PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

LRS EAU
ENP - ALGERIE

NATIONAL POLYTECHNIC SCHOOL

HYDRAULIC DEPARTMENT

WATER SCIENCE RESEARCH LABORATORY

FINAL PROJECT THESIS

PRESENTED WITH A VIEW TO OBTAIN THE DIPLOMA OF STATE ENGINEER IN HYDRAULICS

Mapping groundwater vulnerability to nitrate contamination using Machine Learning techniques

---

MR BENAROUSSI OUASSIM AND MRS DJELLAL MEROUA

PRESENTED AND PUBLICLY DEFENDED ON JULY 03, 2022

JURY COMPOSITION :

President	Mrs C. TCHEKIKEN	MCB	ENP Algiers
promoter	Mr S.E. TACHI	MCA	ENP Algiers
Co-promoter	Mrs M. CHETIBI	MCB	ENP Algiers
Examiner	Mrs S.CHARIFI	MCB	ENSH Blida
Examiner	Mr A. AFROUKH	PHd student	ENP Algiers
Guest	Mrs H.BOUANANI	Doctor	ENP Algiers
Guest	Mr R.MESSAHLI	MAA	ENP Algiers

ENP 2022

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

LRS EAU
ENP - ALGERIE

ÉCOLE NATIONALE POLYTECHNIQUE
DÉPARTEMENT HYDRAULIQUE
LABORATOIRE DE RECHERCHE SCIENCE DE L'EAU

MÉMOIRE DE PROJET DE FIN ÉTUDES
POUR L'OBTENTION DU DIPLÔME D'INGÉNIEUR D'ÉTAT EN HYDRAULIQUE

Cartographie de la vulnérabilité des eaux souterraines à la contamination par les nitrates à l'aide de techniques d'apprentissage automatique

---

MR BENAROUSSI OUASSIM ET MME DJELLAL MEROUA

PRÉSENTÉ ET SOUTENU PUBLIQUEMENT LE 03 JUILLET 2022

COMPOSITION DU JURY :

Présidente	Mme C. TCHEKIKEN	MCB	ENP Alger
Promoteur	Mr S.E. TACHI	MCA	ENP Alger
Co-Promoteur	Mme M. CHETIBI	MCB	ENP Alger
Examinateur	Mme S.CHARIFI	MCB	ENSH Blida
Examinateur	Mr A. AFROUKH	Doctorant	ENP Alger
Invité	Mme H.BOUANANI	Docteur	ENP Alger
Invité	Mr R.MESSAHLI	MAA	ENP Alger

ENP 2022

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

LRS EAU
ENP - ALGERIE

NATIONAL POLYTECHNIC SCHOOL

HYDRAULIC DEPARTMENT

WATER SCIENCE RESEARCH LABORATORY

FINAL PROJECT THESIS

PRESENTED WITH A VIEW TO OBTAIN THE DIPLOMA OF STATE ENGINEER IN HYDRAULICS

Mapping groundwater vulnerability to nitrate contamination using Machine Learning techniques

---

MR BENAROUSSI OUASSIM AND MRS DJELLAL MEROUA

PRESENTED AND PUBLICLY DEFENDED ON JULY 03, 2022

JURY COMPOSITION :

President	Mrs C. TCHEKIKEN	MCB	ENP Algiers
promoter	Mr S.E. TACHI	MCA	ENP Algiers
Co-promoter	Mrs M. CHETIBI	MCB	ENP Algiers
Examiner	Mrs S.CHARIFI	MCB	ENSH Blida
Examiner	Mr A. AFROUKH	PHd student	ENP Algiers
Guest	Mrs H.BOUANANI	Doctor	ENP Algiers
Guest	Mr R.MESSAHLI	MAA	ENP Algiers

ENP 2022

ملخص

لقد اصبح تقييم قابلية تعرض المياه الجوفية للتلوث في الاقليم الشرقي لمتيجا ضروري من اجل مراقبة الموارد المائية والحفاظ عليها. تهدف هذه الدراسة إلى نمذجة قابلية تعرض المياه الجوفية للتلوث بسبب تواجد عنصر النيترات استنادا إلى الحد الأقصى للتركيز المقبول في مياه الشرب (٥٠ ملم / لتر) باستخدام ١٠ عوامل مؤثرة، وهي هطول الأمطار ، طبقة الفادوز ، عمق المياه الجوفية ، الميل ، النفاذية ، البعد عن النهر، كثافة الصرف، استخدام الأراضي ، مؤشر الغطاء النباتي و المؤشر الطبوغرافي. تم تقسيم مجموعة البيانات عشوائيا بين التدريب (٧٠ %) والتحقق (٣٠ %). قارنا بين نتائج نماذج التعلم الآلي "$RandomForest$" و "$AdaBoost$" ، استنادا إلى (ROC) ، فإن المساحة تحت المنحنى (AUC) تساوي ٨٦ % و ٩٤ % على التوالي. بالإضافة إلى ذلك ، كشف كلا النموذجين أن هطول الأمطار والنفاذية وعمق للمياه الجوفية هي العوامل الرئيسية التي تحدد تعرض المياه الجوفية للنترات (NO_3) في الاقليم الشرقي لمتيجا، كما تنبأ بمؤشرات لكل العوامل بناء على أهميتها. ونتيجة لذلك تم تحقيق خريطة لتحديد المناطق المعرضة لتلوث المياه الجوفية .

الكلمات المفتاحية: قابلية تعرض المياه الجوفية للتلوث، الاقليم الشرقي لمتيجا، نترات، $AdaBoost$ ، $RandomForest$.

Résumé

L'évaluation de la vulnérabilité des eaux souterraines à la contamination dans l'aquifère de Mitidja orientale est devenue très importante pour le contrôle et la préservation des ressources en eau. Cette étude vise à modéliser la vulnérabilité spatiale des eaux souterraines aux nitrates sur la base de la concentration maximale acceptable dans l'eau potable (50 mg/L) en utilisant 10 paramètres d'influence, qui sont les précipitations, la zone non saturée, la profondeur des eaux souterraines, la pente, la perméabilité, la distance à la rivière, la densité de drainage, l'occupation du sol, NDVI et TWI. Les données ont été divisé aléatoirement entre la formation (70%) et la validation (30%). Nous avons comparé les résultats des modèles d'apprentissage automatique Random Forest et AdaBoost. Sur la base de la courbe ROC (Receiver Operating Characteristic), l'aire sous la courbe (AUC) est respectivement de 86% et 94 %. En outre, les deux modèles ont révélé que les précipitations, la perméabilité et la profondeur des eaux souterraines sont les principaux facteurs déterminant la vulnérabilité des eaux souterraines aux nitrates (NO_3^-) dans la Mitidja orientale et ils ont également prédit des indices pour chaque paramètre en fonction de leur importance. En conséquence, la carte de vulnérabilité des eaux souterraines a été élaborée.

Mots-clés : Vulnérabilité des eaux souterraines, Mitidja orientale, Nitrate, AdaBoost, Random Forest.

abstract

The evaluation of groundwater vulnerability to contamination in the eastern Mitidja aquifer has become very important for water resources control and preservation. This study aims to model the spatial groundwater vulnerability to nitrate based on the maximum acceptable concentration in drinking water (50 mg/L) by using 10 influencing parameters, which are rainfall, vadose zone, depth to groundwater, slope, permeability, distance to river, drainage density, land use, NDVI and TWI. The dataset was randomly divided between training (70%) and validation (30%). We compared between the results of Random Forest and AdaBoost machine learning models, based on the Receiver Operating Characteristic (ROC) curve, Area Under Curve (AUC) equals 86% and 94%, respectively. In addition, both ML models revealed that rainfall, permeability, and depth to groundwater are the key factors determining groundwater vulnerability to nitrate (NO_3) in the eastern Mitidja and it also predicted indexes for each parameter based on their importance. As a result, the groundwater vulnerability map was elaborated.

Keywords: Groundwater vulnerability, eastern Mitidja, Nitrate, AdaBoost, Random Forest.

Acknowledgement

We would like to express our sincere gratitude to our promoter Mr S.E Tachi and co-promoter Mrs M.Chetibi for their contribution, patience and motivation through all the stages of our project thesis. Besides our advisors, we would like to thank Dr Hakim Belaroui and Dr Hamza Bouguerra for sharing their brilliant suggestions and helpful advices with us. We are also grateful to Mr Alliche Abdelmalek and Mrs H.Bouanani for their availability and help. Special thanks to all the people who contributed to the success of our thesis.

We would also like to thank the members of the jury for the honor they gave us by accepting to evaluate our work.

Finally, we acknowledge our warmest thanks to our parents and family members who always supported and encouraged us.

Contents

List of Figures

List of Tables

General introduction

Groundwater is a valuable resource. It is stored beneath the earth's surface and is the most available source for drinking water, industry and irrigated agriculture in dry and semi-arid areas (Nampak, Pradhan, and Abd Manap, 2014).
Every day, thousands of people die from diseases associated with inadequate supplies of clean water (Price and Back, 2013).

These waters are often exposed to contamination by several pollutants with different nature such as biological pollutants: bacteria, viruses, chemical and organic (chlorides, nitrates, heavy metals, pesticides) or physical pollutants: radioactivity through the soil to the unsaturated zone.

Groundwater management has become a major problematic around the world. This may be related to climate change, rapid population increase and overuse of groundwater for irrigation. Therefore, in order to ensure long-term management of groundwater resources, a thorough assessment of these resources at the local level is essential (Hasiniaina, Zhou, and Guoyi, 2010).

Nitrate (NO_3^-) is the most abundant pollutant in groundwater, while Nitrogen in his various forms of nitrate, nitrite, or ammonium in groundwater is a needed nutrient for enhancing the crop yield and plant growth, but still known as a prevalent inorganic contaminant that plays a vital role in groundwater quality. Indeed,NO_3^- concentrations increase when agricultural activity intensifies, due to the abusive use of nitrogen fertilizers (Ki et al., 2015).
Consequently, consumption of water polluted with NO_3^- may be associated with human health toxicity when it converts to nitrite; infants are at greater risk due to their immature digestive systems.

The groundwater of the Mitidja aquifer supply drinking water to many localities of the capital and several cities of the four wilayas located in large parts in the plain (Algiers, Blida, Boumerdes, and Tipaza). They also ensure the irrigation of tens of thousands of hectares of agricultural land and the

supply of almost all industrial units. This is why a particular interest in terms of qualitative and quantitative monitoring is given to this water table, which has been the subject of many hydrogeological studies.
The Mitidja is a plain situated in northern Algeria. It consists of aquifers that are the main source of drinking water for the whole center part of the country. Many studies of this aquifer have shown that the chemical compositions of groundwater are the result of different components in relation to the type of geological reservoir and anthropogenic factors.

To the best of our knowledge, no previous studies have assessed the vulnerability of groundwater using machine learning in Algeria.
This study proposes a novel method of mapping groundwater vulnerability using Machine Learning algorithms. Adaptive Boosting and Random Forest methods have been used to predict NO_3^- concentration in groundwater to assess the groundwater vulnerability in Algeria (Mitidja East).

The main objective of this study was to use two accurate models (Adaboost and RF) to assess the specifc groundwater vulnerability to NO_3^- of the aquifer of the Mitidja East basin, using ten (10) parameters that may contribute to NO_3^- pollution.
The results of this study will aid in:

• Determining the most important factors that control the groundwater vulnerability to NO_3^- pollution of the aquifer of the Mitidja East basin.

• Identifying the most vulnerable areas to NO_3^- pollution in the Mitidja East basin according to the two ML models (Adaboost and RF).

Chapter 1

Groundwater vulnerability to contamination

1.1 Introduction

The degradation of groundwater quality in several parts of the world calls for an urgent act in order to ensure sustainable utilization and protection of available groundwater resources. Groundwater vulnerability assessment gets complicated by considering the effects of land use and climate change, but it is still essential for effective planning, decision making, and sustainable management of these resources. The various methods for aquifer vulnerability assessment are grouped into three categories, namely: overlay and index methods (qualitative methods); process based methods (quantitative methods) and statistical methods.

1.2 Concept of Aquifer Vulnerability

The notion of groundwater vulnerability was first established in France between the 1960s and 1970s to find a solution to groundwater contamination. The statisticians defined it as the probability of pollutants percolating and diffusing from the ground surface into the groundwater system. The vulnerability is determined by a number of factors depending on the nature, concentration, and transfer time of contaminants. It is assessed by taking into consideration the soil's properties, the saturated zone, and the unsaturated zone (Machiwal et al., 2018).

1.3 Vulnerability types

There are two types of aquifer vulnerability (Ibe, Nwankwor, and Onyekuru, 2001): The first type is intrinsic vulnerability, which is the vulnerability of groundwater to contaminants caused by human activities such as the release of industrial waste, incineration of fossil fuels, particularly coal, and the utilization of as-loaded water for irrigation. It is evaluated based on the geological, hydrological, and hydrogeological characteristics of an area. In any case, the latter is considered static and invariable. On the other hand, the second type is called the "specific vulnerability," which is the groundwater's susceptibility to certain pollutants; it takes into account the characteristics of the pollutant and the intrinsic properties of the basin (Ribeiro, Pindo, and Dominguez-Granda, 2017). It is thought that the specific vulnerability is dynamic and closer to reality, compared to the first type.

1.4 Groundwater vulnerability assessment methods

Groundwater vulnerability assessment methods can be grouped into three major groups:

1.4.1 Process based methods (quantitative method)

The process methods prioritize the protection of both source and resource by including various physical, chemical, and biological parameters in order to consider the transport and the fate of contaminants at both spatial and temporal scales, and in either saturated or unsaturated zones (Ju et al., 2018).

1.4.2 Statistical methods

They are based on a variable that is affected by the concentration of the pollutant or the risk of contamination. These approaches combine data on the distribution of pollutants in the study region and provide information on the probability of pollution.

1.4.3 Overlay and index methods (Qualitative methods)

Those methods have been used in groundwater vulnerability studies since the 1990s. The big advantage of GIS technology is that it can easily accomplish overlay and index operations in the spatial domain. Those methods were grouped into two categories:

Methods for Porous-Media Aquifers

Porous media aquifers are groundwater layers formed from sand, gravel, or even large boulders, which allow groundwater to move through pore spaces between the individual grains. For this type of aquifers, There are several methods for assessing the vulnerability of groundwater to contamination; we will discuss the most popular ones:

The Drastic Method: The Drastic method was first developed in 1985 by the United States Environmental Protection Agency (Aller, 1985), based on calculating a groundwater vulnerability index by combining seven hydrogeological weighted parameters, comprising: Depth to Water (D), Net Recharge (R), Aquifer Media (A), Soil Media (S), Topography (T), Impact of Vadose Zone (I), and Hydraulic Conductivity (C), represented by the following equation:

$$DASTIC_{index} = \sum_{j=1}^{7} W_i.R_i \tag{1.1}$$

Where R_i and W_i are, respectively, the rating and the weight of the i^{th} parameter.
Each of these parameters is assigned a rate and a weight according to their relative importance. The required data for various parameters is used to perform GIS processing and prepare the vulnerability map. The resultant DRASTIC index values represent the state of the aquifer hydrogeological vulnerability, which are regrouped into four classes according to the following table:

Table 1.1: The criteria for vulnerability assessment by the Drastic method

Degree of vulnerability	very high	High	average	Low
Vulnerability index	$>$200	141-200	101-140	$<$101

Several researchers have updated the original DRASTIC method by adjusting rating ranges, relative weights, and vulnerability classes, as well as adding and/or ignoring some of its parameters (Singh et al., 2015; Kumar and Pramod Krishna, 2020).

The GOD Method: This method was developed in 1987 (Foster and Hirata, 1988), and it considers three parameters, namely groundwater occurrence (G), overall lithology (O), and groundwater depth

(D) for vulnerability assessment. It is very suitable for large spatial scales, even though studies dealing with this method are limited. The GOD index is evaluated using the following equation:

$$GOD_{index} = G \times O \times D \tag{1.2}$$

Its results can be grouped into five classes, as shown by the following table:

Table 1.2: The criteria for vulnerability assessment by the GOD method

Degree of vulnerability	very high	High	average	Low	very low
Vulnerability index	0.7-1	0.5-0.7	0.3-0.5	0.1-0.3	0-0.1

The AVI Method: The Aquifer Vulnerability Index method was developed and applied in Canada in 1993, and it depends on two physical parameters, i.e., thickness (d) and hydraulic conductivity (K) of every sedimentary unit above the uppermost aquifer, to determine soil resistance to vertical flow and evaluate the vulnerability of groundwater by using the following equation:

$$AVI_{index} = \sum_{i=1}^{n} \frac{d_i}{k_i} \tag{1.3}$$

Its results can be grouped into five classes, as shown by the following table:

Table 1.3: The criteria for vulnerability assessment by the AVI method

Degree of vulnerability	very high	High	average	Low	very low
Vulnerability index	>10000	1000-10000	100-1000	10-100	0-10

The SINTACS Method: This method was developed in Italy in the early 1990s in order to adapt the mapping to a larger scale in view of the great diversity of hydrogeological Italy, it uses the same seven parameters as the Drastic method, and as the acronym SINTACS comes from the Italian names of the used factors; **S**oggicenza (water table depth); **I**nfiltrazione (net recharge); **N**on saturo (vadose zone); **T**ipologia della copertura (soil media); **A**cquifero (aquifer media); **C**onducibilità (hydraulic conductivity); **S**uperficie topografica (topographic slope), where each of these parameters is assigned

a rate and a weight according to their relative importance (Corniello, Ducci, and Monti, 2004). The SINTACS Index is calculated using the equation:

$$SINTACS_{index} = \sum_{i=1}^{7}(P_i \times W_i) \tag{1.4}$$

P is the rating of each parameter, and W is the relative weight. The final vulnerability index score was divided into five vulnerability classes, represented in the following table:

Table 1.4: The criteria for vulnerability assessment by the SINTACS method

Degree of vulnerability	very high	High	average	Low
Vulnerability index	>200	186-200	106-186	<106

Methods for Karst Aquifers

Limestone and dolomite are soluble rocks. Their dissolution creates distinctive lanforms such as springs, caves, and sinkholes characterized by a unique hydrogeology that is highly productive but extremely vulnerable to contamination. For this type of aquifers, There are several methods for assessing the vulnerability of groundwater to contamination; we will discuss the most popular ones:

The EPIC method: After several tests in Switzerland, the EPIC method was found to be the ultimate solution (Doerfliger, Jeannin, and Zwahlen, 1999) for quantifying the vulnerability of groundwater contamination of the Krast aquifers. The four influencing flow and transport parametes taken into account by the method are epikarst (E), protective cover (P), infiltration condition (I), and karst network development (K), based on the geological, geomorphological, and hydrogeological characteristics. This method shows a consistent results and flexible fast applications to determine the vulnerability of different tested areas such as Supramonte karstic system in north-central Sardinia, Italy and karstic aquifer of southern Belgium.

The GLA method: GLA is an abbreviation of Geologisches Landesamt which is a method developed in Germany (Hölting et al., 1995) All of natural processes in protective cover are responsible for reducing contaminant concentration from the study area and its effectiveness is still mostly influenced the by travel time. The protective cover comprises all the strata between the ground surface and the

groundwater table: soil, subsoil, and unsaturated bedrock; its effectiveness is mostly influenced by the thickness and the formation properties of each stratum. The total protective function is obtained as follows:

- The first step is to determine the protective function of each stratum and multiply it by the thickness of that stratum.
- Secondly, the resulting values are added and multiplied by a factor reflecting the amount of recharge.
- The final value of PTS represents the degree of protection:

PTS $\leq$ 500 indicates a very low degree of natural protection.
PTS $>$ 4000 indicates a very high degree of protection.

The PI method: The PI method is a GIS-based approach for mapping the groundwater vulnerability to contamination (Goldscheider et al., 2000). It is applicable to all kinds of aquifer systems, especially karst aquifers, and is based on a source-pathway-target model. The PI vulnerability map can be used for resource protection for all the surface limited between the ground surface and the groundwater table.

The abbreviation PI stands for the two parameters:

- **Protective cover (P) :** The protective cover (p) is primarily determined by the thickness and hydraulic conductivity of all the strata between the ground surface and the groundwater table.

  This factor is calculated using a slightly modified version of the GLA method and is classified into five classes: P = 1 represents an extremely poor level of protection, P = 5 suggests a highly effective protective cover.

- **Infiltration Conditions (I) :** In the karst areas, it is possible for a permanent surface stream to disappear into a swallow hole.

  In this case, the protective cover is completely bypassed at the swallow hole and bypassed in part by the surface runoff in the catchment area of the sinking stream.

  The I factor describes the degree to which the protective cover is bypassed as a result of surface and near-surface flow concentration, particularly within the catchment zone of a sinking stream. When it's equal to 1, it indicates that the protective cover is completely effective and is not bypassed. The I factor is 0 when the protective cover is fully bypassed by a swallow hole through which surface water enters the karst aquifer directly.

1.5 Conclusion

This chapter provided an understanding of the concept and types of groundwater vulnerability to contamination. It developed the most popular methods for predicting groundwater vulnerability. These methods have several advantages and disadvantages depending on their complexity and type of application.

Process-based methods are pure mathematical models that use equations to represent the behavior of substances in the subsurface environment, and they are intended to assess the specific vulnerability. Intrinsic vulnerability is assessed using both overlay and index methods that integrate the physical elements that determine vulnerability into a weighted index or numerical score, and statistical methods that establish associations with locations of known contamination (Council et al., 1993).

Chapter 2

Groundwater contamination

2.1 Introduction

Pollution is a difficult problem in different countries around the world, many of which lack standards to keep pollution within reasonable limits, and others lack the means to enforce water quality standards (Li et al., 2021). Most natural groundwater contains some dissolved substances. They pose no risk to human health or the ecosystem when the concentrations are low. Otherwise, it would be referred to as contamination.

2.2 Groundwater contamination sources

Groundwater contamination is a change in the quality and character of water that makes its usage harmful and disturbs the aquatic ecosystem. Its sources are classified into two major categories: (Li et al., 2021)

Natural sources: such as seawater, low-quality surface water, and mineral deposits. This may become more harmful due to human activity.

Anthropogenic sources: Human activity may lead to the infiltration of undesired substances such as (Muralikrishna and Manickam, 2017):

- Urban discharges resulting from sewage systems of households, public spaces, and companies, as well as rainwater runoff in urban areas or malfunctions in the collection of urban wastewater (poor connections by private individuals, leaks, and saturation of the networks).

- Agricultural discharges: chemical compound leaching (fertilizers, pesticides, etc.) through rainwater percolation.
- Industrial waste of extremely variable composition since they are likely to receive the residues or the losses during the entire manufacturing process (Water from mine drainage, iron and steel industries, chemical industries. . .).

Groundwater contamination sources can also be grouped according to their extent:

- **Point sources:** classified as accidental because they originate from leaks in pipelines, reservoirs, or the infiltration of leachates from a solid deposit; they occur from a defined surface that is small compared to the catchment area of groundwater.
- **Diffuse (non-point) sources:** Diffuse pollution occurs when solid or liquid products are dispersed across a vast area on the soil's surface.

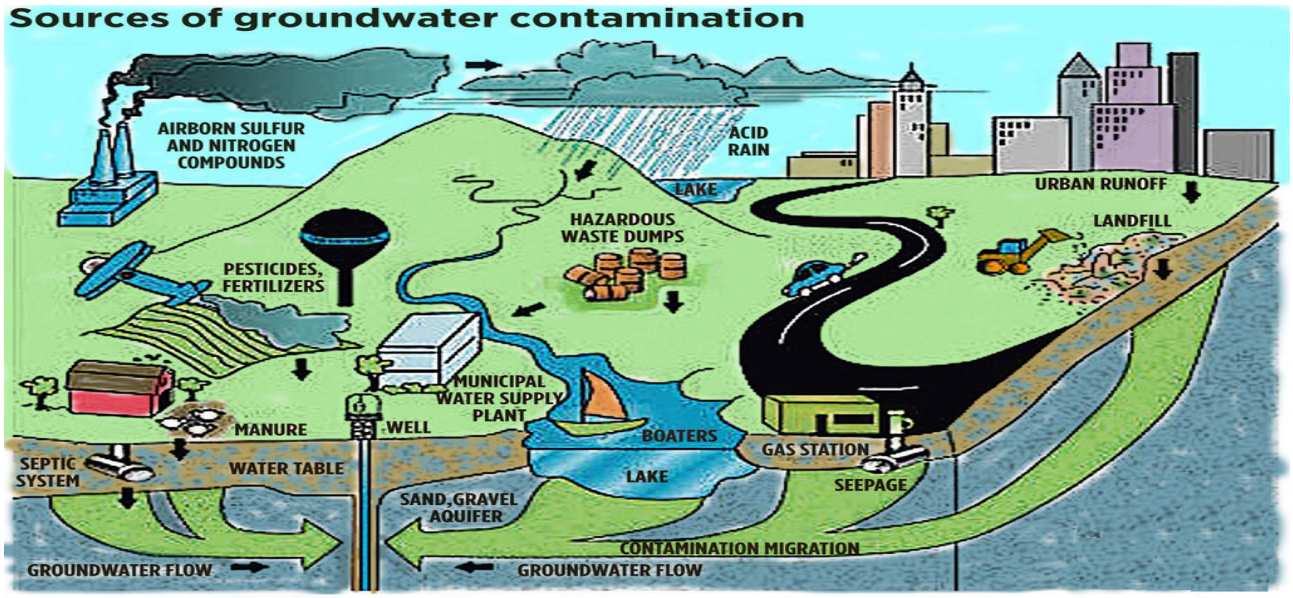

Figure 2.1: Sources of groundwater contamination. Source: California State Water Resources Control Board website

2.3 Contamination types

Contaminants can be grouped in three major types (Sajil Kumar, 2020):

Biological contaminants: This type includes microbial organisms, such as bacteria, viruses, and protozoa. Where more than 400 different bacteria and 100 viruses have been identified.

Radioactive contaminants: including nuclear power plant waste, nuclear weapons testing, and incorrect medical radioisotope disposal.

Chemical contaminants: Chemical contamination has been particularly a frequent subject in groundwater studies due to its significant threat to the human population's health and its challenging and costly remediation once polluted.

2.4 Transport of contaminants

After being adsorbed onto the surface of aquifer solids, the movement of contaminants may differ from moving slowly or very fast while migrating freely with the flowing pore water. They end up crossing many kilometers from their source in a short period of time, while many chemical reactions occur along the way, which causes a contaminant to disappear, or appear from nowhere (Mackay, Roberts, and Cherry, 1985). The spread of contaminants in groundwater is affected by the following processes:

Advection: Advection represents the movement of dissolved contaminants with flowing groundwater that moves at different rates in each stratum, depending on groundwater quality, quantity, and the arrangement of particles in the subsurface.

Diffusion: Diffusion is a process based on the contaminant concentration gradient, which means that even if the fluid is static, the dissolved contaminant spreads from areas of higher to lower concentrations.

Dispersion: Dispersion occurs when two miscible fluids are brought into contact. There is a sharp interface at the beginning which vanishes into a transition zone as the differences between physical properties (concentration) tend to be leveled with time.

2.5 Groundwater contamination with Nitrate

The soil fertility of the Mitidja plain has made it one of Algeria's most important agricultural areas, with a massive load of fertilizers, leading to nitrate contamination. Nitrogen in its various forms of nitrate, nitrite, or ammonium in groundwater is a nutrient needed for enhancing crop yield and plant

growth, but is still known as a prevalent inorganic contaminant that plays a vital role in groundwater quality.

2.6 Nitrogen cycle in groundwater

Several natural processes are responsible for the transformation of nitrogen in the atmosphere or in the soil and its bio-disponibility. These processes are: fixation, nitrification, and denitrification (Canter, 2019; Filion, 2017).

- **Fixation:** It represents the process of converting atmospheric nitrogen (N_2) into ammonium (NH_4^+) and ammoniac gas (NH_3), usable nitrogen for plants and animals. This process occurs in anaerobic conditions by cyanobacteria. The typical chemical reaction is:

$$2\mathrm{N_2(g)} + 3(\mathrm{CH_2O}) + 3\mathrm{H_2O} \longrightarrow 4\mathrm{NH_4^+} + 3\mathrm{C0_2}$$

  In soils with high pH, ammonium is transformed into ammonia gas:

$$\mathrm{NH_4^+} + \mathrm{OH^-} \longrightarrow \mathrm{NH_3(g)} + \mathrm{H_2O}$$

  It's based on a reaction of reduction that takes place via organic substances noted as CH 2 O in the first equation.

- **Nitrification:** It is the oxidation process that coverts fixation products into nitrates (NO_3-) and nitrites (NO_2^-), this happens by enzymatic catalysis linked to bacteria in soils and water. The chain reaction is:

$$\mathrm{NH_4^+} \rightleftharpoons \mathrm{NO_2^-} \rightleftharpoons \mathrm{N0_3^-}$$

$$2\mathrm{NH_4^+} + 3\mathrm{O_2} \longrightarrow 2\mathrm{NO_2^-} + 2\mathrm{H_2O} + 4\mathrm{H^+}$$

$$2\mathrm{NO_2^-} + \mathrm{O_2} \longrightarrow 2\mathrm{NO_3^-}$$

- **Denitrification:** It returns nitrogen to the atmosphere in its molecular form, N_2, a greenhouse gas that contributes to the destruction of the ozone layer in the stratosphere.

  This reduction reaction of NO_3^- through bacteria transforming organic matter is:

$$4\mathrm{NO_3^-} + 5\mathrm{CH_2O} + 4\mathrm{H^+} \longrightarrow 2\mathrm{N_2(g)} + 5\mathrm{CO_2(g)} + 7\mathrm{H_2O}$$

Human activities disturb these processes through the use of fertilizers that add ammonia compounds (NH_4^+, NH_3) and nitrates (NO_3^-) to the soil, resulting in a nitrogen excess, that will end up in groundwater.

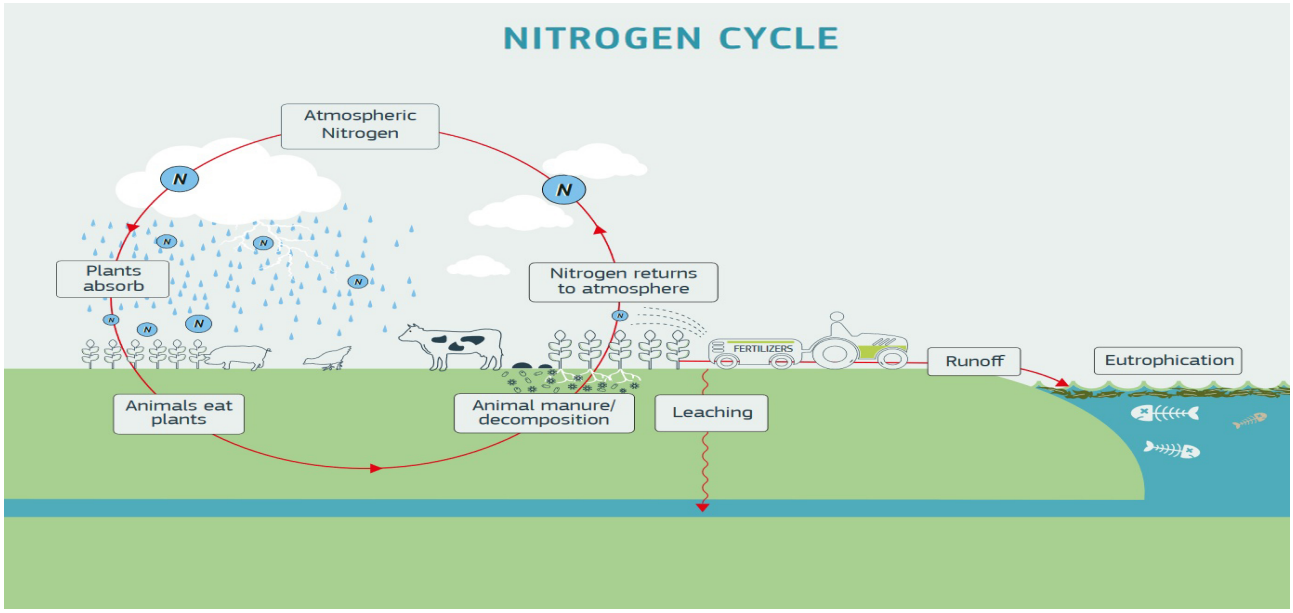

Figure 2.2: Nitrogen cycle

2.7 Nitrate effects on human health

Nitrate becomes toxic for human health when it is transformed into nitrite; infants are more exposed to this risk because of the immaturity of their digestive systems (Ji et al., 2020). Nitrites travel to the blood veins and cause the formation of methemoglobin, a form of hemoglobin unable to transport oxygen. When its concentration is elevated in red blood cells, functional anemia and tissue hypoxia may occur. In addition, nitrites can react with secondary amines resulting from alimentation to form nitrosamines (carcinogens) in the stomach. When the nitrate concentration is within the quality limit of 50 mg/l in water, this risk to infants' health is considered negligible.

2.8 Chemical analysis of nitrate

Water samples were collected from pumping boreholes after a minimum of several minutes of pumping. All the samples were stored in a temperature lower than 4 °C and later transferred to the laboratory, where nitrate was analyzed by colorimetry with a UV–visible spectrophotometer.

2.9 Reduction of water pollution by nitrates

The following two methods are used to reduce groundwater pollution by nitrates:

- Dilute the resource containing nitrates with an uncontaminated resource.
- Apply specific treatments, such as biological denitrification by using bacteria that transform nitrates into nitrogen gas, or ion exchange denitrification by using a resin that absorbs nitrates from the water passing through it, releasing chloride ions in exchange.

2.10 Algerian companies contributions to groundwater quality preservation

In May 2018, the former Algerian Minister of Environment and Renewable Energy, Fatma-Zohra Zerouati, affirmed that Algeria has adopted a work plan to preserve its natural resources. On the same day, the National Agency for Integrated Water Resources Management (AGIRE), the National Water Resources Agency (ANRH), and a private company signed two cooperation agreements to preserve water resources and fight against industrial pollution that threatens public health and the environment.

In January 2020, the Minister of Water Resources, Arezki Berraki highlighted the importance of the National Agency for Water Resources (ANRH) in the protection of water resources, insisting on its modernization through the development of an effective action plan to guarantee sufficient water resources for future generations (*Algeria press service*).

2.11 Conclusion

This chapter has shown the importance of understanding the problem of groundwater contamination through the process of leaching, transport, and cycling of contaminants. The chapter also highlighted the contamination of groundwater by nitrates (NO_3^-), whose concentrations are increasing with the

intensification of agricultural activities due to the excessive use of nitrogen fertilizers. Therefore, the consumption of water polluted by NO_3^- can have negative effects on human health, which implies that the whole population is concerned.

Chapter 3

Study area

3.1 Introduction

Our case study is the Mitidja aquifer, it has an important role in the drinking water supply of the capital and its surroundings, it ensures the irrigation of its agricultural lands and supplies many industrial units. The need to assess the vulnerability of the Eastern Mitidja groundwater began when nitrate concentrations started to exceed the potability threshold (50 mg/l).

3.2 Geographic location of Eastern Mitidja

Mitidja is an Algerian west-to-east lowland. It is oriented towards the WSW-ENE axis that stretches for approximately 90 kilometers from the west of the region of "Hadjout" to the valley of "Boudouaou" in the east, with an average width of 15 kilometers, connecting the provinces of Algiers, Blida, Tipaza, and Boumerdes. In this thesis, the study focuses on the Eastern Mitidja plain, which is in constant subsidence. The region covers over 625,6 square kilometers lying between longitudes 2°57'30"E and 3°25'00"E and latitudes 36°47'00"N and 36°30'00"N. It is bordered in the south by the peaks of the Blidean mountain chain, with altitudes going from 1200 to 1600 m. In the north, it is isolated from the sea by the Sahel, which originates from Eastern Algiers and leads to the narrow gorge of Reghaia. Its altitude varies between -3 and 262 m on a hill zone of a few kilometers wide that is delimited by the Mazafran wadi in the west and by the Boudouaou wadi in the east. Furthermore, the Mitidja East plain is divided into two sub-watersheds (EL Harrach and EL Hamiz), both of which are part of the larger watershed of the Algerian coast.

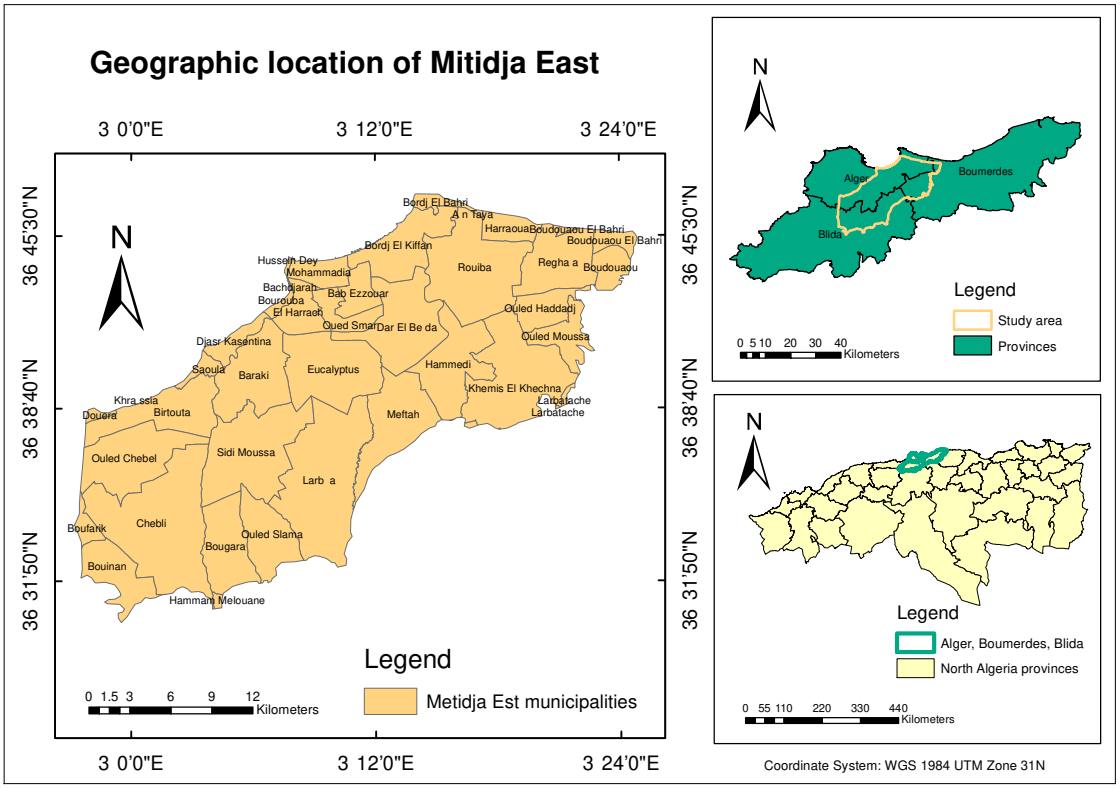

Figure 3.1: Location map of the study area

3.3 Climatology

Climatic conditions play a determining role in the surface and subsurface flow regimes. It provides data for the approach of the balancing terms, which are precipitation and temperature. Mitidja is influenced by the Mediterranean climatic regime due to its geographical location. Its plain experiences the same regional subhumid coastal climate as all coastal lowlands (Derdous et al., 2020). As one goes away from the coast, the climate becomes more continental, and temperatures significantly decrease. The winters are rainy and mild; the summers are hot and dry.

3.3.1 Precipitation

The average interannual rainfall in Mitidja varies between 500 and 900 mm. The rainfall stations characteristics and average interannual rainfall from 2014 to 2018, of 13 pluviometric stations are

shown in table3.1:

Table 3.1: Pluviometric Stations and Average interannual rainfall

Station	Code	X	Y	Av Interannual rainfall (mm)
MEURDJA	21401	539650	365200	635.03
CRESCIA	21406	525400	375300	739.55
BARAKI	21421	535050	376500	637.15
MEFTAH D BORGEA UD	21424	545950	367500	641.04
DOUAR MAKLI	21427	532450	355750	703.02
TUILERIE (Altairac)	21443	539550	379050	631.85
CHEBLI HA SERKADJI	21445	530000	360850	666.32
ARBATACHE	20606	560200	370850	673.13
ALGER Pep	20607	556500	372500	640.72
DAR EL BEIDA	20611	547350	379600	602.24
ROUIBA MONTOYO	20614	552000	382050	599.35
OULED ALI	20627	553000	371650	622.99
REGHAIA PONT R,N,05	20632	557820	382230	620.03

The pluviometric stations are mainly located in the Algerian coast watershed (02), and its two sub-watersheds (EL Harrach and EL Hamiz).

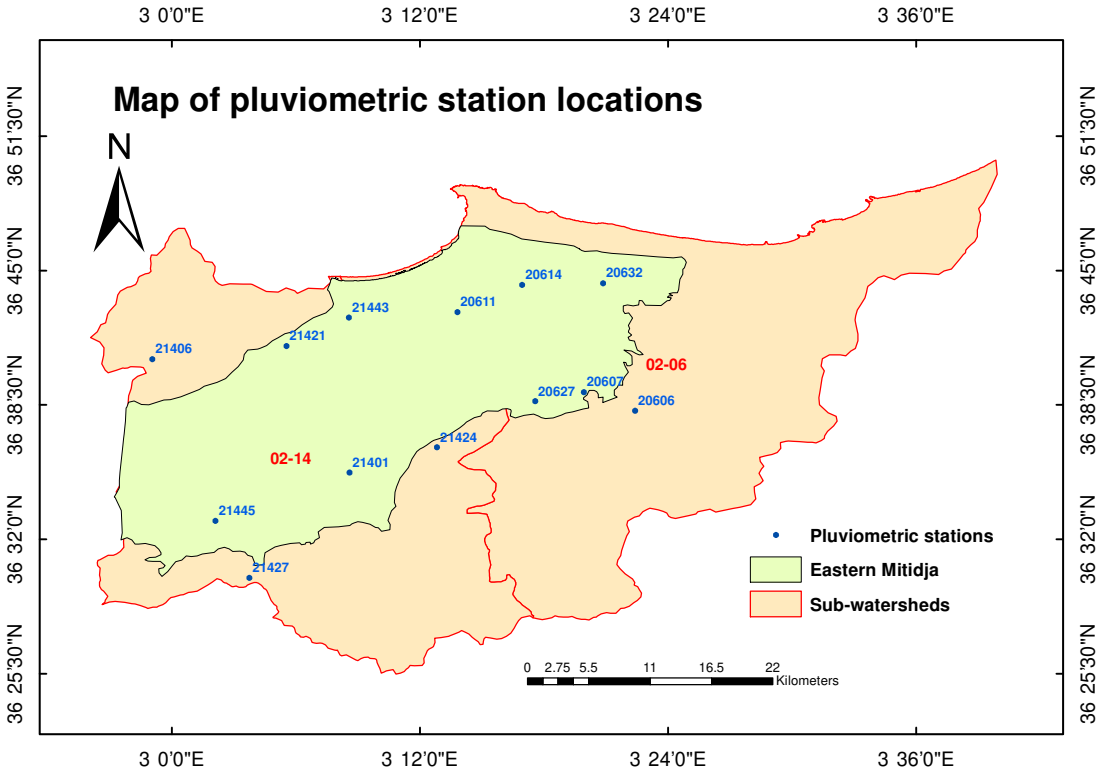

Figure 3.2: The Pluviometric stations

The collected data from ANRH had missing monthly measurements. We had to fill them by using the World Climate Website (*WorldClim*).

3.3.2 Temperature

Summers in the Mediterranean climate of the region combine heat, sunlight, and dryness, with temperatures frequently reaching 32 or 33 °C in July and August. During the winter, the climate is mild and the temperatures rarely drop to 0 °C. The table 3.2represents the annual temperature variation of the year 2018:

Table 3.2: Temperature in 2018

Temperature (°C)	Jan	Feb	Mar	Apr	Mai	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
Max	23.8	20.8	28.1	29.1	25	30.8	32.1	33.5	31.8	29.4	26.1	23.8	33.5
Med	15.1	13	16.6	18	18.7	22.5	26.6	26.8	25.2	21.4	18.5	16.1	19.9
Min	9.2	6.5	8.6	11.8	13.1	18.3	22.3	22.9	19.1	8.9	13.5	9.9	6.5

source:*infoclimat* website.

The graph 3.3 represent the variation of the temperature of our study area in 2018.

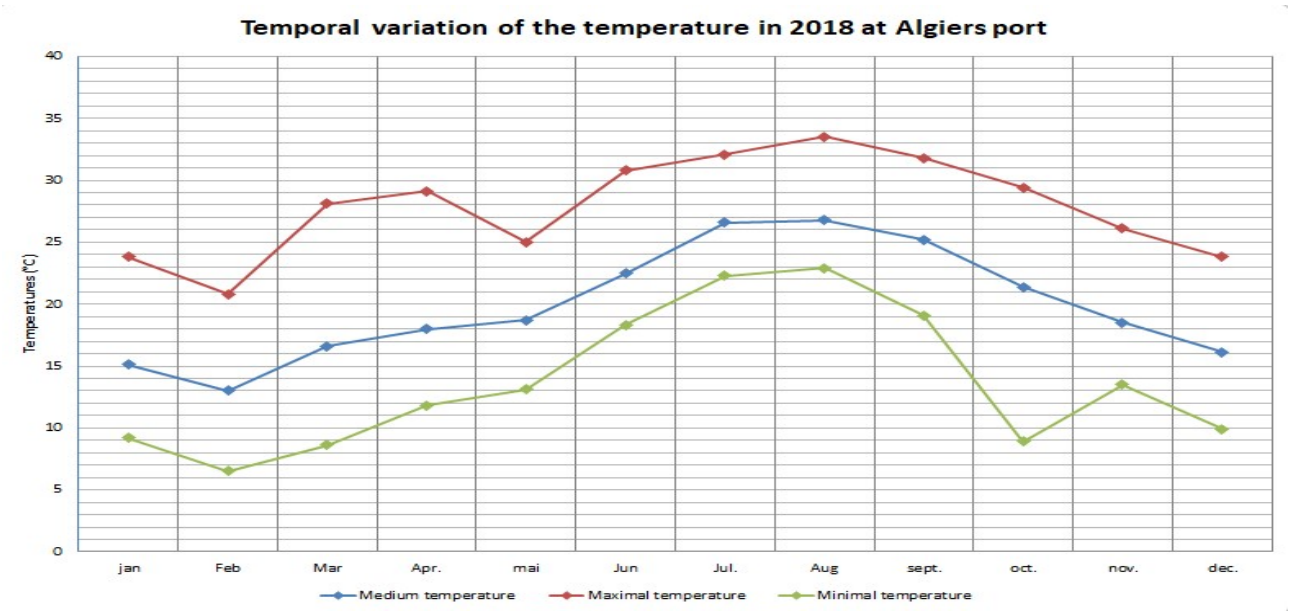

Figure 3.3: temporal variation of the temperature

3.3.3 Relief

The study area is an alluvial plain interrupted by a series of elevations that have a WSW direction to an ENE, sometimes reaching an altitude of 260 meters. The hydrography includes three main wadis with irregular flow:

- Wadi El Harrach in the west.
- Wadi Reghaia in the east.

- Wadi El Hamiz in the center.

3.4 Geological study

The geological structure of Mitidja has been identified since the geological studies of "Glangeaud and Ayme" (1935) (Winn, 1973; Sekkal, 1986). It is crucial to be well informed of this part in order to understand the hydrogeological mechanisms.

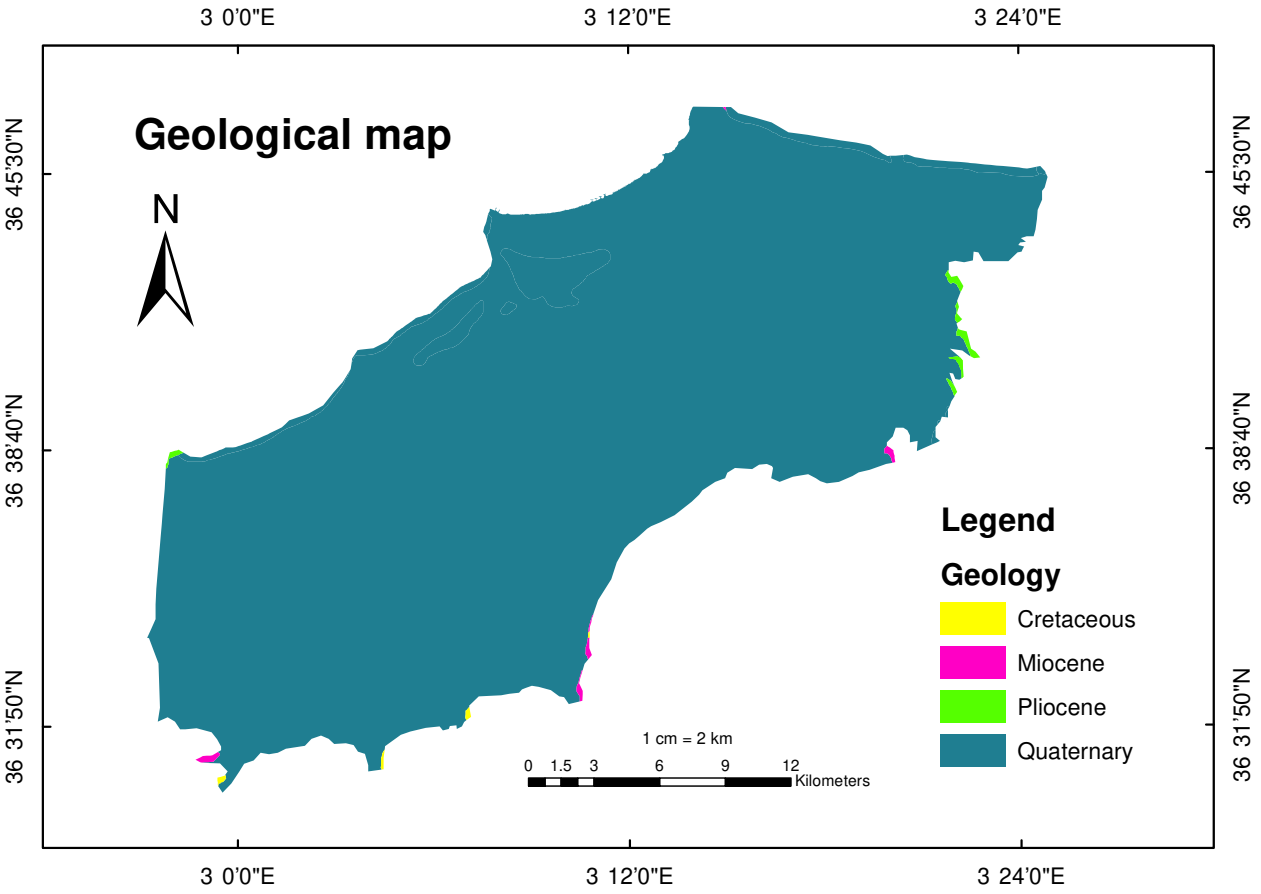

Figure 3.4: Geological map of Eastern Mitidja

source:ANRH, 2018 edition

3.4.1 Stratigraphy

The Atlas and the Sahel are positive ascending blocks, limiing Mitidja from the north to the south. The plain is a very large coastal alluvial basin, formed by permanent subsidence since the Miocene, followed by sedimentation during the Pleistocene. Then, at the end of the Tertiary, a slight folding occurred, which gave it an original synclinal structure.

Miocene and older rock formations: This type of rock can be found in the Atlas at the southern limit of Mitidja, between Meftah and Khmiss El Khechna, and in the heart of the Sahel Anticline, while its upper limit rises to be in contact with the Astian series that overlies the basin (Aymé, Glangeaud, and Magne, 1954). In this formation, the most frequent rock types are clays, mudstones, and shales. Tertiary limestones, sandstones, conglomerates, lava, and volcanic rocks are in contact with recent alluvial deposits of the basin. Metamorphic rocks of the eruptive basement appear in the basin's north and east as tiny isolated outcrops.

Pliocene marine formation (Plaisancian-Astian): Marine Pliocene formations constitute the majority of the long Sahel bulge and a portion of the Mitidja filling.

Plaisancian: The majority of the Mitidja basin is covered with a thick, continuous succession of gray or blue marls, sometimes sandy, known as Plaisancian marls (Glangeaud, 1952). These typical deep-water sediments were deposited above an erosion surface that was 200 meters thick on average. Marl's outcrops can be seen on the surface inside the anticline of the Sahel hills and Reghaia, where they form the basin's eastern closure (Rivoirard, 1952). Other deposits filling the basin dissimulate the Plaisancian formation.

Astian (represents the middle Pliocene): It starts with a glauconite level, which is a clay mineral association halfway between Mica and Smectite, followed by molasse, yellow marls, sandstone facies, calcareous or calcairo-gravel facies, dispersed in isolated units ranging from 1 to 15 m, as it becomes increasingly clayey towards the S-W. It also contains sandstones that often include a high proportion of broken shell fragments and are classified as detrital sandy limestones. The sandstones, sometimes silty, contain yellow rounded gravels and reveal an interlocking stratification. The Astien formation, which has an average thickness ranging from 100 to 130 m, may be found at depths going from 250 to 300 meters below the natural ground level and seen beneath the newer deposits of the basin, up to the Atlas peak. However, there are only a few isolated outcrops along the southern border

of the basin, where the Astian is in contact with both the older Atlas rocks and the newer alluvial deposits.

Continental Pliocene (El Harrach Formation): Clayey materials dominate in alternation with yellow sticky marls, gravelly clays, and some sand and gravel beds assimilated to "El-Harrach." It is composed of a reasonably consistent series of yellow or grey polymeric clays with gravel layers ranging from 2 to 6 meters in length. The layer's thickness, which approximately equals 100 meters in the basin's middle region, gradually declines as it approaches the Atlas.

Mitidja Formation (Continental Quaternary): The Mitidja formation is mostly made of alluvial elements such as sands, sandy clays, gravelly clays, gravels, and clayey or silty sandstone pebbles that originate from several sources and have significant granulometric changes in the horizontal and vertical directions.

3.5 Hydrogeological study

Hydrogeology is defined as the science of the occurrence, distribution, and movement of water below the Earth's surface. In the broadest sense, hydrology addresses occurrence, distribution, movement and chemistry of all waters of the earth. The results of geological and geophysical studies have shown the existence of aquifer formations in the Mitidja. In the following study, we shall present the hydrogeological characteristics by detailing each one.

Mitidja's aquifers: The lithology and hydrodynamic characteristics allow for the distinction of two aquifer units:

- The Astian reservoir.
- The Quaternary Alluvium.

The Astian reservoir: The Astian formation represents the deepest aquifer. It mainly consists of sands, sandstone and more or less limestone, delimited by the argilous Plaisancian substratum and covered by the Marls of El Harrach, at depths ranging from 200 to more than 400 meters in the western zone. The aquifer has the configuration of a flat-bottomed with a northern limit that rises to form the flank of the Sahel. The Astian is in direct contact with the alluvial deposits; so, alluvium in the eastern half of the plain forms a unique aquifer along with the Quaternary. The rainfall can

aliment this reservoir. It mainly rains on the outcrop of the Sahel. The outlet is the alluvial aquifer by draining through the yellow marls (vertical permeability of marls: in the Western Sector).

The quaternary reservoir: In this formation, quaternary alluvium and the base of recent alluvium are mainly considered as its components. The substratum is composed of the yellow marl of El Harrach. This aquifer consists of gravel and pebbles interspersed with silt and clay. The marls of the El-Harrach formation define its lower boundary, while its upper limit is independent.

From a hydraulic behavioral aspect, the water table in the Harrach-Hamiz basin is entirely free, allowing direct infiltration of rainfall in which rain can directly infiltrate. The alluvial groundwater of the Mitidja is supplied by:

- Infiltration from rainfall on the plain.
- Infiltration from rivers and runoff in the foothills of the Atlas.
- Leakage from water distribution networks.
- Infiltration of excess irrigation water.

3.5.1 Physico-chemical analysis of the Mitidja aquifer

Physico-chemical analysis of water has enabled the acquisition of a number of parameters that are often used in the estimation of water quality. The chemical composition was determined in 2018, using an average of forty-five wells.

Data accuracy evaluation: The Ion Balance is a simple calculation used to determine the accuracy of water analysis data based on the principle of electrical neutrality, meaning that the sum of equivalent concentrations of anions and cations should be nearly equal. It is calculated by the following formula:

$$IB(\%) = \frac{\sum cations(\frac{meq}{l}) - \sum anions(\frac{meq}{l})}{\sum cations(\frac{meq}{l}) + \sum anions(\frac{meq}{l})} \tag{3.1}$$

The IB results for each well range from -5% to 4.9%, within the acceptable interval of this parameter. This means that the accuracy of the water analysis is high. The statistical results of chemical analysis are shown in the table below.

Table 3.3: Cations and anions concentrations (mg/l)

	Ca^{+}	Mg^{2+}	Na^{+}	K^{+}	Cl^{-}	SO_4^{-}	HCO_3^{-}	NO_3^{-}
Min	40.40	2.61	19	0.1	36	0.5	91.5	3
Max	249.46	93.91	324	9	490.1	445	823.5	99
med	148.2	48.219	101.71	1.6	151.8	181.98	363.63	53

Source: ANRH, Physico-chemical analysis of 2018.

The average electrical conductivity of the analyzed samples is less than 1500 S/cm, with a maximum of 2400 S/cm, translating to a light salinity of the Eastern Mitidja aquifer. The World Health Organization (WHO) sets 1200 µS/cm as the maximum permissible concentration for electrical conductivity, while the average measured concentration is 1466 µS/cm, which means that the salinity of the groundwater in the East Mitidja plain is relatively high. The average sodium and calcium concentrations do not exceed 150 mg/l, which is well below the maximum permissible limits for potable water. However, the average chloride, bicarbonate, and sulfate concentrations exceed 250 mg/l, which is not permissible. The nitrate concentration in the study area varies between 3 and 99 mg/l, while its average is higher than the maximum permissible of (50 mg/l).

The spatial distribution of nitrate concentrations in the Mitidja aquifer

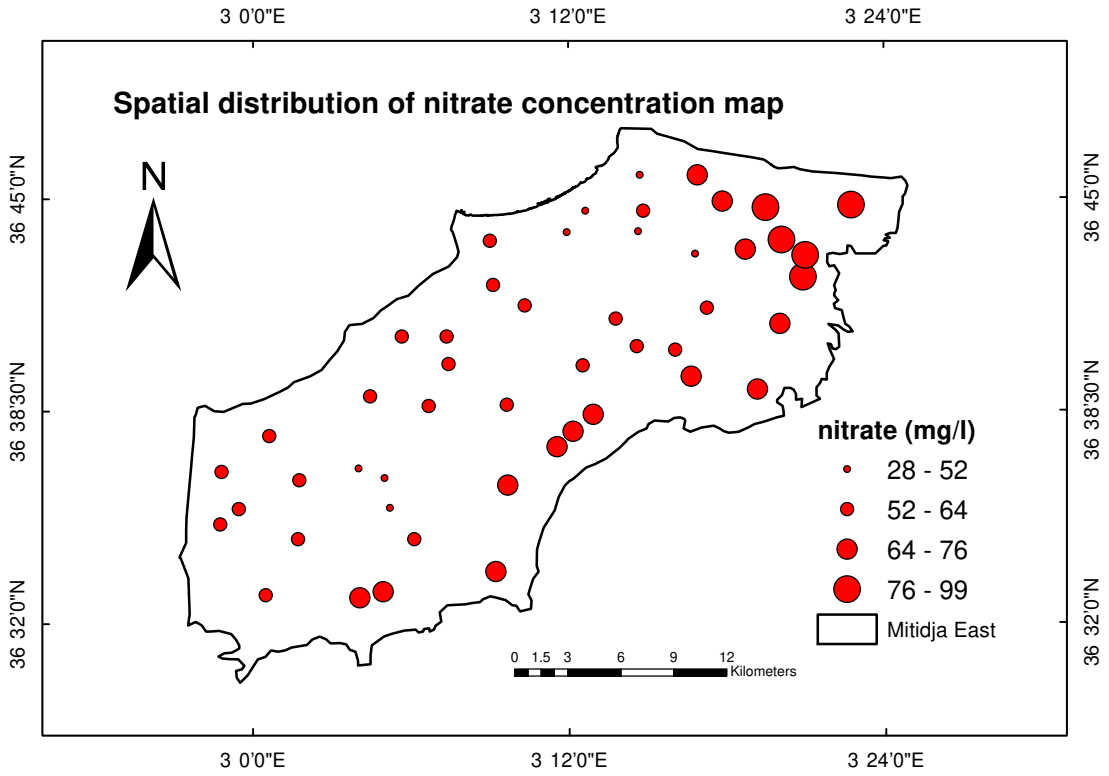

Figure 3.5: Spatial distribution of nitrate (NO_3^-) concentrations in Eastern Mitidja

The dataset utilized in this work includes both real well data and data generated with the Inverse distance weighted (IDW) interpolation using the Geographic information system (Gis).

The map shows that the highest nitrate concentration is mostly located in the south and the south east of Mitidja, regrouping all of Bougara, Ouled slama, Larbaa, Meftah, Khmis el khechna, Reghaia, Boudouaou and Haraoua municipalities. The north west of Mitidja shows a moderated concentration, while Sidi moussa, Rouiba, Dar el beida and Bordj el kiffen are uncontaminated regions.

3.6 Conclusion

This chapter has allowed us to learn about the geographical, geological, hydrogeological, and climatological properties of our study area. The identification of exploitable areas contaminated by nitrates leads us to seek to deepen the assessment tools used in our method of evaluating and mapping the groundwater vulnerability to contamination in the Eastern Mitidja.

Chapter 4

Methodology

4.1 Introduction

The parameters employed in our study were chosen from previous but recent groundwater assessment studies (Lahjouj, El Hmaidi, Bouhafa, et al., 2020; Ahada and Suthar, 2018a). A general review of machine learning approaches is also presented in this chapter.

4.2 Factors

Groundwater can be exposed to all forms of contamination, whether it is urban, agricultural,industrial,or accidental. Its vulnerability depends on a certain number of geological, hydrological,hydrogeological,and topographic factors.

In our study, eleven parameters are represented and mapped using the geographic information system (GIS).

4.2.1 Factors determined from satellite images

The digital elevation model (DEM) with a cell (grid) size of 27.7 by 27.7 meters was used to produce the drainage density, distance to river, TWI, and topographic maps of the study area. The Fill tool was first performed on this DEM in order to correct spurious peaks and sinks.

Distance to river

Due to human activities, rivers can easily be infected by runoff pollution. The distance to the river factor has a significant impact on the contamination of groundwater, knowing that most water exchanges take place in the adjacent areas of the river, which allows the infiltration of higher volumes of its contaminated water into the aquifer (Band et al., 2020).

The distance to river map requires first the stream order of the study area then the Euclidean Distance tool.

NDVI

The Normalized difference vegetation index (NDVI) indicates the health of vegetation, vegetation quantity, and the resistance force against erosion of the top soil in a particular area. The denser the vegetation cover is, the more water is kept from infiltration into the aquifer (Ghosh and Lepcha, 2019).

After downloading 2018 Landsat 8 data from the US Geological Survey (USGS) website, Both the Red and Near Infrared bands were atmospherically corrected with the formula published in the USGS Landsat 8 Product and by using the Raster Calculator, of the Spatial Analyst tools :

$$pA = \frac{(Mp \times Qcal + Ap)}{sin(Theta - se)} \tag{4.1}$$

where:
Mp: Band specific REFLECTANCE-MULT-BAND (from the metadata).
Qcal: value provided in the x-band of the pixel
Ap:REFLECTANCE-ADD-BAND (from metadata).
Theta-se : Local sun elevation angle; the scene center sun elevation angle in degrees is provide in the metadata(SUN-ELEVATION).
pA :Top of the Atmosphere (TOA) reflectance.

This conversion of raw digital numbers into reflectance values has a huge impact on the correction of constant error effects on NDVI by reducing topographic shading (Guyot and Gu, 1994).

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \tag{4.2}$$

Drainage density

Drainage density is indicative of infiltration, which depends on both the climate and physical characteristics of the drainage basin. This factor is obtained by the total length of all the streams and rivers divided by the total area of the drainage basin. In general, the drainage density represents the quantity of runoff to stream discharge (Carlston, 1963).

The determination of the drainage density map requires a stream order map and the line density tool from the arc toolbox.

Land use

Land use is a parameter that represents a potential anthropogenic factor related to NO3-. It may create impervious surfaces through urban construction and have a specific and cumulative effect on water quality, quality of wildlife habitat, climate, and human health.
The land use map of Mitidja was extracted from the esa world cover website (*ESA worldcover*) that provides the global land cover at 10 m resolution, based on Sentinel-1 and 2 data.

Slope

The slope is a topographic parameter that shows changes in the land relief and controls the runoff. A high slope contributes to water flow. Therefore, it reduces the probability of groundwater contamination. The slope map was determined using the DEM fill and the slope operation in the arc toolbox.

Topographic Wetness Index (TWI)

The Topographic Wetness Index is used to determine topographic controls on hydrological processes and quantify soil moisture. The TWI is determined using the DEM fill, hydrology operations (flow direction/ flow accumulation) of the arc toolbox and the slope map, then using functions on map algebra in the following order:

$$Radians of slope = \frac{(Slope(DEM) \times 1.570796)}{90} \tag{4.3}$$

$$Tanslope = con(slope > 0; tan(slope); 0.001) \tag{4.4}$$

$$Flowaccumulationscaled = (flowaccumulation + 1) \times cellsize \tag{4.5}$$

$$TWI = Ln\left(\frac{Flowaccumulationscaled}{Tanslope}\right) \tag{4.6}$$

Elevation

Land elevation is the height of the land above the mean sea level. It impacts soil chemical and physical properties through precipitation, temperature, and vegetation effects. The elevation map was prepared from the digital elevation model (Dem).

4.2.2 Factors determined from measured data

The source of most of the measured data was the National Water Resources Agency (ANRH). This data was then interpolated using the Inverse Distance Weighted (IDW) method, which is a mathematical method based on the distance influence of known cell values on the interpolated points.

Depth to groundwater

The depth to groundwater has an important effect on the determination of groundwater vulnerability to contamination. It represents the depth from the land surface to the water level in the well. The greater the water table depth is, the more residence time for water as it passes through soil layers and gets disinfected (Saranya and Saravanan, 2021).
The seasonal vertical fluctuation of the piezometric level causes rinsing of the particles from the unsaturated zone and entrains the adsorbed substances into the aquifer (Ahada and Suthar, 2018b).

Permeability

Permeability describes the ability of water (or other liquid) to move through the soil. The connection between pore spaces allows groundwater to flow through the sediment or rock and transmit fluids that may contain contamination.

Rainfall

Rainfall is a climatic parameter that can be considered to be an aquifer input influencing groundwater pollution. Rainfall recharges groundwater, which engenders the leaching of soil NO_3^- (Band et al., 2020).

Vadose zone

The vadose zone is an unsaturated layer placed just above the water table. It mainly controls water movement from the land surface to the aquifer. The thickness and permeability of vadose zone materials significantly influence the transmission of recharge water and contaminants by filtering undesirable substances.

The vadose zone map was determined in three steps, using 35 lithological units:

- First, we extracted the thickness of each lithological formation, and then we obtained numerous classes.

Table 4.1: The thickness of formations in 35 lithological units

vadose zone (m)	the thickness of each lithological formation (%)
28.5802	59% marl 7% marl and gravel 10% peat 24% marl and pebble
29.0093	17% gravel 29% clay 54% gravel clay mixes
25.6206	22% sandstone 4% clay 40% limestone 16% pebbles 8% yellow molasse 10% marl
28.6145	78% sandstones 22% clay
31.4646	19% sandy clay 81% sand
20.328	48% clay 52% gravel
19.3685	13% gravel 87% clay
34.5367	32% Dune stone 68% Grey marl
28.2569	100% yellow clay
21.464	29% sand and gravel 71% marl
32.3853	8% yellow limestone 52% sand and gravel 2% limestone 38% blue marl
32.2228	8% dunes 42% sandstones and pebbles 50% sands boudingues gres
35.8439	17% clay 83% alternating clay and gravel
41.9488	33% pebbles and sand 20% clayey sand 36% clayey sand 11% blue marl
39.8549	15% marine sand 30% sandstone 18% clay 8% limestone 10% molasse 19% clayey sand

37.7533	100% gravel
13.2681	75% yellow clay 25% gravel
15.1531	72% clay and sand 28% clay
44.0377	26% clay 28% sand and sandstones 46% sand and gravel
43.2927	12% clay 48% coarse gravel 40% sandy clay
43.116	27% clay 32% gravel 41% yellow sand and sandstone
43.2608	37% sand and gravel 25% clay and pebbles 9% clay 29% gravel
44.5164	47% clay 53% alternating clay and gravel
43.6848	43% clay 57% soft conglomerates
43.8141	100% alternating clay and gravel
43.4436	100% alternating clay and gravel
43.275	67% manes 33% gravel
43.4809	34% clay 7% sand 30% marl 29% gravel
53.1692	7% silt 9% gravel 9% clay 58% gravel 17% pebbles
47.8426	8% silt 17% marl 17% sandy marl 25% sand and gravel 33% yellow sandy clay
49.5565	12% sand 4% marl 21% gravel 22% grey marl 41% yellow clay 12% sand 4% marl 21% gravel 22% grey marl 41% yellow clay
44.1563	50% clay 34% gravel 16% yellow marl and some gravel

source: Lithological units, ANRH.

- Secondly, we classified the previously obtained components according to their permeability into three major categories: permeable, semi-permeable, and impermeable formations. Three

maps were created using the Geographic information system (Gis).

- Finally, we superposed the three maps in order to represent the impact of the vadose zone.

4.3 Machine Learning

Since their evoluion, humans have employed a range of tools to do various jobs more effectively. The ingenuity of the human brain has permitted the invention of multiple technologies. These technologies have simplified man's life by allowing him to address a wide range of needs and requirements, such as industry and information technology. Machine learning is one of these gadgets (Mahesh, 2020).

4.3.1 Types of Machine Learning

There are so many different types of Machine learning systems that it is useful to classify them into broad categories: supervised learning, unsupervised learning, semi-supervised learning, and Re-reinforcement Learning .

Unsupervised learning

Unsupervised learning is a type of algorithm in which the training data is unlabeled. In other word, it allows the model to work on its own to discover patterns and information patterns from untagged data. This approach is used to divide the data into groups of homogeneous elements (MIFDAL, 2019).

Semi-supervised learning

Semi-supervised learning takes a middle ground between Supervised and Unsupervised learning algorithms. It makes use of a little amount of labeled data to boost a larger set of unlabeled data.The advantage of using this type of learning is that labeling data can be expensive and often takes a lot of time and memory space (MIFDAL, 2019).

Reinforcement Learning

This is a very different learning system, which is done without any supervision, by observing with the environment. In general, Reinforcement Learning is able to reward desired behaviors and/or punishing undesired ones and interpret its environment, take actions and learn through trial and error. The objective is to determine the optimal behavioral strategy to maximize the total reward. To do this, a simple feedback of the results is essential for the machine to define its way of behaving. This outcome feedback is called the reinforcement signal. Although it is not evident to manually build or program relevant scenarios for improved results, machine learning is highly advantageous in circumstances when there are a huge number of conditions for predictions (Salmani, 2013).

supervised learning

Supervised Learning is a Machine learning paradigm in which the data comes with additional attributes that we want to predict. This type of learning acquires the input-output relationship information of a system based on a given set of paired input-output training samples (Ju et al., 2018).
A typical supervised learning task is classification, in which the object is identified with a specific category. The contamination to nitrate is a good example of this: it is trained with 70 % of the nitrate data of many wells along with their class (polluted or not polluted), and it must learn how to classify the other wells.

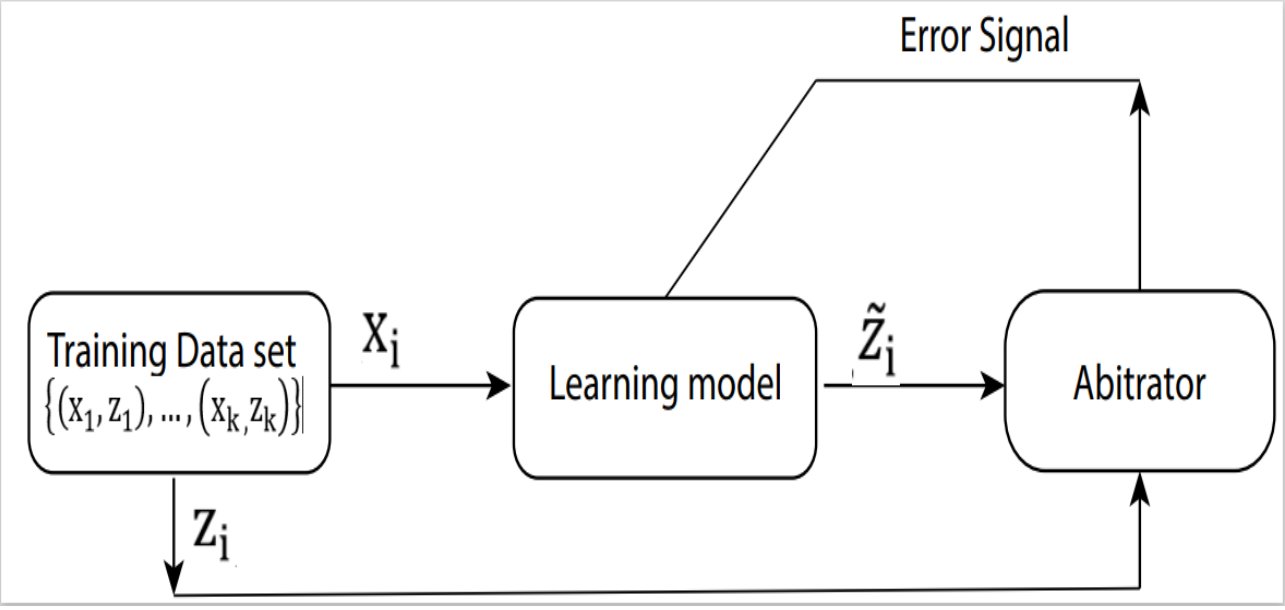

Figure 4.1: Block diagram that illustrates the form of Supervised Learning

The graphic above is a functional diagram illustrating the structure of supervised learning. In this figure:

(x_i,z_i): a supervised training sample.
'x': the system's input.
'z': the system's output.
'i': the index of the training sample.

During a supervised learning process, the 'x_i' values are fed to the learning system as a training input in order to generate an output 'i'. An abitrator compares the learning system output 'i' to the ground truth labeling 'z_i' and computes the difference between them. The difference, termed "error signal" in this diagram, is then sent to the learning system for adjusting the parameters of the learner. The purpose of this learning process is to arrive at a set of optimal learning system parameters capable of minimizing the differences between and 'z_i' for the entire index 'i'.

4.4 The used models

In this study, the two Ml used models are considered from among the most commonly used algorithms that have demonstrated their performance in the Mapping of groundwater vulnerability to nitrate. They are as follows:

4.4.1 Decision Trees

Decision Trees (DTs) are a versatile supervised learning method used for classification and regression. In a decision tree, the data set is divided into two types of nodes (decision nodes and leaf nodes). Each decision node contains a condition to split the data. Each leaf node helps to decide the class of the new data. Each branch represents the outcome of the, A node that has no children is a leaf.

4.4.2 Random Forest

A Random Forest is an ensemble of decision trees, generally trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result and the predictive accuracy. Random forest creates decision trees from randomly selected data samples, gets predictions from each tree, and selects the ultimate solution by voting on the best decision (Géron, 2019a).

4.4.3 Adaptive Boosting (AdaBoost)

An AdaBoost classifier is a meta-estimator algorithm that is trained with the "Boosting" method; it consists of combining several weak learners into a strong learner. The idea is that the classifiers will have to focus on the output data that is difficult to classify correctly, thus improving the overall performance (Géron, 2019b).

4.5 Evaluation criteria

4.5.1 Confusion matrix

A confusion matrix is a visual evaluation tool used to predict results on a classification problem, which can give a better idea of what your classification model is getting right and what types of errors it is making.
A Confusion Matrix has a dimension of 2×2. A set of algorithm performance indicators, such as the positive region check rate and the negative class recall rate. These criteria are relevant to all classification methods (Xu, Zhang, and Miao, 2020).
“true positive” for correctly predicted event values.
“false positive” for incorrectly predicted event values.
“true negative” for correctly predicted no-event values.
“false negative” for incorrectly predicted no-event values.

4.5.2 Accuracy

Accuracy is one metric for evaluating how often the algorithm classifies a data point correctly. It is defined as follows: The total number of right predictions (true positives + true negatives) divided by the total number of predictions (Xu, Zhang, and Miao, 2020).

4.5.3 Receiver Operating Characteristics (ROC) curve

An ROC curve is a graphical plot typically used in binary classification to study the output of a classifier. It illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied (Body et al., 2017).

An ROC curve that follows the diagonal line y=x produces false positive outcomes at the same rate as real positive results. As a result, we expect a reasonably accurate diagnostic test to have a ROC curve in the upper left triangle above the y=x line (Body et al., 2017).

4.6 Conclusion

In this chapter, we presented all the parameters by determining their influence on the vulnerability of groundwater to contamination and how their maps were generated. We also presented the Random Forest and Adaboost machine learning models to map groundwater vulnerability as well as their evaluation methods, including the confusion matrix and the ROC (Receiver Operating Characteristic) curve.

Chapter 5

Results and Discussion

5.1 Introduction

The following chapter includes the modeling process, the interpretation of the ten influence factor maps, as well as the result of groundwater vulnerability maps.

5.2 The interpretation of influence factor maps

5.2.1 Depth to groundwater

The Figure 5.1 shows that the groundwater depth is shallower in the south-east of the aquifer, varying from 4.1 to 23 meters. The groundwater gets deeper in the north-east, heading up to 39 meters, while it becomes the deepest on the west side, reaching a maximum of 56.4 meters.

5.2.2 Distance to river

Three main wadis and their affluents cross the eastern plain of the Mitidja, including Harrach, Hamiz, and Reghaia, which flows directly to the north, into the Mediterranean Sea. The Figure 5.2 shows the distance of every place or pixel in the plain from these wadis, and the distance varies from 0 to 6 kilometers.

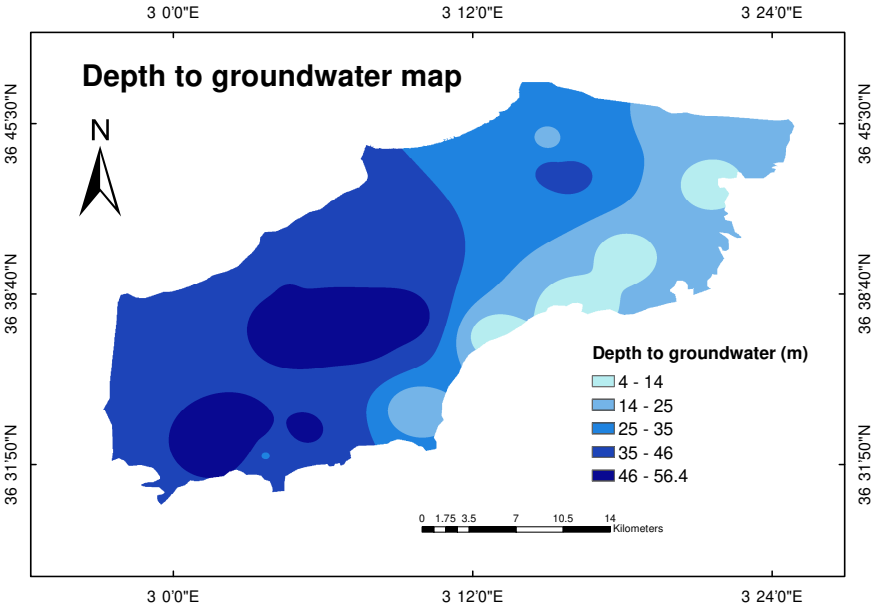

Figure 5.1: Depth to groundwater map

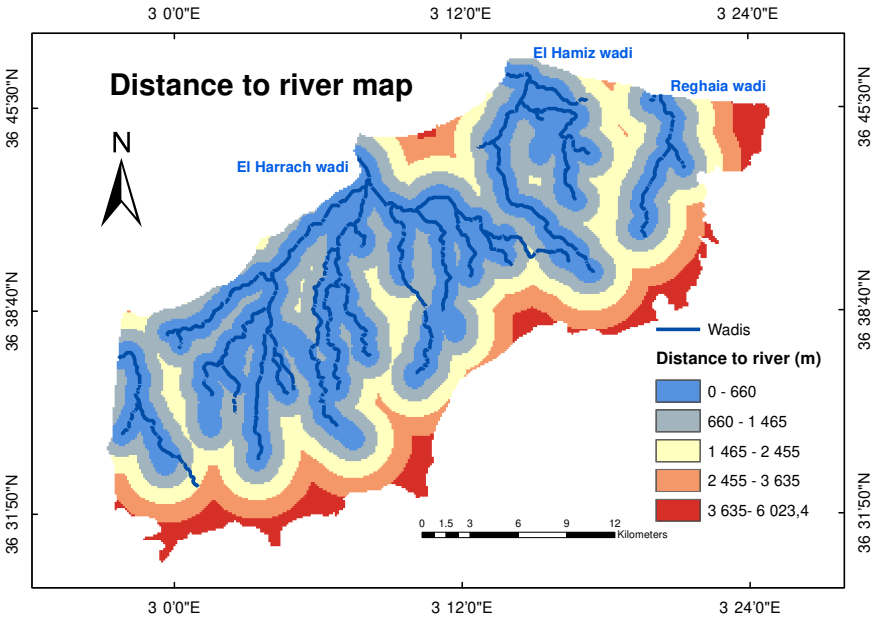

Figure 5.2: Distance to river map

5.2.3 Drainage density

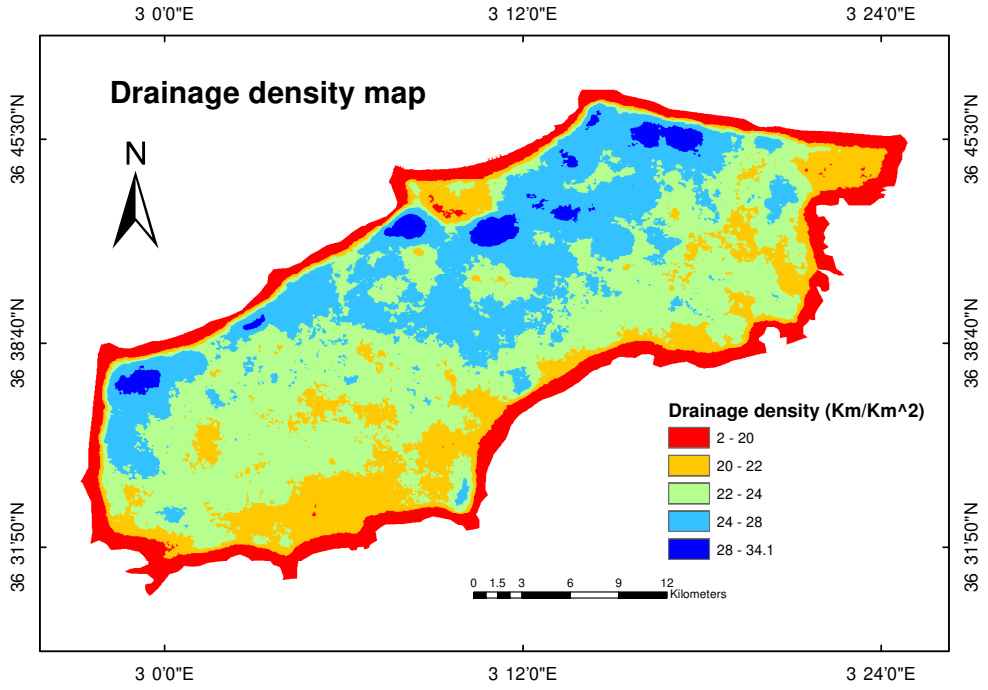

Figure 5.3: Drainage density map

The Figure 5.3 indicates that drainage density in the north is mostly high, varying from 24 to 34 km/km^2. It decreases to 20 km/km^2 down to the south, while all the borders of the plain have the lowest drainage density, with a minimum of 2 km/km^2.

5.2.4 Elevation

According to the elevation map, Figure 5.4, the altitude of the majority of the plain ranges from 0 to 100 meters above sea level, while it reaches 262 meters in the southeastern and southwestern limits.

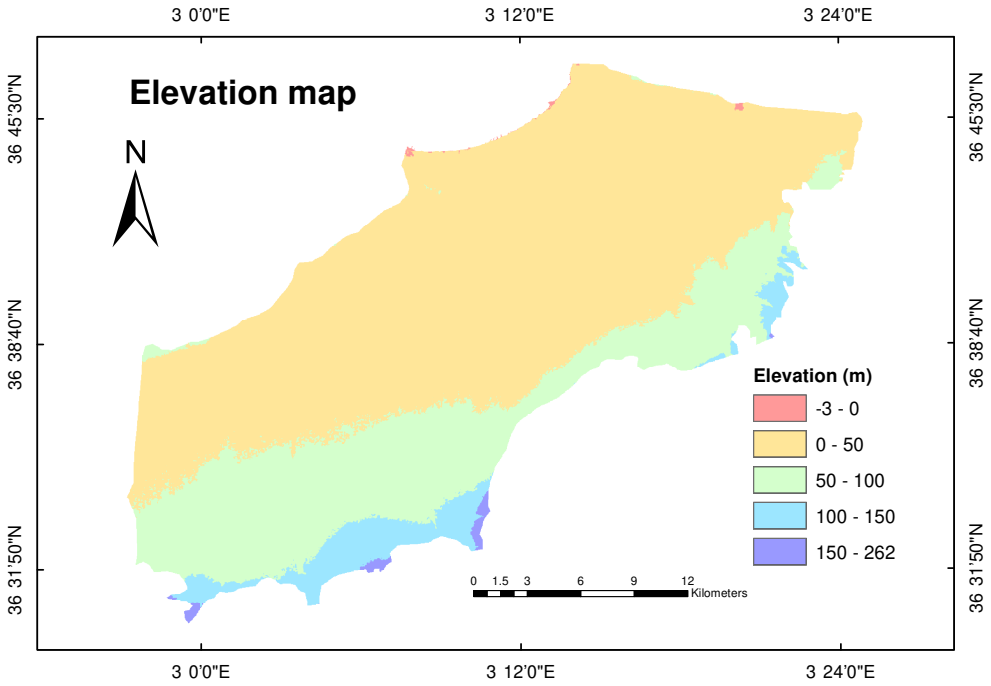

Figure 5.4: Soil Elevation map

5.2.5 Normalized Difference Vegetation Index (NDVI)

The NDVI map, Figure 5.5, is divided into four classes; the first class of negative values indicates the presence of clouds or water; the second class from 0 to 0.2 refers to urbanization or bare soil, which is more concentrated in the north east (Algiers and Boumerdes) as it decreases and gets more dispersed in the south. Moderate values from 0.2 to 0.3 represent shrub and grassland, while higher values indicate forests and croplands that cover most areas of the west.

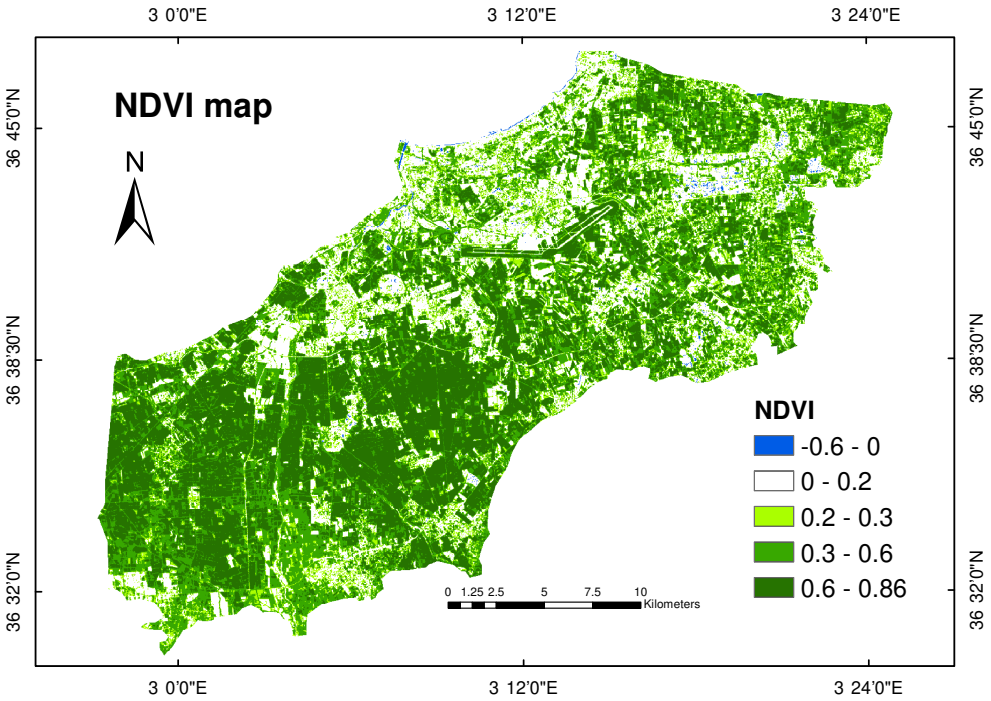

Figure 5.5: Normalized Difference Vegetation Index (NDVI) map

5.2.6 Land use

The Figure 5.6 shows different important categories of land use in the Mitidja plain, where bare and urbanized land represent 38% of the total surface, mainly located on the coast (Algiers), some in the eastern part (Boumerdes), and the southwest (Blida). The rest of the plain covers about 61% of its surface with cultivated lands, grasslands, shrubs, and trees.

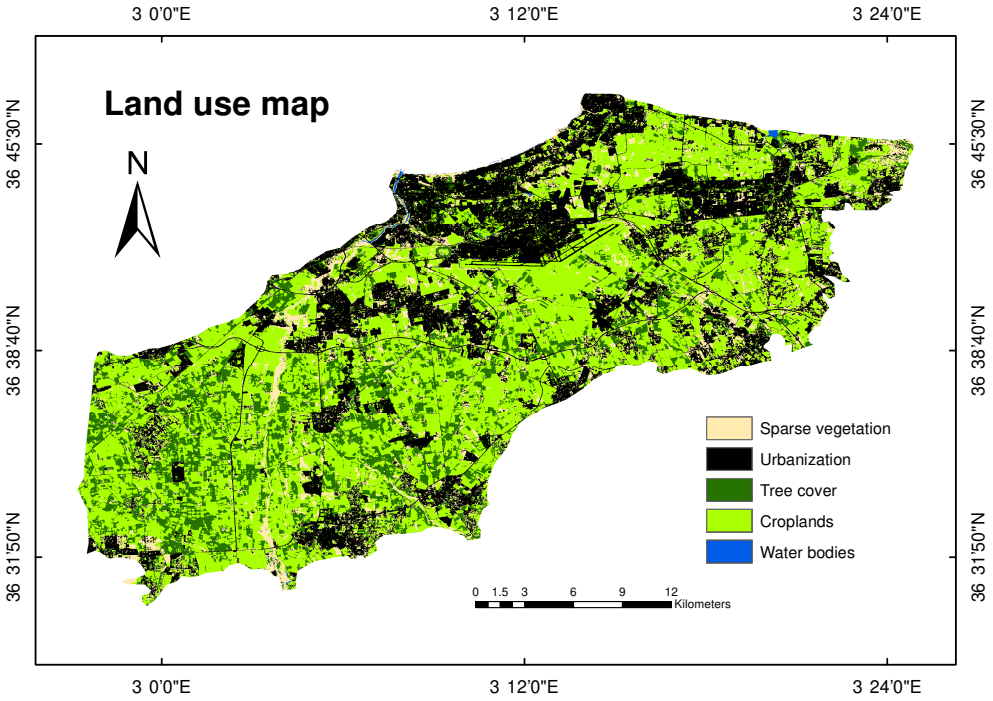

Figure 5.6: Land use map

5.2.7 Permeability

The Figure 5.7 shows that the two most permeable areas of the plain are located in the northeast (Bordj el Bahri, Harraoua, part of west Aintaya and the north of Rouiba) and the north central part (Baraki and El Harrach), with a permeability ranging from 10^{-3} to 10^{-2} m/s. The rest of the plain surface is considered semi-permeable with a permeability varying from 5.10^{-5} to 10^{-3} m/s over most of the northern and northeastern regions, while it drops to 10^{-5} in the southwestern part of the plain.

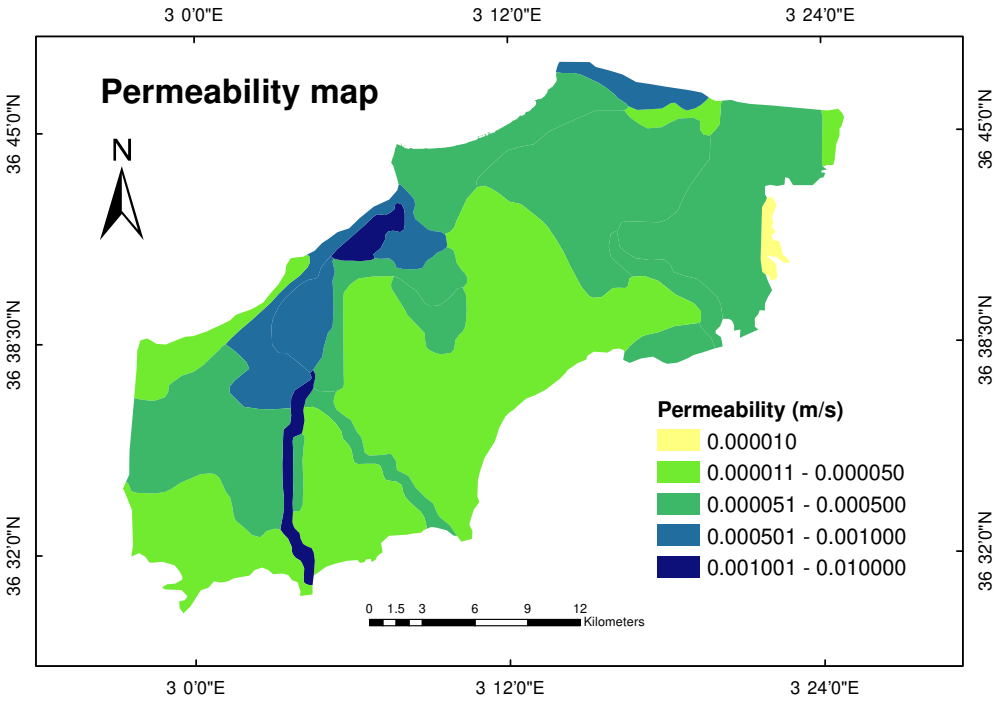

Figure 5.7: Permeability map

5.2.8 Rainfall

The Figure 5.8 shows the inter-annual rainfall variation, which gradually increase from 599 mm in the northeast to 724 mm in the west and south at Meftah and Larbaa.

5.2.9 slope

The Figure 5.9 that the slope is very gentle in most of the Mitidja plain, varying from 0 to 5 degrees, while it gets strong, reaching 9 to 28 degrees at the southern limits.

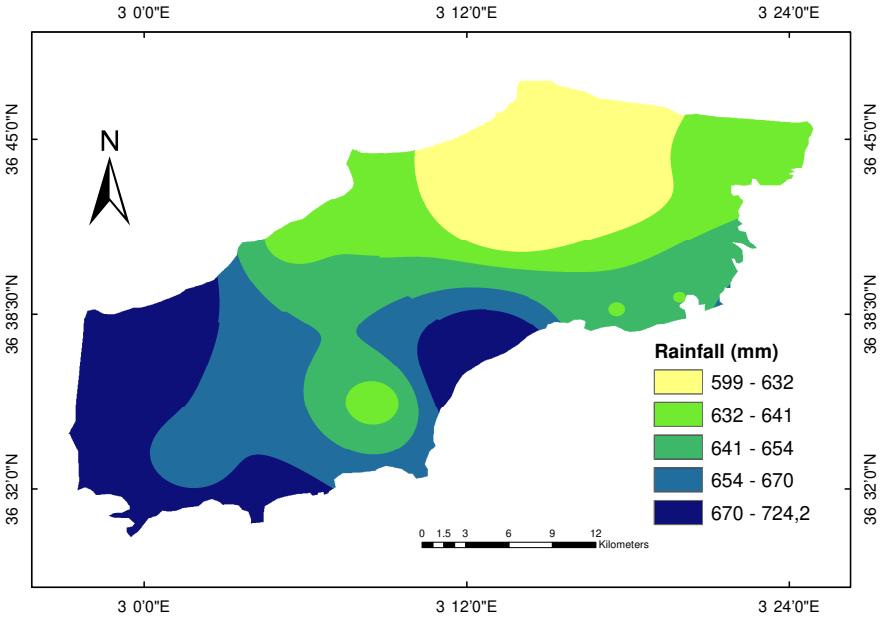

Figure 5.8: Interannual Rainfall map

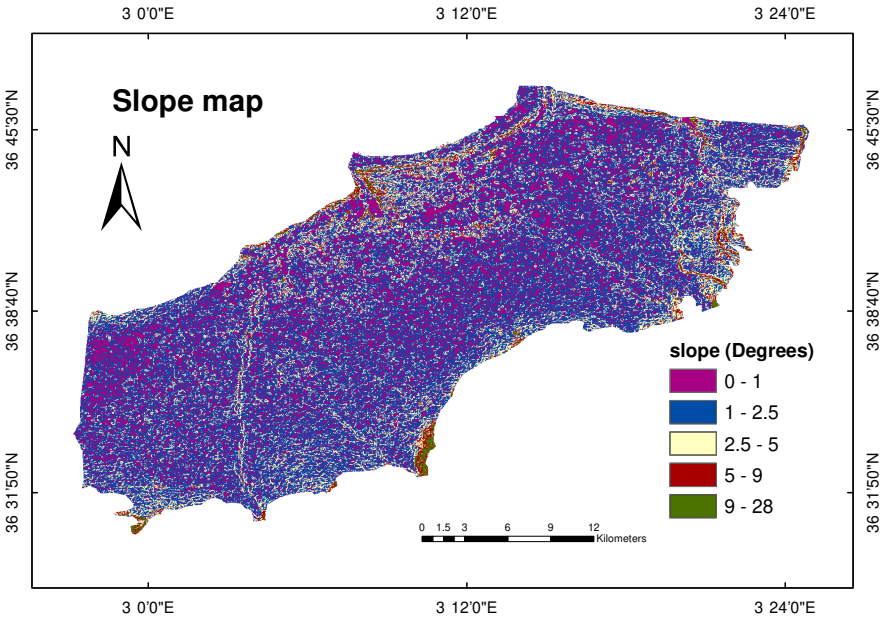

Figure 5.9: Slope map

5.2.10 Topographic Wetness Index (TWI)

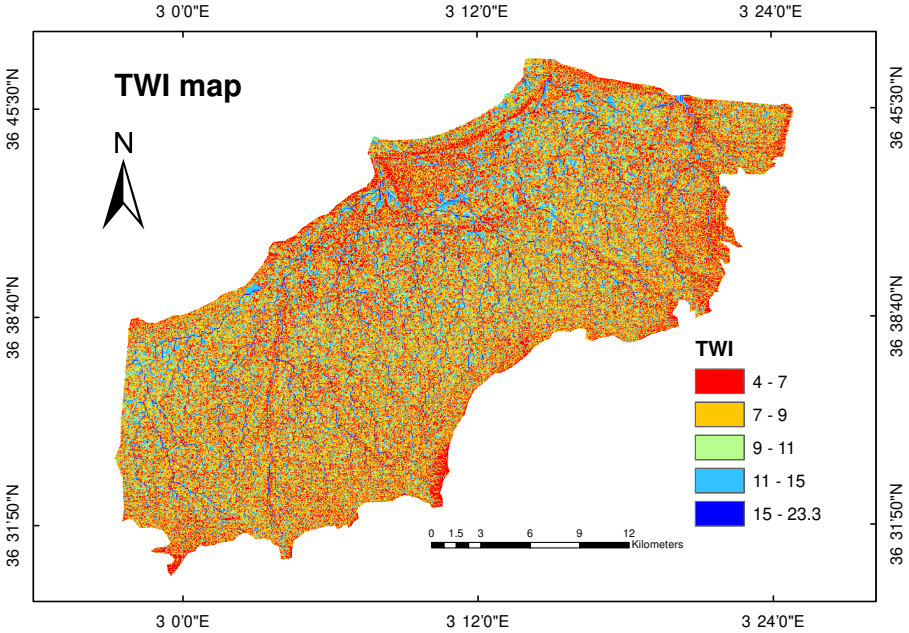

Figure 5.10: Topographic Wetness Index (TWI) map

The topographic wetness index, Figure 5.10, determines the amount of soil moisture, where water tends to drain and where it can remain and saturate the soil. This map is divided into five classes; areas with high TWI ranging from 11 to 23.3 accumulate water to varying degrees, with the highest values referring to wadis. The green class is indicative of moderate moisture, while areas with lower TWI ranging from 4 to 9 lack moisture and will not accumulate water.

5.2.11 Vadose zone

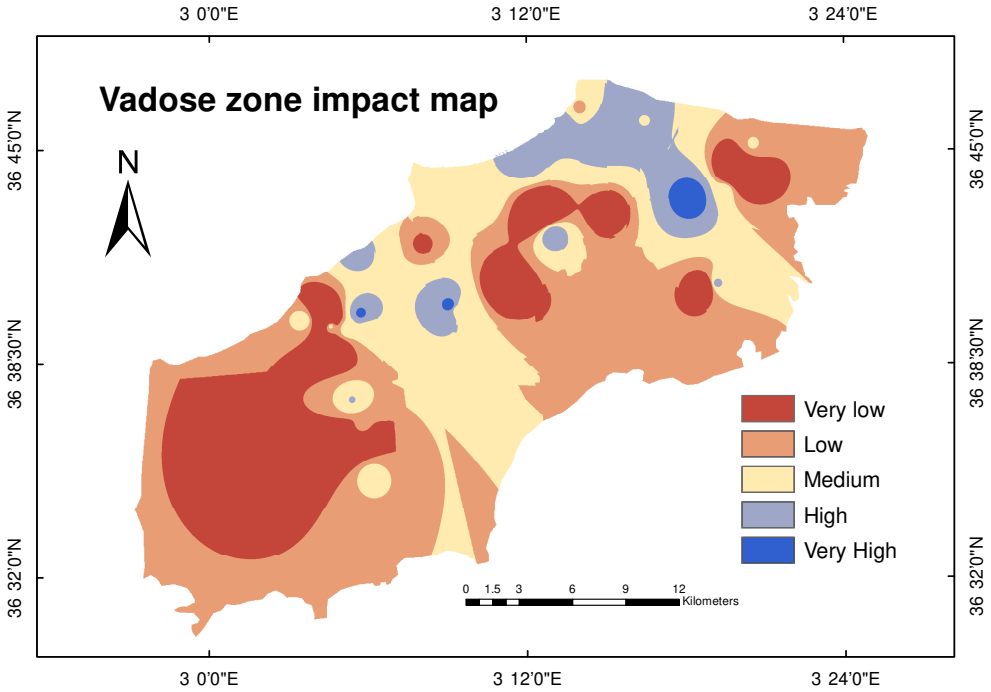

Figure 5.11: Impact of vadose zone map

The Figure 5.11 shows that areas of high to very high impact refer to the permeability of the vadose zone, permitting the leaching of nitrate contamination into the groundwater, mainly located in the east at Rouiba, Bordj El Kiffan and south of Bordj El Bahri, also in some areas of the center at Dar El Beida and west of the Eucalyptus, the medium impact refers to the semi-permeable vadose zone, which is located in the northeast down to Larbaa, Ouledmoussa and others.., while the low to very low impact is located in the west and extreme northeast at Reghaia, Boudouaou, as well as in the center at Meftah, KHmiss El khechna.

5.3 Modeling approach

The modeling approach regroups all of the modeling steps, starting from data processing to training and validation of the ML model.

5.3.1 Data Pre-processing

Data pre-processing guarantees data quality and its usability because it has a great influence on the model learning capabilities.

5.3.2 Data standardization

All of the input datasets were extracted from 40 dispersed locations in the Mitidja plain, and they were standardized to eliminate the dimensional influence, using the following formula:

$$X_y = \frac{(x_y^0 - {x_{ymi}}_n^0)}{({x_{yma}}_x^0 - {x_{ymi}}_n^0)} \tag{5.1}$$

Where:
"x_y" is the actual value of the parameter.
"x_{ymin}" is the minimum value of the parameter.
"x_{ymax}" is the maximum value of the parameter.

5.3.3 Pearson correlation matrix

The Pearson correlation coefficient (correlation matrix) is a statistical measure that describes the linear correlation between the parameters; it is calculated by the covariance to standard deviation ratio, varying from -1 to +1, with a positive correlation indicating that the ranks of both parameters increase together, while a negative correlation indicates that the rank of one variable increases while the other one decreases (Abdallah, 2007).

In our case study, we obtained the correlation matrix shown in Figure 5.12:

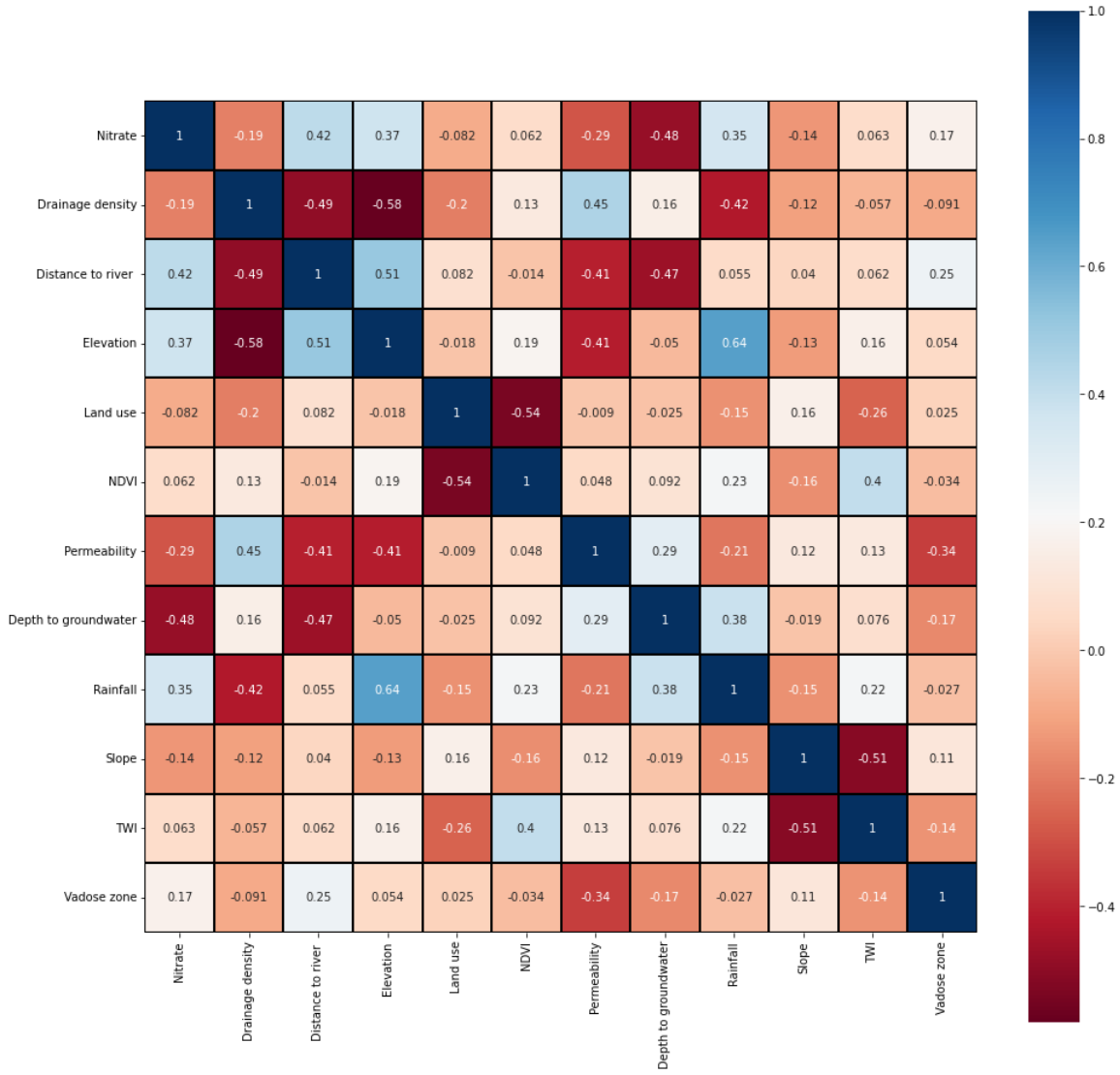

Figure 5.12: Pearson correlation matrix of the parameters

The correlation matrix shows a strong correlation (equals to 0.64) between the soil elevation and rainfall. The amount of precipitation increases with soil elevation. This is due to meteorological conditions as well as topography (Sasaki and Kurihara, 2008).

Having highly correlated parameters causes a co-linearity problem that may decrease the model's accuracy and quality. As a solution, we decided to eliminate the soil elevation parameter rather than the Rainfall parameter.

5.4 Modelling

The modelling process of the two ML models (AdaBoost and Random Forest) is based on splitting the data set. 70% of the dataset was used for the training, and the remaining 30% was used in order to test the ML models.

5.4.1 Training and validation

The dataset was splited randomly into two sub-datasets. The rescaled NO_3^- concentrations were used as an output label, while the ten parameters were input variables.

5.4.2 Testing

Confusion matrix

The results below are obtained from the application of the two Ml algorithms with our data.

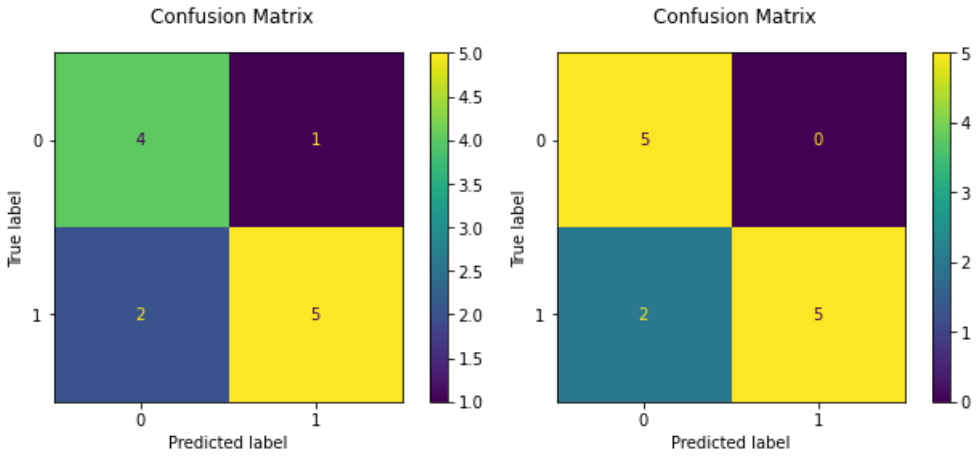

Figure 5.13: Confusion matrix of the both Random Forest and AdaBoost ML models

The results of the confusion matrix are resumed in table 5.1:

Table 5.1: Results of confusion matrix on the both Ml models

	Adaboost	Random Forest
TP	5	5
TN	5	4
FP	0	1
FN	2	2

From the table above, the actual positive values are successfully predicted in the two Ml models, while the difference between them lies in one value that is predicted as positive instead of negative one.

Receiver Operating Characteristics (ROC) curve

The ROC results for the ten input parameters using the Random Forest and Adaboost algorithms are shown in Figure 5.14.

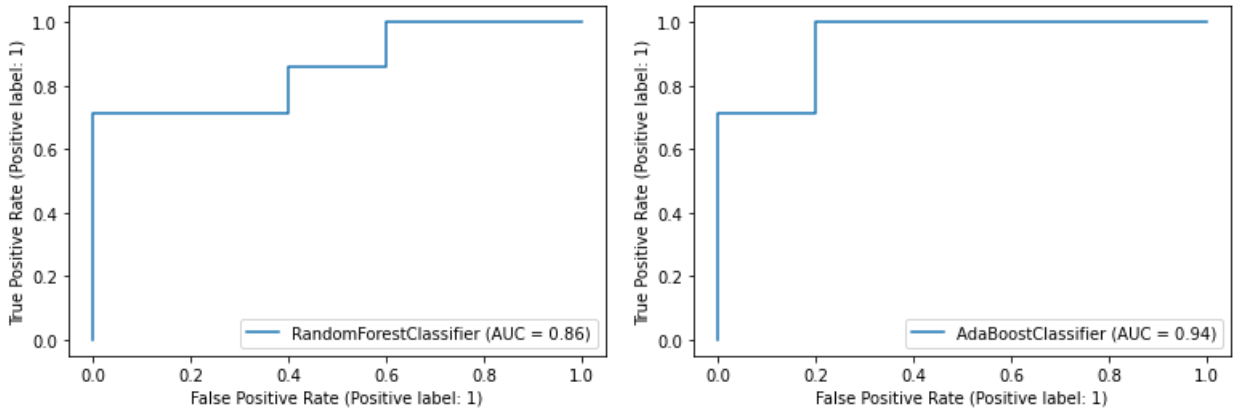

Figure 5.14: ROC curve of both Adaboost and RF ML models

The Adaboost classifier algorithm showed higher accuracy with 94% compared to the Random

Forest classifier algorithm with 86%. Therefore, RF and Adaboost models produce excellent prediction performance, with a slight advantage to Adaboost.

Based on these results, the Adaboost model performs so well in determining groundwater vulnerability. This is due to its ability for learning non-linear relationships between NO_3^- concentrations and explanatory variables used in this study.

5.5 Mapping groundwater vulnerability to nitrate

The vulnerability mapping requires the important features determined by the ML model, which are regrouped in the table 5.2:

Table 5.2: Importance Features of both Ml models

Factors	Adaboost	Random Forest
Drainage density	8	6
Distance to rive	5	5
Land use	10	10
NDVI	9	9
Permeability	3	3
Depth to groundwater	1	2
slope	4	8
TWI	7	7
Vadose zone	6	4
Rainfall	2	1

The results show that permeability, depth to groundwater, and rainfall are the most influential parameters on groundwater vulnerability to contamination.

The areas of the five classes of the vulnerability to contamination map are represented in table 5.3:

Table 5.3: Statistics of the groundwater vulnerability surface area

Vulnerability classes	Adaboost (% surface area)	Area(Km^2)	Random Forest (% surface area)	Area (Km^2)
very low	13.84	85.69	16.83	104.24
Low	25.90	160.37	31.43	194.68
Medium	22.77	140.97	30.87	191.21
High	23.57	145.93	18.77	116.23
very High	13.93	86.26	2.10	13.02

5.5.1 Vulnerability map obtained using using Random Forest

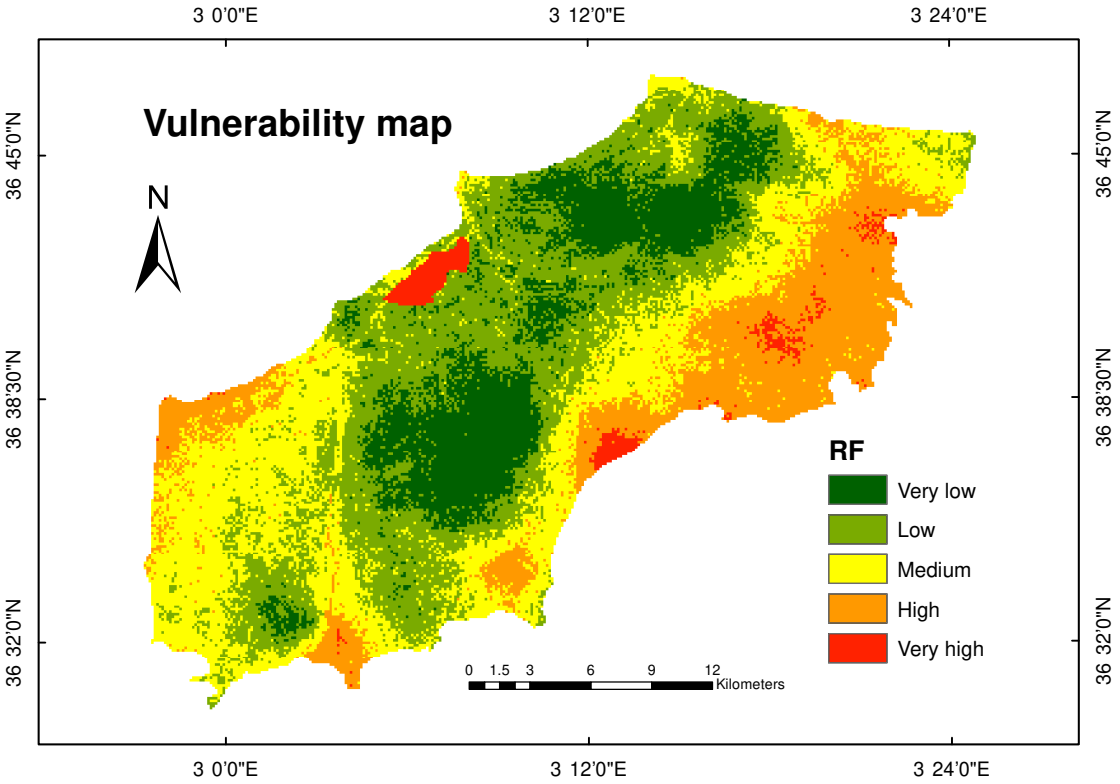

Figure 5.15: Vulnerability map using Random Forest ML model

The Vulnerability map obtained using the Random Forest ML model (Figure. 5.15) is divided into five classes. It shows that high to very high vulnerable areas cover respectively 18.77% and 2.10% of the Mitidja plain, located mostly in the northwest, including Douera and Birtouta, small areas of Harrach and Baraki, and the south at Bougara, Chebli, Larbaa, and HammamMelouane, Meftah, southeast of Hammedi, Khmis el KHechna, OuledMoussa, OuledHadjadj, Reghaia, and Boudouaou. The medium vulnerability area includes Birtouta, Boufarik, Bouinan, OuledChebel, Chebli, Bougara, south of OuledSlama, and Larbaa in the west, and Meftah, northwest Hammedi, Boudouaou, Boudouaou El Bahri, parts of Reghaia, south east of Harraoua, Rouiba, and Bordj El Bahri in the east. While 48.26% of the total area shows low to very low vulnerability of the groundwater.

5.5.2 Vulnerability map obtained using using Adaboost

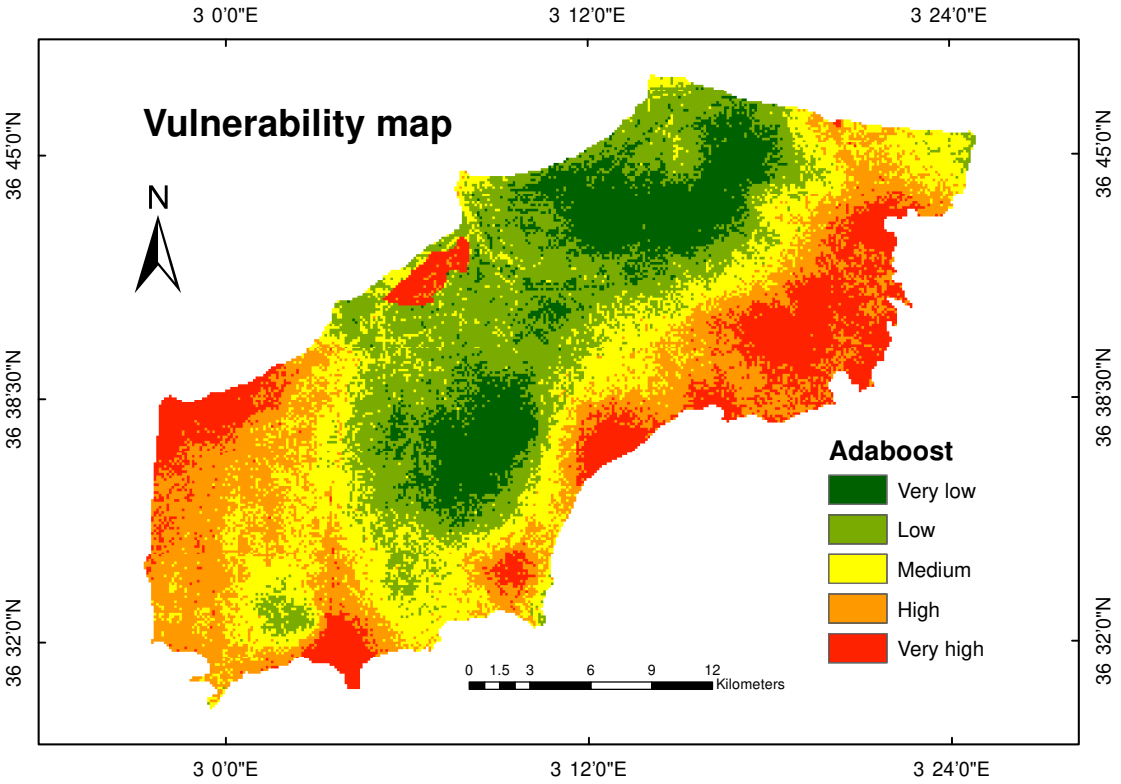

Figure 5.16: Vulnerability map using AdaBoost ML model

The vulnerability map obtained using the Adaboost ML model (Figure.5.16) shows that high to very high vulnerable areas cover respectively 23.57% and 13.93% of the Mitidja plain, located mostly in the south at Reghaia, OuledMoussa, OuledHaddadj, Khemis El Khechna, south Hammadi, south Meftah, south Larbaa, and the west at Bougara, Boinan, Boufarik east, north west and south Chebli, OuledChebel, Birtouta, west of Harrach, and North Baraki. The medium vulnerability covers 22.77% of Mitidja's total area, stretching from the north at Bordj El-Bahri, east Harraoua, north Reghaia, southeast of Rouiba, north of Hammedi, Meftah, northeast of Larbaa, OuledSlama, Bougara, north-east and center of Chebli, and northeast and center of Chebli. The low vulnerability class occupies 39.74% of the total area.

5.5.3 The most influencing factors

Further interpretation of the most influencing factors will be made based on the vulnerability map prepared using the AdaBoost machine learning model, as it offered a better accuracy of results compared to the Random forest.

Rainfall

The rainfall is the most influential parameter on groundwater vulnerability to contamination. The map shows high rainfall quantities varying from 654 to 724.2 mm, raining on the west and south parts of the plain where high vulnerability is occurred.

Depth to groundwater

The depth of groundwater is the second most influential factor on groundwater vulnerability to contamination, and the south-east is highly vulnerable to contamination seems to be influenced by the shallower level of groundwater varying from 4 to 25 meters.

Permeability

The most permeable regions of 10^{-3} to 10^{-2} m/s are located in the north Mitidja plain between the east of El Harrach and the west of Beraki down to the limits of SidiMoussa and Bougara with Chebli.

5.6 Interpretation of human activities in the vulnerable areas

In order make a future plan and adopt prevention methods to protect Mitidja's groundwater; we have interpreted the land occupation of the high to very high vulnerable areas.

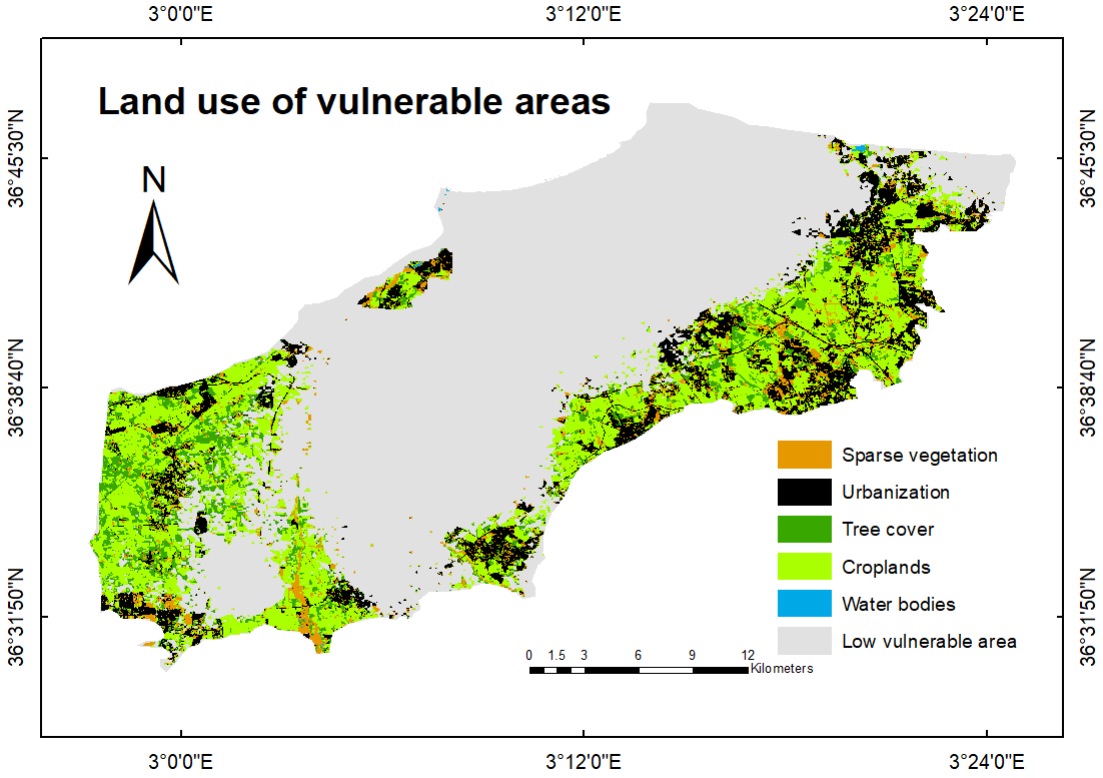

Figure 5.17: Landuse of high to very high vulnerable areas

The areas of each class of land use are presented in table 5.4:

Table 5.4: Landuse of vulnerable areas

Landuse	Area(%)	Area (Km^2)
Tree cover	14.83	34.37
Croplands	49.36	114.41
Urbanization	24.27	56.26
Sparse vegetation	11.41	26.44
Water bodies	0.13	0.31

The results show that most of the vulnerable areas are occupied by croplands with 49% of the total

vulnerable zone, farmers must be more cautious about the amount of fertilizers used in these areas, as they represent the main source of nitrate contamination. Urbanization occupies 24.27% of the total vulnerable area, which refers to human activities and urban discharges that must be better managed and transported away from vulnerable areas.

5.7 Conclusion

This chapter brings together the modeling process, starting with data preprocessing, which is an important step to ensure a high learning capability of the machine learning models.
In order to create a more accurate groundwater vulnerability map, which is a useful tool for monitoring, managing and protecting groundwater from contamination sources, the results of the Random Forest and AdaBoost machine learning models were compared using the Receiver Operating Characteristic (ROC) curve, with an area under the curve (AUC) of 86% and 94%, respectively.
The groundwater vulnerability map generated using the Adaboost ML model is more accurate with 94% AUC. It shows that it is high in areas with shallow water tables and permeable geological formations with high recharge rates. The map was then interpreted to determine anthropogenic activities in highly vulnerable areas based on the land use map.

General conclusion

Improving water management strategies required the development of a method for assessing the vulnerability of groundwater.The present study aimed to use two accurate models (Adaboost and RF) to assess the groundwater vulnerability to NO_3^- of the eastern Mitidja aquifer side.

The Nitrate concentration is classified into two classes according to the World Health Organisation (WHO) threshold value (50mg/l).

In each machine learning model (Adaboost and RF), ten (10) explanatory factors linked to groundwater vulnerability were included as features. These factors were Rainfall, Depth to groundwater, permeability, slope, Vadose zone, distance to rive, land use, Drainage density, TWI and NDVI. Furthermore, the study indicated that all explanatory variables used as NO_3^- groundwater pollution control factors are weighted differently. In fact, using the Adaboost ML model, the most influencing factors on groundwater were rainfall, depth to groundwater, and permeability.

The Area under the curve (AUC) of models, the Adaptive boosting and the Random forest are 86% and 94%, respectively. The performance results of the Adaboost model was better than the random forest model in assessing the groundwater vulnerability to Nitrate concentration (NO_3^-) in the Mitidja aquifer, Eastern side of basin. The more accurate vulnerability map would help decision-makers and environmental planners to produce sustainable development policies more efficiently regarding potential groundwater contamination, knowing that the cities of Algiers, Boumerdas, Blida, Tipaza are highly consuming the groundwater resources after the water scarcity since 2021.

The results revealed that about 23.57% and 13.93% of the total surface area are under high to very high vulnerability to nitrate (NO_3^-) and cover the towns of Reghaia, OuledMoussa, OuledHaddadj, Khemis El Khechna, south Hammadi, south Meftah, south Larbaa, and the west at Bougara, Boinan, Boufarik east, north west and south Chebli, OuledChebel, Birtouta, west of Harrach, and North

Baraki. While about 22.77% of the surface area of Bordj El-Bahri, east Harraoua, north Reghaia, southeast of Rouiba, north of Hammedi, Meftah, northeast of Larbaa, OuledSlama, Bougara, northeast and center of Chebli, and northeast and center of Chebli are in medium vulnerability.

Bibliography

Abdallah, Chadi (2007). "Application of remote sensing and geographical information system for the study of mass movements in Lebanon". PhD thesis. Université Pierre et Marie Curie-Paris VI.

Ahada, Chetan PS and Surindra Suthar (2018a). "A GIS based DRASTIC model for assessing aquifer vulnerability in Southern Punjab, India". In: *Modeling Earth Systems and Environment*.

Ahada, Chetan PS and Surindra Suthar (2018b). "A GIS based DRASTIC model for assessing aquifer vulnerability in Southern Punjab, India". In: *Modeling Earth Systems and Environment*.

Algeria press service. `https://www.aps.dz/`.

Aller, Linda (1985). *DRASTIC: a standardized system for evaluating ground water pollution potential using hydrogeologic settings*. Robert S. Kerr Environmental Research Laboratory, Office of Research and ...

Aymé, Antonin, Louis Glangeaud, and Jean Magne (1954). "Sur la stratigraphie du Crétacé de la feuille de Tablat". In: *COMPTES RENDUS HEBDOMADAIRES DES SEANCES DE L ACADEMIE DES SCIENCES*.

Band, Shahab S et al. (2020). "Comparative analysis of artificial intelligence models for accurate estimation of groundwater nitrate concentration". In: *Sensors*.

Body, Richard et al. (2017). "Troponin-only Manchester Acute Coronary Syndromes (T-MACS) decision aid: single biomarker re-derivation and external validation in three cohorts". In: *Emergency Medicine Journal*.

Canter, Larry W (2019). *Nitrates in groundwater*.

Carlston, Charles William (1963). *Drainage density and streamflow*.

Corniello, Alfonso, Daniela Ducci, and Gennaro Maria Monti (2004). "Aquifer pollution vulnerability in the Sorrento peninsula, southern Italy, evaluated by SINTACS method". In: *Geofísica Internacional*.

Council, National Research et al. (1993). *Ground water vulnerability assessment: Predicting relative contamination potential under conditions of uncertainty*. National Academies Press.

Derdous, Oussama et al. (2020). “A monitoring of the spatial and temporal evolutions of aridity in northern Algeria”. In: *Theoretical and Applied Climatology*.

Doerfliger, N, P-Y Jeannin, and François Zwahlen (1999). “Water vulnerability assessment in karst environments: a new method of defining protection areas using a multi-attribute approach and GIS tools (EPIK method)”. In: *Environmental geology*.

ESA worldcover. `https://worldcover2020.esa.int/`.

Filion, Éric (2017). “La problématique des nitrates dans l’eau souterraine en milieu agricole: du sol à l’aquifère: cas de l’aire d’alimentation de l’eau souterraine de la municipalité de Sainte-Luce, Québec, Canada”. PhD thesis. Université du Québec à Rimouski.

Foster, Stephen SD and Ricardo César Aoki Hirata (1988). “Groundwater pollution risk assessment; a methodology using available data”. In: *Groundwater pollution risk assessment; a methodology using available data*.

Géron, Aurélien (2019a). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*.

Géron, Aurélien (2019b). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. " O’Reilly Media, Inc.".

Ghosh, Paban and Kabita Lepcha (2019). “Weighted linear combination method versus grid based overlay operation method–A study for potential soil erosion susceptibility analysis of Malda district (West Bengal) in India”. In: *The Egyptian Journal of Remote Sensing and Space Science*.

Glangeaud, Louis (1952). “Tectonophysique comparee des chaines telliennes et rifaines”. In: *Bulletin de la Société Géologique de France*.

Goldscheider, NICO et al. (2000). “The PI method–a GIS-based approach to mapping groundwater vulnerability with special consideration of karst aquifers”. In: *Z Angew Geol*.

Guyot, Gérard and Xing-Fa Gu (1994). “Effect of radiometric corrections on NDVI-determined from SPOT-HRV and Landsat-TM data”. In: *Remote Sensing of Environment*.

Hasiniaina, Fanomezantsoa, Jianwei Zhou, and Luo Guoyi (2010). “Regional assessment of groundwater vulnerability in Tamtsag basin, Mongolia using drastic model”. In: *J Am Sci*.

Hölting, Bernward et al. (1995). “Konzept zur Ermittlung der Schutzfunktion der-Grundwasserüberdeckung”. In:

Ibe, KM, GI Nwankwor, and SO Onyekuru (2001). “Assessment of ground water vulnerability and its application to the development of protection strategy for the water supply aquifer in Owerri, Southeastern Nigeria”. In: *Environmental monitoring and assessment*.

infoclimat. `https://www.infoclimat.fr/`.

Ji, Yujie et al. (2020). “Seasonal variation of drinking water quality and human health risk assessment in Hancheng City of Guanzhong Plain, China”. In: *Exposure and health*.

Ju, Lei et al. (2018). “An adaptive Gaussian process-based iterative ensemble smoother for data assimilation”. In: *Advances in water resources*.

Ki, Min-Gyu et al. (2015). “Temporal variability of nitrate concentration in groundwater affected by intensive agricultural activities in a rural area of Hongseong, South Korea”. In: *Environmental Earth Sciences*.

Kumar, Akshay and Akhouri Pramod Krishna (2020). “Groundwater vulnerability and contamination risk assessment using GIS-based modified DRASTIC-LU model in hard rock aquifer system in India”. In: *Geocarto International*.

Lahjouj, Abdelhakim, Abdellah El Hmaidi, Karima Bouhafa, et al. (2020). “Mapping specific groundwater vulnerability to nitrate using random forest: case of Sais basin, Morocco”. In: *Modeling Earth Systems and Environment*.

Li, Peiyue et al. (2021). “Sources and consequences of groundwater contamination”. In: *Archives of environmental contamination and toxicology*.

Machiwal, Deepesh et al. (2018). “Assessment and mapping of groundwater vulnerability to pollution: Current status and challenges”. In: *Earth-Science Reviews* 185, pp. 901–927.

Mackay, Douglas M, Paul V Roberts, and John A Cherry (1985). “Transport of organic contaminants in groundwater”. In: *Environmental science & technology*.

Mahesh, Batta (2020). “Machine learning algorithms-a review”. In: *International Journal of Science and Research (IJSR).[Internet]*.

MIFDAL, Rachid (2019). “MIFDAL,2019”. In:

Muralikrishna, Iyyanki V. and Valli Manickam (2017). “Chapter Eleven - Principles and Design of Water Treatment”. In: *Environmental Management*. Ed. by Iyyanki V. Muralikrishna and Valli Manickam. Butterworth-Heinemann.

Nampak, Haleh, Biswajeet Pradhan, and Mohammad Abd Manap (2014). “Application of GIS based data driven evidential belief function model to predict groundwater potential zonation”. In: *Journal of Hydrology*.

Price, Michael and William Back (2013). *Introducing groundwater*.

Ribeiro, Luís, Juan Carlos Pindo, and Luis Dominguez-Granda (2017). “Assessment of groundwater vulnerability in the Daule aquifer, Ecuador, using the susceptibility index method”. In: *Science of the total environment*.

Rivoirard, R (1952). “Data of hydrology Algeria”. In: *Overview of the hydrogeology of the Mitija, Geology and water problems in Algeria*.

Sajil Kumar, PJ (2020). “Hydrogeochemical and multivariate statistical appraisal of pollution sources in the groundwater of the lower Bhavani River basin in Tamil Nadu”. In: *Geology, Ecology, and Landscapes*.

Salmani, Kimia (2013). “Multiple-instance active learning with online labeling”. PhD thesis.

Saranya, Thiyagarajan and Subbarayan Saravanan (2021). “Evolution of a hybrid approach for groundwater vulnerability assessment using hierarchical fuzzy-DRASTIC models in the Cuddalore Region, India”. In: *Environmental Earth Sciences*.

Sasaki, Hidetaka and Kazuo Kurihara (2008). “Relationship between precipitation and elevation in the present climate reproduced by the non-hydrostatic regional climate model”. In: *SOLA*.

Sekkal, Rached (1986). “Hydrologie de la nappe de la Mitidja (Algérie): étude hydrodynamique des champs captants de la ville d’Alger”. PhD thesis. Université Scientifique et Médicale de Grenoble.

Singh, Anjali et al. (2015). “A modified-DRASTIC model (DRASTICA) for assessment of groundwater vulnerability to pollution in an urbanized environment in Lucknow, India”. In: *Environmental earth sciences*.

Winn, Robert Maurice (1973). “Hydrogeology of the Albian Formation, Algeria”. PhD thesis. Texas Tech University.

WorldClim.

Xu, Jianfeng, Yuanjian Zhang, and Duoqian Miao (2020). “Three-way confusion matrix for classification: A measure driven view”. In: *Information sciences*.