. Ordinary Kriging equations are then used to predict groundwater quality at unsampled locations. Model performance is evaluated through cross-validation statistics to assess accuracy. Spatial predictions are visualized using maps to interpret the distribution of water quality parameters. The chapter concludes with a discussion on the methodology’s effectiveness, limitations, and implications for groundwater management.

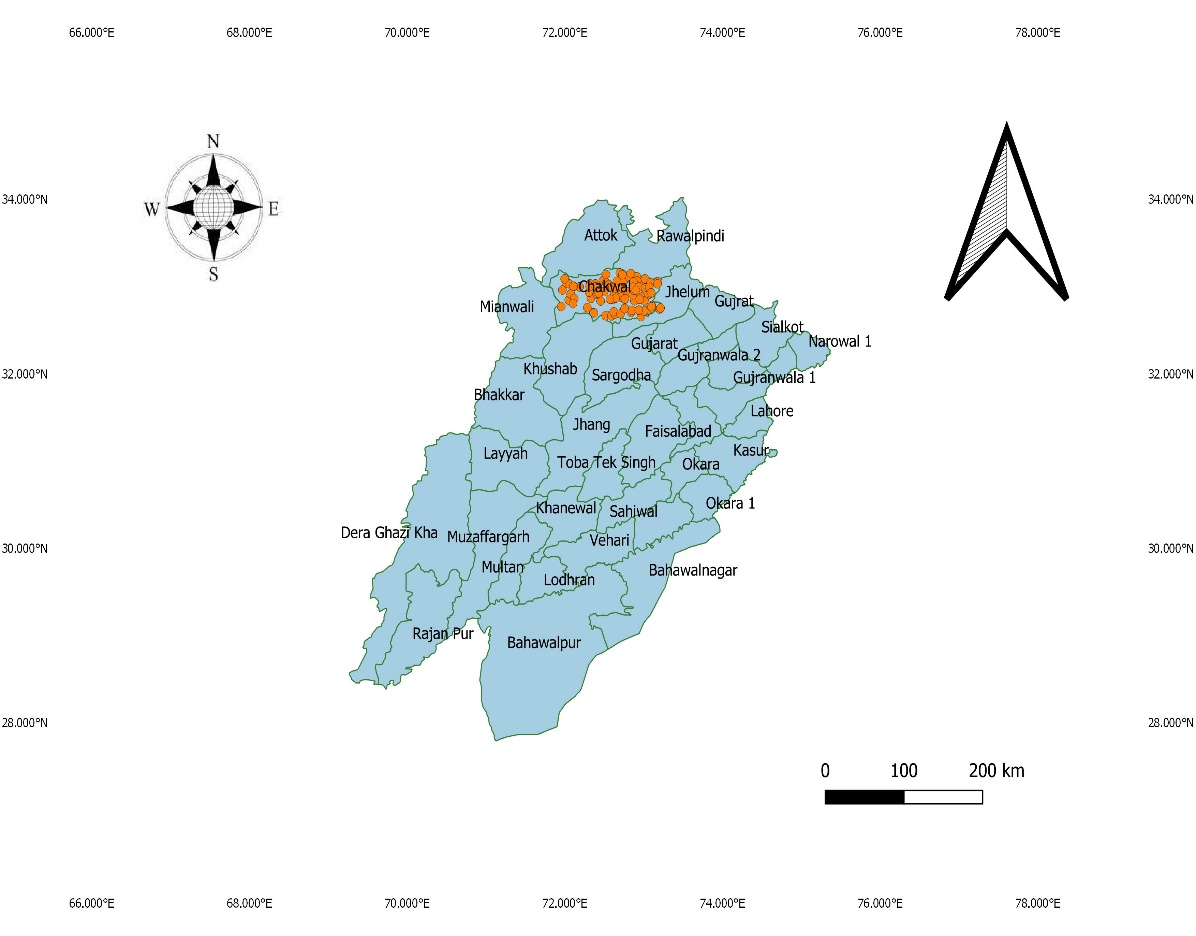
## 3.2 Study Area and Data Collection

### 3.2.1 Description of the study area

To study the spatial statistical techniques, we have used the spatial data of 9 ground water quality data from Chakwal district. It is located in Pakistan's Punjab province on the Pothohar Plateau, the Chakwal district boasts a diverse terrain ranging from flat ground to undulating landscapes with exposed rocks and complex gullies. With geographical coordinates between latitude 33° 01'21"N and longitude 72° 45'49"E, the district spans 1,652,441 acres and is home to an estimated population of 1.49 million, as per the Provisional Census Report (2017). Climatically, Chakwal lies within the semi-arid zone, experiencing the coldest temperatures in December and the warmest in June, with summer temperatures ranging from 15 to 40 degrees Celsius and winter temperatures between 4 to 25 degrees Celsius. Annual rainfall averages between 558 to 625 millimeters, with 70% occurring during the monsoon season from June to September. The district boasts various tourist destinations and captivating landscapes like Uchali, Khabar, and Kallar Kahar. However, agriculture heavily relies on rain and faces vulnerability to drought due to limited irrigation resources, with only 8% of farmland aerated through tube wells, canals, and wells. Wheat is the primary crop cultivated, alongside maize, peanuts, garbanzo beans, durra, millet, and soya beans in smaller areas. The region confronts water scarcity and soil erosion, posing threats to crop output and necessitating the development of groundwater resources to meet agricultural and domestic water needs. Effective water management strategies are essential for sustaining the agricultural economy and addressing the challenges posed by soil erosion and water scarcity.

### 3.2.2 Groundwater Sampling Locations

The groundwater sampling locations are displayed in Figure 3.1.



**Figure 3.1**: Spatial distribution of the ground water data of nine water quality parameters

### 3.2.3 Data sources and types of WQPs

Data have been retrieved from the Pakistan Council of Research in Water Resources (PCRWR) which is a national organization in Pakistan that conduct researches on ground water and provides recommendations on water resources management and quality. Data is based on 106 spatial locations of district Chakwal comprising on 9 WQPs (TDS, Calcium, Magnesium, Hardness, Chloride, Potassium, Sodium, Sulfate and Fluoride).

## 3.3 Exploratory Data Analysis

For the exploratory data analysis, we have to compute statistical summary of the spatial data where minimum values, maximum values, mean values and coefficient of variations of all WQPs will be computed. The formulas for mean values and CV are as follows:

,

and

where SD represents the standard deviations of the particular response variable and represents the mean value of the spatial response variable. CV is being calculated as a measure of the dispersion. Beside these, we have also required the pairwise correlations of all spatially distributed WQPs. The sample correlation coefficient between ith and jth spatial response variables is illustrated as:

=

The correlation matrix for a sample data corresponds to covariance matrix of same data with correlation as a substitute of covariance’s.

R= =

where:

R represents the multivariate correlation matrix which is a square matrix that displays the pairwise correlations between variables in a dataset. represents the correlation coefficient between and variables such that

## 3.4 Dendrogram and Factor Analysis

A dendrogram which is a visual display of the results of hierarchical cluster analysis has also been applied in this study. Clusters are constructed on the basis of similarity or dissimilarity, common distance measures used is include Euclidean distance. If and are two data points in n-dimensional space, the Euclidean distance between these two data points is represented by as

Like cluster analysis and principal analysis, factor analysis is also a multivariate technique in which variables are grouped on the basis of similarity. Inspite of studying many variables, in factor analysis, few factors are constructed which are linear combination of the original variables. (Ahmad et al, 2017).

## 3.5 Kriging Theory and Concepts

It is a known fact that an efficient prediction method assists in reducing the prediction error (Hussain et al. 2015; Mubarak et al. 2016; Gundogdu and Guney 2007). Kriging is a geostatistical prediction technique method which predicts spatially distributed variables by considering the spatial distance and the degree of variation between given data points. Unlike regression, kriging considers the spatial autocorrelation in the data. Due to these characteristics, Kriging is known as the best linear unbiased prediction method (Isaaks and Srivastava 1989). The predicted values resulted from the ordinary kriging at unsampled locations using ordinary Kriging is a linear combination of sampled locations. Therefore, to determine the coefficient of the linear combination, auto covariance plays a significant role (Webster and Oliver 2007). In Ordinary Kriging, mean is assumed constant in the local neighborhood of each estimation point. Ordinary kriging is a used for spatial prediction of the response variable at unsampled location using the data of sampled locations. The main theme is to estimate the value of a response variable at a given point by considering the weighted average of the given values whereas the weights are computed through the spatial correlation of the data.

The equation of the Ordinary Kriging is expressed in simplest form as:

where:

represents the estimated value of response variable at location

represents the given known values of response variable at location

represents the weights applied to each known value of with a constraint that

Before implementing the ordinary kriging, suitable variogram model is selected in such a way that it capture the maximum correlation in the spatial data.

### 3.5.1 Data transformation and normalization

In order to use kriging approaches, normality of the variables is first assessed. In case of asymmetrical data, then several transformations are available to transform the data. In this research, we have considered BOX-Cox transformation (Box and Cox, 1964).

where is the original spatial response variable and is transformed variable. The transformation parameter is represented by (lambda). It is important to note that Box-Cox transformation is only applicable on the positive response variable.

To carry out the analysis, we’ll use R software (Team R.C 2000) and Minitab (Alin, A., 2010).

## 3.6 Spatial Prediction and Mapping

The prediction values resulted from the ordinary kriging after incorporating suitable variogram model are used in mapping the contour plots (Ahmad et al. 2017). These plots are the multivariate graphical representations used to visualize two or more variables concurrently. For the WQPs, these are used to predict the unsampled locations which helps the policy makers for decision making.

## 3.7 Conclusion

This chapter has provided the insights about description and the geography of the study area, visual display of groundwater sampling locations, data sources and types of WQPs. The data analysis techniques (Exploratory data analysis, multivariate techniques, ordinary kriging approaches along with spatial maps), used in the report have been narrated in detail.

# CHAPTER 4: ANALYSIS AND DISCUSSION

## 4.1. Introduction

In this chapter, spatial data of 9 WQPs have been processed and results are presented in detail. Initially, data have been treated for exploratory data analysis where measures of central tendency are computed on the basis of mean values of each WQP whereas measures of variation are illustrated on the basis of coefficient of variation (CV) which is a relative measure of dispersion computed as the ratio of standard deviation to mean. The measures of shape are discussed using histogram and box & whisker diagram. At next stage, data have been processed for multivariate analysis where multivariate correlation matrix is computed to evaluate pairwise strength of association between WQPs. On the basis of scree plot, we computed three factors and selected most significant WQPs. Dendrogram also suggested same WQPs as significant and linearly correlated.

At the final stage, we have processed these WQPs for the spatial analysis using ordinary kriging which is one of the most suitable spatial prediction technique. Kriging provided detailed maps of key WQPs. The kriging model exhibited high accuracy, with prediction errors within acceptable ranges. The spatial distribution maps revealed significant patterns and potential contamination sources, highlighting areas requiring further investigation and remediation. Comparative analysis with other interpolation methods confirmed the superior performance of ordinary kriging in capturing spatial variability and providing reliable predictions for WQPs.

## 4.2. Exploratory Data Analysis

In this section, the WQPS are processed for exploratory data analysis where minimum value, maximum value, mean value, CV are computed along with the permissible limits of WHO of WQPs of Groundwater Samples. Moreover, we also evaluated the shape of the using the histogram. Beside these, we draw Box and whisker diagram to present five number summary (Minimum value, 1st quartile, median, 3rd quartile and maximum value) and to evaluate the outliers present among the values of in each WQPs. In Table 4.1, we have observed that the minimum and maximum values of TDS (mg/l) are 192 mg/l and 2529.00 mg/l respectively whereas mean value is 731.1792 mg/l. The CV is 179.1795%, showing a relatively high variability in TDS values. The permissible limit of WHO for TDS is ≤1,000 mg/l, indicating that all samples in the dataset comply with this limit. The minimum and maximum values of calcium concentration are 8.00 mg/l and 154.00 mg/l respectively whereas the mean calcium concentration is 53.0566 mg/l, representing the average calcium level of the samples. The coefficient of variation (CV) is 162.7475%, indicating relatively high variability in calcium concentrations. The permissible limit of WHO for calcium is ≤75 mg/l, indicating that all samples in the dataset comply with this limit. The minimum and maximum magnesium concentration observed is data are 7.20 mg/l and 315.00 mg/l respectively whereas the mean magnesium concentration is 40.0160 mg/l. The coefficient of variation (CV) is 108.2571%, indicating relatively high variability in magnesium concentrations. The permissible limit of WHO for magnesium is ≤150 mg/l, indicating that all samples in the dataset comply with this limit.

**Table 4.1:** Descriptive Statistics (Minimum value, maximum value, mean value, coefficient of variation (CV) and permissible limits of WHO of Water Quality Parameters of Groundwater Samples

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables** | **Min. Value** | **Max. Value** | **Mean** | **CV** | **WHO limits** |
| TDS (mg/l) | 192 | 2529 | 731.18 | 179.18 | ≤1,000 |
| Calcium (mg/l) | 8 | 154 | 53.06 | 162.75 | ≤75 |
| Magnesium (mg/l) | 7.2 | 315 | 40.01 | 108.26 | ≤150 |
| Hardness (mg/l) | 3.5 | 850 | 278.50 | 200.54 | ≤60 |
| Chloride (mg/l) | 7 | 359 | 88.54 | 117.54 | ≤250 |
| Potassium (mg/l) | 0.4 | 18 | 3.50 | 111.35 | ≤12 |
| Sodium (mg/l) | 10 | 700 | 148.79 | 117.78 | ≤200 |
| Sulfate (mg/l) | 6 | 630 | 101.40 | 96.91 | ≤250 |
| Fluoride (mg/l) | .07 | 3.16 | 0.62 | 95.06 | ≤1.5 |

The minimum and maximum value of hardness observed is 3.50 mg/l and 850.00 mg/l whereas the mean hardness is 278.5047 mg/l. The coefficient of variation (CV) is 200.5401%, indicating relatively high variability in hardness concentrations. The permissible limit of WHO for hardness is ≤60 mg/l, indicating that all samples in the dataset exceed this limit. The minimum and maximum values of chloride concentration are 7.00 mg/l and 359.00 mg/l respectively whereas the mean chloride concentration is 88.5404 mg/l, representing the average chloride level of the samples. The coefficient of variation (CV) is 117.5372%, indicating relatively high variability in chloride concentrations. The permissible limit of WHO for chloride is ≤250 mg/l, indicating that all samples in the dataset comply with this limit. The minimum and maximum potassium concentration are 0.40 mg/l and 18.00 mg/l respectively whereas the mean potassium concentration is 3.4972 mg/l. The coefficient of variation (CV) is 111.346%, indicating relatively high variability in potassium concentrations. The permissible limit of WHO for potassium is ≤12 mg/l, indicating that all samples in the dataset comply with this limit. In case of the WQP sodium, the minimum and maximum sodium concentration levels are 10.00 mg/l and 700.00 mg/l respectively whereas the mean sodium concentration is 148.7925 mg/l. The coefficient of variation (CV) is 117.7833%, indicating relatively high variability in sodium concentrations. The permissible limit of WHO for sodium is ≤200 mg/l, indicating that all samples in the dataset comply with this limit.

For the WQP Sulfate, we observed that the minimum and maximum sulfate concentration are 6.00 mg/l and 630.00 mg/l respectively whereas the mean sulfate concentration is 101.3962 mg/l. The coefficient of variation (CV) is 96.90961%, indicating relatively high variability in sulfate concentrations. The permissible limit of WHO for sulfate is ≤250 mg/l, indicating that all samples in the dataset comply with this limit. In case of the WQP, Fluoride we have observed the minimum and maximum concentration levels as 0.07 mg/l and 3.16 mg/l respectively whereas the mean fluoride concentration is 0.6160 mg/l. The coefficient of variation (CV) is 95.06173%, indicating relatively high variability in fluoride concentrations. The permissible limit of WHO for fluoride is ≤1.5 mg/l, indicating that all samples in the dataset comply with this limit. Overall, these descriptive statistics provide valuable insights into the distribution, variability, and compliance of the WQPs with the permissible limits recommended by the WHO. The interpretation can help identify variables with significant variations and those that may require attention or further investigation to meet the WHO's water quality standards.

## 4.3. Multivariate Analysis of WQPs

In this section, all WQPs are processed for the multivariate analysis. Initially, we applied Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphercity to determine that either existing data is suitable for multivariate Factor analysis or not by checking if there are enough correlations between WQPs. The null and alternative hypothesis of these tests is as follows:

: All WQPs are uncorrelated

: All WQPs are correlated

Table 4.2. Results of Kaiser-Meyer-Olkin (KMO) and Bartlett's Test

|  |  |  |
| --- | --- | --- |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | .603 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 2277.227 |
| df | 91 |
| Sig. | .000 |

The KMO Measure of sampling adequacy is reported to be 0.603 (see, table 4.2) where its values ranges from 0 to 1, and a higher value closer to 1 indicates that the dataset is more suitable for factor analysis. Thus our estimated value suggested that the data may be acceptable for factor analysis. Since, the null hypothesis in the Bartlett's test of sphericity evaluates that there are no correlations among the WQPs; therefore, the significant result (p-value < 0.05) suggested that the WQPs are correlated between each other’s and thus factor analysis may be appropriate.

To proceed the multivariate factor analysis, initially, we computed the Pearson correlation coefficients among all WQPs graphically and highly correlated WQPs are identified (see Figure 4.10).



**Figure 4.1**: Graphical display of multivariate correlation matrix of nine WQPs (upper panel) whereas graphical display of insignificant pairwise correlated WQPs (lower panel)

In Figure 4.1, we have observed that the highest pairwise correlated WQPs are TDS-Sodium (0.89), TDS- Chloride (0.88), Chloride-Sodium (0.83), Sulfate-Sodium (0.76), TDS-Sulfate (0.82), Chloride-Sulfate (0.69), Calcium-Hardness (0.79). Since correlation coefficients ranges from -1 to +1, thus above seven pairs have highest positive correlations. Later on, we designed the dendrogram to represent the clusters among the



**Figure 4.2:** Dendrogram representing the results of hierarchical clustering analysis based on the similarity between different water parameters.

Afterwards, we verified the results of the correlation matrix by using cluster analysis as shown in the dendrogram (see, **Figure 4.2**). The X-axis shows various water parameters, while the Y-axis represents the level of similarity, ranging from 36.76 to 100. The first cluster of WQPs suggested TDS (mg/l), Sodium (mg/l), Chloride (mg/l), Sulfate (mg/l) and Potassium (mg/l). On the basis of scree plot (see, **Figure 4.3**), factor analysis is performed with three factors and we observed that first factors



**Figure 4.3:** Scree plot to evaluate the number of factors required in factor analysis

**Table 4.3:** Factor Score Coefficients of nine WQPs along with communality

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **WQPs** | **Factor 1** | **Factor 2** | **Factor 3** | **Communality** |
| TDS | **0.972** | 0.113 | -0.006 | 0.957 |
| Chloride | **0.925** | -0.054 | 0.032 | 0.859 |
| Sodium | **0.913** | -0.297 | -0.033 | 0.923 |
| Sulfate | **0.874** | 0.085 | -0.025 | 0.772 |
| Hardness | 0.259 | 0.895 | 0.040 | 0.870 |
| Calcium | -0.007 | 0.888 | -0.205 | 0.831 |
| Fluoride | -0.148 | 0.591 | -0.479 | 0.601 |
| Magnesium | 0.150 | 0.447 | 0.657 | 0.653 |
| Potassium | 0.283 | -0.242 | -0.541 | 0.432 |
| **Variance** | **3.588** | **2.308** | **1.009** | **6.898** |
| **% Var** | **0.399** | **0.257** | **0.111** | **0.766** |

Table 4.3 represents the factor score coefficients for three extracted factors (Factor1, Factor2, and Factor3) obtained from a factor analysis. Each row represents a water parameter, and the corresponding values in each column represent the coefficients or loadings of that parameter on the respective factor. Since factor analysis is a statistical technique used to identify underlying factors that explain the correlations among a set of observed variables. The factors are latent variables that capture the common variance among the observed variables and help in reducing the dimensionality of the data. Based on the factor score coefficients, the extracted factors are given such as: In Factor1, positive loadings are TDS (0.972), Chloride (0.925), Sodium (0.913), Sulfate (0.874), Hardness (0.259), Magnesium (0.150) and Potassium (0.283) whereas Negative loadings: Calcium (-0.007), Fluoride (-0.148). In Factor2, positive loadings include Hardness (0.895), Calcium (0.888), Fluoride (0.591), Magnesium (0.447), TDS (0.113), Sulfate (0.085) whereas Negative loadings includes Chloride (-0.054), Sodium (-0.297), Potassium (-0.242). The Positive loadings of 3rd factor are including WQPs such as Magnesium (0.657), Hardness (0.040), Chloride (0.032) whereas for Negative loadings: TDS (-0.006), Sodium (-0.033), Sulfate (-0.025), Calcium (-0.205), Fluoride (-0.479), Potassium (-0.541).

Thus factor analysis determined that in Factor 1, there is positive loadings for TDS, Chloride, Sodium, Sulfate, Hardness, Magnesium and Potassium). Conversely, it has negative loadings for two WQPs such as Calcium and Chloride. In Factor 2, we have observed that Hardness, Calcium, Fluoride, Magnesium, TDS and Sulfate have positive loadings whereas Chloride, Sodium and Potassium have negative loadings. The third factor showed that Magnesium, Hardness and Chloride have positive loadings whereas TDS, Sodium, Sulfate, Calcium, Fluoride and Potassium have negative loadings. Moreover, we have observed that overall these three factors represented the 76.6% of the variability of the data where factor 1 contributed 39.9% alone. Thus factor 1 is most important factor among these three factors and we selected first four WQPs: TDS (0.972), Chloride (0.925), Sodium (0.913), Sulfate (0.874) as these have high and positive loadings.

## 4.4 Spatial Analysis of WQPs

To carry out the spatial analysis, first step is to evaluate the shape of each WQPs where we draw the histogram and box & whisker diagram of four selected parameters: TDS (mg/l), Chloride (mg/l), Sodium (mg/l) and Sulfate (mg/l). As shown in Figures 4.4-4.7, we have observed that all these four WQPs have asymmetrical distribution. In ordinary kriging, the normality of the spatially distributed variables is not a strict requirement but it is often desirable. The basic assumption for the spatial data in ordinary kriging is that data must be spatially homogeneous and that and that the underlying spatial structure can be modeled using a variogram. Thus, Variogram is a tool used to quantify the spatial variability and correlation structure in response variable.



**Figure 4.4**: Histogram (left panel) and Box & Whisker Diagram (right panel) of TDS (mg/l)



**Figure 4.5**: Histogram (left panel) and Box & Whisker Diagram (right panel) of Chloride (mg/l)



**Figure 4.6**: Histogram (left panel) and Box & Whisker Diagram (right panel) of Sodium (mg/l)



**Figure 4.7**: Histogram (left panel) and Box & Whisker Diagram (right panel) of Sulfate (mg/l)

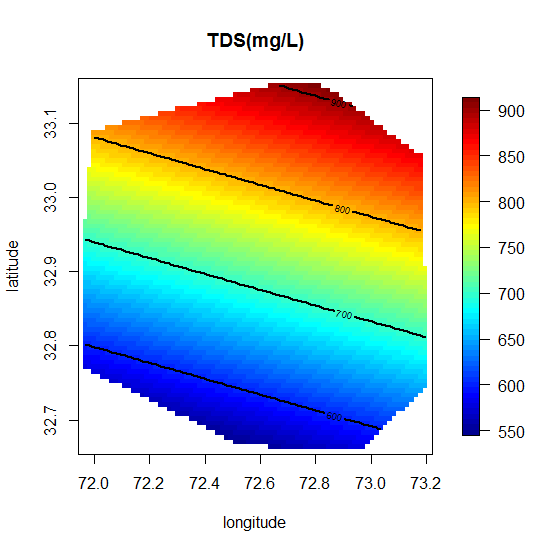
Variogram reflects our understanding about the geometry and continuity of response variable and determines that how data values are related with distances. In ordinary kriging, we have considered three variogram models: Spherical model, exponential model and Gaussian variagram models and fitted all these three variogram models on the response variables of WQPs. Results showed that the spherical model best fitted the data (see, table )which is frequently used in environmental sciences especially for the spatially correlated water data.

Table 4.4: Estimated parameters of the variogram model for all WQPs (TDS, Chloride, Sodium and Sulfate)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **WQPs** | **Fitted model** | **Nugget ()** | **Sill ()** | **Range ()** | **R2** |
| TDS | Spherical | 0.164 | 0.565 | 947 | 0.73 |
| Sulfate | Spherical | 0.067 | 0.641 | 1147 | 0.69 |
| Sodium | Spherical | 0.139 | 0.348 | 1475 | 0.67 |
| Chloride | Gaussian | 0.014 | 0.397 | 1344 | 0.79 |

Later on, we used the estimated parameters of the variogram models to fit the ordinary kriging and generated the contour maps (see, Figure 4.8-4.11). In Figure 4.8, spatial prediction map of WQP TDS, the contour plot illustrates the spatial distribution of Total Dissolved Solids (TDS) in the study area, revealing a high-risk zone with longitude ranging from 72.0° to 73.2° and latitude from 33.0° to 33.1°.

The closely spaced contour lines indicate significant variations in TDS concentrations within this region. To assess water quality against the World Health Organization's (WHO) permissible limit of TDS being less than 1000 mg/L for drinking water, it is evident that the TDS levels in this high-risk area exceed the recommended limit. This indicates potential concerns regarding the suitability of groundwater for human consumption, as elevated TDS levels may be indicative of pollution or high mineral content. Complying with the WHO permissible limit is crucial to ensure the safety and well-being of the local population, necessitating immediate attention, appropriate actions, and water quality management strategies to address the high TDS levels and provide access to safe drinking water sources.



**Figure 4.8:** Spatial Prediction map of Water Quality Parameter:TDS (mg/l) using Ordinary Kriging



**Figure 4.9:** Spatial Prediction map of Water Quality Parameter: Chloride (mg/l) using Ordinary Kriging

In Figure 4.9, the spatial prediction map of WQP Chloride is constructed where the contour plot displays the spatial distribution of chloride levels in the study area, indicating a high-risk zone with longitude ranging from 72.0° to 73.2° and latitude from 33.0° to 33.1°. The closely spaced contour lines suggest substantial variation in chloride concentrations within this region. To assess water quality against the World Health Organization's (WHO) permissible limit of chloride being less than 250 mg/L for drinking water, it is evident that the chloride levels in this high-risk area exceed the recommended limit. This raises concerns about the suitability of groundwater in the region for human consumption, as elevated chloride levels can impact taste and potentially contribute to health risks if consumed in excess.



**Figure 4.10:** Spatial Prediction map of Water Quality Parameter: Sodium (mg/l) using Ordinary Kriging



**Figure 4.11:** Spatial Prediction map of Water Quality Parameter: Sulfate (mg/l) using Ordinary Kriging

In Figure 4.10, the spatial prediction Map of WQP Sodium has been constructed, the contour plot illustrates the spatial distribution of sodium levels in the study area, revealing a high-risk zone with longitude ranging from 72.0° to 73.2° and latitude from 33.0° to 33.1°. The closely spaced contour lines indicate significant variations in sodium concentrations within this region. To assess water quality against the World Health Organization's (WHO) permissible limit of sodium being less than 200 mg/L for drinking water, it is evident that the sodium levels in this high-risk area exceed the recommended limit. In Figure 4.11, the spatial prediction Map of WQP Sulfate, the contour plot presents the spatial distribution of sulfate levels in the study area, indicating a high-risk zone with longitude ranging from 72.0° to 72.2° and latitude from 32.7° to 33.1°. The WHO permissible limit of sulfate being less than 250 mg/L for drinking water, but it is evident that the sulfate levels in this high-risk area exceed the recommended limit. This raises concerns about the suitability of groundwater in the region for human consumption, as elevated sulfate levels can have potential health implications, particularly when ingested in large amounts.

Table 4.5: Comparison of Ordinary Kriging and IDW model for all WQPs (TDS, Chloride, Sodium and Sulfate)

|  |  |  |
| --- | --- | --- |
| **WQPs** | **Fitted model** | **RMSPE** |
| TDS | Ordinary Kriging | 1.568 |
| IDW | 2.564 |
| Sulfate | Ordinary Kriging | 3.254 |
| IDW | 4.686 |
| Sodium | Ordinary Kriging | 3.254 |
| IDW | 3.891 |
| Chloride | Ordinary Kriging | 2.235 |
| IDW | 3.015 |

Table 4.5 addressed that, for all WQPs, the ordinary kriging outperforms the IDW in terms of least RMSPE. Thus, in this research project the use of ordinary kriging is better as compared to IDW method.

## 4.5 Conclusion

This chapter provided the detailed analysis of the nine WQP of groundwater data. To carryout the analysis, we have used exploratory data analysis, multivariate analysis (correlation matrix, dendrogram, factor analysis). In spatial analysis, we have evaluated the shape of the selected significant WQPs and used the Box-Cox transformation and later on, estimated the parameters of the variogram model and plugged in these estimated parameters in the ordinary kriging equation and mapped the prediction results. At the end, we compared the results of Ordinary kriging with the IDW interpolation technique and reported the findings. Moreover, our results also adressed the research question and the objectives of the study.

# CHAPTER 5: CONCLUSION

In this report, a multivariate data of groundwater quality parameters of 106 samples and nine WQPs are statistically analyzed. The main purpose of this research is to assess the efficacy of multivariate statistical and geostatistical techniques in the joint exploration and prediction of WQPs, as well as the relative performance of ordinary kriging and the Inverse Distance Weighting (IDW) method in identifying contaminated areas and assessing groundwater quality in various spatial locations. Initially, data has been processed for the descriptive analysis where findings showed that the mean value of TDS is 731.1792 mg/l whereas the permissible limit of WHO for TDS is ≤1,000 mg/l, the mean calcium concentration is 53.0566 mg/l, representing the average calcium level of the samples whereas the permissible limit of WHO for calcium is ≤75 mg/l, indicating that all samples in the dataset comply with this limit. The mean magnesium concentration is 40.0160 mg/l whereas the permissible limit of WHO for magnesium is ≤150 mg/l. The mean hardness is 278.5047 mg/l whereas the permissible limit of WHO for hardness is ≤60 mg/l, indicating that all samples in the dataset exceed this limit. The mean chloride concentration is 88.5404 mg/l, whereas the permissible limit of WHO for chloride is ≤250 mg/l, indicating that all samples in the dataset comply with this limit. The mean potassium concentration is 3.4972 mg/l whereas the permissible limit of WHO for potassium is ≤12 mg/l, indicating that all samples in the dataset comply with this limit. The mean sodium concentration is 148.7925 mg/l. The coefficient of variation (CV) is 117.7833% whereas the permissible limit of WHO for sodium is ≤200 mg/l, indicating that all samples in the dataset comply with this limit. For the WQP Sulfate, the mean concentration is 101.3962 mg/l whereas the permissible limit of WHO for sulfate is ≤250 mg/l, indicating that all samples in the dataset comply with this limit.

In case of the WQP, Fluoride the mean fluoride concentration is 0.6160 mg/l whereas the permissible limit of WHO for fluoride is ≤1.5 mg/l, indicating that all samples in the dataset comply with this limit. Afterwards, correlation matrix is constructed and observed that the highest correlations is among these pairs: TDS-Sodium (0.89), TDS- Chloride (0.88), Chloride-Sodium (0.83), Sulfate-Sodium (0.76), TDS-Sulfate (0.82), Chloride-Sulfate (0.69), Calcium-Hardness (0.79). We run the KMO and Bartlett's test which suggested that multivariate factor analysis is suitable for this data. The dendrogram of hierarchical cluster analysis and Factor analysis is considered sequentially which suggested four most significant and vital WQP including TDS, sulfate, sodium and chloride.

As a second part of the analysis, these four WQPs are initially transformed using Box-Cox transformation to deal the asymmetrical WQPs. To capture the spatial correlation of the parameters, variogram model is estimated and the parameters (sill range and nugget) are plugged into the Ordinary Kriging model. The performance of the ordinary kriging technique is compared with the Inverse Distance Weighting (IDW) methods where Ordinary kriging outperforms the IDW. The prediction values resulted from the Ordinary kriging are than processed for contour plots which highlighted the contaminated areas. Maps showed that the high-risk zone in the study area is ranging from 72.0° to 73.2° longitude and from 33.0° to 33.1° latitude. Our findings suggested that ensure safe and sustainable water supplies for that contaminated area. These findings are suitable for the policy makers and water management organisations in improving groundwater management. On the basis of literature review, we can recommend that the geostatistical approaches cokring can also be applied on such data where our focus is to consider one of the spatial variables as primary variable and others as secondary variables. Moreover, cross-variography can also be applied to capture the distance-based correlations among the spatially measured variables. In case of large data sets, the best suggested multivariate geostatistical approach is partitioned kriging or Low-rank kriging. As far as the computational methods and softwares are concerned; there is a variety of R packages which can perform the prediction in efficient way. Most recommended packages are geoR, gstat, Rgdal, Rgeos, Raster graphics and sp.

# REFERENCES

Ahmad, M., Mustafa, G., Ali, N. and Laiq, M., 2023. Statistical prediction of fluoride concentration in groundwater of district Multan, Pakistan, using kriging methods. *Fluoride*, *56*(2).

Ahmad, M., Chand, S. and Rafique, H.M., 2017. Geostatistical cokriging and multivariate statistical methods to evaluate groundwater salinization in Faisalabad, Pakistan. *Water Treat*, *84*, pp.93-101.

Ahmad, M., Chand, S. and Rafique, H.M., 2016. Predicting the spatial distribution of sulfate concentration in groundwater of Jampur-Pakistan using geostatistical methods. *Desalination and Water Treatment*, *57*(58), pp.28195-28204.

Ahmad, M. and Chand, S., 2015. Spatial distribution of TDS in drinking water of tehsil Jampur using ordinary and Bayesian Kriging. *Pakistan Journal of Statistics and Operation Research*, pp.377-386.

Alin, A., 2010. Minitab. Wiley interdisciplinary reviews: computational statistics, 2(6), pp.723-727.

Box, G.E. and Cox, D.R., 1964. An analysis of transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *26*(2), pp.211-243.

Chapman, D.V., 2021. *Water quality assessments: a guide to the use of biota, sediments and water in environmental monitoring*. CRC Press.

Din, I.U., Ali, W., Muhammad, S., Shaik, M.R., Shaik, B., ur Rehman, I. and Tokatli, C., 2024. Spatial distribution and potential health risk assessment for fluoride and nitrate via water consumption in Pakistan. *Journal of Geochemical Exploration*, *259*, p.107413.

Fox, J., Weisberg, S., Price, B. and Fox, M.J., 2018. Package ‘carData’.

Foster, S.S.D. and Chilton, P.J., 2003. Groundwater: the processes and global significance of aquifer degradation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1440), pp.1957-1972.

Goovaerts, P., 1997. *Geostatistics for natural resources evaluation* (Vol. 483). Oxford University Press.

Gundogdu, K.S. and Guney, I., 2007. Spatial analyses of groundwater levels using universal kriging. *Journal of earth system science*, *116*, pp.49-55.

Hussain, I., Mubarak, N., Shabbir, J., Hussain, T. and Faisal, M., 2015. Spatial interpolation of sulfate concentration in groundwater including covariates using Bayesian hierarchical models. *Water Quality, Exposure and Health*, *7*, pp.339-345.

HUIZAR A, R.A.F.A.E.L., MENDEZ G, T.E.O.D.O.R.O. and MADRID R, R.A.F.A.E.L., 1998. Patterns of groundwater hydrochemistry in Apan-Tochac sub-basin, Mexico. *Hydrological sciences journal*, *43*(5), pp.669-685.

Idrees, N., Kamal, S. and Ahmad, M., 2020. Appraisal of groundwater alarming zones in district Okara-Pakistan using dimension reduction and Kriging procedures. *Desalination and Water Treatment*, *187*, pp.311-320.

Isaaks, E.H. and Srivastava, R.M., 1989. Applied geostatistics.

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

Kitanidis, P.K., 1997. *Introduction to geostatistics: applications in hydrogeology*. Cambridge university press.

Li, H., Smith, C.D., Wang, L., Li, Z., Xiong, C. and Zhang, R., 2019. Combining spatial analysis and a drinking water quality index to evaluate monitoring data. *International journal of environmental research and public health*, *16*(3), p.357.

Mubarak, N., Hussain, I., Faisal, M., Hussain, T., Shad, M.Y., AbdEl-Salam, N.M. and Shabbir, J., 2015. Spatial distribution of sulfate concentration in groundwater of South-Punjab, Pakistan. *Water Quality, Exposure and Health*, *7*, pp.503-513.

McGrath, D., Zhang, C. and Carton, O.T., 2004. Geostatistical analyses and hazard assessment on soil lead in Silvermines area, Ireland. *Environmental pollution*, *127*(2), pp.239-248.

Morris, B.L., Lawrence, A.R., Chilton, P.J.C., Adams, B., Calow, R.C. and Klinck, B.A., 2003. Groundwater and its susceptibility to degradation: a global assessment of the problem and options for management.

Nas, B., 2009. Geostatistical Approach to Assessment of Spatial Distribution of Groundwater Quality. *Polish Journal of Environmental Studies*, *18*(6).

Ribeiro Jr, P.J. and Diggle, P.J., 2006. Analysis of geostatistical data. *The geoR package, version*, pp.1-6.

Rouhani, S. ed., 1996. *Geostatistics for environmental and geotechnical applications* (Vol. 1283). ASTM International.

Smith, A.H., Lingas, E.O. and Rahman, M., 2000. Contamination of drinking-water by arsenic in Bangladesh: a public health emergency. *Bulletin of the world health organization*, *78*(9), pp.1093-1103.

Smith, V.K. and Desvousges, W.H., 1986. *Measuring water quality benefits* (Vol. 3). Springer Science & Business Media.

Subyani, A.M. and Al-Dakheel, A.M., 2009. Multivariate geostatistical methods of mean annual and seasonal rainfall in southwest Saudi Arabia. *Arab J Geosci*, *2*(1), pp.19-27.

Triki, I., Trabelsi, N., Zairi, M. and Dhia, H.B., 2014. Multivariate statistical and geostatistical techniques for assessing groundwater salinization in Sfax, a coastal region of eastern Tunisia. *Desalination and Water Treatment*, *52*(10-12), pp.1980-1989.

Team, R.C., 2000. R language definition. Vienna, Austria: R foundation for statistical computing, 3(1), p.116.

Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R. and Davies, P., 2010. Global threats to human water security and river biodiversity. *nature*, *467*(7315), pp.555-561.

*WHO chronicle,* 2011. Guidelines for drinking-water quality. Edition, Fourth, *38*(4), pp.104-8.

Webster, R. and Oliver, M.A., 2007. *Geostatistics for environmental scientists*. John Wiley & Sons.

Yeh, M.S., Lin, Y.P. and Chang, L.C., 2006. Designing an optimal multivariate geostatistical groundwater quality monitoring network using factorial kriging and genetic algorithms. *Environmental Geology*, *50*, pp.101-121.

Zhou, F., Liu, Y. and Guo, H., 2007. Application of multivariate statistical methods to water quality assessment of the watercourses in Northwestern New Territories, Hong Kong. *Environmental Monitoring and Assessment*, *132*, pp.1-13.

Zektser, I.S. and Everett, L.G., 2004. Groundwater resources of the world and their use.

R Code

######Correlation Matrix##################

install.packages('relaimpo')

install.packages('ggcorrplot')

library(relaimpo)

library(survey)

library(mitools)

library(ggplot2)

library(ggcorrplot)

library(stringr)

data=read.table("F:data.csv",header=TRUE,sep=",")

print(data)

names(data)

corr <- round(cor(data), 2)

corr

p.mat <- cor\_pmat(data)

p.mat

ggcorrplot(corr, method = "circle",lab=T)

ggcorrplot(corr, hc.order = TRUE,

type = "lower", p.mat = p.mat)

library(maps)

library(geoR)

Rdata=read.table("F:data.csv",header=TRUE,sep=",")

attach(Rdata)

names(Rdata)

par(mfrow = c(1,2),oma = c( 2.5, 2.5, 2.5, 2.5 ), mar=c(2.5,2.5,1.5,1),mgp=c(1.5,0,0))

hist(TDS,main="",xlab="TDS (mg/l)")

#####Spatial Analysis#####################

TDS.geo=as.geodata(data,coords.col = 1:2, data.col =4)

plot(TDS.geo$coord[,1] ~ TDS.geo$coord[,2], data = TDS.geo,xlab="longitude",ylab="latitude", main = "TDS Concentration")

with(TDS.geo, text(TDS.geo$coord[,2], TDS.geo$coord[,1], formatC(TDS.geo$data,dig=2), adj = 0.1))

######Ordinary Kriging################

library(sp)

library(MASS)

library(geoR)

library(mgcv)

data=read.table("F:/data.txt")

kc <- krige.conv(SLP.geo, loc=loci,krige=krige.control(cov.pars=c(1.90,0.05)))

image(kc,xlab="longitudes",ylab="latitudes",main="Ordinary Kriging for TDS")

contour(kc,xlab="longitudes",ylab="latitudes",main="Ordinary Kriging for TDS")

bin1 <- variog(SLP.geo,lambda= 0.06524205)

ini.vals <- expand.grid(seq(0,1,l=5), seq(0,1,l=5))

ols <- variofit(bin1, ini=ini.vals, fix.nug=TRUE, wei="equal")

###### Transformation using Box-Cox approach#########

library(car)

powerTransform(okara[,4]) #### Estimated value of lambda = 0.06524205

plot(TDS.geo,lambda= -0.2202624)

summary(TDS.geo,lambda= -0.2202624)

hist(TDS.geo,lambda= -0.2202624)