الجمهورية الجزائرية الديمقراطية الشعبية

People`s Democratic Republic of Algeria

وزارة التعليم العالي والبحث العلمي

Ministry of Higher Education and Scientific Research

Mohamed Khider University of Biskra

Faculty of exact sciences, natural sciences and life sciences

Department of Agronomic sciences

جامعة محمد خيضر بسكرة

كلية العلوم الدقيقة وعلوم الطبيعة والحياة

قسم العلوم الزراعية

THESIS

Submitted in candidacy for the degree of

DOCTORAT

Specialty: Hydropedology in arid regions

By: **BOUDIBI SAMIR**

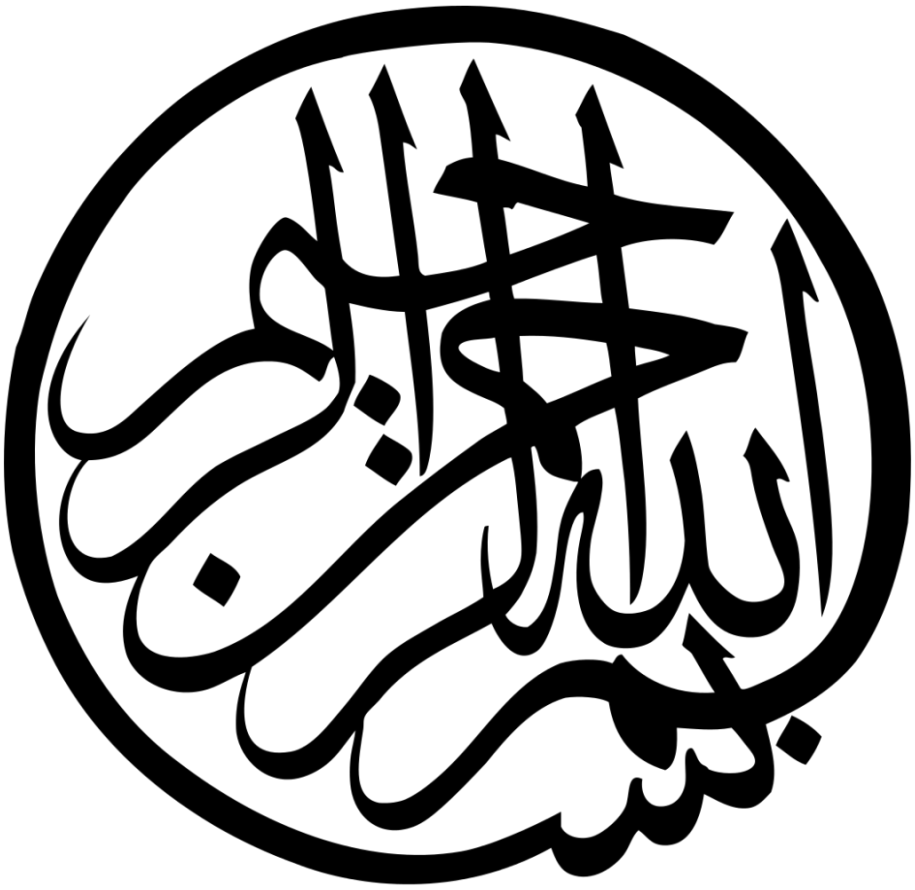
TITLE

Modeling the Impact of Irrigation Water Quality on Soil Salinization in an Arid Region, Case of Biskra

Board of Examiners:

Pr. Masmoudi Ali	Professor	U. Biskra	Chairman
Dr. Sakaa Bachir	Senior researcher	C.R.S.T.R.A.	Supervisor
Pr. Zapata Antonio Jesús	Professor	U. Almeria	Co-supervisor
Dr. Labadi Abdallah	MCA	U. Biskra	Examiner
Dr. Boumaraf Belkacem	MCA	U. Biskra	Examiner

2020/2021



IN THE NAME OF ALLAH, THE MOST MERCIFUL, THE MOST COMPASSIONATE

« *Our earth is the only place where we can live in harmony, peace, and cooperation for the betterment of humanity. It is the duty of each individual to try with utmost ambition to care for the earth environmental issues so that a sustainable future can be handed over to new generations. This is possible only through the scientific principles, where the earth systems and sciences are the major branches* **»**

My mentor: Pr. Zekâi Şen

Dedication

I dedicate this work to my dear parents,

To my wife, brother and my lovely sisters

To all my family and friends

Acknowledgement

Thanks **God**, the merciful and the passionate, for providing me the opportunity to undertake this research study and to persevere and complete it satisfactorily.

My gratitude goes to my supervisor **Dr. Sakaa Bachir** and my co-supervisor **Pr. Zapata-Sierra Antonio Jesus** for their continuous guidance to accomplish this work and their efforts to improve this dissertation.

My thanks go to the members of the board of examiners **Pr. Masmoudi Ali**, **Dr. Labadi Abdallah** and **Dr. Boumaraf Belkacem** for accepting to evaluate this work.

I take pride in acknowledging the insightful guidance of my mentor **Zekai Şen** for sparing his valuable time and motivating me to write this thesis in English. His assistance and advice were indispensable in completing this research project.

I would also like to thank my friends and colleagues at the scientific and technical research center (CRSTRA) and at Mohamed Khider University of Biskra for their help and support during the field and laboratory work.

I would specially thank my friend **Benguega Zinneddine** for his accompaniment during all the fieldwork period using his own car to collect groundwater and soil samples.

My acknowledgement would be incomplete without thanking the biggest source of my strength, my family. The blessings of my parents, the love and care of my wife, brother and my lovely sisters.

To everyone who cleared me with a kind word, I extend my sincere appreciation.

Table of contents

Part A

Chapter I : Literature review / Groundwater quality and soil salinization

Chapter II : Description of the study area

Part B

Chapter III: Mapping and assessment of groundwater quality

Chapter IV: Spatial variability and risk assessment of nitrate groundwater pollution

Chapter V: Modeling and mapping of soil salinity

Abstract

In El Outaya plain (southeastern Algeria), the climate is arid, the rainfall is scarce, and the evapotranspiration is high. Consequently, the use of groundwater is mandatory to maintain the irrigation and drinking water supply. During the last few years, the expansion of irrigation using groundwater resources has had an important positive impact on agricultural production but it has also introduced the challenge of groundwater sustainability in the plain. The deterioration of groundwater quality used for irrigation decreases the agricultural yields and causes soil salinization.

In order to evaluate the groundwater quality and its impact on soil salinization it was necessary to gather data from the field and the satellite images (Remote Sensing) over two seasons (dry and wet seasons). The field data consist in 136 groundwater and 272 soil samples (top soil and sub soil) collected from 68 farms spread over the plain's cultivated area. The remote sensing data consist of two satellite images used to calculate some environmental covariables, which help to predict soil salinity. Hydrogeochemical evaluation of groundwater and its suitability for irrigation and drinking purposes was carried out using different classical methods and water quality index approach. Artificial neural networks, machine learning and stochastic techniques performed the prediction and modeling of soil salinity using groundwater's geochemical properties, physical soil properties and remote sensing covariates as inputs.

Water quality index (WQI) depicts that the majority of groundwater samples fall within the "very poor" and "unsuitable" classes for drinking categories. Electrical conductivity (EC) and sodium adsorption ratio (SAR) are classified using Reverside diagram, which points out a high to very high risk of groundwater salinity and medium to high risk of alkalinity for soil salinization. The groundwater quality was accurately assessed using geographic information system (GIS) and ordinary kriging (OK) method, which can be helpful for groundwater managers and decision makers in arid areas.

The different techniques used for modeling and predicting soil salinization show the superiority of multilayer perceptron neural network (MLP-NN) in the accuracy of estimating soil salinity. In addition, the combination of groundwater salinity (ECw) and sand percentage from the field covariates and topographic wetness index (TWI), land surface temperature (LST) and elevation as inputs to MLP-NN and auxiliary variables to cokriging (SCOK) can estimate and improve the prediction of soil salinity in the study area.

Key words: El Outaya plain, Groundwater quality, soil salinization, artificial neural network, geostatistic, remote sensing.

Résumé

Dans la plaine d'El Outaya (sud-est de l'Algérie), le climat est aride, les précipitations sont rares et l'évapotranspiration est forte. Par conséquent, l'utilisation des eaux souterraines pour maintenir l'irrigation et l'approvisionnement en eau potable est obligatoire. Au cours des dernières années, l'expansion de l'irrigation utilisant les ressources en eau souterraine a eu un impact positif important sur la production agricole, mais elle a également introduit le défi de la durabilité des eaux souterraines dans la plaine. La détérioration de la qualité des eaux souterraines utilisées pour l'irrigation diminue les rendements agricoles et provoque la salinisation des sols.

Afin d'évaluer la qualité des eaux souterraines et son impact sur la salinisation des sols, il a été nécessaire de recueillir des données sur le terrain et des images satellites (télédétection) sur deux saisons (saison sèche et saison humide). Les données de terrain consistent en 136 échantillons d'eaux souterraines et 272 échantillons de sol (0-20 cm et 40-60 cm) recueillis auprès de 68 exploitations agricoles réparties sur la zone cultivée de la plaine. Les données de télédétection consistent en deux images satellites utilisées pour calculer certaines covariables environnementales, qui aident à prédire la salinité du sol. L'évaluation hydrogéochimique des eaux souterraines et de leur aptitude à l'irrigation et à la consommation a été réalisée en utilisant différentes méthodes classiques et l'approche de l'indice de qualité de l'eau. Des réseaux neuronaux artificiels, l'apprentissage de la machine et des techniques stochastiques ont permis de prédire et de modéliser la salinité des sols en utilisant comme intrants les propriétés géochimiques des eaux souterraines, les propriétés physiques des sols et les covariables de télédétection.

L'indice de qualité de l'eau (IQE) montre que la majorité des échantillons d'eaux souterraines appartiennent aux catégories "très mauvaise" et "mauvaise" à la consommation. La CE et le taux d'adsorption du sodium (SAR) classés à l'aide du diagramme de Reverside indiquent un risque élevé à très élevé de salinité des eaux souterraines et un risque moyen a élevé d'alcalinité pour la salinisation des sols. La qualité des eaux souterraines a été évaluée avec précision à l'aide du système d'information géographique (SIG) et de la méthode du krigeage ordinaire (OK), qui peut être utile aux gestionnaires des eaux souterraines et aux décideurs dans les zones arides.

Les différentes techniques utilisées pour la modélisation et la prévision de la salinisation des sols montrent la supériorité du réseau neuronal perceptron multicouche (MLP-NN) dans la précision de l'estimation de la salinité des sols. En outre, la combinaison de la salinité des eaux souterraines (ECw) et du pourcentage de sable des covariables de terrain et de l'indice d'humidité topographique (TWI), de la température de la surface des terres (LST) et de l'altitude en tant que entrées du MLP-NN et des variables auxiliaires du cokrigeage (SCOK) peut estimer et améliorer la prévision de la salinité des sols dans la zone d'étude.

Mots clés : Plaine d'El Outaya, qualité des eaux souterraines, salinisation des sols, réseau neuronal artificiel, géostatistique, télédétection.

ملخص

في سهل لوطاية (جنوب شرق الجزائر) ، المناخ جاف ، والأمطار نادرة والتبخر عالي. وبالتالي فإن استخدام المياه الجوفية لري المزروعات وإمدادات مياه الشرب إلزامي. خلال السنوات القليلة الماضية، كان لتوسيع المساحات المسقية باستخدام موارد المياه الجوفية تأثير إيجابي مهم على الإنتاج الزراعي ، لكنه أدى أيضًا إلى ظهور تحدي آخر وهو استدامة المياه الجوفية في السهل. تدهور جودة المياه الجوفية المستخدمة في الري يؤدي إلى انخفاض المحاصيل الزراعية ويسبب تملح التربة.

من أجل تقييم جودة المياه الجوفية وتأثيرها على تملح التربة كان من الضروري جمع البيانات من الميدان ومن صور الأقمار الصناعية (الاستشعار عن بعد) على مدى موسمين (المواسم الجافة والرطبة). تتكون البيانات الميدانية من 136 عينة من آبار المياه الجوفية و 272 عينة من التربة تم جمعها من 68 مزرعة موزعة على مساحة السهل المستخدمة من طرف الفلاحين. تتكون بيانات الاستشعار عن بعد من صورتين قمر صناعي تستخدمان لحساب بعض المتغيرات البيئية التي تساعد على التنبؤ بملوحة التربة. تم إجراء التقييم الهيدروجيوكيميائي للمياه الجوفية ومدى ملاءمتها لأغراض الري والشرب باستخدام طرق تقليدية مختلفة ومنهج مؤشر جودة المياه. استخدمنا الشبكات العصبية الاصطناعية والتعلم الآلي والتقنيات العشوائية بالتنبؤ لنمذجة ملوحة التربة باستخدام الخصائص الجيوكيميائية للمياه الجوفية وخصائص التربة الفيزيائية والمتغيرات المشتركة للاستشعار عن بعد كمدخلات.

يوضح مؤشر جودة المياه أن غالبية عينات المياه الجوفية تقع ضمن فئات "سيئة للغاية" و "غير المناسبة" للشرب. تشير نسبة EC ونسبة SAR إلى وجود مخاطر عالية إلى عالية جدًا من ملوحة المياه الجوفية ومخاطر متوسطة إلى عالية من الخاصية القلوية لتمليح التربة. تم تقييم جودة المياه الجوفية بدقة باستخدام

نظام المعلومات الجغرافية وطريقة kriging العادية، والتي يمكن أن تكون مفيدة لمسيري المياه الجوفية وصناع القرار في المناطق الجافة.

أظهرت التقنيات المختلفة المستخدمة لنمذجة تملح التربة والتنبؤ بها تفوق الشبكة العصبية متعددة الطبقات في دقة تقدير ملوحة التربة. بالإضافة إلى ذلك ، فإن الجمع بين ملوحة المياه الجوفية ونسبة الرمل من المتغيرات المشتركة في الميدان ومؤشر الرطوبة الطبوغرافية ودرجة حرارة سطح الأرض والارتفاع كمدخلات في MLP-NN والمتغيرات المساعدة لـ SCOK يمكن تقدير وتحسين التنبؤ لملوحة التربة في منطقة الدراسة.

الكلمات المفتاحية: سهل لوطاية، جودة المياه الجوفية ، تملح التربة ، الشبكة العصبية الاصطناعية ، الإحصاء الجغرافي ، الاستشعار عن بعد.

List of figures

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter V

List of tables

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter V

General Introduction

General introduction

Water use has grown worldwide at a rate more than twice the rate of population increase in the 20th century due to demographic pressures, the rapid socio-economic development and urbanization, to the point where reliable water services can longer be delivered in many regions, especially in arid and semi-arid regions (FAO, 2008). Agriculture is the sector with greatest water scarcity, where it accounts for more than 70 % of global freshwater withdrawals, and more than 90 % of its consumption, while domestic water withdrawal represents globally about 10 % of all water uses (FAO 2008).

Globally, groundwater is an important source for irrigated agriculture; it provides about 50% of all drinking water and 43% of agricultural irrigation. 20 % of the total cultivated land is considered as irrigated agriculture and contributes to 40 % of the global food production (FAO 2020). Food Agriculture Organization of the United Nations (FAO) estimates that the developing countries such as Algeria will increase their irrigated land by 34 %, but the quantity of water used for irrigation will increase only by 14 %, thanks to new irrigation technologies (precision agriculture) and improved irrigation practices (FAO 2020).

The most arid and semi-arid regions where average annual rainfall does not exceed 200 mm depends solely on groundwater resources to satisfy the growing demands of water needs. The groundwater resources in these regions are limited in the shallow aquifers and the most of the stored groundwater in local and regional sedimentary aquifers are non-renewable or semi-renewable. Under the exponential growth of population and the very limited recharge conditions in arid and semi-arid regions, any extensive use of groundwater will necessarily affect negatively the quantity and the quality of these resources. Furthermore, the concept of groundwater sustainability has been defined as the level of groundwater development that meets the needs of the present generation without compromising the ability of future generations to meet their needs (Abderrahman 2006). Practically, when speaking of aquifer over-exploitation, there is more concern about the consequences of intensive groundwater abstraction than in its potential level. Thus, a more appropriate definition is probably an economic one: that the overall cost of the negative impacts of groundwater exploitation exceeds the net benefits of groundwater use (Abderrahman 2006).

The groundwater resources in arid and semi-arid regions require sensitive planning and management approaches to deal with the overexploitation consequences such as the decrease in groundwater levels and the deterioration of their quality. Sustainable groundwater resources

management is defined as the group of measures, which avoids an irreversible or quasi-irreversible destruction of groundwater and any other natural resource depending on it such as soil and ecosystems (Amer 2008).

The irrigated areas have expanded rapidly in recent years. It has reached more than 324 million hectares in 2014 (Bradai 2014) and the equipped land for irrigation is estimated of 338 million hectares in 2018 (FAO 2020). However, the projected growth in world population over the next 30 years will require an increase of more than 20% in agricultural production in developed countries and 60% in developing countries to maintain current levels of food consumption (Bradai 2014), which will lead to the development of new irrigated land, particularly in arid and semi-arid regions. This development will put more pressure on groundwater where it is considered the main source of irrigation in these regions.

In Algeria, the agricultural sector contributes over 13 % to the country's GDP and employs at least 20 % of the population in rural areas (USDA 2018, Bizri 2018). The agricultural land is 413,602 km^2 (17.36 % of the total area) and the arable land is 75050 km^2 roughly 3.15 % of the total area (FAOSTAT 2018). About 51 % of the total arable land is dedicated to field crops, mostly cereals and pulses, 6 % to arboriculture, and 3 % to industrial crops (USDA 2018). The agriculture is primarily rain-fed and suffers from drought, which have worsened due to climate change, in recent years. The water sector in Algeria has to date paid scant attention to the issue of climate change and is often unaware of its impact on future water resources where models for climate change indicate that rainfall could decrease by more than 20% by 2050, which would result in even greater worsening water shortages in different basins of Algeria (Hamiche et al. 2015). The water resources in in the northern regions of Algeria are estimated at 7.4 billion m^3 of surface water and 2.6 billion m^3 of groundwater. The Algeria Sahara is one the vulnerable region to climate change, which have negative impacts on several socioeconomic sectors of the region like water resources. The water reserves available in these regions are estimated at a total of 5.37 billion m^3 where 0.37 billion m^3 of surface water and 5 billion m^3 of underground water (Hamiche et al. 2015).

The agricultural water demand corresponds to the potential irrigation water requirements assessed based on recommended crop rotations and theoretical crop water requirements. These water needs are mainly provided by irrigation, especially in the southern part of the country where rainfall is scarce and groundwater is the main source of irrigation. Most of the aquifers in these regions are overexploited due to the high water demand of the different sectors and the low or non-renewable groundwater. The overexploitation of aquifers, the poor agricultural

practices and return flow from irrigated lands are the major causes of groundwater salinization and water quality deterioration, which affect the sustainable use of this valuable resource.

Agricultural irrigation represents the main use of global water resources (Pulido-Bosch et al. 2018). Irrigation contributes to food security, poverty reduction and improvement quality of life for a large part of the world's population, but if the irrigation water is not appropriate, the damages will be catastrophic. The risk of salinity is one the biggest challenges facing irrigated agriculture. Irrigation has an impact on the environment, and scientific evidence suggests that it inevitably leads to the salinization of soils and aquifers. These impacts are enhanced under the arid and semi-arid conditions (Pulido-Bosch et al. 2018). The impacts of irrigation practices on groundwater quality can be classified into direct such as the consequences of fertigation (Applying water and accompanying agrochemicals such as fertilizers, herbicides, and pesticides, to irrigated cropland), or indirect such as the effects of irrigation withdrawals on the water chemistry of aquifers, which are generally manifested by a continuous degradation of the quality of the groundwater being pumped (Bouzourra et al. 2014; Pulido-Bosch et al. 2018).

Low groundwater quality (e.g. saline) contributes significantly to soil salinization in irrigated areas. Arid and semi-arid regions are characterized by a negative water balance (Bradai 2014). Evaporation of water accumulated in low areas, either from shallow aquifers or from irrigation using saline water leads to soil salinization. The latter is one of the major threats of land degradation occurring around the world, posing risks to agricultural production, environmental health and economic prosperity. The combination of high evapotranspiration rates, low rainfall amounts results in salt accumulation in the top soil layers where it affects the physicochemical properties of the soil, leads to texture degradation and limits vegetation growth eventually rendering productive fields barren (Alexakis 2016).

In Algeria, the extension of irrigation in agriculture has led to a significant increase in agricultural yields. In recent years, the development of irrigated crops in the south of the country has been mainly based on groundwater resources, which ensured higher income for farmers, but it is also responsible for soil salinization caused by poor water quality (Abdennour et al. 2020; Kuper et al. 2016). In Biskra province (southeastern Algeria), farmers are currently using groundwater of poor quality, usually rich in salts, for irrigation, and no information is available about its geochemical pathway (Abdennour et al. 2020).

El Outaya plain is located in the northern part of Biskra province, on the southern flank of the Aures Mountains, as part of the Saharan Atlas (Boudjema 2015). It is characterized by a very

severe arid climate with low rainfall amounts and very high evapotranspiration rates. In recent years, the plain has experienced a rapid expansion of agriculture, and consequently, the increase in water demand either for agriculture or for human consumption. Groundwater is the main source of irrigation and drinking purposes despite the existence of the Gazelles' fountain dam in northern part of the plain, which is used to irrigate El Hzima perimeter in the southern part of the plain. However, the continuous increasing of water demand will make more pressure on groundwater and will affect negatively its quality. The deterioration of the groundwater quality in the plain will endanger the health of the population, degrade the soil quality (e.g. salinization) and reduce the agricultural production.

The present work is intended to achieve the following objectives:

- Characterization of the chemical facies of groundwater in El Outaya plain, the origin of the mineralization as well as the spatio-temporal evolution of its parameters and assess the suitability of this water for irrigation and drinking purposes.
- Elucidate the impact of groundwater salinity on soil salinization, prediction and mapping of soil salinity by using of remote sensing, artificial neural networks (ANN), machine learning and geostatistical approaches.

Consideration of the objectives mentioned above has led us to pursue the following strategy, which is structured in two parts comprising five chapters.

The first part includes two chapters:

- The first chapter presents a literature review on groundwater quality, soil salinization and the current methods and approaches used for predicting soil salinity.
- The second chapter presents a brief description of the study area (localization, geology, hydrogeology, climatology and pedology).

The second part includes three chapters:

- In the third chapter, the study of the geochemical evolution of the groundwater is exposed with its suitability for irrigation and drinking purposes by means of water quality index and geostatistical approaches.
- In the fourth chapter, the focus is on the nitrate-N pollution of groundwater by applying different interpolation techniques.
- Finally, in the fifth chapter, use the groundwater geochemical parameters are considered in addition to the soil properties and remote sensing data as covariables for modeling

and improving the prediction of soil salinization by applying cokriging, multilayer perceptron neural networks and support vector machines.

Chapter I

Literature review / Groundwater quality and soil salinization

1. Geochemistry characterization and assessment of groundwater quality

1.1. Geochemistry

Groundwater geochemistry is the science that explores the processes controlling the chemical composition of groundwater. This resource may contain hazardous substances that affect health when consumed or which deteriorate the environment when used for irrigation (soil salinization). The groundwater quality may change by natural processes or it may be affected by human activities of which the impact is not always immediately evident (Appelo and Postma 2005).

The geochemical evolution and the origin of groundwater can be determined by the construction of Piper (1944) trilinear diagram and Durov (1948) plot.

1.1.1 Piper's trilinear diagram

The Piper's diagram (Figure 1) is a multifaceted plot in which percentage concentrations in milliequivalents of the major cations and anions are plotted in two triangular fields, which were then projected further into the central diamond field (Ravikumar et al. 2015). In this diagram, the diamond-shaped field that decides the water type and the hydrochemical facies in groundwater samples.

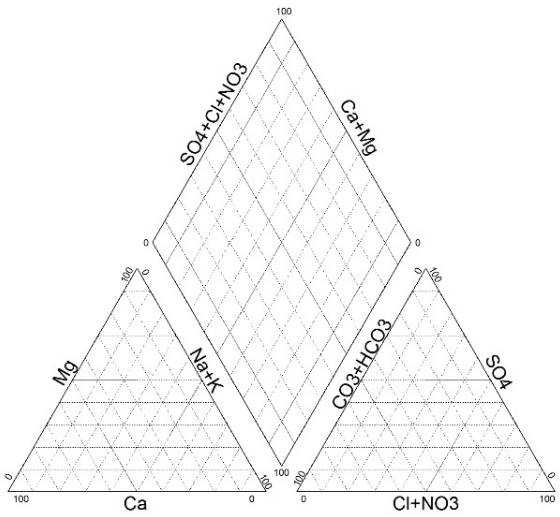


Figure 1: Piper's diagram

1.1.2. Durov's plot

The Durov's diagram (Figure 2) is a composite plot consisting of two ternary diagrams, where the percentages of milliequivalents of the cations of interest have been plotted against those of the anions; the sides form a central rectangular and binary plot of total cation concentrations

against total anion concentrations (Ravikumar et al. 2015). The intersection of the lines extended from the points of the ternary diagrams and projected on the sub-divisions of the binary plot of the Durov's diagram defines the hydrochemical processes involved as well as the type of water

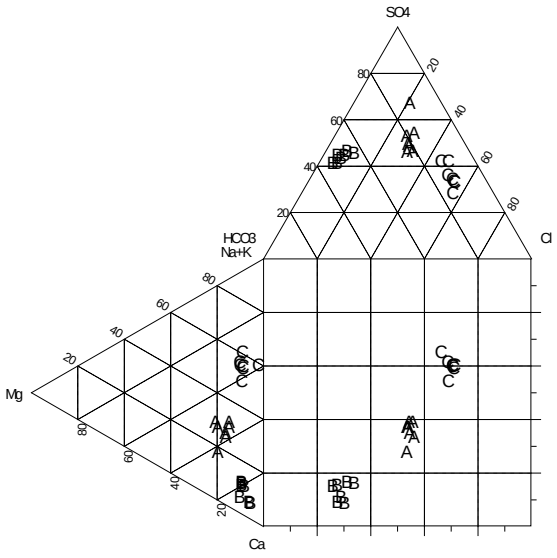


Figure 2: Durov's diagram

1.2. Aptitude for irrigation

Many methods and indices are used by researchers to assess the suitability of groundwater for irrigation and the most used are based on Electrical conductivity (EC), Sodium absorption ratio (SAR), Sodium percentage (NA %) Residual sodium carbonate (RSC) and Permeability index (PI).

1.2.1. Assessment criteria

1.2.1.1. Electrical conductivity (EC)

The total concentration of soluble salts in irrigation waters can be adequately expressed in terms of EC for diagnosis and classification aims. High values of EC means a large quantity of ions in solution, which makes it more difficult for plants to absorb water and mineral elements (USSLS 1954). The main chemical ions responsible of water salinity are Na^+ (sodium), Cl^- (Chloride), Ca^{2+} (calcium), SO_4^{2-} (sulfate) and K^+ (potassium).

The water salinity in terms of EC (µS/cm) is divided into 5 categories (USDA 2011, Zaman et al. 2018; Bradaï 2016):

Low salinity water (EC < 250): This category of water can be used for irrigation of most crops on most soils with little likelihood that soil salinity will develop. Some leaching will be required for salinity Class C1 water, but this occurs under normal irrigation practices, except for soils with extremely low permeability.

Medium salinity water (250 < EC < 750): This class can be used if a moderate amount of leaching can occur. Plants with moderate salt tolerance can be grown in most cases without special practices for salinity control.

High salinity water (750 < EC < 2250): This category cannot be used on soils, which possess restricted drainage and, thus, poor leaching abilities. Even with adequate drainage, special management for salinity control may be required and plants with good salt tolerance should always be selected.

Very high salinity water (2250 < EC < 5000): This kind of water is not suitable for irrigation under ordinary conditions; but may be used occasionally under very special circumstances. Here, the soils must be permeable, drainage must be adequate to good and irrigation water must be applied in excess in order to provide considerable leaching. Only very salt tolerant crops should be selected.

Excessive salinity water (EC > 5000): This water is unsuitable for irrigation.

1.2.1.2. Sodium absorption ratio (SAR)

The alkali hazard involved in the use of irrigation water is determined by the relative concentrations of the cations. If the concentration of sodium is high, the alkali hazard is high; and conversely, if the calcium and magnesium are the dominant cations, the alkali hazard is low (USSLS 1954; USDA 2011).

The sodium hazard of irrigation water is expressed as the sodium adsorption ratio (SAR in meq/l). Sodium contributes directly to the total salinity and may also be toxic to sensitive crops. The main problem with a high sodium concentration is its effect on the physical properties of soil (soil structure degradation) (Zaman et al. 2018).

The classification of irrigation waters with respect to sodium adsorption ratio (SAR) is based primarily on the effects, where exchangeable sodium accumulation has on the physical conditions of the soil (USDA 2011; Zaman et al. 2018).

Low Sodium Water (SAR < 10): It can be used for irrigation on almost all soils with little danger of the soil developing harmful levels of exchangeable sodium. However, sodium sensitive crops such as stone fruit trees and avocados may accumulate injurious concentrations of sodium.

Medium Sodium Water (10 < SAR < 18): It will present an appreciable sodium hazard in fine textured soils having high cation exchange capacity, especially under low leaching conditions, unless gypsum is present in the soil. This water may be used on coarse-textured soils or organic soils with good permeability.

High Sodium Water (18 < SAR < 26): It may produce harmful levels of exchangeable sodium in most soils. Its use will require special soil management methods, good drainage, a high leaching ability and high organic matter conditions. Gypsiferous soils, however, may not develop harmful levels of exchangeable sodium from such waters. Management methods may require use of chemical amendments, which encourage the replacement of exchangeable sodium. That said, use of those amendments may not be feasible with waters of very high salinity.

Very High Sodium Water (SAR > 26): It is generally unsatisfactory for irrigation purposes except at low and perhaps medium salinity. Specifically, where the soil water solution is rich in calcium or the use of gypsum or other soil amendments may make the use of this category of irrigation water feasible.

1.2.1.3. Residual sodium carbonate (RSC)

The excess sum of carbonate and bicarbonate in groundwater over the sum of calcium and magnesium also influences the suitability of groundwater for irrigation. This can be expressed as residual sodium carbonate (RSC), which has been widely used to predict the additional sodium hazard, which is associated with $CaCO_3$ and $MgCO_3$ precipitation (Raju et al. 2009; Zaman et al. 2018).

According to the RSC (meq/l) values, the irrigation water is divided into three classes: Safe water (RSC < 1.5), marginal (1.5 < RSC < 2.5) and unsuitable (RSC > 2.5).

1.2.2. Assessment methods

1.2.2.1. Riverside diagram

The Riverside diagram (Figure 3) is a combination between the 5 salinity classes (µS/cm) given by the letters C1 to C5 in the horizontal axis and the 4 classes of SAR (meq/l) given by the

letters S1 to S4 in the vertical axis. Thus, the class C1S1 represents excellent water because of the low values of EC and SAR, while the class C5S5 corresponds to the worst water with the highest values of EC and SAR.

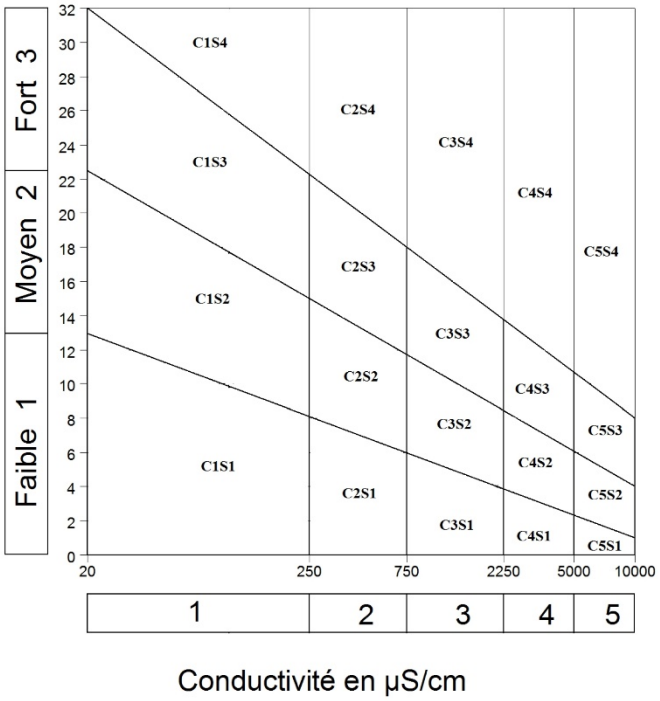


Figure 3: Riverside diagram

The interpretation of the different classes is as follows (USSLS 1954; Bradaï 2016):

C1S1 class: Represents the water of good quality for irrigation and must be used with precaution to sensitive plants.

C1S2 and C2S1 Classes: Correspond to water of good to medium quality, which must be used, with precautions to poorly drained soils and sensitive crops.

C1S3, C2S2 and C3S1 classes: They correspond to water of medium to mediocre quality, which requires drainage and leaching doses.

C1S4, C2S3, C3S2 and C4S1 classes: These classes represent water of mediocre to poor quality, which is used with precaution in the heavy soil and sensitive crops. For the slight soil, drainage and leaching doses and/or supply of gypsum are required.

C2S4, C3S3, C4S2 and C5S1 classes: They correspond to water of very poor quality used only for the slight and drained soils, and tolerant crops with leaching doses and/or gypsum contribution.

C3S4, C4S3 and C5S2 classes: Represent water of very poor quality used only in very exceptional conditions.

C4S4, C5S3 and C5S4 classes: These classes are not recommended for irrigation.

1.2.2.2. Wilcox diagram

The sodium percentage (Na%) is one of the most used parameters to evaluate the suitability of all natural waters for agricultural purposes. The classification of Wilcox (1948) is based on plotting the water salinity in terms of EC (µS/cm) on the horizontal axis against the Na % on the vertical axis.

From the Wilcox diagram (Figure 4), five classes can be assigned to irrigation water: Excellent, good, permissible, doubtful and unsuitable.

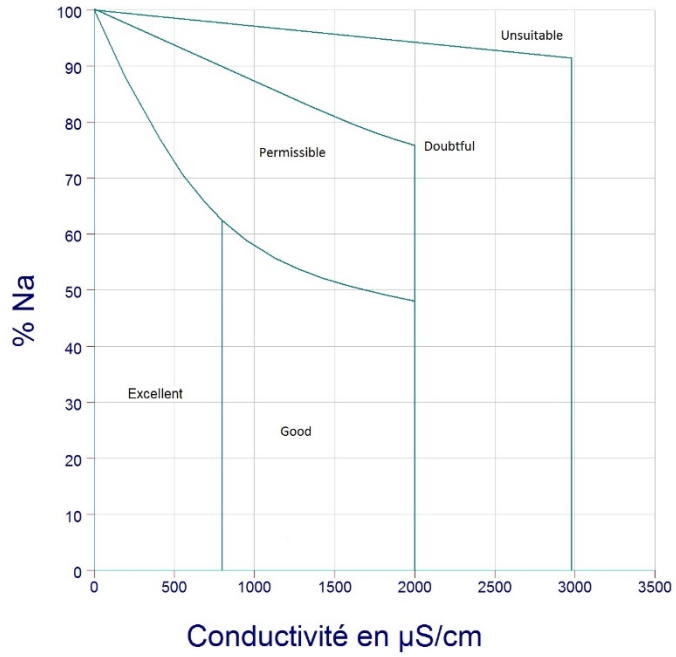


Figure 4: Wilcox diagram

1.3. Aptitude for human consumption

Groundwater is the main source of drinking for many people worldwide, especially in arid and semi-arid regions, where rainfall is scarce and the access to surface water of good quality is difficult. It is estimated that approximately one third of the world's population use groundwater

for drinking purpose (Srinivas et al. 2017). Therefore, the protection of this valuable resource from contamination and the determination of its suitability for human consumption is a priority.

1.3.1. Water quality indices

The aim of water quality indices (WQI) is to give a single value as the water quality of one or the other system that translates the list of constituents and their concentrations present in a sample into a single value (Abbasi and Abbasi, 2014). Then the quality of different samples can be compared based on the WQI of each sample.

The development of WQIs to assess the water quality in a sample using a single value have gained more importance in the last three decades, but the concept was introduced in 1848 in Germany. Since then different European countries have developed and applied various water classification systems, which are usually of two types; those concerned with the amount of pollution present and those concerned with living communities of microscopic and macroscopic organisms (Abbasi and Abbasi, 2014).

The WQIs can be formulated in two ways one in which the index numbers increase with the degree of pollution (increasing scale indices) and the other in which the index numbers decrease with the degree of pollution (decreasing scale indices). One may classify the former as ‘water pollution indices’ and the latter as ‘water quality indices’ (Abbasi and Abbasi, 2014).

The most often associated steps for developing any WQI are as follows:

- The selection of the parameter;
- The transformation of the parameters of different units and dimensions to a common scale;
- Assignment of weightages to all the parameters;
- Aggregation of sub-indices to produce a final index score.

The development of sub-index of each parameter is based on the standards of drinking water.

1.3.2. Water quality and World Health Organization (WHO) standards

The deficiency or the excess of one chemical constituent in drinking water can cause health problems. The WHO and the contribution of many individuals from various countries have participated in the development of guidelines determining the safe interval of each water parameter for human consumption.

The main aim of the guidelines for drinking water quality is the protection of public health. The guidelines provide the recommendations of WHO for managing the risk from hazards that may compromise the safety of drinking water. The nature and form of drinking water standards may vary among countries and regions (WHO 2011).

A guideline value normally represents the concentration of a constituent that does not result in any significant risk to health over a lifetime of consumption. The most constituents arising in drinking water are of health concern after long-term exposure and the exception is the nitrate (WHO 2011). The chemicals in drinking water can come from various sources: natural, industrial and human dwellings, and agricultural activities.

Table 1: WHO standard of cations, anions, pH and TDS in drinking water

Parameters	WHO Standards
TDS (mg/l)	1000
pH	6.5 - 8.5
EC (µS/cm)	2500
NO_3^- (mg/l)	50
SO_4^{-2} (mg/l)	250
Cl^- (mg/l)	250
HCO_3^- (mg/l)	-
Ca^{2+} (mg/l)	100
Mg^{2+} (mg/l)	50
Na^+ (mg/l)	200
K^+ (mg/l)	12

2. Soil salinity

Soil salinity is a measure of the concentration of all the soluble salts in soil solution, and is usually expressed as electrical conductivity (EC). Soil is said to be saline when the EC of a soil extracts from a saturated paste (ECe) equals, or exceeds 4 (dS/m) at 25 °C. The soluble salts that occur in soils consist of mostly various proportions of the cations sodium (Na^+), calcium (Ca^{2+}), and magnesium (Mg^{2+}), and the anions chloride (Cl^-) and sulfate (SO_4^{2-}). Constituents that ordinarily occur only in minor amounts are the cation potassium (K^+) and the anions

bicarbonate (HCO_3^-), carbonate (CO_3^-), and nitrate (NO_3^-). Hyper-saline soil solution may also contain boron (B), selenium (Se), strontium (Sr), lithium (Li), silica (Si), rubidium (Rb), fluorine (F), molybdenum (Mo), manganese (Mn), barium (Ba), and aluminum (Al) (USSLS 1954; Zaman et al. 2018).

2.2. Saline and Alkali soils

Saline soils refers to soil that contains sufficient soluble salts to interfere with its productivity. Similarly, alkaline soils can be defined in terms of productivity under the influence of exchangeable sodium. For agricultural use, these soils are considered a class of problematic soils that require special remedial measures and management practices. Soluble salts produce harmful effects on plants by increasing the salt content of the soil solution and by increasing the degree of saturation of exchange materials in the soil with exchangeable sodium (USSLS 1954).

2.3. Sources of soluble salts in soils

The most common sources (Zaman et al. 2018) are listed below:

- Inherent soil salinity (weathering of rocks, parent material);
- Brackish and saline irrigation water;
- Sea water intrusion into coastal lands as well as into the aquifers due to over extraction; and overuse of fresh water;
- Restricted drainage and a rising water-table;
- Surface evaporation and plant transpiration;
- Sea water sprays, condensed vapors which fall onto the soil as rainfall;
- Wind borne salts yielding saline fields;
- Overuse of fertilizers (chemical and farm manures);
- Use of soil amendments (lime and gypsum);
- Use of sewage sludge and/or treated sewage effluent;
- Dumping of industrial brine onto the soil.

2.4. Primary and secondary salinization

Soil salinization is a global problem that affects almost every continent; it is not static but dynamic. Salinization can affect the ecosystem to such a degree that it cannot provide environmental services at their full potential. It is a global, regional and national problem that

concerns every society. Many factors contribute to the development of soil salinization conditions (Shahid et al. 2010).

Large agricultural land are abandoned each year due primary and secondary salinization. Presence of parent materials and salt minerals in soil, weathering of rocks (chemical or physical) and rising water table and the subsequent evaporation of the soil water are among the main sources of primary salinization (Gorji et al. 2020; Shahid et al. 2010). In contrast to primary salinization, human activities such as irrigated agriculture and poor agricultural practices are the main cause of secondary salinization (Zaman et al. 2018).

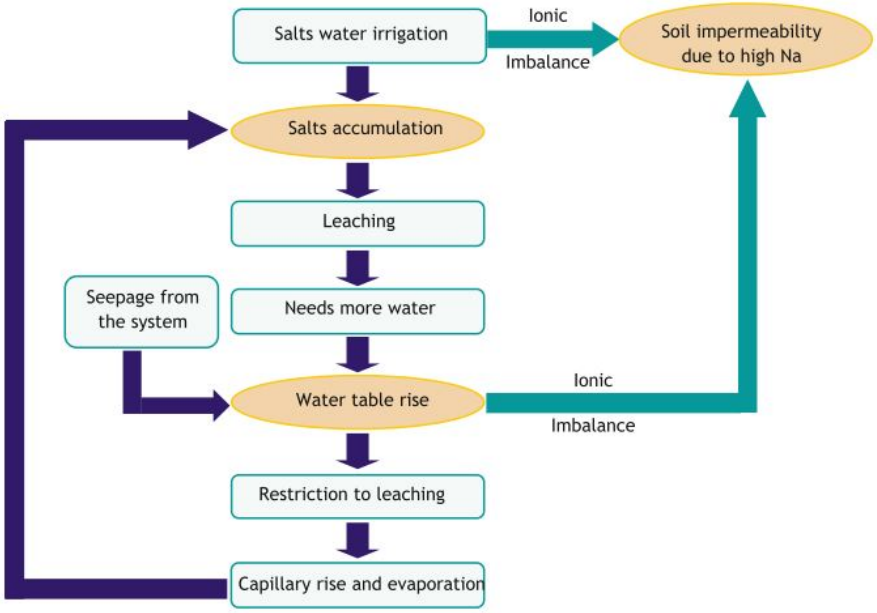


Figure 5: A hypothetical soil salinization cycle (Shahid et al. 2010; Zaman et al. 2018)

2.5. Damages caused by soil salinity and solutions

The scarcity of water in arid semi-arid regions requires the use of saline groundwater to meet part of the water needs of crops. Inappropriate use of this poor water quality, particularly in soils with limited drainage, leads to capillary upwelling and subsequent evaporation of water from the soil. This results in the development of surface and subsurface salinity, thus reducing the value of soil resources. Environmental researchers all around the world have developed many techniques to fight against the phenomenon of salinization in agricultural land. The most used methods (Zaman et al. 2018) are:

- Lowering of shallow water tables with safe use or disposal of pumped salt water;

- Reducing evaporation and salt build-up on surface soils through conservational agriculture practices such as mulching;
- Leaching of excess salts from the surface soil into the subsoil;
- Selection of salt tolerant crops.

2.6. Visual assessment of soil salinization

The development of soil salinization in agricultural fields affects considerably the soil properties and crops growth. The consequences can be observed visually and the most used visual indicators to assess soil salinization (Shahid and Rahman 2011) are:

- The development of white salt crust on the soil surface;
- Reducing the plant vigor;
- The surface of the ground has a fluffy appearance;
- The rate of germination is reduced;
- Slat stains can be observed on the soil surface
- Damage and burning of leaves;
- Waterlogging;
- The presence of naturally developing halophytes (indicator plants).

2.7. Field and laboratory assessment of soil salinization

The visual indicators cannot be used for the quantification of soil salinization. This is only possible through field measurement of EC. The EC determination can be done using two methods, the saturated paste extract or the soil: water suspension. The commonly used soil to water ratios (Zaman et al. 2018) are:

- 1:1 ratio corresponds to 10 g of soil + 10 ml of distilled water;
- 1:2.5 ratio corresponds to 10 g of soil + 25 ml of distilled water;
- 1:5 ratio corresponds to 10 g of soil + 50 ml of distilled water.

2.8. Socioeconomic and environmental impacts of soil salinization

The soil salinization has considerable impacts on the socioeconomic and the environmental aspects of the affected regions (Shahid and Rahman 2011; Zaman et al. 2018).

2.8.1. Socioeconomic impacts

- Socioeconomic disturbances due to farm abandonment and reduction of farmers;
- Economic losses due to low production;
- High costs for soil reclamation;

- Increase in the rate of poverty as a result of income loss due to the reduction of crops productivity in saline land;
- Degradation of soil quality requires more inputs (financial pressure on farmers).

2.8.2. Environmental impacts

- Fragmentation of the ecosystem;
- Erosion and soil degradation due to poor vegetation;
- Sand encroachment in productive areas
- The contamination of groundwater (return flow);
- Reduction of reservoir storage capacity.

3. Current approaches and techniques of groundwater quality and soil salinity modeling and mapping

Modeling and mapping of soil salinity levels are indispensable for the prevention and mitigation of land degradation in arid and semi-arid regions. The following section provides a brief review of the advanced techniques currently used for soil salinity modeling and mapping.

3.2. Stochastic techniques

Geostatistical (Stochastic) techniques are used for mapping surface characteristics from limited sample data and predicting values at unsampled locations. They are widely used in fields where spatial data are studied (Zaman et al. 2018).

In the early 1950s, Daniel Krige (The father of geostatistics) developed empirically statistical method to predict ore grades from spatially correlated sample data in the South African gold mines. This technique was effectively the first use of kriging, which he called later simple elementary kriging (Oliver 2010; Oliver and Webster 2014). The first use of the term “kriging” was in 1963 by the French researcher Matheron, where he developed the empirical ideas of D.G. Krige, in particular the concept that neighboring samples could be used to improve the accuracy prediction, and placed them within the theoretical framework of the regionalized variable theory underlying geostatistics, which provides the basis for solving the most critical problem in environmental science of the need to predict the unsampled location from dispersed data (Oliver 2010). Now Geostatistical methods are applied in many environmental fields, such as pedology, hydrogeology, geology, agriculture, hydrology, meteorology, remote sensing, and, importantly here, groundwater and soil sciences.

Many environmental scientists used different geostatistical techniques, such as ordinary kriging, simple kriging, universal kriging, indicator kriging and cokriging, for predicting and mapping groundwater and soil salinity.

3.3. Remote sensing (RS) and geographical information systems (GIS)

Groundwater quality and soil salinity mapping and modeling can be performed by integrating remote sensing (RS) and geographical information systems (GIS).

The ability of remote sensing to generate information on spatial and temporal domain is one the greatest advantages for hydrogeological investigations and monitoring (Khan and Jharya 2018). The quantification of land use and land cover (LULC) types, such as agricultural land and urban area associated with human activities, through remote sensing images is one of the useful applications to understand the driving factors influencing the groundwater quality of an area (Singh et al. 2014).

Remote sensing (RS) technology has been widely used to assess soil salinity; while aerial photographs have been used to map salt-affected soils (Gorji 2020). RS techniques provide spatial and temporal data for varying spatial domains and conditions, which is a key element in assessing and detecting soil salinity. Soil salinity can be properly detected and quantified using numerical indices derived from different spectral bands of satellite images and relating them to ground measurements of soil electrical conductivity (EC) (Gorji 2015). RS data and techniques offer more economic and efficient tools for monitoring and mapping soil salinity through salt features that are visible at the soil surface or from indicators such as the presence of halophytic plants (Allbed and Kumar 2013).

Geographical information system (GIS) is a system that involves the storage, analysis, retrieval, and display of data, which are spatially referenced to the earth (Khan and Jharya 2018; Zaman et al. 2018). The most familiar type of spatial data is the map. GIS is actually a means of electronically storing map information. A GIS map has many advantages over older maps, one of the main ones being stored electronically and they can be easily analyzed by computer.

3.4. Artificial neural networks (ANNs) and support vector machines (SVMs)

In the recent years, machines learning techniques (ML) have gain popularity in modeling soil properties. Artificial neural networks (ANNs) and support vector machines (SVMs) are among the most widely used ML techniques for modeling soil salinity.

The Artificial Neural Network (ANN) is a simplified model of a biological neural system that consists of a massive parallel-distributed information processing system that exhibits certain performance characteristics resembling the biological neural networks of the human brain (Gong 2016; Haykin 1999). However, the following characteristics are considered as common functions in real and artificial networks: (Shanmuganathan 2016):

- Learning and adaptation;

- Generalization;
- Massive parallelism;
- Robustness;
- Associative storage of information;
- Spatiotemporal information processing.

The neurons are the backbone of any ANN because they establish relationship between the input and the desired output variables. This process of establishing relationship between inputs and outputs is called training. The ANN problem is solved according to the selected training algorithm (Levenberg-Marquardt, Scaled conjugate gradient, Gradient descent …etc.) (Gong 2016; Achieng 2019). The construction of a neural network could be achieved according to the following steps (Shanmuganathan 2016):

- Identification of the studied problem and the available knowledge;
- Selecting the appropriate neural network to solve the problem;
- Data preparation for training the network (statistical analysis, discretization, and normalization);
- Training the network using training data;
- Testing the trained network and validation of the results.

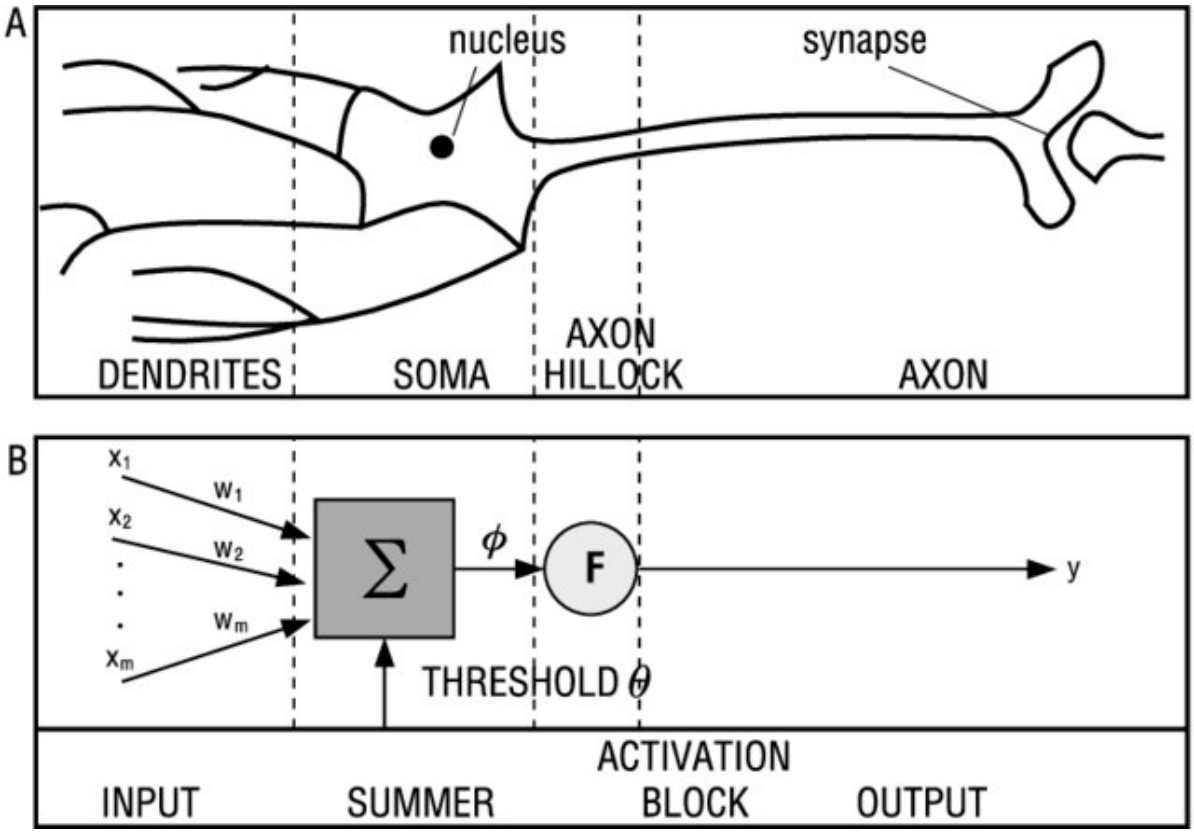


Figure 6: (A) A biological neuron; (B) An artificial neuron (Świetlik et al. 2004)

Support vector machines are powerful class of tools that are becoming increasingly popular in the fields of classification, data mining, pattern recognition, artificial intelligence, and optimization (Yang 2019). Recently SVM has been used to solve non-linear regression estimation and time series prediction by introducing ε- insensitive loss function (Samui 2008). SVM regression is a supervised, non-parametric, statistical learning technique, generally has an adequate balance between predictive accuracy and the ability to generalize trained models to unseen data (Deiss et al. 2020). The advantages of SVM models are their ability to handle high-dimensional multivariate spaces and to handle noisy models and multimodal class distributions of soil properties (Deiss et al. 2020).

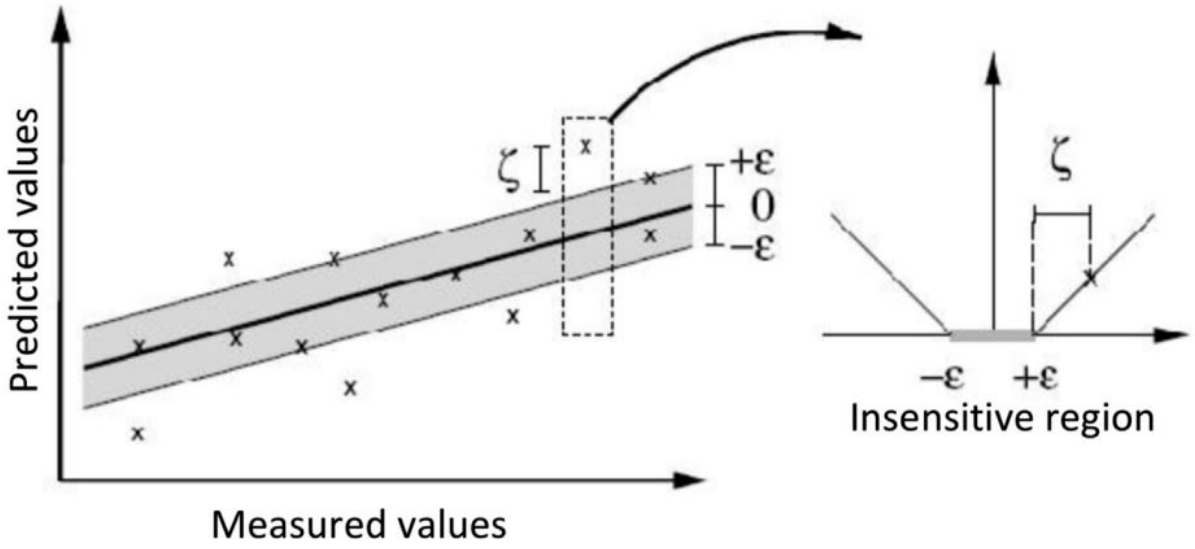


Figure 7: A theoretical representation of how SVM models deal with the ε-insensitive loss function (Deiss et al. 2020)

Today ANNs and SVMs have been recognized as powerful tools in the modeling of various complex environmental problems of different domains.

3.5. Previous studies

The described techniques (geostatistics, RS, GIS, ANNs and SVMs) have been integrated and used, with success, in many studies for mapping, predicting and modeling groundwater and soil properties, among them are the followings:

Zhou et al. (2020) integrated RS derivatives and machine learning algorithms to predict and map soil organic carbon (SOC) and soil total nitrogen (STN) in the southern part of Central Europe. The study used the derivatives of different RS sensors as inputs to four machine-learning techniques including support vector machine (SVM), random forest (RF), boosted

regression trees (BRT) and bagged CART. The results of this study proved the usefulness and the accuracy of using RS derivatives to predict SOC and STN. Koulla et al. (2019) used geostatistics and ANNs for modeling soil salinity, where they mentioned the accuracy and the superiority of ANNs with digital elevation model (DEM) derivatives as inputs. Gorji et al. (2020) applied RS images for mapping soil salinity in western part of Urmia Lake in Iran, where they employed different indices coupled with EC measurements to assess the performance of Landsat-8 and sentinel-2 in soil salinity mapping. Wagh et al. (2016) developed ANN model to determine the suitability of groundwater for irrigation in India. They used the different physicochemical groundwater parameters as inputs to the model. Shahabi et al. (2016) predicted soil salinity by employing ordinary kriging (OK), multiple linear regression (MLR) and ANNs using terrain features and remote sensing indices. The results of the study showed the robustness of ANNs in predicting soil salinity.

Chapter II

Description of the study area

1. Geographical localization

Biskra region is located in southeastern Algeria at 425 Km of the capital (Algiers) (Figure 8). It is bounded in the north by the Saharan Atlas, which represents a SW – NE directional relief. It extends to the Chott Melghir area in the southeast and to the Eastern erg in the southwest. Biskra region constitutes a transition zone between two different morpho-structural domains, the folded domains in the north and the flat and desert expanses of the Sahara in the south.

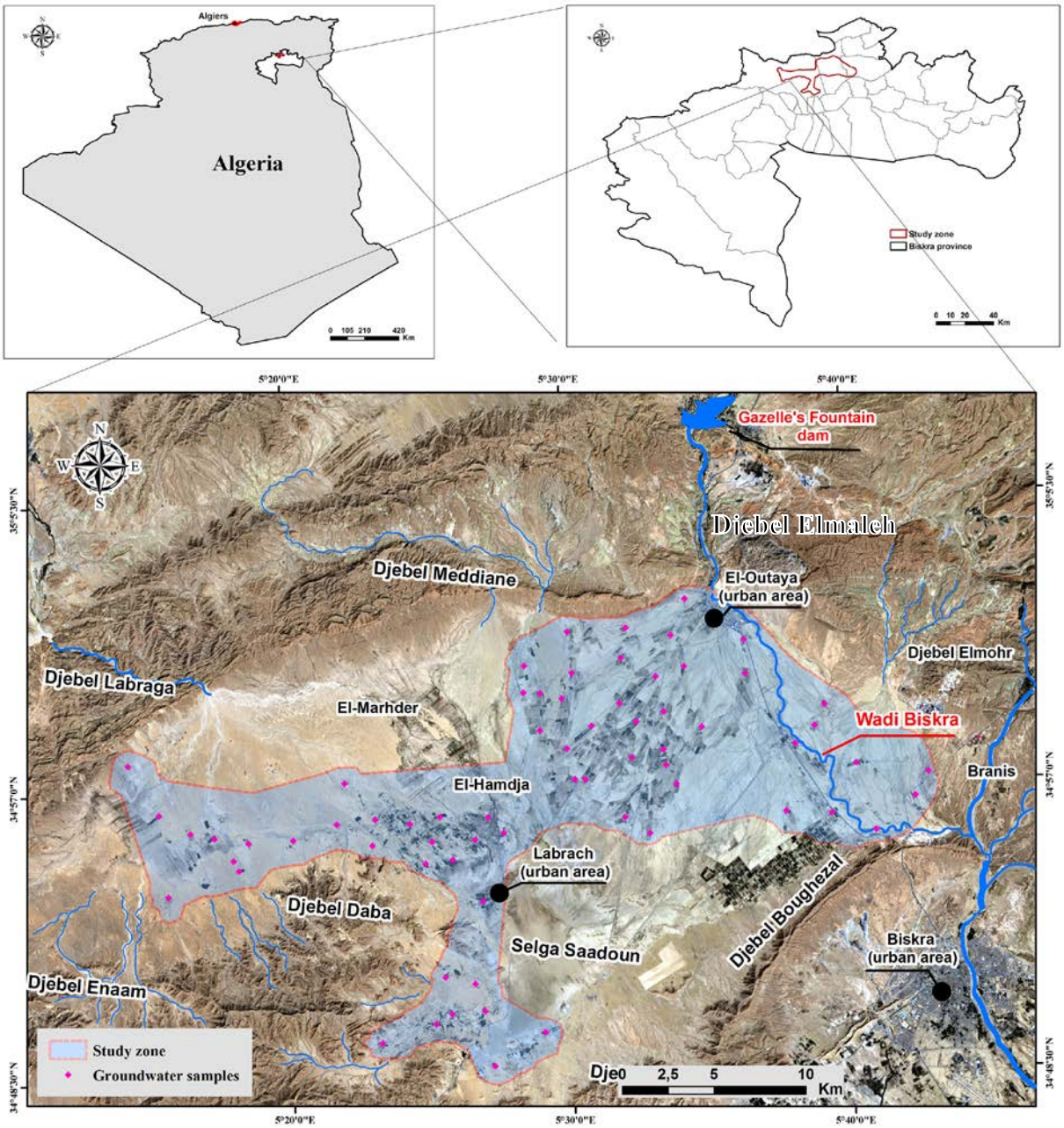


Figure 8: Geographical location of the study area

Biskra province occupies a surface area of 21671.2 km^2, with a population density of 43 hab/Km^2 and a demographic development of 2.3 %.

El Outaya plain, which is the study area, is located at 25 km in the northwest of the capital city of Biskra. The studied zone (Figure 8) extends over 405 km^2 of area in the basin of El Outaya. It lies between latitude 34°48' N - 35°03' N and longitude 5°09' E - 5°45' E. It is bounded at the north by Djebel Labraga, Djebel Meddiane and Djebel Elmaleh, at the south by Djebel Boughezal and Djebel Fouinissa, Bled Chicha and Djebel Elmohr at the east and Djebel Daba, Djebel Enaam at the west.

2. Climatological context

Climatic conditions play a determining role in the recharge of aquifers and precipitation is the essential factor. In arid areas, they intervene, especially, by the showers that generate floods and effective rainfall. The different aspects of precipitation are highly influenced by other climatic conditions and the most important are evapotranspiration and temperature.

2.1.Precipitation

Rainfall in Biskra region is characterized by irregularities both inter-annual and intra-annual. The annual rainfall recorded over a period of 18 years (2000-2018) gives an average rainfall of about 118.75 mm. In addition, there is a peak of about 297 mm during the year 2014 while the low values are registered in 2017 and 2018 with 42 mm and 42.33 mm, respectively (Figure 9).

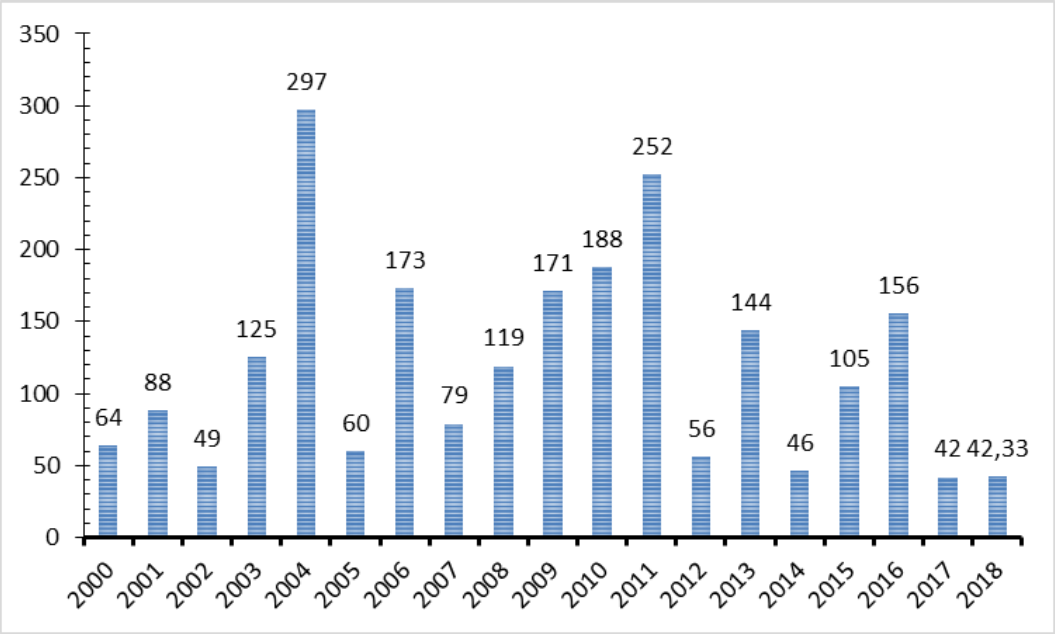


Figure 9: Inter-annual variation of precipitation in Biskra region (2000 – 2018)

Analysis of the average monthly rainfall shows that the autumn months are the most rainy with a cumulative of 41.9 mm (35.37 %), followed by the spring months (March, April, May) with a cumulative rainfall of 41.3 (34.87 % mm). The winter season contributes at about 24.4 % (28.9 mm). The three months of summer commonly defined (June, July and August) contribute only with 6.4 mm (5.4 %) of the annual rainfall (Table 2).

Analysis of the monthly variations in precipitation shows that the values of the standard deviations and coefficients of variation vary considerably. Thus, the values of the standard deviations exceed the mean values, leading to coefficients of variation exceeding 100%, which explains the high inter-annual and intra-annual variability and irregularity of precipitation.

Table 2: Monthly average (P), standard deviation (SD) values and variation coefficient of precipitation (CV) in Biskra region (Period of 2000 – 2018)

Month	Jan.	Feb.	Mar.	Apr.	Mai	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
P (mm)	14.6	5.5	13.9	15.7	11.7	3.3	0.7	2.4	15.2	14.9	11.8	8.8
SD	22.7	8.4	21.2	21.0	16.2	6.7	1.2	3.5	13.4	20.0	13.8	10.7
CV %	155.8	154.2	151.8	133.5	138.3	202.8	182.6	143.5	88.3	134.3	116.9	122.4

2.2.Temperature

The average monthly temperatures recorded in Biskra region over 18 years (2000 - 2018) showed that the thermal regime is characterized by high temperatures in summer and relatively low temperatures in winter. The highest temperatures are recorded during the months of July and August, where they reach a maximum of 35.1 °C on average. On the other hand, January is the coldest month with an average temperature of 8.7 °C.

The analysis of monthly temperature over the same period (2000 - 2018) shows the low variation of temperatures. This is characterized by low values of the standard deviations (SD) of each month, as well as the value of the coefficient of variation (CV %) (Table 3).

Table 3: Monthly average (P), standard deviation (SD) values and variation coefficient of temperature (CV) in Biskra region (period of 2000 – 2018)

Month	Jan.	Feb.	Mar.	Apr.	Mai	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
T (°C)	8.7	13.3	17.5	21.6	26.5	31.2	35.1	34.1	29.1	24.1	16.9	12.6
SD	5.4	1.5	1.2	1.1	1.5	1.6	1.0	0.9	1.0	1.8	0.8	1.0
CV %	61.7	11.5	6.9	5.3	5.6	5.2	3.0	2.5	3.5	7.3	4.7	7.7

The observation of relative to the ombrothermic diagram; shows that the climate of the region of Biskra is characterized by a single dry season extending throughout the year (Figure 10). This is one of the parameters for the climate of arid zones, in addition to the high evaporation and irregularity in the rainfall regime.

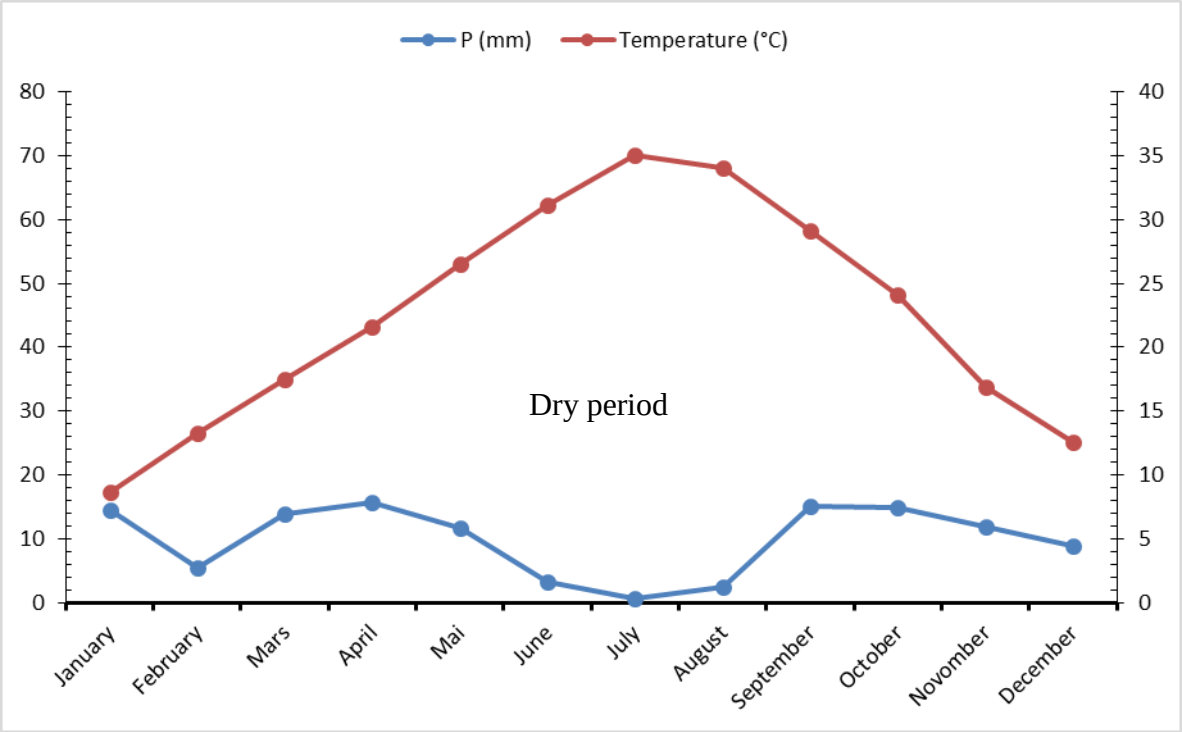


Figure 10: Ombrothermic diagram of Biskra region (period of 2000 – 2018)

2.3.Evapotranspiration

Evapotranspiration is defined as the maximum possible value of evaporation under given climatic conditions. It is the result of two phenomena: one physical: evaporation, the other biological: transpiration. For the estimation of this parameter, the Thornthwaite formula is used. This method allows one to calculate the potential evapotranspiration from the data of precipitation (P in mm), temperature (T °C), the correction factor that depends on the latitude and longitude (F(m,ϕ)) and the monthly thermal index according to the following equations:

$$ETP\ (mm) = 16 * \left(\frac{10 * T}{I}\right)^{a} * \mathrm{F(m,\phi)}$$

$$I = \sum_{1}^{12} i = \sum_{1}^{12} \left(\frac{T}{5}\right)^{1.514}$$

$$a = 0.016\, I + 0.5$$

The ETP (mm) values estimated from climate data using the Thornthwaite method over a period of 18 years are generally higher than the precipitation, only for the month of January. The average annual ETP value is very considerable (1200.41 mm).

Table 4: ETP (mm) estimation results using Thornthwaite method (2000 – 2018)

Month	Jan.	Feb.	Mar.	Apr.	Mai	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
T °C	8.7	13.3	17.53	21.6	26.5	31.2	35	34	29.1	24.1	16.9	12.6
i(moy)	2.4	4.41	6.7	9.2	12.5	16	19.1	18.3	14.4	10.8	6.3	4
F(m,ϕ)	0.9	0.8	1	1	1.2	1.2	1.21	1.2	1	0.98	0.9	0.9
P (mm)	14.6	5.5	14	15.7	11.7	3.3	0.7	2.4	15.2	14.86	11.8	8.8
ETP (mm)	9	16.5	39	69.5	127	189.7	257.9	228	137.1	82.4	29.9	14.2

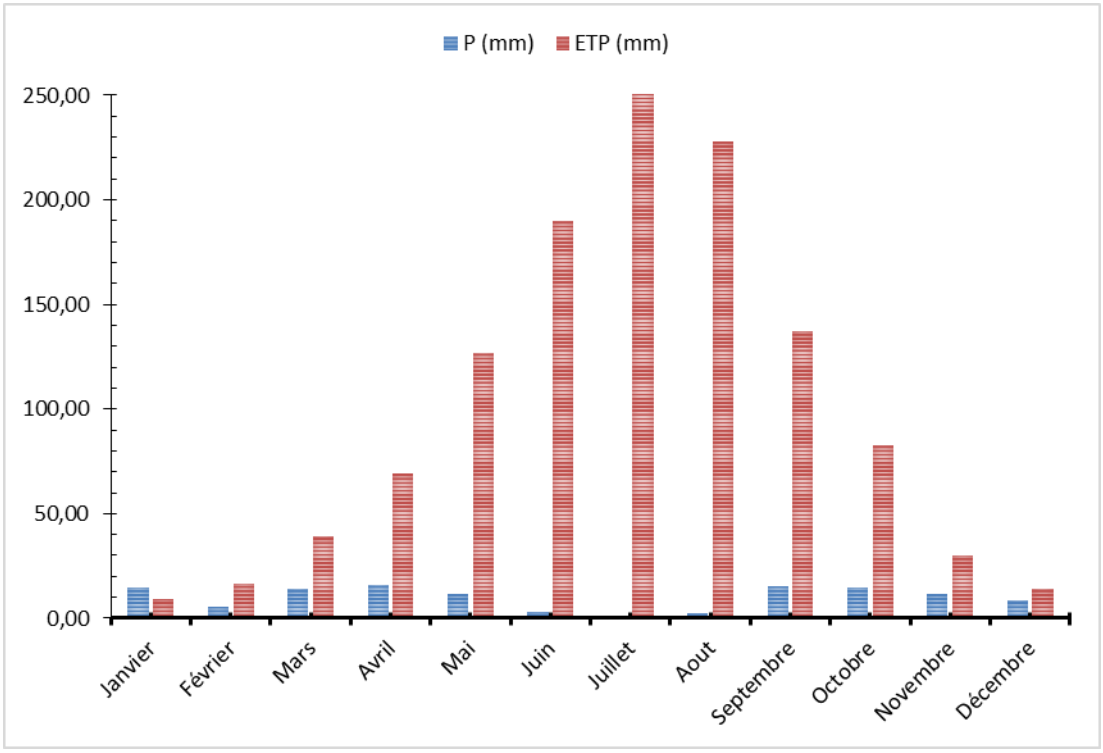


Figure 11: Average monthly estimated evapotranspiration (ETP) compared to the average monthly precipitation (P) in Biskra region (period of 2000 – 2018)

3. Hydrographic network

The Biskra region is crossed by several wadis, which constitute a simple hydrographic network that works only in winter or during exceptional rains and they are a part the large Saharan watershed of Chott Melghir (Figure 12):

- Wadi Djeddi is the most important wadi in the region, presents the drainage area of 9130 km^2. It constitutes a runoff water collector of a significant area of the southern hillside of the Saharan Atlas.
- Wadi El Arab is located in the eastern part of the region and it takes its source from the mountains that constitute the eastern part of the Aurès and flows into the depression zone of Chott Melghir.
- Wadi El Abiod is located in the middle of the region on which Foum El Gherza dam is built.
- Wadi Biskra which crosses El Outaya plain contains many streams that collet runoff water from the southwest Aurès Mountains.

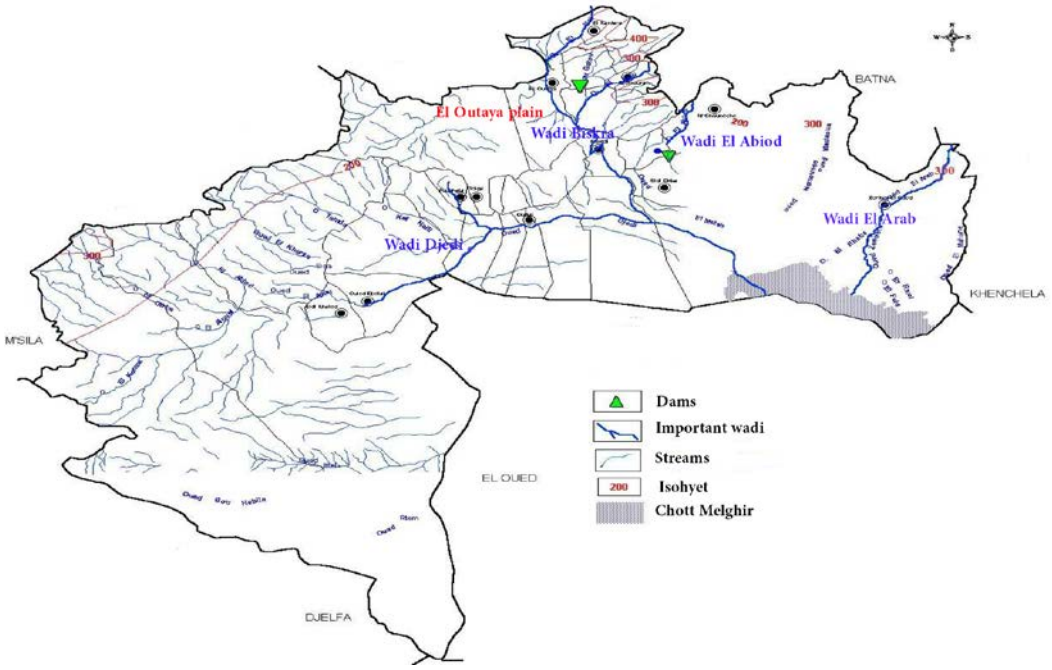


Figure 12: Hydrographic network of Biskra province

4. Geological and hydrogeological context

The geology and hydrogeology of the study area were described by several authors: Guiraud (1990); Chebbah (2007); Haouchine (2010); Sedrati (2011); Boudjema (2015) and Chebbah (2016); Chebbah et al. (2008).

4.1. Geology

El Outaya Plain is characterized by a very complex and highly tectonized geology. It occupies a vast post-Miocene middle syncline, affected by minor anticlinal wrinkles. It appears as a stack in the form of sandy-clay lenses, filling a basin slightly elongated from the east to the west. The

Mio-Pliocene filling constitutes a regular topographic surface from north to south and the Quaternary deposits are is formed of detrital filling materials.

4.1.1. Lithostratigraphy

The Biskra region is characterized by sedimentary terrains, from the Barremian at the base to the Quaternary layer at the top. The stratigraphic scale below show the main strata found in this region from the oldest to most recent.

4.1.1.1. Mesozoic

The Mesozoic forms the main dominant reliefs throughout the study area, as in the whole Saharan Atlas. It is largely composed of Cretaceous terrains, where clay-carbonate sedimentation dominates.

4.1.1.1.1. Triassic

The Triassic is the oldest geological formation in the study area. It outcrops in the north of El Outaya (Djebel El Maleh). It is characterized by the conservation of rock salt in outcrop, which forms the mass of this mountain. It is associated with shreds of purple marls abundant only at its eastern edge, in the middle of which are found the frequent crystals of this facies, such as: aragonite, anhydrite, dolomite, bi-pyramid quartz and hematite.

- Lias

The Lias is unknown in normal position and when it is encountered, it is packed in the Triassic in the form of shale flakes (Menaa, Aurès center) or dolomites (El Outaya). West of the Aurès (Eastern Saharan Atlas), it is represented by yellow and blue limestones with Belemnites and Foraminifera of the Lower Toarcian. On the other hand, on the northern margin of the Saharan platform, it is formed of oolitic limestones and dolomites and becomes anhydrite towards the south.

4.1.1.1.2. Jurassic

a- Kimmeridgian

Corresponds to the oldest terrains recognized in normal position in the Aurès massif where it forms a buttonhole between the valleys of Oued Abdi and Oued Labiod (anticline of Djebel El Azereg). It splits into two quite distinct sets:

- A basal carbonate ensemble with 250 to 300 m thick, which relates to the Lower Kimmeridgian.
- A marly ensemble at the top with 400 m thick, monotonous and admitting in its median part intercalations of marly limestones with lamellibranches.

b- Portlandian and Berriasian

This has about 250 to 300 m thickness that occurs under two distinct facies forming between them a progressive lateral passage with a calcaro-marly facies to the NW and a dolomitic facies to the SW of the Aurésian massif.

4.1.1.1.3. Cretaceous

Two series stand out in the Cretaceous formations: a basal sandstone series from the Lower Cretaceous (Berriasian - Albian) and a marl-limestone series from the Upper Cretaceous (Cenomanian - Maestrichtian).

a- Lower Cretaceous

- Valanginian: It outcrops only in the Gebel El Azreg anticline, in the shape of an aureole all around the Jurassic and is essentially composed of greenish and yellowish marls as well as small quartzite benches.
- Hauterivian: It outcrops in the Aurès (Djebel El Azreg anticline), drawing a regular aureole where it forms around depressions occupied by Valanginian marls, limestone or sandstone escarpments with limestone pastures alternating with sandstone banks. The thickness of this stage is considerable and can reach 1000 m.
- Barremian: In the Aurès, this stage is essentially quartz and presents a thickness of 850 to 900m. It outcrops to the southeast of El Outaya where it is formed by red sandstones with clayey interlayers. To the south of the plain, at Koudiat El Leham (Debel Boughezal anticline) can be seen on 20 m thick slightly sandstone red clays without marine fossils but with fossilized wood, demonstrating the continental and subaerial origin of this formation.
- Aptian: It outcrops in all the anticlines of the Aurès where it is characterized by a lagoonal sedimentation regime with some marine invasions. At Chaïba (south of El Outaya), it is represented by clays, versicolored marls, anhydrites, dolomitic limestones and dolomites over a thickness of about 60m. At Koudiat El Leham, it is represented by orbitoline limestones and annelid tubes, over a thickness of about 20 meters.

- Albian: It begins in the Aurès with alternating sandstone and marl often reaching 150 to 200m thick, becoming variegated to reddish towards the south-west. Thereafter, limestones tend to replace sandstones, while pelagic fauna make their appearance. In the South of El Outaya, at Koudiat El Leham, and northeast of Chaïba, a lower Albian under a continental facies, formed by alternating sandy sandstones and marls, and an upper Albian under a sub-reefal facies, formed by alternating dolomitic limestones and lumachelles.

b- Upper Cretaceous

The Upper Cretaceous forms the main part of the Mesozoic outcrops in the region. It is represented by crystalline and dolomitic limestones in very thick layers, marly and gypsum lagoon intercalations very numerous in the West (Djebel Gouara and on the northern flank of Djebel Boughezal).

- Cenomanian: It outcrops on the northern flank of Djebel Boughezal, its thickness is 300 to 400 m of grey or white limestone regularly alternating with greyish marls, sometimes gypsum. It is characterized by an abundant fauna (echinoids, ostracians) especially in the upper part which is mostly marly-limestone while the lower part is generally marly.
- Turonian: It outcrops northwest of El Outaya. It is uniformly represented throughout the mountainous area by a massive level of 200 to 300m thick composed of crystalline limestone, marly limestone and dolomitic limestone.
- Senonian: It exists in all the synclines of the Aurès as well as around the anticlines. It is very thick (2000 m on average) and occupies considerable surface areas on outcrops. In Djebel El Maleh (NE of El Outaya), the core of Triassic drilling has brought to outcrop a small island of black marl and chalk limestone fossiliferous enough located at the bottom of this mountain. To the north and east of these marls, outcrops massive limestones, because of their position above the Campanian and their facies, attributed to the Maestrichtian. Figure 6 shows that the Senonian is formed of fissured dolomitic hard limestone.

  On the northern slope of the Boughezal Anticline, the Senonian is represented by important Maestrichtian limestone masses, about 400 to 500 m thick.

4.1.1.2. Cenozoic

a- Paleogene (Nummulitic)

- Lower Eocene: It outcrops to the east-southeast of Djebel El Maleh. It is represented by several dozen meters of yellowish greenish marls slightly gypsum and clayey limestones with lamellibranches. A bank of micro-conglomerate limestone with Pecten, marks the implantation of marine sedimentation. This formation also outcrops north of Tolga and east of El Outaya (Djebel Ahmar) where it is represented by limestones rich in black flint and often fissured. This is confirmed by Figure 13 and Figure 14 boreholes, which show that the Eocene was often formed of fissured white limestone.

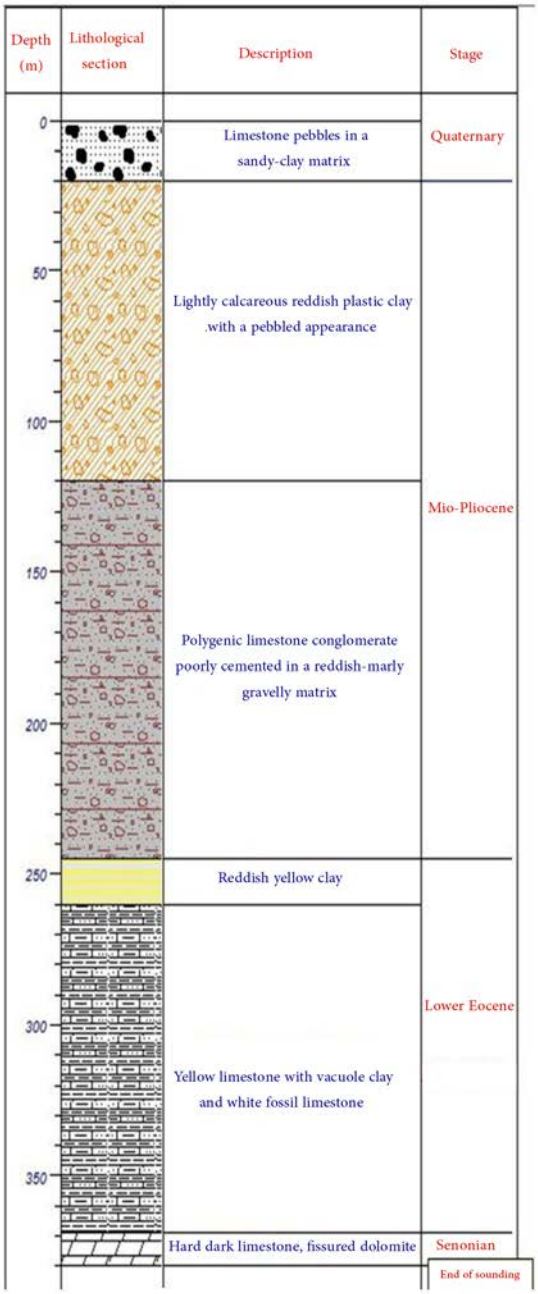


Figure 13: Sounding in the northeast of El Outaya (Djebel El Maleh)

- Middle Eocene:

It outcrops to the east of El Outaya, on the northern flank of Djebel Ahmar. It is characterized by a lagoon sedimentation of different types: clays, gypsum, anhydrites and limestones; its thickness varies from 100 m to 400 m. The lithological section of the El Mazouchia borehole (Figure 14) shows that the Middle Eocene is composed of alternating limestones and marls at the top and limestones at the base, all over a thickness of about 270 m.

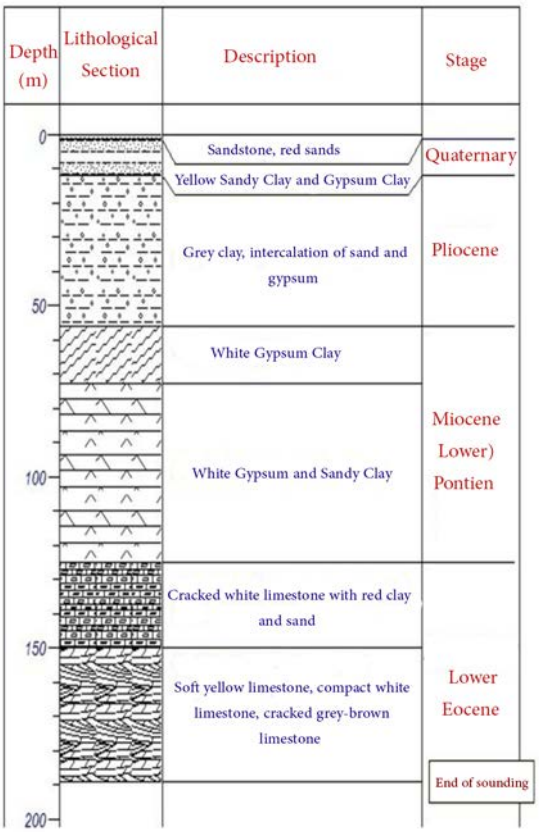


Figure 14: Sounding in Labrach zone

- Upper Eocene: The Upper Eocene does not outcrop in the plain of El Outaya but it is found in some places in the Aurès.

b- Neogene: The Neogene is subdivided into two sub-stages: the Marine Neogene (Lower Miocene) and the Continental Neogene (Upper Miocene and Pliocene).

- Lower Miocene: It outcrops to the north of El Outaya, on the northern flank of Djebel Mediane, DjebelMagraoua and to the east of the Branis region. It is represented by red limestones with gypsum and flint. In the basin of El Outaya, it is formed in the east by limestone layers on the first slopes of the Aurès massif, and in the west by molasses rich in fossils, covered by marls with little fossiliferous. Further north, at

Sebâa Mgataâ, the Miocene is formed by marly-limestone molasses rich in pectinids and molluscs. Above, come marls without fossils with diffuse gypsum. The total thickness is about 200 to 300 m.

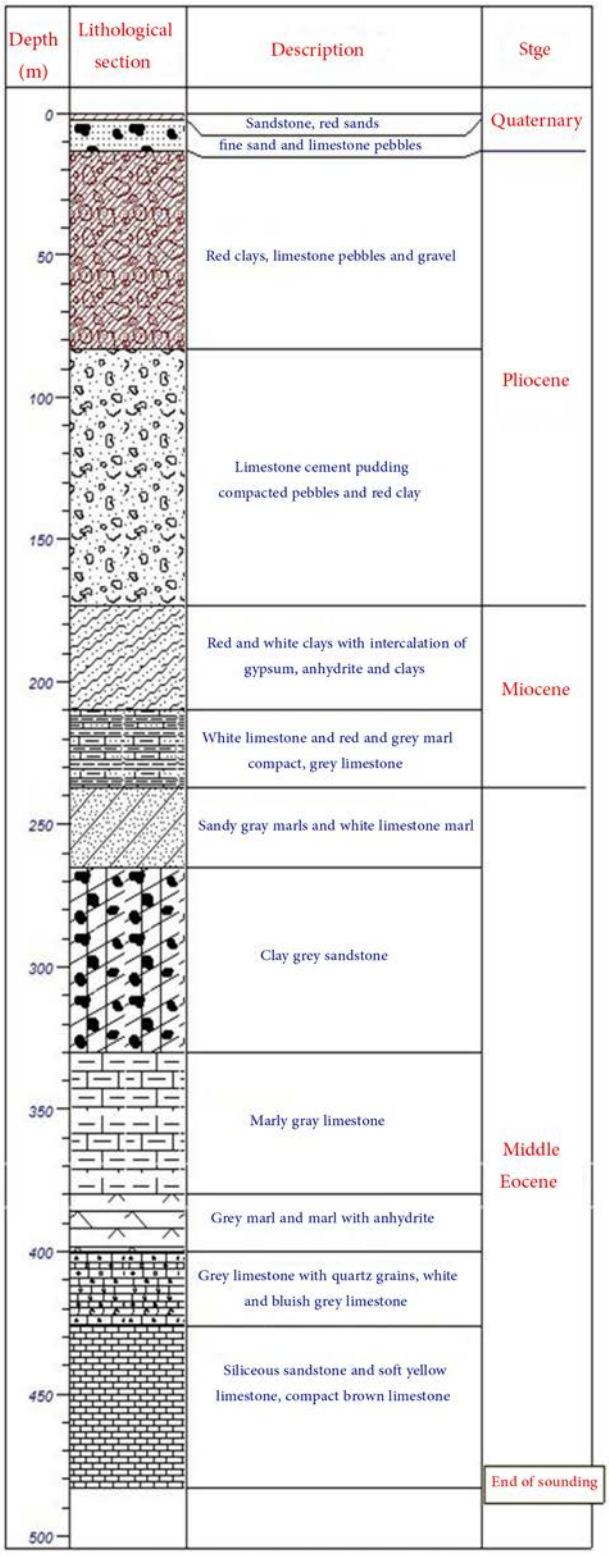


Figure 15: Sounding in Bled El Mazouchia

- Upper Miocene: it outcrops north of El Outaya in the form of reddish brown marly limestone layers. To the west, on the southern flank of Gebel Maghraoua and especially on Gebel Moddiane, the Upper Miocene consists of conglomerates and breccias up to several hundred meters thick. In the East of El Outaya, red layers have been observed overlying on the fossiliferous marine Miocene; these are red clays, coarse sandstones with rare gravel banks, poorly cemented puddings and coarse puddings that he attributes to the Upper Miocene. The thickness of this ensemble is about 500 m.
- Pliocene: The Pliocene forms a series of very continuous outcrops to the north and east of the El Outaya plain. Good sections can be observed immediately to the northwest of the plain. East of El Outaya, the Pliocene only outcrops discontinuously in the strongly tectonized zone that separates Djebel El Maleh from Djebel El Mohr.

c- Quaternary: Recent soils are as widespread on the periphery of the Aurès as they are rare in the center of the massif; there they are represented by scree slopes and terraces, while on the periphery they form large alluvial layers of considerable thickness. It is necessary to distinguish in the region of El Outaya, on the one hand the piedmont established on the Neogene, very dissected and drained by several wadis, and on the other hand, the plain; the passage from one area to another being generally abrupt.

- Piedmonts:To the north and east of El Outaya, various small juxtaposed basins show an important stony overburden, between which the alluvial plains are limited to narrow strips along the main wadis. To the north Gazelles fountain dam, the plain formed of sandy-clay alluvium with conglomerates at the base and is covered in its eastern part by a glacis at the outlet of the wadi Hassi ben Tamtam. To the northwest of El Outaya, at Bled Salaouine, two glacis with gypsum or gypso-calcareous cover appear. The origin of the gypsum, more than 1 m thick, which participates in these incrustations, seems to be sought both in the leaching of Miocene bedrock and in the wind contributions, given the proximity of the Lower Sahara.
- El Outaya plain: This plain, established for the most part on neogenous terrains, is formed by alluvium in its major part. The western part is invaded in its almost totality by dunes still mobile, while in the eastern part, the wadis are encased in small alluvial cones near the reliefs, then in alluvium which they do not submerge any more, except in rare zones. These alluvial deposits show in the neighborhoods of El Outaya, few meters from the surface, a horizon rich in calcareous tabular concretions.

4.1.2. Tectonic

El Outaya plain is located in an area separating two geographically and geologically distinct domains: the Saharan Atlas to the north and the Sahara to the south. The passage between these two domains is done through a set of flexures, folds and faults oriented from west to east, called the "South Atlas Flexure". The South Atlasic Flexure developed during the Pliocene and post-Pliocene paroxysmal phase of the Aurès surrection. This phase is also responsible for all the deformations of the neogenic continental (Mio-Pliocene).

There are two systems of faults with different directions: the NW-SE oriented faults and the SW-NE faults; the first type is known throughout the Saharan Atlas. Locally, El Outaya Plain appears to be bounded to the north by an anomalous contact of general east-west direction; a major NW-SE fault crosses the entire region. Other faults of the same importance and practically the same direction affect all the carbonate terrains. This plain is located in a syncline affected by a set of folds and faults, as well as a succession of small synclines and anticlines (Figure 16).

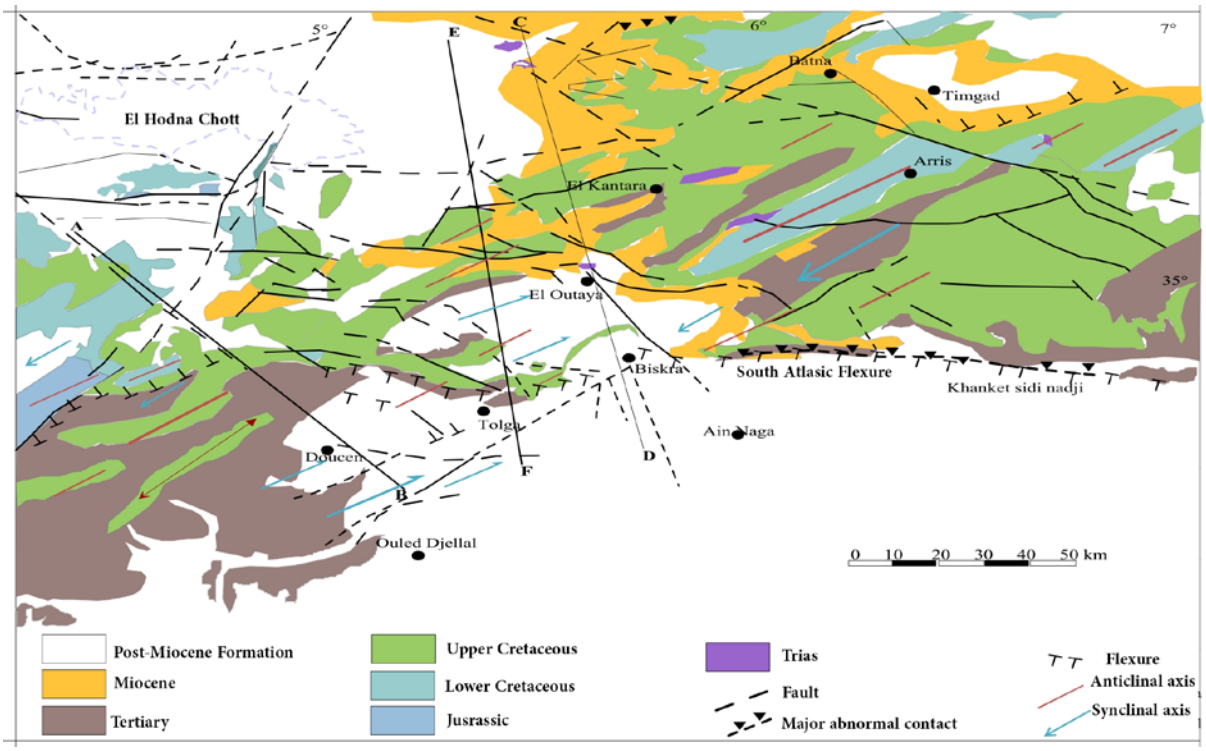


Figure 16: Tectonic sketch (Extracted from the hydrogeological map of Biskra, Salem et al. 1979)

4.2. Hydrogeology

The South of Algeria contains many aquifers with great extensions. Their depth differs from a region to another and can reach more than 500 m (Sedrati 2011).

4.2.1. The main aquifers in Biskra region

The hydrogeological studies carried out in Biskra region have shown the existence of several aquifers of different ages (Haouchine 2010; Sedrati 2011; Boudjema 2015). According to the hydrogeological cross sections (Figure 17) and the hydrogeological map (Figure 19), these aquifers occur in the Quaternary, Mio-Pliocene, Lower Eocene, Upper Senonian (Maestrichtian) and Albian formations (The direction of the hydrogeological sections are presented in Figure 16). Senonian and Turonian aquifers are also present in the region of Biskra, but they are rarely exploited because of their low extension. Boudjmaa (2015) attested that El Outaya plain presents a lack of knowledge concerning the hydrogeological aspect despite the important number of wells drilled in the last years.

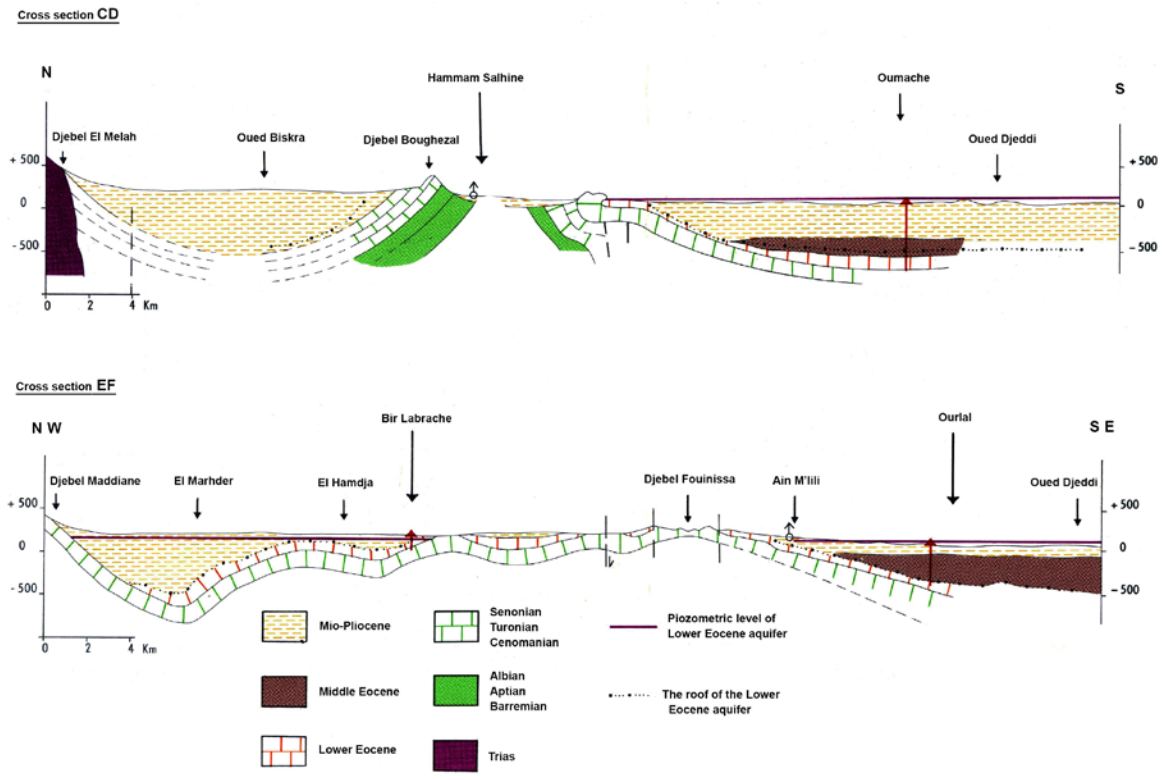


Figure 17: Hydrogeological cross sections (Extracted from the hydrogeological map of Biskra)

The different studies describing the hydrogeological aspect of Biskra region have mentioned the presence of three different hydrogeological sets: the Quaternary aquifer, the Terminal Complex (TC) aquifer and the Intercalary Continental (IC) aquifer.

4.2.2. Quaternary

This quaternary formation is essentially represented by the water table of the Biskra wadi and the Djeddi wadi, which represent the most important resource for the irrigation of the palm groves of Sidi Khaled and Ouled Djellal and a large part of the drinking water supply to the city of Biskra.

The Quaternary formations cover a large part of the region and contain the Biskra water table, which is very common in Tolga palm groves and at the level of the wadis (Inféro-Flux water table). Its depth ranges between 10 m and 40 m and its reservoir is composed of pebbles, sand dunes and sandy and stony alluvium of the current beds of the wadis, which indicates heterogeneity of the reservoir. To the west of Wadi Djeddi, the aquifer is composed of sandy and clayey alluvium, while to the east the puddings are gradually relayed by a sandy or clayey deposit. The bedrock of this aquifer is formed by a thick clay layer, with intercalation of few layers of sand, gravel and marl, appearing as sand lenses in discordance with the clay layers.

The aquifer recharge, particularly in El Outaya plain, is provided by runoff from the Aures massifs (Ain Touta). Concerning the Inféro-Flux aquifers, the recharge is ensured during the floods of the wadis. The runoff causes the recharging of the alluvium. Irrigation water and waste water are another source of aquifer recharge.

4.2.3. Terminal Complex (TC)

The Terminal Complex includes different formations ranging from the Upper Senonian to the Mio-Pliocene. This name (Terminal Complex) represents the fact that several aquifers are grouped together within the same reservoir. These aquifers are the carbonated Eocene (lower Eocene limestone water table) and the sandy Mio-Pliocene (sand water table).

4.2.3.1. Mio-Pliocene water table

The Mio-Pliocene aquifer occupies, generally, the depressions forming the plains bordered by the massifs of Djebel Guedare and Djebel Boughezal in the North West and Djebel Ammar Khadou in the North East. It consists of alternating clays, sands, gravels and conglomerates with gypsum. It lies on a substratum formed by the impermeable layers of the marine Miocene and the middle Eocene.

The aquifer outcrops to the northeast of the Biskra region, north of the Mkhedma Oasis, where it consists of red pudding sands. This same formation is found south of Wadi Djeddi and southwest of the Doucen towards Ouled Djellal. Note that near the borders of the Ziban Mountains the sandy clays are in contact with the limestones of the lower Eocene. This indicates exchanges between sand and fissured limestone aquifers, especially when the thickness of the clays is low. In the eastern part of the region, this water table is divided into two aquifers separated by a thick layer of clay and sandy clay, one deep called the Pontian and the other moderately deep which is the Mio-Pliocene water table known in this region.

In the western part of El Outaya plain, the Mio-Pliocene aquifer is less important compared to the eastern part. It consists of a set of pebbles, scree, and sandy to sandy clay. In the central part of the plain, the aquifer is formed by sandy, sandy clay, sometimes marly and pebbled puddings. The thickness of this layer is also less important than that of the eastern part. The southern part of the plain is formed by saliferous sands, with clayey intercalations, which deteriorate water quality in this region.

The substratum of this aquifer is constituted of limestones of the Lower Eocene in the western part, on the other hand, in the central and eastern parts, it is formed by marly Miocene. In proximity of the points, where the aquifer is in contact with the limestones of the Eocene, the latter sinking is under a thick formation of the Mio-Pliocene in the northern part and in the center towards the south of the plain.

The general flow is in north-south direction. The recharge of the aquifer is ensured mainly in the northern part, by the infiltration of the waters of wadi Biskra during the flood period, or by the formations of the Mio-Pliocene (Aurès Region). The eastern part also contributes to the recharge of the water table by the Pliocene puddings, on the other hand, in the western part, the recharge is ensured by the limestones of the lower Eocene.

4.2.3.2. Lower Eocene water table

This aquifer is located in its entirety in the region of Biskra. The western part is more exploited than the eastern part, because of its smaller depth. The latter reaches a maximum of 200 m in the West while it exceeds 400 m in the East.

The roof of the aquifer is composed of Mio-Pliocene clayey-sandy formations to the north and Middle Eocene gypsum marls to the south, while its reservoir is essentially constituted of

fissured white limestone from the Lower Eocene and dolomitic marly limestones from the Upper Senonian.

Its depth varies from 10 m to 60 m in the Tolga zone. Towards Oumeche and Mllili, the depth increases to 150 m. Around Doucen, the depth exceeds 200m. It exceeds 400 m at the level of the El Outaya plain.

The flow is characterized by two directions:

- The first one directed from North to South,
- The second is in the direction South-West North-East.

4.2.4. Intercalary Continental (IC)

It is a very important aquifer composed essentially of sandstone and marl of Albian and Barremian age, with a depth that varies between 1,600 m and 2,500 m and an average flow of 80 l/s gushing out (Figure 18 and 19). Its exploitation is very expensive, because of its depth, in addition the temperature of the water can exceed 60°C.

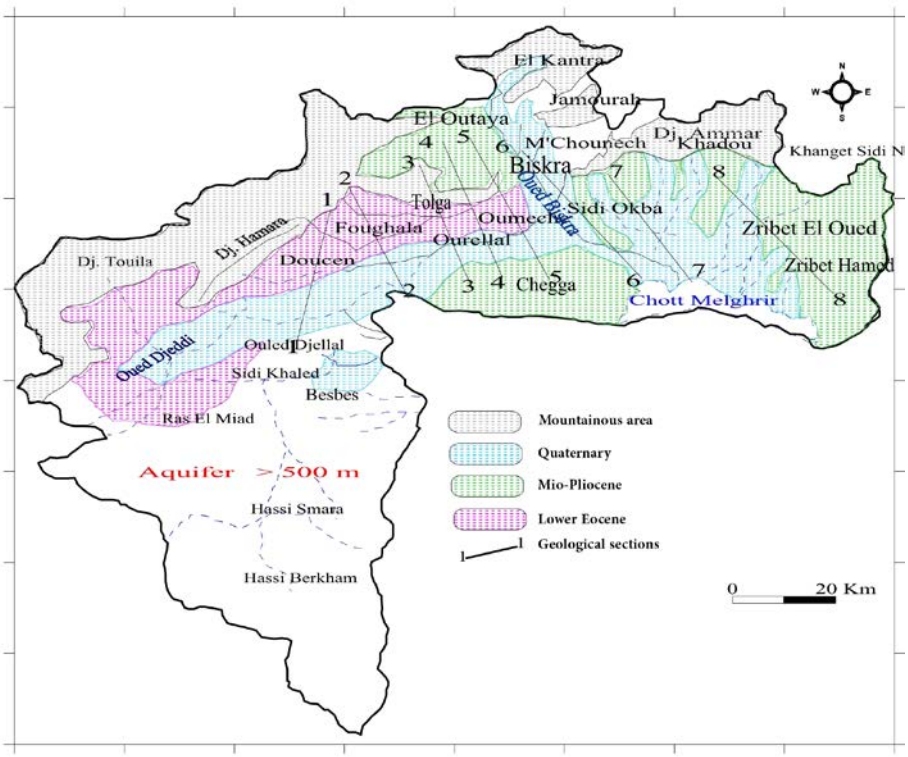


Figure 18: Schematic representation of the extension of the main exploited aquifers in Biskra region (Sedrati 2011)

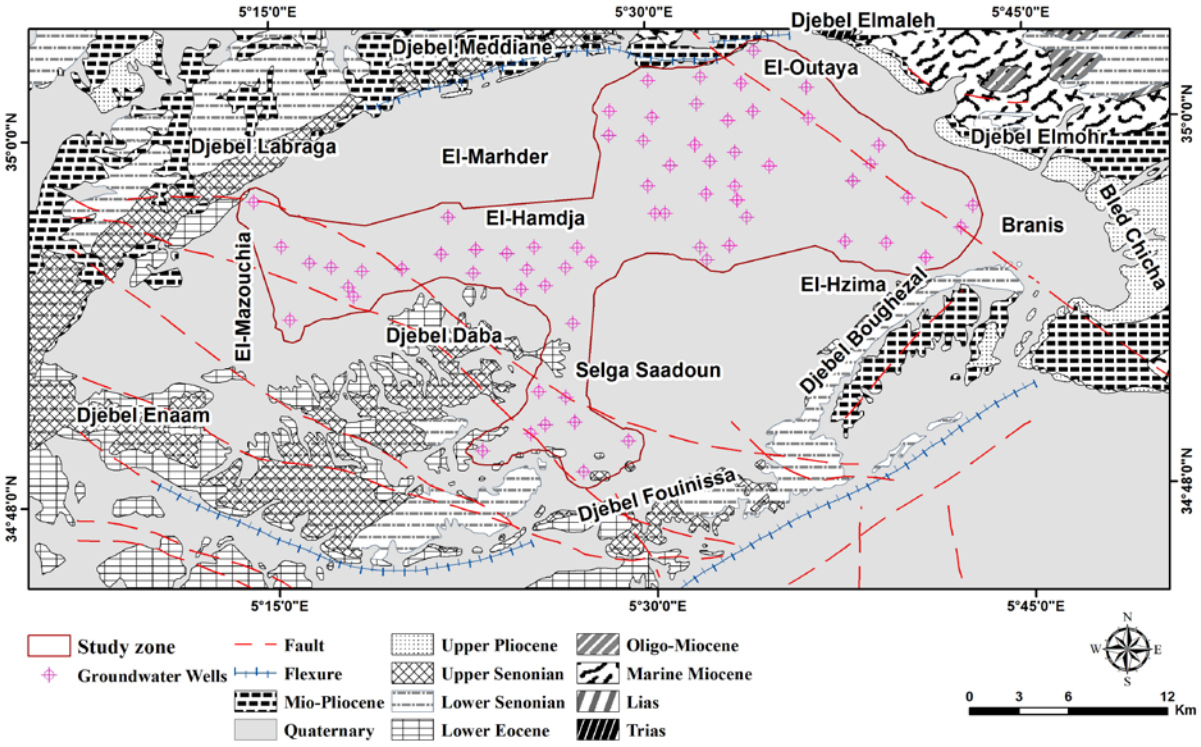


Figure 19: Hydrogeological map (Digitalized and adapted from the hydrogeological map of Biskra)

5. Pedology

5.1. Soils of Algeria

Soil is an element of the natural environment linked to structure, vegetation, climate and human activities. The nature of the soil is also an essential factor in the agricultural development of a region.

The Durand's classification of soils in Algeria is based firstly on the climatic factor that played the essential role in the formation of the soil (wind, rain, temperature), secondly on the degree of soil evolution (number of differentiated horizons), and finally on the degree of leaching. Soils in Algeria can be grouped, according to the climatic factor dominating their formation, into three large groups of zonal soils, each with its own characteristics determined by the environmental conditions controlling the pedogenesis processes (Benchetrit 1956):

- Saharan soils, where the dominant factor in soil formation is wind;
- Soils of arid and semi-arid regions, where the zonal soil would be due to the wind-rainfall climatic equilibrium;
- Soils in humid tellian regions, where moisture and vegetation become the main factors in the formation of zonal soils.

In addition to these zonal soils, a whole series of azonal soils are classified separately. It would be more accurate to call them local soils, because these soils are not formed under the influence of zonal factors, but rather of local factors that create microclimates within a climatic zone that give the soil special conditions for its evolution.

5.2. Soils of the arid and semi-arid regions

The soils of arid and semi-arid region can be fixed by steppe vegetation, but their evolution is slowed down by the lack of percolating water. Rainfall is not strong enough to modify the absorbent complex, which remains in its original state, hence, the stability of the profile of these soils. The bedrocks are generally calcareous: eolian alluvium deposited on pulverulent limestone. However, they constitute a differentiating factor giving, according to their nature; soils containing limestone or gypsum (calcareous soils), and soils, which do not contain any soils in equilibrium, but the latter are rare (Benchetrit 1956).

Calcareous soils: They have only one thinly differentiated horizon. They are more or less rich in limestone, their absorbent complex is saturated by the Ca^{+} ion and their pH is always between 7 and 8. They contain low quantities of organic matter (between 0.3 and 1%), but no soluble salts. In these soils, the limestone is mainly presents in the sand fraction and has little influence on their dynamics. They are, therefore, typical calcareous soils. Beside the typical calcareous soils, two subtypes can be observed as gypsum soils and soils formed at the expense of scree slopes of the two previous ones, always presenting the same characteristics: light texture, good permeability but low water retention capacity due to their relative poverty in colloids.

Soils in equilibrium: They have the same characteristics as calcareous soils, but formed on non-calcareous bedrock. Their absorbent complex is saturated in Ca^{+} ion, there is no movement of substances and they have only one differentiated horizon. These soils are rare in Algeria, because calcareous or gypsum rocks cover about 90% of the surface of the arid and semi-arid zones.

Finally, in in arid semi-arid regions, in intra-zonal position, ablation or accumulation wind soils are found as leached soils. In general, the soils in these regions are not normally rich in soluble salts. On the other hand, solontchak (saline soils) are often found in poorly drained areas or those supplied with water by a superficial water table.

According to the soil map of Algeria, Biskra (Figure 20), produced by Durand and Barbut (1938) and explained by Benchetrit (1956), El Outaya plain is characterized by basic alluvium

surrounded by wind ablation soils. Saline soils (Slontchak type) are present in the synclinal depression of Selga Saadoun. The western part of the plain is characterized by wind accumulation soils (dunes) due to the wind erosion.

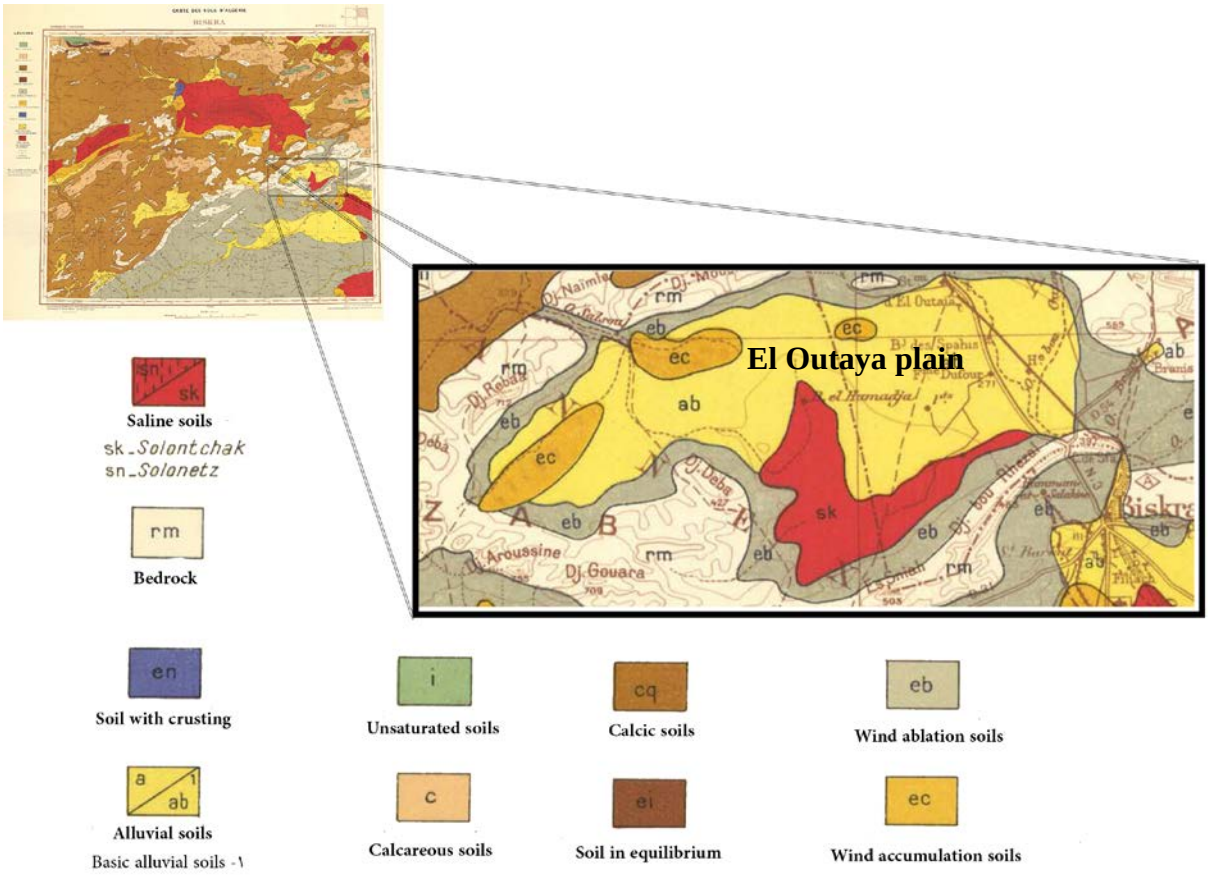


Figure 20: Soil map of Algeria adapted to the study area **(**Durand and Barbut, 1938)

Chapter III

Mapping and assessment of groundwater quality

1. Introduction

Groundwater is the major water resource for domestic, industrial and irrigation uses in arid and semi-arid regions as well as in El Outaya basin (Biskra, southeastern Algeria). However, growing population and fast expansion of agriculture are the main causes of the overexploitation of aquifers, which leads to the deterioration of groundwater quality, continuous salinization of the top soil and reducing crop productivity (Amichi et al., 2012; Bradaï et al., 2016). Agriculture irrigation consumes large quantities of groundwater, where date palm cultivation and cereals are the most common crops and surface irrigation is the main system of irrigation in the area (ABHS, 2016). Thus, evapotranspiration rates and low precipitation influence significantly groundwater resources and increase irrigation needs (Ahmadi and Sedghamiz, 2007; Shaji et al., 2018). The salinization of groundwater is one of the greatest problem affecting the sustainable use of groundwater (Ahmed et al., 2013). Application of irrigation with saline water and accompanying agrochemicals and the intense abstraction of groundwater are the main causes of the degradation of groundwater quality (Pulido-Bosch et al., 2018). Therefore, monitoring and assessment of groundwater quality seems necessary for optimum exploitation of this precious and scarce resource and for the social and agricultural sustainable development.

Water Quality Index (WQI) is an easy and effective communication tool for the evaluation of the groundwater quality for human consumption and it aims at giving a unique value to the quality of a water sample, which signifies its overall quality category, in order to compare the quality of different samples based on the obtained values. (Abbasi and Abbasi, 2014; Ahamad et al., 2018). Several authors have used WQI to assess the suitability of groundwater for domestic usage (Sadat-Noori et al., 2013; Rawat et al., 2017; Singh et al., 2014; Deepa and Venkateswaran, 2018).

Geostatistics (Journel, 1986) study of the regionalized variables spatial distribution is used for the application. Kriging is a geostatistical interpolation technique that deals with many variations such as, universal kriging (UK), simple kriging (SK), ordinary kriging (OK) and cokriging (CK). In this study, the adopted geostatistical variant for interpolation is OK, which is frequently the most used alternative (Arslan, 2012; Bradaï et al., 2016). Arslan (2012) and Bradaï et al. (2016) used OK and IK to interpolate groundwater salinity in Bafra plain (Turkey) and Lower Cheliff plain (Algeria), respectively. They used the cross validation method to check the accuracy of the results. Xie et al. (2011) applied OK to estimate the spatial distribution of soil heavy metal concentrations at unsampled points in Beijing (China). Táany et al. (2009)

employed OK to assess the spatio-temporal variability of groundwater level fluctuations in the Amman–Zarqa basin.

Hydrogeochimical methods and geostatistics are widely used to study the groundwater quality (Sajil Kumar et al., 2013). In the current chapter, 136 groundwater samples have been collected during two campaigns (dry and wet season). The aim of the chapter is to assess the quality of groundwater and its suitability for drinking and irrigation purposes in El Outaya basin based on hydrogeochemical analysis (WQI, Piper diagram, Saturation indices (SI), Sodium absorption ratio (SAR), Sodium percentage (Na^+%), Riverside diagram and Wilcox diagram) and geostatistical analysis (analyze the spatial and temporal distribution of WQI and EC using OK method).

2. Materials and methods

2.1. Groundwater sampling, laboratory and data analysis

Over two field campaigns, a total of 136 groundwater samples were collected using a global positioning system (GPS) device from 68 wells located in agricultural farms distributed along the study area. The first campaign was after the dry season (3rd to 24th September 2017) and the second campaign was after the wet season (20th April to 10th May, 2018). The samples were collected after a few minutes of pumping in 1L (one liter) polyethylene bottles are rinsed with distilled water and with the same groundwater before filling. The bottles after labelling are transported in coolers and conserved into the laboratory at a temperature of 4°C in order to avoid any chemical modification. As mentioned above, in climatic aspect of the study area, the dry season is spread out over the whole year so the terms “dry season” and “wet season” are used to make the difference between the first campaign and second campaign, respectively.

Physicochemical parameters such as temperature, pH and electrical conductivity (EC) are measured in the field using a portable multi-parameter (WTW multi 3430). The total dissolved solids (TDS) is determined by the relation TDS = 0.64 EC (Kawo and Karuppannan, 2018; Brown et al., 1970). The major cations (Ca^{2+}, Mg^{2+}, Na^+ and K^+) and the major anions (Cl^-, NO_3^-, HCO_3^- and SO_4^{2-}) are analyzed in the laboratory of the scientific and technical research center on arid regions (Biskra, Algeria) adopting standard procedures (Rodier et al., 2009). UV-visible spectrophotometer is used to measure NO_3^- and SO_4^{-2}. In this chapter, the analyses of 65 groundwater samples are employed due to the absence of certain parameter values in 3 samples.

Titration method (Mohr method) is used to determine the concentration of Cl^-. The total hardness, Ca^{2+} and Mg^{2+} are measured by EDTA titrimetric method. HCO_3^- concentrations is determined by complexometric titration method. Flame photometer (JENWAY) is considered to measure Na^+ and K^+. The reliability of the chemical analysis is checked using ion balance errors (IBE), which were generally < ±10%, most of them (80%) were < ±5% (Sun et al., 2016; Kim et al., 2015; Soumya and Kamble, 2017).

Descriptive statistics are determined using SPSS software v22.0. DIAGRAMMES software v6.51 is used for making Piper, Riverside and Wilox diagrams. Saturations indices are calculated using speciation code PHREEQC 2.8. VARIOWIN software v2.2 is taken into consideration for making and fitting omnidirectional semivariograms. ArcGIS 10.2 software with Geostatistical Analyst Extension is applied to prepare various maps. Multivariate statistical analysis is performed using statistica v13.

2.2. Multivariate statistics

Multivariate statistics refer to the methods that analyze the simultaneous effect and relationships of more than two variables (Marinković, 2008). Multivariate statistics involves principal component analysis (PCA) and factor analysis (FA).

PCA is a multivariate statistical technique used to transform a set of variables into new uncorrelated variables or principal components (PCs) with minimum loss of the original information. In general, it is objective to reduce a large number of variables into few factors, which can be easily interpreted (Dhanasekarapandian et al., 2016). PCs are the orthogonal variables, generated by multiplying the original correlated variables with eigenvector, which gives the loadings. These loadings of PCs are strong if they are > 0.75, moderate if they are between 0.5 and 0.75, and weak if they are between 0.4 and 0.5 (Trikey et al., 2017). PCs are expressed as follows:

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + \cdots + a_{im}x_{mj} \quad (1)$$

where z is the component loading, a is the component score, x is the measured value of a variable, i is the component number, j is the sample number and m is the total number of variables (Abbasi and Abbasi, 2014).

Factor Analysis (FA) is conducted after PCA, where the aim was to reduce the contribution of less significant variables generated from PCA. The new group of variables (Varifactors)

obtained by FA are extracted through rotating the axis defined by PCA, where the concept is expressed as follows:

$$z_{ji} = a_{f1}f_{1i} + a_{f2}f_{2i} + a_{f3}f_{3i} + \cdots + a_{fm}f_{mi} + e_{fi} \quad (2)$$

Herein, *z* is the measured value of a variable, *a* is the factor loading, *f* is the factor score, *e* is the residual term accounting for errors or other sources of variation, *i* is the sample number, *j* is the variable number, and *m* the total number of factors (Singh et al., 2014).

2.3. Water quality index

Water quality index (WQI) is a very useful technique for the assessment of groundwater quality for drinking purposes. It reflects the combination influence of every water quality parameter on the overall quality of water (Vasanthavigar et al., 2010; Deepa and Venkateswaran, 2018). The method is adopted for the calculation of WQI used by several authors (Yidana and Yidana, 2010; Ketata et al., 2012; Kumar and James, 2013; Shabir and Ahmad, 2015; Dhanasekarapandian et al., 2016; Deepa and Venkateswaran, 2018). The computing procedure of the WQI contains three steps. In the first step, each of the 10 parameters (TDS, pH, Cl^-, NO_3^-, HCO_3^-, SO_4^{2-}, Ca^{2+}, Mg^{2+}, Na^+ and K^+) are assigned a weight (wi) from 1 to 5 depending on their importance in the overall quality of water and possible health effects. In the second step, the relative weight (Wi) is calculated by using the following equation (Eq. 3):

$$Wi = {wi}/{\sum_{i=1}^{n} wi} \quad (3)$$

where Wi is the relative weight, wi is the assigned weight of each parameter, and n is the number of parameters.

In the third step, a quality rating scale (qi) of each parameter is computed according to the following equation:

$$qi = {C_i}/{S_i} \times 100 \quad (4)$$

Herein, qi is the quality rating scale, C_i, S_i are respectively, the concentration and the World Health Organization (WHO) standards for drinking water (WHO, 2011) of each parameter in each groundwater sample (mg/l).

For computing WQI, the sub index (SI_i) is determined first for each parameter (Eq. 5), which is then used to calculate the WQI for each sample (Eq. 6)

$$SI_i = \sum Wi \times qi \quad (5)$$

$$WQI = \sum SI_i \quad (6)$$

The calculated relative weights, the assigned weights, and the WHO standards are given in Table 10 (for HCO_3^-, WHO standards were not available, because of that, the US Public Health Service value is used). The computed WQI are classified into five categories as follows: Excellent water (WQI < 50); Good water (50 ≤ WQI < 100); Poor water (100 ≤ WQI < 200); Very poor water (200 ≤ WQI < 300); Water unsuitable for drinking (WQI ≥ 300).

2.4. Ordinary Kriging (OK)

Geostatistical methods are reliable techniques for mapping and interpolating groundwater qualitative parameters (Rawat et al., 2016; Maroufpoor et al., 2017). The key function in geostatistics is the variogram (γh), which expresses the spatial dependence between neighbouring observations (Eldeiry and Garcia, 2011; Arslan, 2012). It can be defined as half the average squared difference between the attribute values at all points separated by a distance h as follows:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (7)$$

Herein, $\gamma(h)$ is the experimental semivariance value for all pairs at a lag distance h; $Z(x_i)$ and $Z(x_i + h)$ are the studied variable values at locations x_i and $x_i + h$, reserctively; $N(h)$ is the number of observation pairs separated by a distance h (Isaaks and Srivastava, 1989; Arslan, 2012).

Kriging methods are powerful interpolation stochastic techniques used to find the linear unbiased estimate (Ahmadi and Sedghamiz, 2007). Among the various methods of Kriging, in this study ordinary Kriging (OK) approach is used to interpolate WQI and EC values for unsampled locations. OK so called the best linear unbiased estimate (BLUE), where it aims at minimizing the error variance (Isaaks and Srivastava, 1989). The form of OK estimator is defined as follows:

$$Z^*(x_0) = \sum_{i=1}^{n} \lambda_i \, . Z(x_i) \qquad with \qquad \sum_{i=1}^{n} \lambda_i = 1 \quad (8)$$

where $Z^*(x_0)$ is the estimated value at location x_0, $Z(x_i)$ is the measured value at location x_i, λ_i is the weighting factor assigned to $Z(x_i)$ and n is the number of observations (Journel, 1986; Bradaï et al., 2016). The weights λ_i are determined in such a way as to satisfy the optimising conditions of unbiasedness (Eq. 7) and minimum variance (Eq. 8), (Delhomme 1978).

$$E\{Z^*(x_0) - Z(x_0)\} = 0 \quad (9)$$

$$Var\{Z^*(x_0) - Z(x_0)\} = minimum \quad (10)$$

The fitting of theoretical semivariogram (spherical, experimental or Gaussian) is an important step in Kriging methods. The important characteristics of the semivariogram are range (a), sill (C_0 + C), and nugget effect (C_0). In this study, the indicative goodness of fit (IGF) of the software VARIOWIN is used to fit the theoretical semivariogram. It indicates a good fit when it is as close as possible to zero (Pannatier, 1996).

Prediction performances are checked using cross validation method. To provide an accurate prediction models, the mean standardized error (MSE) should be near to zero, the root mean square error (RMSE) should be as small as possible (used when comparing models), and the root mean square standardized error (RMSSE) should be near to one (Johnston et al., 2003; Arslan, 2012).

$$ME = \frac{1}{n}\sum_{i=1}^{n}[Z^*(x_i) - Z(x_i)] \quad (11)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}[Z^*(x_i) - Z(x_i)]/\sigma(x_i) \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[Z^*(x_i) - Z(x_i)]^2} \quad (13)$$

$$RMSSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[Z^*(x_i) - Z(x_i)/\sigma(x_i)]^2} \quad (14)$$

where $Z^*(x_i)$ is the estimated value, $Z(x_i)$ is the measured value and $\sigma(x_i)$ is the estimation variance.

The normal distribution of WQI and EC is assessed using the Shapiro-Wilk (SW) test. The nugget to sill ratio ($C0/(C + C0)$) was considered to define the different classes of spatial dependence for the data as follows: a ratio ≤ 25% referring to a strong spatial dependence, a ratio between 25% and 75% indicates a moderate spatial dependence, and a ratio > 75% for a weak spatial dependence (Cambardella et al., 1994).

3. Results and discussion

3.1. Hydrochemistry of groundwater

3.1.1. Descriptive statistics

A statistical summary of physicochemical groundwater parameters for the dry season as also wet season is given in Table 5. The pH of groundwater samples is found slightly alkaline (pH > 7). It ranges from 7.21 to 7.94 with a mean value of 7.64 for the dry season and from 7.19 to

7.84 with a mean of 7.65 for the wet season. The average value of EC is the same for both seasons (average of 4 dS m^{-1}). During the dry season, EC varies from 1.14 dS m^{-1} to 7.61 dS m^{-1}, while during the wet season, EC varies from 1.13 dS m^{-1} to 7.07 dS m^{-1}. The coefficient of variation of 39.89 % and 39.21 % for dry and wet season, respectively, explains the spatial heterogeneity of groundwater EC. The TDS values range from 729.6 mg/l to 4870.4 mg/l (mean of 2560.1 mg/l) and from 723.2 mg/l to 4524.8 mg/l (mean value of 2559.6 mg/l) for dry and wet seasons, respectively. However, more than 90 % of the groundwater samples are above the permissible limit (1000 mg/l) recommended by the WHO (2011).

The relative abundance of major cations and major anions in groundwater samples for the two campaigns is in the sequence as $Na^+ > Ca^{2+} > Mg^{2+} > K^+$ and $Cl^- > SO_4^{-2} > HCO_3^- > NO_3^-$. The Na^+ and Ca^{2+} are the dominant cations, with sodium ranging from 108.79 mg/l to 1406.82 mg/l (average = 595.38 mg/l) and from 56.63 mg/l to 1365.09 mg/l (average = 565.21 mg/l) for the dry season and the wet season, respectively. The calcium is the second dominant cation, which varies between 88 – 508 mg/l with an average of 249.29 mg/l for the dry season and between 80 – 612 mg/l with an average of 240.62 mg/l for the wet season. The Cl^- and SO_4^{-2} are the dominant anions, with chloride ranging between 195.25 – 2236.50 mg/l (average = 908.25 mg/l) for the dry season and between 213 – 1704 mg/l (average = 859.65 mg/l) for the wet season. The sulphate varies from 202.11 to 2122.11 mg/l, with a mean of 858.17 mg/l for the dry season and from 180.45 to 1920 mg/l, with a mean 800.76 mg/l for the wet season. The concentrations of NO_3^- ranges between 6.72 - 97.56 mg/l and 6.80 - 97.84 mg/l with mean values of 42.57 mg/l and 39.73 mg/l for the dry season and wet season, respectively. The high concentrations of nitrates (11% of groundwater samples exceeding the permissible limit (50 mg/l) recommended by the WHO (2011) for both seasons) may be due to the excessive use of agricultural fertilizers and the existence of septic tanks noticed near the sampled wells.

3.1.2. Correlation coefficients

Pearson's correlation matrix is used to analyze the interrelationship between groundwater physicochemical variables and their diverse sources. Tables 6 and 7 represent correlation coefficient matrices at a significance level 0.05 of the dry season and wet season, respectively. The correlation coefficients among the studied parameters have shown approximately, a similar trend. According to these tables, EC parameter has highest correlation with Na+ (r = 0.861 and r = 0.878 for the dry and wet season, respectively) and Cl- (r = 0.883 and r = 0.943 for the dry and wet season, respectively). It also has high correlation with Mg2+ (dry season 0.833 and wet season 0.812), Ca2+ (in dry season r = 0.738 and in wet season r = 0.798) and SO4- (r = 0.714

and 0.703 for dry and wet season, respectively). This indicates that these five elements are the main contributors to the salinity of groundwater in the study area (Figure 21).

Table 5: Statistical summary of physicochemical parameters

Parameters	Period	Minimum	Maximum	Mean	Standard deviation	Coefficient of variation (%)
CE (dS m^{-1})	Dry season	1.14	7.61	4.00	1.60	39.89
	Wet season	1.13	7.07	4.00	1.57	39.21
TDS (mg/l)	Dry season	729.60	4870.40	2560.1	1021.6	39.89
	Wet season	723.20	4524.80	2559.6	1003.5	39.21
pH	Dry season	7.21	7.95	7.64	0.14	1.87
	Wet season	7.19	7.84	7.56	0.15	2.04
NO_3^- (mg/l)	Dry season	6.72	97.56	42.57	19.64	46.14
	Wet season	6.80	97.84	39.73	19.76	49.73
SO_4^{-2} (mg/l)	Dry season	202.11	2122.11	858.17	416.79	48.57
	Wet season	180.45	1920.00	800.76	423.46	52.88
Cl^- (mg/l)	Dry season	195.25	2236.50	908.25	458.07	50.43
	Wet season	213.00	1704.00	859.65	403.37	46.92
HCO_3^-(mg/l)	Dry season	197.64	549.00	305.83	58.86	19.24
	Wet season	139.08	387.96	235.03	34.23	14.56
Ca^{2+} (mg/l)	Dry season	88.00	508.00	249.29	116.25	46.63
	Wet season	80.00	612.00	240.62	126.72	52.67
Mg^{2+} (mg/l)	Dry season	2.40	340.80	128.86	66.85	51.88
	Wet season	60.00	309.60	138.75	59.14	42.62
Na^+ (mg/l)	Dry season	108.79	1406.82	595.38	241.24	40.52
	Wet season	56.63	1365.09	565.21	237.24	41.97
K^+ (mg/l)	Dry season	2.73	19.98	8.60	2.98	34.70
	Wet season	2.73	17.26	8.08	2.56	31.72

There is no correlation between NO_3^- – EC in both seasons (r = 0.012 in dry season and r = -0.017 in wet season). The poor correlation of NO_3^- with the other constituents indicates its exogenous origin of the hydrogeological environment in the study area, which could be from agricultural activities. High positive correlation coefficients are observed between Na^+ – Cl^- (0.866 and 0.886 in dry and wet season, respectively), Mg^{2+} – Cl^- (0.750 in dry season and 0.734 in wet season), Mg^{2+} – SO_4^- with correlation values of 0.738 in the dry season and 0.745 in the

wet season, Ca^{2+} – SO_4^{-} with correlation values of 0.728 and 0.743 in dry and wet season, respectively. These associations are linked to the chloride minerals, mainly halite (NaCl), and the sulfate minerals such as gypsum ($CaSO_4 - 2H_2O$), anhydrite ($CaSO_4$) and epsomite ($MgSO_4 - 7H_2O$), which are widespread in arid regions, where they are the principal cause of salinity in groundwater.

Table 6: Correlation matrix between groundwater variables of the dry season

Variables	NO_3^-	SO_4^-	Cl^-	HCO_3^-	Ca^{2+}	Mg^{2+}	Na^+	K^+	pH	EC
NO_3^-	1.000	0.281	-0.186	-0.275	0.011	0.209	-0.090	-0.139	0.086	0.012
SO_4^-		1.000	0.431	0.183	**0.728**	**0.738**	0.514	0.047	-0.436	**0.714**
Cl^-			1.000	0.556	0.642	**0.750**	**0.866**	0.431	-0.570	**0.883**
HCO_3^-				1.000	0.306	0.395	0.520	0.071	-0.391	0.522
Ca^{2+}					1.000	0.680	0.446	0.040	-0.629	**0.738**
Mg^{2+}						1.000	0.635	0.064	-0.620	**0.833**
Na^+							1.000	0.537	-0.354	**0.861**
K^+								1.000	-0.002	0.345
pH									1.000	-0.526
EC										1.000

Underlined coefficients exhibit a significant correlation at the 0.05 level, n =65
Underlined and bold coefficients exhibit a strong correlation coefficient r > 0.700

Table 7: Correlation matrix between groundwater variables of the wet season

Variables	NO_3^-	SO_4^-	Cl^-	HCO_3^-	Ca^{2+}	Mg^{2+}	Na^+	K^+	pH	EC
NO_3^-	1.000	0.265	-0.159	-0.366	0.133	0.295	-0.199	-0.124	-0.059	-0.017
SO_4^-		1.000	0.484	0.110	**0.743**	**0.745**	0.485	0.061	-0.477	**0.703**
Cl^-			1.000	0.400	0.687	**0.734**	**0.886**	0.248	-0.512	**0.943**
HCO_3^-				1.000	0.097	0.207	0.415	-0.048	-0.232	0.352
Ca^{2+}					1.000	**0.804**	0.484	0.008	-0.584	**0.798**
Mg^{2+}						1.000	0.545	-0.103	-0.659	**0.812**
Na^+							1.000	0.392	-0.325	**0.878**
K^+								1.000	0.167	0.239
pH									1.000	-0.526
EC										1.000

Underlined coefficients exhibit a significant correlation at the 0.05 level, n =65
Underlined and bold coefficients exhibit a high correlation coefficient r > 0.700

On the other hand, human activities, such as inappropriate cultural practices that we have observed in the fields and the excessive exploitation of groundwater can contribute to the salinity of this precious resource.

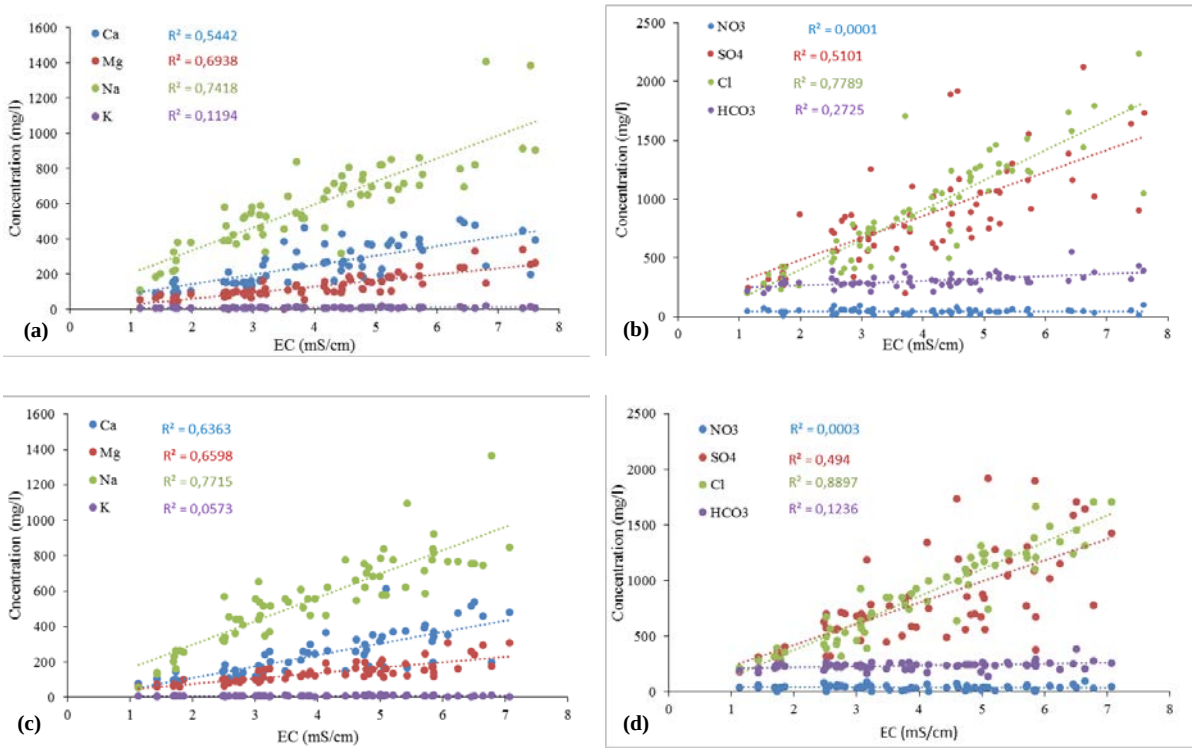


Figure 21: Relationship between different chemical elements (cations and anions) and groundwater EC (a, b: dry season; c, d: wet season)

3.1.3. Multivariate statistical analysis

Factor analysis with principal component extraction (FA/PCA) is carried out on 10 physicochemical parameters data of two periods (dry sand wet season) for a better understanding of groundwater hydrochemistry. To minimize the complexity of the factor loadings and to make the structure simpler for interpretation, Equamax rotation method is taken into consideration. The orthogonal Equamax rotation is a combination of Varimax method, which simplifies factors, and the Quartimax method, which simplifies variables (Reemtsma and Ittekkot, 1992).

Eigenvalues describe the maximized variance associated with the different factors and provide measures of the significance of these factors, where all the factors with eigenvalues of one or greater are considered significant (the factor with the highest eigenvalue is the most significant) (El Alfy et al., 2017). Equamax rotation is used to extract the main significant factors. Eigenvalues variation plotted against different factors is given in Figure 22.

Three factors are retained from the FA/PCA with Equamax rotation (Table 9, Figure 23), which explain 80.832 % and 81.475 % of the total variance and a cumulative eigenvalue of 8.083 and 8.148 for the dry and wet season, respectively (Table 8). In the dry season, the first factor, accounting for 52.939 % of the total variance and an eigenvalue of 5.294, is characterized by strong positive loadings with sulfate (0.793), chloride (0.833), calcium (0.851), magnesium (0.913), sodium (0.732), and EC (0.924). In the wet season (accounting for 51.347 % of the total variance and an eigenvalue of 5.135), it is characterized by very high positive loadings with sulfate (0.814), chloride (0.855), calcium (0.882), magnesium (0.930), sodium (0.731) and EC (0.942). This factor represents the natural hydrogeochemical evolution of groundwater, which can be linked to the dissolution of rocks and minerals by chemical weathering.

The second factor explains 16.292 % of the total variance with an eigenvalue of 1.629 in the dry season and 17.983 % of the total variance with an eigenvalue of 1.798 in the wet season. This factor is characterized by strong positive loadings with nitrate (0.875 in the dry season and 0.823 in the wet season). F2 may be related to the contamination of groundwater by agricultural fertilizers and domestic septic tanks.

The third factor explains 11.601 % of the total variance with an eigenvalue of 1.160 in the dry season and 12.146 % of the total variance with an eigenvalue of 1.215 in the wet season. It had strong loading with potassium (0.908 in the dry season and 0.924 in the wet season. This result can be explained by the impact of the use of potassium fertilizers on groundwater quality.

Table 8: Eigenvalues, percent of variance, and cumulative

	Dry season			Wet season		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
Eigenvalue	5.294	1.629	1.160	5.135	1.798	1.215
% of variance	52.939	16.292	11.601	51.347	17.983	12.146
Cumulative	5.294	6.923	8.083	5.135	6.933	8.148
Cumulative %	52.939	69.232	80.832	51.347	69.330	81.475

Table 9: Loading for the equamax rotated factor matrix of the three factors model explaining 80.832 % and 81.475 % of the total variance in the dry season and wet season, respectively

Variables	Dry season			Wet season		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
NO_3^-	0.063	0.875	-0.078	0.157	0.823	-0.153
SO_4^-	0.793	0.411	-0.047	0.814	0.255	0.001
Cl^-	0.833	-0.271	0.376	0.855	-0.337	0.254
HCO_3^-	0.536	-0.585	-0.023	0.285	-0.782	-0.162
Ca^{2+}	0.851	0.048	-0.158	0.882	0.135	-0.052
Mg^{2+}	0.913	0.162	-0.010	0.930	0.125	-0.174
Na^+	0.732	-0.129	0.585	0.731	-0.377	0.453
K^+	0.129	-0.065	0.908	0.067	-0.013	0.924
pH	-0.736	-0.223	-0.285	-0.697	0.078	0.394
EC	0.924	-0.013	0.299	0.942	-0.171	0.236

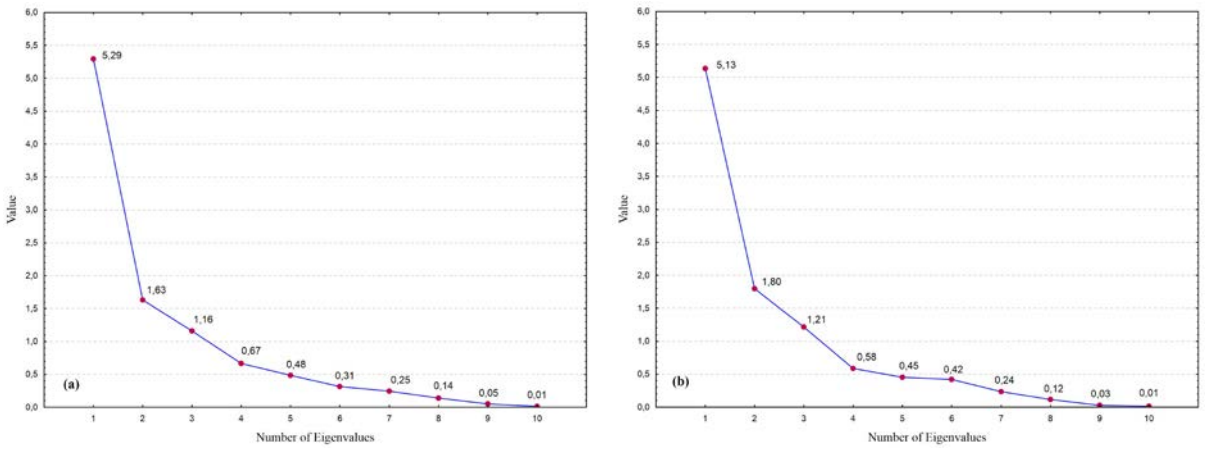


Figure 22. Scree plot of eigenvalues (a: dry season, b: wet season)

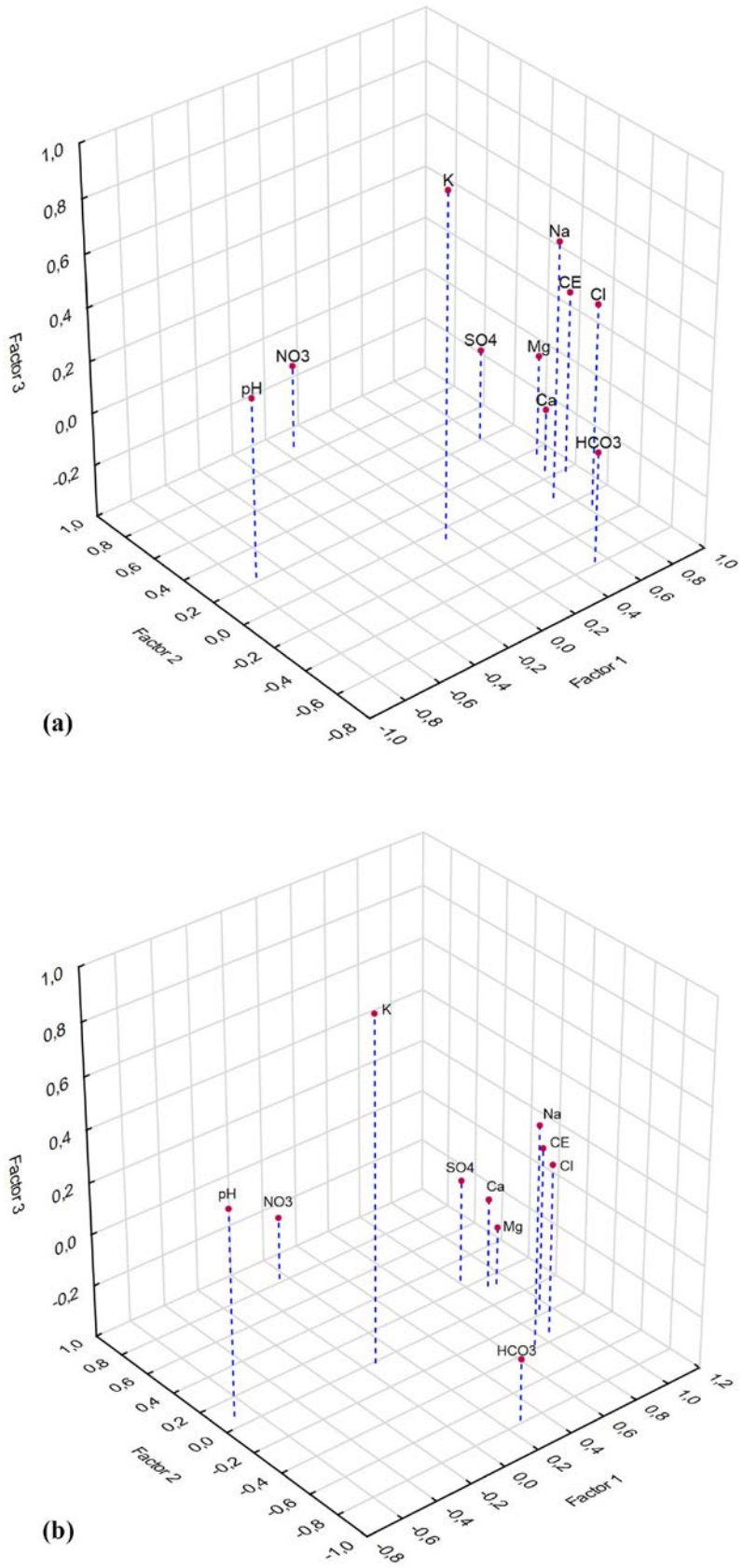


Figure 23: 3D representation of loadings of Factor 1, factor 2 and factor 3 after equmax rotation (a: dry season, b: wet season)

The coordinate analysis of cases after equamax rotation helps to determine the contribution of each water samples in each factorial axis. Score < -0.5 indicates high negative contribution, -0.5 < score < 0 indicates low negative contribution, 0 < score < 0.5 yields low positive contribution and score > 0.5 shows high positive contribution (Figure 24).

The mineralization characteristic of the F1 axis is identified by 19 samples and 17 samples as high positive contributions in the dry season, and wet season, respectively, and most of them are located in the middle of the study area. However, 27 samples for the dry season and 23 samples for the wet season have high negative contribution, where the majority are located in the western part of the study area (Figures 25a and 25d).

The contamination by nitrate is determined by factor 2 by 20 samples for the dry season and 16 samples for the wet season with high positive contribution, while 18 samples and 22 samples display high negatives contribution for the dry and wet season, respectively (Figures 25b and 25e).

Groundwater samples with levels of potassium justifying factor 3 with high positive contribution are 16 samples for the dry season and 12 samples in the wet season. However, 22 samples (dry season) and 18 samples (wet season) make high negative contribution (Figures 25c and 25f).

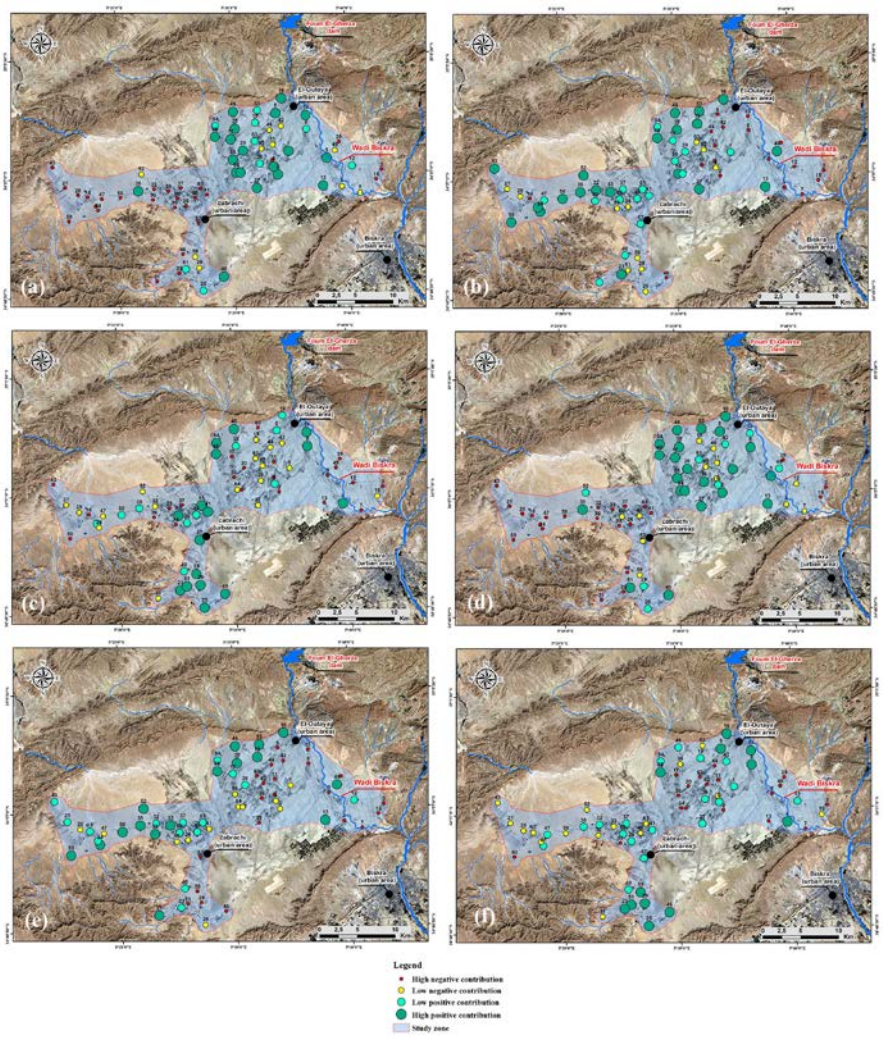


Figure 24: Spatial distribution of scores of the different factors (a, b and c are scores of F1, F2 and F3, respectively in the dry season, d, e and f are scores of F1, F2 and F3, respectively in the wet season)

3.1.4. Hydrochemical facies

The hydrochemical facies of the groundwater samples is determined by plotting the concentrations of cations and anions (meq/l) in the trilinear Piper diagram, which includes two triangles and one diamond shaped field (Piper, 1944; Sreedhar and Nagaraju, 2017). The triangles of cations (Figure 25) indicates that water samples fall into two groups: the first is water with a dominance of sodium and the second group with non-dominant cation. The triangle of anions shows the dominance of chloride and sulphate. The water samples can be grouped into two main hydrochemical facies based on their position on the diamond field. The first hydrochemical facies is the $(Na^+ + K^+) - (Cl^- + SO_4^{-2})$ type, where the alkalis exceed the alkaline

earths and the strong acids exceed the weak acids. The second is the (Ca^{2+} + Mg^{2+}) − (Cl^{-} + SO_4^{-2}) type, where the alkaline earths exceed the alkalis.

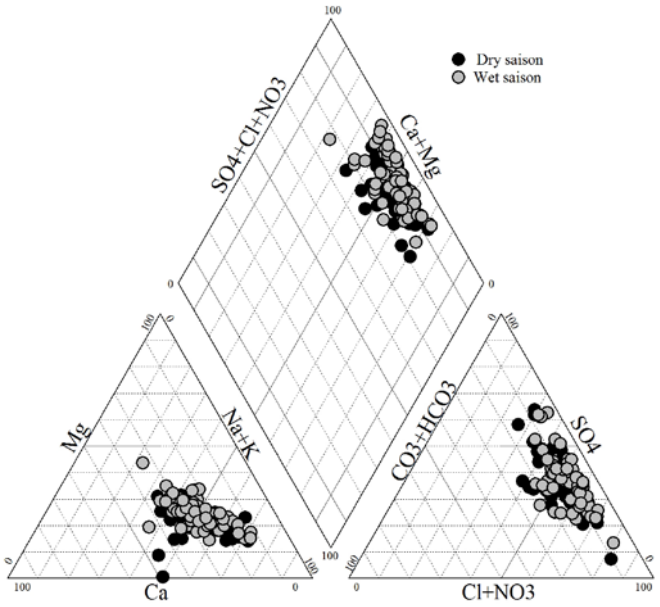


Figure 25: Piper trilinear diagram showing the main hydrochemical facies

3.1.5. Saturation index

Chemical equilibrium for six mineral species (anhydrite, aragonite, calcite, dolomite, gypsum and halite) is examined by calculating the saturation index (SI), which is expressed as follows:

$$SI = \log(IAP/Ks) \quad (15)$$

where IAP is the ion activity product and Ks is the solubility constant of the mineral. If the groundwater is exactly saturated with a particular mineral species then SI = 0 (equilibrium state); while SI > 0 indicates oversaturation with respect to the mineral species (precipitation state) and SI < 0 indicates under saturation (dissolution state) (Zaidi et al., 2017; Kant et al., 2018).

According to the saturation indices, calculated for the two seasons and plotted on graphs (Figure 26), all groundwater samples are oversaturated with respect to carbonate minerals (aragonite, calcite and dolomite), whereas, evaporate minerals (anhydrite, gypsum and halite) are under saturated. The precipitation of carbonate minerals could explain the limited enrichment of groundwater in bicarbonates, whereas the dissolution of gypsum and anhydrite (Eq. 16 and Eq. 17) could explain the enrichment of sulphate. The increase of the concentration of calcium and magnesium due to reverse cation exchange as well as the dissolution of gypsum and anhydrite oversaturate the groundwater with respect to calcite and dolomite and causes its precipitation.

The dissolution of the halite (Eq. 18), which is considerably under saturation, is the primary source of high concentrations of sodium and chloride in the groundwater samples.

$$CaSO_4 \cdot 2H_2O \rightarrow Ca^{2+} + SO_4^{2-} + 2H_2O \quad (16)$$

$$CaSO_4 \rightarrow Ca^{2+} + SO_4^{2-} \quad (17)$$

$$NaCl \rightarrow Na^{+} + Cl^{-} \quad (18)$$

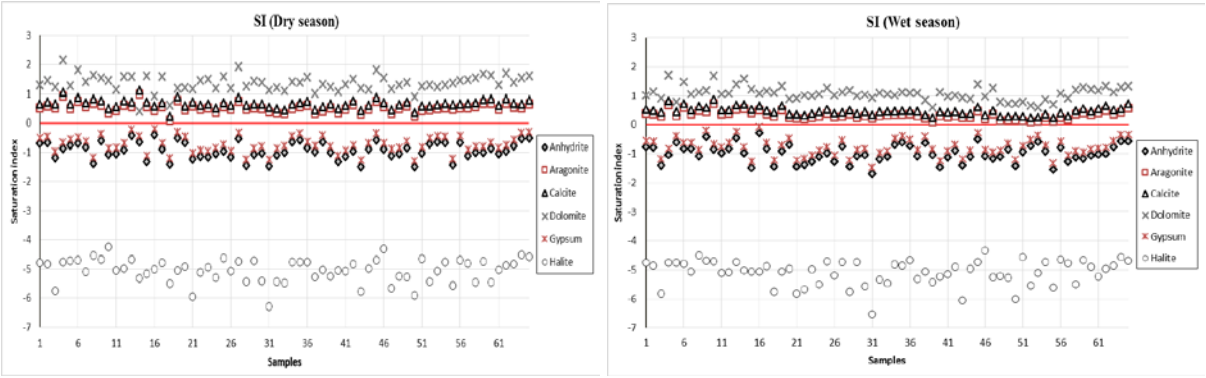


Figure 26: Plots of saturation indices (SI) of some mineral species (dry and wet season)

3.1.6. Base ion exchange

The ion exchange reactions between the groundwater and aquifer environment during residence or movement can be inferred by using the chloro-alkaline indices (CAI 1 and CAI 2). They are proposed by Schoeller (1977 in Zaidi et al., 2017) and they are calculated using the following formulas:

$$CAI\ 1 = \frac{Cl^- - (Na^+ + K^+)}{Cl^-} \quad (19)$$

$$CAI\ 1 = \frac{Cl^- - (Na^+ + K^+)}{(Cl^- + HCO_3^- + SO_4^{2-} + NO_3^-)} \quad (20)$$

where all the ions are in meq/l

A positive index (CAI) indicates an exchange between sodium and potassium in groundwater with calcium and magnesium in the aquifer environment (direct exchange). On the other hand, a negative index (CAI) indicates an exchange between calcium and magnesium in groundwater with sodium and potassium in the aquifer environment (reverse exchange). About 54 % of the samples in the dry season (Figure 27a) and 57 % f samples in the wet season (Figure 27b) show negative values for the both indices. The results indicate a slight dominance of the reverse ion exchange in the study area.

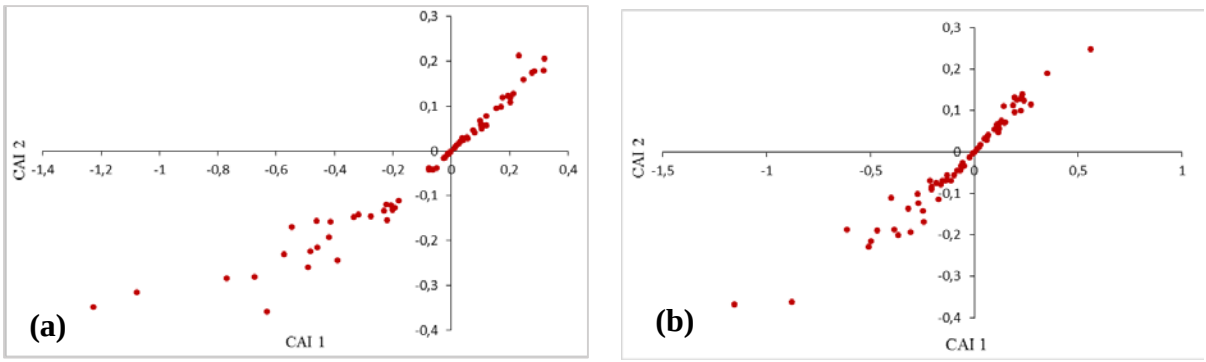


Figure 27: Chloro-alkaline indices (CAIs). a: dry season, b: wet season

3.1.7. Ionic relationships

Stoichiometric relations among dissolved species can explain the origin and the process of groundwater salinization (Sun et al., 2016). According to the ionic relationships (Figure 28), the ionic behavior of groundwater has been decided during the dry season and the wet season, which are almost the same. From Figures 29a and 29b, the increase of calcium and magnesium does not correlate with bicarbonate, which indicates the low contribution of calcite and dolomite dissolution to groundwater mineralization (Aouidane et al. 2017). The relationship between calcium and bicarbonate (Figures 29b) shows that the samples are positioned above the 1:1 line, indicates an excess of calcium over bicarbonate, which can be explained by the dissolution of gypsum with base exchange. Figures 29c shows the relationship between sodium and chloride. The samples have a stoichiometric distribution around the 1:1 line, which indicates the dominance of halite dissolution. According to the good correlation between calcium and sulfate (Figures 29d), the mineralization of groundwater is influenced by the gypsum and anhydrite dissolution. The majority of samples are distributed below the 1:1 line and show a relative deficit in calcium relative to sulfate, which could be to cation exchange where calcium ions are replaced sodium ions. Figures 29e shows a weak correlation between the alkalis ($Na^+ + K^+$) and and the alkaline earth elements ($Ca^{2+} + Mg^{2+}$), ($R^2 = 0.34$ in the dry season and 0.29 in the wet season) with samples falling on the both sides of the 1:1 line.

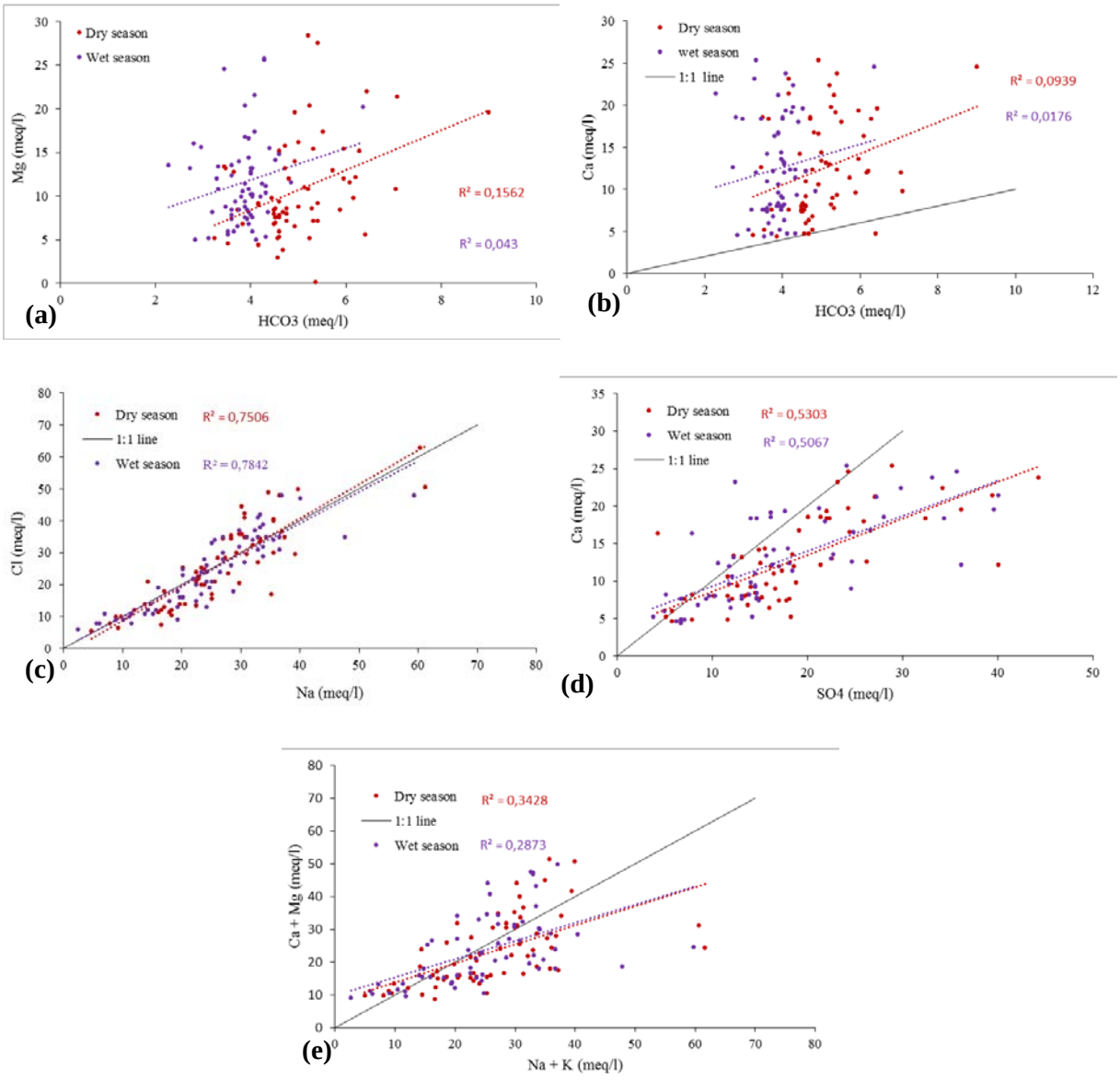


Figure 28: The bivariate diagrams: **(a)** HCO3- vs Mg relationship, **(b)** HCO3- vs Ca2+ relationship, **(c)** Na+ vs Cl- relationship, **(d)** Cl- vs Na+/Cl- relationship, **(d)** Na+ + K+ vs Ca2+ + Mg2+ relationship, **(e)** SO42- vs Ca2+ relationship.

The relationship between (Ca + Mg) - (HCO3 +SO4) and (Na + K) – Cl gives a clear view of the cation exchange process in groundwater (Farid et al., 2015; Aouidane et al., 2017). As shown in Figure 29, the groundwater samples in the study area define a straight line with determination coefficients of 0.80 and 0.57 for the dry and wet season, respectively, which confirms the participation of sodium, potassium, calcium and magnesium in cation exchange reactions. This graph supports the hypothesis that groundwater in El-Outaya is the influenced by two primary processes: rock dissolution as well as base exchange.

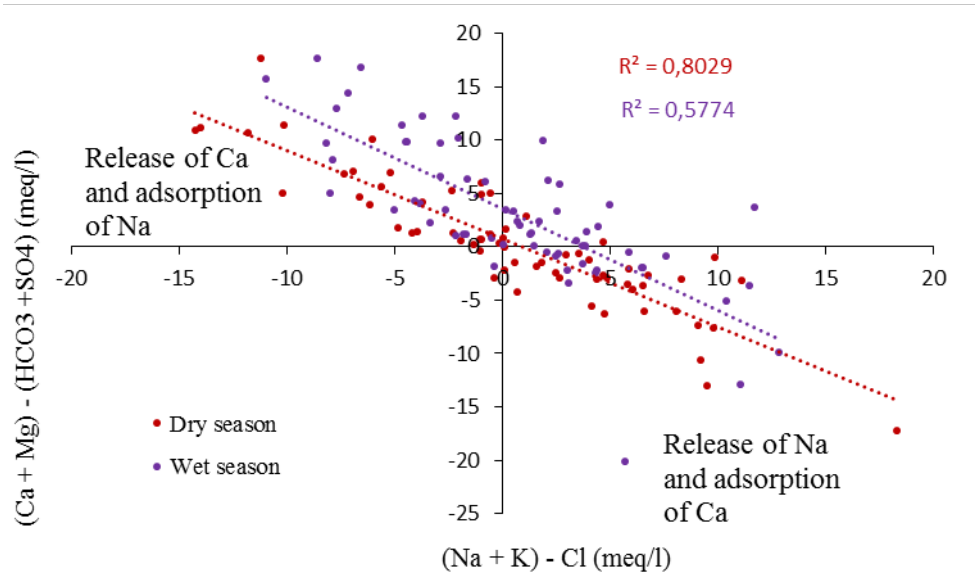


Figure 29: Correlation diagram between [(Ca2+ + Mg2+) - (HCO3- + SO42-) and (Na+ + K+) – Cl-] (meq/l)

3.2. Spatial variability of hydro-chemical parameters

The spatial variability maps of pH, Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Cl^{-}, HCO_3^{-} and SO_4^{2-} have been generated using inverse distance weighted interpolation method. Spatial distribution of nitrate will be studied separately in the next chapter.

3.2.1. Potential Hydrogen (pH)

Spatial variability of pH concentrations are shown in Figure 30, which indicates that pH values of groundwater in the study area vary from 7.21 to 7.95 in the dry season and from 7.19 to 7.84 in the wet season. This shows that groundwater in El-Outaya plain is mainly neutral to lightly alkaline but falls in the safe limit in terms of drinking water.

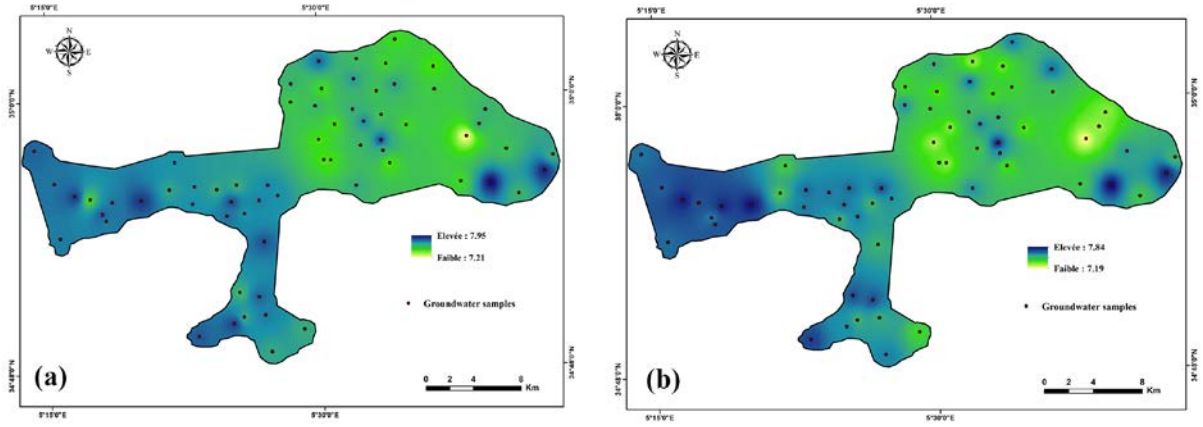


Figure 30: Spatial variability of pH (a: dry season, b: wet season)

3.2.2. Calcium (Ca^{2+} mg/l)

Figure 31: shows the spatial distribution of Ca^{2+}, where its concentrations vary between 88 mg/l and 508 mg/l in the dry season and between 80 mg/l – 612 mg/l in the wet season. The maps indicate almost the same spatial variation with a slight increase of concentrations in the western part of the study area and a slight decrease of concentrations in the middle of the study area.

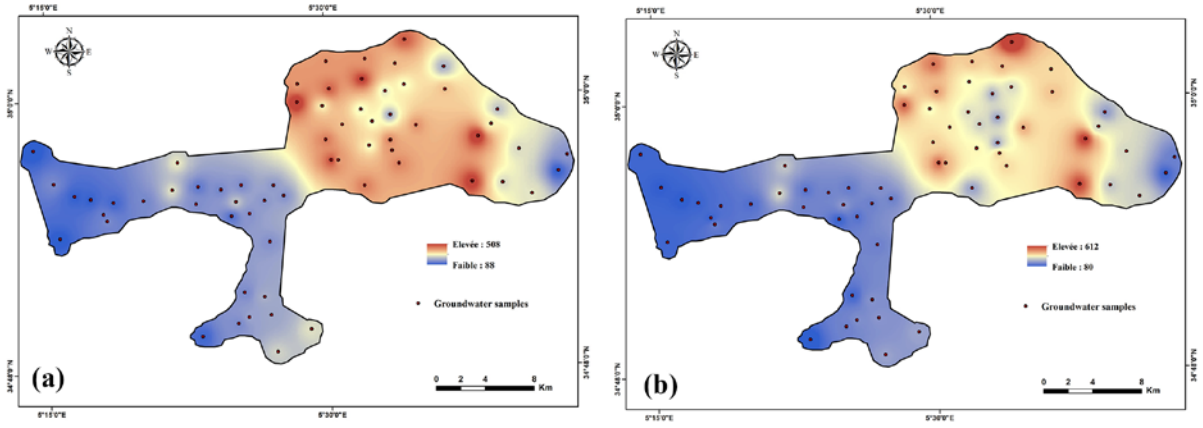


Figure 31: Spatial variability of Ca^{2+} (a: dry season, b: wet season)

3.2.3. Magnesium (Mg^{2+} mg/l)

Figure 32 is for the spatial variability maps of Mg^{2+}. It is obvious from these maps that it has been observed mostly the middle part of the plain high concentrations, while low concentrations are noticed in the western part of the plain.

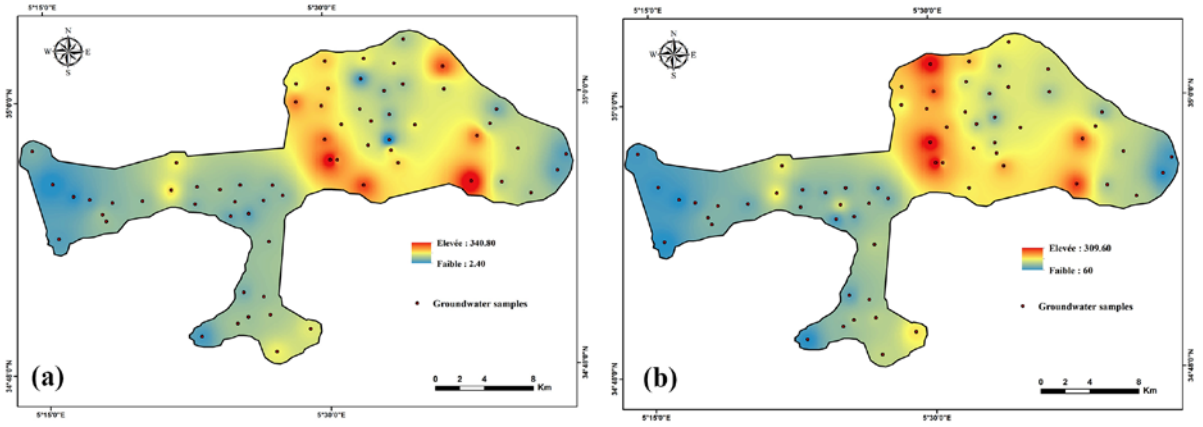


Figure 32: Spatial variability of Mg^{2+} (a: dry season, b: wet season)

3.2.4. Sodium (Na^{+} mg/l)

The spatial distribution of sodium ion concentrations in groundwater of El-Outaya plain is shown in Figure 33. The maps indicate that the study area is characterized generally by high

concentrations of Na^{+}. The highest values are observed in the northeastern part of the plain, near of Djebel El-Maleh, and in southeastern part of the plain (depression of Salga Saadoune).

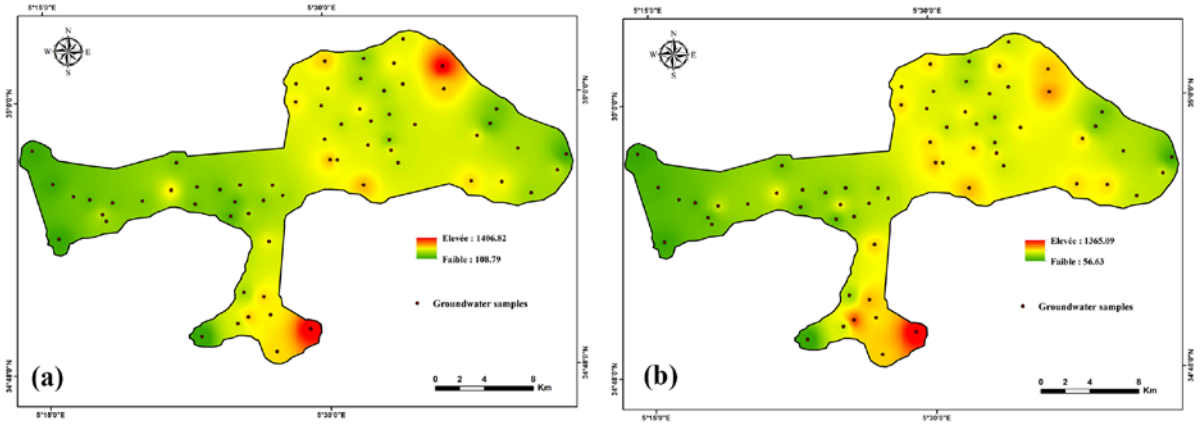


Figure 33: Spatial variability of Na^{+} (a: dry season, b: wet season)

3.2.5. Potassium (K^{+} mg/l)

The spatial distribution maps of potassium (Figure 34) indicate the almost the same spatial variation between the dry season (concentrations are varied from 2.73 mg/l to 19.98 mg/l) and the wet season (concentrations are varied from 2.73 mg/l to 17.26 mg/l).

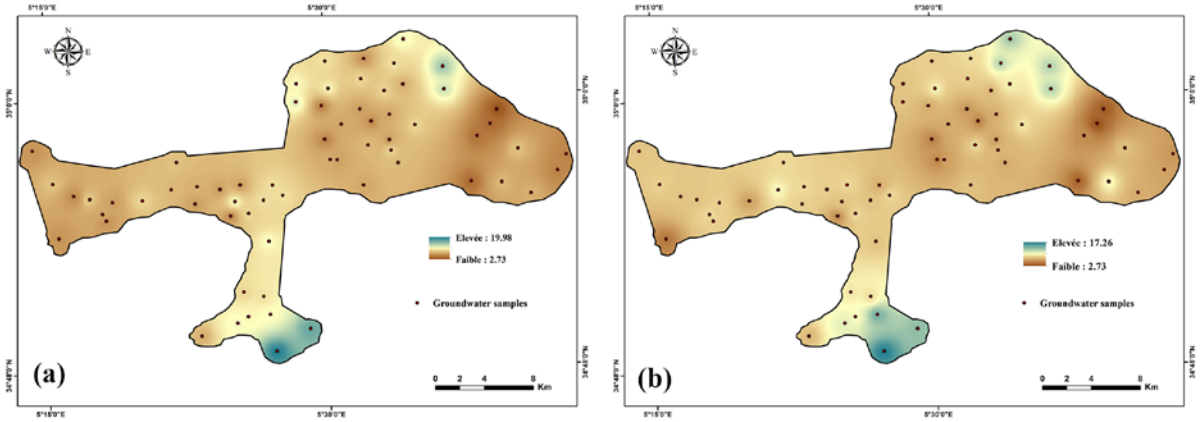


Figure 34: Spatial variability of K^{+} (a: dry season, b: wet season)

3.2.6. Chloride (Cl^{-} mg/l)

The spatial variability maps of chloride in El-Outaya plain (Figure 35) indicate its variability from a region to another, where the high concentrations are observed near the depression of Selga Saadoune (southwestern region) and near the urban area of El-Outaya city. The low concentrations of Cl^{-} are observed in the western region (Elmazouchia) and in the the extreme eastern part of the study area (Branis).

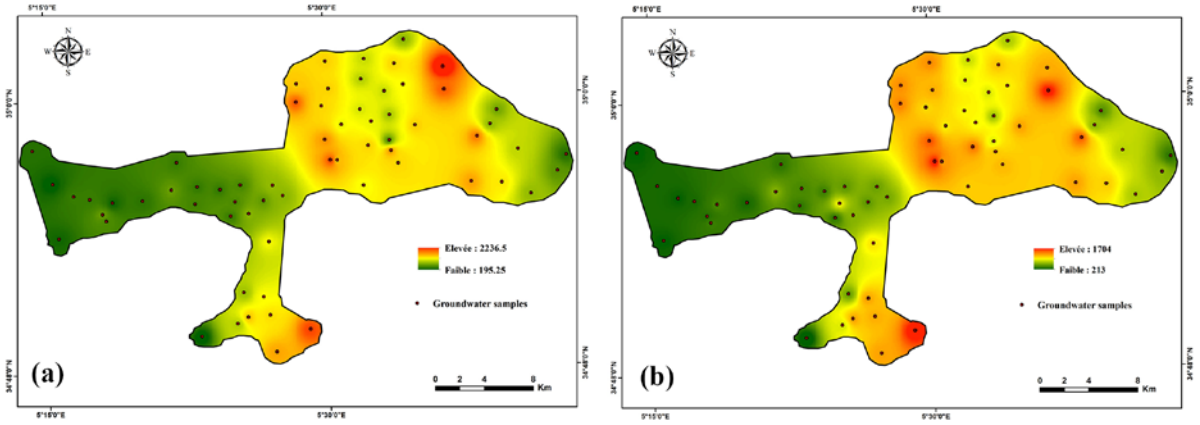


Figure 35: Spatial variability of Cl^- (a: dry season, b: wet season)

3.2.7. Sulfate (SO_4^{2-} mg/l)

According to the spatial distribution maps of sulfate concentrations in El-Outaya plain (Figure 36), the groundwater in the area of investigation contains Cl^- exceeding largely the permissible limit of WHO. The highest concentrations are in the middle of the study area and near to El-Hzima region, and the lowest concentrations are in the western part of the plain (Elmazouchia region).

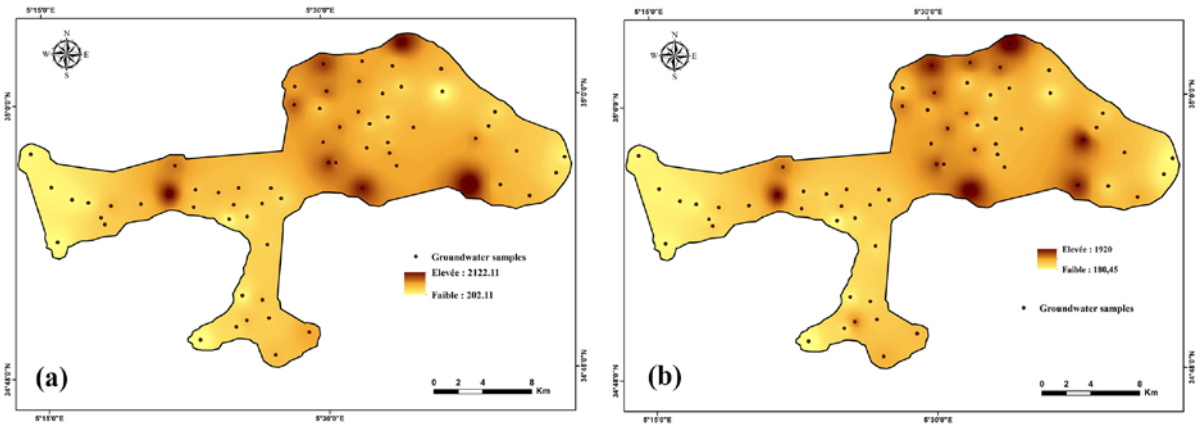


Figure 36: Spatial variability of SO_4^{2-} (a: dry season, b: wet season)

3.2.8. Bicarbonate (HCO_3^- mg/l)

Figure 37 indicates the spatial variation maps of bicarbonate concentrations in the dry and wet season. From these maps, it is possible to observe that mostly the eastern part of the study area has highest concentrations of HCO_3^-. In the dry season, the concentrations range is from 197.64 mg/l to 549 mg/l. However, a decrease in HCO_3^- concentrations is observed during the wet season (139.08 mg/l to 387.96 mg/l).

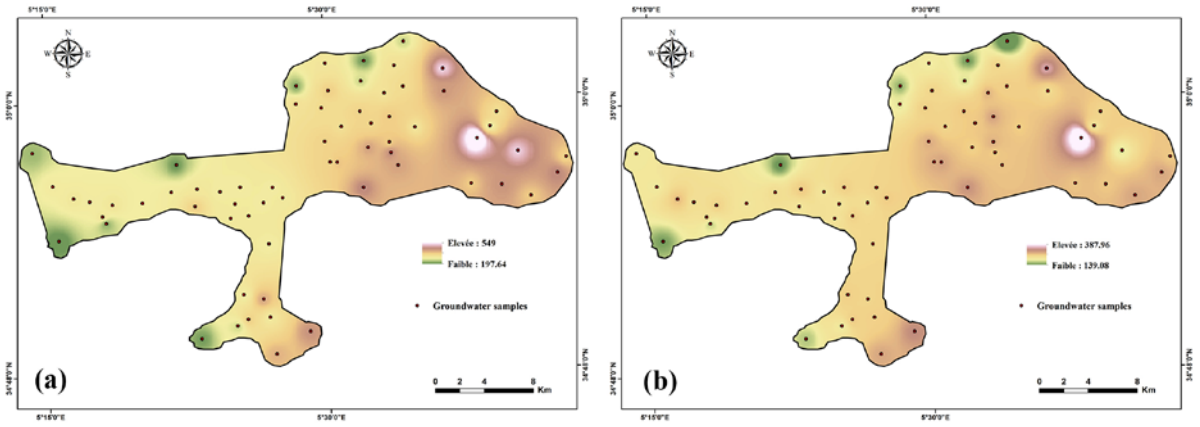


Figure 37: Spatial variability of HCO_3^- (a: dry season, b: wet season)

3.3. Assessment of groundwater for drinking purpose

3.3.1. Data distribution

The water quality for drinking purposes is assessed based on WQI. The spatial distribution maps of WQI in the study area for both seasons are prepared using OK. The OK method works better with normally distributed data (Arslan, 2012). The statistic value of Shapiro-Wilk test is highly significant with 0.145 and 0.269 for dry season and wet season, respectively. The skewness and the kurtosis values are close to zero with 0.361 and -0.462 for the dry season, 0.137 and -0.748 for the wet season, respectively (Table 11), which validate the normal distribution of the calculated WQI (Figure 38).

Table 10: Relative weights and WHO standards of chemical parameters

Parameters	WHO Standards	Weight (wi)	Relative weight (Wi)
TDS (mg/l)	1000	4	0.1290
pH	8.5	3	0.0968
NO_3^- (mg/l)	50	5	0.1613
SO_4^{-2} (mg/l)	250	4	0.1290
Cl^- (mg/l)	250	4	0.1290
HCO_3^- (mg/l)	120	1	0.0323
Ca^{2+} (mg/l)	100	2	0.0645
Mg^{2+} (mg/l)	50	2	0.0645
Na^+ (mg/l)	200	4	0.1290
K^+ (mg/l)	12	2	0.0645
		$\sum wi = 31$	$\sum Wi = 1$

Table 11: Statistics and SW test of WQI

WQI	Min	Max	Mean	SW test	Skewness	Kurtosis
Dry season	86.39	406.09	230.6	0.145	0.361	-0.462
Wet season	77	378.58	220.7	0.269	0.137	-0.748

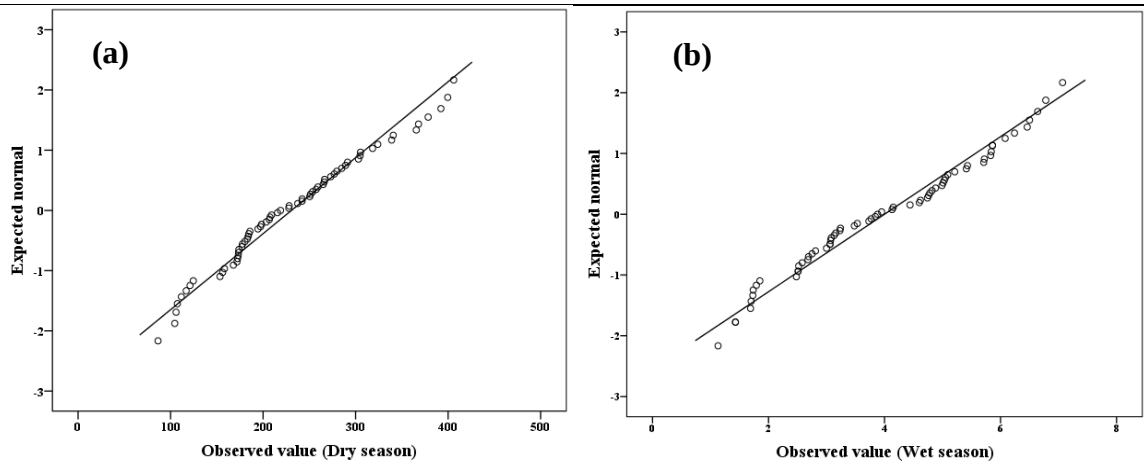


Figure 38: Scatter plot of the WQI normal distribution [(a) dry season, (b) wet season]

3.3.2. Variography and cross validation

The experimental omnidirectional variograms of the WQI (Figure 39) indicate that the exponential models suited better the WQI for both seasons. Table 12 shows the parameters of the variogram models. The nugget effect (C_0) is at values 100 and 635.2; the partial sill (C) range is between 5720.7 and 4596.8; and the range is 11137 (m) and 14275 (m), for the dry season and wet season, respectively. The values of the nugget to sill ratio indicate a high spatial dependence of the calculated WQI. The cross validation (Table 13) yields an accuracy of the predictions for both seasons. ME and MES are close to zero. RMSES are close to one with a value of 1.027 for the dry season and 0.981 for the wet season.

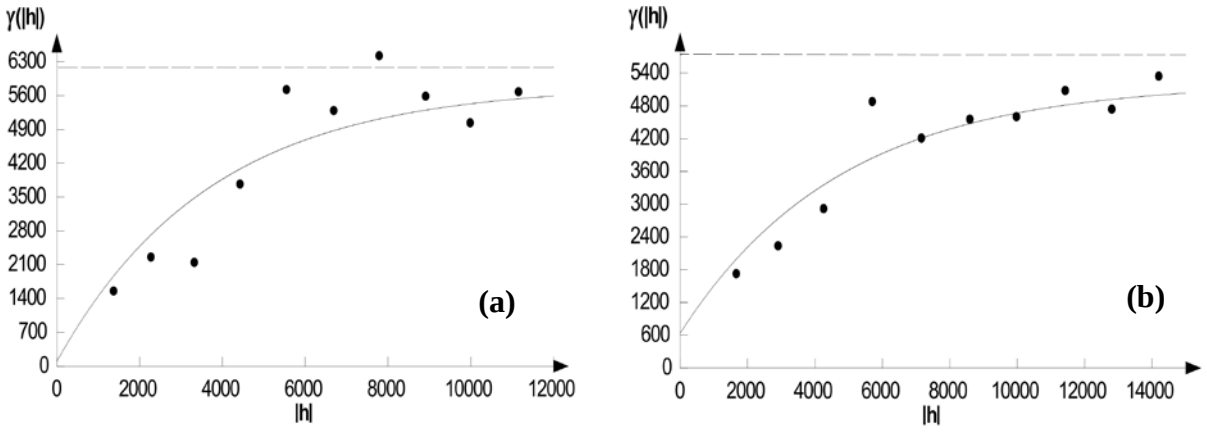


Figure 39: Omnidirectional variogram of the predicted WQI [(a) dry season, (b) wet season]

Table 12: Parameters of variogram models of WQI

Period	Model	Nugget effect	Partial sill	Range (m)	Nugget ratio (%)	IGF
Dry season	Exponential	100	5720.7	11137	1.72	$3.53 * 10^{-2}$
Wet season	Exponential	635.2	4596.8	14275	12.14	$1.46 * 10^{-2}$

Table 13: Cross validation for WQI

WQI	ME	MES	RMSES
Dry season	-0.014	0.0038	1.027
Wet season	0.104	0.004	0.981

3.3.3. Spatial distribution of WQI

It is obvious from the spatial distribution maps (Figure 40) that the groundwater in the study area is not of acceptable quality for human consumption. The prediction maps show almost the same overall spatial distribution of WQI for both seasons, only a slight difference is observed in the shape and the size of the area representing each class. In general, the groundwater quality decreases from the west (El Mazouchia and Oued Besbes) to the middle of the study area and from the east (Branis) to the north west of the study area. The very poor and unsuitable for drinking classes, represented by 58% of the groundwater samples for both seasons, are located in the middle of the study area and in the urban zone of El Outaya.

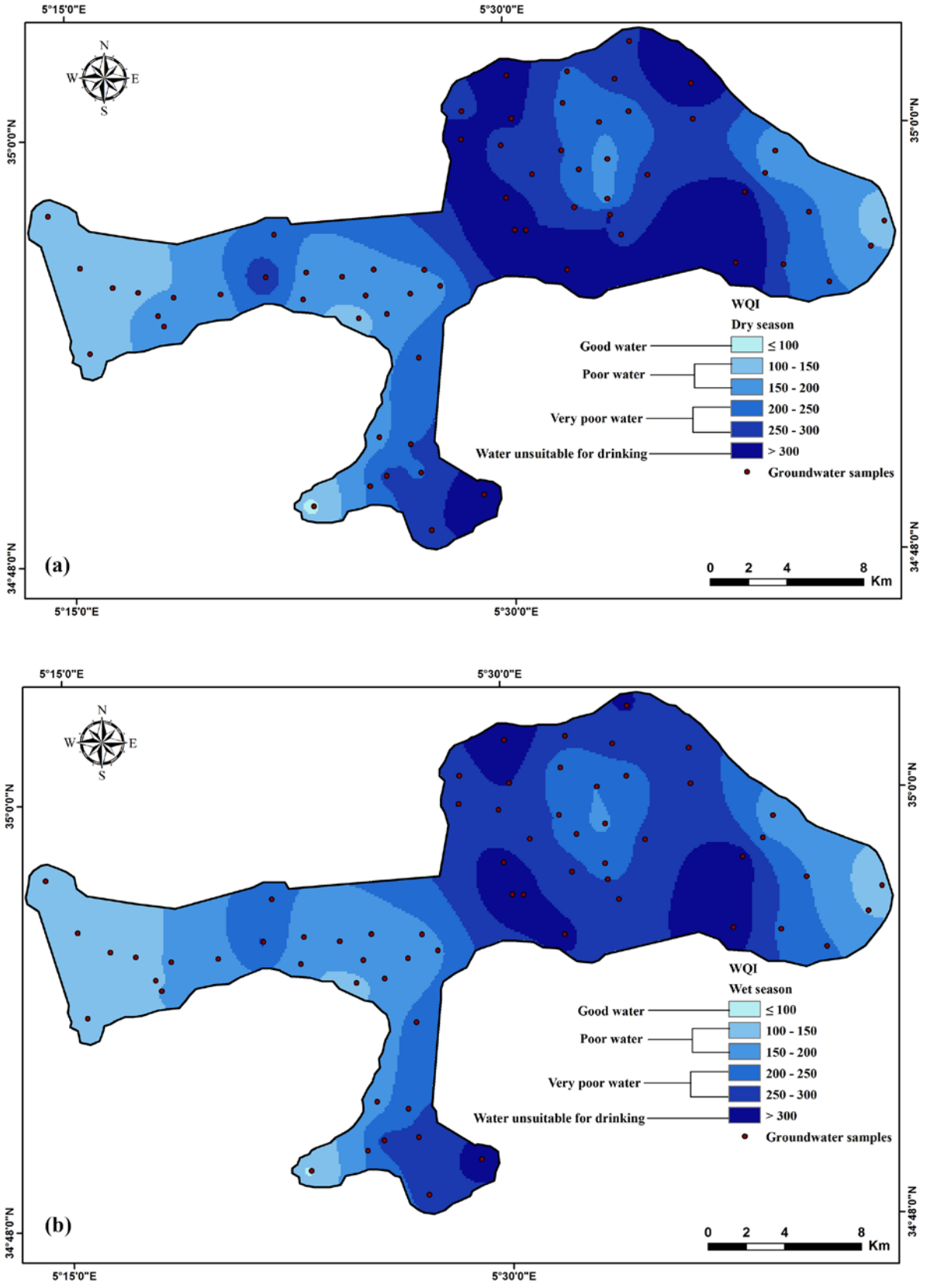


Figure 40: Prediction maps of the WQI [(a) Dry season, (b) Wet season]

3.4. Assessment of groundwater for irrigation purpose

3.4.1. Alkali hazard (SAR) and salinity hazard (EC)

Sodium adsorption ratio (SAR) is an important parameter in assessing the suitability of groundwater for irrigation purposes and it is defined by the following equation (USSLS 1954; Bouderbala, 2017):

$$SAR = Na^{+}/\sqrt{(Ca^{2+} + Mg^{2+}/2)} \quad (21)$$

where concentrations are expressed in meq/l.

The calculated SAR values range between 2.14 and 17.51 with a mean of 7.70 for the dry season and from 1.15 to 16.92 with a mean of 7.25 for the wet season. The United States Salinity Laboratory Staff (USSLS 1954) constructed a diagram, called Riverside diagram, for classifying the irrigation water. The classification based on the integrated effect of SAR and EC. The Riverside diagram (Figure 41a) shows that the majority of the groundwater samples for both seasons fall in classes C4S2 and C4S3. The class C4S2 (21.54% for the dry season and 26.15% for the wet season) indicates a high salinity hazard and medium alkalinity hazard, respectively. C4S3 (36.92% for the dry season and 27.69% for the wet season) indicates high salinity and high alkalinity hazard. Such categories can be used for irrigation on coarse textured or organic soils with good permeability and more tolerable crops should be selected. 7.69 % of the groundwater samples for the dry season and 9.23 % for the wet season fall in the C3S1 category, indicating water of medium salinity hazard and low alkalinity hazard. The class C3S2 (medium salinity and medium alkalinity) is represented by 6.15 % of the samples for the dry season and 4.62 % for the wet season. The classes of very high salinity and alkalinity (C5S3 and C5S4) are represented by 27.69 % and 32.31 % for the dry and wet season, respectively. Such water is unsuitable for irrigation purpose (Raju and Ram 2011; Bouderbala 2017).

3.4.2. Sodium percentage (Na%)

High concentration of Na^{+} in irrigation water displaces the calcium and magnesium in the soil and affects the soil permeability (reduction of soil infiltration capacity) and the growth of plants. The combination of sodium with carbonate forms alkaline soils, while its combination with chloride forms saline soils. Sodium percentage is calculated using the following formula (Raju and Ram, 2011; Alharbi, 2018):

$$Na^{+}\% = \frac{Na^{+} + K^{+}}{Ca^{2+} + Mg^{2+} + Na^{+} + K^{+}} * 100 \quad (22)$$

where concentration are expressed in meq/l.

The sodium percentage in the study area ranges from 33.46 % to 71.63 % and from 27.45 % to 78.46 % with an average of 53.01 % and 58.67 % for the dry season and wet season, respectively. Wilcox's diagram is widely used to assess the suitability of groundwater for irrigation by correlating Na^+ % and EC. Wilcox diagram in Figure 41b illustrates that the majority of the groundwater samples are unsuitable for irrigation for the two seasons (69.23 % for the dry season and 72.31 % for the wet season). 16.92 % of the samples belongs to the doubtful category in the dry season and 13.85 % in the wet season. Only 13.85 % of the groundwater samples are good to permissible for irrigation in both the seasons, which indicates a high risk of deterioration of the soil structure in the study area using this groundwater for irrigation.

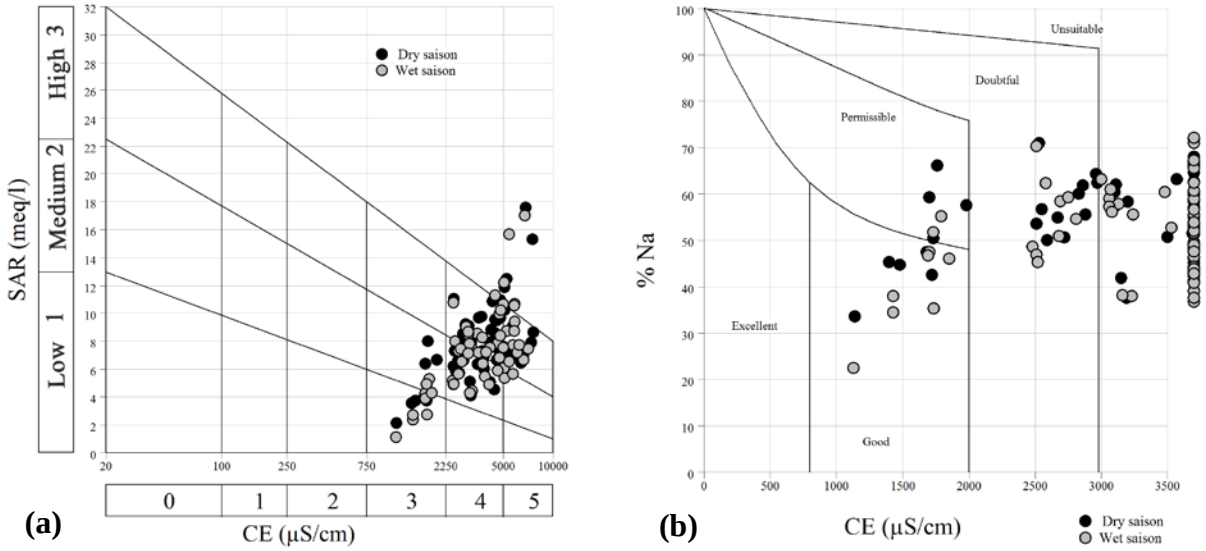


Figure 41: (a) Riverside diagram, (b) Wilcox diagram

3.5. Mapping of groundwater salinity (EC)

3.5.1. Data distribution

The prediction maps of EC are prepared using OK. Shapiro-Wilk test (with high significant values of 0.217 for the dry season and 0.080 for the wet season), kurtosis and skewness values (Table 14) validate the normal distribution of the groundwater EC for both seasons (Figure 42).

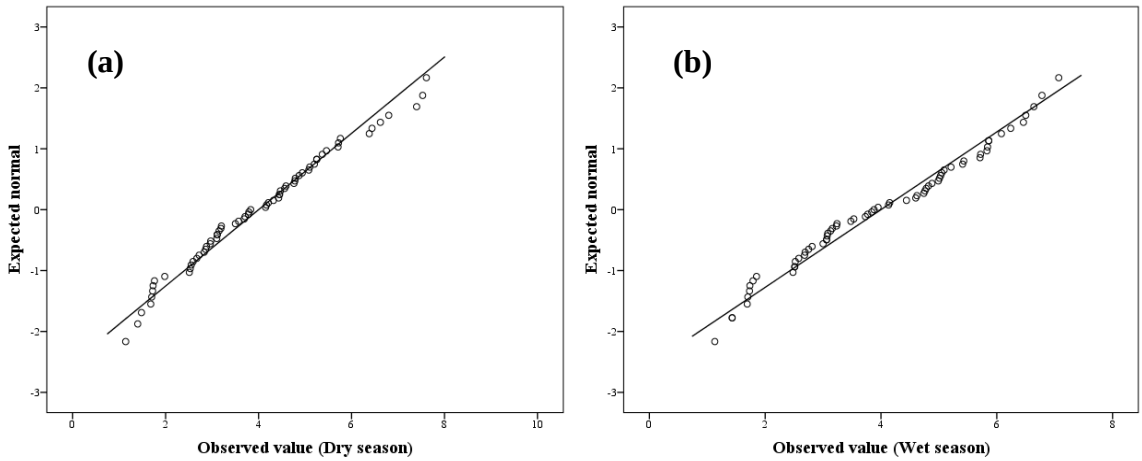


Figure 42: Scatter plot of the groundwater EC normal distribution [(a) dry season, (b) wet season]

Table 14: Normal distribution test of groundwater EC

EC	SW test	Skewness	Kurtosis
Dry season	0,217	0,314	-0,439
Wet season	0,080	0,059	-1,019

3.5.2. Variography and cross validation

Figure 43 shows that Gaussian model is the best fitted omnidirectional variogram of groundwater EC (for both seasons) obtained based on IGF. Table 15 summarizes the characteristics of the best fitted variograms. It is observed from Table 15 that the groundwater EC has a moderate spatial dependence in the dry season with a nugget to sill ratio of 27.68 and a high spatial dependence in the wet season with a nugget to sill ratio of 23.90. Low values of nugget effect indicate that groundwater EC is spatially correlated. However, nugget effect is the random error. The high values of the range shows a spatial continuity of the groundwater EC (Ahmadi and Sedghamiz, 2007; Bradaï et al., 2016). The results of cross validation (Table 16) implies accurate predictions. The ME and MES were 0.009 and 0.004, 0.013 and 0.007 for the dry season and wet season; respectively. RMSES was 1,073 in the dry season and 1.021 in the wet season.

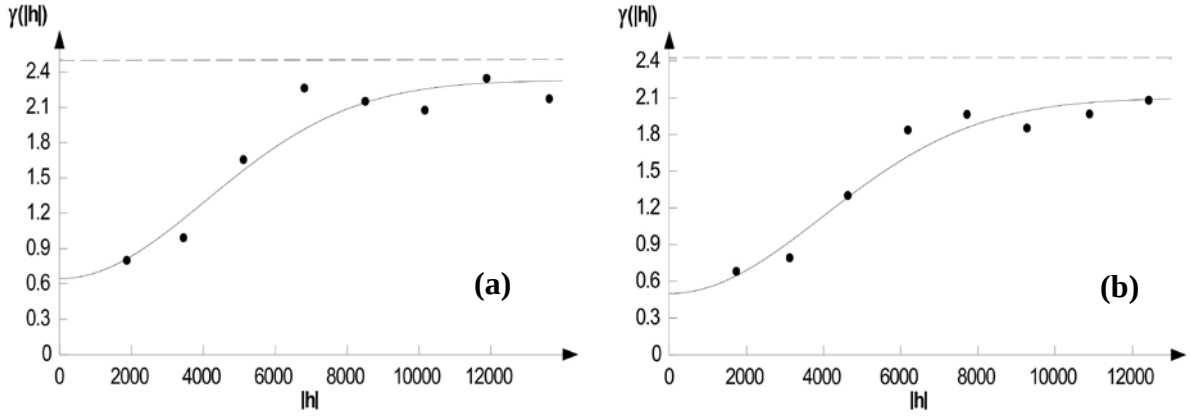


Figure 43: Omnidirectional variograms of the predicted EC [(a) dry season, (b) wet season]

Table 15: Characteristics of variogram models for groundwater EC

Period	Model	Nugget effect $(dS\ m^{-1})^2$	Partial sill $(dS\ m^{-1})^2$	Range (m)	Nugget ratio (%)	IGF
Dry season	Gaussian	0.645	1.685	9960	27.68	$8.85 * 10^{-3}$
Wet season	Gaussian	0.502	1.598	9737	23.90	$4.78 * 10^{-3}$

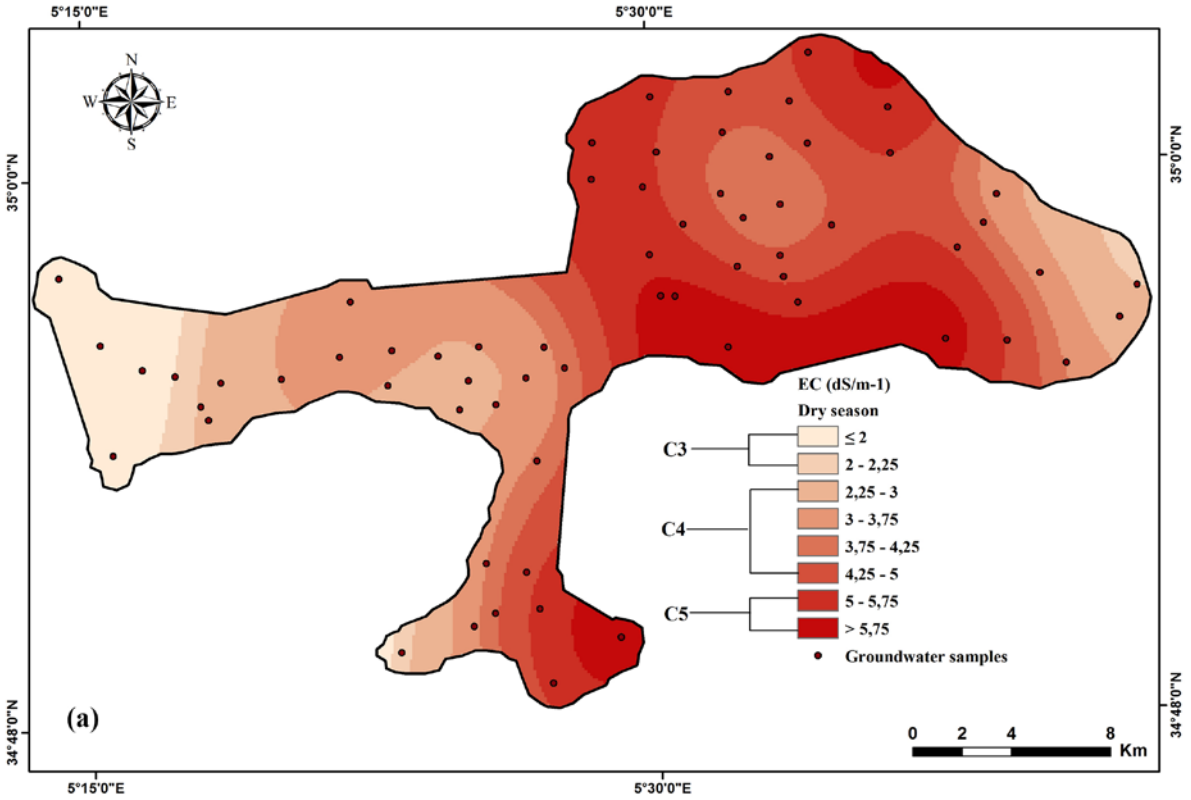
Table 16: Cross validation for groundwater EC

EC	ME	MES	RMSES
Dry season ($dS\ m^{-1}$)	0.009	0.004	1.073
Wet season ($dS\ m^{-1}$)	0.013	0.007	1.021

3.5.3. Prediction maps of groundwater EC

In general, the obtained maps of groundwater EC using OK (Figure 44) show the same overall spatial distribution of salinity for the two campaigns. The difference is in the shape and size of the area representing each irrigation class. The groundwater with salinity between 2.25 and 5 $dS\ m^{-1}$ (C4) is the dominant class in the west and east of the study area, which occupies 58.06 % and 51.39 % of area in the dry and wet season respectively. The groundwater with very high salinity risk (C5) is localized in the middle of the study area, which is characterized by intense agriculture activities, and in the urban zone of El Outaya near Djebel Elmaleh. The area of this class has increased form 32.08 % in the dry season to 37.95 % in the wet season. The dominance of groundwater with high and very high salinity in the study area could be due to the intense exploitation of the aquifer, the effects of highly saline irrigation return flows on the groundwater

and the dissolution of Triassic evaporitic materials in Djebel Elmaleh. The least salted water (class C3 with EC < 2.25 dS m^{-1}) is located at the east (El Mazouchia and Oued Besbes) and at the ouest (Branis) of the study area. It occupies 9.87 % and 10.67 % of area in the dry and wet season, respectively. The low values of groundwater EC in the east and the west of the study area may be due the low thickness of the Neogene aquifer and the interaction between shallow and deep groundwater, enhanced by over pumping activities. In the south east of the study area (Selga Saadoun), a depression (low elevation) with very high salinity is identified, which may be due to the result of the infiltration and recharge of groundwater with water accumulated from upstream farms. The spatial distribution of groundwater EC shows generally increasing trends from the east to the center and from the west to the center of the study area.



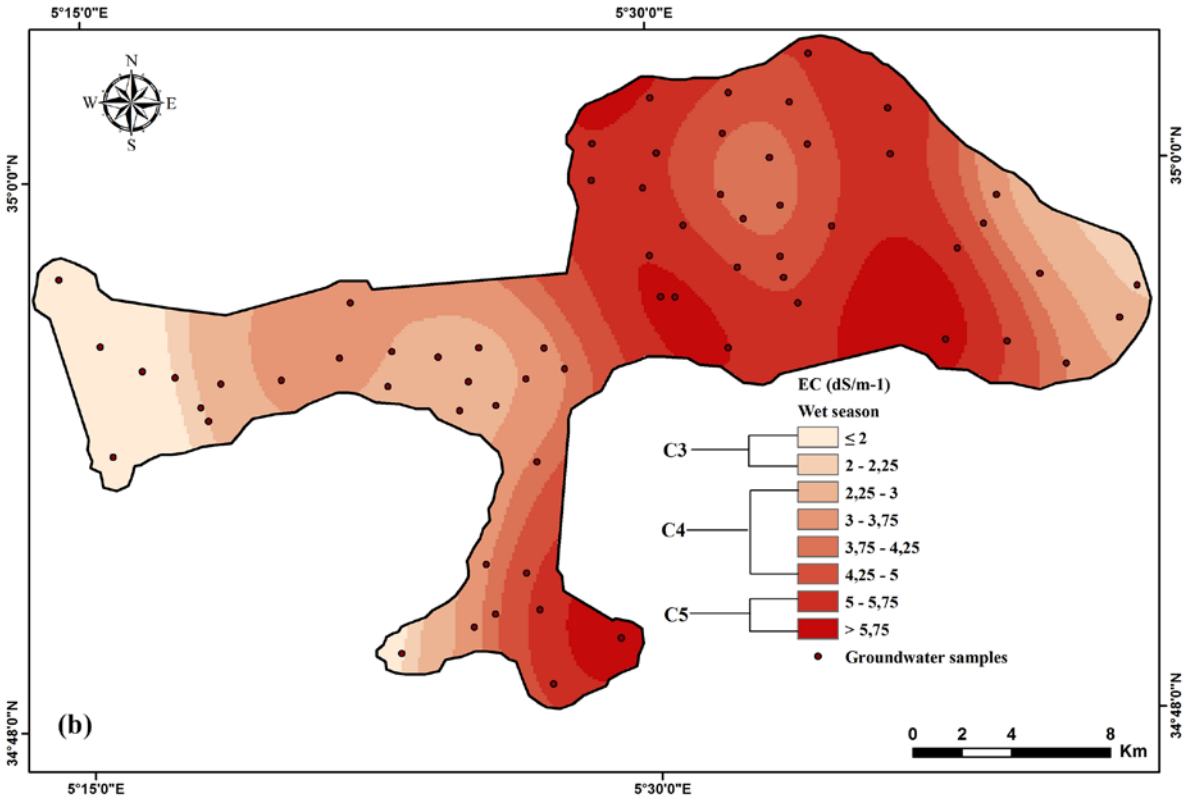


Figure 44: Spatial distribution maps of groundwater EC [(a) Dry season, (b) Wet season]

4. Conclusion

Hydro-geochemical analyses and Ordinary Kriging are used to assess the suitability of groundwater for human consumption and agricultural uses in El Outaya basin. The results of hydro-geochemical analysis reveal that Na^+ and Ca^{2+} are the dominant cations and Cl^- and SO_4^{-2} are the dominant anions in the groundwater. The trilinear Piper diagram shows the prevalence of $(Na^+ + K^+) - (Cl^- + SO_4^{-2})$ and $(Ca^{2+} + Mg^{2+}) - (Cl^- + SO_4^{-2})$ hydrochemical facies. The saturation indices indicate that the dissolution of evaporitic rocks (halite, calcite and anhydrite) and reverse ion exchange are the main hydrochemical processes that control the groundwater chemistry as a result of rock-water interaction.

WQI shows that the majority of groundwater samples are very poor or unsuitable for human consumption. The spatial distribution maps of WQI reveal that the groundwater quality decreases from the west (El Mazouchia and Oued Besbes) to the middle of the study area and from the east (Branis) to the north west of the study area. According to the Riverside diagram, the most of groundwater samples are falling under C4S2 and C4S3 classes, which can be used for irrigation on coarse textured or organic soils with good permeability provided that more tolerable crops should be selected. About 30 % of the groundwater are unsuitable for irrigation

purpose (C5S3 and C5S4 classes). Wilcox diagram illustrates that the majority of the groundwater samples are unsuitable for irrigation, with a high risk of deterioration of the soil structure in the study area using this groundwater for irrigation.

According to the prediction maps of groundwater EC, the spatial distribution of salinity shows general increasing trends from the east to the center and from the west to the center of the study area. The dominance of groundwater with high and very high salinity in the study area could be due to the intense exploitation of the aquifer, the effects of highly saline irrigation return flows on the groundwater, the low rate of precipitation and the dissolution of Triassic evaporitic materials in Djebel Elmaleh. It is concluded that the use of the groundwater for irrigation in El Outaya plain is bound to deteriorate the soil structure, increase soil salinization, damage the cultivated crops and reduce yields. It is suggested that appropriate leaching and drainage systems can be applied to minimize soil damage, and desalination of groundwater by reverse osmosis, which can decrease the EC and TDS of the groundwater more than 95% (Belkacem et al., 2007).

Chapter IV

Spatial variability and risk assessment of nitrate groundwater pollution

1. Introduction

Groundwater is the main source of industrial, agricultural and drinking supply in many regions all above the world, especially in arid and semi-arid areas, including El-Outaya plain in southeastern Algeria. In general, groundwater in the study area is intended for agricultural use, because of its high salinity (Boudibi et al., 2019). Despite that, it is found that a lot of farmers use it as a source of drinking water. However, population growth, intensive agriculture and over-exploitation of groundwater might contaminate this scarce and valuable resource and render it unsuitable composition for human consumption, which is of great risk for human health (Babiker et al., 2003).

Groundwater contamination by nitrate is an important and a globally growing environmental problem due to natural and anthropogenic factors (Fan and Ding, 2010; Chica-olmo et al., 2014). World Health Organization (WHO) has fixed a value of 50 mg/l of NO_3^- or 11.3 mg/l of NO_3-N as a permissible limit of nitrate for potable water (WHO, 2011). The European Water Framework Directive (FDW) has established a categorization of good quality water in terms of nitrate concentration at a level of 37.5 mg/l of NO_3^- or 8.9 mg/l of NO_3-N (Chica-Olmo et al., 2014). Some authors (Fang and Ding, 2010; Eckhardt and Stackelberg, 1995) have attested that groundwater nitrate concentration exceeding a threshold of 13 mg/l of NO_3^- or 3 mg/l in terms of NO_3-N contaminated due to human activities.

Environment researchers have used many methods to assess and study the origins of contamination of groundwater by different pollutants. Geostatistical and deterministic interpolation methods are widely used for this purpose. Kriging is a common term given for a range of stochastic methods to provide an optimal spatial prediction (Oliver and Webster, 2014; Krivoruchko, 2012). Kriging methods are powerful sophisticated weighted-average techniques (Şen, 2016; Hsu et al., 2017). They are commonly used in geosciences, hydrogeology, and especially, in monitoring and managing groundwater resources. These methods involve two phases, namely, theoretical variogram, which requires fitting mathematical models to experimental calculated semi-variograms, and a process of Kriging (André, 2017; Abzalov, 2016). Geostatistical Kriging interpolation techniques include two broad families as linear Kriging algorithms, including ordinary Kriging (OK), simple Kriging (SK), universal Kriging (UK) and Bayesian Kriging (BK); nonlinear kriging algorithms, also indicator kriging (IK), disjunctive kriging (DK), probability Kriging (PK), cokriging (COK), lognormal Kriging (LK) and multi-Gaissian Kriging (MGK) (Şen, 2016; Asa et al., 2012; Arslan, 2012; Oliver, 2010). On the other hand, deterministic interpolation techniques can be divided into two categories as

global interpolators (GI) and local interpolators (LI). For GI, the predictions are calculated using the entire dataset, such as global polynomial interpolator (GPI), and for LI the predictions are calculated from sample points within neighborhoods, as an example, inverse weighted distance (IDW), which generates surfaces from measured points using mathematical functions based on the extent of similarity, and radial basis functions (RBF), where the mathematical functions are based on the degree of smoothing (ESRI, 2019a; Adhikary, 2014). In this study, the results of three geostatistical techniques, two linear Kriging methods (OK and EBK) are investigated together one non-linear Kriging method, (IK), and a deterministic technique (RBFs).

The investigated interpolation techniques are the object of many studies in several disciplines, for instance, Xie et al. (2011) used OK. RBFs are employed to study the spatial variability of soil heavy metal pollution in Beijing, China. Arslan (2012) has evaluated OK and IK techniques for the spatial and temporal mapping of groundwater salinity in Bafra plain, Turkey. Adhikary and Dash (2014) worked on the prediction of spatial variability of groundwater depth using RBFs, on the other hand, OK, IDW and UK in Delhi, India. Samsonova et al. (2017) conducted EBK and OK analyses for revealing heterogeneities in the distribution of organic carbon on agricultural lands in Bryansk Opole region, Russia. Hsu et al. (2017) used OK, EBK and other interpolation techniques to estimate street level air temperature (T °C), and to study the potential effect of some environmental characteristics on the spatial distribution of T °C. Li et al. (2019) has applied EBK for managing groundwater by quantifying its spatial and temporal evolution in Guanajuato, México.

All these studies agreed that geostatistical techniques are the most accurate in predicting the spatial distribution of geospatial variables. During the fieldwork for this study, it is noticed that groundwater of the Mio-Pliocene aquifer is the main source of agricultural irrigation in El-Outaya plain, which means that the subject of the study area depends on different farming activities (application of different chemical treatments. chemical and organic fertilization), but farmers neglect its high salinity and they are still using it for drinking purposes. The field observations led to the question about the contamination of this valuable resource and its suitability for human consumption in terms of NO3-N concentration.

The aim of this chapter is on the one hand to predict the spatial variability of nitrate-N concentration using OK. EBK and RBFs, and on the other hand, IK is used to generate probability maps of groundwater pollution by [NO3-N] taking into consideration the abovementioned thresholds. A simple method is implemented to transform probability maps to

categorical maps. Using the [NO3-N] of the dry season, the accuracy of the different geostatistical and deterministic interpolation techniques is then assessed and the best one is determined. The best technique is used to predict the spatial distribution of nitrate-N concentrations of the wet season campaign.

2. Material and methods

2.1. Groundwater samples collection and analyses

Figure 45 shows the location of the 136 groundwater samples collected from 68 sites over two campaigns (dry season of 2017 and wet season of 2018) from boreholes distributed in different agricultural farms in the study area. The well locations are obtained using a global positioning system (GPS) device, and water samples are collected from each well.

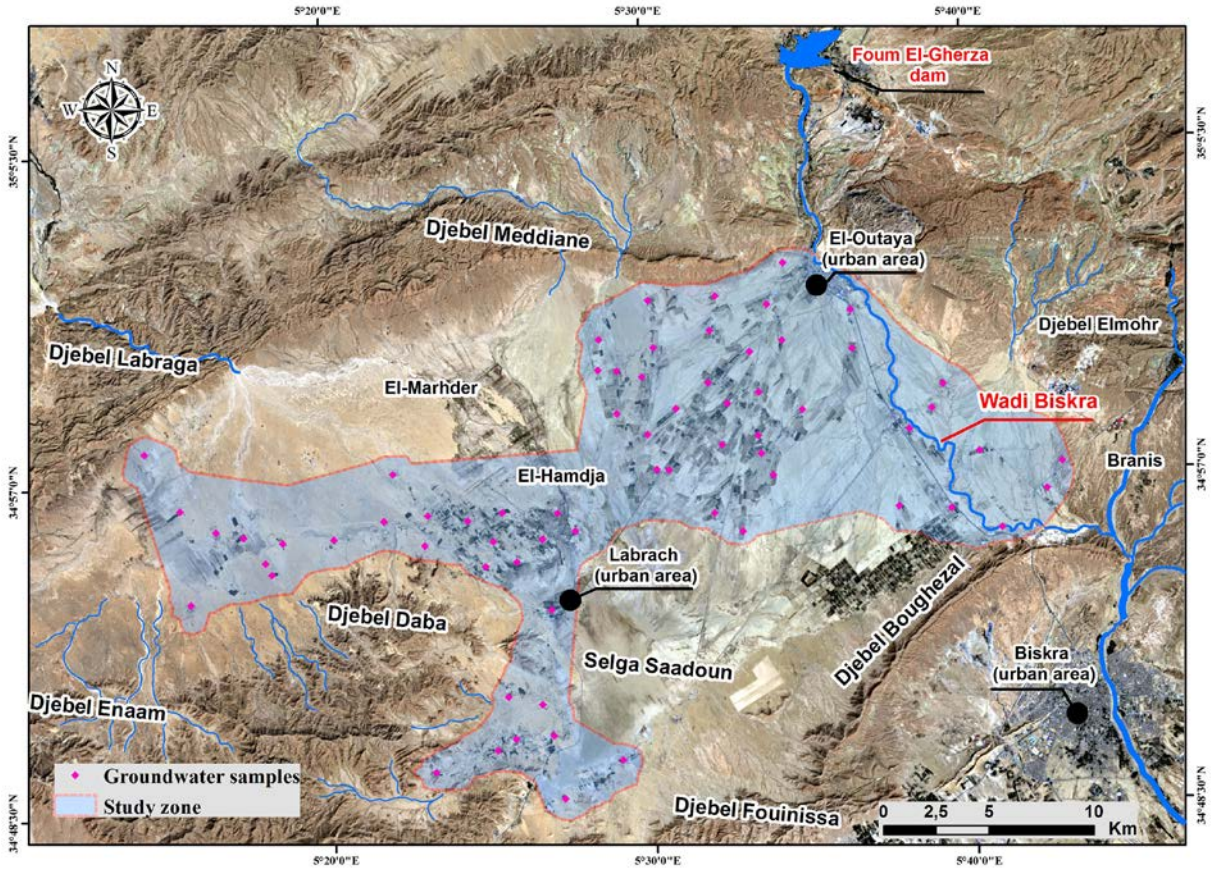


Figure 45: Location of the collected groundwater samples and analyzed for [NO3-N]

The samples are collected in polyethylene bottles are rinsed using distilled water and washed three times by the groundwater of the same sample. The majority of the boreholes are found in operation. Otherwise, the well water is pumped for few minutes before sampling. The samples are transported in coolers and conserved in the laboratory at a temperature of 4°C.

UV-visible spectrophotometer is used to measure nitrate-N concentration (mg/l) in groundwater samples. SPSS software v22.0 is employed to calculate descriptive statistics, and VARIOWIN software v2.2 for making and fitting semi-variograms (OK method). ArcGIS 10.2 software with Geostatistical Analyst Extension is used to run different interpolation methods and to generate various prediction maps.

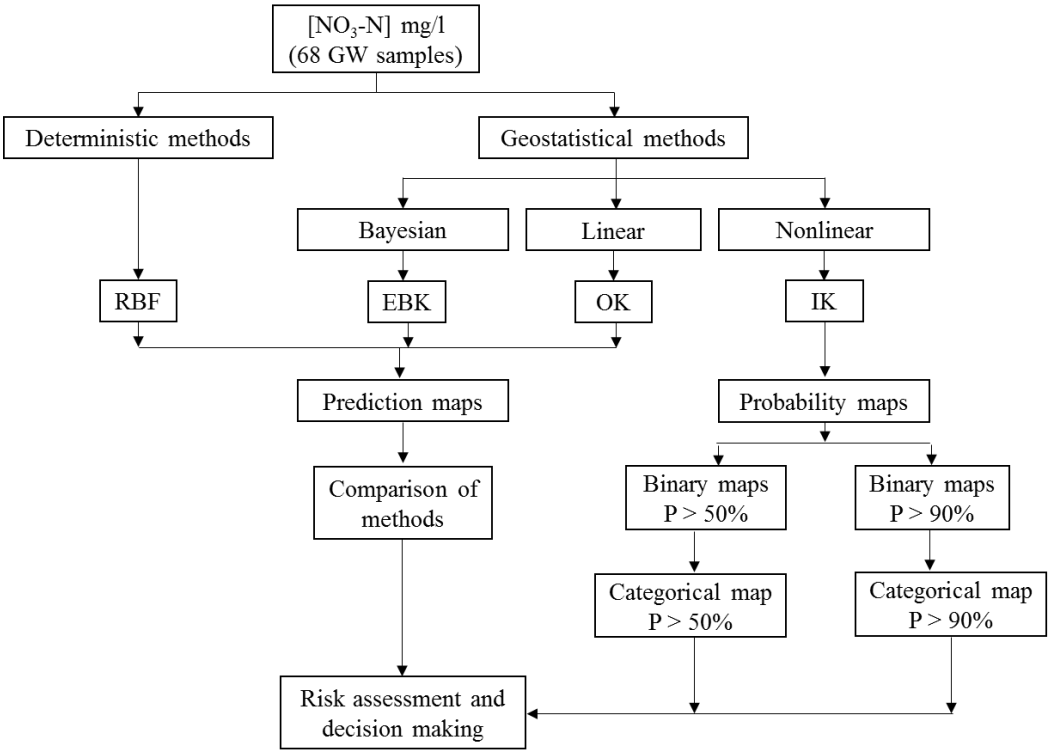


Figure 46: Flowchart of the used methodology

2.2. Kriging (EBK, OK and IK)

Kriging is a powerful geostatistical interpolation technique based on the regionalized variable theory, which describes a variable that is spread out in space and/or in time and which shows a certain structure (Delhomme, 1978). Kriging methods (simple, ordinary, disjunctive, indicator, universal, etc.) depend on mathematical and statistical models. The addition of a statistical model that includes probability separates Kriging methods from the deterministic methods (global polynomial interpolation, inverse weighted interpolation, radial basis function, etc.) for spatial interpolation (ESRI, 2019a).

The semi-variogram is the central tool in geostatistical methods and is very useful for visualization, interpretation and modeling of spatial continuity of the studied variables (Arslan, 2012; Şen, 2020). It is defined as half the average squared difference between the attribute values at all points separated by a lag distance *h* as follows (Arslan, 2012):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (23)$$

where $\gamma(h)$ is the experimental semi-variance value for all pairs at a lag distance *h*; $Z(x_i)$ and $Z(x_i + h)$ are the studied variable values at locations x_i and $(x_i + h)$, respectively; N(h) is the number of observation pairs separated by a distance h (Isaaks and Srivastava, 1989; Arslan, 2012).

As already mentioned in the previous sections, in this chapter, three Kriging methods, namely, [linear (OK), nonlinear (IK), and Bayesian (EBK)] are applied to the nitrate-N concentration.

2.2.1. Ordinary Kriging

Ordinary Kriging (OK) is the most used geostatistical technique for interpolation. The Kriging equations and the function of the semi-variogram are used to derive the weights of OK (Xie, 2011). OK is given as follows:

$$Z^*(x_0) = \sum_{i=1}^{n} \lambda_i \, . Z(x_i) \qquad \text{with} \qquad \sum_{i=1}^{n} \lambda_i = 1 \quad (24)$$

where $Z^*(x_0)$ is the random variable at the location x_0, Z(xi) is the measured value at a location x_i, λ_i is the weighting factor assigned to $Z(x_i)$ and n is the number of observations (Journel, 1986; Bradaï et al., 2016).

OK is a linear unbiased estimate, because its estimates are weighted by linear combinations of the available data, and unbiased, because it tries to have the mean error equal to 0. It aims at minimizing the variance of the errors, and hence, distinguishes OK from other methods (Isaaks and Srivastava, 1989). Mathematically, these conditions are expressed as follows (Delhomme, 1978):

$$E\{Z^*(x_0) - Z(x_0)\} = 0 \quad (25)$$

$$Var\{Z^*(x_0) - Z(x_0)\} = \text{minimum} \quad (26)$$

2.2.2. Emperical Bayesian Kriking

Emperical Bayesian Kriging (EBK) is an automatic based interpolator that differs from other Kriging methods by adjusting the parameters of the spatial process and accounting for the errors to estimate the semi-variogram model and by using the intrinsic random function as the Kriging model (Krivoruchko, 2012; Samsonova, 2017). The classical Kriging methods require the manual intervention to calculate the parameters of the semi-variogram and EBK calculates these

parameters automatically through a process of sub-setting and simulations, where power, linear and thin plate spline are the supported semi-variograms in ArcGis geostatistical analyst (without transformation) (ESRI. 2019b). In this chapter, the following semi-variogram power model is used:

$$\gamma(h) = C_0 + b|h|^a \quad (27)$$

where, C_0 is the nugget, b is the slope and a is the power.

Krivoruchko (2012), Krivoruchko (2019) and ESRI (2019b) described the internal process of EBK modeling as follows:

1) A semi-varigram and the parameters of the spatial process are estimated using the data.

2) Using this semi-variogram and parameters, new values are simulated at each location of the input data.

3) New semi-variogram and parameters are estimated from the simulated data. A weight for the simulated model is calculated using Bayes' rule. It shows how likely the observed values can be generated from then estimated model.

4) Repetition of the two previous steps, where in each repetition, the semi-variogram estimated in the first step is used to simulate a new set of values at the input locations. The simulated data is used to estimate a new semi-variogram model and its weight.

5) Predictions and prediction standard errors are produced at the un-sampled locations using these weights.

2.2.3. Indicator kriging

Indicator kriging (IK) is the most commonly used non-linear geostatistical technique for the estimation of probability of exceeding a defined threshold, z_c, in an un-sampled location x_i (Liu. 2004). IK differs from other geostatistical methods by estimating the conditional probability distribution directly, and it does not require any assumption regarding the distribution of the regionalized variable (Bierkens, 1993). IK approach requires the transformation of the spatial variable, Z(xi), into an indicator variable, $I(x_i ; z_c)$, with a binary distribution as follows:

$$I(x_i; z_c) = \begin{cases} 1. & \text{if } Z(x_i) > z_c \\ 0. & \text{if otherwise} \end{cases} \qquad i = 1. \ldots . n \quad (28)$$

The expectation of $I(x_i ; z_c)$, conditional to n surrounding data is given as:

$$E\left[I\left(x_i; z_c|(n)\right)\right] = Prob\{Z(x_i) \leq z_c|(n)\} = F\left(x_i; z_c|(n)\right) \quad (29)$$

$$Prob\{Z(x_i) > z_c|(n)\} = 1 - F\left(x_i; z_c|(n)\right) \quad (30)$$

where, $F(x_i;z_c|(n))$ is the conditional cumulative distribution function (CCDF) of $Z(x_i) \leq z_c$.

For each threshold (z_c) an experimental semi-variogram is calculated for the CCDF and fitted to a theoretical model to estimate the semi-variogram parameters, which are used to get an estimate, $I^*(x_0 ; z_c)$, at an unsampled location x_0 using the following equation:

$$I^*(x_0; z_c) = \sum_{i=n}^{n} \lambda_i(z_k)\, I(x_i; z_c) \quad (31)$$

where $I(x_i ; z_c)$ represents the values of the indicator at the sampled locations, x_i, and λ_i is the weighting factor of $I(x_i ; z_c)$ used in estimating $I^*(x_0 ; z_c)$ (Lee. 2007; Triantafilis. 2004).

The application of this formula, which is an OK equation, generates values from 0 to 1 for un-sampled locations. These values represent the conditional probability that the regionalized variable at un-sampled locations exceeds the defined threshold.

2.3. Radial Basis Function (RBF)

Radial basis functions (RBF's) are series of exact interpolators used to produce smooth surfaces from a large number of data points and require the surface to pass through the measured points (ESRI, 2019a). The basic equations used in RBF's are dependent on the distance between the predicted values and the measured ones (Xie, 2011). The interpolation equation, $Z(x)$, can be expressed as the sum of two components as follows (Xie, 2011; Mitášová and Mitáš, 1993):

$$Z(x) = T(x) + \sum_{j=1}^{N} \lambda_j\, R(d_j) \quad (32)$$

The trend function T(x) is given by:

$$T(x) = \sum_{i}^{M} a_i f_i(x) \quad (33)$$

where $f_i(x)$ is a set linearly independent functions, $R(d_j)$ is a radial basis function and d_j is the distance between the measured and the prediction locations. Once $R(d_j)$ is known, coefficients a_i and λ_i are calculated by solving the following system of linear equations (Xie, 2011; Mitášová and Mitáš, 1993):

$$Z(x_k) = \sum_{i}^{M} a_i f_i(x_k) + \sum_{j=1}^{N} \lambda_j R(d_{jk}) \quad \text{for } k = 1.2. \ldots . N \tag{34}$$

$$\sum_{i}^{N} \lambda_k f_i(x_k) = 0 \quad \text{for } k = 1.2. \ldots . M \tag{35}$$

There are five different basis functions (BF), which are thin-plate spline (TPS), spline with tension (ST), completely regularized spline (CRS), multi-quadratic function (MQ) and inverse multi-quadratic function (IMQ). Each RBF has different results in a different interpolation surface (ESRI., 2019a). In this chapter, the RBFs are used with the choice of the best one (MQ) based on the cross-validation results. MQ ranks the best in accuracy (Chen, 2002). The MQ basis function is given by:

$$R(d_j) = \sqrt{d^2 + c^2} \tag{36}$$

where d is the distance from the measured to the prediction location and c is a smoothing factor (Xie, 2011).

2.4. Variography and cross validation

The semi-variogram is a special geostatistical tool applied for modelling the spatial continuity of the regionalized variables (Abzalov, 2016). The experimental semi-variogram is calculated using Eq. 1 only for some lag distances but geostatistical modeling requires values at any lag distance, because of this reason an accurate theoretical model should be fitted to the experimental semi-variogram (Ma, 2019). Stable, circular, power, exponential, spherical and Gaussian are the most used theoretical models. The three important parameters of theoretical models are the sill (C), the range (a), and the nugget effect (C_0). In this chapter, spherical model is employed for OK method, power model for EBK and exponential models for IK method. The choice of the best fitted model is based on the indicator goodness of fit (IGF) value of VARIOWIN software (IGF indicates a good fit when it's close to zero) for OK, and the cross validation results for EBK and IK. The spatial dependence is investigated using the nugget to sill ratio (r), where r > 75% indicates weak spatial dependence; 25% < r < 75% moderate spatial dependence and r < 25% indicates strong spatial dependence (Cambardella, 1994)

Cross validation is the commonly used method for comparing and checking the accuracy of interpolation techniques. Where all the available data are used to predict the spatial distribution of the regionalized variable in the study area, once the prediction model is built, each of the

measured value at a location x_i is removed from the data set, and then estimated using the prediction model. To provide accurate interpolation results, the mean error (ME) and the mean error standardized (MES) should be close to 0, the root mean square error (RMSE) should be as small as possible and near to the average standard error (ASE), the root mean square error standardized (RMSES) should be close to 1. The error measures are given by the following equations:

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n}\left[e_{(x_i)}\right] \quad (37)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[e_{(x_i)}]^2} \quad (38)$$

$$\mathrm{ASE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\sigma^2{}_{(x_i)}} \quad (39)$$

where, $e_{(xi)}$ is the difference between the measured and the predicted value at a location x_i, and $\sigma^2{}_{(xi)}$ is the variance (Asa, 2012). In MES and RMSES equations, $e_{(xi)}$ are divided by the standard deviation.

3. Results and discussion

3.1. Descriptive statistics and data distribution

The statistical summary of nitrate-N concentration of the 136 GW samples is illustrated in Table 17. [NO_3-N] ranged from 1.52 to 22.03 mg/l and 1.54 to 22.09 mg/l, with an average value of 9.88 mg/l and 9.08 mg/l for the dry and wet season, respectively. The results show a high degree of variability with a coefficient of variation of 46.35% in dry period and 48.78% in wet period. The descriptive statistics shows almost the same parameter values for both seasons. The frequency distribution of [NO_3-N] is given in Figure 47a. The groundwater (GW) samples are grouped into four water quality categories based on Nitrate-N concentration (The same number of samples for each group is observed for both seasons). Excellent quality category is observed with 10.29% of GW samples, where. [NO_3-N] is less than or equal to the so-called human affected value, ([NO_3-N] $\leq$ 3 mg/l) (Babiker, 2004). Good quality category with 25% of the GW samples (3 mg/l < [NO_3-N] $\leq$ 8.9 mg/l) occurs according to 8.9 mg/l, which is the threshold of good quality class established by the European Water Framework Directive (Chica-Olmo, 2013). Medium quality category is the dominant class with 35.29% of the GW samples (8.9

mg/l < [NO_3-N] ≤ 11.3 mg/l), where 11.3 mg/l is the permissible limit of WHO (WHO. 2011). Poor quality category with 29.41% of GW samples exceeding the permissible limit of WHO ([NO_3-N] > 11.3 mg/l).

The statistic Kolmogorov-Smirnov (KS) test is used the check the normal distribution of the [NO_3-N]. The KS value is significant with 0.064 for the dry season. The skewness and the kurtosis are 0.39 and 0.32, respectively. The skewness and the excess kurtosis are positive and close to 0 which confirm the normal distribution of nitrate-N concentration (Figure 47b). For the wet season, KS value is not significant.

Table 17: Statistical analysis of groundwater [NO_3-N] (mg/l)

Statistical parameters	[NO_3-N] (mg/l) Dry season	[NO_3-N] (mg/l) Wet season
Number of samples	68	68
Mean	9.88	9.08
Median	9.79	9.01
Minimum	1.52	1.54
Maximum	22.03	22.09
Variance	20.96	19.59
Standard deviation	4.58	4.43
Coefficient of variation (%)	46.35	48.78
Skewness	0.39	0.40
Kurtosis	0.32	0.44
Kolmogorov-Smirnov test	0.064	0.025

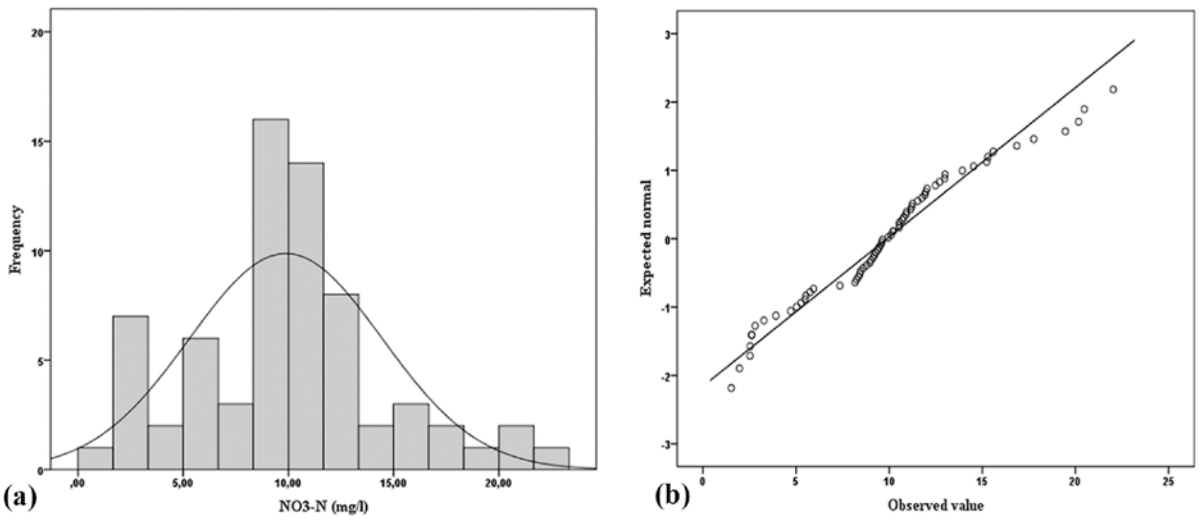


Figure 47: (a) Frequency plot of [NO3-N] values; (b) scatter plot of [NO3-N] normal distribution

3.2. Spatial distribution of nitrate-N concentration

Three interpolation methods are employed to estimate the un-sampled locations and to generate variability maps of [NO_3-N] in the study area by means of two stochastic methods: linear (OK) and Bayesian (EBK), deterministic method (RBF).

3.2.1. Variography and cross-validation

Figure 48b presents the omnidirectional semi-variogram of the groundwater nitrate-N used for OK, which indicates that the spherical semi-variogram is the best fitted model with an IGF of $2.21*10^{-2}$. The nugget effect is 3.95 $(mg/l)^2$, the partial sill is 17.14 $(mg/l)^2$ and the range is 7223.81 m, and this high value indicates the spatial continuity of groundwater nitrate-N concentrations (Bradaï, 2016). The nugget to sill ratio is 18.73 %, indicating a strong spatial dependence of [NO_3-N].

Figure 48a shows the estimated spectrum semi-varioram models (blue lines) using the automated interpolator EBK, which are implemented in ArcGis geostatistical analyst to predict [NO_3-N] at un-sampled locations. The blue crosses represent the empirical semi-variogram. The solid red line is the median of the distribution and the red dashed lines are the 25^{th} and 75^{th} percentiles. The models do not have a range and sill, because the power function does not have an upper bound.

The cross validation, which represents the accuracy of the different interpolation methods, shows accurate prediction results. For stochastic methods (EBK and OK) the ME and MES values are close to 0, RMSES values are almost equal to 1 and the RMSE and ASE values are close to each other. For the deterministic method (RBF), the ME and RMSE indicate either satisfactory prediction results. The numerical results of the different calculated error measures are presented in Table 18.

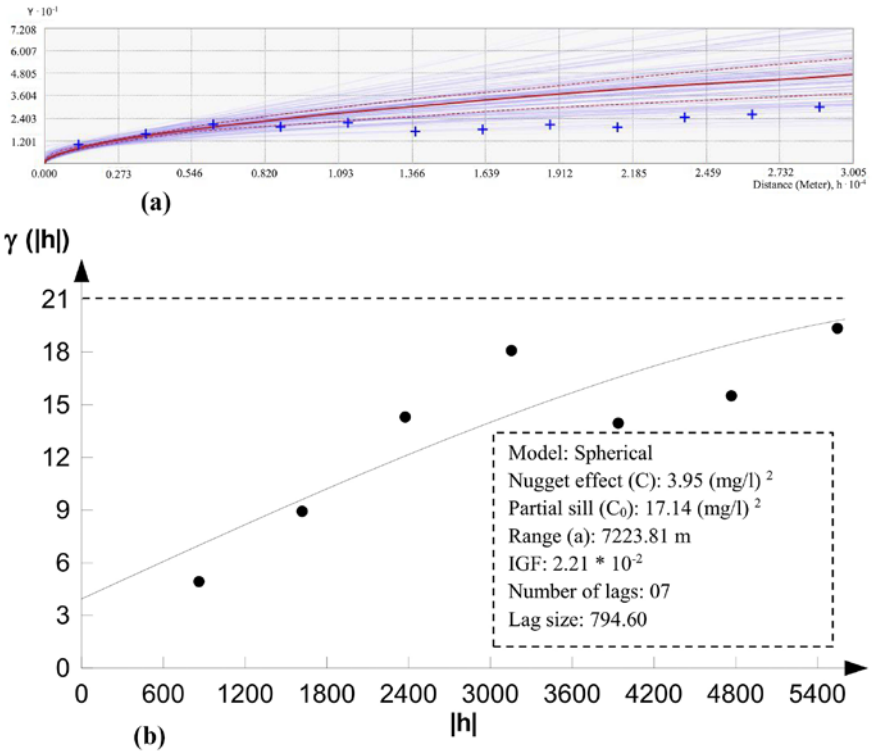


Figure 48: Semivariograms of the predicted [NO3-N]: (a) EBK , (b) OK

Table 18: Cross validation of EBK, OK and RBF

Prediction method	Model	ME	RMSE	ASE	MES	RMSES
EBK	Power	-0.0021	3.245	3.50	0.0006	0.945
OK	Spherical	0.0423	3.28	3.57	0.0083	0.9406
RBF	Kernel function (Multiquadtratic)	0.0687	3.23	/	/	/

3.2.2. Prediction maps

The obtained prediction maps using OK (Figure 49a), EBK (Figure 49b) and RBF (Figure 49c) show the same overall spatial distribution of groundwater nitrate-N concentration in the study area. The only difference is in the shape and the size of the area of different categories. Table 19 represents the surface areas of the four categories of [NO_3-N] estimated using the three interpolation techniques. The groundwater with intermediate quality ($8.9 < [NO_3\text{-N}] \geq 11.3$) is the dominant class with surface areas of 39.49%, 39.22% and 36.20% for EBK, OK and RBF, respectively, which are located in the middle and the west of the study area. The category with good quality ($3 < [NO_3\text{-N}] \geq 8.9$) represents 32.63% of the study area by using EBK, 32.18% by using OK and 33.25% by using RBF, this class is located in the east and a partially in the south and south-west of the study area. The category with excellent groundwater quality ([NO_3-

N] ≤ 3) occupies a small surface area in the west of the study area with 1.02% for EBK method, 1.75% for OK and 3.6% for RBF. This class is overestimated by OK and RBF. The category that represents the contaminated groundwater, where nitrate-N concentration exceeding the permissible limit of WHO is located in the north, the south and the west part of the study area, with surfaces areas of 26.86%, 26.85% and 26.94% for EBK, OK and RBF, respectively.

The comparison of the three predicted maps using EBK, OK and RBF shows that they have the same general structure despite of the difference in theoretical aspect of the used methods, which validate the obtained results and the spatial distribution of nitrate-N concentration in El-Outaya plain.

Table 19: Surface areas of different nitrate-N categories estimated using EBK, OK and RBF

Categories	EBK		OK		RBF	
	Area (ha)	Area (%)	Area (ha)	Area (%)	Area (ha)	Area (%)
≤ 3	411.5	1.02	710.7	1.75	1458.1	3.60
3 < NO_3-N ≥ 8.9	13215.9	32.63	13031.4	32.18	13467.8	33.25
8.9 < NO_3-N ≥ 11.3	15994.8	39.49	15884.9	39.22	14661.0	36.20
NO_3-N > 11.3	10876.9	26.86	10872.1	26.85	10912.3	26.94
Total	40499.1	100	40499.1	100	40499.1	100

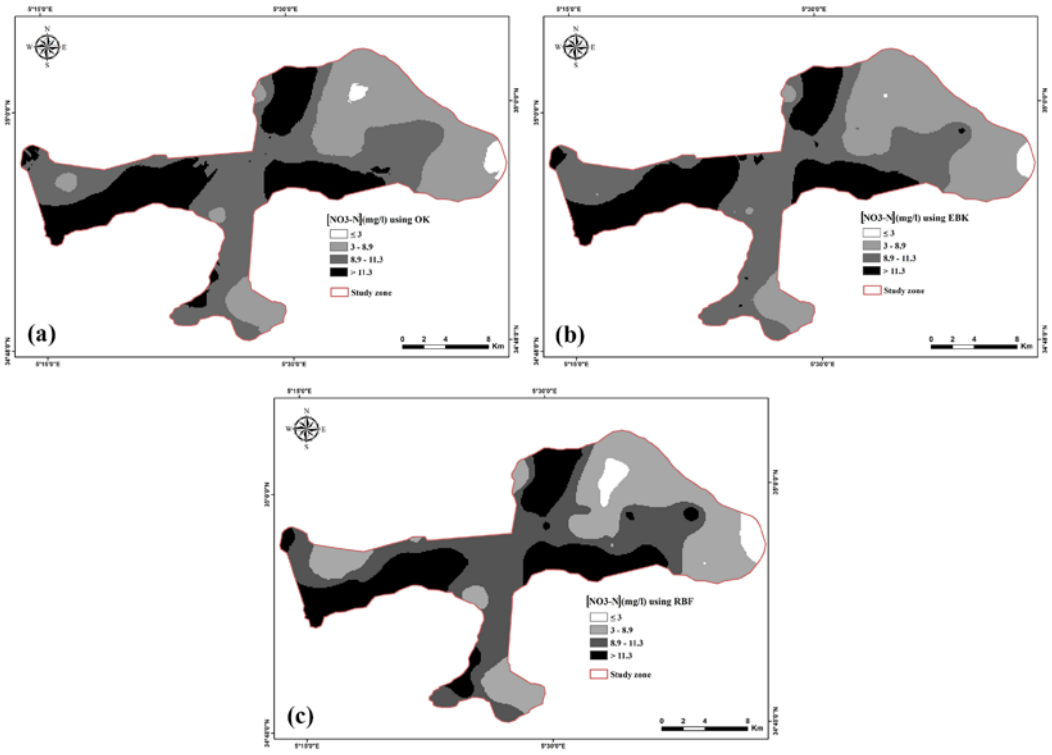


Figure 49: Prediction maps of [NO3-N] classes: (a) OK, (b) EBK and (c) RBF

3.3. Nitrate-N probability mapping using IK

IK is employed to generate groundwater nitrate-N concentration probability maps. For this purpose, three thresholds are selected as the first one (z_1 = 3 mg/l) is the so-called human affected value; the second one (z_2 = 8.9 mg/l) is the high limit of good quality established by the European Water Framework Directive (EWFD) and the third threshold (z_3 = 11.3 mg/l) is the permissible limit of WHO. On the bases of these thresholds, one can retain four groundwater quality categories as excellent quality ($[NO_3\text{-}N] < Z_1$), good quality ($z_1 < [NO_3\text{-}N] \leq z_2$), medium quality ($z_2 < [NO_3\text{-}N] \leq z_3$) and poor quality ($[NO_3\text{-}N] > z_3$).

3.3.1. Variography and coss validation

Table 20 shows the characteristics of the best-fitting semi-varigram models (model, nugget, partial sill and the range) and the cross validation results (ME, RMSE and ASE) of the different thresholds using ArcGis Geostatistical Analyst. As the table indicates, the exponential model is selected as the best fit for the three thresholds, with ranges of 4793 m, 2935 m and 2798 m in ascending order. The nugget to sill ratios of the three models indicate a strong spatial dependence for the first ($[NO_3\text{-}N] > 3$ mg/l) and the second ($[NO_3\text{-}N] > 8.9$ mg/l) thresholds, and a moderate spatial dependence for the third ($[NO_3\text{-}N] > 11.3$ mg/l) threshold. The error measures validate the chosen model for each threshold, where, ME are very close to 0 and the difference between RMSE and ASE is 0.2751, 0.4379 and 0.4652 for the three threshold in ascending order, respectively.

Table 20: Crossvalidation and parameters of the semivariogram models of $[NO_3\text{-}N]$ thresholds using IK

Threshold	Model	Nugget effect $(mg/l)^2$	Partial sill $(mg/l)^2$	Range (m)	ME	RMSE	ASE
3	Exponential	0	0.0922	4793	-0.0026	0.2389	0.2751
11.3	Exponential	0.0179	0.1708	2935	0.00014	0.4252	0.4379
8.9	Exponential	0.086	0.1145	2798	0.0029	0.4264	0.4652

3.3.2. Probability maps

The probability maps are given in Figure 50. They represent the spatial probability that $[NO_3\text{-}N]$ could exceed the quality thresholds (z_1, z_2 and z_3) and enlighten about the nitrate-N groundwater pollution. Furthermore, to estimate the affected and the non-affected areas for each threshold, one can proceed with the transformation of probability maps to binary maps based

on two levels of significance as the fifty-fifty risk to exceed a given threshold ($P(Z(x_i) > z_c) > 50\%$) and the high risk to exceed thresholds ($P(Z(xi) > zc) > 90\%$). Binary maps of a level over 50% for surpassing z_1, z_2 and z_3 are given in Figure 51a, 52b and 52c, respectively, and those of a level over 90% for surpassing z_1, z_2 and z_3 are given in Figure 51d, 52e and 52f, respectively. Thus, at levels over 50%, the percentages of affected area for each threshold are 92.85% for z_1, 70.93% for z_2 and 20.35 for z_3, while at a level over 90% (high risk to exceed the defined referenced value), the percentage of affected area was 77.10% for z_1, 21.80% for z_2 and 0.88% for z_3.

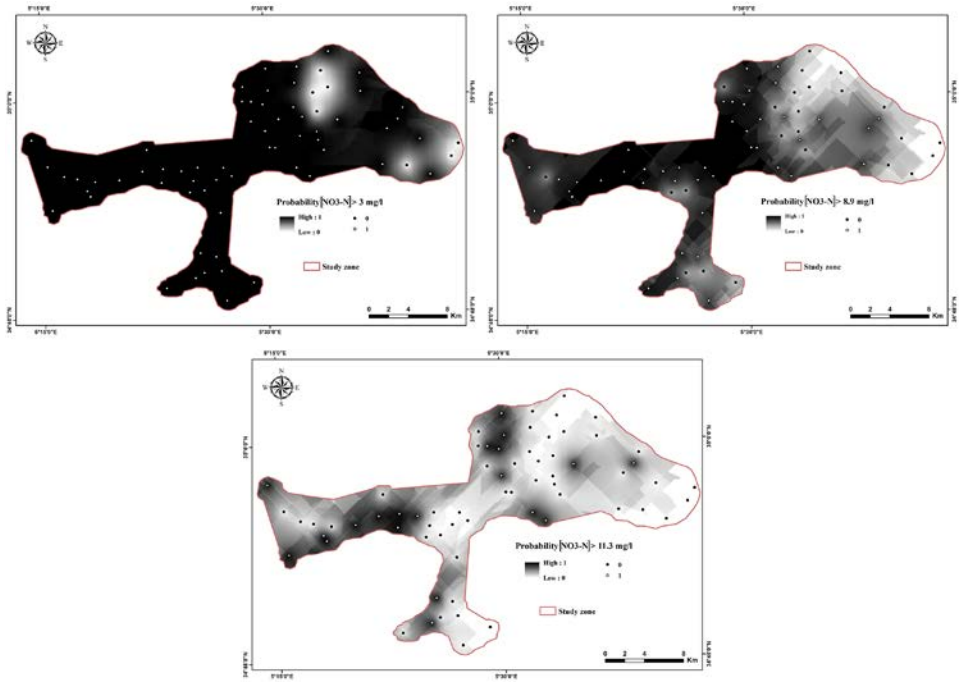


Figure 50: Probability maps of [NO3-N] obtained using IK: (a) probability map of [NO3-N] > 3 mg/l, (b) probability map of [NO3-N] > 8.9 mg/l, (c) probability map of [NO3-N] > 11.3 mg/l

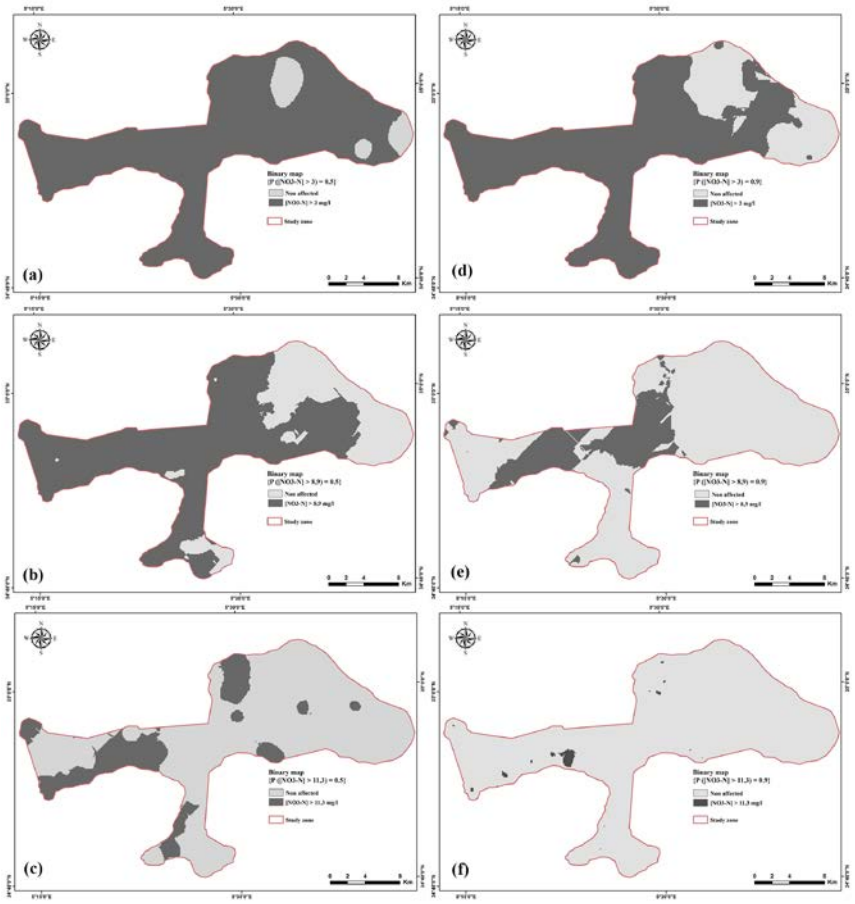


Figure 51: Binary maps: (a), (b) and (c) binary maps obtained using a probability over 50%; (d), (e) and (f) binary maps obtained using a probability over 90%.

3.3.3. Categorical maps

The conversion of binary maps into categorical maps is for aim to show the spatial distribution of the different groundwater quality categories in the study area according to two levels of contamination risk as the categorical map with medium risk (Figure 53a) at a level of 50% and another one with high risk (Figure 53b) at a level of 90%.

Figure 52 provides a general explanation of the methodology used to elaborate categorical maps from probability maps in two steps. As the first step, probability maps are transformed to binary maps taking into consideration the two levels of significance (50% and 90%), so at a given threshold, each cell of the affected area takes a value of 1, and a value of 0 is assigned to the cells of the non-affected area. In the second step, the generated binary maps are summed together to obtain categorical maps that contain four nitrate-N concentration at different levels of significance (Figure 53).

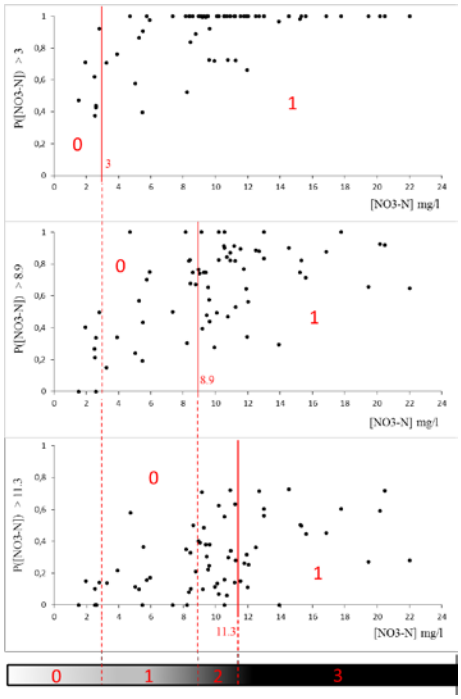


Figure 52: Scatter plots of measured [NO3-N] and probability values at each threshold, with explanation of the methodology used to transform indicator predictions to binary values and categorical values

The final categorical maps show that 20.30% of the aquifer's surface in the study area presents groundwater of poor quality for human consumption, which exceed the standard value of WHO (11.3 mg/l) at a level of significance of 50% (Figure 53a) and only 0.88% at a level of significance of 90% (Figure 53b). The excellent water quality presents 7.15 % of the surface at a level of 50% significance, while at a level of 90% for exceeding the human affected value (3mg/l), the surface increased to 22.9%. The surface of good quality category increased from 50.69% at a level of significance of 50% to 55.24% at a level of 90%, while the surface of intermediate quality dropped from 50.69% to 21.05% at levels of significance of 50% and 90%, respectively. The results are reported in Table 21.

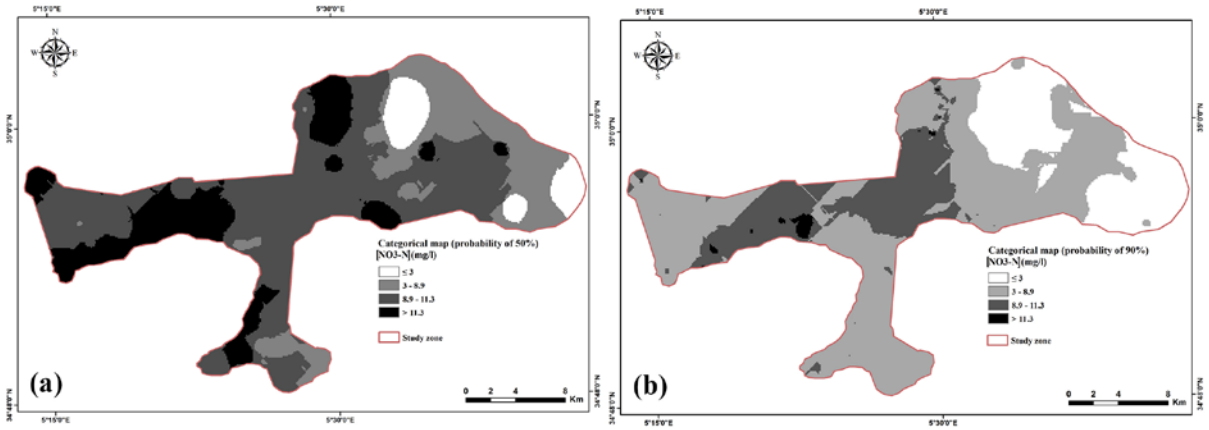


Figure 53: Categorical [NO3-N] maps: (a) with a probability P ≥ 50% to exceed thresholds, (b) with a probability P ≥ 90% to exceed thresholds

Table 21: Surface areas of different nitrate-N categories estimated using IK

Categories	IK (Probability over 50%)		IK (Probability over 90%)	
	Area (ha)	Area (%)	Area (ha)	Area (%)
≤ 3	2896.0	7.15	9274.6	22.90
3 < NO_3-N ≥ 8.9	8853.6	21.86	22371.0	55.24
8.9 < NO_3-N ≥ 11.3	20529.8	50.69	8523.6	21.05
NO_3-N > 11.3	8219.7	20.30	329.9	0.81
Total	40499.1	100.00	40499.1	100.00

The categorical map obtained at a level of fifty-fifty risk for surpassing thresholds (Figure 53a) using IK shows the same general structure of the spatial distribution of the four groundwater quality categories generated using (EBK. OK and RBF). The categorical map obtained at a level of 90% for surpassing thresholds is very important for decision makers, where it shows areas that need immediate intervention to reduce nitrate-N pollution.

3.4. Comparison of methods

In order to compare the results of the different interpolations techniques, the interpolation surfaces of nitrate-N concentration in the study zone are converted to raster maps (the size of each pixel is 108×108 m). From the continuous raster maps of EBK, OK, and RBF, ME, RMSE ASE, the determination coefficient (R^2) and the area (%) occupied by each category are calculated. For IK, the obtained raster maps are discrete maps, thus the comparison is mainly based on the calculated surface area occupied by each class.

Table 22 presents the errors measure of EBK, OK and RBF obtained using the pixel values of raster maps and Figure 54a depicts the scatterplot of the measured versus the predicted [NO_3-N] with determination coefficients (R^2) of the same methods. Low ME and RMSE and high R^2 values for the three methods indicate their accuracy and applicability to predict the contamination by NO_3-N. From Table 22, the ASE values are greater than the RMSE values for OK (The difference between ASE and RMSE, D = 0.72) and EBK (D = 0.268), which implies that the variability is overestimated, while it is underestimated by RBF (the ASE values are less than RMSE values, D = -0.3193). The R^2 values are 0.984, 0.951 and 0.998 for EBK, OK and RBF, respectively. The RMSE and the R^2 yields for the three interpolation techniques

give advantage to RBF, and therefore, RBF is considered as the best method to predict the spatial distribution NO_3-N contamination. This result is in agreement with the outcomes of Arslan, (2014).

Table 22: Errors measure of predictions (raster values)

Method	Errors			
	ME	RMSE	ASE	D = RMSE - ASE
OK	-0.0081	1.1814	0.4614	0.7200
EBK	0.0099	0.7546	0.4866	0.2680
RBF	-0.0166	0.2197	0.5390	-0.3193

From Figure 54b, a good correspondence is observed between the percentages of surface areas obtained by the different interpolation techniques at each category. The category of intermediate groundwater quality occupies the highest surface area with 39.49%, 39.22%, and 36.20% for EBK, OK and RBF, respectively. The surface area of this category is estimated by 50.69% using IK at a level P ≥ 50%. The lowest surface area is occupied by the excellent quality 1.2, 1.75, and 3.60% for EBK, OK and RBF. IK at a level P ≥ 50%, where estimates the surface area of this category appear by 7.15%. The surface areas occupied by the good and poor qualities are almost the same for EBK, OK and RBF, while IK underestimates these categories.

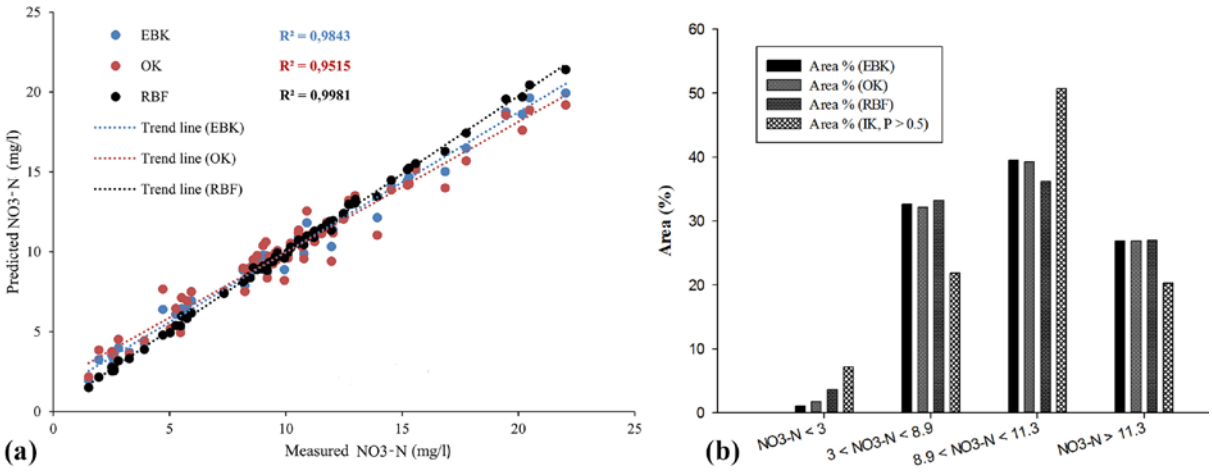


Figure 54: (a) Scatterplot of observed versus the predicted NO_3-N concentrations obtained by EBK. OK and RBF. (b) Bar chart showing area of different categories (%) obtained by EBK. OK. RBF and IK (P ≥ 50%).

Figure 55 shows the spatial variability of wet season nitrate-N concentrations. This map is obtained using the best interpolation technique among those tested, namely the RBF method. The maps of NO_3-N concentrations in the dry and wet season, generated using RBF method,

indicate almost the same spatial distribution of the different classes. The difference is in the shape and the area occupied by each class, where a decrease of the area percentages are observed as of medium and poor quality classes against an increase of the area percentages excellent and good quality classes. The area of excellent groundwater class increased from 3.60% to 5.21%, the good quality class increased from 33.25% to 41%, the medium quality class dropped form 36.20% to 30.07% and the poor quality class dropped from 26.94% to 23.72%. These results are unexpected at the end of the first campaign, with expectation in the increase of nitrate-N contamination due agricultural activities. This can be due to the near absence of the rainfall in the study area and therefore the contaminant will not reach the studied aquifer.

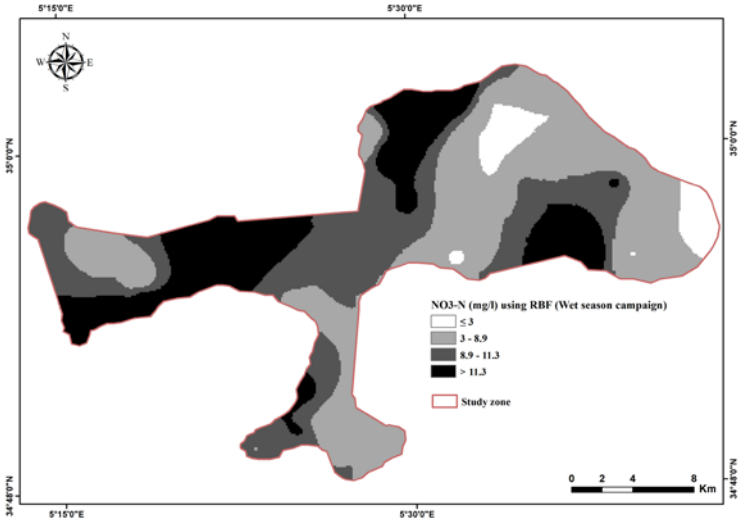


Figure 55: Prediction map of [NO3-N] classes of the wet season using the best method (RBF)

4. Discussions

The results of this study indicate that RBF is the best interpolation technique to predict the spatial variability of [NO_3-N] in groundwater and to investigate the surface areas with high risk to human health in El-Outaya plain. These results and the finally adopted map is generated by using RBF, which have important involvement for the aquifer monitoring and public health as long as these water resources are used for human consumption.

The final adopted map shows that the category of poor groundwater quality with 26.94% of the surface area is located in the middle of the study area, which is associated with intensive agriculture, excessive use of fertilizers and animal manure, especially chicken manure that has a higher amount of nitrogen. It is located also in the western part of the study area, where the agriculture is extensive, but it is noticed in this chapter that the presence of septic tanks near

the groundwater wells play major role. The category of excellent and good quality with 3.60% and 33.25% of the surface area, respectively, are located mainly in the eastern part of the study area, where the agriculture is extensive and the presence of Wadi Biskra, which participates in recharging the aquifer of Mio-Pliocene, and therefore, the dilution of nitrate-N concentration in the groundwater. The dominant category (intermediate quality) with 36% of surface, is located along the study area. RBF is an accurate interpolation technique to identify water pollution in El-Outaya plain.

The presented transformation of probability maps to categorical maps based on IK are important for decision makers to identify regions with high nitrate-N contamination probability, and high risk to human health, and could be applied to other risk assessment studies. In this way, decision makers and local agricultural manager may establish good agricultural practical protocols for farmers to prevent and reduce groundwater pollution by nitrate-N. Establishment of action programs are advised to raise awareness among farmers about proper application of pesticides and fertilizers, and designation of pollution vulnerable zones.

5. Conclusion

The application of robust interpolation technique is crucial for the prediction of spatial variability of contaminants in groundwater. This chapter presents a comparison of the linear and nonlinear geostatistical techniques (EBK, OK and IK) results and deterministic method (RBF). The cross-validation measures are used to choose the best model for prediction. The measurement errors of the pixel values of raster maps are used to check the accuracy of predictions and comparison of the interpolation methods.

All the interpolation methods have a high accuracy to predict groundwater nitrate-N contamination in the study area. The root mean square error (RMSE) and the determination coefficient (R^2) gave advantage to RBF. IK has been used to map the probability of excessive groundwater nitrate-N in the study area. The transformation of these probability maps to categorical maps shows the tendency of the groundwater quality at two levels to exceed a given threshold. IK at a medium risk exceeds thresholds ($P \geq 50\%$) indicating that the groundwater in the study area tends to have intermediate quality with an estimated area of 50.69% and at a high risk to exceed a given threshold ($P \geq 90\%$), and the groundwater tends to be of good quality over an estimated area of 55.24%. After all what have been presented in this chapter modern agricultural practices should be implemented in El-Outaya plain and nitrate-N concentrations of groundwater should be monitored continuously.

Chapter V

Modeling and mapping of soil salinity

1 Introduction

Soil salinization is defined as the process of salt enrichment in the soil, which can be caused by two different factors, i.e. natural (primary salinization) such as salty groundwater table near the soil surface (upwelling) and/or human-induced (secondary salinization) such as the continuous irrigation with saline water that leads to the increase of salts' accumulation in the irrigated lands (Pouladi et al. 2020; Koulla et al. 2019). Soil salinity is one the major environmental hazards that causes land degradation and affect agricultural productivity and the sustainability of land resources all over the world. The secondary salinization affects almost 45 million ha (20 %) of the total cultivated and 74 million ha (33 %) of irrigated agricultural areas in the world (Shrivastava and Kumar 2015). Furthermore, the salinized areas are still increasing at a rate of 10 % each year due to different problems including high surface evaporation, low precipitation, chemical or physical weathering of native rocks, irrigation with high saline water and poor cultural practices (traditional irrigation techniques practiced with irrelevant drainage systems) (Gorji et al 2020). Some researchers estimated that 50 % of the world's arable land will be affected by salinity (Machado and Serralheiro 2017).

In the Mediterranean region, the salt-affected soil is estimated as 1 million ha and is the main cause of desertification (Machado and Serralheiro 2017). Algeria is one of the Mediterranean countries where agricultural land is largely affected by this phenomenon, especially in the southeastern part of the country such as Biskra province, which is characterized by an arid climate and groundwater is the main source of irrigation. The farmers of this region are using groundwater of poor quality, usually unsuitable for irrigation and characterized by high salinity (Boudibi et al. 2019, Abdennour et al. 2020). Such practice has enhanced the development of agricultural sector, but it is also a principle factor of soil salinization. Therefore, it is urgent and necessary to ascertain the best modeling technique to improve the prediction of soil salinity by using of auxiliary variables, which can enhance the ability of decision makers and land use planners to assess accurately the soil salinity and reducing the cost and time consuming of sampling campaigns.

Consequently, many tools, modeling techniques and spatial interpolation methods have been developed and applied by researchers for the spatiotemporal monitoring and assessment of soil properties especially soil salinity in terms of electrical conductivity (ECs). Numerous geostatistical techniques have been applied for the spatiotemporal predictions of soil salinity and soil properties, including ordinary Kriging (Dai et al. 2014; Shahabi et al. 2016; Koulla et al. 2019), Indicator Kriging (Bradaï et al. 2016) and cokriging (Shen et al. 2019). Recently,

many researchers also applied new machine learning methods to predict soil characteristics. Were et al. (2015) used artificial neural networks (ANN), support vector machine (SVM) and random forest (RF) for predicting and mapping soil organic carbon stocks in Kenya. Achieng (2019) applied artificial and deep neural network for modeling of soil moisture and Pouladi et al. (2019) used multilayer perceptron neural network (MLP-NN) for predicting soil salinity in Iran. All the mentioned studies have shown the high accuracy of machine learning techniques in modeling soil properties.

Therefore, in this chapter, it is aimed to improve the prediction of soil salinity by the integration of new field and environmental covariates and explore the most suitable auxiliary variables for soil salinity prediction using multiple linear stepwise regression (MLSR). Finally, the evaluation and comparing the performance of simple Kriging (SK), simple cokriging (SCOK), multilayer perceptron neural network (MLP-NN) and support vector machine (SVM) modeling techniques in predicting the variability of ECs in El Outaya plain, southeastern Algeria using mean absolute error (MAE), root mean square error (RMSE), correlation coefficient (R), determination coefficient (R^2) and Taylor diagram.

2 Material and Methods

2.1 Soil sampling and ECs analysis

A total of 272 soil samples are collected from 68 irrigated farms in El Outaya plain over two campaigns, the dry season and wet season (The same period of groundwater sampling). At each location from each of the 68 farms, two soil depths (topsoil of 0-20 cm and subsoil of 40-60 cm) are sampled using a hand auger and placed into polyethylene bags. In the laboratory, the samples are air dried, crushed and sieved through a 2 mm sieve. The soil particle fractions (clay, sand and silt) are determined using the pipette method. The soil electrical conductivity (ECs) is measured in a 1:5 soil water diluted extract method using the multi-parameter (WTW multi 3430). The measured ECs was used to express the soil salinity. The flow diagram of modeling and mapping of soil salinity is presented in Figure 56.

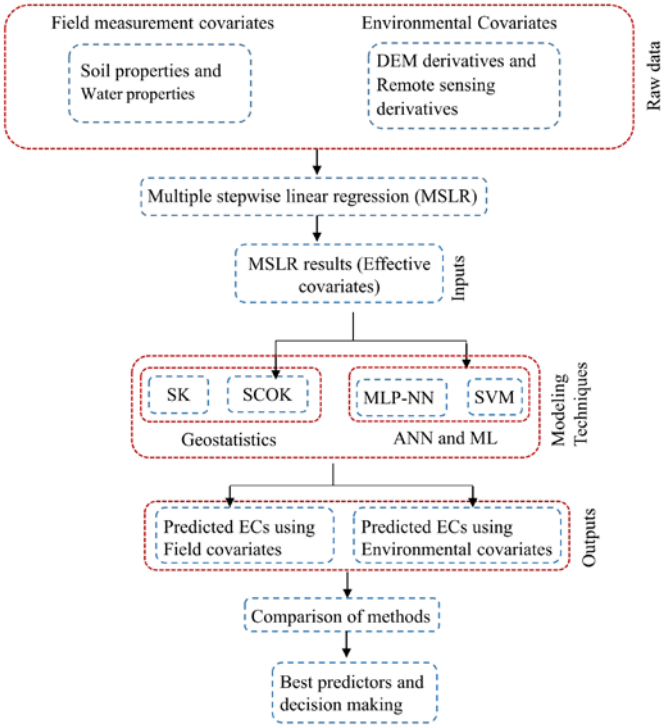


Figure 56: Flowchart of the used methodology

2.2 Covariates collection

In this chapter, the covariates used for predicting and modeling soil salinity are all continuous variables and regrouped into two categories, the field measurement covariates and the environmental covariates. The field measurements covariates including irrigation water properties and irrigated soil properties, and the environmental covariates including DEM derivatives and remote sensing derivatives are considered for modeling soil salinity in El Outaya plain. The used methodology is summarized in Figure 56. Therefore, 80 % of the samples, distributed through 54 sampling locations, is used for training and 20 %, distributed over 14 sampling locations, for testing the different modeling techniques (Figure 57).

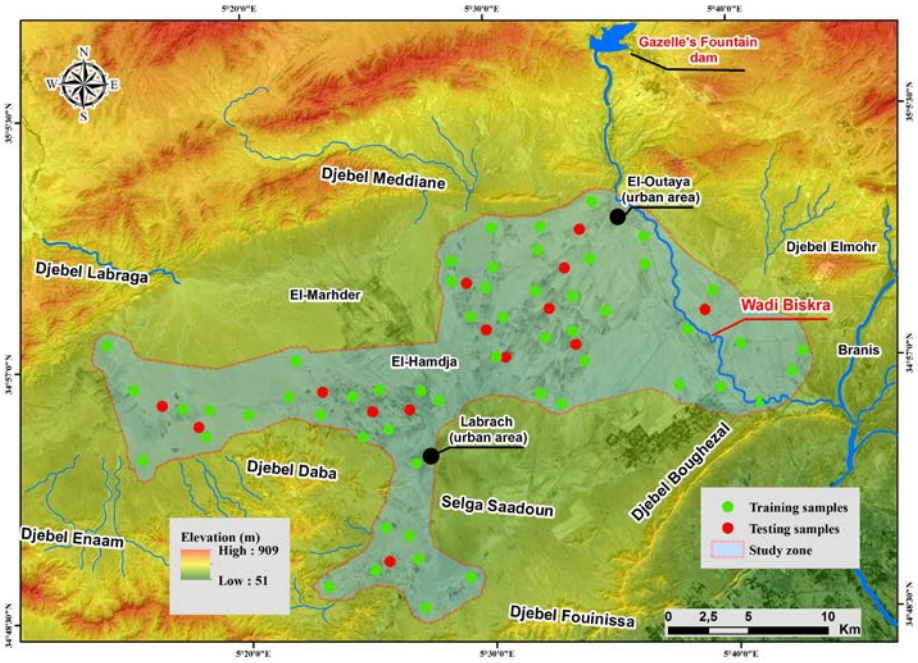


Figure 57: location of the training and testing samples in the study area

2.2.1 Field measurement covariates

The irrigation water properties are obtained from 136 groundwater samples over two seasons (See chapter 3). The used water properties are the electrical conductivity of groundwater (ECw), the potential Hydrogen (pH), cations plus anions of groundwater, the sodium percentage (Na %), the sodium adsorption ratio (SAR). The soil properties used as covariates are the soil particle fractions (clay %, sand % and silt %) of the topsoil and subsoil.

2.2.2 Environmental covariates

The digital elevation model (DEM) derivatives for modeling soil salinity included elevation, aspect, curvature, slope and topographic wetness index (TWI) (Figure 58). These DEM derivatives are calculated using ArcGis 10.2. The 30 m resolution Shuttle Radar Topography Mission (SRTM) DEM data covering the entire study area is obtained from the United States Geological Survey (USGS) website.

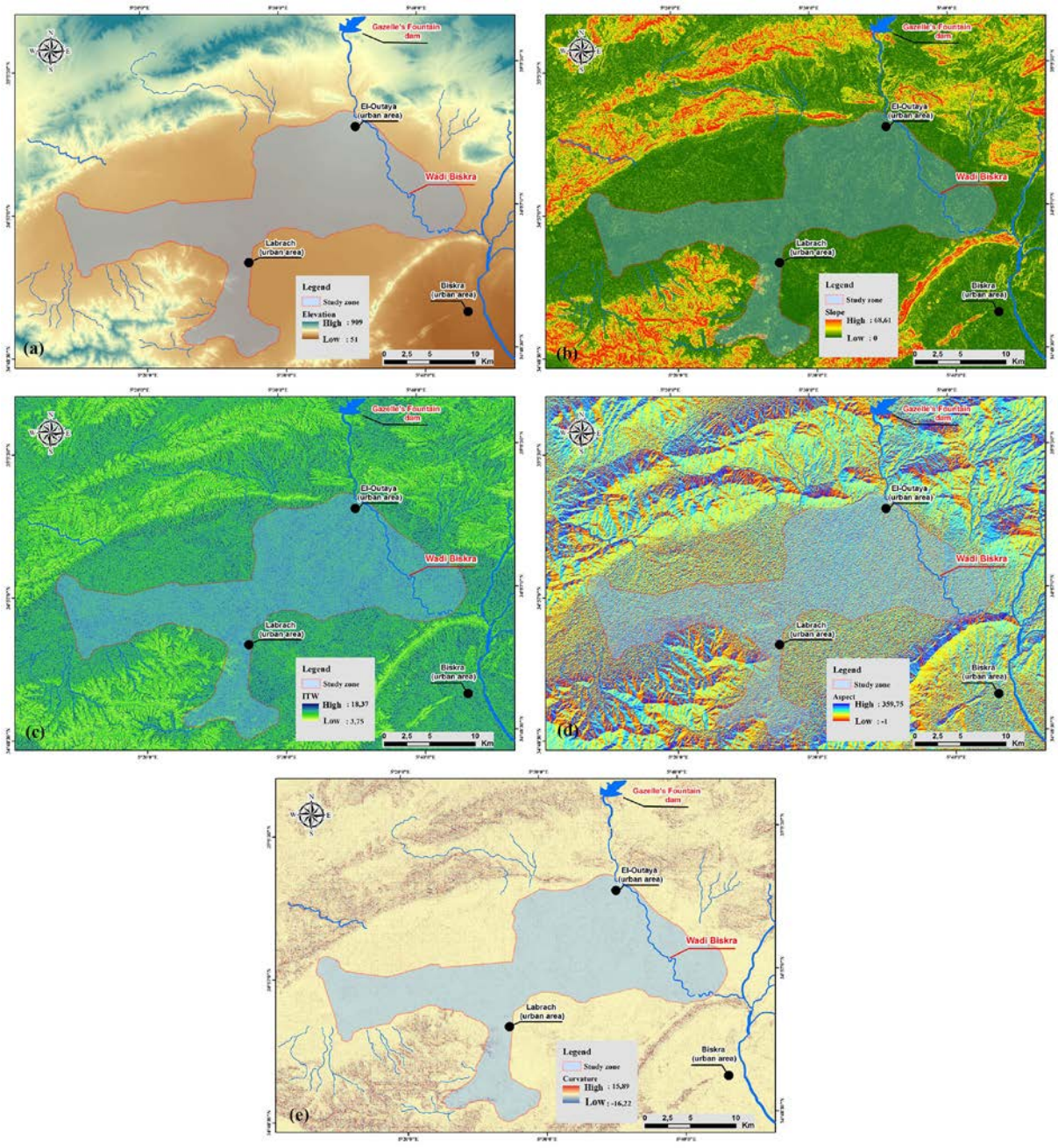


Figure 58: The raw covariates derived from DEM: (a) Elevation (b) Slope (c) TWI (d) Aspect (e) Curvature

The remote sensing derivatives are generated using Landsat-8 OLI satellite images. Two Landsat-8 satellite images with 30 m of resolution dated the same period of sampling, one in the dry season (September 2017) and another one in the wet season (Mai - June 2018), are obtained by downloading from the USGS website. The Landsat-8 OLI images are atmospherically corrected and radiometrically calibrated using ENVI 5.3 software through the FLAASH atmospheric model. After mosaicing and clipping the satellite images, three derivatives are evaluated including the Land Surface Temperature (LST), the Normalized

Difference Vegetation Index (INDVI) and Soil Adjusted Vegetation Index (SAVI) using map algebra tool of ArcGis software (Figure 59).

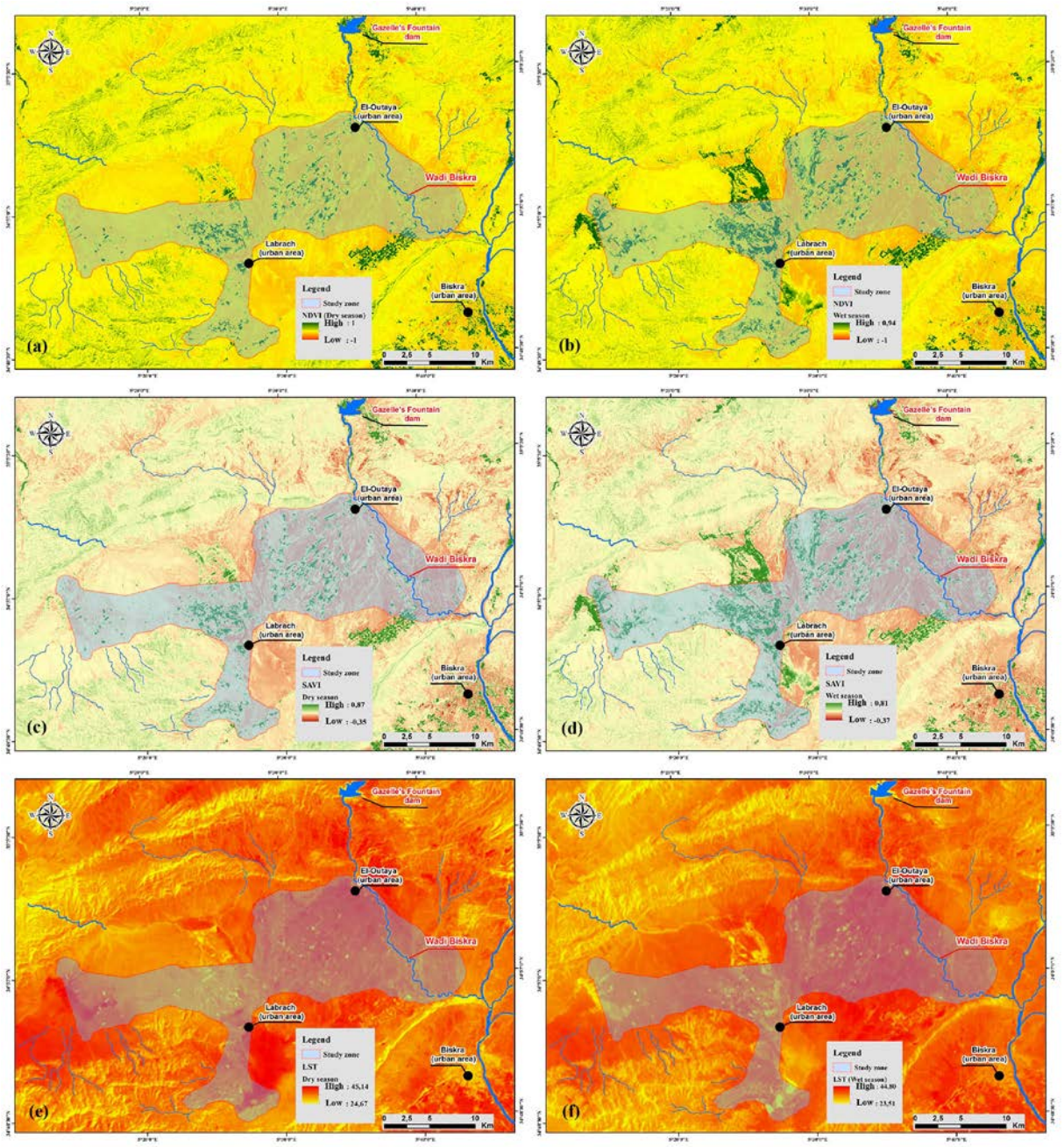


Figure 59: The raw covariates derived from satellite images: (a) and (b) NDVI, (c) and (d) SAVI, (e) and (f) LST of the dry season and wet season, respectively.

The NDVI is a standardized index allows the generation of relative biomass map. The default equation of NDVI is as follows (Shen et al. 2019):

$$NDVI = (NIR - R)/(NIR + R) \quad (40)$$

where NIR and R are the near infrared and Red bands of the Landsat-8 images, respectively.

The SAVI is a vegetation index used in regions with low vegetative cover (arid regions), where it attempts to minimize the influence of soil brightness (ESRI 2020). It is calculated as follows (Huete 1988):

$$SAVI = \left(\frac{NIR - R}{NIR - R + L}\right) * (1 + L) \quad (41)$$

Herein, L is the soil brightness correction factor, which depends on the amount of green vegetation cover (0.5 in this study).

The LST is one of the most important aspects of the land surface calculated from remote sensing. It is defined as the skin temperature of the ground (Avdan and Jovanovska 2016). The LST is computed as follows (Avdan and Jovanovska 2016; Yin et al. 2020):

$$LST\ (°C) = \frac{BT}{(1 + (\left(\frac{\lambda BT}{\rho}\right) * \ln(\varepsilon)))} \quad (42)$$

where BT is at sensor brightness temperature calculated using equation 4 (°C), λ is the wavelength of emitted radiance (λ (band 10) = 10.8 µm), ε is the emissivity calculated using Equation (44) and ρ = 1.4388 $*10^{-2}$ m k = h*c/δ (h = Plank's constant = 6.626 * 10^{-34}, c = Velocity of light = 2.998*10^{8} m/s and δ = Boltzmann constant = 1.38*10^{-23}).

$$BT = \frac{K_2}{\ln\left[\left(\frac{K_1}{L_\lambda}\right) + 1\right]} - 273.15 \quad (43)$$

where L_λ represents the top of atmosphere spectral radiance (Equation 45), K_1 and K_2 are the band specific thermal conversion constants from the metadata file.

$$\varepsilon = 0.004 * \left(\left(\frac{\text{NDVI} - NDVI_{min}}{NDVI_{max} - NDVI_{min}}\right)^2\right) + 0.986 \quad (44)$$

$$L_\lambda = M_L * Q_{cal} + A_L \quad (45)$$

where M_L is the band-specific multiplicative rescaling factor, Q_{cal} represents the band 10 of the satellite image and A_L is the band-specific additive rescaling factor.

2.3 Covariates preprocessing

Regression techniques such as multiple linear stepwise regressions (MLSR) are often used to predict a dependent variable (response) by using of one or more independent variables (predictors). In case where there is high number of covariates, it is essential to select the most

important covariates for modeling (Zounemat-kermani and Scholz 2014). MLSR is a combination of forward selection and backward elimination of input variables to generate the best combinations to the model (Zounemat-kermani et al. 2020). The probability of F statistic (pF) for inclusion is set at 0.05 (if pF is less than 0.05, the auxiliary variable is included in the linear equation) and the pF statistic for exclusion is set at 0.1 (if pF is greater than 0.1, the auxiliary variable is excluded (Gu et al. 2017). Some covariates are found correlated between themselves, which are referred to as multicollinearity and the covariates should be independent variables (Jalal et al. 2020). MLSR is an available method for eliminating multicollinearity problem between the covariates (Shen et al. 2019). In this study, MLSR is used for determining the optimal input combinations for SCOK, MLP-NN and SVM modeling techniques.

2.4 Modeling techniques

2.4.1 Simple kriging (SK)

Simple kriging (SK) estimator is considered as the weighted linear combination in which, the areal mean of the regionalized variables is known and must be second-order stationary (mean and variance constancy) (Maroufpoor et al. 2017; Şen 2016). Simple Kriging equation can be written as:

$$Z_{SK}^{*}(x) = \sum_{i=0}^{n} w_i(x)\, Z(x_i) + \left[1 - \sum_{x=1}^{n} w_i(x)\right] \mu \qquad (46)$$

Herein, $Z^{*}_{SK}(x)$ is the linear regression estimator, $Z(x)$ is the regionalized variable at the location x, μ is the location-dependent expected value of $Z(x)$, w_i are the weights and μ is the mean of the SK process.

2.4.2 Simple Cokriging (SCOK)

Cokriging (COK) estimator is a developed multivariate extension of Kriging, where the aim is to predict one variable (target variable) from data plus those of one or more variables (axillary variables) taking into account the additional correlated information in these variables (Webster and Oliver 2007). Simple cokriging (SCOK) is a multivariate extension of simple Kriging. The SCOK estimator is made up of the mean of the target variable of interest plus a linear combination of weights λ_i with the residuals of the auxiliary variables (Wackernagel 2003). SCOK is expressed as:

$$Z_{i0}^{*}(x_0) = m_{i0} + \sum_{i=1}^{N}\sum_{\alpha}^{ni} w_{\alpha}^{i}\,(Z_i(x_\alpha) - m_i) \quad (47)$$

where Z^{*}_{i0} is the estimated target variable i_0 at an unsampled location x_0, N is the number of auxiliary variables, ni is the number of samples of i^{th} auxiliary variable, m_{i0} is the global mean of a target variable, and m_i is the global mean of i^{th} auxiliary variable.

2.4.3 Multilayer perceptron neural network (MLP-NN)

The multilayer perceptron neural network (MLP-NN) is one type of a general class of Artificial Neural Networks (ANNs) called feed-forward neural network, which are the most common and widely applied to solve environmental problems (Pouladi et al. 2019; Sakaa et al. 2020). In this study, MLP-NN is used for the prediction of soil salinity (ECs). The MLP-NN is structured with one input layer, one or more hidden layers (one layer in this study) and an output layer (Figure 60). The neurons of each layer are connected to the neurons of the subsequent layer through synaptic weights but not in the same layer (Erdik et al. 2009). The number of neurons in the hidden layer is determined after training and testing of a series of networks having from 1 to 20 neurons in order to obtain the best structure with minimum error (e.g. square of errors, SOS) and maximum correlation coefficient.

In MLP-NN, the input variables (x_n) are fed into the input layer neurons (x_i) which sends them to the hidden layer (y_j). The neurons of the hidden layer sum up the received value of the input layer after the multiplication of each input value by its weight (w_{ij}) and adding a bias (b_j). The weighted sum (S_j) becomes the input to activation function in the hidden layer, the latter generates the output of each neuron in the hidden layer, which in turn, becomes the input of activation function in the output layer, where the output variable (Z_1) is further generated (Erdik et al. 2009; Kisi et al. 2017; Sakaa et al 2020).

$$S_j = b_j + \sum_{i=1}^{n} X_i * w_{ij} \quad (48)$$

S_j is the weighted sum, b_j is the bias, x_i stands for the input standardized value, n is the number of input variables in the input vector and w_{ij} are the synaptic weights.

The overall equation that describes the input/output relationship of three-layered MLP-NN is given as follows (Bishop 1995):

$$Z_k = f_k\left(b_k + \sum_{j=1}^{p} w_{jk} * f_j\left(\sum_{i=1}^{n} X_i * w_{ij} + b_j\right)\right) \quad (49)$$

where Z_k is one of the output variables, b_k and bj are the bias (they can be absorbed into the weights), w_{ij} and w_{jk} denote successively the weights between the input layer and the output layer neurons and between the hidden layer and the output layer neurons, p is the number of neurons in hidden layer, n is the number of neurons in the input layer, f_k and f_j are the activation functions of the neurons of the hidden layer and the neurons of output layer, successively.

In this study, MLP-NNs are trained with three different back-propagation algorithms, which are the Broyden–Fletcher– Goldfarb–Shanno (BFGS Quasi-Newton), conjugate gradient (CG) and gradient descent (GD). The BFGS Quasi-Newton yielded the best results for all the trained models. Several activation functions are used in MLP-NNs, such as logistic (sigmoid), hyperbolic tangent (Tangh), linear and exponential.

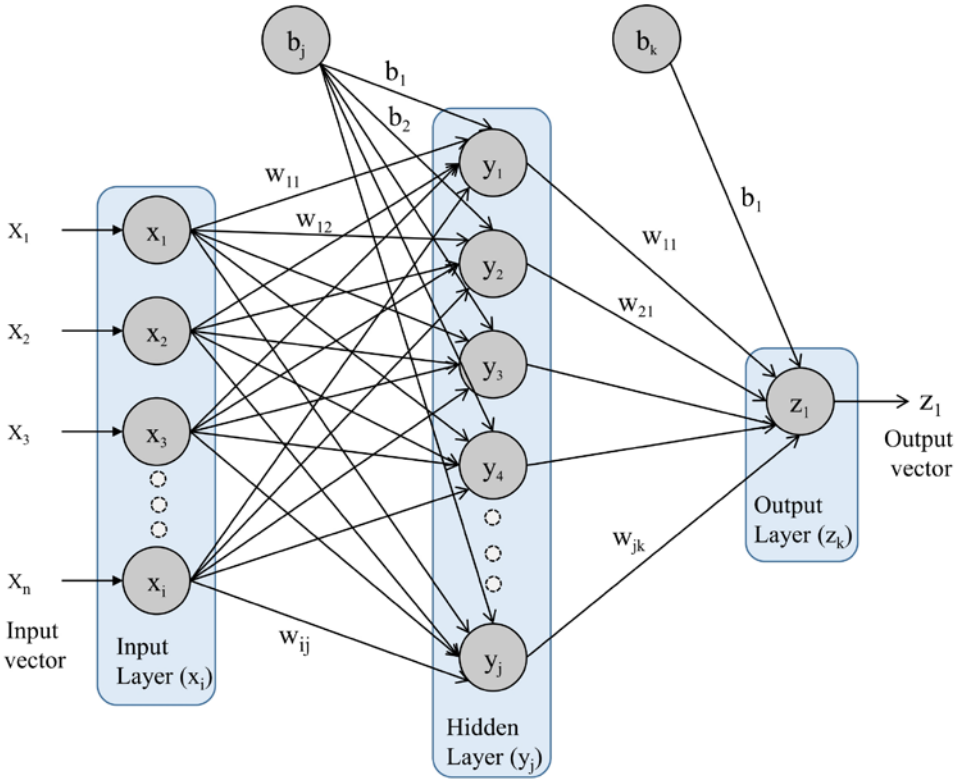


Figure 60: Schematic representation of three-layered feed-forward MLP-NN

2.4.4 Support vector machine (SVM)

Support vector machine (SVM) is a supervised learning technique proposed by Cortes and Vapnik (1995) for classification problems by introducing the soft margin classifier, and then the algorithm expanded to regression cases by Vapnik in 1995 (Vapnik 1995). Since it maps the original space into a high dimensional feature space, it is considered as two layers learning machine (input layer and high dimensional feature space layer) (Wu et al. 2006). The detailed description of SVM can be found in many publications (Vapnik 1995; Vapnik 1998; Cherkassky and Ma 2004; Jalal et al. 2020).

The SVM for regression model is defined by $y = f(x) + \delta$ where δ is the independent random error, x is a multivariate input and y is the output function (Jalal et al. 2020).

For a given training dataset $[(x_1, y_1), \ldots, (x_N, y_N)]$, x_N and y_N are the input and output values and N is the size of the training dataset. The optimum SVM regression estimator (f) can be assumed as follows:

$$f(x) = w * \Phi(x) + b \tag{50}$$

where Φ is a non-linear mapping function, w is a weight vector and b is the bias term.

Regression estimates by minimization of the empirical risk function (R_{emp}) given bellow:

$$R_{emp} = \frac{1}{N}\sum_{i=1}^{N} L_{\varepsilon}(y_i - f(x)) \tag{51}$$

where $L_{\mathcal{E}}$ is the loss function used for minimization of R_{emp} called ε-sensitive loss proposed by Vapnik (1998):

$$L_{\varepsilon}(y - f(x)) = \begin{cases} 0 & for\ |y - f(x)| < \varepsilon \\ |y - f(x)| - \varepsilon & Otherwise \end{cases} \tag{52}$$

Using the ε-sensitive loss objective function and introducing the regularization parameter (C), and the slack variables (ξ and ξ^*), the SVM regression is formulated as minimization of the following functional (Cherkassky and Ma 2004; Herceg et al. 2019):

$$minimise\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{53}$$

Subjected to the constraints $\begin{cases} y_i - f(x) - b \le \varepsilon + \xi_i^* \\ f(x) + b - y_i \le \varepsilon + \xi_i \\ \xi_i^*, \xi_i \ge 0 \end{cases}$ (54)

The solution of this optimization problem is given by:

$$f(x) = \sum_{i=1}^{N} (a_i - a_i^*) K(x_i, x) + b \quad (55)$$

where a and a^* are Lagrange multipliers and $K(x_i, x)$ is the kernel function (Herceg et al. 2019).

The performance of the SVM regression model is highly influenced by the type of the kernel function and setting the parameters C (capacity factor), ε (the error-insensitive zone) (Samui 2008). The commonly used kernel functions are linear, sigmoid, radial basis function and polynomial. In the current study, the kernel function and the different parameters are chosen by a trial and error method, where the kernel with high correlation coefficient and least mean square error (MSE) results is used.

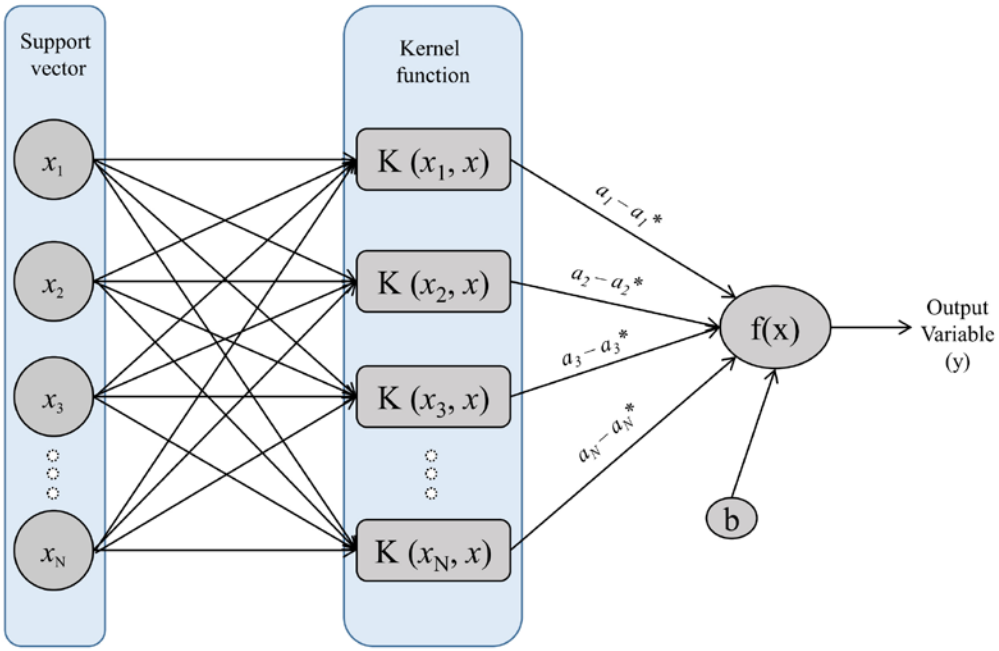


Figure 61: schematic representation of SVM model

2.5 Data normalization

Normalization of data is common in artificial intelligence modeling to prevent the models from domination by the large values (Ceryan 2014). Before running the modeling techniques, the inputs and outputs are normalized in the range 0 – 1 using the following equation:

$$X_n = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (56)$$

where, X_n is the normalized value, X_i is the original value, X_{min} and X_{max} are the minimum and maximum of the original data, respectively.

After modeling, the results are converted from the standard mode to the real values using the following equation:

$$X_i = X_{min} + (X_{max} - X_{min}) * X_n \quad (57)$$

2.6 Evaluation and comparing models performance

The accuracy of the different modeling techniques (SK, SCOK, MLP-NN and SVM) can be evaluated using different statistical measures of goodness of fit that describe the errors associated with the different models (Zounemat-kermani et al. 2014). In this study, three common statistical parameters are used to evaluate the performance of the prediction models. These are the correlation coefficient (R), which is defined as the degree of the relationship between the observed and the predicted values, the root mean square error (RMSE) and the Mean absolute error (MAE) are expressed as follows:

$$R = \frac{\sum_{i=1}^{n}(EC_i^O - \overline{EC_i^O})(EC_i^P - \overline{EC_i^P})}{\sqrt{\sum_{i=1}^{n}(EC_i^O - \overline{EC_i^O})}\sqrt{\sum_{i=1}^{n}(EC_i^P - \overline{EC_i^P})}} \quad (58)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(EC_i^O - EC_i^P)}{n}} \quad (59)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|EC_i^P - EC_i^O\right| \quad (60)$$

Where EC_i^O is the observed EC values, EC_i^P is the predicted EC

values, $\overline{EC}$ is the average of EC values and n is the total number of samples.

3 Results and Discussions

3.1 Input selection

The raw covariates are chosen using a prior knowledge about the entire modeling problem but the key covariates are selected using MLSR, which helps to exclude the covariates that do not have a significant effect on the modeling techniques. In this study, eight field measurement covariates (ECw, pH, Na%, SAR, cations + anions, clay%, sand% and silt%) and eight

environmental covariates (Elevation, aspect, curvature, slope, TWI, LST, NDVI and SAVI) are candidates for inclusion in MLSR and predictors of ECs in the dry season (ECs1) and ECs in the wet season (ECs2) of the top soil (TS) and the sub soil (SS).

Table 23 and Table 24 show the results of applying MLSR procedure for the field measurement covariates and the environmental covariates, respectively, including p values, variance inflation factors (VIF), tolerance, unstandardized coefficients (B), standardized coefficients (*B)* and the determination coefficient (R^2).

If a *P* value is less than the threshold ($\alpha = 0.05$), the null hypothesis (i.e., the covariates have no significant effect on the output variable, ECs) is rejected and the difference is statistically significant. On the other hand, if P value is greater than α, the null hypothesis is confirmed and there is no statistically significant difference. The tolerance and the variance inflation factor (VIF) indices of the multi-collinearity test are used to ensure the accuracy of MLSR results. The tolerance index indicates the percentage of variations in each covariate, which is not identified by other covariates. VIF = 1 indicates that the covariate of interest has no correlation with the other covariates; 1 < VIF < 5 shows moderate correlation and VIF > 5 indicates high correlation (Zounemat-kermani et al. 2020).

From the MLSR outputs of the dry season covariates, one can conclude that sand % and ECw_1 are the optimal field predictors of ECs1 (TS) and ECs1 (SS) with R^2 of 0.528 and 0.695, respectively. Additionally, the LST_1, TWI and Elevation are the optimal environmental predictors of ECs1 (TS) with R^2 of 0.482 and LST_1 with TWI the appropriate predictors of ECs1 (SS) with R^2 of 0.349. The regression equations of ECs for the dry season are as follows:

ECs1 (TS) = -0.052 * Sand + 0.398 ECw_1 + 2.932

ECs1 (SS) = -0.049 * Sand + 0.316 ECw_1 + 3.069

ECs1 (TS) = 0.552 * LST_1 + 0.712 * TWI + 0.054 * Elevation – 36.361

ECs1 (SS) = 0.696 * LST_1 + 0.381 * TWI – 28.125

From the MLSR outputs of the wet season covariates, one can see that the silt% with ECw_2 and pHw_2 is the optimal combination of field covariates to predict ECs2 (TS) with a coefficient R^2 = 0.599, while the optimal combination to predict ECs2 (SS) is sand percentage plus ECw_2 with R^2 = 0.706. For the environmental covariates, LST_2, TWI and elevation are the optimal predictors of the ECs of both depths, with R^2 = 0.620 and 0.530 for ECs2 (TS) and ECs2 (SS), respectively. The regression equations of ECs for the wet season are as follows:

$ECs2\ (TS) = 0.056 * Silt + 0.406 * ECw_2 + 2.445 * pHw_2 – 20.395$

$ECs2\ (SS) = -0.046 * Sand + 0.238\ ECw_2 + 3.115$

$ECs2\ (TS) = 0.699 * LST_2 + 0.345 * TWI + 0.035 * Elevation – 35.950$

$ECs2\ (SS) = 0.590* LST_2 + 0.329 * TWI + 0.035 * Elevation – 31.351$

These optimal predictors of ECs (field measurement and environmental covariates) issued from MLSR are used as inputs for SCOK, MLP-NN and SVM modeling techniques.

Table 23: Results of the MLSR for the field measurement covariates

Independent variable	Covariates	P value	Tolerance	VIF	B	β	R^2
ECs1 (TS)	ECw_1	0.002	0.955	1.048	0.398	0.283	0.528
	Sand	0.000	0.955	1.048	-0.052	-0.611	
ECs1 (SS)	ECw_1	0.000	0.921	1.085	0.316	-0.706	0.695
	Sand	0.000	0.921	1.085	-0.049	0.287	
ECs2 (TS)	ECw_2	0.000	0.683	1.464	0.406	0.394	
	pHw_2	0.049	0.912	1.490	2.445	0.194	0.599
	Silt	0.000	0.671	1.097	0.056	0.659	
ECs2 (SS)	ECw_2	0.001	0.905	0.105	0.238	0.239	0.706
	Sand	0	0.905	0.105	-0.046	-0.735	

Table 24: Results of the MLSR for the environmental covariates

Independent variable	covariates	P value	Tolerance	VIF	B	β	R^2
ECs1 (TS)	LST_1	0.004	0.636	1.572	0.552	0.339	
	TWI	0.001	0.988	1.012	0.712	0.302	0.482
	Elevation	0.002	0.634	1.577	0.054	0.366	
ECs1 (SS)	LST_1	0.000	0.999	1.001	0.696	0.547	0.349
	TWI	0.043	0.999	1.001	0.381	0.206	
ECs2 (TS)	LST_2	0.000	0.884	1.131	0.699	0.593	0.620
	TWI	0.012	0.996	1.004	0.345	0.199	
	Elevation	0.000	0.881	1.135	0.035	0.328	
ECs2 (SS)	LST_2	0.000	0.884	1.131	0.590	0.518	0.530
	TWI	0.026	0.996	1.004	0.329	0.196	
	Elevation	0.000	0.881	1.135	0.035	0.338	

3.2 Descriptive statistics of the input and output data

3.2.1 Output data

The descriptive statistics of soil salinity of 272 samples from 68 sites (54 for training and 14 for testing) over two periods are summarized in Table 25. The soil salinity of training dataset

varied from 0.12 to 12.37, 0.12 to 10.23, 0.15 to 7.93 and 0.14 to 7.42 mS/cm for ECs1 (TS), ECs1 (SS), ECs2 (TS) and ECs2 (SS), respectively. The ECs of testing dataset varied from 0.12 to 8.72, 0.13 to 7.10, 0.16 to 5.43 and 0.13 to 6.5 mS/cm for ECs1 (TS), ECs1 (SS), ECs2 (TS) and ECs2 (SS), respectively. The mean value of ECs varied between the dry and wet season and between the different depths, where the dry season had higher values than the wet season and the top soil had the highest mean values. The mean of ECs1 (TS), with a value of 2.06 mS/cm, is higher than ECs1 (SS) and ECs2 (TS) with 1.87 and 1.54 mS/cm, respectively. Furthermore, the mean value of ECs2 (TS) is higher than ECs2 (SS), with a mean value of 1.73 mS/cm. These results maybe due to the accumulation of salt in the top soil as a function of evaporation especially in the dry season. The coefficient of variation (CV) of ECs varied from 104.34 to 130.55 indicates a strong variation of salinity in the study area. This strong spatial variation of soil salinity maybe due to the high variation of soil texture (slight textures in the eastern and heavy in the western part of the study area), land use and irrigation schedules.

Table 25: descriptive statistics of ECs dataset used for training and testing

Output	Data set	Number	Mean	Min	Max	CV	SD	Variance
ECs1 (TS)	Training	54	2.06	0.12	12.37	131.55	2.71	7.36
	Testing	14	2.08	0.12	8.72	130.76	2.72	7.41
	Total	68	2.06	0.12	12.37	130.58	2.69	2.27
ECs1 (SS)	Training	54	1.77	0.12	10.23	115.81	2.05	4.2
	Testing	14	2.3	0.13	7.10	104.34	2.4	5.79
	Total	68	1.87	0.12	10.37	112.83	2.11	4.46
ECs2 (TS)	Training	54	1.54	0.15	7.93	132.46	2.04	4.18
	Testing	14	1.52	0.16	5.43	117.76	1.79	3.2
	Total	68	1.54	0.15	7.93	128.57	1.98	3.93
ECs2 (SS)	Training	54	1.6	0.14	7.42	111.25	1.78	3.16
	Testing	14	2.21	0.13	6.5	107.69	2.38	5.68
	Total	68	1.73	0.13	7.42	110.4	1.91	3.67

3.2.2 Input data

Table 26 summarizes the descriptive statistics of the field and the environmental covariates resulting from MLSR as the optimal predictors of ECs in the study area.

Table 26 descriptive statistics of the input covariates

Inputs	Mean	Min	Max	CV	SD	Variance
ECw_1 (mS/cm)	4.23	1.14	10.68	45.39	1.92	3.67
ECw_2 (mS/cm)	4.23	1.13	12.01	45.39	1.92	3.69
pHw_2	7.550	7.19	7.84	2.12	0.16	0.025
Sand % (TS)	49.34	1.55	92.5	64.45	31.80	1011.48
Sand % (SS)	51.61	1.35	89.48	58.88	30.39	923.88
Silt % (TS)	31.46	2.23	89.79	74.79	23.530	553.92
LST1 (°C)	38.18	34.6	41.35	4.35	1.660	2.75
LST2 (°C)	38.89	33.66	42.08	4.31	1.680	2.83
TWI	8.95	7.32	12.73	12.73	1.14	1.31
Elevation(m)	200.57	173	254	9.19	18.43	339.76

3.3 Geostatistical models

3.3.1 Simple kriging (SK)

In this study, SK is applied for mapping the spatial distribution of soil salinity based only on the field measurements of ECs. The Kolmogorov and Smirnov test (KS test) revealed that the data of ECs for 54 training sites of both seasons are not normally distributed; therefore, data are subjected to the normal score transformation. Fitting a theoretical semivariogram model to the empirical semivariogram is the main processing step in the Kriging techniques. Different models are tried (Spherical, Gaussian, exponential …etc.) and the best fit is found as determined based on the cross validation results (Lowest RMSE).

Figures 63a, b, c and d show the best semivariogram models selected for ECs1 (TS), ECs1 (SS), ECs2 (TS) and ECs2 (SS), respectively, and their different parameters. A spherical model with nugget effect value of 0.051, partial sill of 1.587 and a range of 32309 m is suitable for ECs2 (TS), while a Gaussian model with a nugget effect of 0.212, partial sill of 1.2 and a range of 26216 m is considered for ECs2 (SS). The best fit models for ECs1 (TS) and ECs1 (SS) are Gaussians with nugget effects of 0.238, 0.265, partial sills of 1.23, 1.571 and ranges of 29944 m and 29961 m, respectively. The nugget to sill ratio, which is used for the determination of the level of spatial dependence (Cambardella et al. 1994), indicates a strong spatial dependence of soil salinity in the four models.

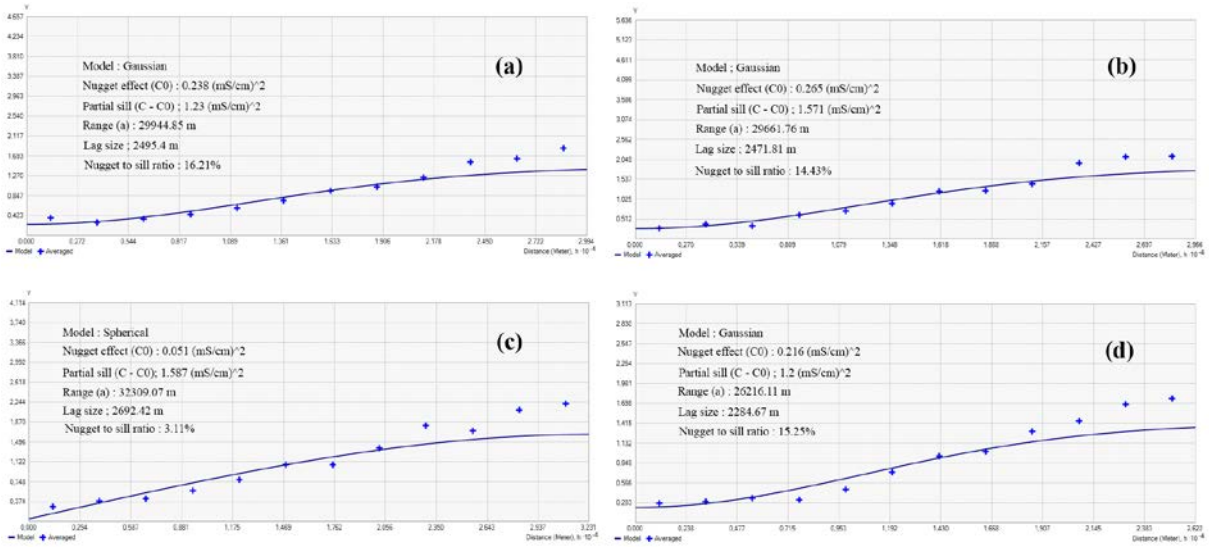


Figure 62: Semivariograms and fitted models of (a) ECs1 (TS), (b) ECs1 (SS), (c) ECs2 (TS) and (d) ECs2 (SS)

The results of the performance criterions for training and testing stages are provided in Table 27. SK technique showed results that are more accurate in the testing stage. The MAE, RMSE, R are calculated as 0.61, 0.85, 0.874 for ECs2 (TS), 0.81, 1.2, 0.927 for ECs2 (SS), 1.15, 1.72, 0.761 for ECS1 (TS) and 0.75, 1.17, 0.934 for ECs1 (SS), respectively.

Table 27: Prediction performance measures of SK for training and testing datasets

		Training			Testing		
Period	Variable	MAE	RMSE	R	MAE	RMSE	R
Dry season	ECs1 (TS)	1.2	2.08	0.631	1.15	1.72	0.761
	ECs1 (SS)	0.95	1.56	0.641	0.75	1.17	0.934
Wet season	ECs2 (TS)	0.87	1.43	0.710	0.61	0.85	0.874
	ECs2 (SS)	0.86	1.27	0.700	0.81	1.2	0.927

The resultant maps of soil salinity from SK (Figure 63) show the spatial distribution of different classes of soil salinity in the top and sub soil of El Outaya plain over the dry and wet season. The maps indicate approximately the same overall distribution of ECs classes in the study area. For the interpretation of the results of the salinity levels, the soil salinity classification of Soil Test Handbook for Georgia has adopted (Abuelgasim and Ammad 2018), which is compatible with our analysis method for the measuring ECs (Table 28).

From Figure 63 and Table 28, it is apparent the dominance of two main salinity classes, the slightly saline to low salinity class and excessively high salinity class. The slightly saline / Low salinity is located generally in the western part of the study area with area percentages of 41.14

%, 40.57 %, 39.76 % and 38.31 % for ECs2 (TS), ECs2 (SS), ECs1 (TS) and ECs1 (SS), respectively. The excessively high salinity class is located in the eastern part of the plain, where the agriculture is more intensive, with area percentages of 36.78 %, 37.70 %, 44.60 % and 42.13 % for ECs2 (TS), ECs2 (SS), ECs1 (TS) and ECs1 (SS), respectively.

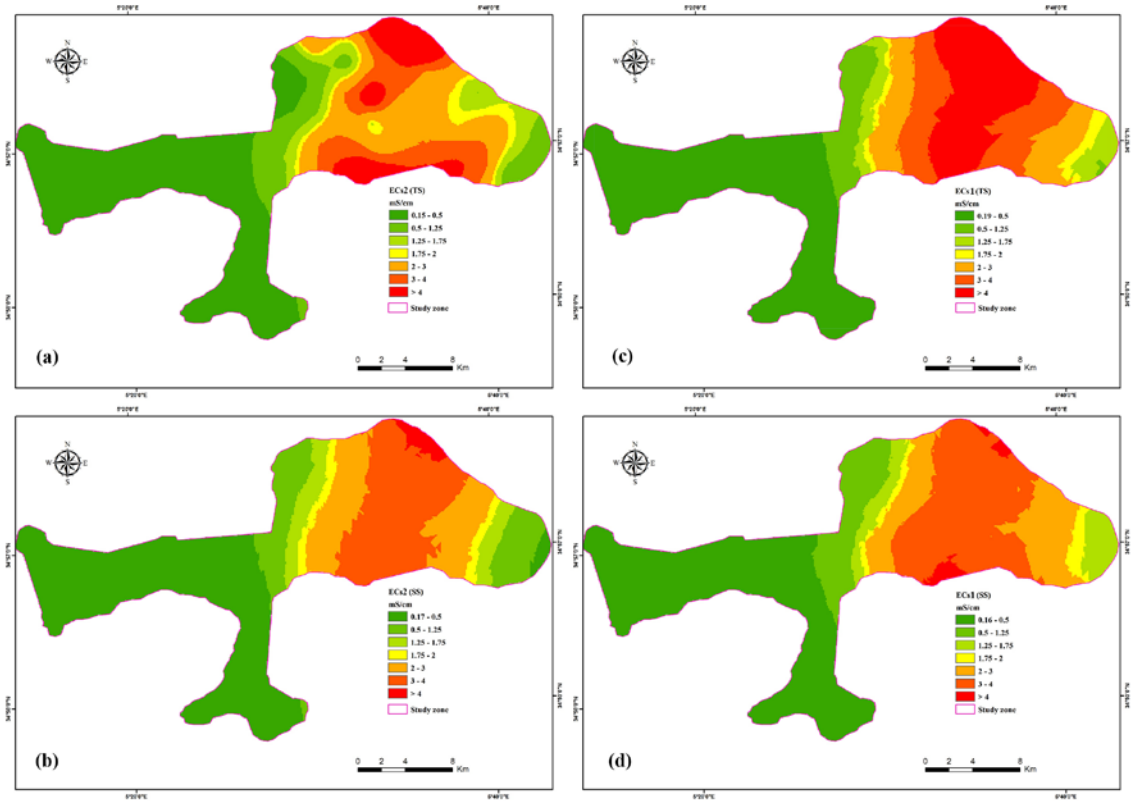


Figure 63: Prediction maps of soil salinity using SK: (a) ECs2 (TS), (b) ECs2 (SS), (c) ECs1 (TS), (d) ECs1 (SS)

Table 28 The area percentages of different soil salinity classes of SK maps

Classes of ECs (mS/cm)	Salinity level	Area (%)			
		ECs2 (TS)	ECs2 (SS)	ECs1 (TS)	ECs1 (SS)
0 < ECs ≤ 0.15	Non saline	0	0	0	0
0.15 < ECs ≤ 0.50	Slightly Saline/Low Salinity	41.14	40.57	39.76	38.31
0.5 < ECs ≤ 1.25	Moderately Saline/Medium	11.02	12.19	7.30	10.04
1.25 < ECs ≤ 1.75	Strongly Saline/High Salinity	7.00	6.25	5.15	6.47
1.75 < ECs ≤ 2	Very High Salinity	4.05	3.29	3.19	3.04
ECs > 2	Excessively High Salinity	36.78	37.70	44.60	42.13
	Total	100.00	100.00	100.00	100.00

The scatter plot of the measured soil EC and the estimated mean errors (Figure 64) using SK reveals an overestimation of the high ECs values and an underestimation of the low ECs value.

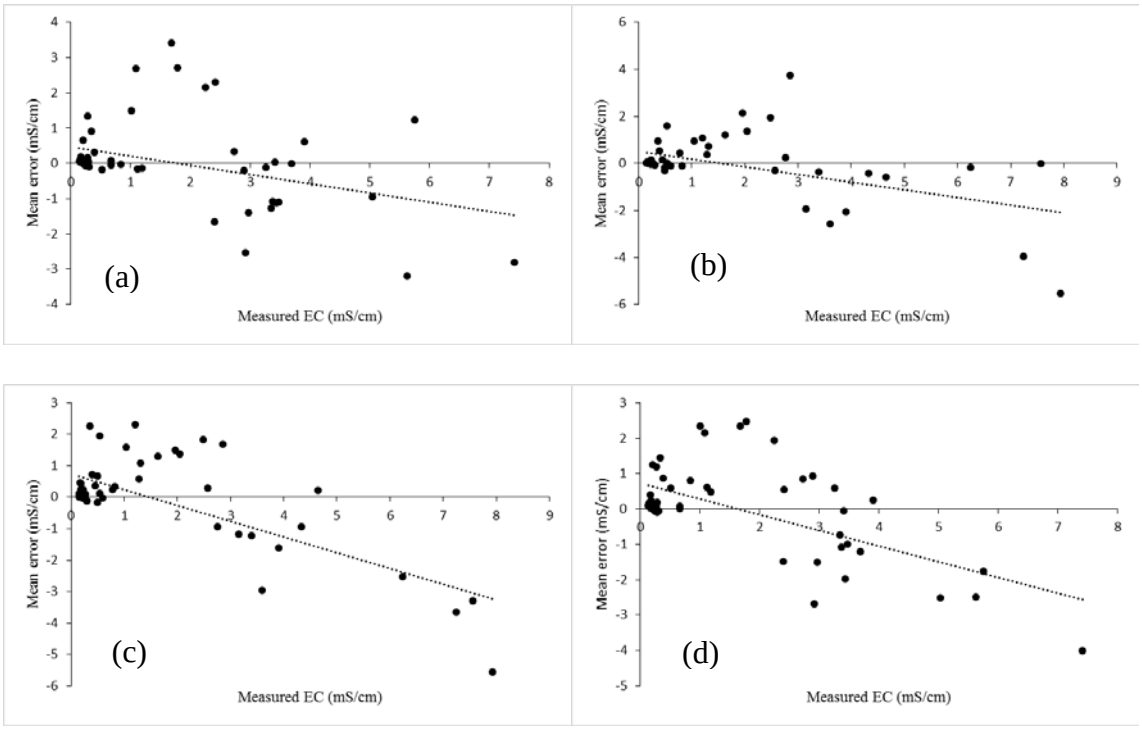


Figure 64: Scatter plots of measured soil EC and the estimated mean errors using SK: (a) ECs1 (TS), (b) ECs1 (SS), (c) ECs2 (TS), (d) ECs2 (SS)

3.3.2 Simple Cokriging (SCOK)

3.3.2.1 SCOK with field covariates

Figure 65 shows the experimental auto semivariograms of ECs1 (TS), ECs1 (SS), ECs2 (TS), ECs2 (SS) and their fitted models. Their different cross semivariograms with the field covariates are shown in Figure 66. All the experimental semivariograms and cross semivariograms are fitted with spherical models. The semivariogram model presents information about the auto-correlation of a spatial variable while the cross covariogram provides information of cross correlation of two related spatial variables (Wang et al. 2013).

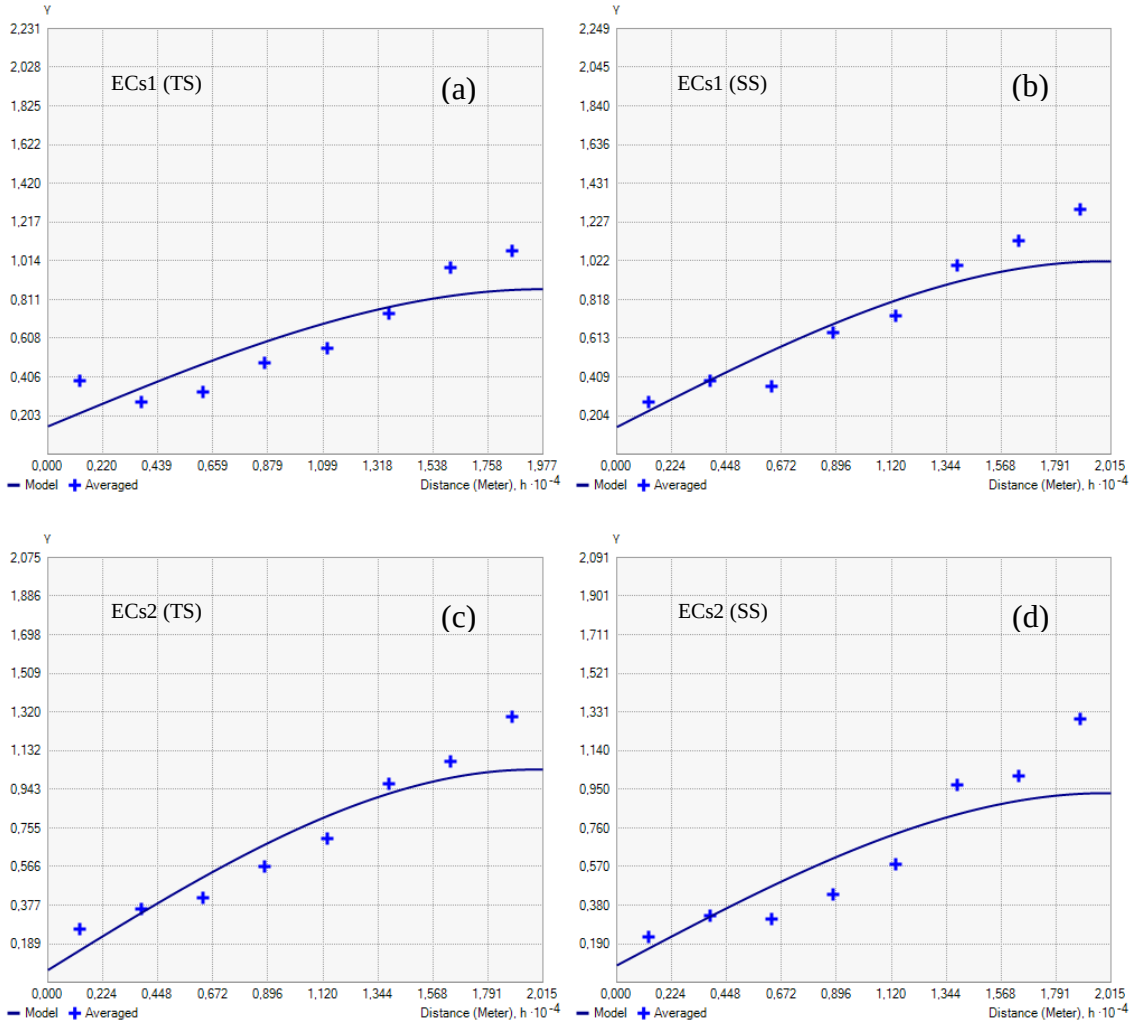
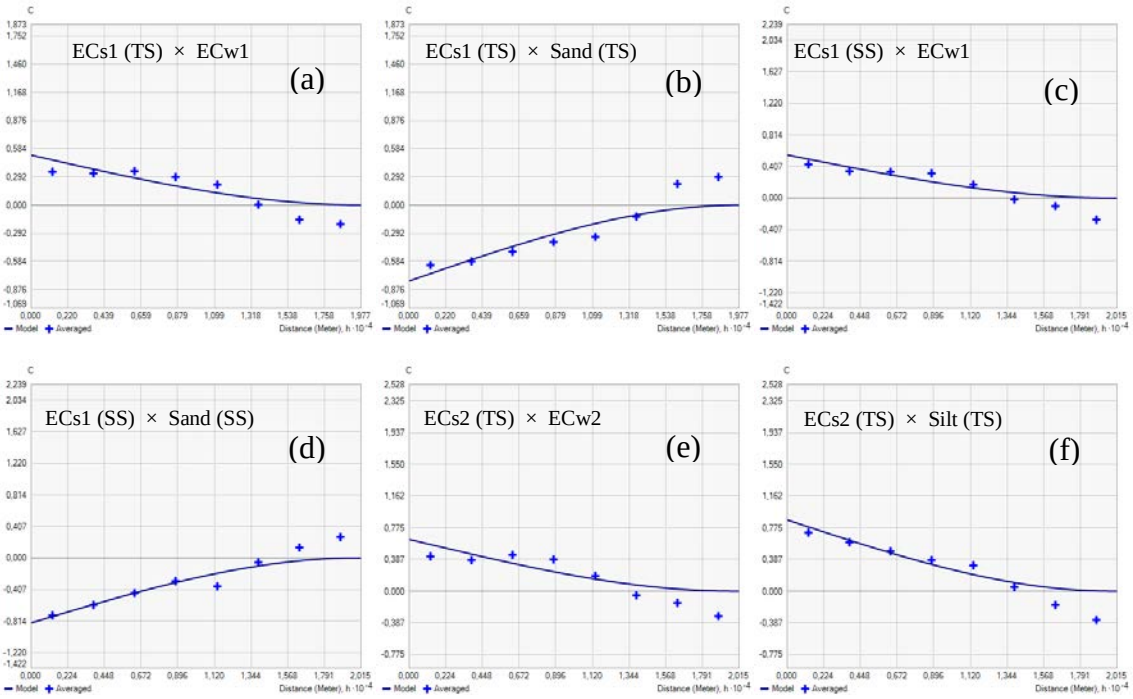


Figure 65: Auto semivariograms of ECs using field covariates: (a) ECs1 (TS), (b) ECs1 (SS), (c) ECs2 (TS), (d) ECs2 (SS). All using spherical models.



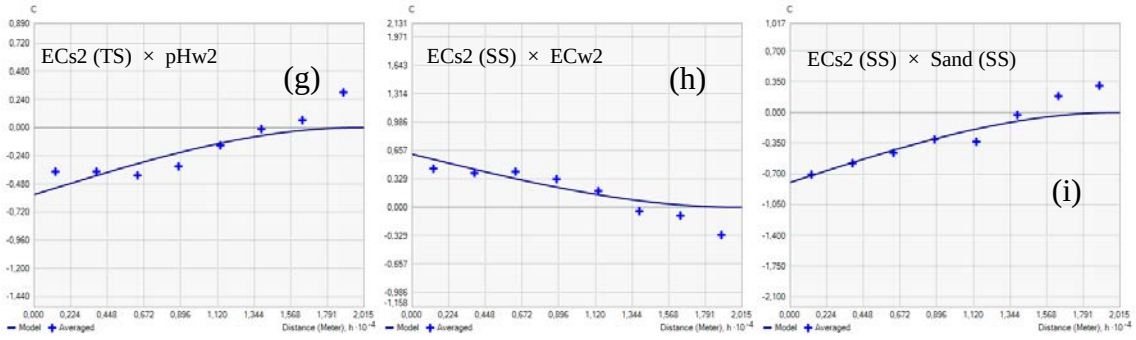


Figure 66: Cross covariograms between ECs and field covariates using SCOK: (a) ECs1 (TS) – ECw1, (b) ECs1 (TS) – Sand (TS), (c) ECs1 (SS) – ECw1, (d) ECs1 (SS) – Sand (SS), (e) ECs2 (TS) – ECw2, (f) ECs2 (TS) – Silt (TS), (g) ECs2 (TS) – pHw2, (h) ECs2 (SS) – ECw2, (i) ECs2 (SS) – Sand (SS). All using spherical models.

The parameters of the different semivariograms and cross covariograms are tabulated in Table 29. The models are fitted using different parameters (Nugget effect (C0), Sill (C) and range (a)). The nugget to sill ratios (C0/(C+C0)) for all the auto semivarigram models of ECs indicate strong spatial dependences. The intrinsic factors are the main cause of strong dependence, while weak dependence is caused by extrinsic factors (Wang et al. 2013).

Table 29: Parameters of the auto semivariogram models for ECs and their cross covariogram models (using field covariates).

Period	Variable	Model	C0 $(mS/cm)^2$	C $(mS/cm)^2$	a (m)	(C+C0)/C0 (%)
Dry season	ECs1 (TS)	Spherical	0.146	0.719	19747.51	16.87
	Cov ECs1 (TS) - ECw1	Spherical	-	0.517	-	-
	Cov ECs1 (TS) - Sand (TS)	Spherical	-	-0.786	-	-
	ECs1 (SS)	Spherical	0.143	0.875	19897.06	14.04
	Cov ECs1 (SS) - ECw1	Spherical	-	0.555	-	-
	Cov ECs1 (SS) - Sand (SS)	Spherical	-	-0.837	-	-
Wet season	ECs2 (TS)	Spherical	0.059	0.979	19769.94	5.68
	Cov ECs2 (TS) - ECw2	Spherical	-	0.633	-	-
	Cov ECs2 (TS) - Sand (TS)	Spherical	-	0.871	-	-
	Cov ECs2 (TS) - pHw2	Spherical	-	-0.571	-	-
	ECs2 (SS)	Spherical	0.143	0.875	19897.06	14.04
	Cov ECs2 (SS) - ECw2	Spherical	-	0.593	-	-
	Cov ECs2 (SS) - Sand (SS)	Spherical	-	-0.837	-	-

3.3.2.2 SCOK with environmental covariates

Figure 67 and Figure 68 show respectively the fitted auto semivariograms and cross covariograms models of ECs prediction using SCOK with the environmental covariates. ECs1 (TS) and ECs2 (SS) experimental variograms are fitted using Gaussian models. ECs1 (SS) and ECs2 (TS) are matched using a spherical model (the best model was chosen based on cross validation results). Their different parameters are listed in Table 30. The nugget to sill ratio indicates strong spatial dependence for ECs1 (SS) and ECs2 (TS), and a moderate spatial dependence for ECs1 (TS) and ECs2 (SS).

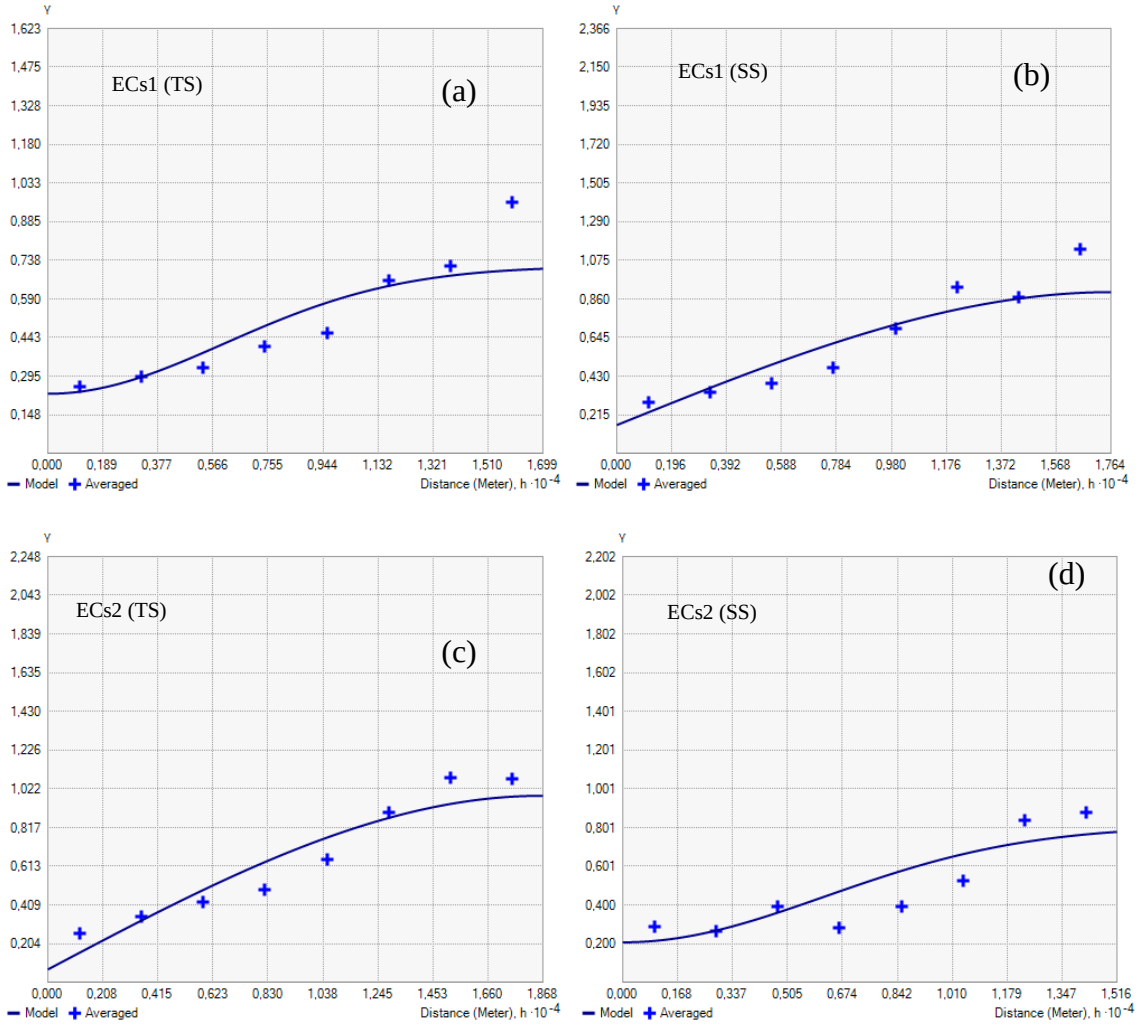


Figure 67: Auto semivariograms of ECs using environmental covariates: (a) ECs1 (TS), (b) ECs1 (SS), (c) ECs2 (TS), (d) ECs2 (SS)

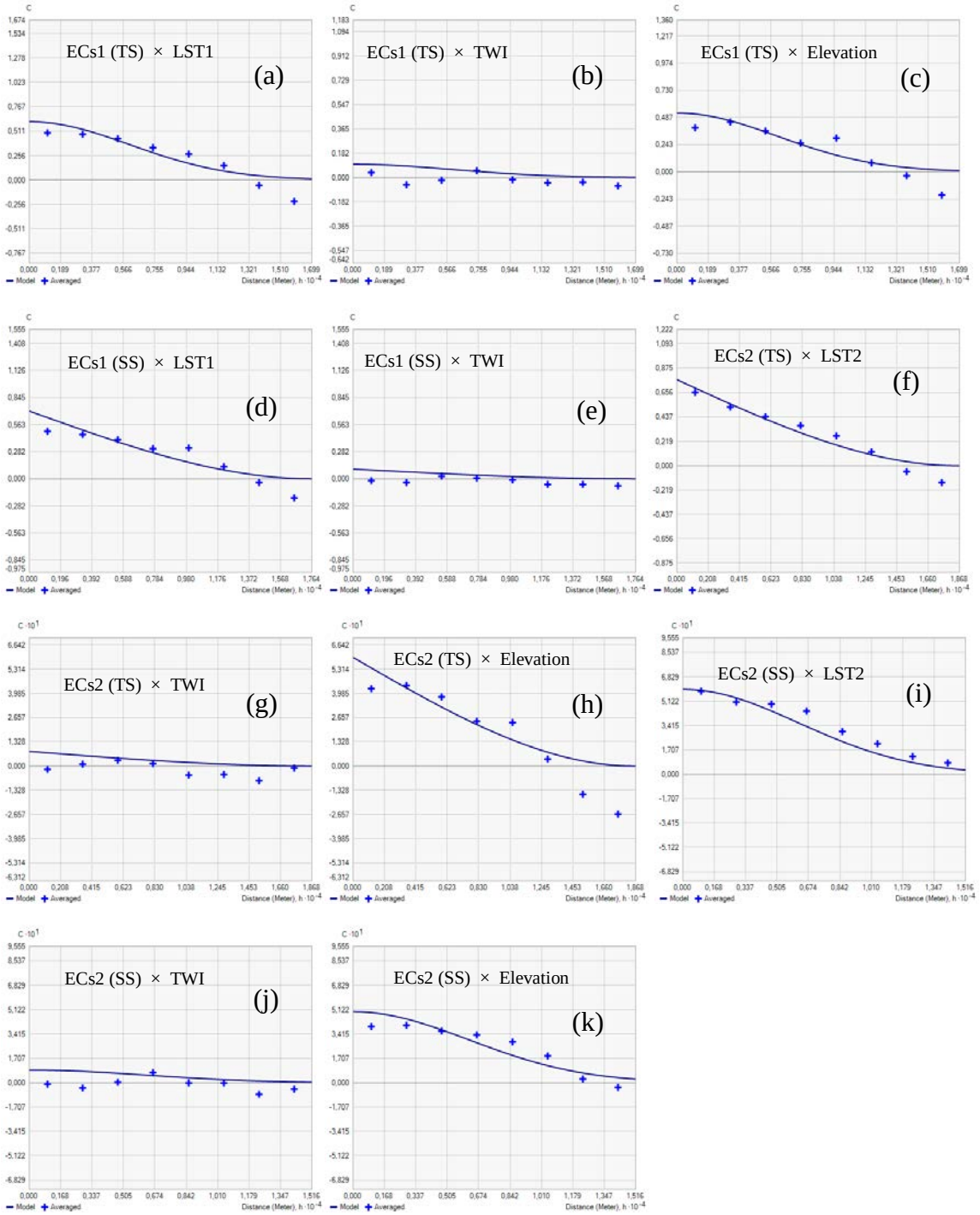


Figure 68: Cross covariograms between ECs and environmental covariates using SCOK: (a) ECs1 (TS) – LST1, (b) ECs1 (TS) – TWI, (c) ECs1 (TS) – Elevation, (d) ECs1 (SS) – LST1, (e) ECs1 (SS) – TWI, (f) ECs2 (TS) – LST2, (g) ECs2 (TS) – TWI, (h) ECs2 (TS) – Elevation, (i) ECs2 (SS) – LST2, (j) ECs2 (SS) – TWI, (k) ECs2 (SS) – Elevation.

Table 30: Parameters of the auto semivariogram models for ECs and their cross covariogram models (using environmental covariates).

Period	Variable	model	C0 $(mS/cm)^2$	C $(mS/cm)^2$	a (m)	C0/ (C+C0) (%)
Dry season	ECs1 (TS)	Gaussian	0.277	0.487	14799.79	31.79
	Cov ECs1 (TS) - LST1	Gaussian	-	0.609	-	-
	Cov ECs1 (TS) - TWI	Gaussian	-	0.1	-	-
	Cov ECs1 (TS) - Elevation	Gaussian	-	0.526	-	-
	ECs1 (SS)	Spherical	0.157	0.65	17644.21	17.50
	Cov ECs1 (SS) - LST1	Spherical	-	0.662	-	-
	Cov ECs1 (SS) - TWI	Spherical	-	0.11	-	-
Wet season	ECs2 (TS)	Spherical	0.067	0.916	18565.18	6.81
	Cov ECs2 (TS) - LST2	Spherical	-	0.77	-	-
	Cov ECs2 (TS) - TWI	Spherical	-	0.09	-	-
	Cov ECs2 (TS) - Elevation	Spherical	-	0.696	-	-
	ECs2 (SS)	Gaussian	0.244	0.724	15549.37	25.20
	Cov ECs2 (SS) - LST2	Gaussian	-	0.64	-	-
	Cov ECs2 (SS) - TWI	Gaussian	-	0.1	-	-
	Cov ECs2 (SS) - Elevation	Gaussian	-	0.493	-	-

The MAE, RMSE and R performance statistics of training and testing periods for SCOK with field covariates and SCOK with environmental covariates are given in Tables 31 and 32. All models reflect accurate results in the testing period in both seasons. For ECs1 (SS), SCOK with ECw1 and sand (SS) covariates gave more accurate results than SCOK with LST1 and TWI covariates, with the smallest MAE of 0.44, RMSE of 0.59 and the highest R of 0.972, respectively. With MAE of 0.86, 0.48 and 0.57; RMSE of 1.38, 0.56 and 0.77; and R of 0.867, 0.917 and 0.981, SCOK of ECs1 (TS), ECs2 (TS) and ECs2 (SS) using LST2, TWI and elevation as covariates showed more accurate results than SCOK using field covariates.

Table 31: Prediction performance measures of SCOK for training and testing datasets using field covariates

		Training			Testing		
Period	Variable	MAE	RMSE	R	MAE	RMSE	R
Dry season	ECs1 (TS)	1.01	1.71	0.77	0.99	1.53	0.827
	ECs1 (SS)	0.82	1.27	0.783	0.44	0.59	0.972
Wet season	ECs2 (TS)	0.72	1.28	0.775	0.65	0.87	0.863
	ECs2 (SS)	0.71	1.13	0.787	0.55	0.78	0.961

Table 32: Prediction performance measures of SCOK for training and testing datasets using environmental covariates

		Training			Testing		
Period	Variable	MAE	RMSE	R	MAE	RMSE	R
Dry season	ECs1 (TS)	1.24	1.99	0.672	0.86	1.38	0.867
	ECs1 (SS)	0.91	1.45	0.716	0.65	0.92	0.961
Wet season	ECs2 (TS)	0.76	1.26	0.786	0.48	0.56	0.917
	ECs2 (SS)	0.91	1.45	0.716	0.57	0.77	0.981

From the spatial distribution maps of soil salinity generated using SCOK with field covariates (Figure 69) and SCOK with environmental covariates (Figure 70), one can distinguish the dominance of two salinity classes, the first one is slightly to low salinity class and the second one is the excessively high salinity class. As mentioned on the maps the first class occupies generally the eastern part of El Outaya plain and the second class occupies the western part of the study area. The produced maps of ECs1 (TS) using SCOK with field covariates and SCOK with environmental covariates show from slightly to low salinity class occupation areas as 38.62 % and 39.69 %, and the excessively high salinity occupies areas of 43.13 % and 35.9 %, respectively; whereas the generated maps of ECs1 (SS) indicate the decreasing of slightly to low salinity class with surface areas 36.27 % and 35.28 % and the increasing of the excessively high salinity to 44.74 % and 43.95 %, respectively. For ECs2 (TS) and ECs2 (SS) maps with field covariates and environmental covariates, the surface areas occupied slightly to low salinity class are 41.86 %, 43.10 % and 38.45 %, 43.10 %, respectively, while the surface areas occupied by the excessively high salinity class are 38.42 %, 35.00 % and 41.96 % and 34.52 %, respectively.

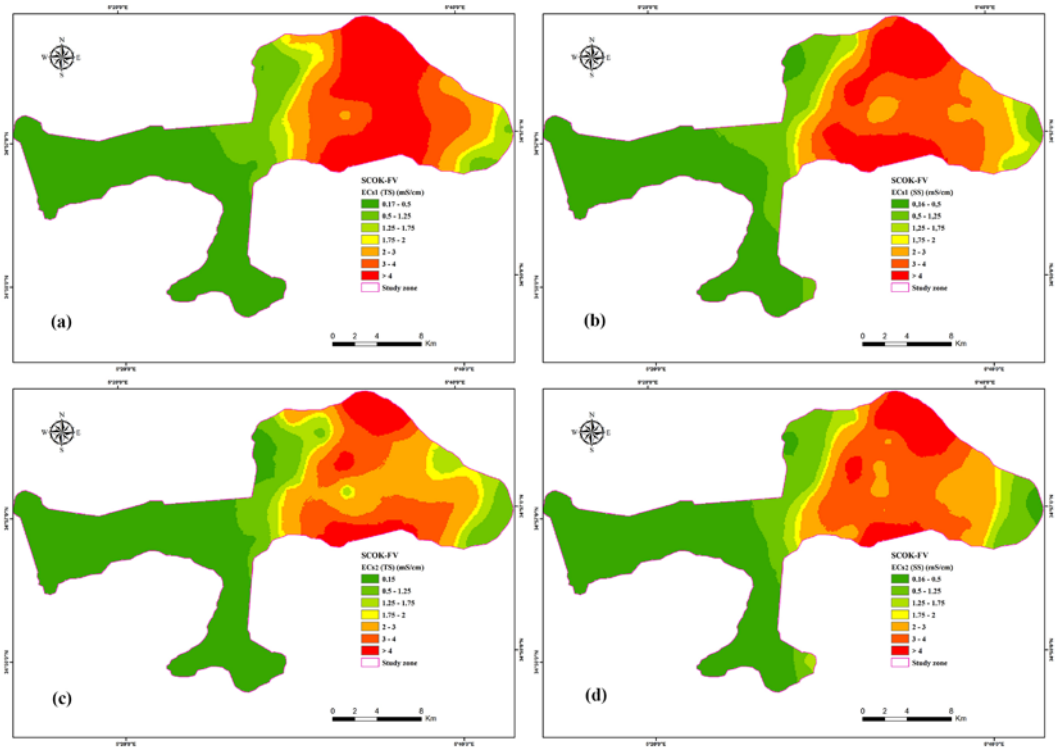


Figure 69: Prediction maps of SCOK using field covariates: (a) ECs1 (TS), (b) ECs1 (SS), (c) ECs2 (TS), (d) ECs2 (SS)

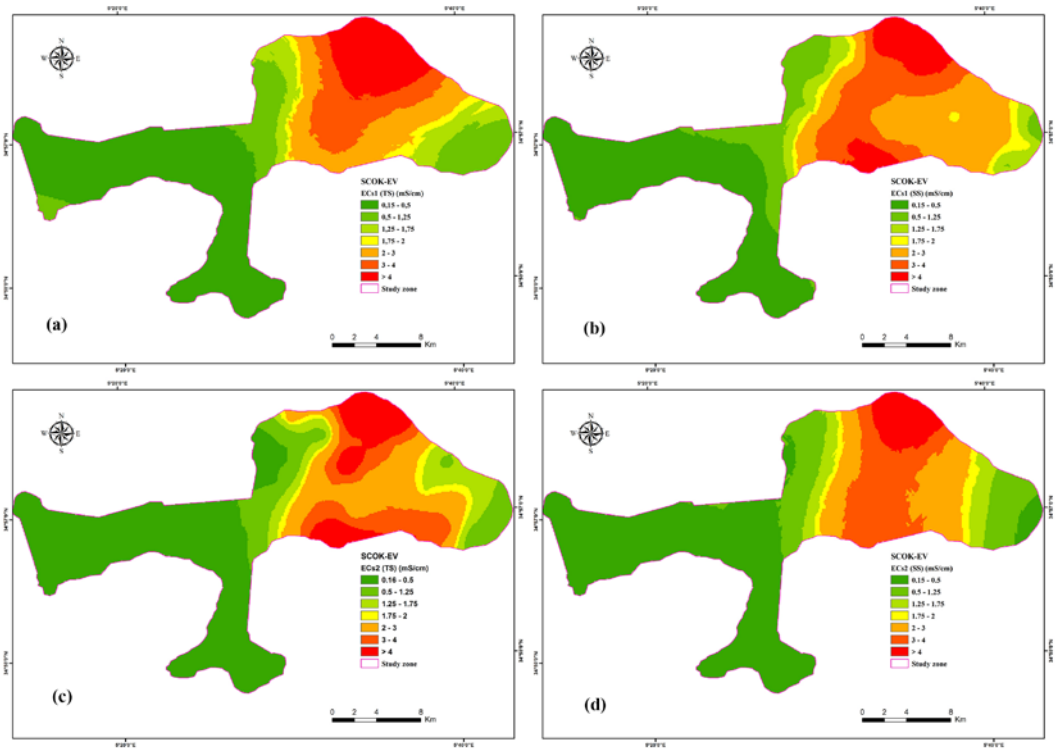


Figure 70: Prediction maps of SCOK using environmental covariates: (a) ECs1 (TS), (b) ECs1 (SS), (c) ECs2 (TS), (d) ECs2 (SS)

3.4 Artificial neural network and machine learning models

3.4.1 Multilayer perceptron neural network (MLP-NN)

Different MLP-NN models with Broyden–Fletcher– Goldfarb–Shanno (BFGS Quasi-Newton) back propagation algorithm are constructed for predicting soil salinity (ECs) in El Outaya plain using field covariates and environmental covariates as inputs. Table 33 and Table 34 show the properties of the MLP-NN models used for predicting ECs with field covariates and environmental covariates, respectively. The optimal MLP-NN structure, the number of hidden neurons and the activation functions are identified using a trial and error procedure by varying the number of hidden neurons from 1 to 20 and checking the performance criteria generated automatically by the software (the least SOS and highest R).

As shown in Table 33, MLP (2-10-1) with logistic activation function in the hidden layer and exponential (Expo) activation function in the output layer, MLP (2-6-1) with identity and exponential activation functions, MLP (3-8-1) with logistic activation function in the hidden and output units, and MLP (2-10-1) with Tanh hidden activation function and exponential output activation function are the best structures for predicting ECs1 (TS), ECs1 (SS), ECs2 (TS) and ECs2 (SS) using field covariates, respectively.

Table 33: Properties of the MLP-NN used for modeling ECs with field covariates

Inputs	Output	ANN structure	Training error (SOS)	Testing error (SOS)	Hidden activation	Output activatio-n
ECsw1, sand (TS)	ECs1 (TS)	MLP 2-10-1	0.008	0.007	Logistic	Expo
ECsw1, sand (SS)	ECs1 (SS)	MLP 2-6-1	0.007	0.001	Identity	Expo
ECw2, silt (SS), pHw2	ECs2 (TS)	MLP 3-8-1	0.005	0.008	Logistic	Logistic
ECw2, sand (SS)	ECs2 (SS)	MLP 2-10-1	0.005	0.003	Tanh	Expo

The properties of MLP-NN models for predicting ECs using environmental covariates as inputs are listed in Table 12. The optimal structures of MLP-NN for predicting ECs1 (TS), ECs1 (SS), ECs2 (TS) and ECs2 (SS) are MLP (3-5-1) with identity activation function in the hidden layer and logistic output activation function, MLP (2-8-1) with Tanh activation function in the hidden unit and exponential activation function in the output unit, MLP (3-5-1) with logistic function in the hidden layer and exponential activation function in the output layer, and MLP (3-5-1) with logistic activation function in both layers.

Table 34: Properties of the MLP-NN used for modeling ECs with environmental covariates

Inputs	Output	ANN structure	Training error (SOS)	Testing error (SOS)	Hidden activation	Output activation
LST1, TWI, elevation	ECs1 (TS)	MLP 3-5-1	0.012	0.008	Identity	Logistic
LST1, TWI	ECs1 (SS)	MLP 2-8-1	0.008	0.013	Tanh	Exponential
LST2, TWI, elevation	ECs2 (TS)	MLP 3-5-1	0.012	0.003	Logistic	Exponential
LST2, TWI, elevation	ECs2 (SS)	MLP 3-5-1	0.012	0.014	Logistic	Logistic

Table 35 and Table 36 show the calculated performance criteria for the evaluation of MLP-NN models field and environmental covariates, respectively. The performance criteria showed accurate results for training and testing periods with more accuracy in the testing stage.

From Table 35, MLP-NN modeling using field covariates as inputs are more accurate in predicting ECs1 (SS) and ECs2 (SS) with MAE of 0.34 and 0.38, RMSE of 0.49 and 0.6, and R of 0.977 and 0.968 in the testing period, respectively. For MLP-NN modeling using environmental covariates as inputs, the performance criteria of the testing stage (Table 36) indicate more accurate results in predicting ECs2 (TS) and ECs2 (SS), with MAE of 0.43 and 0.95, RMSE of 0.6 and 1.22 and R of 0.946 and 0.961, respectively.

Table 35: Performance measures of MLP-NN models using field covariates

	Training performance			Testing performance		
Output	MAE	RMSE	R	MAE	RMSE	R
ECs1 (TS)	0.86	1.53	0.837	0.87	1.42	0.889
ECs1 (SS)	0.7	1.21	0.824	0.34	0.49	0.977
ECs2 (TS)	0.41	0.79	0.922	0.75	1.01	0.813
ECs2 (SS)	0.5	0.76	0.903	0.38	0.6	0.968

Table 36: Performance measures of MLP-NN models using environmental covariates

	Training performance			Testing performance		
Output	MAE	RMSE	R	MAE	RMSE	R
ECs1 (TS)	1.26	1.87	0.749	0.9	1.52	0.845
ECs1 (SS)	1.02	1.29	0.773	1.24	1.63	0.702
ECs2 (TS)	0.82	1.22	0.798	0.43	0.6	0.946
ECs2 (SS)	0.84	1.13	0.768	0.95	1.22	0.961

3.4.2 Support vector machine (SVM)

The parameters of the kernel functions of the SVM models (C, ε and ʊ) and the stopping errors generated automatically by the software (MSE) during the training and testing periods are shown in Table 37 and Table 38. The optimal kernel functions are determined by varying the kernel parameters in order to obtain the lowest MSE and highest R. As noted in Table 37, the radial basis functions (RBF) are the kernels applied to train SVMs and predicting ECs using field covariates. The values of the error margin (ε), the kernel scale (C) and the regularization parameter (ʊ) have high influence on the results of the SVM models (Jalal et al. 2020). The values of C, ε and ʊ that gave the best SVM models are 0.15, 0.1 and 1 for ECs1 (TS) and ECs1 (SS), 0.3, 0.1 and 0.2 for ECs2 (TS), and 14, 1 and 1 for ECs2 (SS), respectively. The properties of the kernels used to train SVMs with environmental covariates are shown in Table 38. RBFs showed the lowest stopping errors for ECs1, while polynomial kernels showed the best results for ECs2.

Table 37: Characteristics of the kernels and SVM models using field covariates

Output	Kernel function	C	ε	ʊ	training errors (MSE)	Testing Errors (MSE)
ECs1 (TS)	RBF	15	0.1	1.0	0.02	0.021
ECs1 (SS)	RBF	15	0.1	1.0	0.011	0.01
ECs2 (TS)	RBF	0.3	0.1	0.2	0.046	0.043
ECs2 (SS)	RBF	14	1.0	1.0	0.019	0.015

Table 38: Characteristics of the kernels and SVM models using environmental covariates

Output	Kernel function	C	ε	ʊ	training errors (MSE)	Testing Errors (MSE)
ECs1 (TS)	RBF	7.0	0.15	0.33	0.02	0.021
ECs1 (SS)	RBF	1.0	0.20	0.60	0.027	0.034
ECs2 (TS)	Polynomial	3.0	1.0	0.30	0.029	0.014
ECs2 (SS)	Polynomial	4.0	0.10	0.33	0.031	0.041

The results of SVM models' performance in terms of MAE, RMSE and R during the training and testing periods are presented in Table 39 and Table 40. From Table 39, SVM models for predicting ECs1 (SS) and ECs2 (SS) by using of ECw and sand % as inputs show clearly

superior performance in the testing stage with MAE of 0.61 and 0.58, RMSE of 0.77 and 0.88, and R of 0.966 and 0.939, respectively.

Table 39: The calculated performance criteria for SVM models with field covariates

		Training			Testing		
Input combination	Output	MAE	RMSE	R	MAE	RMSE	R
ECw1, Sand (TS)	ECs1 (TS)	0.93	1.72	0.785	1.25	2.01	0.645
ECw1, Sand (SS)	ECs1 (SS)	0.7	1.05	0.854	0.61	0.77	0.966
ECw2, pHw2. Silt (TS)	ECs2 (TS)	0.94	1.68	0.768	0.83	1.36	0.698
ECw2, Sand (SS)	ECs2 (SS)	0.61	1.01	0.824	0.58	0.88	0.939

The calculated performance criteria in terms of MAE, RMSE and R for SVM modeling with environmental covariates (Table 40) indicate a superior performance for the prediction of ECs2 (TS) and ECs2 (SS) with the lowest MAE (0.69 and 1.2) and RMSE (0.93 and 1.54), and the highest R (0.903 and 0.922).

Table 40: The calculated performance criteria for SVM models with environmental covariates

		Training performance			Testing performance		
Input combination	Output	MAE	RMSE	R	MAE	RMSE	R
Elevation, LST1, TWI	ECs1 (TS)	1.38	1.74	0.777	1.35	1.76	0.77
LST1, TWI	ECs1 (SS)	1.25	1.66	0.615	1.56	1.86	0.631
Elevation, LST2, TWI	ECs2 (TS)	0.86	1.33	0.818	0.69	0.93	0.903
Elevation, LST2, TWI	ECs2 (SS)	0.96	1.27	0.716	1.2	1.54	0.922

4 Discussion and comparison of the different models

For the purpose of more clarity and a better comparison between the different modeling techniques, the Taylor diagram during the testing period was utilized (Figure 71). The Taylor diagram is a graphical framework that allows the comparison of different models in the same time and provides concise statistical summary of how well patterns match each other in terms of their correlation, the ratio of their standard deviations, and their root-mean-square difference (Santos et al. 2019; Koulla et al. 2019). The overall comparisons between the different modeling techniques applied to predict the soil salinity in terms of ECs in El Outaya plain illustrated the superiority of MLP-NN over the other modeling methods, either by using the field covariates (ECS1 (TS), ECs1 (SS) and ECs2 (SS)) or the environmental covariates (ECs2 (TS)).

As illustrated in Figure 71a, b and d, for the prediction of ECs1 (TS), ECs1 (SS) and ECs2 (SS), the MLP-NN models. ECw and sand % as inputs provided better performance results than the other models and could explain 79.1 % and 95.5 % of the top soil and the sub soil ECs variability in dry season, respectively, whereas could explain 94 % of the sub soil ECs variability in the wet season. However, SCOK with the same covariates could explain 75.2 % of the ECs1 (TS) variability, 92.5 % ECs1 (SS) variability and 92.5 % ECs2 (SS) variability in the study area. SVM using ECw1 and sand % as inputs to predict ECs1 (TS) and SVM using LST, TWI and elevation as inputs to predict ECs1 (SS) were the worst models with determination coefficients of 0.41 and 0.40, respectively.

Taylor diagram (Figure 71c) showed that MLP-NN using LST2, TWI and elevation is the best model to predict the top soil salinity in the wet season, where it could explain 90 % of ECs2 (TS) variability. SVM using ECw2, silt % and pHw was the worst model with $R^2 = 0.49$.

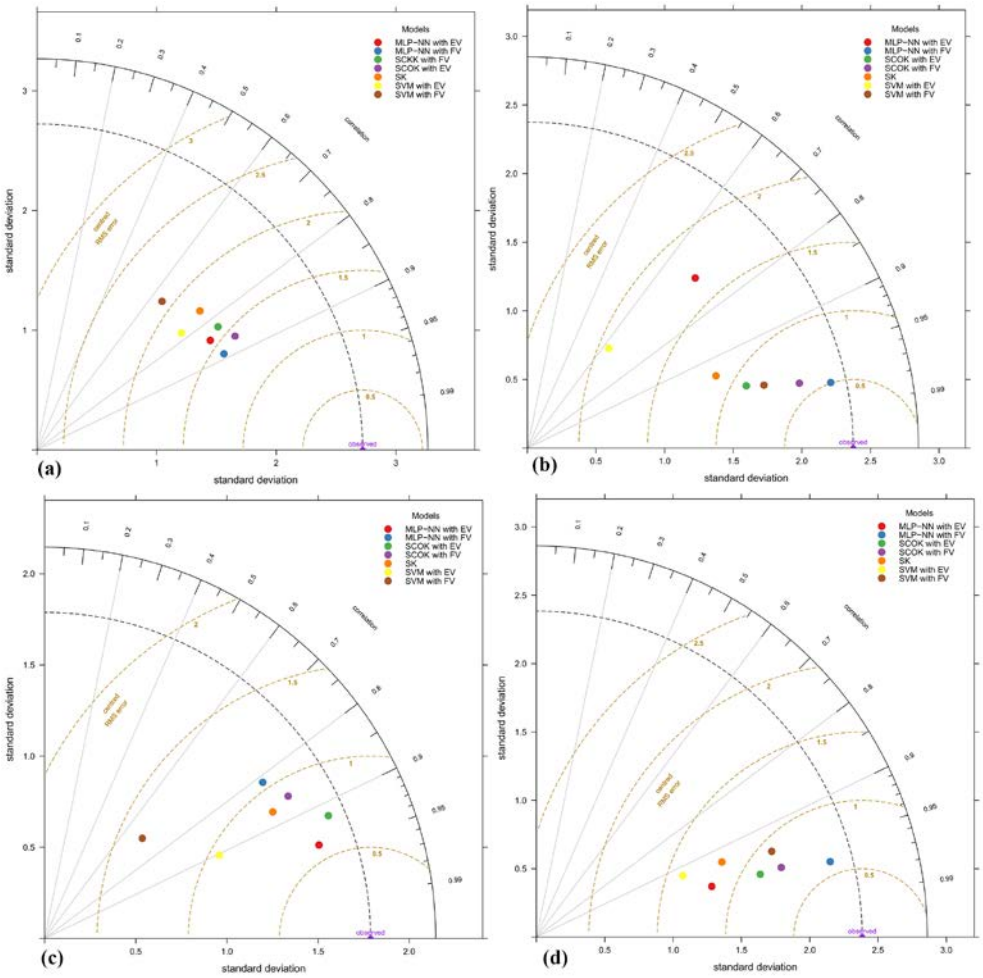


Figure 71: Taylor diagram for the testing results of the modeling techniques: (a) ECs1 (TS), (b) ECs1 (SS), (c) ECs2 (TS), (d) ECs2 (SS)

5 Conclusion

The present chapter aimed to improve the prediction of top soil and sub soil salinity over two seasons in El Outaya plain by using of two categories of covariates, the field covariates that affect directly the soil salinization (soil and water properties) and the environmental covariates (remote sensing and topographic covariates). Multiple linear stepwise regression, simple Kriging (SK), simple cokriking (SCOK), multilayer perceptron neural network (MLP-NN) and support vector machine (SMV) are used in this thesis. The results showed that the spatial prediction of soil salinity could be improved using cokriging with field and environmental auxiliary variables. Compared to SK, SCOK provides the lowest mean absolute errors (MAE) and root mean square errors (RMSE), and the highest correlations (R) between the measured and the predicted soil salinity. The spatial prediction maps of ECs indicated low concentrations in the western part of El Outaya plain, whereas the high concentrations are located in the eastern part where the agriculture is more intensive, the irrigation water is more saline and the soil are characterized by heavy textures.

The comparison of the performance of the different modeling techniques depicted the superiority of MLP-NN in the accuracy of estimating soil salinity with lower MAE and RMSE, and higher Pearson's correlation coefficients. Thus, the results of SCOK are close to those of MLP-NN. In addition, the combination of ECw and sand % from the field covariates and TWI, LST and elevation as inputs to MLP-NN and auxiliary variables to SCOK can estimate and improve the predictions of soil salinity in the study area.

General conclusion and recommendations

General conclusion

The literature has demonstrated that agricultural irrigation is the main use of global water resources and groundwater provides about 43 % of agricultural irrigation. In arid and semi-arid regions, the water demand is increasing due to the scarcity of rainfall and drought, which lead to the overexploitation of groundwater. The bad use of this valuable resource for irrigation accompanied with poor agricultural practices cause the deterioration of its quality. The irrigation with low groundwater quality (e.g. saline) contributes significantly to soil salinization in irrigated areas, which is the major threats of land degradation occurring around the world, posing risks to agricultural production, environmental health and economic prosperity.

The situation of El Outaya plain in the southwest of Algeria gives it the characteristic of aridity, where rainfall is scarce, evapotranspiration rates are higher than precipitations and the dry season is spread out over the whole year. In addition, the rapid extension of agriculture and urbanization put more pressure on the limited groundwater resources of the plain, particularly the Mio-Pliocene, which is the main exploited aquifer in the study area. The apparition of white crust on the top soil of irrigated areas in the plain is a visual indication of the impact of groundwater salinity on soil salinization.

The main issue addressed in this thesis, namely the hydrogeochemical analyses and assessment of groundwater quality for human consumption and irrigation and the impact of the salinity on soil salinization in El Outaya plain is complex and needs to be addressed at different distinct levels. In order to answer this question, this research study conducted several aspects.

The hydrochemical analyses based on major cations and anions, and the salinity based on electrical conductivity (EC) are the most used criteria for assessing the impact of irrigation water on the soil and the irrigated crops. The results of hydro-geochemical analysis reveal that sodium and calcium are the dominant cations, while chloride and sulfate are the dominant anions in the groundwater. The trilinear Piper diagram shows the prevalence of $(Na^+ + K^+) - (Cl^- + SO_4^{-2})$ and $(Ca^{2+} + Mg^{2+}) - (Cl^- + SO_4^{-2})$ hydrochemical facies. The dissolution of evaporate rocks (halite, calcite and anhydrite) and reverse ion exchange are the main hydrochemical processes that control the groundwater chemistry as a result of rock-water interaction. The most of groundwater samples are falling under C4S2 and C4S3 classes, which can be used for irrigation on coarse textured or organic soils with good permeability and more tolerable crops should be selected. About 30 % of the groundwater are unsuitable for irrigation

purpose (C5S3 and C5S4 classes), which indicates a high risk of deterioration of the soil structure in the study area using this groundwater for irrigation.

The spatial distribution of groundwater salinity in terms of EC shows general increasing trends from the east to the center and from the west to the center of the study area. The dominance of groundwater with high and very high salinity in the study area could be due to the intense exploitation of the aquifer, the effects of highly saline irrigation return flows on the groundwater, the low rate of precipitation and the dissolution of Triassic evaporitic materials in Djebel Elmaleh. It is concluded that the use of the groundwater for irrigation in El Outaya plain will deteriorate the soil structure, increase soil salinization, damage the cultivated crops and reduce yields.

The groundwater of the Mio-Pliocene aquifer is the main source of agricultural irrigation in El-Outaya plain, which means that the subject of the study area depends on different farming activities (application of different chemical treatments. chemical and organic fertilization), but farmers neglect its high salinity and they are still using it for drinking purposes. The field observations led to the question about the contamination of this valuable resource and its suitability for human consumption in terms of NO3-N concentration. The application of robust interpolation technique is crucial for the prediction of spatial variability of contaminants in groundwater. All the interpolation techniques used for this purpose, including the linear and nonlinear geostatistical techniques (empirical Bayesian kriging (EBK)), oridinary kriging (OK) and indicator kriging (IK)) results and deterministic method (radial basis function (RBF)) have a high accuracy to predict groundwater nitrate-N contamination in the study area, with relative superiority of RBF. IK technique has been used to map the probability of excessive groundwater nitrate-N in the study area. The transformation of these probability maps to categorical maps using the developed method shows the tendency of the groundwater quality at two levels to exceed a given threshold. IK at a medium risk exceeds thresholds ($P \geq 50\%$) indicating that the groundwater in the study area tends to have intermediate quality with an estimated area of 50.69% and at a high risk to exceed a given threshold ($P \geq 90\%$), and the groundwater tends to be of good quality over an estimated area of 55.24%. The presented transformation of probability maps to categorical maps based on IK is important for decision makers to identify regions with high nitrate-N contamination probability, and high risk to human health, and could be applied to other risk assessment studies.

The farmers of El Outaya plain are using groundwater of poor quality, usually unsuitable for irrigation and characterized by high salinity, which has enhanced the development of

agricultural sector but it is also a principle factor of soil salinization. In the aim of modeling, spatiotemporal monitoring and assessment of top soil and sub soil salinity (in terms of soil electrical conductivity (ECs)) over two seasons in El Outaya plain by using of two categories of covariates, the field covariates that affect directly the soil salinization (soil and water properties) and the environmental covariates (remote sensing and topographic covariates), Multiple linear stepwise regression, simple kriging (SK), simple cokriking (SCOK), multilayer perceptron neural network (MLP-NN) and support vector machine (SMV) were used. The spatial prediction of soil salinity was improved using cokriging with field and environmental auxiliary variables, where low concentrations are located in the western part of El Outaya plain whereas the high concentrations located in the eastern part, which is characterized by an intensive agriculture, the irrigation water is more saline and the soil are characterized by heavy textures. The comparison of the performance of the different modeling techniques depicted the superiority of MLP-NN in the accuracy of estimating soil salinization and the combination of groundwater salinity in terms of ECw and sand percentage from the field covariates and Topographic Wetness Index (TWI), Land Surface Temperature (LST) and elevation as inputs to MLP-NN and auxiliary variables to SCOK used accurately to estimate and improve the predictions of soil salinity in the study area.

Recommendations

The following recommendations made based on the results obtained from this study must be taken into consideration for future application by decision makers:

- The construction of new wells without hydrogeological study by the responsible institutions should be stopped totally and the new wells should be constructed in appropriate areas respecting the distance between wells and the capacity of the captured aquifer;
- The number of old and traditional wells should be inventoried and studying the possibility of replacing them with new ones respecting the international standards;
- Raising awareness among farmers about the dangers of saline water for irrigation and encourage them to apply appropriate leaching and drainage systems to minimize soil damages due to secondary salinization, and desalination of groundwater by reverse osmosis, which can decrease the EC the groundwater more than 95%;
- Assist and encourage farmers to use modern technologies for irrigation, fertilization and harvesting such as smart irrigation and precision agricultural systems;

- Establishment of good agricultural practical protocols for farmers to prevent and reduce groundwater pollution by nitrate-N and other contaminants related to poor agricultural practices;
- Establishment of action programs are advised to raise awareness among farmers about proper application of pesticides and fertilizers, and designation of pollution vulnerable zones;
- Although, Remote sensing combined to machine learning techniques performed well, further studies about using hyperspectral satellite images and Artificial Neural Networks (ANNs) in soil salinity and groundwater quality management are recommended.

References

References

Abbasi T, Abbasi SA (2014) Water Quality Indices (WQI). Elsevier, Amsterdam

Abdennour MA, Douaoui A, Barrena J, et al (2020) Geochemical characterization of the salinity of irrigated soils in arid regions (Biskra, SE Algeria). Acta Geochim 1–17. doi: 10.1007/s11631-020-00426-2

Abderrahman WA (2006) Groundwater Management for Sustainable Development of Urban and Rural Areas in Extremely Arid Regions : A Groundwater Management for Sustainable Development of Urban and Rural Areas in Extremely Arid Regions : A Case Study. Int J Water Resour Dev 21:403–412. doi: 10.1080/07900620500160735

ABHS (2016) Actualisation des inventaires des ouvrages et infrastructures de prélèvement d'eau du domaine public hydraulique à usage agricole - wilaya de Biskra. 27

Abuelgasim A, Ammad R (2018) Mapping Soil Salinity in Arid and Semi-Arid Regions Using Landsat 8 OLI Satellite Data. Remote Sens Appl Soc Environ 13:415–425. doi: 10.1016/j.rsase.2018.12.010

Abzalov M (2016) Variography. In: Applied Mining Geology. Springer International Publishing, Switzerland, pp 239–262

Achieng KO (2019) Computers and Geosciences Modelling of soil moisture retention curve using machine learning techniques : Artificial and deep neural networks vs support vector regression models. Comput Geosci 133:104320. doi: 10.1016/j.cageo.2019.104320

Adhikary PP, Ch. Dash J (2017) Comparison of deterministic and stochastic methods to predict spatial variation of groundwater depth. Appl Water Sci 7:339–348. doi: 10.1007/s13201-014-0249-8

Ahamad A, Madhav S, Singh P, et al (2018) Assessment of groundwater quality with special emphasis on nitrate contamination in parts of Varanasi City, Uttar Pradesh, India. Appl Water Sci 8:115. doi: 10.1007/s13201-018-0759-x

Ahmadi SH, Sedghamiz A (2007) Geostatistical Analysis of Spatial and Temporal Variations of Groundwater Level. Environ Monit Assess 129:277–294. doi: 10.1007/s10661-006-9361-z

Ahmed MA, Abdel Samie SG, Badawy HA (2013) Factors controlling mechanisms of groundwater salinization and hydrogeochemical processes in the Quaternary aquifer of the

Eastern Nile Delta, Egypt. Environ Earth Sci 68:369–394. doi: 10.1007/s12665-012-1744-6

Alharbi TG (2018) Identification of hydrogeochemical processes and their influence on groundwater quality for drinking and agricultural usage in Wadi Nisah, Central Saudi Arabia. Arab J Geosci 11:359. doi: 10.1007/s12517-018-3679-z

Allbed A, Kumar L (2013) Soil Salinity Mapping and Monitoring in Arid and Semi-Arid Regions Using Remote Sensing Technology : A Review. Adv Remote Sens 2:373–385

Amer KM (2008) Groundwater resources sustainability in Qatar: Problems and perspectives. In: GROUNDWATER FOR SUSTAINABLE DEVELOPMENT: PROBLEMS, PERSPECTIVES AND CHALLENGES. Taylor & Francis Group, London, pp 25–39

Amichi H, Bouarfa S, Kuper M, et al (2012) HOW DOES UNEQUAL ACCESS TO GROUNDWATER CONTRIBUTE TO MARGINALIZATION OF SMALL FARMERS ? THE CASE OF PUBLIC LANDS. Irrig Drain 61:34–44. doi: 10.1002/ird.1660

André D (2017) Geographical Models with Mathematica. Elsevier Ltd, Kidlington

Aouidane L, Belhamra M (2017) Journal of African Earth Sciences Hydrogeochemical processes in the Plio-Quaternary Remila aquifer. J African Earth Sci 130:38–47. doi: 10.1016/j.jafrearsci.2017.03.010

Appelo CAJ, Postma D (2005) Geochemistry, groundwater and pollution, Second edi. A.A. Balkema Publishers

Arslan H (2012) Spatial and temporal mapping of groundwater salinity using ordinary kriging and indicator kriging : The case of Bafra Plain , Turkey. Agric Water Manag 113:57–63. doi: 10.1016/j.agwat.2012.06.015

Asa E, Saafi M, Membah J, Billa A (2012) Comparison of Linear and Nonlinear Kriging Methods for Characterization and Interpolation of Soil Data. J Comput Civ Eng 26:11–18. doi: 10.1061/(ASCE)CP.1943-5487.0000118.

Avdan U, Jovanovska G (2016) Algorithm for Automated Mapping of Land Surface Temperature Using LANDSAT 8 Satellite Data. J Sensors 8

Babiker IS, Mohamed MAA, Terao H, et al (2004) Assessment of groundwater contamination by nitrate leaching from intensive vegetable cultivation using geographical information system. Environ Int 29:1009–1017. doi: 10.1016/S0160-4120(03)00095-3

Benchetrit M (1956) Les sols d'Algérie. Rev géographie Alp 44:749–761. doi: https://doi.org/10.3406/rga.1956.1790

BIERKENS MFP, BURROUGH PA (1993) The indicator approach to categorical soil data. Journalof Soil Sci 44:361–368

BISHOP CM (1995) Deep neural networks for pattern recognition

Bouderbala A (2017) Assessment of water quality index for the groundwater in the upper Cheliff plain, Algeria. J Geol Soc India 90:347–356. doi: 10.1007/s12594-017-0723-7

Boudibi S, Sakaa B, Zapata-Sierra AJ (2019) Groundwater quality assessment using GIS , ordinary kriging and WQI in an arid area. Ponte 75:204–226. doi: 10.21506/j.ponte.2019.12.14

Boudjema A (2015) Hydrogéologie, vulnérabilité et modélisation de la nappe du Mio-Pliocène d'El Outaya, (Biskra, Sud-Est algérien). Mohamed Khider Biskra

Bradaï A, Douaoui A, Bettahar N, Yahiaoui I (2016) Improving the Prediction Accuracy of Groundwater Salinity Mapping Using Indicator Kriging Method. J Irrig Drain Eng 142:11. doi: 10.1061/(ASCE)IR.1943-4774.0001019

Brown RM, McClelland NI, Deininger RA, Tozer RG (1970) A Water Quality Index: Do we dare? Water Sew Works 117:339–343

Cambardella CA, Moorman TB, Parkin TB, et al (1994) Field-Scale Variability of Soil Properties in Central Iowa Soils. Soil Sci Soc Am J 58:1501–1511. doi: 10.2136/sssaj1994.03615995005800050033x

Ceryan N (2014) Application of support vector machines and relevance vector machines in predicting uniaxial compressive strength of volcanic rocks. J African Earth Sci 100:634–644. doi: 10.1016/j.jafrearsci.2014.08.006

Chebbah M (2016) A Miocene-restricted platform of the Zibane zone (Saharan Atlas, Algeria), depositional sequences and paleogeographic reconstruction. Arab J Geosci 9:1–14. doi: 10.1007/s12517-015-2132-9

Chebbah M (2007) Lithostratigraphie, Sédimentologie et Modèles de Bassins des dépôt néogènes de la région de Biskra, de part et d'autre et de l'Accident Sud Atlasique (Zibans, Algérie). Mentouri University - Constantine

Chebbah M, MARMI K, Lamouroux C (2008) Sédimentologie Des Faciès Et Paléogéographie

Du Bassin D ’ El Outaya Au Néogène (Atlas Saharien , Algérie). Bull du Serv Géologique Natl 19:235–244

Chen W, Tanaka M (2002) A Meshless , Integration-Free , and Boundary-Only RBF Technique. Comput Math with Appl 43:379–391

Cherkassky V, Ma Y (2004) Practical selection of SVM parameters and noise estimation for SVM regression. Neural Networks 17:113–126. doi: 10.1016/S0893-6080(03)00169-2

Chica-olmo M, Luque-espinar JA, Rodriguez-galiano V, et al (2014) Science of the Total Environment Categorical Indicator Kriging for assessing the risk of groundwater nitrate pollution : The case of Vega de Granada aquifer (SE Spain). Sci Total Environ 470–471:229–239. doi: 10.1016/j.scitotenv.2013.09.077

Cortes C, Vapnik V (1995) Photonit neural networks and learning mathines the role of electron-trapping materials. Support Networks 20:273–297. doi: 10.1109/64.163674

Dai F, Zhou Q, Lv Z, et al (2014) Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau. Ecol Indic 45:184–194. doi: 10.1016/j.ecolind.2014.04.003

Deepa S, Venkateswaran S (2018) Appraisal of groundwater quality in upper Manimuktha sub basin, Vellar river, Tamil Nadu, India by using Water Quality Index (WQI) and multivariate statistical techniques. Model Earth Syst Environ 4:1165–1180. doi: 10.1007/s40808-018-0468-3

Deiss L, Margenot AJ, Culman SW, Demyan MS (2020) Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. Geoderma 365:114227. doi: 10.1016/j.geoderma.2020.114227

Delhomme JP (1978) Kriging in the hydrosciences. Adv Water Resour 1:251–266. doi: 10.1016/0309-1708(78)90039-8

Dhanasekarapandian M, Chandran S, Devi DS, Kumar V (2016) Spatial and temporal variation of groundwater quality and its suitability for irrigation and drinking purpose using GIS and WQI in an urban fringe. J African Earth Sci 124:270–288. doi: 10.1016/j.jafrearsci.2016.08.015

Ding JFÆY (2010) Assessment of groundwater contamination by NO 3 2 using geographical information system in the Zhangye Basin , Northwest China. Environ Earth Sci 60:809–816. doi: 10.1007/s12665-009-0218-y

Durand MJ-H, Barbut MM (1938) Carte de reconnaissance des sols d'Algerie: Biskra. Service Geographique de l'Armée (in French)

Durov, S.A. (1948) Natural Waters and Graphic Representation of Their Composition. Doklady Akademii Nauk SSSR, 59, 87-90.

El Alfy M, Lashin A, Abdalla F, Al-Bassam A (2017) Assessing the hydrogeochemical processes affecting groundwater pollution in arid areas using an integration of geochemical equilibrium and multivariate statistical techniques. Environ Pollut 229:760–770. doi: 10.1016/j.envpol.2017.05.052

Eldeiry A. ., Garcia LA (2011) Using indicator kriging technique for soil salinity and yield management. J Irrig Drain Eng 137:82–93. doi: 10.1061/(ASCE)IR.1943-4774.0000280

Erdik T, Savci ME, Şen Z (2009) Expert Systems with Applications Artificial neural networks for predicting maximum wave runup on rubble mound structures. Expert Syst with Appl J 36:6403–6408. doi: 10.1016/j.eswa.2008.07.049

ESRI (2019a) Deterministic methods for spatial interpolation. https://desktop.arcgis.com/fr/arcmap/latest/extensions/geostatistical-analyst/deterministic-methods-for-spatial-interpolation.htm (accessed 20.03.2020)

ESRI (2019b) What is Empirical Bayesian kriging?. https://desktop.arcgis.com/fr/arcmap/latest/extensions/geostatistical-analyst/what-is empirical-bayesian-kriging-.htm (accessed 20.03.2020)

ESRI (2020) Soil Adjusted Vegetation Index (SAVI). https://pro.arcgis.com/en/pro-app/arcpy/spatial-analyst/savi.htm (Accessed in 05.09.2020)

FAO (2008) Coping with water scarcity An action framework for agriculture and food security. Rome

FAO (2020) Water. http://www.fao.org/water/en/ (Accessed in 09.2020)

Farid I, Zouari K, Rigane A, Beji R (2015) Origin of the groundwater salinity and geochemical processes in detrital and carbonate aquifers: Case of Chougafiya basin (Central Tunisia). J Hydrol 530:508–532. doi: 10.1016/j.jhydrol.2015.10.009

Fourier J (2019) Geostatistical Variography for Geospatial Variables. In: Quantitative Geosciences: Data Analytics, Geostatistics, Reservoir Characterization and Modeling. Springer

Nature, Cham, pp 301–330

Gong Y, Zhang Y, Lan S, Wang H (2016) A Comparative Study of Artificial Neural Networks , Support Vector Machines and Adaptive Neuro Fuzzy Inference System for Forecasting Groundwater Levels. Water Resour Manag 30:375–391. doi: 10.1007/s11269-015-1167-8

Gorji T, Tanik A, Elif S (2015) Soil salinity prediction, monitoring and mapping using modern technologies. Procedia Earth Planet Sci 15:507–512. doi: 10.1016/j.proeps.2015.08.062

Gorji T, Yildirim A, Hamzehpour N, et al (2020) Soil salinity analysis of Urmia Lake Basin using Landsat-8 OLI and Sentinel- 2A based spectral indices and electrical conductivity measurements. Ecol Indic 112:106173. doi: 10.1016/j.ecolind.2020.106173

Gu Y, Bao Z, Lin Y, et al (2017) The porosity and permeability prediction methods for carbonate reservoirs with extremely limited logging data: Stepwise regression vs. N-way analysis of variance. J Nat Gas Sci Eng 42:99–119. doi: 10.1016/j.jngse.2017.03.010

Guiraud, R., 1990. Evolution post-triasique de l'avant pays de la chaîne alpine en Algérie d'après l'étude du bassin du Hodna et des régions voisines, Office National de la Géologie, Alger.

Haouchine A (2010) Hydrogeologie en zone semi aride et aride: region de Biskra. University of Science and Technology – Houari Boumediene

Haykin S (1999) Neural Networks: A Comprehensive Foundation, Second edi. Pearson Education, Ontario

Herceg S (2019) Chemical engineering research and design development of soft sensors for isomerization process based on support vector machine regression and dynamic polynomial models. Chem Eng Res Des 149:95–103. https://doi.org/10.1016/j.cherd.2019.06.034

HUETE AR (1988) A Soil-Adjusted Vegetation Index (SAVI) 295. Remote Sens Environ 25:295–309

Isaaks EH, Srivastava MR (1989) Applied Geostatistics. Oxford University Press, New York

Jalal M, Arabali P, Grasley Z, et al (2020) Behavior Assessment, Regression Analysis and Support Vector Machine (SVM) Modeling of Waste Tire Rubberized Concrete. J Clean Prod 273:122960. doi: 10.1016/j.jclepro.2020.122960

Johnston K, Ver Hoef JM, Krivoruchko K, Lucas N (2003) Using ArcGIS Geostatistical

Analyst. ESRI, Redlands (California)

Journel AG (1986) Geostatistics: Models and tools for the earth sciences. Math Geol 18:119–140

Kant N, Singh PK, Kumar B (2018) Hydrogeochemical Characterization and Groundwater Quality of Jamshedpur Urban Agglomeration in Precambrian Terrain, Eastern India. J Geol Soc India 92:67–75. doi: 10.1007/s12594-018-0954-2

Kawo SN, Karuppannan S (2018) Journal of African Earth Sciences Groundwater quality assessment using water quality index and GIS technique in Modjo River Basin , central Ethiopia. J African Earth Sci 147:300–311. doi: 10.1016/j.jafrearsci.2018.06.034

Ketata M, Gueddari M, Bouhlila R (2012) Use of geographical information system and water quality index to assess groundwater quality in el khairat deep aquifer (enfidha, central east tunisia). Arab J Geosci 5:1379–1390. doi: 10.1007/s12517-011-0292-9

Khan R, Jhariya DC (2018) Assessment of Land-use and Land-cover Change and its Impact on Groundwater Quality Using Remote Sensing and GIS Techniques in Raipur City , Chhattisgarh , India. J Geol Soc INDIA 92:59–66. doi: 10.1007/s12594-018-0953-3

Kim H, Kaown D, Mayer B, et al (2015) Identifying the sources of nitrate contamination of groundwater in an agricultural area (Haean basin, Korea) using isotope and microbial community analyses. Sci Total Environ 533:566–575. doi: 10.1016/j.scitotenv.2015.06.080

Kisi O, Alizamir M, Zounemat-Kermani M (2017) Modeling groundwater fluctuations by three different evolutionary neural network techniques using hydroclimatic data. Nat Hazards 87:367–381. doi: 10.1007/s11069-017-2767-9

Koulla N, Mohammed A, Santos CAG, Abdelkader D (2019) Spatial modeling of soil salinity using multiple linear regression, ordinary kriging and artificial neural network methods in the Lower Cheliff plain, Algeria. J Urban Environ Eng 13:34–41. doi: 10.4090/juee.2019.v13n1.034041

Krivoruchko, K., (2012) Empirical Bayesian Kriging. Esri, Redlands, CA, USA. Available online: <http://www.esri.com/news/arcuser/1012/empirical-byesian-kriging.html> (accessed 15.01.2020)

Krivoruchko K, Gribov A (2019) Evaluation of empirical Bayesian kriging. Spat Stat 32:100368. doi: 10.1016/j.spasta.2019.100368

Krzysztof G Meta-Learning in Decision Tree Induction

Kumar PJS, James EJ (2013) Development of Water Quality Index (WQI) model for the groundwater in Tirupur district, South India. Chinese J Geochemistry 32:261–268. doi: 10.1007/s11631-013-0631-5

Lee J, Jang C, Wang S, Liu C (2007) Evaluation of potential health risk of arsenic-affected groundwater using indicator kriging and dose response model. Sci Total Environ 384:151–162. doi: 10.1016/j.scitotenv.2007.06.021

Liu C, Jang C, Liao C (2004) Evaluation of arsenic contamination potential using indicator kriging in the Yun-Lin aquifer (Taiwan). Sci Total Environ 321:173–188. doi: 10.1016/j.scitotenv.2003.09.002

Machado RMA, Serralheiro RP (2017) Soil salinity: Effect on vegetable crop growth. Management practices to prevent and mitigate soil salinization. Horticulturae 3:1–13. doi: 10.3390/horticulturae3020030

Madhav V, Dipak W, Panaskar B, Avinash A (2016) Prediction of groundwater suitability for irrigation using artificial neural network model : a case study of Nanded tehsil ,. Model Earth Syst Environ 2:1–10. doi: 10.1007/s40808-016-0250-3

Maroufpoor S, Fakhri-Fard A, Shiri J (2017) Study of the spatial distribution of groundwater quality using soft computing and geostatistical models. ISH J Hydraul Eng 25:232–238. doi: 10.1080/09715010.2017.1474389

Masad IS, Al-Fahoum A, Abu-Qasmieh I (2019) Automated measurements of lumbar lordosis in T2-MR images using decision tree classifier and morphological image processing. Eng Sci Technol an Int J 22:1027–1034. doi: 10.1016/j.jestch.2019.03.002

Mitášová H, Mitáš L (1993) Interpolation by Regularized Spline with Tension : I. Theory and Implementation. Math Geol 25:641–655

Oliver MA (2010) Geostatistical Applications for Precision Agriculture. Springer Science+Business, New York

Oliver MA, Webster R (2014) A tutorial guide to geostatistics : Computing and modelling variograms and kriging. Catena 113:56–69. doi: 10.1016/j.catena.2013.09.006

Pannnatier Y (1996) VARIOWIN: Software for Spatial Data Analysis in 2D. Springer, New

York

Piper M (1944) A graphic prosedure in the geochemical interpretation of water analyses. Trans Am Geophys union 914–928

Pouladi N, Asghar A, Farzin J, et al (2019) Design and implementation of a hybrid MLP-FFA model for soil salinity prediction. Environ Earth Sci 78:159. doi: 10.1007/s12665-019-8159-6

Pulido-Bosch A, Rigol-Sanchez JP, Vallejos A, et al (2018) Impacts of agricultural irrigation on groundwater salinity. Environ Earth Sci 77:197. doi: 10.1007/s12665-018-7386-6

Rahman A, Tiwari KK, Mondal NC (2020) Assessment of hydrochemical backgrounds and threshold values of groundwater in a part of desert area, Rajasthan, India. Environ Pollut 266:115150. doi: 10.1016/j.envpol.2020.115150

Raju JN, Ram PR, Dey S (2009) Groundwater Quality in the Lower Varuna River Basin, Varanasi District, Uttar Pradesh. J Geol Soc INDIA 73:178–192

Raju NJ, Shukla UK., Ram P (2011) Hydrogeochemistry for the assessment of groundwater quality in Varanasi : a fast-urbanizing center in Uttar Pradesh , India. Environ Monit Assess 173:279–300. doi: 10.1007/s10661-010-1387-6

Ravikumar P, Somashekar RK, Prakash KL (2016) A comparative study on usage of Durov and Piper diagrams to interpret hydrochemical processes in groundwater from SRLIS river basin , hydrochemical processes in groundwater from SRLIS river basin , Karnataka , India. Earth Sci 80:31073–31077

Rawat KS, Mishra AK, Singh SK (2017a) Mapping of groundwater quality using Normalized Difference Dispersal Index of Dwarka sub-city at Delhi National Capital of India. ISH J Hydraul Eng 23:229–240. doi: 10.1080/09715010.2016.1277795

Rawat KS, Singh SK, Jacintha TGA, et al (2017b) Appraisal of long term groundwater quality of peninsular India using water quality index and fractal dimension. J Earth Syst Sci 126:122. doi: 10.1007/s12040-017-0895-y

Rodier J, Legube B, Merlet N (2009) Analyse de L'eau 9eme edition. DUNOD, Paris

Sadat-Noori SM, Ebrahimi K, Liaghat AM (2014) Groundwater quality assessment using the Water Quality Index and GIS in Saveh-Nobaran aquifer, Iran. Environ Earth Sci 71:3827–3843. doi: 10.1007/s12665-013-2770-8

Sajil Kumar PJ, Elango L, James EJ (2013) Assessment of hydrochemistry and groundwater quality in the coastal area of South Chennai, India. Arab J Geosci 28:497–505. doi: 10.1007/s12517-013-0940-3

Sakaa B, Brahmia N, Chaffai H, Hani A (2020) Desalination and Water Treatment Assessment of water quality index in unmonitored river basin using multilayer perceptron neural networks and principal component analysis. Desalin water Treat 1–13. doi: 10.5004/dwt.2020.26108

Salem A, Combes A, Caquel et al (1979) Carte hydrgeologique de Biskra. Direction des études du milieu et de la recherche hydraulique, Alger.

Samsonova VP, Blagoveshchenskii YN, Meshalkina YL (2017) Use of Empirical Bayesian Kriging for Revealing Heterogeneities in the Distribution of Organic Carbon on Agricultural Lands. SOIL Chem 50:305–311. doi: 10.1134/S1064229317030103

Samui P (2008) Slope stability analysis: A support vector machine approach. Environ Geol 56:255–267. doi: 10.1007/s00254-007-1161-4

Santos CAG, Freire PKMM, da Silva RM, Akrami SA (2019) Hybrid wavelet neural network approach for daily inflow forecasting using Tropical Rainfall Measuring Mission data. J Hydrol Eng 24:1–13. doi: 10.1061/(ASCE) HE.1943-5584.0001725

Sedrati N (2011) Origines et caractéristiques physico-chimiques des eaux de la wilaya de Biskra-Sud-Est-Algérien. Badji Mokhtar-Annaba University

Şen Z (2016) Spatial Modeling Principles in Earth Sciences. Springer International Publishing, Switzerland

Shabbir R, Ahmad SS (2015) Use of Geographic Information System and Water Quality Index to Assess Groundwater Quality in Rawalpindi and Islamabad. Arab J Sci Eng 40:2033–2047. doi: 10.1007/s13369-015-1697-7

Shahabi M, Jafarzadeh AA, Neyshabouri MR, et al (2016) Spatial modeling of soil salinity using multiple linear regression , Ordinary kriging and artificial neural network methods Spatial modeling of soil salinity using multiple linear regression , ordinary kriging and artificial neural network methods. Arch Agron SOIL Sci. doi: 10.1080/03650340.2016.1193162

Shahid SA, Abdelfattah MA, Omar SAS, et al (2010) Mapping and Monitoring of Soil Salinization Remote Sensing , GIS , Modeling , Electromagnetic Induction and Conventional Methods – Case Studies. In: Proceedings of the International Conference on Soils and

Groundwater Salinization in Arid Countries. pp 59–97

Shaji E, Gómez-Alday JJ, Hussein S, et al (2018) Salinization and Deterioration of Groundwater Quality by Nitrate and Fluoride in the Chittur Block, Palakkad, Kerala. J Geol Soc India 92:337–345. doi: 10.1007/s12594-018-1017-4

Shanmuganathan S, Samarasinghe S (2016) Artificial Neural Network Modelling. Springer Nature, Switzerland

Shen Q, Wang Y, Wang X, et al (2019) Comparing interpolation methods to predict soil total phosphorus in the Mollisol area of Northeast China. Catena 174:59–72. doi: 10.1016/j.catena.2018.10.052

Shrivastava P, Kumar R (2015) Soil salinity: A serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation. Saudi J Biol Sci 22:123–131. doi: 10.1016/j.sjbs.2014.12.001

Singh SK, Srivastava PK, Singh D, et al (2014) Modeling groundwater quality over a humid subtropical region using numerical indices, earth observation datasets, and X-ray diffraction technique: a case study of Allahabad district, India. Environ Geochem Health 37:157–180. doi: 10.1007/s10653-014-9638-z

Soumya SB, Kamble RB (2017) Precipitation controlled spatial variations in groundwater quality indices-suitability for drinking and irrigation purposes in the basalts of South India. Arab J Geosci 10:171. doi: 10.1007/s12517-017-2887-2

Sreedhar Y, Nagaraju A (2017) Groundwater quality around Tummalapalle area, Cuddapah District, Andhra Pradesh, India. Appl Water Sci 7:4077–4089. doi: 10.1007/s13201-017-0564-y

Srinivas Y, Aghil TB, Hudson Oliver D, et al (2017) Hydrochemical characteristics and quality assessment of groundwater along the Manavalakurichi coast, Tamil Nadu, India. Appl Water Sci 7:1429–1438. doi: 10.1007/s13201-015-0325-8

Sun Z, Song X, Bu H, et al (2016) Origin of groundwater salinity and hydrochemical processes in an unconfined aquifer: case of Yang-Dai River basin in Qinhuangdao (China). Environ Earth Sci 75:1–12. doi: 10.1007/s12665-015-4825-5

Świetlik D, Bandurski T, Lass P (2004) Artificial neural networks in nuclear medicine. Nucl Med Rev 7:58–67

Táany RA, Tahboub AB, Saffarini GA (2009) Geostatistical analysis of spatiotemporal variability of groundwater level fluctuations in Amman-Zarqa basin, Jordan: A case study. Environ Geol 57:525–535. doi: 10.1007/s00254-008-1322-0

Tirkey P, Bhattacharya T, Chakraborty S, Baraik S (2017) Assessment of groundwater quality and associated health risks: A case study of Ranchi city, Jharkhand, India. Groundw Sustain Dev 5:85–100. doi: 10.1016/j.gsd.2017.05.002

Triantafilis J, Odeh IOA, Warr B, Ahmed MF (2004) Mapping of salinity risk in the lower Namoi valley using non-linear kriging methods. Agric Manag 69:203–229. doi: 10.1016/j.agwat.2004.02.010

USDA (2011) Assessing Water Quality for Human Consumption, Agriculture, and Aquatic Life Uses Tom Pick, Water Quality Specialist. 1–31

USSLS (1954) Diagnosis and Improvement of Saline and Alkali soils. Washington

Vapnik VN (1995) The nature of statistical learning theory. Springer, New York

Vasanthavigar M, Srinivasamoorthy K, Vijayaragavan K, et al (2010) Application of water quality index for groundwater quality assessment: Thirumanimuttar sub-basin, Tamilnadu, India. Environ Monit Assess 171:595–609. doi: 10.1007/s10661-009-1302-1

Vladimir vapnik (1998) The Support Vector Method of Function Estimation. In: Nonlinear Modeling : ADVANCED BLACK-BOX TECHNIQUES. Springer Science+Business, Boston, pp 55–85

Wackernagel H (2003) Multivariate Geostatistics: An Introduction with Applications, Third. Springer Berlin Heidelberg

Wang K, Zhang C, Li W (2013) Predictive mapping of soil total nitrogen at a regional scale: A comparison between geographically weighted regression and cokriging. Appl Geogr 42:73–85. doi: 10.1016/j.apgeog.2013.04.002

Webster R, Oliver MA (2007) Geostatistics for Environmental Scientists. John Wiley & Sons, Southern Gate

Were K, Bui DT, Dick ØB, Singh BR (2015) A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecol Indic 52:394–403. doi:

10.1016/j.ecolind.2014.12.028

WHO (2011) Guidelines for drinking water quality, 4th edn. Geneva

Wilcox, L.V. (1948) Classification and Use of Irrigation Waters. U.S. Department of Agriculture, Washington DC, 962.

Wu YX, Guo L, Li Y, et al (2006) Multi-layer support vector machine and its application. In: Proceedings of the Fifth International Conference on Machine Learning and Cybernetics. pp 3627–3631

Xie Y, Chen TB, Lei M, et al (2011) Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: Accuracy and uncertainty analysis. Chemosphere 82:468–476. doi: 10.1016/j.chemosphere.2010.09.053

Yang X-S (2019) Support vector machine and regression. In: Introduction to Algorithms for DataMining and Machine Learning. pp 129–138

Yin CL, Meng F, Yu QR (2020) Infrared Physics & Technology Calculation of land surface emissivity and retrieval of land surface temperature based on a spectral mixing model. Infrared Phys Technol 108:103333. doi: 10.1016/j.infrared.2020.103333

Zaidi FK, Al-Bassam AM, Kassem OMK, et al (2017) Factors influencing the major ion chemistry in the Tihama coastal plain of southern Saudi Arabia: evidences from hydrochemical facies analyses and ionic relationships. Environ Earth Sci 76:472. doi: 10.1007/s12665-017-6817-0

Zaman M, Shahid SA, Heng L (2018) Guideline for Salinity Assessment , Mitigation and Adaptation Using Nuclear and Related Techniques. Springer Nature Switzerland AG, Switzerland

Zereg S, Boudoukha A, Benaabidate L (2018) Impacts of natural conditions and anthropogenic activities on groundwater quality in Tebessa plain, Algeria. Sustain Environ Res 28:340–349. doi: 10.1016/j.serj.2018.05.003

Zhou T, Geng Y, Chen J, et al (2020) High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. Sci Total Environ 729:138244. doi: 10.1016/j.scitotenv.2020.138244

Zounemat-kermani M, Kisi O, Rajaee T (2014) Performance of radial basis and LM-feed

forward artificial neural networks for predicting daily watershed runoff. Appl Soft Comput J 1–12. doi: 10.1016/j.asoc.2013.07.007

Zounemat-kermani M, Scholz M (2014) Modeling of Dissolved Oxygen Applying Stepwise Regression and a Template-Based Fuzzy Logic System. J Environ Eng 140:69–76. doi: 10.1061/(ASCE)EE.1943-7870.0000780.

Zounemat-kermani M, Ramezani-charmahineh A, Razavi R, et al (2020) Machine Learning and Water Economy : a New Approach to Predicting Dams Water Sales Revenue. Water Resour Manag 34:1893–1911. doi: https://doi.org/10.1007/s11269-020-02529-0