Ministry of Higher Education and Scientific Research
University of Hassiba Benbouali, Chlef
Faculty of Civil Engineering and Architecture
Department of Hydraulic

THESIS

Presented for obtaining the diploma of

DOCTORATE

Field: Hydraulic

Specialty: Urban Hydraulic

By

Amina TACHI

Theme:

STUDY OF THE SPATIO-TEMPORAL CORRELATION BETWEEN ELECTRICAL CONDUCTIVITY - TDS IN ALGERIAN GROUNDWATER

Presented on 24/06/2025 in the front of the following doctoral:

Yamina ELMEDDAHI	Professor	University of Hassiba Benbouali	President
Mehdi METAICHE	Professor	University of Bouira	Promotor
Abderahmen MESSOUL	MCB	University of Hassiba Benbouali	Co-Promotor
Housseyn BOUZERIA	MCA	University of Saad Dahlab	Examiner
Taieb HADBI	MCA	University of Hassiba Benbouali	Examiner

ACKNOWLEDGMENTS

*First of all, I would like to express my gratitude to the Almighty **Allah**, who has enlightened my path and provided me with courage, willpower and patience to successfully complete this task.*

I would like to express my sincere thanks to the people who have made a significant contribution to the completion of this doctoral thesis.

*I would like to express my deep gratitude to my thesis director**, Professor Metaiche Mehdi**. This moment is an opportunity for me to renew my greatest esteem and gratitude to you. I would like to express my most sincere thanks for your exceptional commitment and your enlightened guidance throughout the completion of this thesis. I am grateful for the opportunity given to me to benefit from your attentive direction. Sir, without your presence, the work would not have been as successful. It is difficult for me to express all my gratitude, so I simply say thank you.*

*My sincere thanks also go to my Co-supervisor, **Dr. Messoul Abd errahmene**.*

*I am deeply honored by the presence of **Dr. El Meddahi Yamina** who graciously accepted to serve as the Chair of this thesis committee. I extend my sincere gratitude to all the esteemed members of the jury for their commitment and for taking the time to evaluate this work despite their demanding professional responsibilities. Their insightful and constructive remarks will undoubtedly contribute to strengthening the scientific rigor of this research. I wish to express particular appreciation to **Dr. Bouzeria Housseyn** from the University of Blida and **Dr. Hadbi Taieb** from UHB Chlef.*

*I do not forget to express my gratitude to **Dr Bouguerra Hamza**, **Pr. Djabri Larbi**, **Dr Tachi Salah Eddine**, **Dr. Ben nouara Nawel**, whose valuable collaboration greatly contributed to the enrichment of my work. Their vision and advice were invaluable asset and who made a lot of efforts to help me.*

I would like to express my gratitude to all my teachers and colleagues throughout my university career. In addition, I would like to warmly thank all the people who contributed directly or indirectly to the completion of this work.

TACHI AMINA

DEDICATIONS

As I approach the completion of this research work, I would like to express my deep gratitude to those who have been the pillars of my academic and personal journey. Their unwavering support has been the cornerstone of my successes and the driving force that has propelled my quest for knowledge.

It is with pride and respect that I dedicate this modest work:

*To **my father**, who has made countless efforts day and night for my education and well-being. This work is a testament to the considerable sacrifices you have made for my education and training.*

*To **my beloved mother**, you are the supreme example of dedication, as you have never ceased to encourage me and pray for me. May the blessings of God, the Almighty, preserve you and grant your health, longevity and happiness.*

*To **my brothers** and **my husband** and **sisters in law**, companions at every stage of my life. Your encouragement has been the source of my strength.*

*To **my nieces and nephew**, whose innocence and joy of life have brightened my darkest days.*

*To **my friends**, companions in stimulating discussions and moments of relaxation. Your moral support has been an essential element for my personal and professional balance.*

Together, you have formed the solid foundation on which this achievement rests. My words of thanks seem modest compared to the depth of my gratitude. May this dedication be a humble testimony of the love and gratitude that I feel for each of you.

TACHI AMINA

الملخص

تُعَدُّ الموارد المائية ضرورية للحياة، والتنمية الاقتصادية، والصحة، والبيئة. وتشكل المياه الجوفية مصدرًا رئيسيًا للمياه العذبة، كما أن جودتها أساسية لمياه الشرب والري والاستخدامات المنزلية. ويُعتبر تقييم جودة المياه الجوفية ونمذجتها زمانيًا ومكانيًا من بين أهم المهام لتحسين استراتيجيات إدارة الموارد المائية ودعم الأنشطة البيئية. ويُعد مجموع المواد الصلبة الذائبة مؤشرًا مهمًا لتحديد جودة المياه الجوفية ومدى ملاءمتها للري، حيث يتأثر بمحتوى الأملاح المعدنية والتصريف. أما الطرق التقليدية لحساب المواد الصلبة الذائبة الكلية فهي تستغرق وقتًا طويلًا وتكون عرضة للأخطاء، خصوصًا في الدول النامية. غير أن نماذج التعلم الآلي توفر حلولًا أسرع وأكثر جدوى من حيث التكلفة، من خلال التنبؤ بجودة مياه الري في الأنظمة الجوفية اعتمادًا على الخصائص الفيزيائية والكيميائية
في هذا البحث، تم تطبيق أربعة نماذج للتعلم الآلي، وهي: شجرة القرار، الغابة العشوائية، التعزيز التصنيفي، والانحدار
بواسطة آلات الدعم الناقل، للتنبؤ بقيم المواد الصلبة الذائبة وتحديد العلاقة بين الخصائص الفيزيائية-الكيميائية للمياه الجوفية، بما في ذلك الناقلية الكهربائية، والعناصر الرئيسية مثل الكالسيوم، المغنيسيوم، الصوديوم، البوتاسيوم، الكبريتات، البيكربونات، النترات، بالإضافة إلى العسرة الكلية، والتي تؤثر على تراكيز المواد الصلبة الذائبة في ثلاثة أنظمة جوفية مختلفة بالجزائر. وقد جُمعت البيانات من الآبار بواسطة الوكالة الوطنية للموارد المائية في الجزائر، ثم قُسّمت عشوائيًا إلى مجموعتين فرعيتين في كل واحدة من مناطق الدراسة الثلاث: بيانات تدريب النماذج وبيانات اختبار للتحقق من صحتها. في البداية، جرى تطوير هذه النماذج بافتراض تركيبات مختلفة من الخصائص الفيزيائية-الكيميائية بالاعتماد على التحليل الارتباطي وتحليل الحساسية. ولتقليل عدد المتغيرات، جرى تحليل أهميتها باستخدام خوارزميات الغابة العشوائية، مما سمح باختيار أفضل المدخلات لتقدير المواد الصلبة الذائبة. بعد ذلك، جرت نمذجة متغيرات الجودة باستخدام النماذج المختارة للتعلم الآلي؛ وأخيرًا، تم تقييم أداء النماذج باستخدام الجذر التربيعي لمتوسط مربع الأخطاء ومعامل كفاءة ناش–سوتكليف.

وأظهرت المقارنات بين القيم المرصودة والمتوقعة أن نماذج شجرة القرار، التعزيز التصنيفي، الغابة العشوائية، والانحدار بواسطة آلات الدعم الناقل قد نجحت في التنبؤ بتغيرات المواد الصلبة الذائبة، حيث تفوق نموذج الانحدار بواسطة آلات الدعم الناقل على بقية النماذج في جميع الأحواض الجوفية المدروسة، مع اختلاف التركيب في كل منطقة. وقد أثبتت هذه النماذج كفاءة عالية في المحاكاة الزمانية والمكانية لجودة المياه الجوفية، مما يبرز التعلم الآلي كأداة واعدة للنمذجة باستخدام المعايير الهيدروكيميائية من أجل تحقيق إدارة مستدامة للمياه الجوفية.

الكلمات المفتاحية: المياه الجوفية، المواد الصلبة الذائبة الكلية، الانحدار بواسطة آلات الدعم الناقل، شجرة القرار، الناقلية الكهربائية، العناصر الرئيسية، الجزائر

RÉSUMÉ

Les ressources en eau sont essentielles à la vie, au développement économique, à la santé et à l'environnement. Les eaux souterraines sont une source majeure d'eau douce et leur qualité est vitale pour la boisson, l'irrigation et les usages domestiques. L'évaluation et la modélisation de la qualité spatiotemporelle des eaux souterraines comptent parmi les tâches les plus importantes pour améliorer les stratégies de gestion des ressources en eau et soutenir les activités environnementales. Les solides dissous totaux (TDS) sont un paramètre important pour déterminer la qualité des eaux souterraines et l'irrigation. Bien que leur pertinence soit influencée par la teneur en sels minéraux et le débit, les méthodes traditionnelles de calcul des TDS prennent du temps et sont sujettes aux erreurs, en particulier dans les pays en développement. Cependant, les modèles d'apprentissage automatique fournissent des solutions plus rapides et plus rentables en prévoyant la qualité de l'eau d'irrigation des systèmes aquifères à l'aide de paramètres physico-chimiques. Dans le présent travail, quatre modèles d'apprentissage automatique, à savoir Decision tree DT, Random Forest RF, Catboost et support victor regression SVR, ont été appliqués pour prévoir les paramètres TDS et déterminer la relation entre les paramètres physico-chimiques des eaux souterraines, notamment la conductivité électrique (EC), les éléments majeurs (Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, SO_4^{2-}, HCO_3^{-}, NO_3^{-}) et la dureté totale (TH), qui affectent les TDS dans trois aquifères souterrains différents en Algérie. Les données, collectées à partir de forages par l'Agence Nationale des Ressources Hydrauliques (ANRH) en Algérie, ont été aléatoirement réparties en deux sous-ensembles dans chacune des trois zones d'étude considérées: les données d'entraînement du modèle et les données de test pour la validation du modèle. Initialement, ces modèles ont été développés en supposant diverses combinaisons de paramètres physico-chimiques à l'aide d'une analyse de corrélation et de sensibilité. Pour réduire le nombre de caractéristiques, l'importance des caractéristiques a été analysée à l'aide des algorithmes de forêt aléatoire, en sélectionnant les meilleures entrées pour l'estimation des TDS. Les variables de qualité ont ensuite été modélisées à l'aide de modèles d'apprentissage automatique sélectionnés ; Enfin, les performances du modèle ont été évaluées à l'aide de l'erreur quadratique moyenne (RMSE) et de l'efficacité Nash-Sutcliffe (NS). Sur la base de comparaisons entre les valeurs observées et prédites, les résultats ont montré que les modèles DT, Catboost, RF et SVR pouvaient prévoir avec succès les variations de TDS, le SVR surpassant les autres dans tous les aquifères étudiés, avec une combinaison différente dans chaque zone. Ces modèles ont démontré une grande efficacité dans la simulation spatiotemporelle de la qualité des eaux souterraines, mettant en

évidence l'apprentissage automatique comme un outil prometteur pour la modélisation par paramètres hydrochimiques pour parvenir à une gestion durable des eaux souterraines.

Mots-clés : Eaux souterraines, Solides dissous totaux, régression à vecteur de support, Catboost, Arbre de décision, Forêt aléatoire, Conductivité électrique, Élément majeur, Algérie

ABSTRACT

water resources are critical for life, economic development, health and the environment. Groundwater is a major freshwater source, and its quality is vital for drinking, irrigation and domestic purposes. Spatiotemporal groundwater quality evaluation and modeling are among the most significant tasks for improving water resource management strategies and supporting environmental activities. Total dissolved solids (TDS) is an important parameter for determining groundwater quality and irrigation. appropriateness, it is impacted by mineral salt content and discharge. Traditional TDS calculations methods are time-consuming and error-prone, particularly in developing countries. However, machine learning models provide faster, more cost-effective solutions by forecasting irrigation water quality of aquifer systems using physicochemical parameters. In the present work, four machine learning models, namely Decision tree DT, Random Forest RF, Catboost and support victor regression SVR, were applied to forecast TDS parameters and determine relationship between groundwater physicochemical parameters including electrical conductivity (EC), major elements (Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, SO_4^{2-}, HCO_3^{-}, NO_3^{-}) and total hardness (TH), which affect TDS at three different groundwater aquifers in Algeria. The data collected from wells by the National Water Resources Agency (ANRH) in Algeria were randomly split into two subsets within each of the three study areas considered in this work: model-training data and testing data for model validation. Initially, these models were developed by assuming various combinations of physicochemical parameters using correlation and sensitivity analysis. To reduce the number of features, feature importance was analyzed using the random forest algorithms, selecting the best inputs for TDS estimation. The quality variables were then modeled using selected machine learning models; Finally, the model's performances were evaluated using the Root Mean Square Error (RMSE) and Nash – Sutcliffe efficiency (NS). Based on comparisons between observed and predicted values, the results showed that DT, Catboost, RF and SVR models could successfully forecast TDS variations, with SVR outperforming the others in all studied aquifers, with different combination in each area. These models demonstrated high efficiency in the spatiotemporal simulation of groundwater quality, highlighting machine learning as a promising tool for modeling by hydrochemical parameters to achieve sustainable groundwater management.

Keywords: Groundwater, Total Dissolved Solid, support vector regression, Catboost, Decision tree, Random Forest, Electrical conductivity, Major element, Algeria.

TABLE OF CONTENTS

CHAPTER II: THE STUDIED AREAS

CHAPTER III: MATERIALS AND METHODS

CHAPTER IV: RESULTS AND DISCUSSION

Conclusion

LIST OF FIGURES

LISTE OF TABLES

INTRODUCTION

Water is important for life on earth, it is essential resource that supports the survival and development of all living beings; It is not only sustains living organisms but also for maintains the balance and functionality of the ecosystems and climate worldwide, it plays a vital role in sustaining ecological balance and the providing habitats for a wide range of plants and animals, making it an essential component for sustaining life.

Within these resources, groundwater is acknowledged as one of the most precious and necessary natural sources of freshwater globally for human and agriculture intake. Moreover, it represents the second largest water resource in terms of use after surface water. Globally, groundwater is a major source of drinking water for around 2 billion people [1], and about 278.8 million hectares' agricultural lands are irrigated using groundwater resources [2].

Groundwater demand is projected to rise in the future as populations grow and economies develop [3]. Furthermore, the high quality of groundwater, with its low pollution levels and widespread availability, has made it a popular choice for human consumption, leading to its widespread use around the world [4]. Evaluating groundwater quality is a crucial component of water management methods, as it guarantees the efficient employment of groundwater resources [5,6]. Groundwater quality is commonly assessed using various parameters such as pH, turbidity, major ions, temperature, electrical conductivity (EC), and total dissolved solids (TDS), among others, which are indicative of water quality. These parameters are known to exhibit considerable spatial and temporal variations, and hence must be evaluated thoroughly to assess groundwater quality of the various parameters, TDS and EC are crucial in evaluating groundwater quality, as they provide key information on dissolved minerals and the overall water quality, including salinity and hardness [7].

Countries worldwide are currently facing significant challenges related to water resources. [1]. Algeria, located in North Afric, is among the countries most affected by water scarcity, since the Sahara Desert dominates the majority of its territory. It is impacted not only by its severe climate and geography, but also by the alarming overexploitation of its water resources. Improving the quality of water management and services is one of the government's primary goals, as many

forecasts predict that climate change will lead to more frequent droughts and floods, which would adversely affect Algeria [8]. Algeria's water stress index is poor, with a water availability ratio of 411m^3 per capita per year, and forecasting researchers estimate this will decline to 320 m^3 per capita per year by 2030 and 300m^3 per capita per year by 2050, ranking Algeria in the category of absolute water scarcity. [9]. Algeria has diverse water resources, including surface water and groundwater that have been exploited for various purposes, such as domestic, drinking, agricultural, industrial and environmental uses [10]. In recent years, water consumption in Algeria has risen significantly due to rapid population growth, economic development, urban expansion, climate change impacts, and ongoing industrial transformation [10]. Furthermore, the fast expansion in agricultural activities, along with other human and domestic behaviors, has contributed to the degradation of Algerian water bodies utilized for drinking purposes, as well as the irregular and uncontrolled extraction of surface and groundwater. This situation leads to a reduction in the country's per capita water availability to below 300 m^3 per y. [10]. With increasing water demands and the scarcity and unpredictability of surface water, groundwater has emerged as the primary and most important source for agricultural, household, and industrial activities across the country. Consequently, it is important to focus on evaluating groundwater quality.

Total dissolved solids (TDS) serves as an important indicator in establishing the suitability of groundwater for drinking and is frequently used as a proxy for assessing water quality. Precisely estimating TDS amounts is critical for efficient management of freshwater resources, since elevated TDS in groundwater can cause salinization, making it unsuitable for both drinking purposes and irrigation [11]. High TDS levels in groundwater also pose significant challenges in agriculture, as they can adversely affect soil quality and crop growth. Therefore, understanding the relationship between mineral salt composition and TDS is essential for effective management of water resources, as variations in mineral composition and discharge can significantly influence TDS levels [12]. Total dissolved solids (TDS) mainly consist of inorganic salts, containing anions such as nitrate (NO^{3-}), sulfate (SO_4^{2-}), chloride (Cl^-), and bicarbonate (HCO_3^-), as well as cations like magnesium (Mg^{2+}), sodium (Na^+), potassium (K^+), and calcium (Ca^{2+}). These salts stem from natural sources, such as rocks and minerals, and anthropogenic activities, including

agriculture and industry. Aside from inorganic salts, TDS encompasses dissolved organic elements that contribute to overall water quality.

The World Health Organization (WHO) established guidelines [13] for measuring TDS in water, offering a standard to maintaining water quality. While laboratory, research, and computational approaches are frequently used to measure TDS levels, traditional lab testing and manual computations tend to be time-consuming, Prone to errors, and may not provide the accuracy required for precise measurements. Additionally, these procedures are often costly [14, 15].

Modeling the concentrations of total dissolved solids (TDS) is essential for agricultural and hydrological management research [16]. Numerous mathematical models have been extensively applied to predict and estimate TDS, with different methodologies documented in the literature. Among the notable used techniques are Sorensen's method (1977) [17], which calculates TDS by summing ion concentrations in water and the method developed by D. Hem (1985) [18], which estimates TDS using a linear regression equation based on electrical conductivity. These approaches offer valuable insights into the quality of groundwater and its suitability for various applications.

Although mathematical models are capable of predicting TDS concentrations accurately, they are often complex and require significant laboratory sampling and data collection. Moreover, these models typically rely on a large number of parameters, which can be time-consuming to obtain. Despite their potential accuracy, previous studies have indicated that numerical models may sometimes produce inaccurate results due to the complex hydraulic conditions of groundwater [11]. In developing countries, the use of models that rely on data or basic information can be challenging for users, as these may be lacking in such contexts. Moreover, in cases where there are no linear, stochastic, or temporal relationships among various water quality parameters, it can be difficult to develop mathematical models using traditional approaches to predict events in such circumstances [19, 20].

Maintaining groundwater quality sustainably the development of novel and efficient methods for evaluating and monitoring parameters including TDS, EC, and major elements. To forecast future supplies and accurately predict water quality, robust and flexible models are

crucial, as they can help address the challenges involved in groundwater planning and management [21, 22, 23, 24, 20, 25]. By implementing such models, pollution and health risks can be minimized, and water resources can be effectively managed and sustained for the benefit of present and future generations.

During the past two decades, the application of Artificial Intelligence (AI) techniques has expanded in numerous fields, including water quality modeling, where algorithms have recently been explored as viable and effective methods. These algorithms analyze intricate and hidden correlations between input and output data, developing models that accurately reflect these correlations [20]. AI models have been used by researchers to predict the water quality variables, for example, the radial basis function neural network (RBFNN), adaptive neuro fuzzy inference system (ANFIS), multiple linear regression (MLR), support vector machine (SVM), random forest regression (RFR) and decision tree (DT). The literature indicates that these models are highly non-linear and flexible, which allows them to on large datasets across different time scales. Moreover, AI models tend to be less sensitive to missing data than classical methods [26,27,20].

In the context of water quality simulation and prediction, several studies have been carried out using AI techniques, including Lu et Ma (2020) [28], who predict six water quality indicators in the Tualatin river using Extreme Gradient Boosting (XGBoost) and Random Forest (RF) models. Meyers et al. (2017) [29] predicted the water turbidity of a trunk network in the United Kingdom using Artificial Neural Network (ANN), Random Forest (RF) and Support Vector Machine (SVM) models. Di et al. (2019) [30] assessed the quality of water in the Yangtze River in China using machine learning models that employ classification techniques. Similarly, Artificial Intelligence methods have become increasingly reliable in predicting groundwater levels through the application of various machine learning models. Rajaee et al. 2019 [31] utilized various ML models, including artificial neural networks (ANN) such as multilayer perceptron (MLP), backpropagation neural networks (BPNN), and feed-forward neural networks (FFNN), to estimate and predict groundwater salinity.

Comprehensive research has shown that machine learning (ML) models are quite accurate in predicting and assessing water quality, particularly for both surface and groundwater. However, it

is crucial to highlight that the performance of ML models relies not only on their predictive accuracy but also on the type and number of predictors employed in the model [25]

The aim of this thesis is to forecast the total dissolved solids in three different studied areas: The Upper Chellif, Mitidja and Hassi R'mel aquifers, by employing Random Forest, Decision Tree, Support Victor Regression and Categorical Boosting models to determine the optimum parameters for TDS modeling while evaluating the efficiency and the effectiveness of the different machine learning algorithms applied. For clarity of presentation, this thesis is organized into four chapters as follows:

The first chapter is devoted to the state of the art, focusing on an overview of global water resources, particularly in Algeria, and the definition of groundwater.

It also presents general information on groundwater quality, aquifer types, and techniques for measuring different characteristics of groundwater quality.

- The Second chapter provides a brief description of the study areas, including The Upper Chellif, Mitidja, and Hassi Rimel aquifers, as well as a presentation of the database used.
- The Third chapter provides an overview of machine learning (ML) and its construction, presents the adopted methodology for processing various ML models, and describes the data analysis procedures.
- Chapter four discusses the results, their interpretation, and comparisons across different models and study areas.
- Finally, the fifth chapter concludes the evaluation of our ML models forecasting Total dissolved solids (TDS) in the three studied aquifers and provides perspectives for enhancing groundwater resource utilization, improving water quality, and reducing analysis time and cost.

CHAPTER I

THE STATE OF THE ART

I. THE STATE OF THE ART

Introduction

Water resources, particularly groundwater, represent a fundamental component of life, health, and economic development. Their scarcity, uneven distribution, and vulnerability to pollution make them a central issue in both global and Algerian contexts. This chapter reviews the current state of knowledge on water resources, groundwater characteristics, aquifer types, and the key parameters used to assess and manage water quality.

I.1. Water resources in the World

Water, omnipresent and vital to the maintenance of life, is one of the most essential chemical substances on our planet. Its properties are also quite exceptional. Water covers about 72% of the Earth's surface, representing approximately 1.4 billion m³ of water on the planet [32]. However, the majority (97%) of this water is in the form of salt water in the seas and oceans, making it difficult to exploit for human activities. Of the remaining 3% (36 million km^3) of this water commonly called fresh water, more than ¾ (Approximately 2.15%) constitutes glaciers that are largely inaccessible. The remaining ¼ mainly comprises groundwater (less than 1% of the total water on the globe) and a small part in the form of surface water contained in lakes and rivers (i.e. 0.01% of the planet's water) [33, 34]. The main sources of raw water are rainwater, seawater, surface water and groundwater. Each of these sources has characteristics influenced by its interaction with the environment.

- **Rainwate**r: Rainwater is of good quality for human consumption, being soft and without dissolved salts such as magnesium and calcium. However, in industrialized regions, it can be contaminated by atmospheric dust. The scarcity of precipitation and the difficulties of capture limit its use by municipalities [35].
- **Seawater:** Seawater, very saline (33,000 to 37,000 mg/L), is used only in the absence of fresh water. Its treatment is complex and expensive, limiting its use [36].
- **Surface water:** Surface water, from groundwater or runoff, circulates or is stored on the surface of continents (rivers, streams, dams, ponds, backwaters). It is characterized by constant contact with the atmosphere and rapid circulation [37].

  Surface water requires treatment in several stages before being used for drinking and

domestic purposes. It is essential to avoid soil erosion, unsanitary conditions and pollution to guarantee its quality [38].

- **Groundwater:** Groundwater is water that circulates underground, filling rock fractures and pores in granular environments (sand, gravel). Unlike surface water, it moves deep into geological formations [39].

I.2. Water resources in Algeria

Algeria is a semi-arid, and in some regions, an arid country with an average annual rainfall ranging from 200 to 400 mm. It faces significant water scarcity [40, 41, 33]. The spatial distribution of water resources across the country is uneven, with most of them concentrated in the north. About 90% of all surface runoff occurs in the coastal region, which represents only 7% of the national territory. The remaining 10% is distributed between the High Plateaus and the Saharan basins. This scarcity is due to desertification and will be accentuated by the effects of climate change [42]. The average annual water volume in northern Algeria is estimated at between 95×10^9 and 100×10^9 m^3, more than 80×10^9 m^3 evaporate, 3×10^9 m^3 infiltrate, and 12.5×10^9 m^3 flow into watercourses [33]. The overall demand for water has increased considerably. It has quadrupled over the past forty years and now exceeds half of the volume of potentially mobilizable resources [43]. Algeria, like the 17 African countries affected by water stress, is in the category of the poorest countries in terms of water potential. Water management in Algeria poses a serious challenge for the authorities, as available resources are insufficient to meet national needs. The dilapidated state of the water supply networks and the deficient storage capacity hinder the proper distribution of water to consumers. The daily allocation per capita remains low compared to national standards.

I.3. Groundwater Overview

The term groundwater refers to water found beneath most of the Earth's land surfaces, within formations known as aquifers. Groundwater in aquifers is confined either within the fractures of compact rocks or in the interstitial spaces of porous soils [44]. The origin of groundwater is attributed to the infiltration of water into the soil, a process that varies according to the soil's porosity and geological structure, which have evolved over time. Aquifers play a vital role in supplying freshwater and represent the largest reservoirs of potable water. Groundwater is generally protected from surface pollution sources and is therefore of higher physicochemical and microbiological quality than surface water. It is extracted through well drilling or naturally emerges at the surface as springs [45, 44]. Groundwater is of paramount importance in most regions of

the world, where it serves as the main source of domestic, agricultural, and industrial water supply.

However, groundwater contamination has been identified as one of the most serious environmental problems in many countries [42]. According to the World Health Organization (WHO), safe drinking water is defined as water free from harmful materials, chemicals, or pathogens that can endanger health. It should also contain essential minerals in solution and may contain microorganisms as long as they pose no health risks or discomfort to the consumer [45]. Drinking water intended for human consumption mostly originates from groundwater and spring sources. According to [47], groundwater is typically considered bacteriologically pure and is often preferred over surface water for sanitation engineering purposes [48]. However, the natural quality of groundwater extends beyond health and technical considerations related to drinking water. As it also contributes to river flow, its chemical and microbiological properties can significantly influence aquatic ecosystems [42]. Therefore, both bacteriological and physicochemical qualities are essential. Through analysis, the quality of water and its potential uses can be determined, thereby enabling the adaptation of water unsuitable for consumption for purposes such as irrigation or fish farming.

Scientists assess groundwater quality by quantifying the concentrations of various components present in the water. These measurements are frequently represented in milligrams per liter (mg/L) [49]. All of these measurements, analyses, or observations are typically conducted in a laboratory or at a measuring station site. Standards are established to ensure the safety of potable water, aiming to protect consumers from potential harm. Compliance with these standards is crucial for those managing water resources [50]. These standards specify maximum levels for harmful substances that may be present in water [44], and their development is grounded in scientific knowledge and technological advancements, particularly on health risks and chemical analysis [51]. They are tailored to public health criteria, the country's level of development, and technological advancements, can be updated as needed; With the constant emergence of new substances in water and improvements in detection and measurement methods, new standards are regularly issued. Notably, different countries may have their own standards or adopt those recommended by the WHO, which accommodate the limitations of certain developing countries. Current standards primarily focus on ensuring water potability to fulfill objectives related to public health and consumer well-being [44].

I.3.1. The origin of groundwater

Most groundwater originates from meteoric sources, meaning it derives from precipitation such as rain or snow and seeps into the soil. In major aquifers, groundwater may have been recharged during past climatic periods, thus providing valuable insights into paleo climatic conditions.

Connate waters: found deep within the Earth's crust typically at depths ranging from 1 to 2 kilometers, originate from meteoric water reservoirs that interacted with the surrounding rocks. These waters often have higher salinity levels. They play a role in the hydrology of recently buried geological formations or in rocks with very low permeability, retaining water since their formation [52].

Juvenile waters: on the other hand, stem directly from magmatic activity at depth. Distinguishing them from deep-seated waters can be challenging. Magmatic processes may also release gaseous compounds such as CO2 alongside water [52].

I.4. The principal types of G.W aquifers

I.4.1. confined aquifers

Water in a confined aquifer is held under pressure higher than atmospheric pressure by impermeable (aquiclude) or semi-permeable (aquitard) layers. The confined aquifer is therefore isolated from the ground surface by an impermeable layer. As it is not directly recharged from the surface and occurs at considerable depths, it is less susceptible to pollution [53, 54]. When the water level in a well tapping a confined aquifer rises above the ground surface, it is referred to as a flowing artesian well; otherwise, it is simply called an artesian well.

I.4.2. Unconfined aquifer

An unconfined aquifer is a saturated, permeable layer in which the water table forms the upper surface of the saturation zone. It is commonly referred to as a water-table aquifer and is directly recharged by the infiltration of rainfall and surface runoff. Contrary to confined aquifers, the water in an unconfined aquifer is not under pressure, as it is not overlain by an impermeable layer [55].

I.5. The groundwater resources in Algeria

In the northern Algeria, the various aquifers identified are mostly categorized into main types: karst aquifers, alluvial aquifers sedimentary basins; this renewable groundwater resources present in Algeria's northern aquifers are estimated to contain around 2.5 billion m³ of groundwater, distributed across 177 aquifers. Of this, approximately 1.9 billion m³ (over 90%) is already being exploited, and many aquifers are currently experiencing overexploitation, this assessment was carried out by ANRH in 1993 ,2004 and 2009. While, the southern region's potential is estimated around 60000 billion m^3. However, exploiting this latter is difficult challenging and requires renewal effort. Around 6 billion m^3 can be extracted annually of these Saharan aquifers in Algeria.

Geologically; the Algerian Sahara has 02 fossil aquifers: the complete complex aquifer (CC) and the continental intercalary aquifer (CI), whose reserves are very vast; However, their resources are non-renewable [56]. These resources are based on an inventory including of: 9000 natural springs, 23000 boreholes and 60000 wells.

I.6. Technique for measuring different characteristics of Groundwater quality

GW quality is characterized through various physical and chemical parameters, that are relatively easy to determine. These physico-chemical analyzes of GW refer to all procedures to determine values of a sample, whiter they involve analyses, measurements, observations etc, performed in the laboratory or at the measuring station site. Among the parameters taken into account are:

I.6.1. physical parameters

I.6.1.1. Temperature

The temperature of water is a comfort parameter for uses and is of interest with regard to hydrogeological research. Where it is depending on exchange of heat with the surrounding air and solar radiation. It also enables the correction of analysis parameters whose values are correlated to the temperature, particularly conductivity; Additionally, by highlighting water temperature contrasts within an environment, indications about the origin and flow of water can be obtained. Temperature plays an important role in increasing chemical and bacterial activities, as well as water evaporation. it varies based on the outside temperature (air), the seasons, geological properties and the depth of the water level relative to the ground surface. It is one of the most

important characteristics of aquatic systems that influences the concentration of dissolved oxygen [57]. It is important to have accurate knowledge of water temperature, as this factor plays a crucial role in the solubility of salts and gases, the dissociation of dissolved salts, electrical conductivity, and the determination of pH, all of which are essential for understanding the origin and potential mixing of waters [58]. Temperature is also a significant factor influencing biological productivity, as it affects the physical and chemical properties of water such as density, viscosity, gas solubility (especially oxygen), and the rates of chemical and biochemical reactions [59]. To be refreshing, drinking water should have a temperature between 8°C and 15°C; when it reaches 20°C to 25°C, it quenches thirst less effectively. The World Health Organization (WHO) recommends that the temperature of water intended for human consumption should not exceed 25°C [60]. Temperature should be measured in situ, and instruments used to measure conductivity or pH often include an integrated thermometer for this purpose.

I.6.1.2. The Potential of Hydrogen

It ranks among the most important parameters evaluating water quality. This factor characterizes numerous physicochemical balances and depends on many factors, including the origin of the water. This parameter is associated with the concentration of hydrogen ions (H^+) present in the water [61]. In simple terms, it determines the acidity or alkalinity and neutrality of water, pH is a parameter that defines whether water is acidic or basic; the pH of pure water is 7 at 25°C, this value was chosen as the reference for neutral water; while water with a pH below 7 is considered acidic, and water with a pH higher than 7 is basic [60, 62]. The pH depends on natural environmental conditions such as vegetation cover, the nature of rocks and soil substrates, and on human activities such as pollution [63].

I.6.1.3. Total Water Hardness

Total water hardness, also known as the hydrotimetric title, is a broad indicator of water mineralization. It represents the total concentration of metallic cations, excluding alkali metals (Na^+, K^+) and hydrogen ions. Hardness is often caused by the presence of calcium and magnesium ions; and may occasionally include iron, aluminum, manganese, and strontium ions [60].

I.6.1.4. Turbidity

Turbidity is one of the main parameters perceived by consumers [64]. It is related to water transparency; in other words, put differently, turbidity refers to the decrease in the transparency of water due to the presence of suspended matter (SM) such as organic debris, clays and microscopic organisms [60]. To ensure water safety, turbidity levels should be maintained below 5 NTU (NTU, nephelometric turbidity unit). The high turbidity of the water reveals the presence of iron, aluminum or manganese due to oxidation within the distribution network [65], additionally, it promotes floc formation and the attachment and also promotes the attachment and growth of microorganisms, raising suspicion about its bacteriological quality [66]. According to Desjardins (1997), Kettab (1992), and Boeglin (2000) [35, 67, 68] One key property of groundwater is its extremely low turbidity levels, it's extremely low turbidity levels, that should be remembered. Turbidity can be evaluated using various methods, carried out in situ or in the laboratory setting as necessary. Turbidity is measured using a turbidimeter.

I.6.1.5. Electrical Conductivity

Conductivity serves as a method for validating physico-chemical analyses of water, it is expressed in µS/cm. In fact, contrasting conductivity assessments in a region help to detect areas of contamination, mixing, or infiltration.

Electrical conductivity measures the ability of water to conduct an electric current. Measuring conductivity allows for a quick assessment of water mineralization and facilitates monitoring its development [69]; whereas the majority of the substances dissolved in water are in the form of electrically charged ions the conductivity indicates the overall mineral content present in a solution: soft water generally has a low conductivity, and on the contrary, the hard water shows high conductivity. It is also a function of water temperature, whereas, it is greater as the temperature increases [70], and directly related to mineral concentration [71]. High conductivity indicates normal pH, or more commonly, high salinity [72].

I.6.1.6. Total dissolved solids

Total dissolved solids (TDS) represent an essential factor for characterizing natural waters. It is specified as the overall concentration of organic and inorganic substances in a water solution, whether ionic or not [73,74]. The inorganic constituents of TDS consist mainly of the ions: calcium, magnesium, sodium, bicarbonate, chloride and sulfate; while the organic components are

evaluated using the biochemical oxygen demand (BOD) and the chemical oxygen demand (COD) [75]. To determine this water quality parameter (TDS), there are two methods available: direct and indirect methods.

The direct method, known as the gravimetric method, which is recognized as the standard procedure for TDS determination, involves weighing the dry residue due to the evaporation of a measured volume of a filtered water sample [73]. It is performed by grab sampling, where distinct samples are collected at certain times to reflect the water conditions at the time of sampling [76]. Nevertheless, this technique has been associated with several disadvantages [18,77,78,79,73]. The quantification of bicarbonate ions is inaccurate, as half of them volatilize as CO_2 at temperatures around 100°C [79]. During evaporation, precipitated salts crystalline structure may trap some water within them. Consequently, in this situation, the water of crystallization will be weighed along with the precipitated salts [18]. Furthermore, the volatilization of acids, particularly at low pH levels, leads to the loss of some other anions, like chloride and nitrite ions. On the other hand, more mass may be viewed as the result of oxidation or conversion to hydroxides during heating [73]. While the gravimetric method is labor-intensive and time-consuming, its accuracy is sometimes not satisfactory because it primarily depends on the chosen drying temperature, which must allow the release of all the water of crystallization to come out without inducing any chemical decomposition [78]. The indirect method for estimating TDS is based on an electrolyte solution's conductivity, which refers to its ability to conduct electricity when a potential is applied between the electrodes submerged in it [80,18].

I.6.2. Chemical Parameters

The mineralization of most waters is dominated by eight ions, often referred to as major ions. These include the cations: calcium, Magnesium, Sodium, and Potassium, as well as the anions: Chloride, Sulfate, Nitrate, and Bicarbonate. Their concentrations in water vary independently, primarily depending on their solubility.

I.6.2.1. The cations

I.6.2.1.1. Calcium (Ca^{2+})

Calcium is an element existing in all natural waters [81]; and typically dominates the composition of drinking water and its content varies significantly depending on the geological formations

through which the water flows [82]. Calcium is an alkaline earth metal that is abundantly found in nature and particularly in limestone rocks in the form of carbonates. It is an important component of the overall water hardness. It mainly exists in two natural forms: either the dissolution of carbonate formation ($CaCO_3 \rightarrow Ca^{2+} + CO_3^{2-}$), or it can also come from the dissolution of gypsum formations [$CaSO_4 \cdot 2H_2O \rightarrow Ca^{2+} + SO_4^{2-} + 2H_2O$] which dissolve easily, Low calcium concentrations indicate either a base exchange with sodium or the absence of easily alterable calcium-rich minerals. On the other hand, a high calcium content results from the dissolution of the gypsum or anhydrite. Moderately hard water typically contains 250–350 mg/l; water exceeding 500 mg/l present serious drawbacks for domestic purposes and for supplying boilers. However, the WHO standards recommend a maximum concentration of 100 mg/l. Otherwise, calcium does not cause problems for drinking water. The only domestic inconvenience linked with high hardness is scaling [83,84,85,86].

I.6.2.1.2. Magnesium (Mg^{2+})

Magnesium, is an essential substance crucial for life, and it often accompanies calcium, playing a vital role in respiration. It is also considered one of the most widespread elements in nature and imparts a bitter flavor to water [60]. According to Nowayti et al, (2015) [82], Magnesium's origins appear to be related to water interacting with limestone and dolomitic rocks, which are carbonate formations rich in magnesium. Its geological abundance, high solubility, and wide industrial use mean that its content in water can potentially reach high levels [87]. The WHO standards recommend a maximum concentration limit of magnesium in water of 150 mg/l.

I.6.2.1.3. Sodium (Na^+)

Sodium is a stable element in water, but its concentration can be extremely variable. In general, sodium is an element found in all waters because the solubility of its salts is very high. Sodium may originate from the leaching of geological formations containing sodium chloride, the salt can come from numerous source; for instance, the decomposition of mineral salts such as sodium and aluminum silicates, from marine-origin fallout, from intrusion of saltwater into groundwater, various industrial activities [88]. The WHO recommends a maximum sodium concentration of 200 mg/l in drinking water.

I.6.2.1.4. Potassium (K^+)

Potassium is an element found mainly in igneous rocks, particularly volcanic rocks, and clays. It is mostly found in silicate rocks as orthoclase ($KAlSi_3O_8$), micas and felspathoids as leucite ($KAlSi_2O_6$). Typically, Groundwater contains potassium levels below 10 mg/l [89]. The concentration of potassium in water is usually low due to vegetation readily absorbs this element. Nevertheless, anthropogenic factors, particularly agricultural and industrial activities, can cause significant increases in potassium concentrations.

I.6.2.2. The Anions

I.6.2.2.1. Chlorides (Cl^-)

Chlorides are constantly present in natural waters in varying proportions. Their occurrence in groundwater is due to the dissolution of natural salts, such as sylvite (KCl) and halite (NaCl) [90]. In drinking water, the acceptable limit is 250mg/l (WHO standard). Exceeding this value can affect the taste of the water. Elevated chloride concentrations can cause eczema and erythema [91]. Also, water with high chloride content is laxative and corrosive [92].

I.6.2.2.2. Bicarbonate (HCO_3^-)

Bicarbonates are found in natural waters, their presence in water results from the dissolution of limestone rocks and dolomites or the release from deep magma, the atmospheric contribution to the presence of this element is negligible [84]. The dissolution of carbonate minerals and the influence of CO_2 from meteoric waters and soil is often the primary origin of bicarbonates [93]. WHO standards do not establish any value for this parameter, as bicarbonate levels do not affect the suitability of water for drinking purposes.

I.6.2.2.3. Nitrates (NO_3^-)

Nitrates are present in soil, surface water, and groundwater. They enter water mainly through the leaching of nitrogen compounds from soil, the decomposition of organic matter, and the use of natural or synthetic fertilizers [69]. Nitrates are used primarily as fertilizers; many other nitrogen-based fertilizers also transform into nitrates within the soil. Groundwater contamination by nitrates represents a major issue in water quality research [94]. The concentration limit of nitrate in water, established by the WHO, is 50 mg/l.

I.6.2.2.4. Sulfate (SO_4^{2-})

Sulfur is a non-metallic element that occurs naturally in soils and rocks, in both organic and inorganic forms. It exists as protein sulfur in organic form, and as sulfides, sulfates, and elemental sulfur in mineral form. When sulfur combines with oxygen, it gives the sulfate ion, which present in various minerals such as gypsum and barite [95]; It is considered the predominant sulfate form found in groundwater. Sulfates (SO_4^{-2}) originate from runoff or infiltration in gypsum areas. They are also produced from the action of certain bacteria (chlorothiobacteria, rhodothiobacteria, etc..). This activity can oxidize toxic hydrogen sulfide (H_2S) into sulfate [59]. The World Health Organization (WHO) has set the concentration limit for sulfate in drinking water at 250 mg/l.

I.7. Correlation TDS-EC

Analysis of TDS is essential and key because it provides a better understanding of water quality, particularly in groundwater, to assess the impact of seawater intrusion through electrical conductivity evaluation. TDS can be measured or estimated through indirect or direct methods (which can be determined empirically using conductivity). The indirect approach involves summing the measured concentrations of different components in a filtered water sample. However, field measurement of TDS is often more difficult, costly, and time-consuming compared to EC measurement [96]. Electrical Conductivity (EC) is used as a proxy for total dissolved solids (TDS). The use of correlation with EC to estimate TDS in natural waters offers notable advantages over the gravimetric method and chemical analysis because of the simplicity, rapidity, low cost, and accuracy in measurements [97,77]. Consequently, several research evaluates TDS by correlation or empirical factors. The estimation of TDS in water samples is calculated by multiplying EC by an empirical factor. This coefficient can range from 0.55 to 0.90, depending on the type of dissolved ions, their concentration; and the water temperature, the literature frequently depicts the correlation between TDS and EC through the following relationship (Eq 1) [18]:

$$\mathbf{TDS = k \,.\, EC} \qquad \mathbf{(1)}$$

where TDS is measured in mg/l; “k” the empirical factor is unit less EC is measured in µS/cm. As the concentration of ions in water increases, the value of “k” also increases. However, the correlation between TDS and EC is not directly linear; it is influenced by factors such as the activity of specific dissolved ions, the average activity of all ions in the liquid, and the ionic

strength [96]. In closing; Determining the ratio between TDS and EC remains challenging, and ongoing research continues to refine the methodologies used to address this issue. For this purpose, researchers employed machine learning (ML) or artificial intelligence (AI) methods, utilizing advanced algorithms capable of capturing both linear and nonlinear correlations, providing an alternative to traditional empirical regression approaches [96]. Consequently, these techniques may be effectively applied to estimate TDS using EC and temperature.

I.8. The importance of quantifying TDS in Groundwater

Groundwater is one of the world's most significant sources of fresh water. It is widely used across various parts of the world; it is regarded as the principal supply of drinking water, household, industrial, and agricultural needs in many countries, and is frequently referred to as clean water than other water sources due to natural filtration processes occurring within underground soil layers [98]. However, it must meet at least a satisfactory level of water quality [99]. Groundwater quality is determined by multiple factors, including recharge water quality, for the characteristics of geology, the degree of chemical weathering of various rock types, and the rock–water interface [100,101].

This important and essential water source is exposed to the risk of pollution. Groundwater contamination is frequent in both developed and developing countries. Groundwater pollution can lead to the degradation of water quality of drinking water, disrupted water availability, high cleanup costs, costly alternative water source, and potential health issues [102]. The principal sources of groundwater pollution are seawater intrusion, leakage from underground septic tanks, landfills, agricultural operations and chemical use [103].

As a result, research into the management of groundwater is required to be effective, managing and evaluating the resources of groundwater need a grasp of the aquifer's hydro geochemical and hydrogeological characteristics [104]. Because hydro geochemical processes that causing differences in groundwater's chemical composition over time and space, thus, continuous evaluation of these chemical characteristics of groundwater which are crucial for classifying and assessing water quality is important [105,106].

Groundwater chemistry refers to all dissolved solids in the groundwater system, also known as total dissolved solids (TDS), therefore, evaluating water quality for drinking and irrigation purposes necessitates comparing specific parameters against established standard norms [107].

Natural waters mainly include dissolved solids such as chlorides, carbonates, bicarbonates,

sulphates, magnesium, potassium, sodium, phosphates of calcium and nitrates, along with trace amounts of iron, a manganese, and other minerals. Water quality can be evaluated based on physical, chemical and bacteriological parameters for various applications.

TDS is regarded as the most essential parameter for determining water quality, as it directly connected and is influenced by increasing turbidity, hardness, alkalinity, and conductivity of the examined water sample [108]. The WHO standards recommend the concentration limit for TDS in drinking water at 500 mg/l; hence, the variations of these parameters may induce changes in the quality of ground water.

Understanding TDS levels aids in identifying possible health risks, tracing pollution sources, and informing successful water treatment options, all of which support public health and water resource management. Furthermore, studying the influence of TDS on public health is important because high TDS levels can reveal health problems and harmful pollutants [109]. Also, elevated TDS level can cause scaling and corrosion, specifically applications like cooling water systems and boilers, making it essential to observe TDS in water systems [110]. This knowledge helps to evaluate water safety, inform regulatory standards, enhance and safeguard public health.

Even though TDS is not considered a major water pollutant, elevated TDS concentrations often results from the presence of bicarbonates, carbonates, chlorides, calcium, sulphates, magnesium along with iron and manganese; these can come from natural sources depending on the chemical composition of the stratum where the water flows through, as well as from sewage, urban runoff and industrial wastewater [111]. As they may also be formed from chemicals utilized in the water treatment process, as well as pipes and other plumbing hardware [112].

Generally, when the ionic elements increase exceeds the permissible levels in groundwater, it is considered as pollution. Thus, more precisely, groundwater pollution can come from both natural (geogenic) and human-induced (anthropogenic) sources. While groundwater pollution occurs when groundwater contamination exceeds the safe limits for its intended use [107].

Conclusion

In summary, groundwater remains an indispensable yet fragile resource, requiring careful monitoring and sustainable management. The scientific advances reviewed here highlight both the complexity of aquifer systems and the need for reliable quality indicators such as electrical conductivity and total dissolved solids. These insights provide the basis for developing innovative approaches to improve water management strategies.

CHAPTER II

THE STUDIED AREAS

II. THE STUDIED AREAS

Introduction

The study areas of Upper Cheliff, Mitidja Plain, and Hassi R'mel represent key hydrogeological Zones in Algeria, characterized by distinct geographical, climatic, geological, and hydrological conditions. Their importance lies in both agricultural and industrial activities, as well as their role in groundwater supply. Understanding these contrasts is crucial for assessing water quality and sustainable aquifer management.

II.1. THE UPPER CHELIFF

II.1.1. The Geographical location

The cheliff basin, categorized as Basin 01 by the Algerian National Agency for Hydraulic resources (ANRH), is recognized as the biggest basin in northern Algeria. It is bordered by the Tellian Atlas Mountain ranges which extend along the Mediterranean coast. The Cheliff Basin is further divided into three distinct watersheds: The Upper Cheliff Basin, which originates from the Boughzoul dam, the Middle Cheliff and Lower Cheliff–Mina regions.

This research concentrates on the upper cheliff alluvial plain in northwestern Algeria. This region is an intra-mountainous depression that extends east–west for approximately 500 km and is 30 to 80 km wide. The Khemis Miliana plain, located in the Ain Defla province, covers around 21,035 km². Geographically, the plain is bounded by the Zaccar Mountains to the north, the Ouarsenis foothills to the south, the Doui massif to the west, and the Djendel uplands to the east (see **Fig 01** below). Is situated between 36° 12' and 36°17' north latitude and 2°12' to 2°17' east longitude, and includes the municipalities of Bir- Ouled Khelifa and Miliana, located between the Doui and Gontas Mountains.

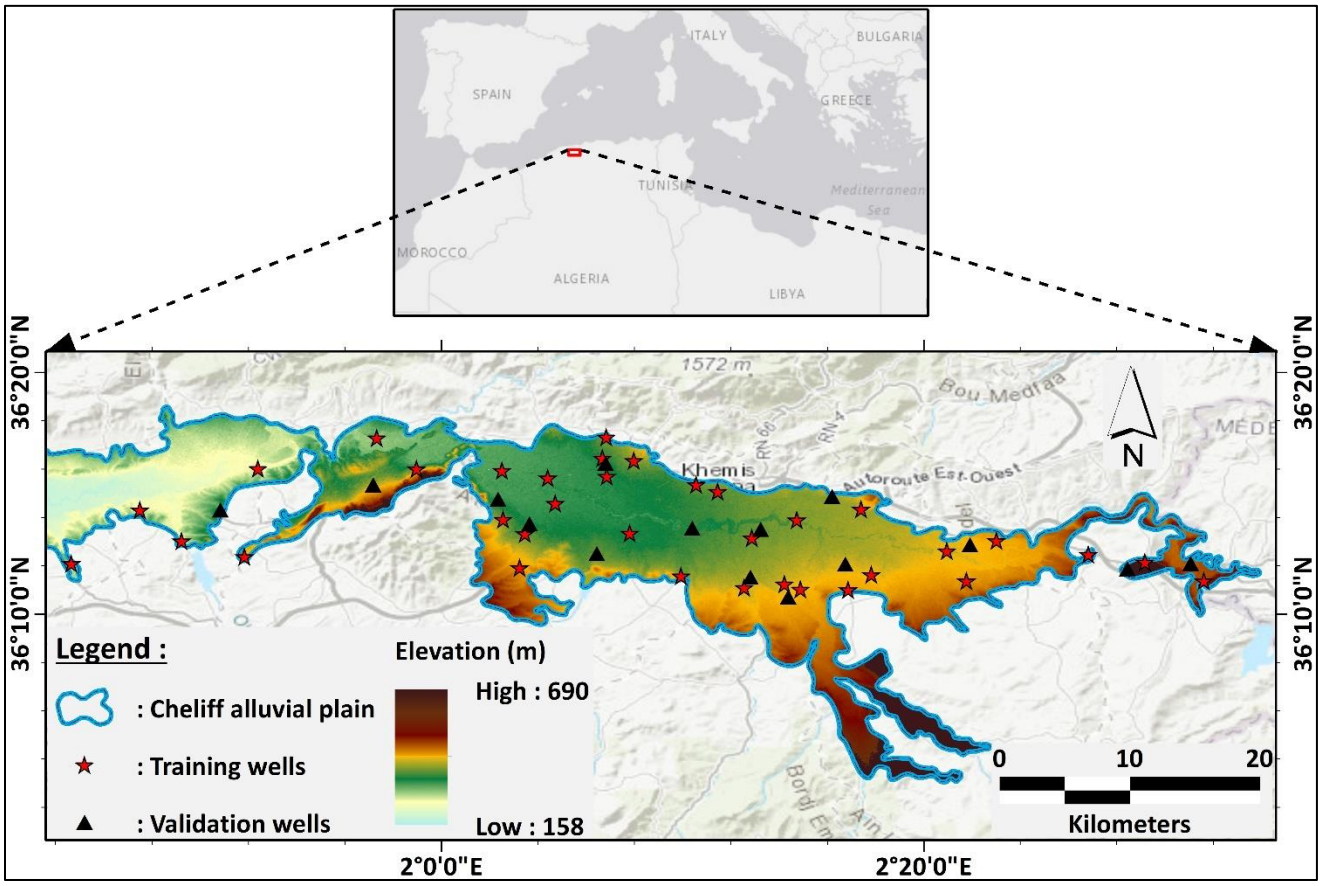


Fig 1. Location map of the Upper Cheliff

II.1.2. Geomorphological Characteristics

The plain features a relatively low slope of approximately 15‰ [113]. To the north, it is bordered by alluvial fans that form the transition zone between the valley and the mountains. The region's northern boundary is the Zaccar massif, while the Quarsenis mountain range serves as its southern border. The elevation ranges from 200 meters in the plains to 1000 meters in the mountain summits. Regarding vegetation; The vegetation cover is considerable, it occupies around 45% of the basin, it is primarily present in both the north and the south of the plain. The Upper Cheliff Plain is mostly utilized for agriculture, with fruit trees on the right bank of the wadi and cereal crops on the left, demanding intense irrigation between March, May and August. However, the basin lacks a dense forest cover, with remaining forests existing only in the southern part of the basin, and these are often degraded. This lack resulting in erosion, and causes damage such as the loss of arable land and siltation of hydraulic infrastructure.

II.1.3. Climate

The Upper Cheliff region has a semi-arid, Mediterranean-type Climate, characterized by very hot and dry summers, cold and harsh winters, despite its proximity to the sea (around 50 kilometers) [114]. The region is often referred to as the "Oven of the Tell" or "the portion of Sahara lost in the Tell"; In reality, one of the most distinguishing features of the environment is the intense heat that reigns in the Cheliff during the summer, which is notably higher than in nearby regions. There are others. Low winter temperatures and dry air are examples of certain climatic characteristics which justify the expression “Chelifien Climate". It has a brief spring (April and May) and an even shorter fall (October). The average annual rainfall is estimated to be 380 mm / year at the station of Ain Defla (upper cheliff plain), with an average annual temperature of 18.5° C at the Khemis station in the Upper Cheliff.

II.1.4. Geology and Hydrogeology

The upper Cheliff alluvial plain features a mixed aquifer formed by Mio-Plio-Quaternary alluvial deposits like sandstone, with a thickness ranging from 50 to 100 meters near the Cheliff wadi. The substratum of this aquifer layer is composed of marls. Groundwater in this area is predominantly extracted from the alluvial aquifer of the Upper Cheliff Plain that is accessed through wells and drilling. The aquifer is refilled by seepage from precipitation, runoff from wadis (e.g.,Cheliff, Deurdeur, Boutane and Souffay), and also excess irrigation water [115].

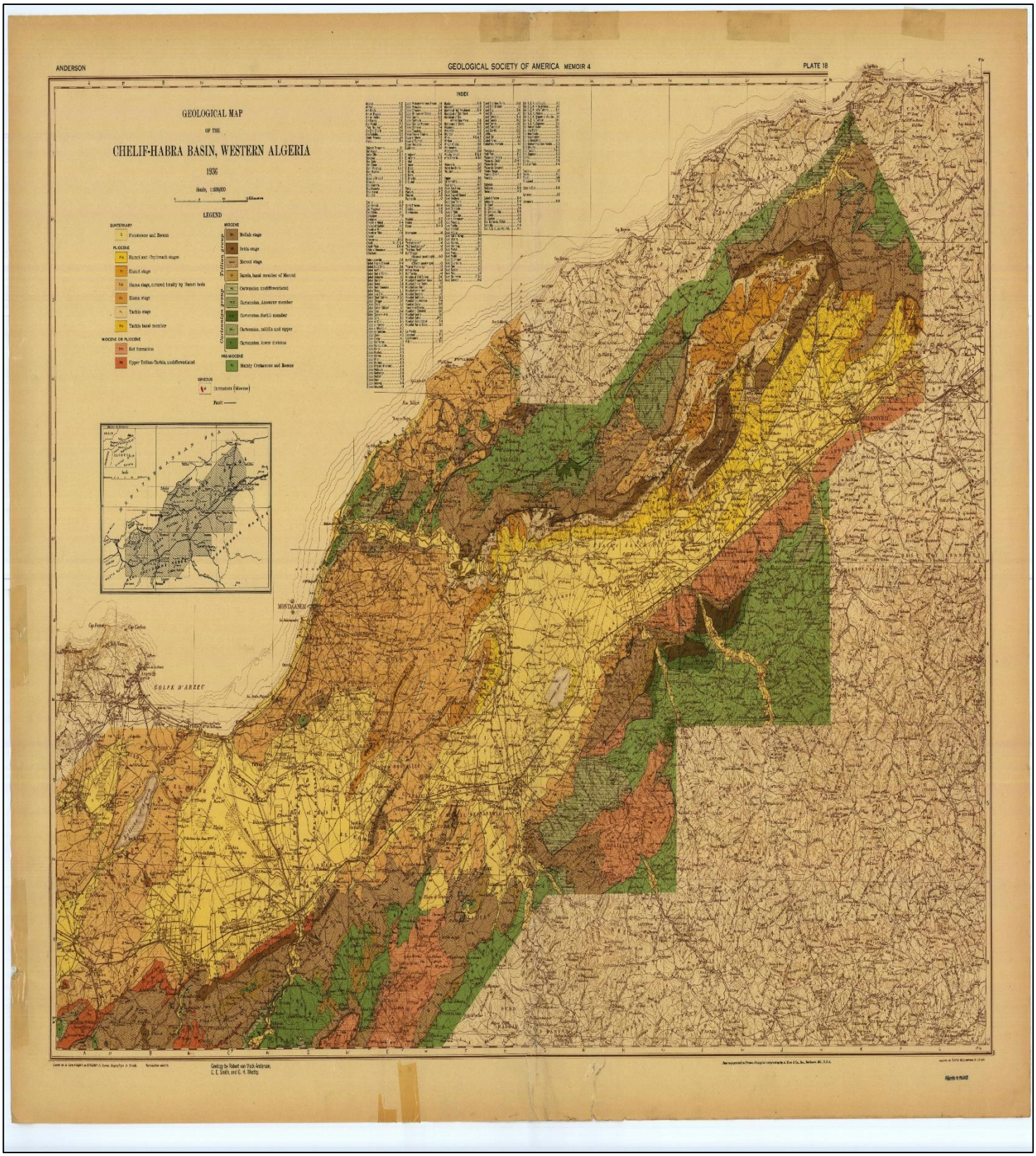


Fig2. Geological map in the Cheliff plain

II.2. MITIDJA PLAIN

II.2.1. The Geographical location

The Mitidja forms part of the Algiers Basin, coded 02 by the National Agency for Hydraulic Resources (ANRH), It is an elongated depression plain extending approximately 90 km from Hadjout in the west to the Boudouaou Wadi valley in the east, with an average width of 15 kilometers (see **Fig 03** below). This alluvial plain covers a geographical area of 1,400 km² and includes the wilayas of Algiers, Blida, Tipaza, and Boumerdes, its elevation ranges from 40 to 1,600 meters above sea level. The aquifer's limits with those of the Mitidja plain, which is bordered to the south by foothills of the Blida Atlas mountain range and to the north by a narrow coastal strip bordering the Mediterranean Sea.

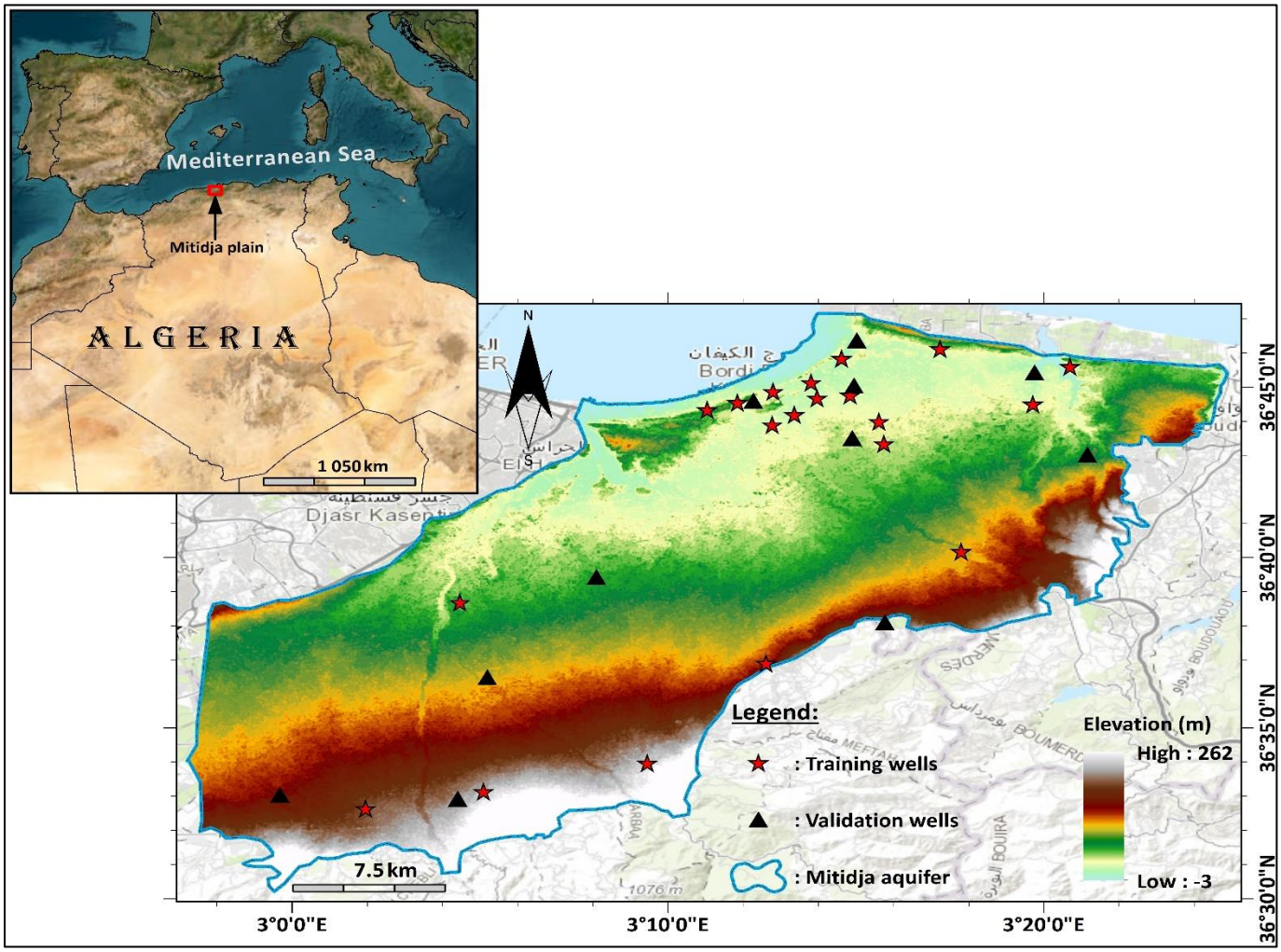


Fig 3. Location map of the Mitidja Plain

II.2.2. Geomorphological Characteristics

The Mitidja Plain is a sloping depression that is approximately 100 km in length and 15 to 20 km in width, situated between the Blida Atlas to the south and the coast to the north. It opens broadly to the sea for around 30 km. In its western part, the coast hills connect with the Chenoua mountain range (905 m) and join at the Fadjana plateau, marking the first elevations of the Atlas Mountains, including the Djebel Thebarrarine to the south (853 m). The Mitidja plain has a moderate slope from south to north, descending from the Atlas Mountains toward the hills. Only 50 m separate Ahmeur- el-Ain and the base of lake Hallola. it spans about 70 km from the western edge of Algiers, the plain is only connected to the Mediterranean through the Nador wadi, and, about 40 km further east, via the Mazafran River. The Mitidja Plain is a set of highly fertile lands, it has the best soils in the area; this soil diversity allows for a wide aptitude of crop varieties: citrus fruits are grown primarily in the central part of the plain, while vines are grown everywhere; Additionally, wheat is associated with fodder and market garden crops, and the region also supports the cultivation of industrial crops.

The vegetation cover plays an important role in the hydrological behavior of the basin, which is Mediterranean in nature [116]. Chrea National Park's forest forms. are based on Atlas cedar, holm oak, cork oak, zean oak Aleppo pine and Berber Tuya.

II.2.3. Climate

The Mitidja plain is influenced by the Mediterranean climatic regime due to its geographical location. It has the same regional sub-humid coastal climate as the nearby coastal lowlands [117]. Nevertheless, as one goes further from the coast, the climate becomes more continental, with significant drop in temperatures. The area's Mediterranean climate produces hot, sunny and dry summers, with temperature often reaching 32 - 33°C in July and August. The winter climate is mild, with temperatures rarely falling below 0°C. The average annual rainfall in Mitidja ranges between 500 and 900 mm.

II.2.4. Geology and Hydrology

Several studies have been conducted on the geology of the Mitidja Plain, based on the main work of [118], which demonstrated that Mitidja is an intra-mountainous basin formed by tectonic subsidence and subsequently filled with Plio-Quaternary deposits. The creation of the synclinal subsidence basin beneath the current Mitidja plain began in the Miocene and proceeded into the Pliocene. During this epoch, marine clayey rocks (referred to as Plaisancien and Astian clays) accumulated in the deep sea over cretaceous carbonate rocks. By the late Pliocene, the basin becom increasingly shallow, with a sequence of marls, sandstones, and clays known as the Astian Formation were deposited. The Mitidja basin was isolated from the sea by a small folding at the end of the Pliocene, accompanied by uplift of the basin. In the post-Pliocene period, fluvio-lacustrine sedimentation was periodically disrupted by periods of severe erosion.

The alternation between erosion and sedimentation was regulated by changes in the base level, owing to sea level fluctuations throughout the Pleistocene, as well as by a little folding within the stratigraphic basin and the tectonic uplift of the Astian Formation.

Geological and physical studies indicated the presence of aquifer formations within the Mitidja plain. The lithology and hydrodynamic characteristics allow the identification of two main aquifer units: The Astian reservoir and the Quaternary Alluvium.

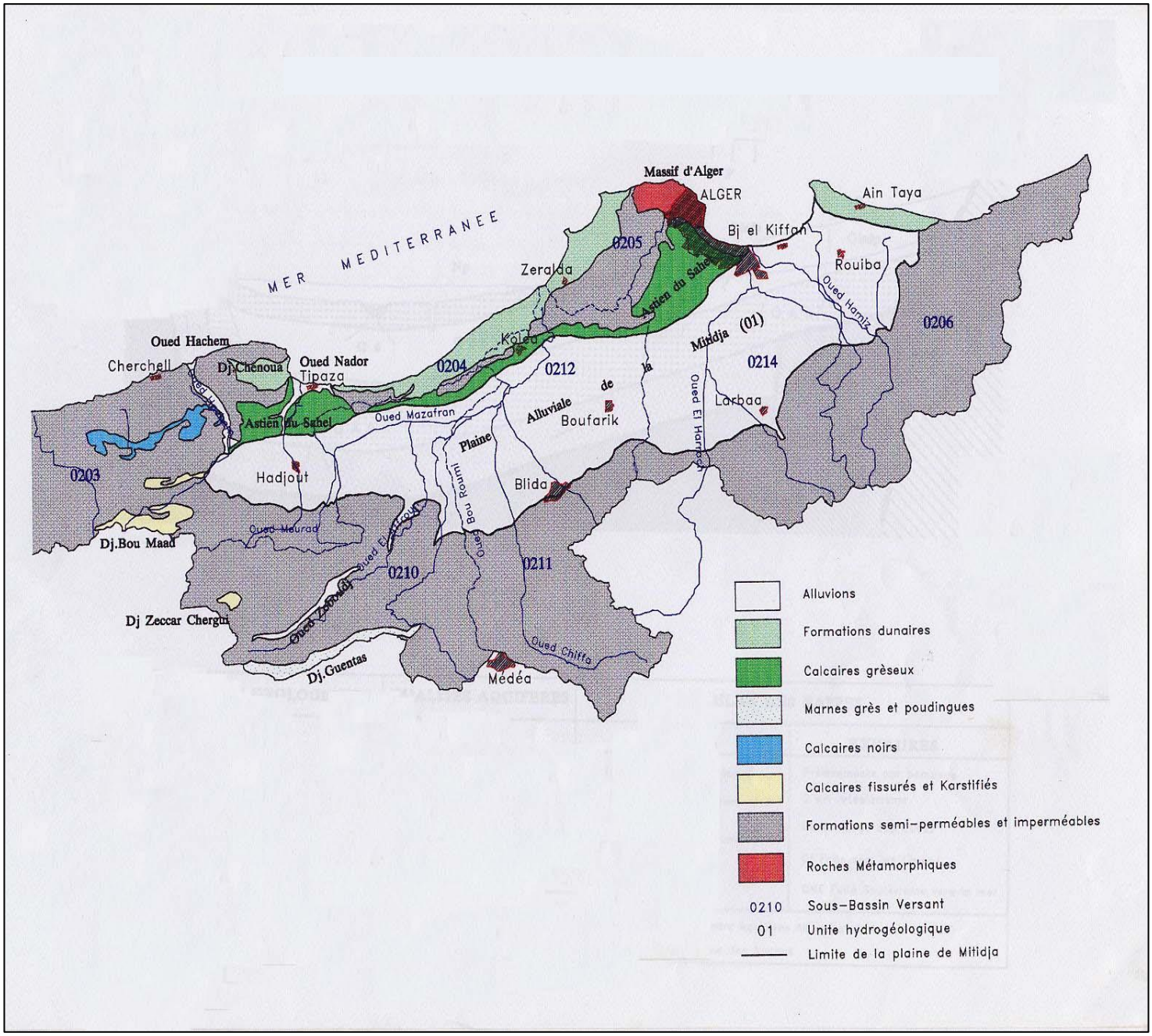


Fig 4. Geological map in the Mitidja Plain

II.3. HASSI R'MEL

II.3.1. The Geographical location

Hassi R'mel is identified as a Sub-Basin No. 7of the Sahara hydrographic region, according to the classification by the National Agency for Water Resources (ANRH). Hassi R'mel, commonly known as the "gateway to the desert", is situated in the central part of the Northern Saharan Platform, within Laghouat Province. It is located approximately 550 km south of Algiers and 120 km south of Laghouat City, 70 km west of Berriane, and 12 km northwest of Ghardaïa. Geographically, it is between 2°55′ and 3°50' east longitude, and between 33°15' and 33°45' north latitude (see **Fig 05** below). This region is limited by the Grand Erg Occidental to the south and west, the Touggourt area to the east, and the Saharan Atlas to the north. The average elevation is around 760 m above sea level. The Hassi R'mel field covers an area of around 3,500 km², It spans about 80 km from north to south and 60 km from east to west with a perimeter of 380 km [119]. It is recognized as one of the world's largest natural gas reserves.

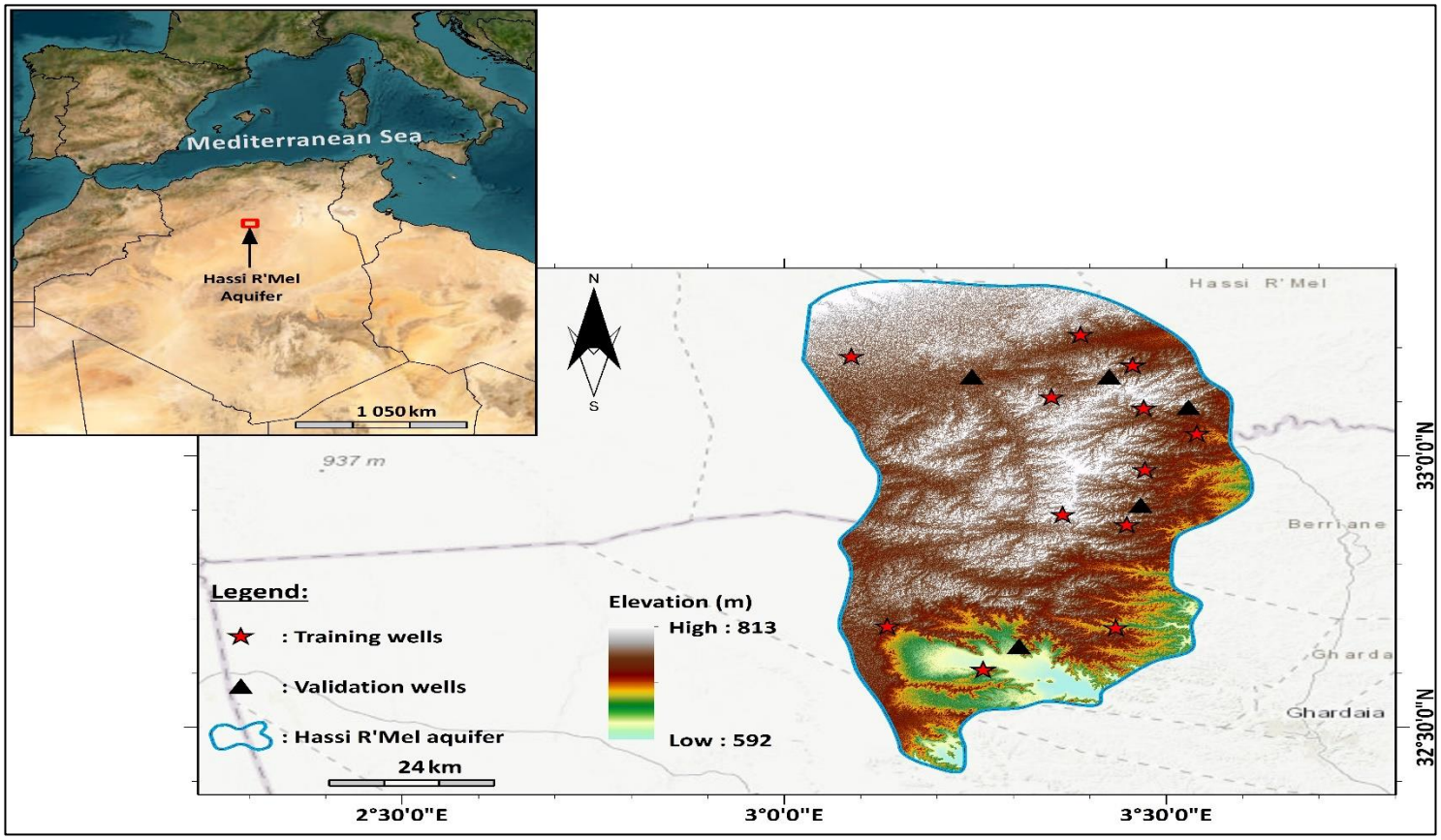


Fig 5. Location map of the Hassi R'mel (Laghouat Province)

II.3.2. Geomorphological Characteristics

Hassi R'mel is located within the High and Saharan Plateaus region. The relief is relatively flat and rocky, characterized by exposures of Senonian limestone rocks from the Upper Cretaceous period [120]. The geography of this plain is characterized by elevations ranging from 500 to 1000 m, with slopes varying from 0 to 3%. The band region includes vast steppe areas, much of which have been deteriorated due to severe drought conditions. The landscape is dominated by a broad rocky plateau interspersed with small depressions locally known as "daïas", small depressions where soil has accumulated and allowed for vegetation growth, mainly composed of shrubs, with a few isolated trees like wild pistachios.

II.3.3. Climate

Hassi R'mel, situated in the Saharan region, experiences a desert climate with pronounced aridity. This region has minimal precipitation, averaging 138.5 mm. The rainy season spans from September to April, while the dry season dominates the rest of the year, the area's average annual temperatures are roughly 21.2°C.

II.3.4. Geology and Hydrogeology

The Hassi R'mel anticlinal structure is directed north-south and located at the western edge of the Triassic province [121,122]. This anticline rests upon Paleozoic reliefs eroded down to Ordovician and Cambrian formations. It lies at the junction of two major structural axes: the first is the northern extension of the Hoggar-Idjerane M'Zab ridge, which curves slightly westward [119]. The other, oriented east-west and encompassing the elevated areas of Tilrhemt and Djemaa, which likely represent the buried extension of the Anti-Atlas reliefs. Hassi R'mel structure softly descends to the north into the Saharan flexure, the extends to the south-west by the small anticlinal of Bissa mountain, as well as to the south by the anticline of Hassi R'mel sud. It is bounded to the south by the Qued Mya depression and to the east by the Djemaa Touggourt [119,123].

The structural analysis of the Hassi R'mel region found three significant paleographic features: remnants of Austrian and Atlas movements, as well as an Ante–Mio–Pliocene morphology.

These geological occurrences resulted the terrain to be divided into three zones characterized by varying elevations and depressions:

1-Southern Zone: Affected by numerous strike-slip faults, some of which are substantial (up to

100 m).

2. Central Zone: A gently sloping plateau comprised mostly of Mesozoic rock formations.

3- Northern zone: Heavily influenced by tectonic activity, especially in post-Jurassic deposits. The presence of relatively deep faults has resulted in the formation of horsts and grabens, facilitating Miopliocene deposition. The research discovered three major aquifers in the region: The first Aquifer "Turonian Free Superficial"; it is Dolomitic limestone and vacuolar dolomite located at depths ranging from 0 to 150 m. a confined Albian-aged aquifer is the second aquifer found between depths of 150 and 450 m, made up of friable fine sandstone. with clayey cement. Regarding the last aquifer "Confined Barremian-Aged", it is Medium sandstone with carbonate and Clayey Cement, occurs at depths of 480 to 540 m [124].

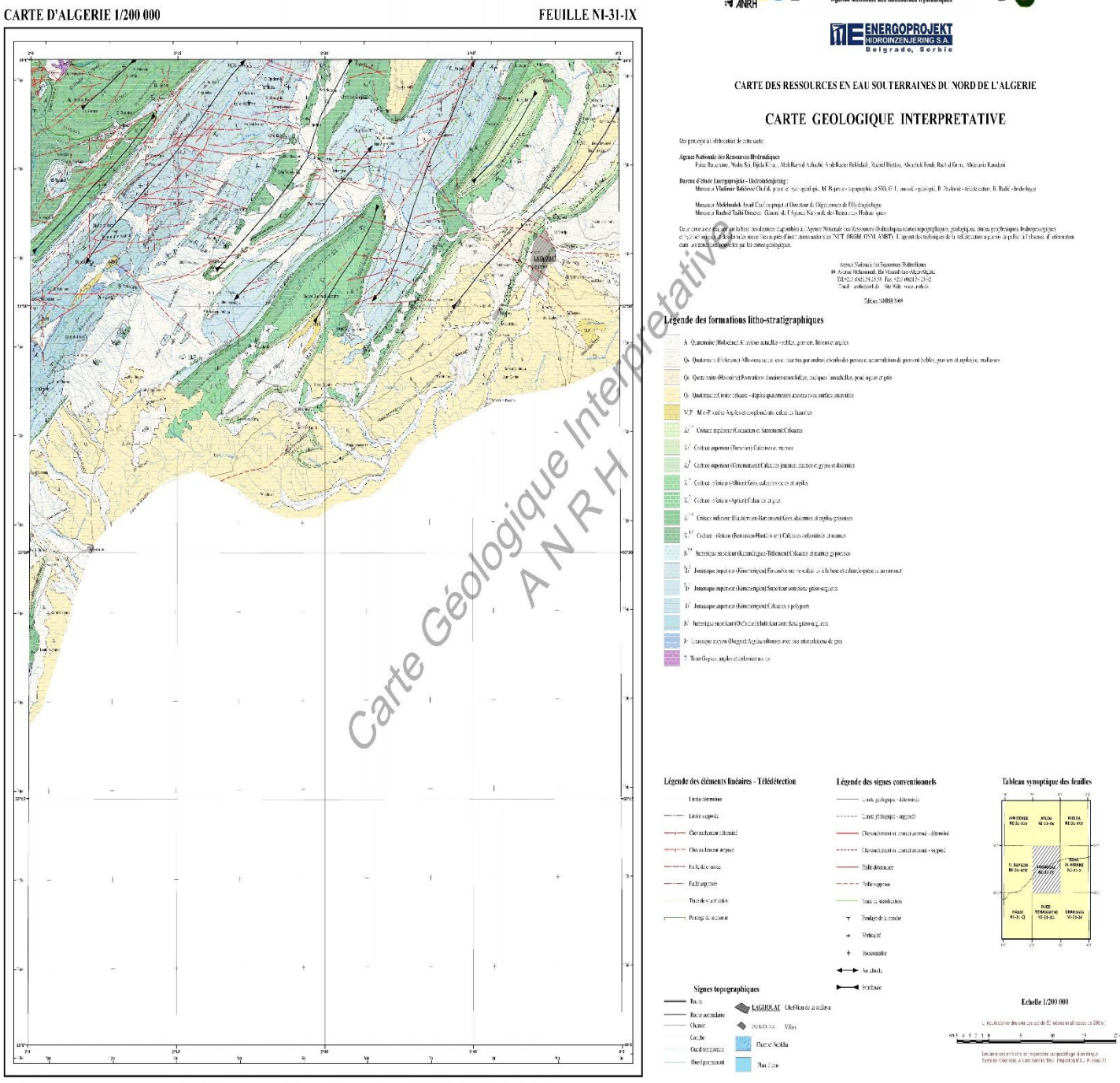


Fig 6. Geological map in the Hassi R'mel (Laghouat Province)

II.4. USED DATA

II.4.1. Protocol for physico-chemical analyzes of groundwater

II.4.1.1. In situ measurements

Four Physical Parameters (Hydrogen Potential (pH), temperature, electrical conductivity, and turbidity) were measured in situ, immediately after sampling using a multi-parameter field kit (including a pH meter and a conductivity meter). Calibration was performed before each measurement cycle. Turbidity was measured by a HACH 2180 N portable turbidimeter, also calibrated before each campaign. Samples were stored in an icebox or cooler at 4°C and sent to the laboratory for analysis on the same day.

II.4.1.2. Analysis of chemical elements

Chemical analyses were performed on the primary elements responsible for water mineralization, commonly referred to as major elements: cations (calcium, magnesium, sodium, and potassium) and anions (chlorides, sulfates, bicarbonates, and nitrates). These analyses were carried out using chemical reagents and, in some cases, specialized instruments.

All physicochemical parameters were measured at the laboratory of the water treatment plant (ADE).

II.4.1.3. Treatment standards

Standards have not been established for high quality water. According to the WHO, drinking water standards are broad, in order to accommodate many developing nations that have often experience poor water quality and that lack the technological means to monitor or perform the necessary treatments to ensure potability.

The following table lists drinking water standards for Algeria and the WHO:

Table 1. Drinking water standards according to Algeria and according to the WHO

Parameter	WHO	Algerian	Unit
PH	6,5 - 9,2	6,5 – 8,5	/
Temperature	-	25	°C
Electrical Conductivity	-	2800	µs/cm
TDS	1500	2000	mg/l
Turbidity	5	2	NTU
Calcium	-	200	mg/l
Magnesium	150	150	mg/l
Sodium	-	200	mg/l
Potassium	-	20	mg/l
Sulfate	250	400	mg/l
Chloride	250	500	mg/l
Nitrate	50	50	mg/l

II.4.2. Available data

To model the relationship between groundwater parameters in the three aquifers selected for this study; a dataset consisting of several parameters was used. Groundwater samples were collected from wells and recorded by the National Agency for Hydraulic Resources (ANRH). These samples were analyzed for several physicochemical variables, including parameters that influence water quality, this series is mainly composed of total dissolved solids, Electrical conductivity and the major elements.

The data contains outlier values for practically all parameters, making it critical to undertake a comprehensive investigation prior to modeling. Exploration and cleansing of data are key processes in the construction of ML models, as they ensure the correctness and dependability of the results.

In this study, data preprocessing involved several steps, beginning with a reliability assessment that included verification of the ionic balance [125].

This step was preceded by an examination of the major ions used to calculate the ionic balance (IB). The ionic balance is the difference between the total of anions and the sum of cations, or the anion-cation balance, which is normally zero, implying that the anions and cations are equal (1). The study yielded an ion balance of less than 5%, indicating that the data are credible.

$$\textbf{IonBalance : } \boldsymbol{IB}\ (\%) = \mathbf{100} \times \frac{\left|\sum \boldsymbol{cations}\left(\frac{\boldsymbol{meq}}{\boldsymbol{l}}\right) - \sum \boldsymbol{anions}\left(\frac{\boldsymbol{meq}}{\boldsymbol{l}}\right)\right|}{\sum \boldsymbol{cations}\left(\frac{\boldsymbol{meq}}{\boldsymbol{l}}\right) + \sum \boldsymbol{anions}\left(\frac{\boldsymbol{meq}}{\boldsymbol{l}}\right)} \qquad \textbf{(2)}$$

II.4.3. Statistical Data

The statistical description of the parameters is presented in Tables (02, 03, and 04), including the minimum, maximum, mean, standard deviation and coefficient of variation, for the datasets used during the training and validation phases. These parameters were used to ensure and assess the accuracy and robustness of the models.

Table 2. Statistical parameters of the training and validation datasets cheliff

Variables	Training dataset					Validation dataset				
	Min	Max	Mean	SD	CV%	Min	Max	Mean	SD	CV%
TDS (mg/l)	452.00	6571.00	1975.91	914.12	46.26	609.00	5686.00	1973.39	962.87	48.79
EC (µS/cm)	610.00	6800.00	2943.10	1257.23	42.72	993.00	8100.00	2941.77	1330.96	45.24
Ca^{2+} (mg/l)	62.96	489.80	246.96	98.02	39.69	117.00	470.00	247.47	82.67	33.41
Mg^{2+} (mg/l)	3.00	224.00	84.69	46.19	54.54	2.45	349.09	91.44	63.74	69.71
Na^{+} (mg/l)	12.00	800.00	195.01	138.35	70.95	13.10	540.00	179.14	120.46	67.24
K^{+} (mg/l)	0.10	48.00	6.07	7.39	121.75	0.20	34.20	5.98	7.26	121.40
Cl^{-} (mg/l)	60.00	1390.00	518.22	313.46	60.49	98.00	1690.00	529.80	315.59	59.57
SO_4^{2-} (mg/l)	40.00	2185.00	320.31	230.23	71.88	67.00	1540.00	330.13	247.38	74.93
HCO_3^{-} (mg/l)	146.40	762.50	334.86	101.09	30.19	148.84	457.50	303.21	71.30	23.52
NO_3^{-} (mg/l)	0.00	270.00	47.57	35.34	74.29	0.00	175.00	59.36	38.31	64.54

Table 3. Statistical parameters of the training and validation datasets mitidja

Variables	Training dataset					Validation dataset				
	Min	Max	Mean	SD	CV%	Min	Max	Mean	SD	CV%
TDS (mg/l)	218.00	4217.00	998.13	466.61	46.75	317.00	2557.00	939.30	387.73	41.28
EC (µS/cm)	300.00	6450.00	1555.79	724.39	46.56	480.00	4170.00	1446.50	578.44	39.99
Ca^{2+} (mg/l)	6.67	624.00	143.57	69.40	48.34	26.67	284.00	138.58	55.13	39.78
Mg^{2+} (mg/l)	1.00	199.15	47.25	31.96	67.64	0.00	240.00	42.79	34.25	80.04
Na^{+} (mg/l)	10.00	525.00	90.49	62.06	68.58	16.00	320.00	75.59	53.08	70.22
K^{+} (mg/l)	0.00	14.00	2.98	2.39	80.37	0.00	20.00	2.50	2.52	100.8
Cl^{-} (mg/l)	24.00	1755.00	178.39	163.30	91.54	30.00	1075.00	147.64	133.07	90.13
SO_4^{2-} (mg/l)	0.00	625.00	175.28	116.41	66.41	23.00	410.00	166.63	93.07	55.85
HCO_3^{-} (mg/l)	45.75	823.50	338.91	125.96	37.17	91.50	610.00	338.25	98.73	29.19
NO_3^{-} (mg/l)	0.00	251.00	48.09	32.30	67.15	0.00	120.00	46.37	24.85	53.59

Table 4. Statistical parameters of the training and validation datasets HASSI

Variables	Training dataset					Validation dataset				
	Min	Max	Mean	SD	CV%	Min	Max	Mean	SD	CV%
TDS (mg/l)	300.39	1526.53	714.66	192.95	27.00	486.86	1299.15	753.41	175.00	23.23
EC (µS/cm)	420.84	2075.03	1015.40	266.29	26.23	684.60	1774.02	1081.49	240.94	22.28
Ca^{2+} (mg/l)	15.71	177.29	93.59	29.98	32.03	54.41	174.55	98.97	22.23	22.46
Mg^{2+} (mg/l)	7.20	74.00	34.45	11.35	32.95	8.09	59.83	36.92	9.50	25.74
Na^{+} (mg/l)	32.00	227.00	83.56	36.29	43.42	34.00	302.50	86.80	54.13	62.36
K^{+} (mg/l)	3.00	23.00	7.11	4.11	57.77	2.80	19.50	6.07	3.19	52.62
Cl^{-} (mg/l)	43.00	361.00	109.46	57.20	52.26	52.00	309.18	111.92	54.54	48.73
SO_4^{2-} (mg/l)	34.00	560.00	197.50	87.72	44.42	55.00	403.00	197.99	77.96	39.38
HCO_3^{-}(mg/l)	59.60	348.00	245.09	55.34	22.58	148.00	382.17	275.88	55.91	20.27
NO_3^{-} (mg/l)	0.00	43.00	6.18	6.73	108.82	1.10	43.00	8.96	11.46	127.85

The statistical analysis of the training and validation datasets from Cheliff, Mitidja, and Hassi aquifers reveals substantial variability in groundwater quality. In Cheliff and Mitidja, both TDS and EC exhibit wide ranges and relatively high coefficients of variation, indicating heterogeneous hydrochemical conditions and possible anthropogenic influences. By contrast, Hassi aquifer shows lower variability (CV < 30% for most parameters), reflecting more stable hydrochemical characteristics. Major ions such as Ca^{2+}, Mg^{2+}, and Cl^{-} display high dispersion in Cheliff and Mitidja, while remaining more consistent in Hassi. Overall, these differences highlight spatial contrasts in groundwater quality and underline the need for site-specific management strategies.

Conclusion

In summary, the three study areas illustrate the diversity of Algeria's hydrogeological systems, ranging from semi-arid Mediterranean basins to Saharan environments. Their specific geomorphological and climatic settings strongly influence groundwater dynamics and quality. These differences provide a valuable basis for comparative analysis and for developing region-adapted water resource management strategies.

CHAPTER III

MATERIALS AND METHODS

III. MATERIALS AND METHODS

III.1. Introduction

Throughout history, humans have used various tools to improve their efficiency in different jobs. The creativity of the human mind has enabled to the creation of numerous technologies, which have made people's lives easier in various aspects by enabling them to handle many demands and requirements, including industry and information technology. Among these innovation is machine learning [126]. Machine learning is a growing field of computing algorithms aimed at mimicking human intelligence by learning the environment. They are regarded as the working horse in the era of big data. It involves solving practical problems by gathering data and building statistical models algorithmically from that data [127]. This chapter presents the main concepts of machine learning, its categories, and the specific models selected for predicting groundwater quality parameters. This chapter presents the methodological framework adopted to assess groundwater quality and forecast total dissolved solids in three Algerian aquifers. It describes the datasets used, the preprocessing steps applied, and the machine learning models implemented. The procedures for feature selection, model training, and performance evaluation are also detailed to ensure reproducibility and scientific rigor.

III.2. Machine learning

Machine learning is a field of contemporary computer science. In the literature, the ML Domain has received various formal definitions, the prominent. Arthur Samuel described ML as "a field of study that gives computers the ability to learn without being explicitly programmed" in his pioneering work [128]. Tom Mitchell used a computer science dictionary to explain it: "A computer program is said to learn from experience (E) with respect to some class of tasks (T) and performance measure (P), if its Performance at tasks in T, as measured by P, improves with experience E" [129]. Additionally, Ethem Alpaydin defined ML as: "Programming computers to optimize a performance criterion using example data or past experience" [130]. These multiple definitions collectively emphasize the concept of training computers to execute tasks intelligently beyond traditional statistics by learning from repeated examples and adapting to their environment. Machine learning (ML) shows its capability for identifying trends in large and diverse

data sets, commonly known as big data. It is more precise and faster than traditional approaches of analyzing large data sets since ML algorithms can learn and adjust their outcomes based on new data without necessitating reprogramming. While conventional analytics techniques often grapple with massive amounts of data, ML excels as databases grow, enabling it to identify trends with considerably higher accuracy [131].

III.3. Types of machine learning

Given the variety of ML systems, it is useful to divide them into four main categories: supervised learning, unsupervised learning, semi-supervised learning and, reinforcement learning

III.3.1. Supervised learning

Supervised learning is a machine learning model where data includes other attributes to be predicted. This type of learning obtains information about the input-output relationship of a system using a provided set of prepared input-output training samples [132].

An example of a typical supervised learning task is linear regression, where the objective is to find a linear relationship between input and the values and the target values. The TDS-EC correlation that we applied in our research is a good example of supervised learning.

III.3.2. Unsupervised learning

Unsupervised learning refers to algorithms that operate on unlabeled data. In simpler terms, enabling the model to uncover independent patterns and structures in untagged data. In this way it is possible to categorize the data into groups of similar elements [133].

III.3.3. Semi-supervised learning

Semi-supervised learning is an intermediate approach between supervised and unsupervised algorithms, leveraging a small portion of labeled data to improve learning from a larger set of unlabeled data. This type of learning offers a notable advantage, as labeling data can be costly, time-consuming, and resource-intensive [133].

III.3.4. Reinforcement learning

Reinforcement learning is a very distinct learning system that is carried out without direct supervision, through interaction with the environment. Generally, this type involves rewarding desired behaviors and/or penalize undesirable ones, while also interpreting the environment, taking actions, and learning via trial and error. The primary goal is to find the best behavioral strategy to

maximize the overall reward. To do this, the machine requires basic knowledge or rules to define the nature of its behavior. This feedback is known as the reinforcement signal. Although it is not obvious to manually create or program scenarios to achieve better results, machine learning is extremely useful when there is a large number of conditions to forecast [134].

III.4. The used models

In this research, we selected four ML models that are considered among the most recognized algorithms, having provided their performance in forecasting TDS according to previous studies

III.4.1. Decision tree

Decision Trees are popular supervised machine learning algorithms commonly used in classification and regression tasks to predict target values. This model is one of the most essential data mining techniques, as it uses training data to develop a set of decision rules. The decision tree was then represented as a tree structure consisting of a root node, branches, and leaf nodes, and it is built by selecting which selecting the feature that provides the most information gain at each step of the algorithm [135,136].

III.4.2. Random Forest

The Random Forest technique, proposed by Breiman (2001) [137], is a supervised machine learning algorithm based on decision trees that may be applied to both classification and regression tasks [25]. This model has been improved and extensively adopted in several studies, resulting in strong performance [138,139,140,141]. Random Forest has also become a strong contender to boosting methods [142]. Predictions are generated by first dividing the data into numerous sub-datasets via bootstrapping, a technique employed by the Random Forest model. Each sub-dataset is subsequently given its own decision tree, and these sub-decision trees are integrated with random forest model to produce final forecasts that represent the model's overall performance [143, 135]. RF is a widely used machine learning model, regarded as one of the most accurate models for general task learning. This model has various appealing characteristics, its attractive lies in its quick and straightforward application during training, as well as its capacity to handle numerous input variables without the risk of overfitting which is a common problem in many models. Random Forest is also known for producing extremely accurate forecasts [144].

III.4.3. Categorical Boosting

Prokhorenkova et al (2018) introduced a novel gradient-boosting-based decision tree (GBDT) technique known as Categorical Boosting (CatBoost) [145]. It uses an advanced ensemble learning technique based on gradient descent. In training, a set of decision trees (DTs) is constructed progressively, with each DT learning from the previous tree and influencing the next tree to enhance model performance, resulting in a robust model. Unlike traditional gradient boosting trees (GBTS), it has two notable characteristics: ordered boosting and effective handling of categorical features [145, 146].

III.4.4. Support victor regression

Support Vector Machine (SVM) is a supervised machine learning algorithm invented by Vapnik in 1995[186], that has been effectively employed to solve classification, regression and pattern recognition tasks [147]. This model employs equation (3) to illustrate the correlation between input and output variables, aiming to reduce the variance between observed and predicted values [44, 45] [148, 149].

$$f(x) = \eta^T \Phi(x) + b \qquad (3)$$

where $f(x)$ the output value; η is the weighting coefficient of input data; b is the bias term; T denotes the transpose symbol; and $\phi(x)$is the nonlinear mapping function.

SVR model is an advanced and more sophisticated data-driven model that aims to reduce structural risk, thereby lowering high-margin errors, as opposed to other machine learning models that simply consider local training errors. It uses a kernel function to transform input data sets to a high-dimensional or infinite-dimensional feature space. As SVMs are kernel-based algorithms, the kernel function used is critical and can have a significant influence on the final models. SVR offers four kernel functions from which to choose: linear, polynomial, sigmoid, and radial basis function (RBF) [46,47] [150, 151].

III.5. Model performance evaluation

The primary method to assess a model's ability to generalize to new instances is through rigorous testing. This includes deploying the model, evaluating its performance, and validating its

predictions. For our study, two statistical criteria were employed:

III.5.1. Nash-Sutcliffe efficiency

The Nash-Sutcliffe efficiency (NS) coefficient is used to evaluate a machine learning model's predictive accuracy, with higher NS values indicating greater accuracy [152]. The NS coefficient it accepts the values ranges from - ∞ to 1, where a score of 1 implies perfect agreement, and 0 indicates that the model fails to explain any of the observed variance.

$$\mathbf{NS} = \mathbf{1} - \frac{\sum\left(\mathbf{TDS_{i_{pred}}} - \mathbf{TDS_{i_{obs}}}\right)^2}{\sum\left(\mathbf{TDS_{i_{obs}}} - \overline{\mathbf{TDS}}_{i_{obs}}\right)^2} \qquad \mathbf{(4)}$$

III.5.2. The root mean square error

The RMSE values reflect the square root of the variation in residual errors between observed and forecasted values. Thus, lower RMSE values indicate higher model accuracy.

$$\boldsymbol{RMSE} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\left(\boldsymbol{TDS}\mathbf{i_{pred}} - \boldsymbol{TDS}\mathbf{i_{obs}}\right)^2} \qquad \mathbf{(5)}$$

Where TDSi_obs and TDSi_pred are the observed and predicted TDS values; TDS̄i_obs is the mean of observed TDS; and N is the total number of observations.

III.6. The model development

III.6.1. Data normalization

Data normalization is crucial to minimize or prevent the influence of variable dimensionality on model performance and to improve model generalization. Normalized values typically range from 0 to 1. To resolve the errors by the differences in measurement units between input and output parameters, we normalize data using the following formula:

$$\mathbf{x_{norm}} = \frac{(\mathbf{0.95} - \mathbf{0.05}) * (\mathbf{x_i} - \mathbf{x_{min}})}{(\mathbf{x_{max}} - \mathbf{x_{min}})} + \mathbf{0.05} \qquad \mathbf{(6)}$$

where x_{norm} is the normalized data set, x_i is the original data set, x_{min} and x_{max} are, respectively, the minimum and maximum values of the original dataset.

III.6.2. Data split

Adhering to standard protocols established for spatial and temporal modeling (see figure 1) the data sets underwent a random split into two subsets. The initial set was utilized in the training process to develop the selected models. This set encompassed samples collected from randomly chosen wells. The subsequent set was held aside to evaluate the performance of the outcomes.

In the upper chellif aquifer, we used 36 wells for training process (70%) and 15 wells (30%) for the validation process.

In the second aquifer (metidja plain), there were 64 wells of data. We divided the wells, assigning 44 wells (75%) to the training process and reserving the remaining 20 wells (25%) for testing.

Regarding the Hassi R'mel aquifer (the third aquifer) we collect data from 17 wells. We allocated 12 wells for the training, covering 70% of the data. Following that, we designated 5 wells for the testing.

The measurements are conducted in the laboratory while electrical conductivity is measured on-site, with a variation in database size from one station to another (see Table 5).

Table 5. Size of database used

N°	Aquifer Designation	Start	End	Wells number	Sample Size	training		validation	
						analysis samples	%	analysis samples	%
01	The upper chellif	2008	2016	51	191	135 (36 wells)	70	56 (15 wells)	30
02	Mitidja	2014	2018	64	363	271 (44 wells)	75	92 (20 wells)	25
03	Hassi R’mel	2011	2020	17	168	118 (12 wells)	70	50 (5 wells)	30

III.6.3. Feature selection and input combination

Identifying relationships within the training data helps determine whether one factor is directly influenced by another, indicating the degree of correlation among various factors. Afterwards, we exclude one of the strongly correlated factors, as they will have similar effect.

In our case studies, which aim to assess the effect of various input data combinations on TDS, we utilized feature importance analysis to rank input variables according to their relevance

to output uncertainty. This technique is important for the development of predictive models. The input combinations were selected based on the parameters contribution most significantly to TDS, as determined by the random forest technique. (see tables 06. 08. 10).

Tables 07, 09, and 11 summarize the resulting input combinations at different phases.

Table 6. Feature importance order of the used parameters in Chellif aquifer

Input Variables	**EC**	**Na^+**	**Cl^-**	**SO_4^{2-}**	**Mg^{2+}**	HCO_3^-	K^+	NO_3^-	Ca^{2+}
Value (%)	**70.79**	**12.67**	**07.53**	**04.80**	**02.18**	0.60	0.58	0.48	0.37
Rank	**1**	**2**	**3**	**4**	**5**	6	7	8	9

This ranking parameter for TDS prediction prioritizes monitoring efforts by focusing on parameters with the most significant impact; their relative influence on the TDS parameter dictates priority. Here, Electrical Conductivity (EC) ranks highest, with a contribution of 70.79%, indicating it is the most significant predictor among the variables listed, followed by sodium, chloride, sulfate, and magnesium with contributions of 12.67%, 7.53%, 4.80%, and 2.18%, respectively.

Table 7. Various input combinations used in the modeling

ML.	**Different Input Combinations**
ML_{01}	EC
ML_{02}	EC & Na^+
ML_{03}	EC, Na^+& Cl^-
ML_{04}	EC, Na^+, Cl^-& SO_4^{2-}
ML_{05}	EC, Na^+, Cl^-, SO_4^{2-}& Mg^+

Table 07 shows various input combinations used for modeling TDS. In this application, five combinations were formed based on the results that it gave from the ranking parameters by priority, starting from the simplest combination (EC alone) in Combination 1, with increasing complexity in the parameters considered incrementally introduced Na^+, Cl^-, SO_4^{2-} and, Mg^+.

Table 8. Feature importance order of the used parameters in Mitidja aquifer

Input Variables	**EC**	**TH**	**Cl^-**	**Ca^{2+}**	**Na^+**	SO_4^{2-}	NO_3^-	PH	HCO_3^-	Mg^{2+}	TAC	K^+	CO_3
Value (%)	**84.48**	**4.34**	**3.89**	**2.43**	**1.75**	0.74	0.51	0.46	0.43	0.38	0.30	0.28	6.96E-07
Rank	**1**	**2**	**3**	**4**	**5**	6	7	8	9	10	11	12	13

In the second study, the results of determining the relative impact on the TDS parameter determine priority have changed. For conductivity, it remained in the first rink with a contribution of 84.48%, then the total water hardness in the second rink with 4.34%, followed by Cl^-, Ca^{2+}, and Na^+ with contributions of 3.89%, 2.43%, and 1.75%, respectively.

Table 9. Various input combinations used in the modeling

ML.	**Different Input Combinations**
ML_{01}	EC
ML_{02}	EC & TH
ML_{03}	EC, TH^+& Cl^-
ML_{04}	EC, TH, Cl^-& Ca^{2+}
ML_{05}	EC, TH, Cl^-& Ca^{2+} & Na^+

As shown in Table 09, this study utilizes five distinct input combinations, beginning with conductivity alone. The parameters are incrementally added in each subsequent combination, culminating in the final set, which includes EC, TH, Cl^-, Ca^{2+} and Na^+.

Table 10. Feature importance order of the used parameters de Hassi R'mel aquifer

Input Variables	**EC**	**SO_4^{2-}**	**Cl^-**	**$Na+^-$**	**Ca^{2+}**	Mg^{2+}	K^+	HCO_3^-	T°	NO_3^-
Value (%)	**65.44**	**19.42**	**05.98**	**03.06**	**02.51**	01.89	00.52	00.51	00.45	00.22

For the third investigation, the findings of the calculation of the relative influence on the TDS parameter to decide priority also altered. where the electrical conductivity is as usual in the first rink with a contribution of 65.44%, and the changes occur in the other parameters: in the second

rink, sulfate is followed by chloride, sodium, and then calcium with 19.42%, 05.98%, 03.06%, and 02.51%, respectively, of contribution.

Table 11. Various input combinations used in the modeling

ML.	Different Input Combinations
ML_{01}	EC
ML_{02}	EC & SO_4^{2-}
ML_{03}	EC, SO_4^{2-} & Cl^-
ML_{04}	EC, SO_4^{2-}, Cl^- & Na^+
ML_{05}	EC, SO_4^{2-} , Cl^- & Na^+ & Ca^+

According to table 11, the five several input combinations are as follows in this study: the first combination was the conductivity alone, and gradually increasing in the parameters considered until we reach the last combination represented by: EC, SO_4^{2-}, Cl^-, Na^+, and C^{a+}.

Conclusion

In summary, the materials and methods applied in this study provide a structured approach for integrating hydrochemical data with advanced machine learning techniques. The systematic design of data preparation, model selection, and validation ensures the reliability of the results. These methods form the foundation for the subsequent analysis and discussion of groundwater quality prediction.

CHAPTER IV

RESULTES AND DISCUSSION

IV. RESULTS AND DISCUSSION

IV.1. Introduction

The TDS consist the total dissolved inorganic salts and certain amounts of organic substances that are dissolved in water, influenced by various parameters. In the current study, TDS parameter is the primary focus, which is regarded as one of the most essential parameters for the overall evaluation of groundwater quality [153,154,115]

In this research, Machine Learning models such us Decision Tree, Random Forest and, Support Victor Machine and CatBoost were utilized to forecast TDS. Therefore, ML models have been widely embraced by researchers and academics across various fields and might be useful in predicting future scenarios of nature and conservation processes.

This work employed sensitivity analysis to select the optimal input combination for predicting TDS. The major components were used in the sensitivity analysis to identify which input data (physicochemical parameters) have the highest impact on the TDS output. As previously mentioned, the major elements have a substantial effect on TDS levels in groundwater. Hence, understanding feature importance in ML models might assist researchers in identifying TDS sources. This allows for the control and improvement of groundwater quality and management through pollution reduction or prevention.

The performance of machine learning models was investigated across three different study areas; each model was trained using a training dataset after selecting the most influential input variables, and then evaluated with the testing dataset [155,115]; and the results were compared using of NS and RMSE statistical measures.

This chapter presents the results of applying machine learning models to forecast Total Dissolved Solids (TDS) in three Algerian aquifers: Upper Cheliff, Mitidja, and Hassi R'mel. The analysis combines statistical performance indicators, feature selection, and sensitivity analysis to identify the most influential physicochemical parameters. Comparative evaluation between models highlights the predictive capabilities of each approach in relation to groundwater quality.

IV.2. Forecasting Total Dissolved Solids in The Upper Cheliff Plain

The total dissolved solids were trained and validated by ML models using the data of the Upper

Chellif aquifer; where the table (12) summarizes the performance of multiple ML models: Decision Tree (DT), CatBoost, Random Forest (RF) and Support Vector Regression (SVR) in predicting TDS for various input combinations (ML 01 to ML05). The performance evaluated using the Nash Sutcliffe efficiency (NS) and the Root Mean Square Error (RMSE), with higher NS values indicating greater performance and lower RMSE values suggesting more accurate forecasts.

Table 12 displays numerical findings for the Upper Chellif Plaine

Table 12. Performance indices of the four machine learning models in the validation phase (the Upper Chellif Plain)

Input Combination	Machine Learning Models							
	DT		**CatBoost**		**RF**		**SVR**	
	NS	**RMSE**	**NS**	**RMSE**	**NS**	**RMSE**	**NS**	**RMSE**
ML_{01}	0.6580	0.0869	0.6679	0.0809	0.7596	0.0688	0.9006	0.0442
ML_{02}	0.7960	0.0634	0.8327	0.0574	0.8076	0.0616	0.9017	0.0440
ML_{03}	0.8770	0.0492	0.9178	0.0402	0.9291	0.0374	0.9240	0.0387
ML_{04}	0.8677	0.0510	0.8914	0.0463	0.9346	0.0359	0.9455	0.0328
ML_{05}	0.8691	0.0508	0.9141	0.0411	0.9252	0.0384	0.9390	0.0347

Based on the performance statistics, Forecasting TDS showed very good accuracy performance for all SVM, DT, RF and CatBoost models. According to Table 7, five input combinations were generated using the Decision Tree (DT) model to predict TDS in the groundwater aquifer of the Upper Chellif Plain. The estimated TDS values for all combinations were evaluated using NS and RMSE values ranging from 0.7960 to 0.8770 and 0.0492 to 0.0634, respectively. The best results were achieved using ML03 (Na^+, Cl^-, EC), which showed NS = 0.8770 and RMSE = 0.0492; ML01 and ML02 also provided acceptable results, with NS values of 0.6580 and 0.7960, and RMSE values of 0.0869 and 0.0634, respectively.

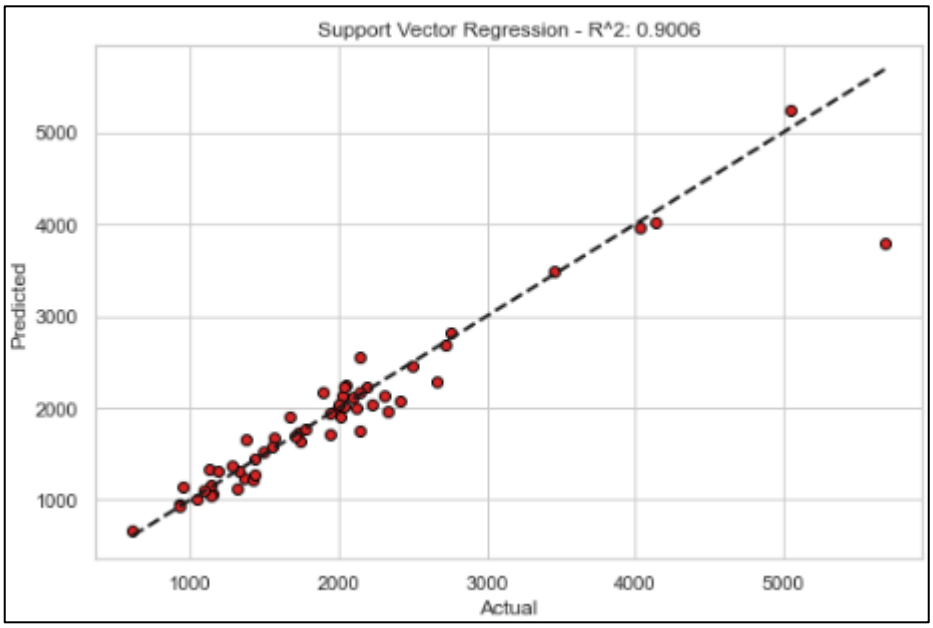
The CatBoost and Random Forest (RF) models followed in performance. Validation results showed that the CatBoost model performed particularly well with ML03 (EC, Na^+, Cl^-), achieving the highest NS value of 0.9178 and the lowest RMSE of 0.0402. ML05 ranked second, with

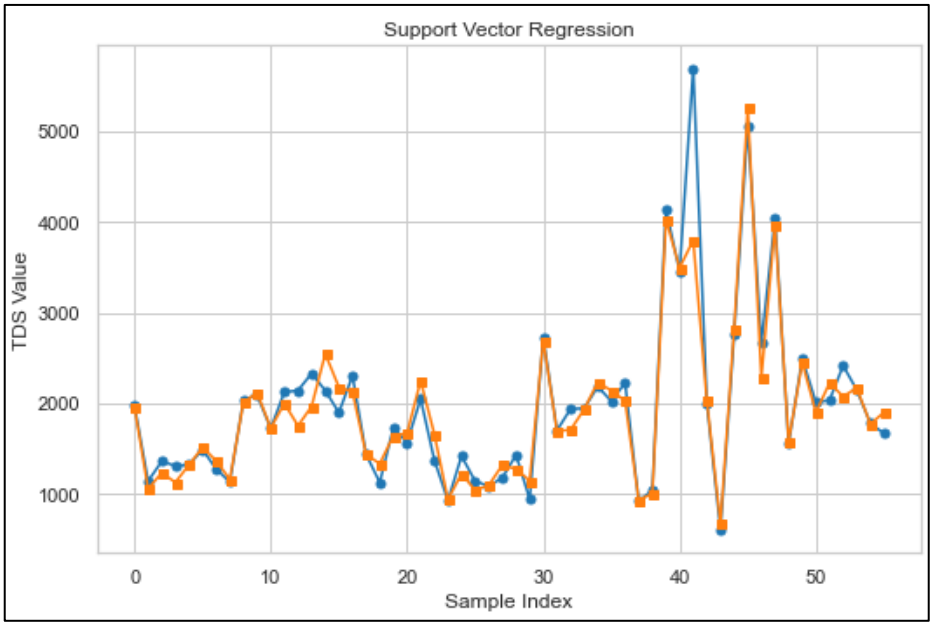
NS = 0.9141 and RMSE = 0.0411. For the RF model, the best performance came from ML04 (EC, Na^+, SO_4^{2-}, Cl^-), with NS = 0.9346 and RMSE = 0.0359. The ML03, ML05, and ML02 models also performed well, with NS values of 0.9291, 0.9252, and 0.8076, and RMSE values of 0.0374, 0.0384, and 0.0616, respectively. When comparing the results in the table, the RF model slightly outperformed the CatBoost and Decision Tree models.

The SVM model achieved the best performance, showing excellent accuracy with NS values ranging from 0.9455 to 0.9006 and RMSE values between 0.0328 and 0.0442, outperforming all other models. Specifically, the best results were observed with ML04 (Na^+, SO_4^{2-}, Cl^-, EC), achieving an NS of 0.9455 and RMSE of 0.0328, followed by ML05 (Mg^{2+}, Na^+, Cl^-, SO_4^{2-}, EC) with NS = 0.9390 and RMSE = 0.0347. The ML03, ML02, and ML01 models also performed very well, with results close to ML04 and ML05, showing NS values of 0.9240, 0.9017, and 0.9006, and RMSE values of 0.0387, 0.0440, and 0.0442, respectively.

The figures below (see Fig 7,8,9,10,11) display the observed and predicted TDS variations during the validation phase for the optimal input combinations of our ML models. These figures compare the predicted TDS values from the Decision Tree, Random Forest, SVM, and CatBoost models with the observed field data. There is a strong agreement between the predicted values and the actual TDS measurements. Based on the scatter plots, line plots, and density plots, no significant differences were observed between the predicted and measured TDS values across the four ML models. Additionally, a clear linear relationship was observed, with the model predictions closely matching the measured TDS values.

As observed in this study, all selected model combinations performed satisfactorily, confirming the effectiveness of machine learning models in predicting TDS values. The results indicate that the SVM model outperformed the Decision Tree, Random Forest, and Catboost models using the same input combinations: ML01 (EC), ML02 (EC, Na^+), ML03 (EC, Na^+, Cl^-), ML04 (EC, Na^+, Cl^-, SO_4^{2-}), and ML05 (EC, Na^+, Cl^-, SO_4^{2-}, Mg^+). The best performance of the SVM model during the validation phase was achieved with combination No. 4 (EC, Na^+, SO_4^{2-}, Cl^-), yielding an NS of 0.9455 and an RMSE of 0.0328.


Support Vector Regression - R^2: 0.9006
Predicted
Actual
1000
2000
3000
4000
5000


Support Vector Regression
TDS Value
Sample Index
1000
2000
3000
4000
5000
0
10
20
30
40
50

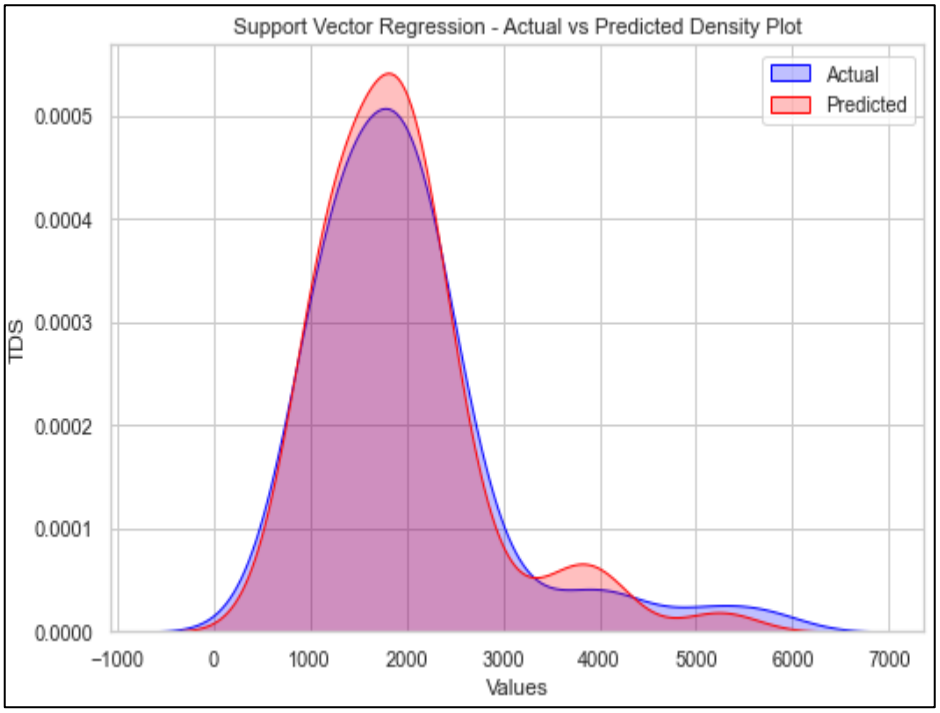
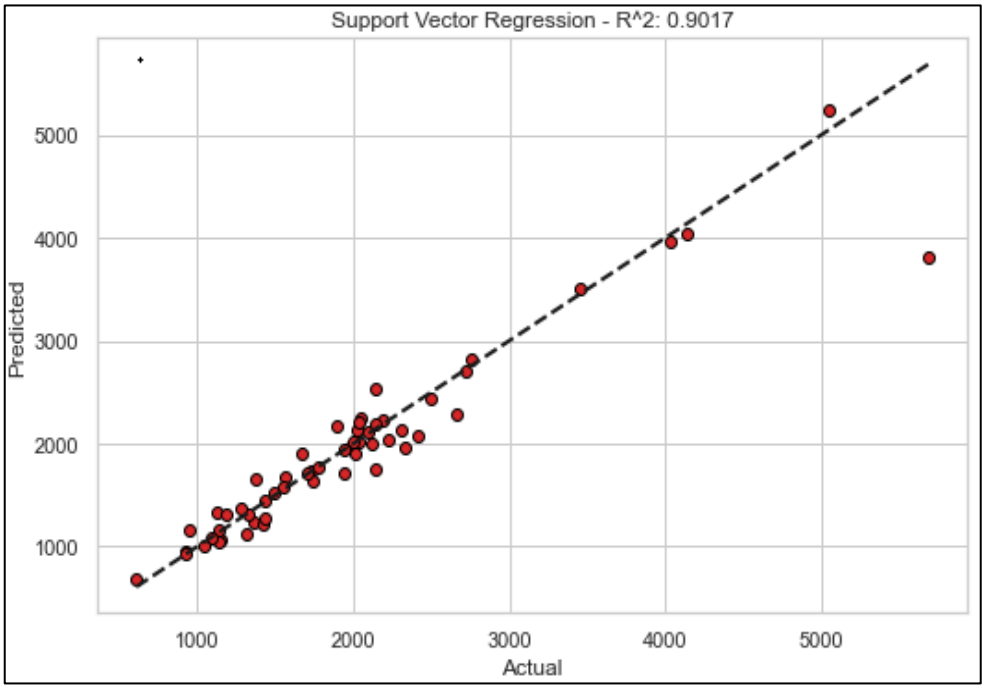
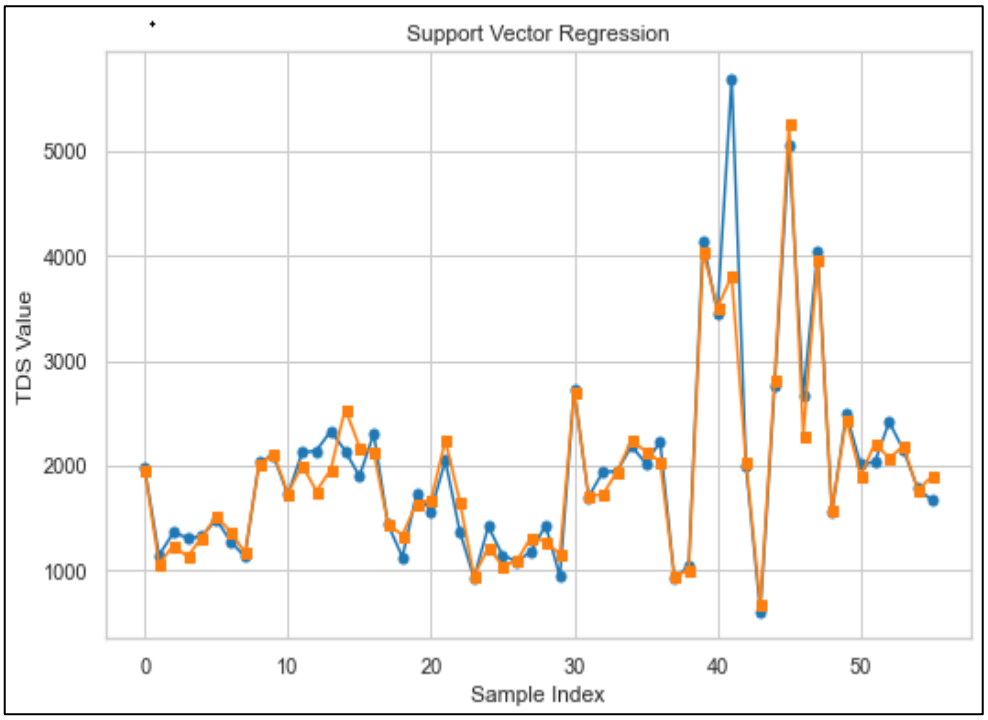


Fig7. Observed and predicted TDS using ML01 combinations: The Upper Cheliff


Support Vector Regression - R^2: 0.9017
Predicted
Actual
1000
2000
3000
4000
5000


Support Vector Regression
TDS Value
Sample Index
1000
2000
3000
4000
5000
0
10
20
30
40
50

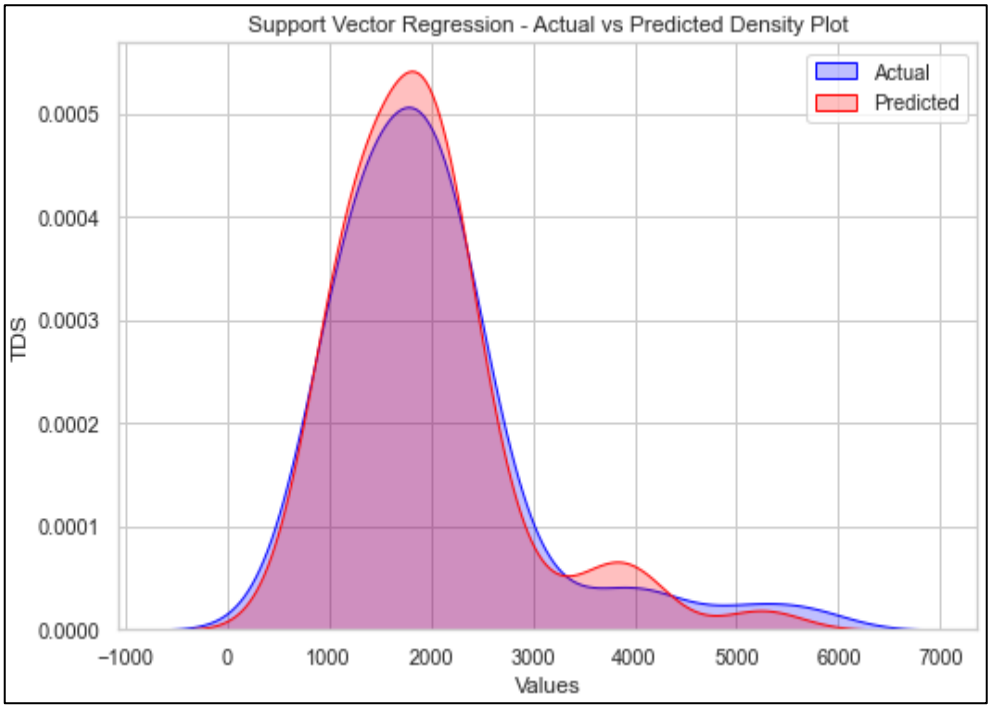
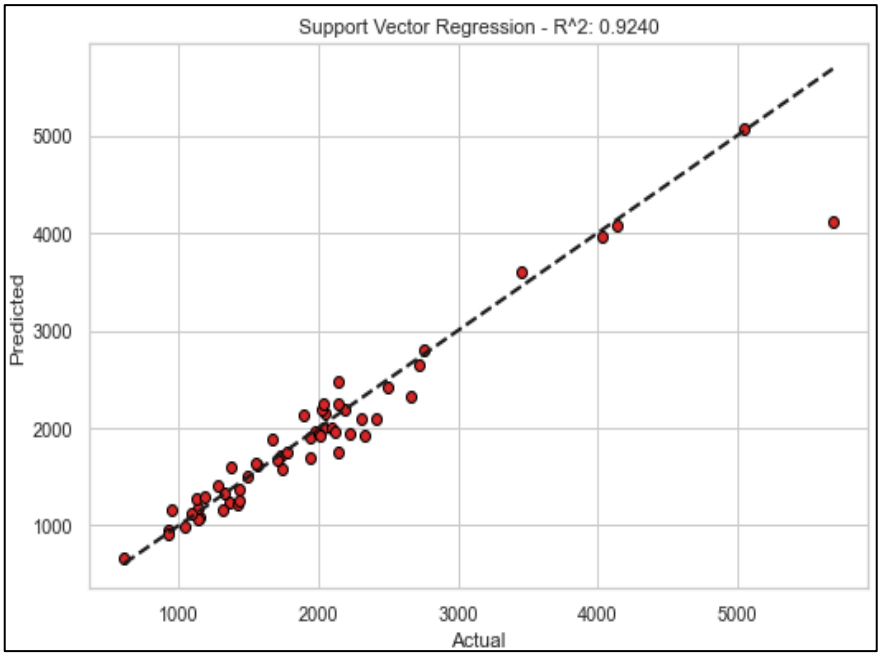
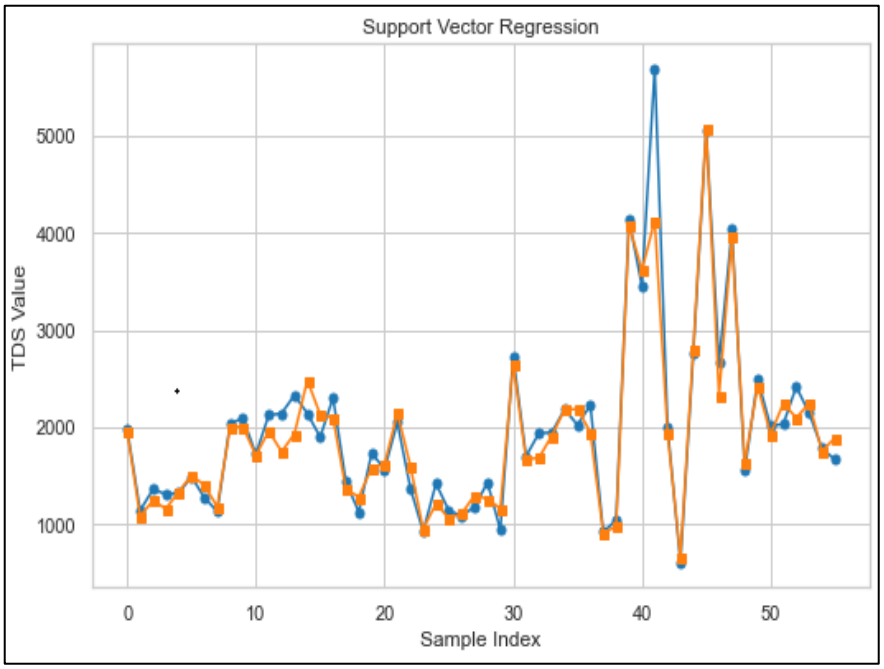


Fig 8. Observed and predicted TDS using ML02 combination: The upper chellif


Support Vector Regression - R^2: 0.9240
Predicted
Actual
1000
2000
3000
4000
5000


Support Vector Regression
TDS Value
Sample Index
1000
2000
3000
4000
5000
0
10
20
30
40
50

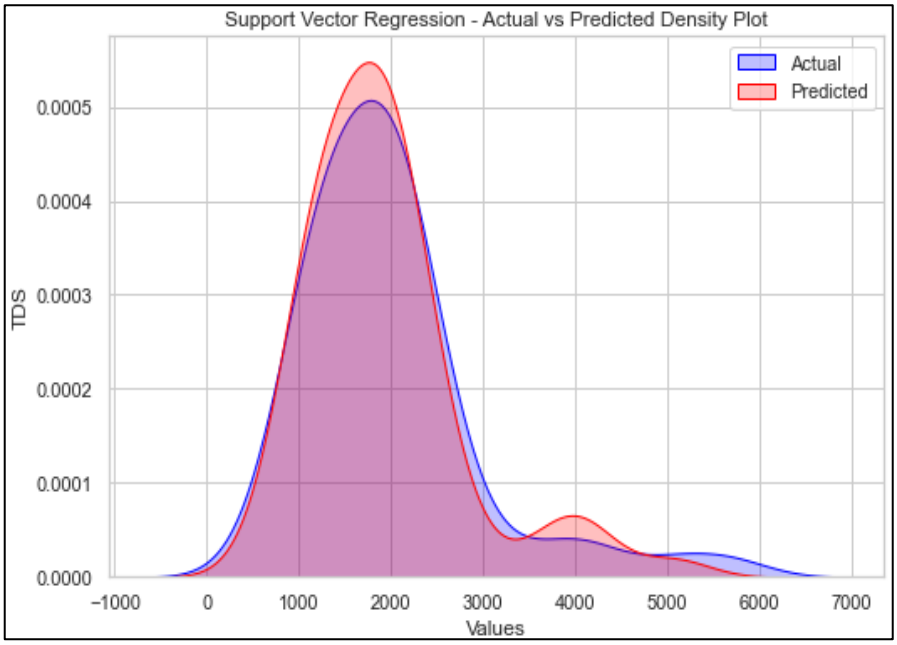
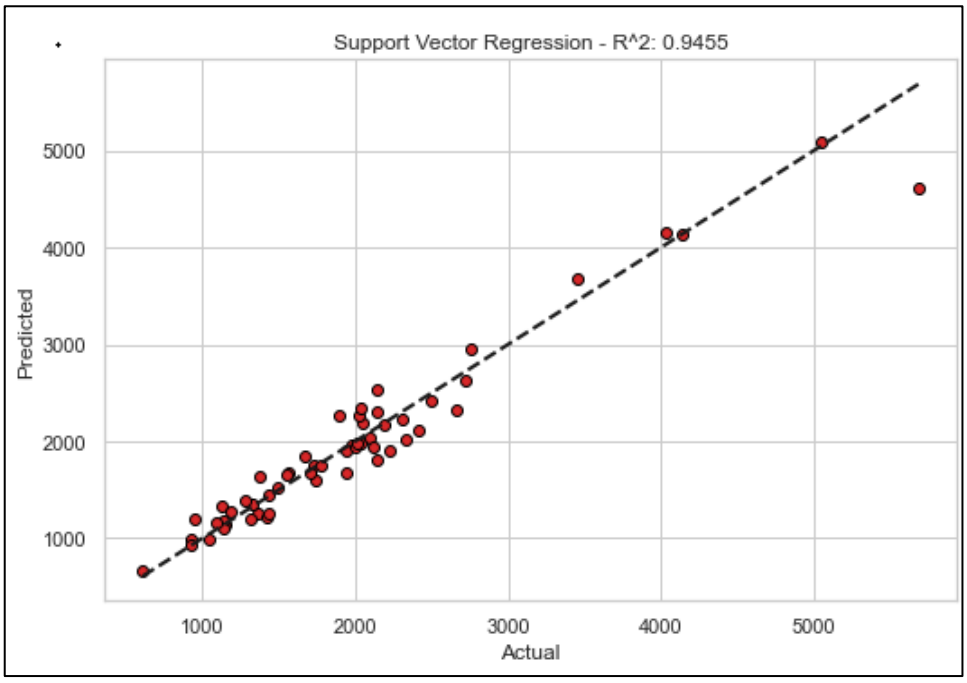
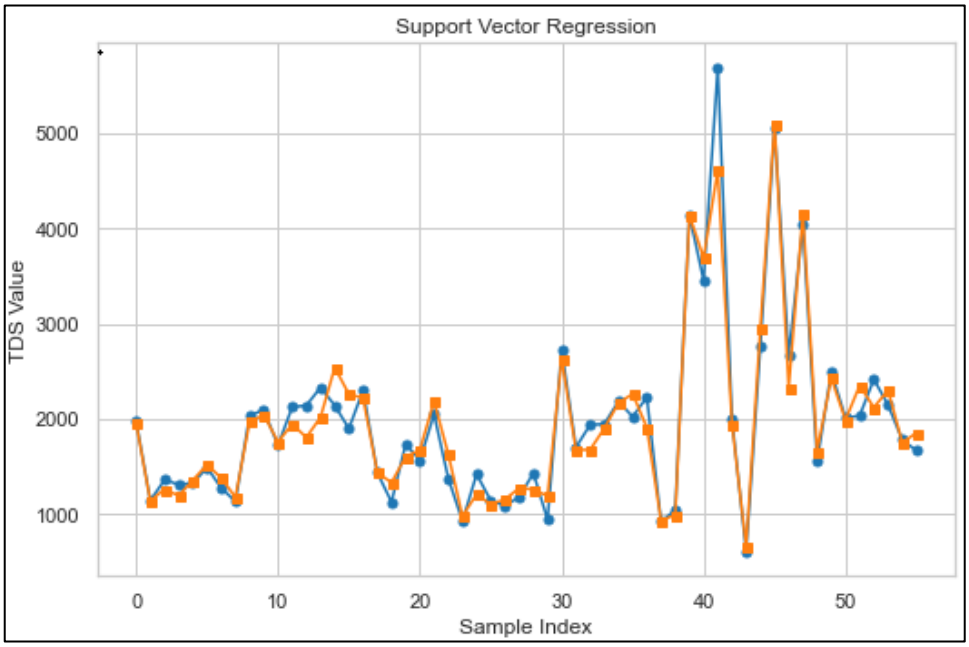


Fig9. Observed and predicted TDS using ML03 combination: The upper chellif


Support Vector Regression - R^2: 0.9455
Predicted
Actual
1000
2000
3000
4000
5000


Support Vector Regression
TDS Value
Sample Index
1000
2000
3000
4000
5000
0
10
20
30
40
50

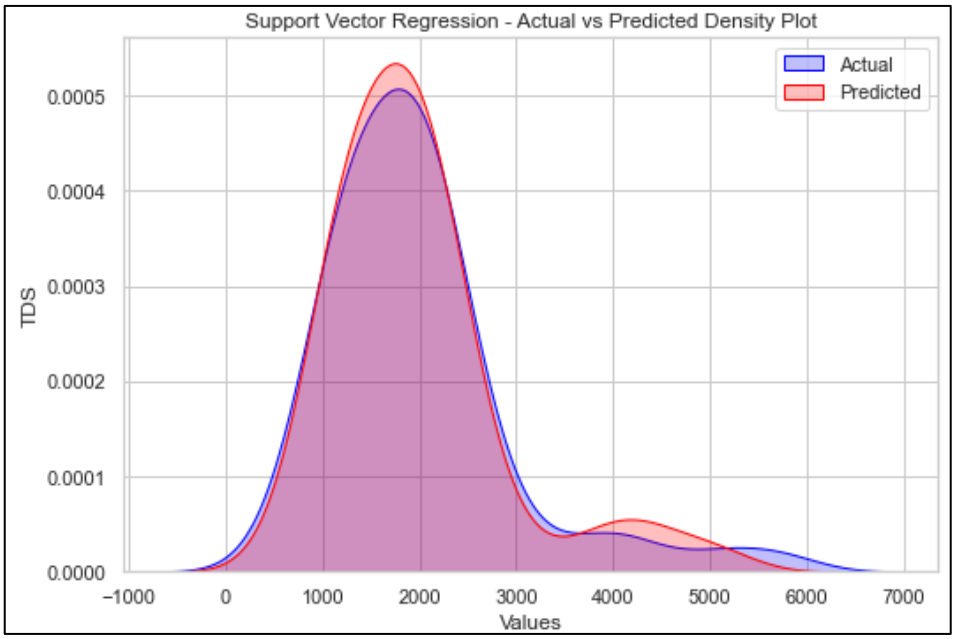
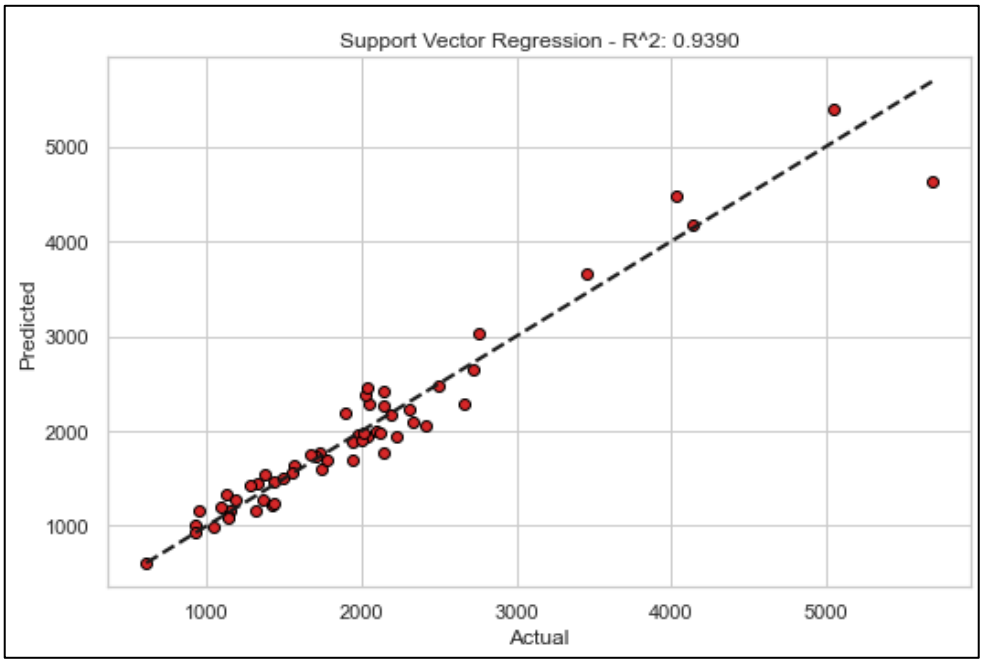
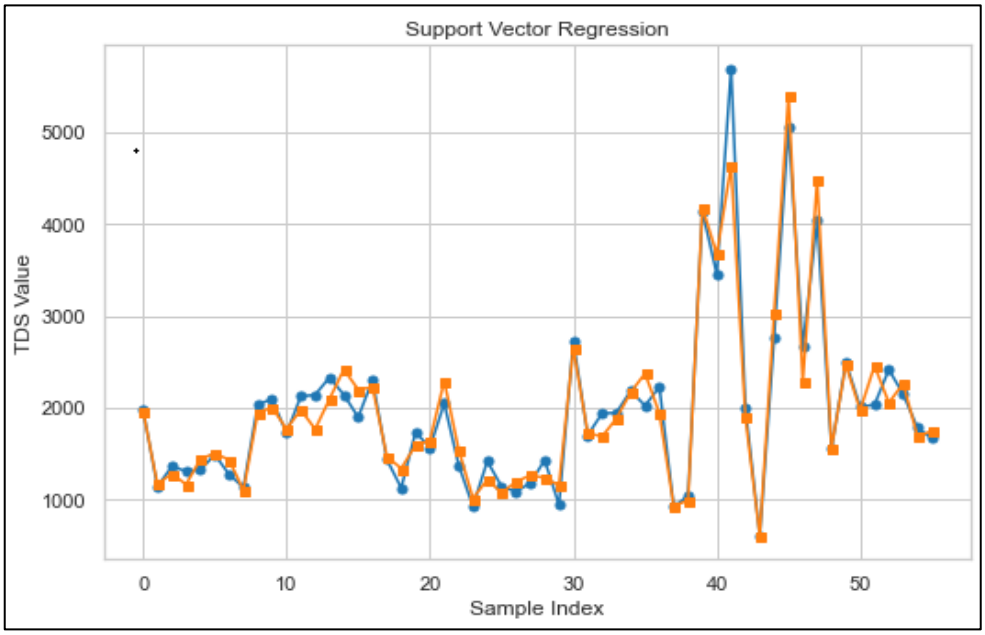


Fig10. Observed and predicted TDS using ML04 combination: The upper chellif


Support Vector Regression - R^2: 0.9390
Predicted
Actual
1000
2000
3000
4000
5000


Support Vector Regression
TDS Value
Sample Index
1000
2000
3000
4000
5000
0
10
20
30
40
50

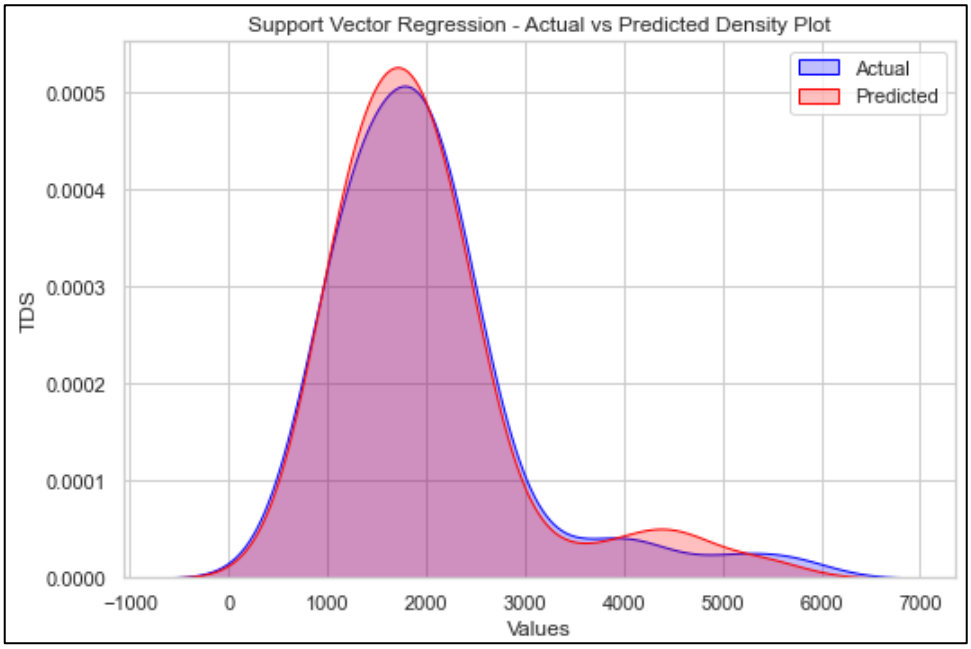
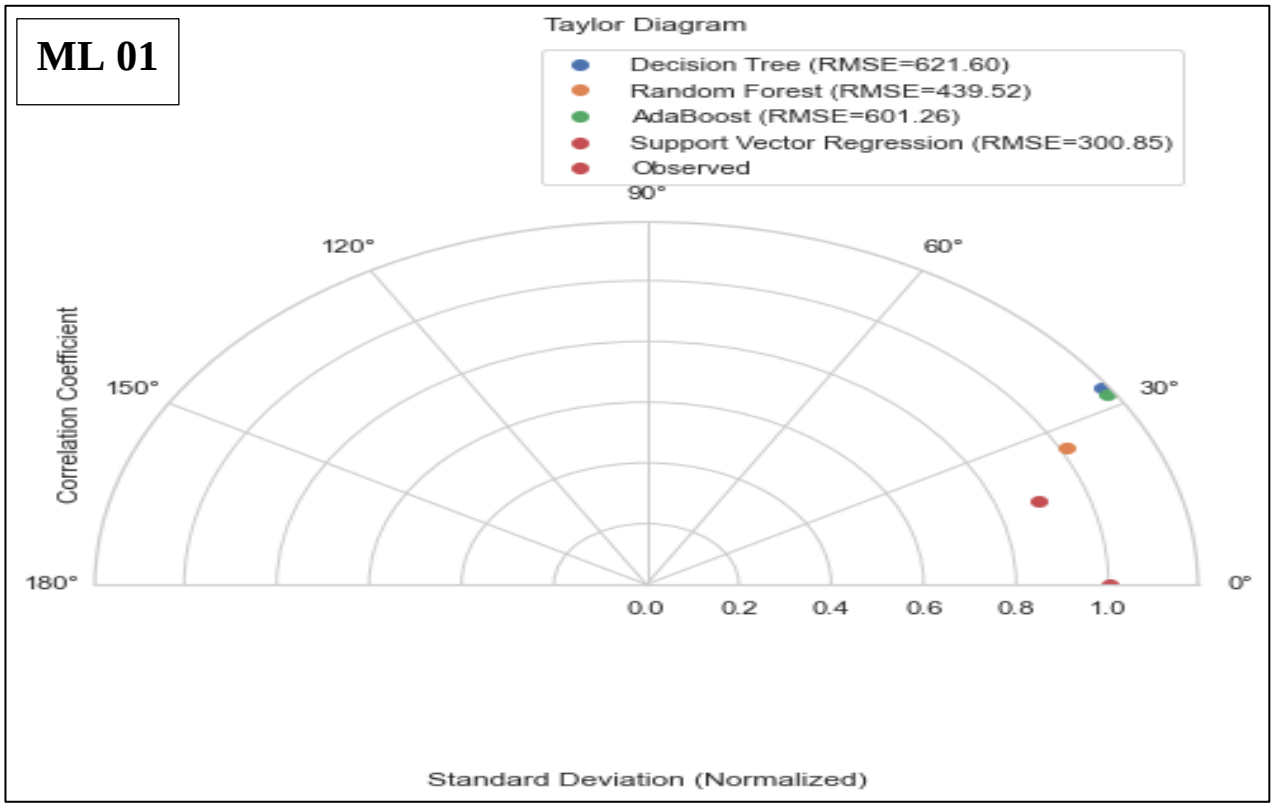
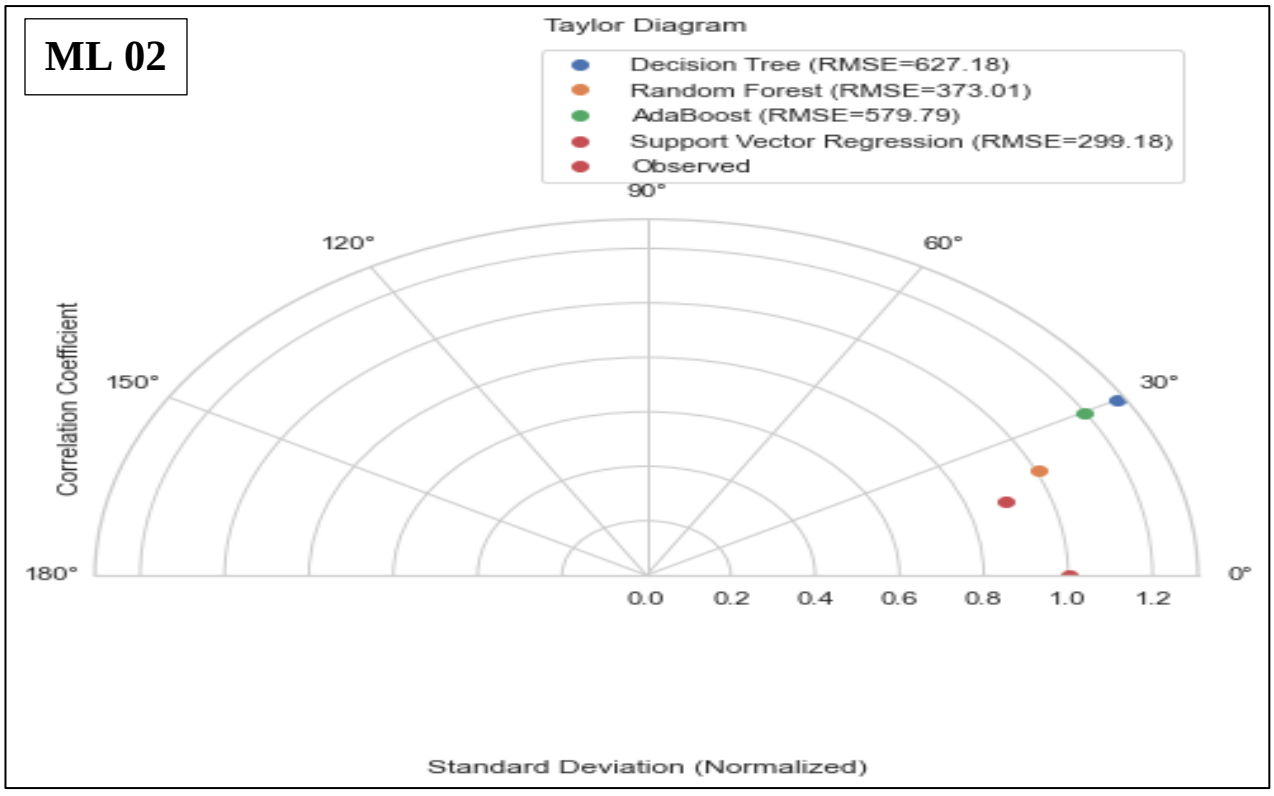
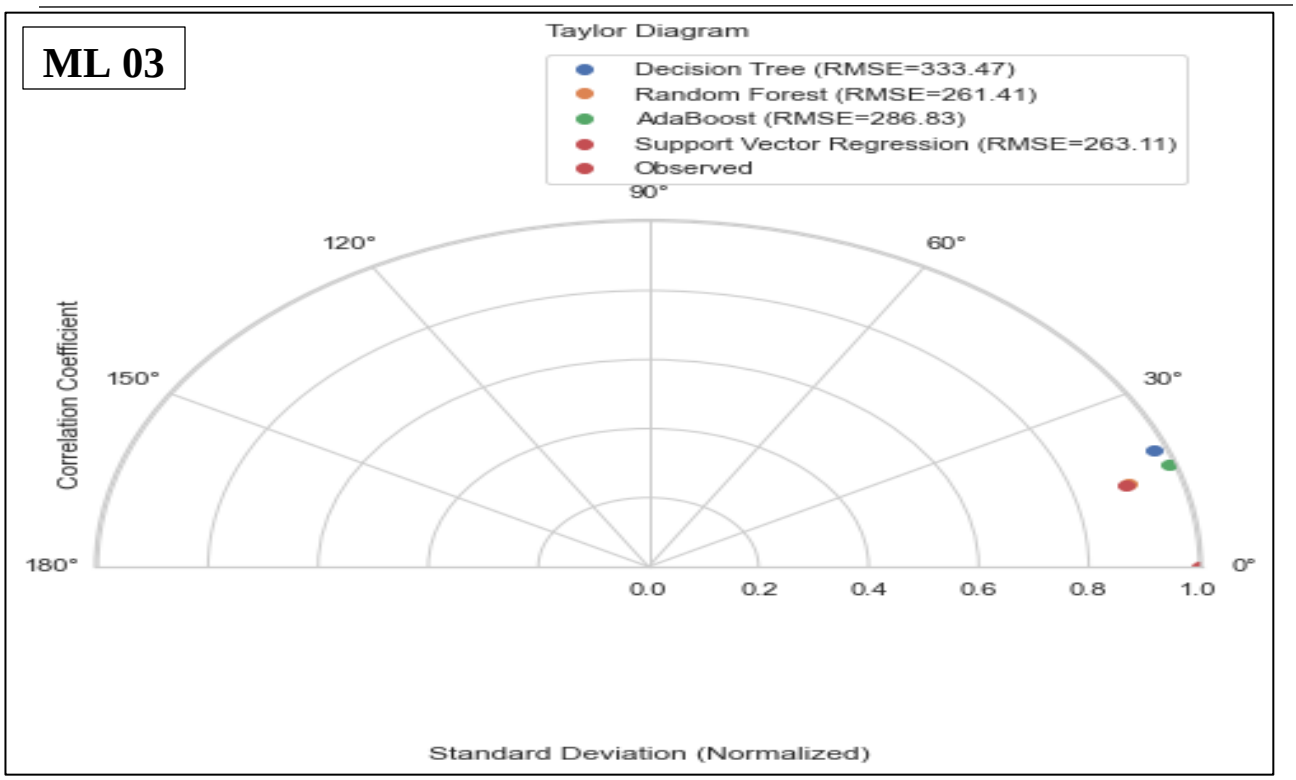


Fig11. Observed and predicted TDS using ML05 combination: The upper chellif

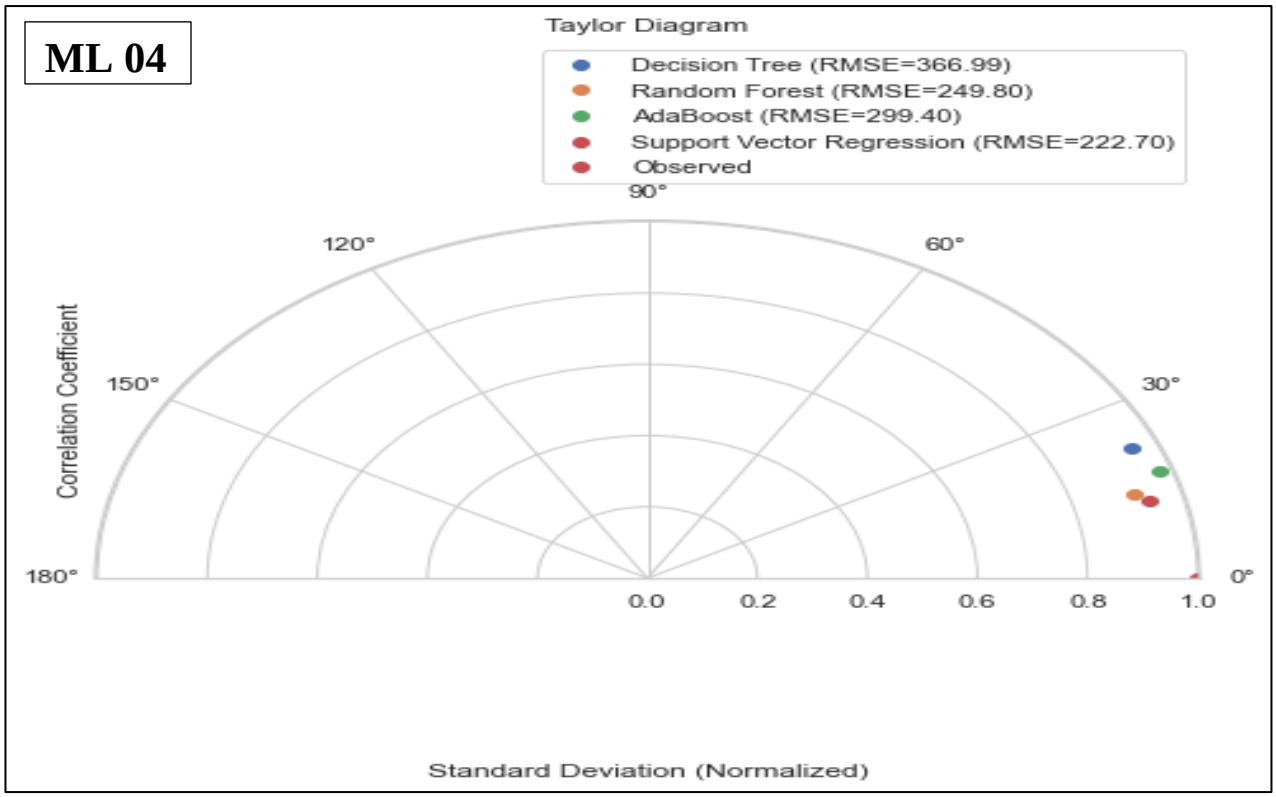

ML 01
Taylor Diagram
Decision Tree (RMSE=621.60)
Random Forest (RMSE=439.52)
AdaBoost (RMSE=601.26)
Support Vector Regression (RMSE=300.85)
Observed
90°
120°
60°
150°
30°
180°
0°
0.0
0.2
0.4
0.6
0.8
1.0
Correlation Coefficient
Standard Deviation (Normalized)


ML 02
Taylor Diagram
Decision Tree (RMSE=627.18)
Random Forest (RMSE=373.01)
AdaBoost (RMSE=579.79)
Support Vector Regression (RMSE=299.18)
Observed
90°
120°
60°
150°
30°
180°
0°
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Correlation Coefficient
Standard Deviation (Normalized)

ML 03


Taylor Diagram
Decision Tree (RMSE=333.47)
Random Forest (RMSE=261.41)
AdaBoost (RMSE=286.83)
Support Vector Regression (RMSE=263.11)
Observed
90°
120°
60°
150°
30°
180°
0°
0.0
0.2
0.4
0.6
0.8
1.0
Correlation Coefficient
Standard Deviation (Normalized)

ML 04


Taylor Diagram
Decision Tree (RMSE=366.99)
Random Forest (RMSE=249.80)
AdaBoost (RMSE=299.40)
Support Vector Regression (RMSE=222.70)
Observed
90°
120°
60°
150°
30°
180°
0°
0.0
0.2
0.4
0.6
0.8
1.0
Correlation Coefficient
Standard Deviation (Normalized)

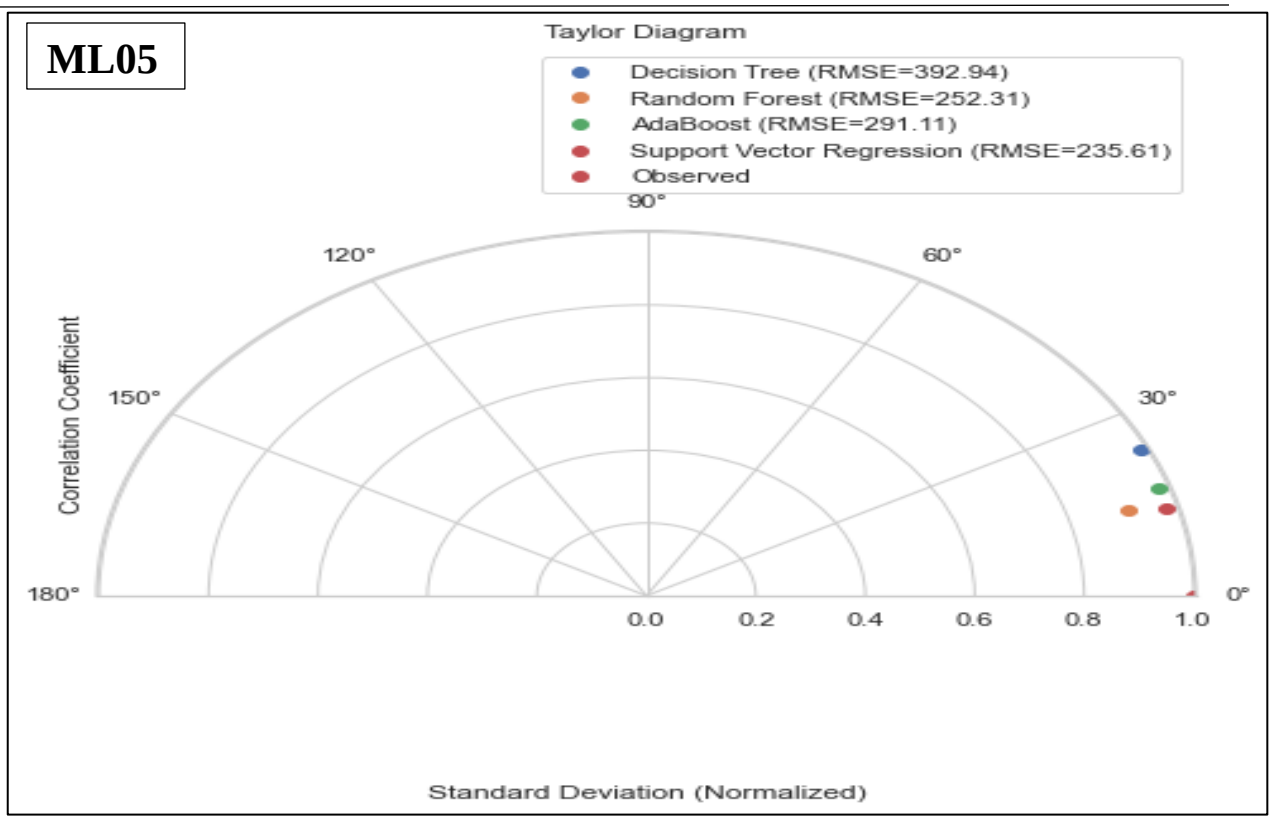


Fig12. The Taylor diagrams for the comparison between the best ML models in all combinations: The upper chellif

The Taylor diagram (see fig 12) evaluates the performances of four machine learning models (DT, RF, CatBoost an SVR) in forecasting TDS parameters, focusing on RMSE, NS and standard deviation. Support Vector Regression (SVR) emerges as best-performing model for forecasting groundwater quality, achieving the lowest RMSE and highest correlation. It is followed by Random Forest (RF), which also demonstrates strong accuracy. CatBoost exhibits moderate performance, while Decision Tree (DT) is the least accurate, with the highest RMSE and lowest correlation.

IV.3. Forecasting Total Dissolved Solids in The Mitidja Plain

The performance of the ML models (DT, CatBoost, RF and, SVR) at the Metidja Plain was evaluated using time-series of TDS predictions and observation by starting with a comparison based on the two model performance evaluation metrics: NS and RMSE. The detailed numerical results can be found in Table 13.

Table 13. Performance indices of the four machine learning models in the validation phase (Metidja plain)

Input Combination	Machine Learning Models							
	DT		CatBoost		RF		SVR	
	NS	RMSE	NS	RMSE	NS	RMSE	NS	RMSE
ML_{01}	0.896	0.1162	0.915	0.1007	0.920	0.0988	0.924	0.0946
ML_{02}	0.852	0.1330	0.916	0.1027	0.908	0.1016	0.926	0.0938
ML_{03}	0.870	0.1284	0.920	0.1044	0.915	0.0975	0.927	0.0932
ML_{04}	0.874	0.1263	0.918	0.1011	0.912	0.0996	0.920	0.0956
ML_{05}	0.875	0.1244	0.916	0.1051	0.911	0.0997	0.918	0.0952

According to the performance data, predicting TDS performed extremely well for all SVM, DT, RF, and CatBoost models.

Table (4) shows that in the first application, five input combinations were created using the Decision Tree (DT) model to predict TDS in groundwater quality in the Metidja Plain. The predicted TDS values for all combinations were assessed using NS and RMSE values of 0.852 to 0.896 and 0.1330 to 0.1162, respectively. The first combination (EC) produced the greatest results, with an NS of 0.896 and an RMSE of 0.1162. Both ML05 and ML04 produced satisfactory results, with NS values of 0.875 and 0.874, and RMSE values of 0.1244 and 0.1263, respectively.

The CatBoost and Random Forest (RF) models ranked next in performance. Validation results revealed that the CatBoost model excelled with ML01 (EC), achieving the highest NS value of 0.915 and the lowest RMSE of 0.1007. In second place was ML04, with an NS of 0.918 and an RMSE of 0.1011. For the RF model, the results are very close to each other, however we consider the best results to have been obtained from ML01. (EC), with an NS of 0.920 and an RMSE of 0.0988. The ML03, ML04, ML05 and ML02 models also performed well, showing NS values of 0.915, 0.912, 0.911 and 0.908, and RMSE values of 0.0975, 0.0996, 0.0997 and 0.1016, respectively. Overall, the RF model beat the CatBoost and Decision Tree models with marginally

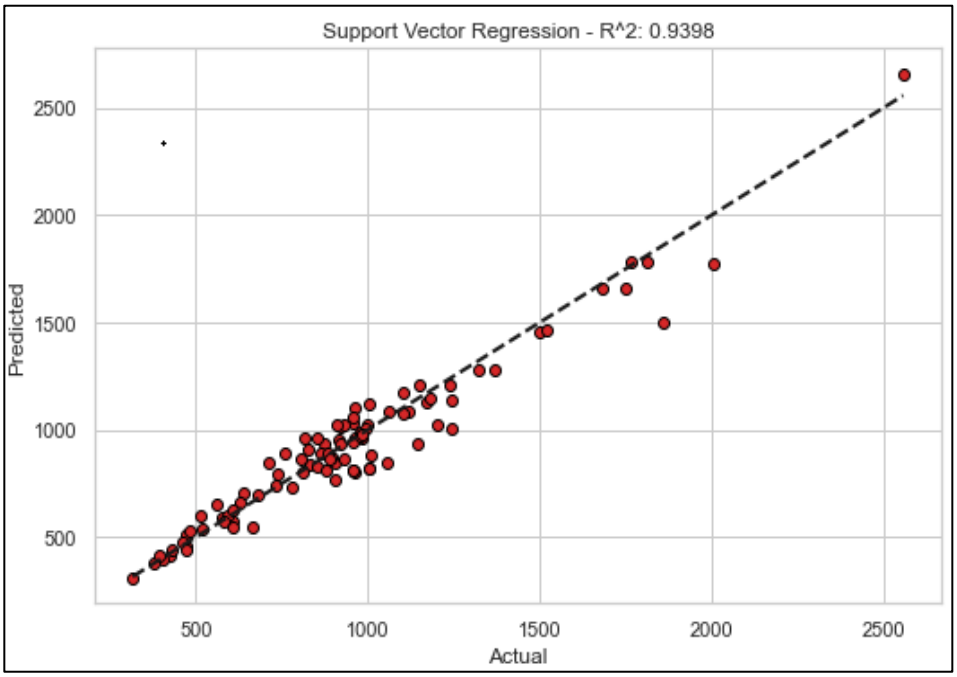
outperformed.

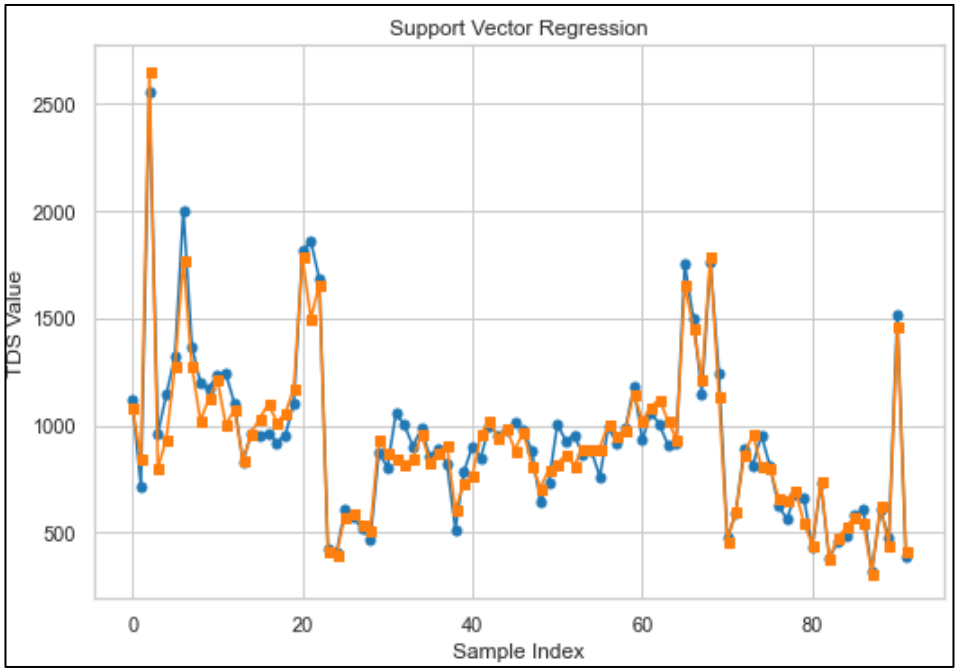
The SVR model outperformed the other models, achieving good accuracy with NS values ranging from 0.927 to 0.918 and RMSE values between 0.0932 and 0.0956. ML03 (EC, TH^{+}, Cl^{-}) had the greatest results, with an NS of 0.927 and an RMSE of 0.0932, followed by ML02 (Mg^{2+}, Na^{+}, Cl^{-}, SO_4^{2-}, EC), which had an NS of 0.926 and an RMSE of 0.0938. The ML01, ML05, and ML04 models also performed well, with results comparable to ML04 and ML05, with NS values of 0.924, 0.918, and 0.920, and RMSE values of 0.0946, 0.0952, and 0.0956, respectively.

We have also visualized these results in scatter plots, box plots, residual plots, prediction error distributions and, density plots (see fig below) and Taylor diagram (fig) for the best ML model in each different combination.

It can be seen from these plots (fig 13,14,15,16,17) that there is high agreement between the values predicted and the values measured by the SVR machine learning model.

There was no apparent bias between the observed and predicted values, and the models effectively captured the extreme TDS values. This strong performance can be attributed in part to the high correlation between TDS and CE, TH, Cl, Ca, Na parameters in different combinations as shown in the Table; as well as the strong predictive capabilities of the data-mining techniques used.


Support Vector Regression - R^2: 0.9398
Predicted
Actual
500
1000
1500
2000
2500


Support Vector Regression
TDS Value
Sample Index
500
1000
1500
2000
2500
0
20
40
60
80

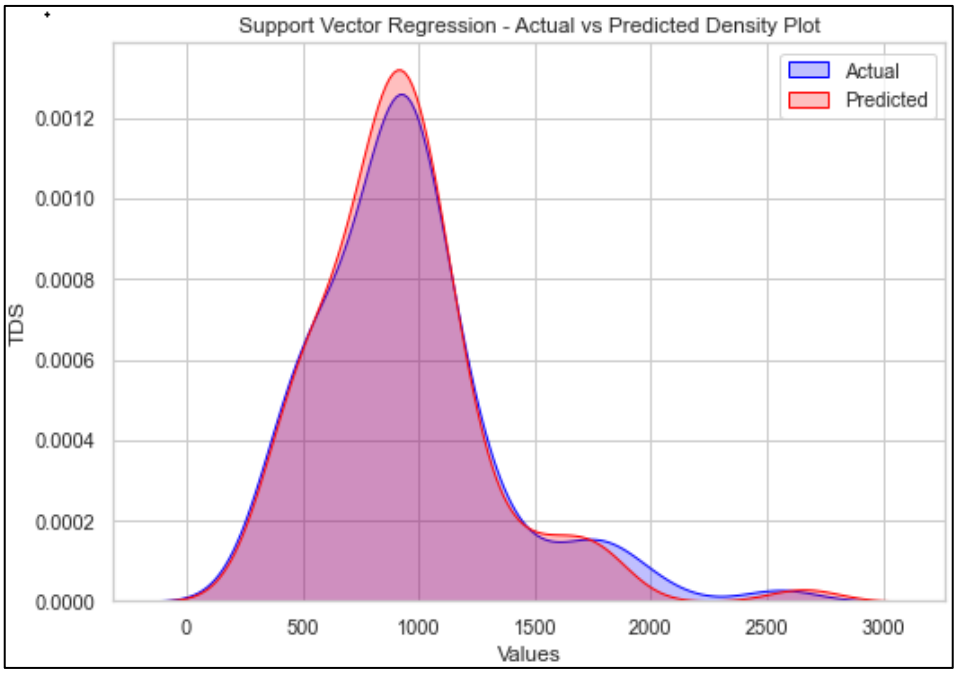
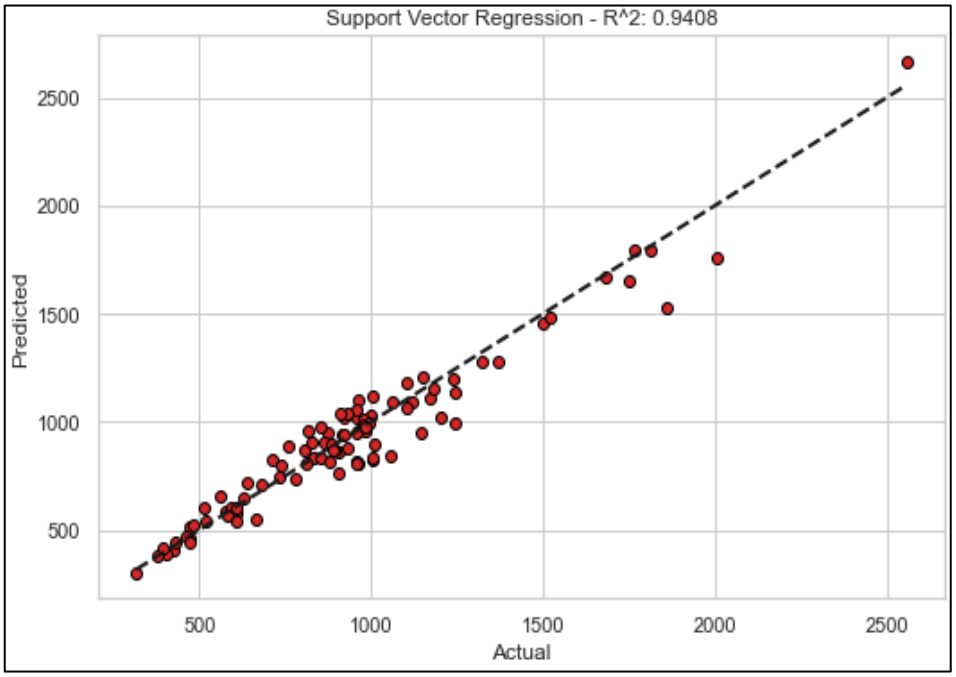
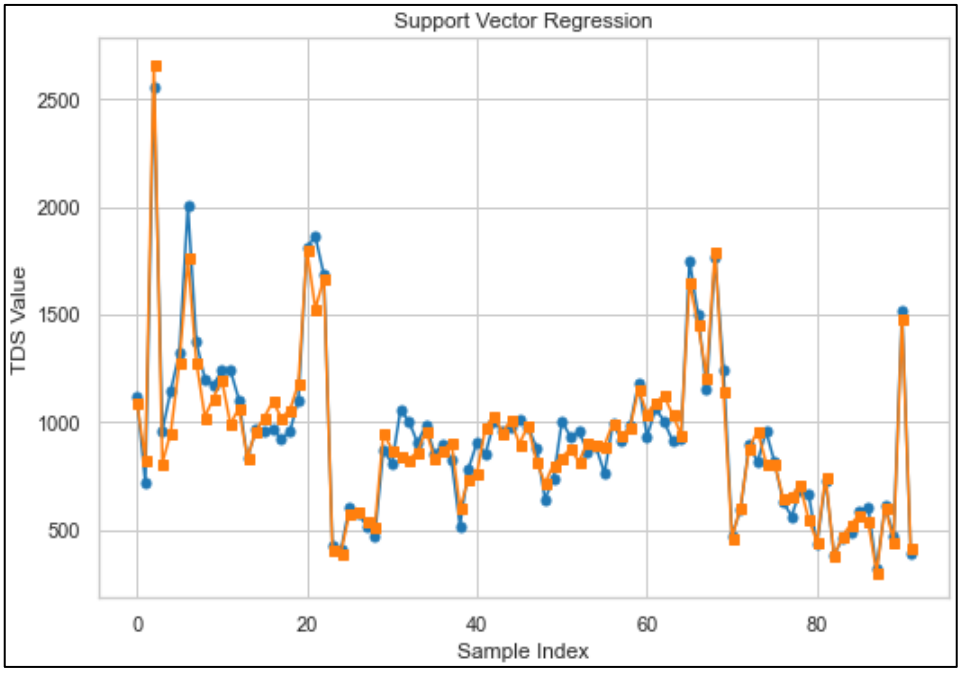


Fig 13. Observed and predicted TDS using ML01 combination: Metidja plain


Support Vector Regression - R^2: 0.9408
Predicted
Actual
500
1000
1500
2000
2500


Support Vector Regression
TDS Value
Sample Index
500
1000
1500
2000
2500
0
20
40
60
80

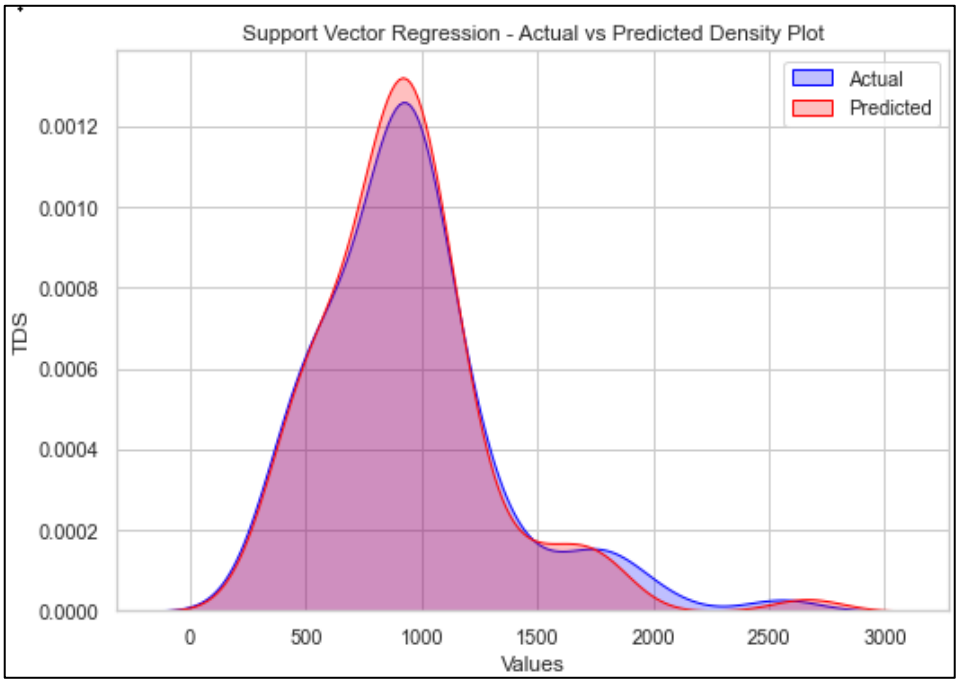
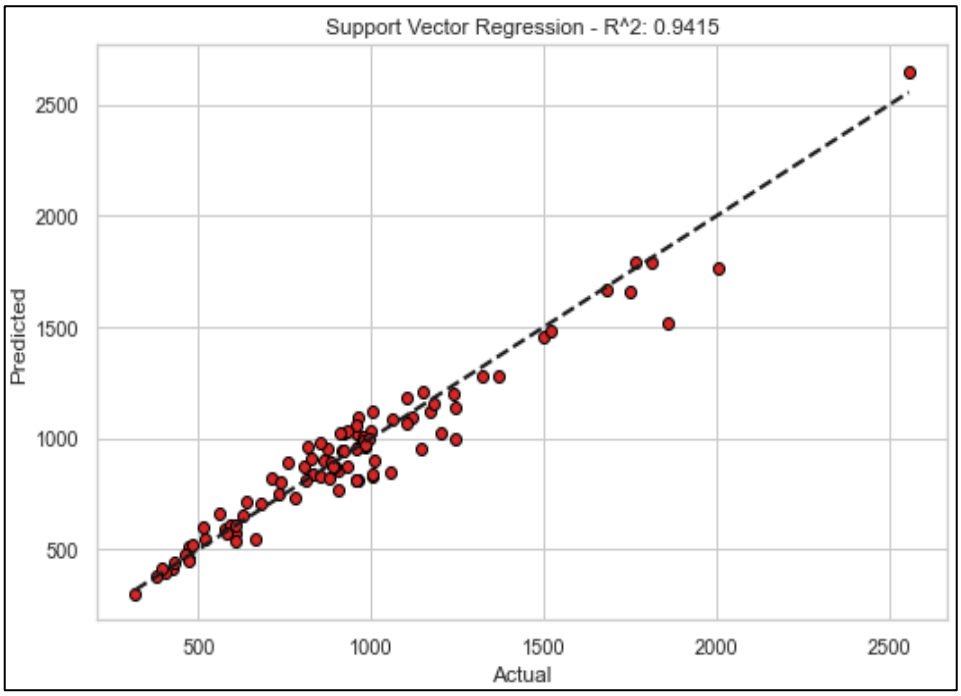
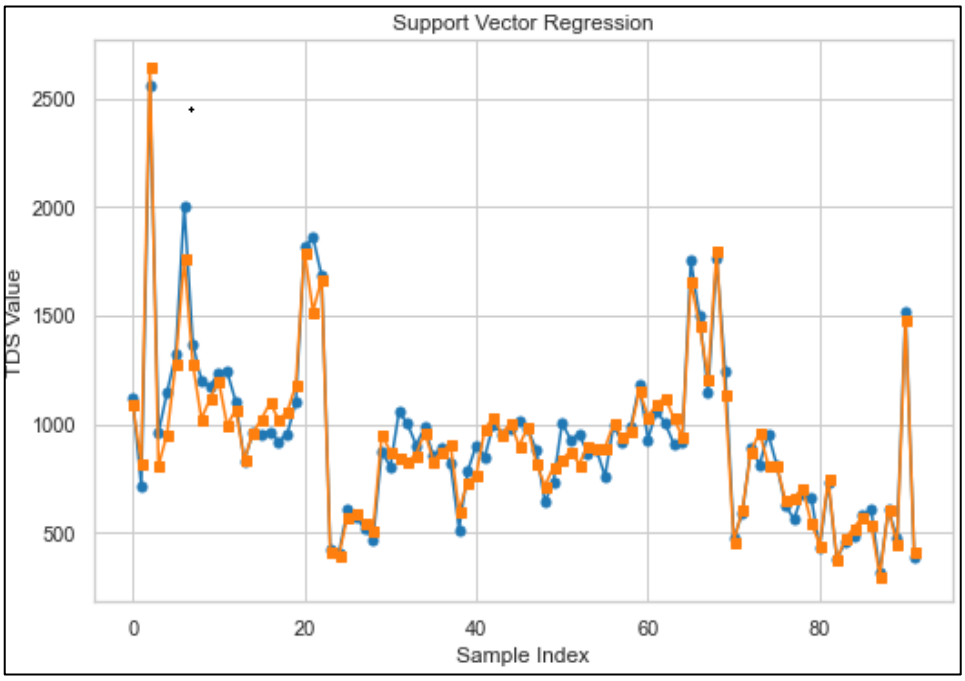


Fig 14. Observed and predicted TDS using ML02 combination: Metidja plain


Support Vector Regression - R^2: 0.9415
Predicted
Actual
500
1000
1500
2000
2500


Support Vector Regression
TDS Value
Sample Index
500
1000
1500
2000
2500
0
20
40
60
80

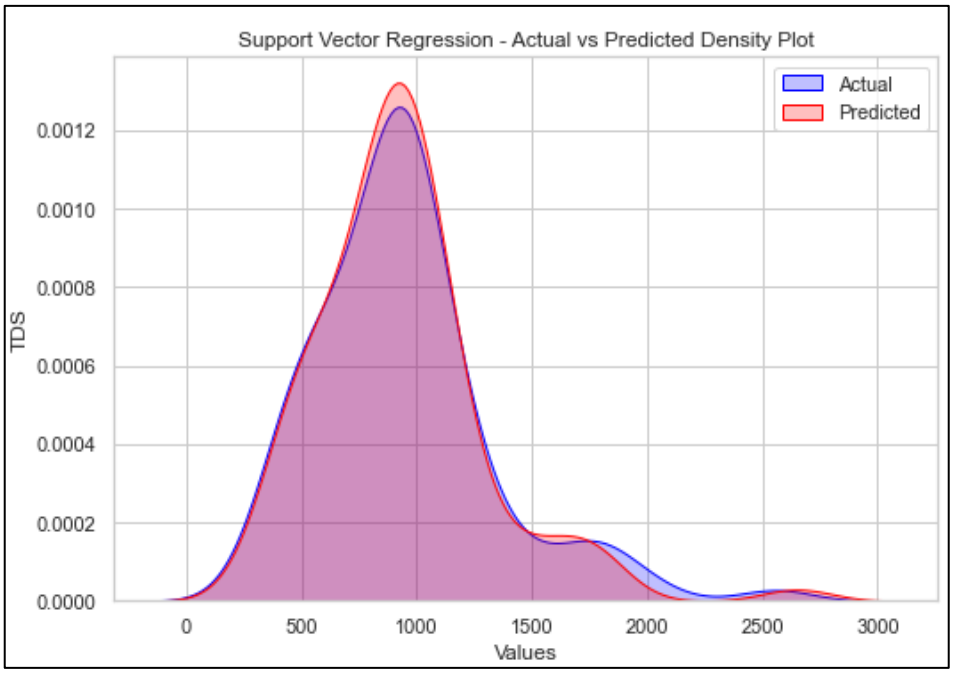
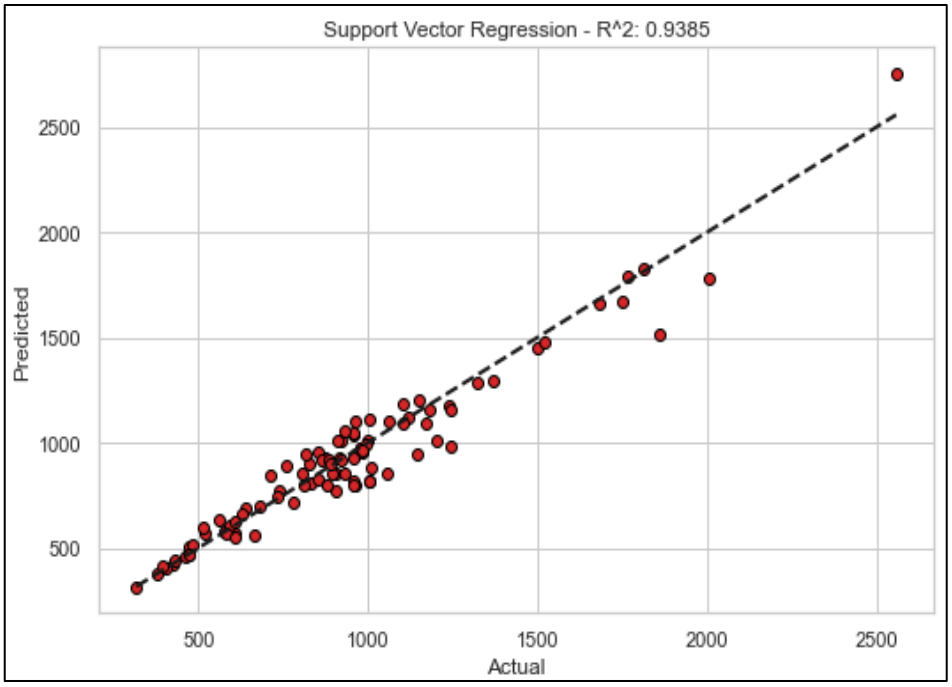
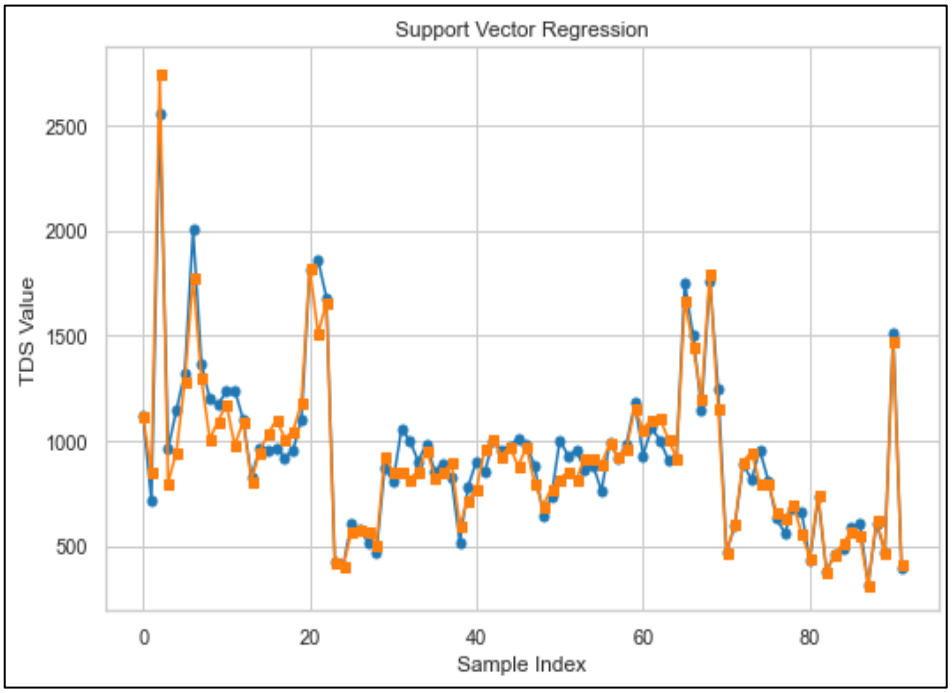


Fig 15. Observed and predicted TDS using ML03 combination: Metidja plain


Support Vector Regression - R^2: 0.9385
Predicted
Actual
500
1000
1500
2000
2500


Support Vector Regression
TDS Value
Sample Index
500
1000
1500
2000
2500
0
20
40
60
80

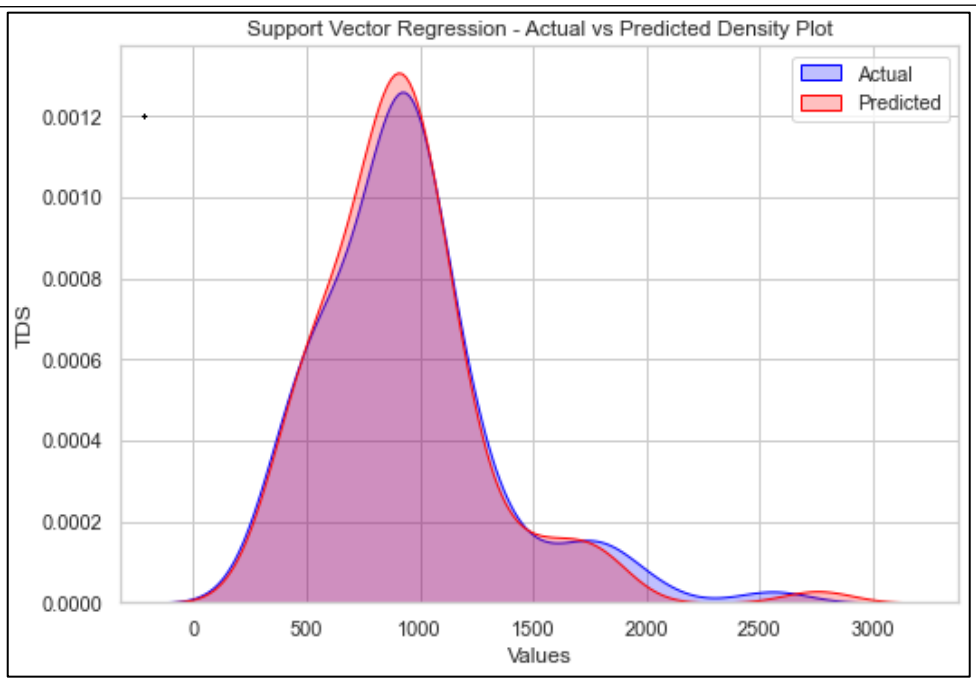
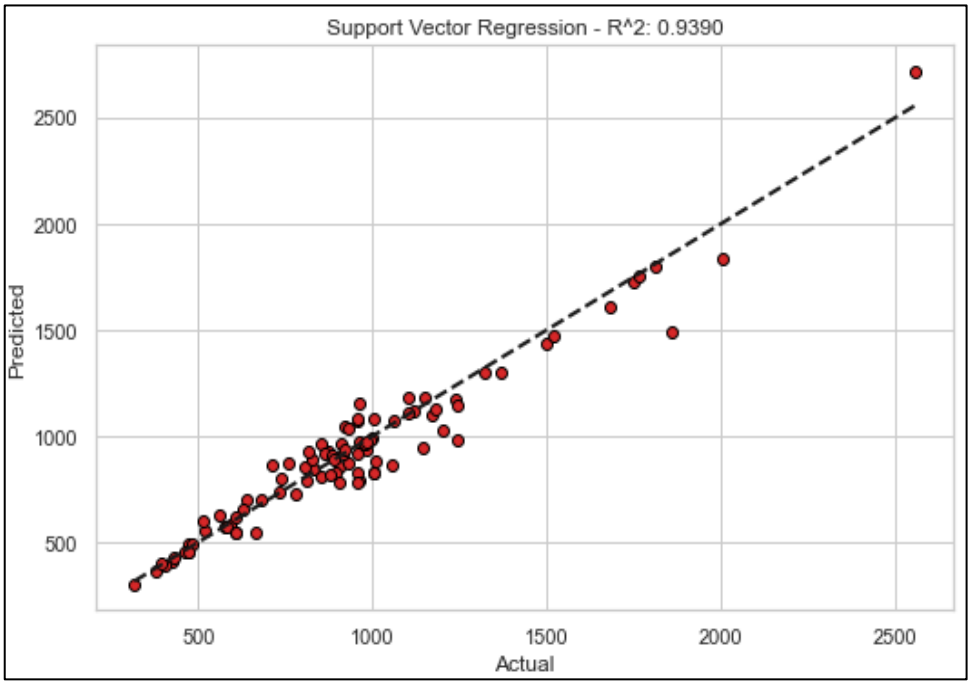
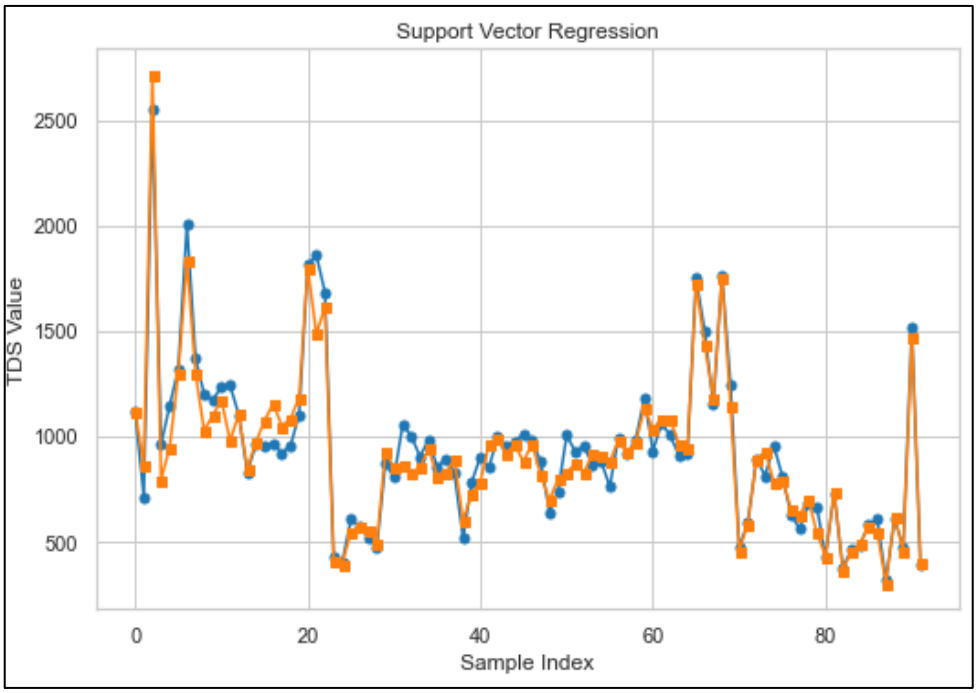


Fig 16. Observed and predicted TDS using ML04 combination: Metidja plain


Support Vector Regression - R^2: 0.9390
Predicted
Actual
500
1000
1500
2000
2500


Support Vector Regression
TDS Value
Sample Index
500
1000
1500
2000
2500
0
20
40
60
80

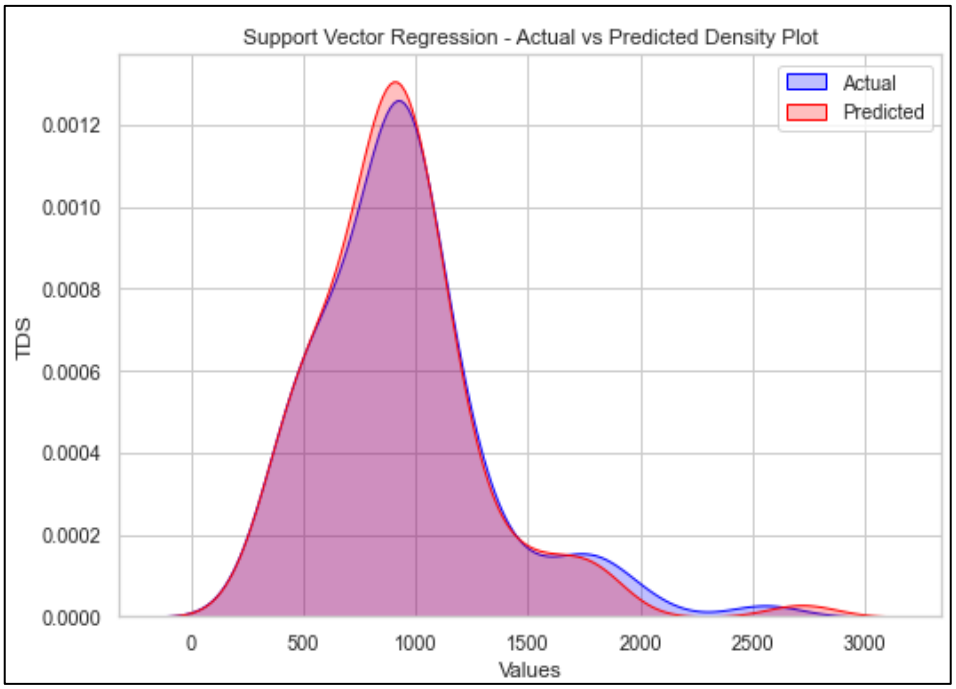
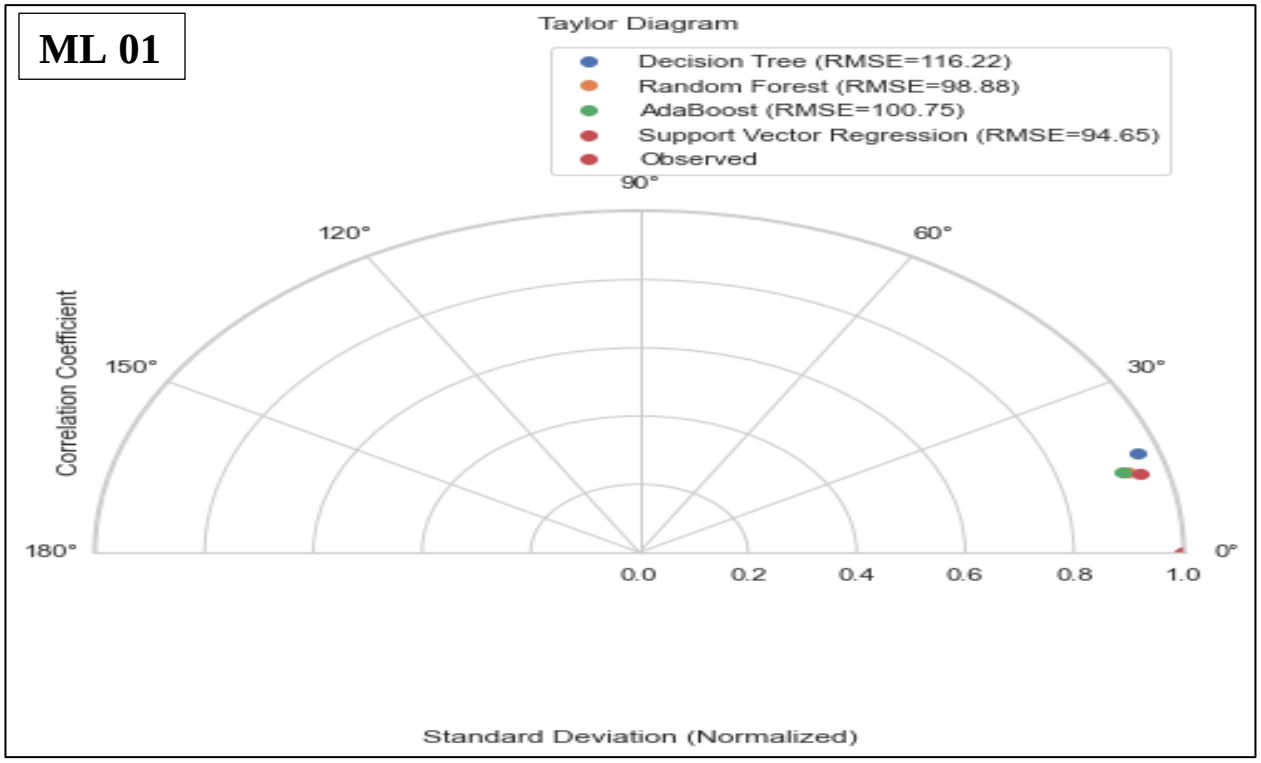
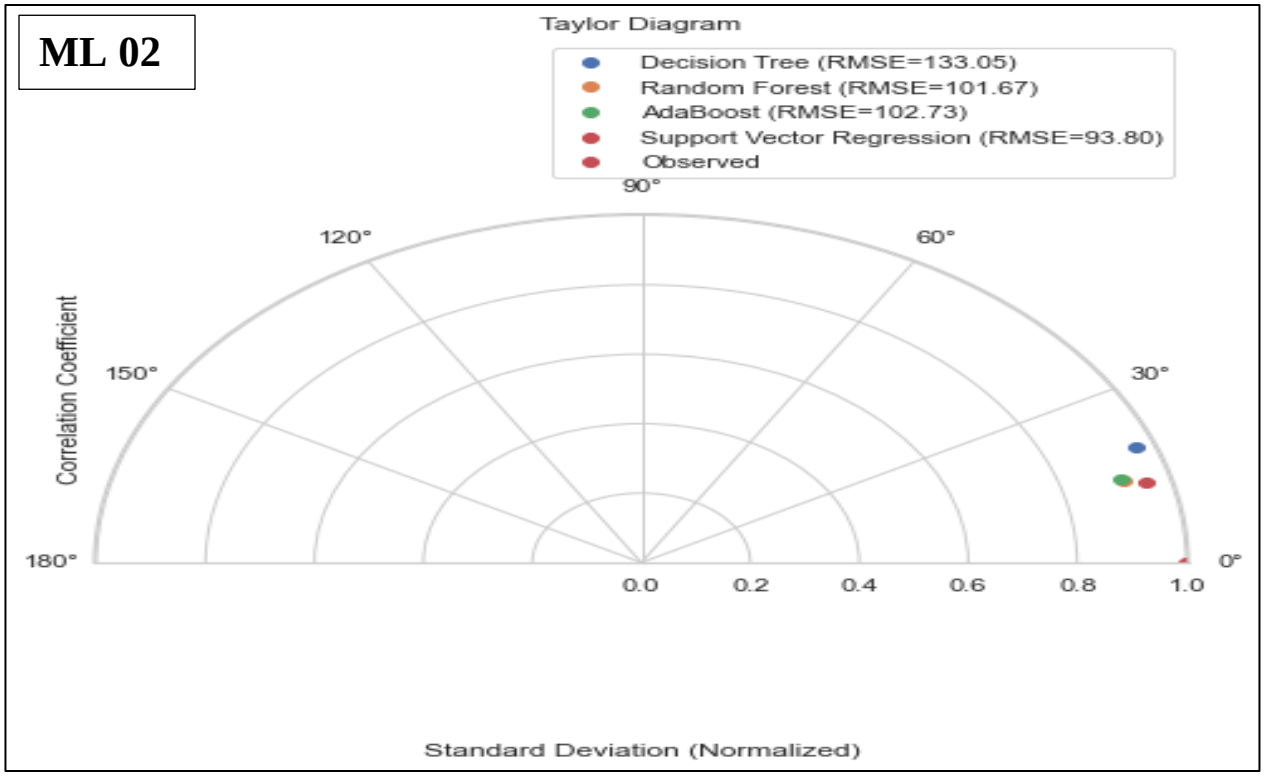


Fig 17. Observed and predicted TDS using ML05 combination: Metidja plain

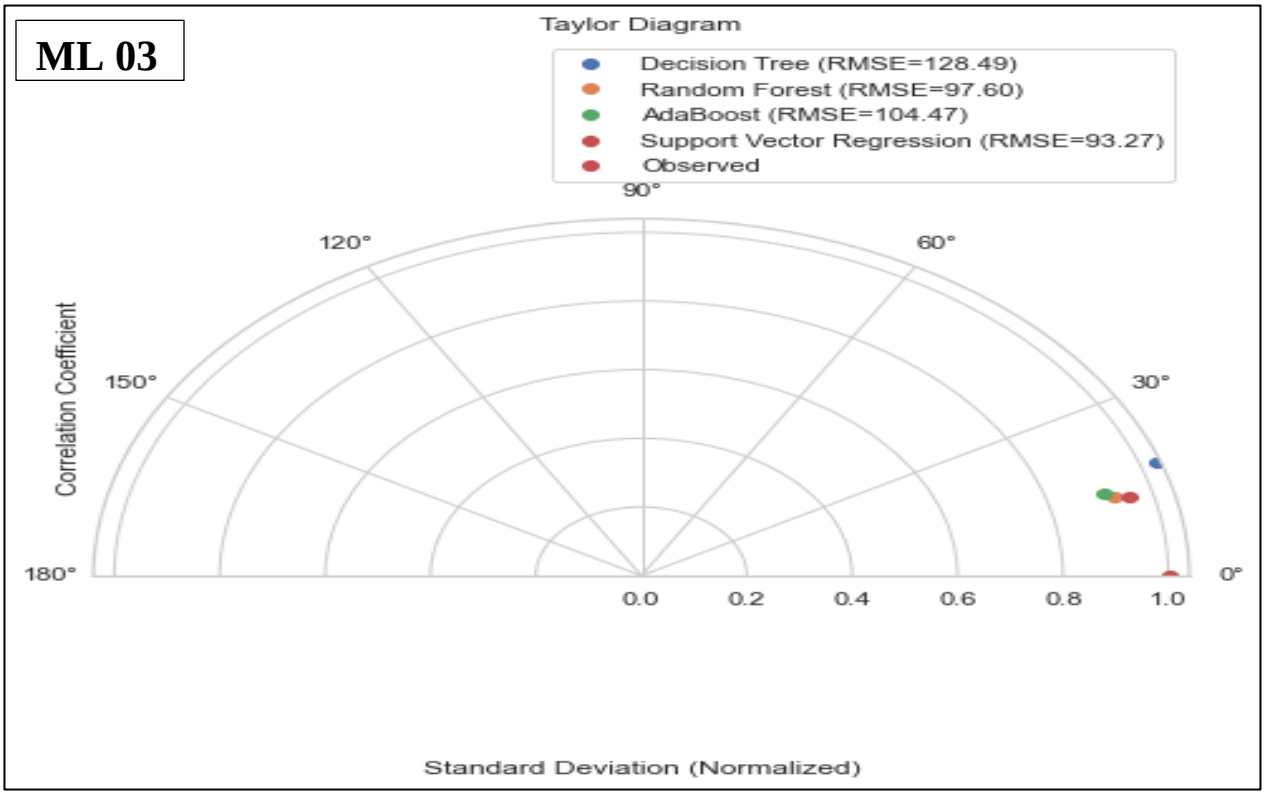
ML 01


Taylor Diagram
Decision Tree (RMSE=116.22)
Random Forest (RMSE=98.88)
AdaBoost (RMSE=100.75)
Support Vector Regression (RMSE=94.65)
Observed
90°
120°
60°
150°
30°
180°
0°
0.0
0.2
0.4
0.6
0.8
1.0
Correlation Coefficient
Standard Deviation (Normalized)

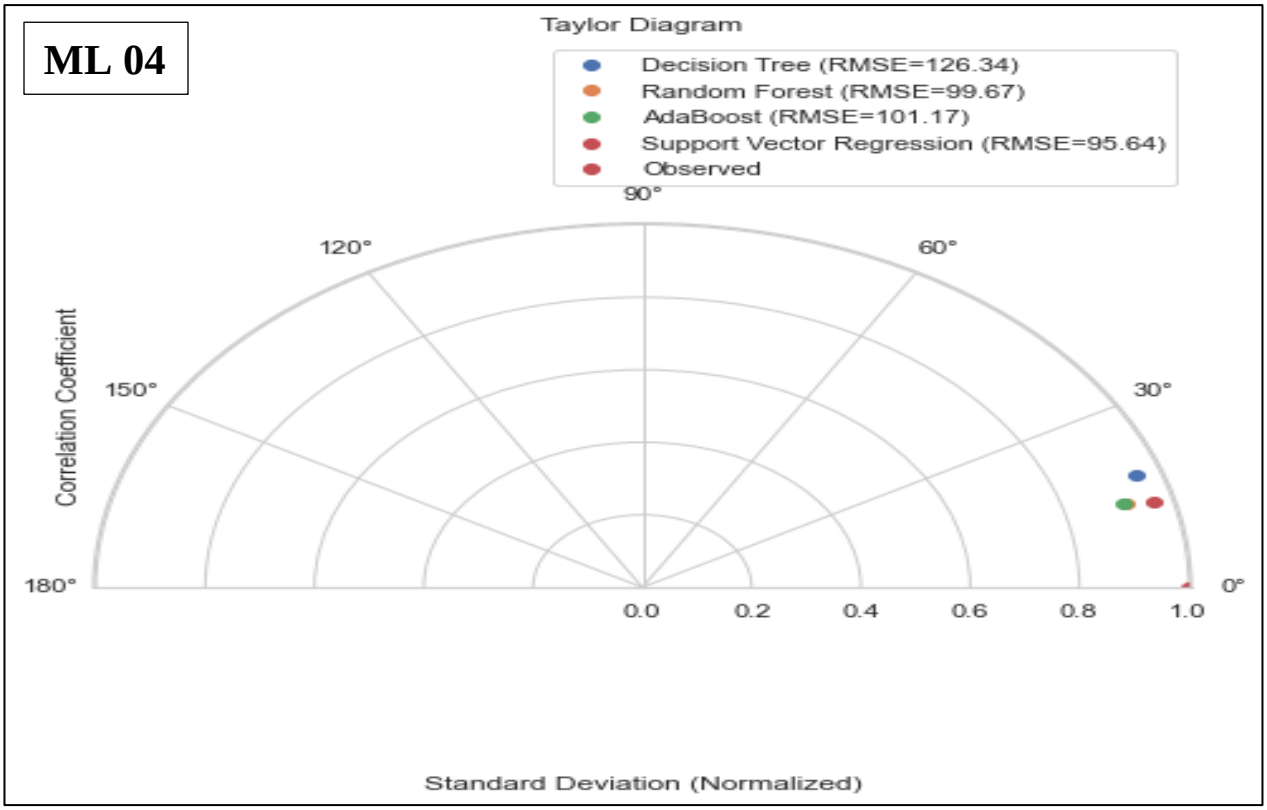
ML 02


Taylor Diagram
Decision Tree (RMSE=133.05)
Random Forest (RMSE=101.67)
AdaBoost (RMSE=102.73)
Support Vector Regression (RMSE=93.80)
Observed
90°
120°
60°
150°
30°
180°
0°
0.0
0.2
0.4
0.6
0.8
1.0
Correlation Coefficient
Standard Deviation (Normalized)

ML 03


Taylor Diagram
Decision Tree (RMSE=128.49)
Random Forest (RMSE=97.60)
AdaBoost (RMSE=104.47)
Support Vector Regression (RMSE=93.27)
Observed
90°
120°
60°
150°
30°
180°
0°
0.0
0.2
0.4
0.6
0.8
1.0
Correlation Coefficient
Standard Deviation (Normalized)

ML 04


Taylor Diagram
Decision Tree (RMSE=126.34)
Random Forest (RMSE=99.67)
AdaBoost (RMSE=101.17)
Support Vector Regression (RMSE=95.64)
Observed
90°
120°
60°
150°
30°
180°
0°
0.0
0.2
0.4
0.6
0.8
1.0
Correlation Coefficient
Standard Deviation (Normalized)

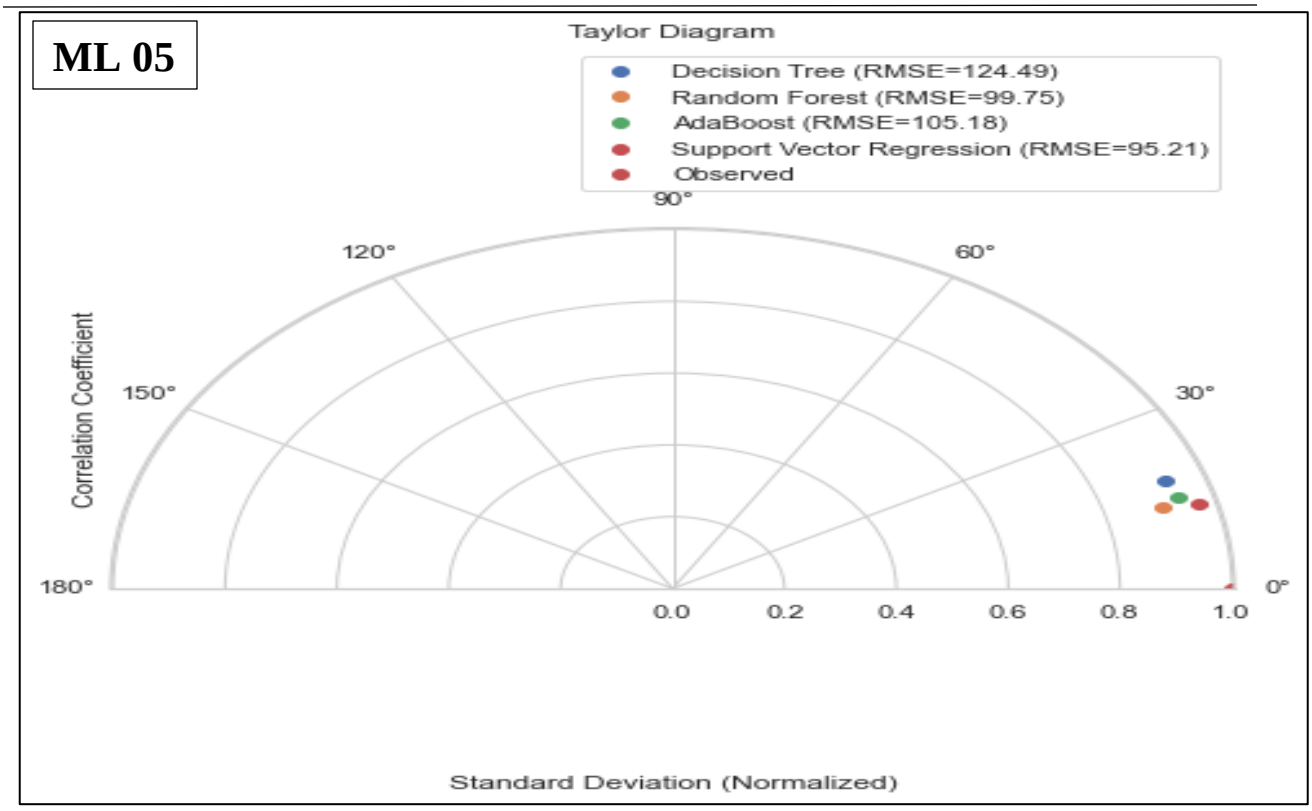


Fig 18. The Taylor diagrams for the comparison between the best ML models in all combinations: Mitidja plain

The observed and modeled TDS variations in the validation phase are presented in Taylor diagram (see fig18) for the best combinations in our ML models. where SVR and RF proved to be the most reliable models for predicting groundwater quality, whereas CatBoost and DT demonstrated very low accuracy.

IV.4. Forecasting Total Dissolved Solids in Hassi R'mel Plain

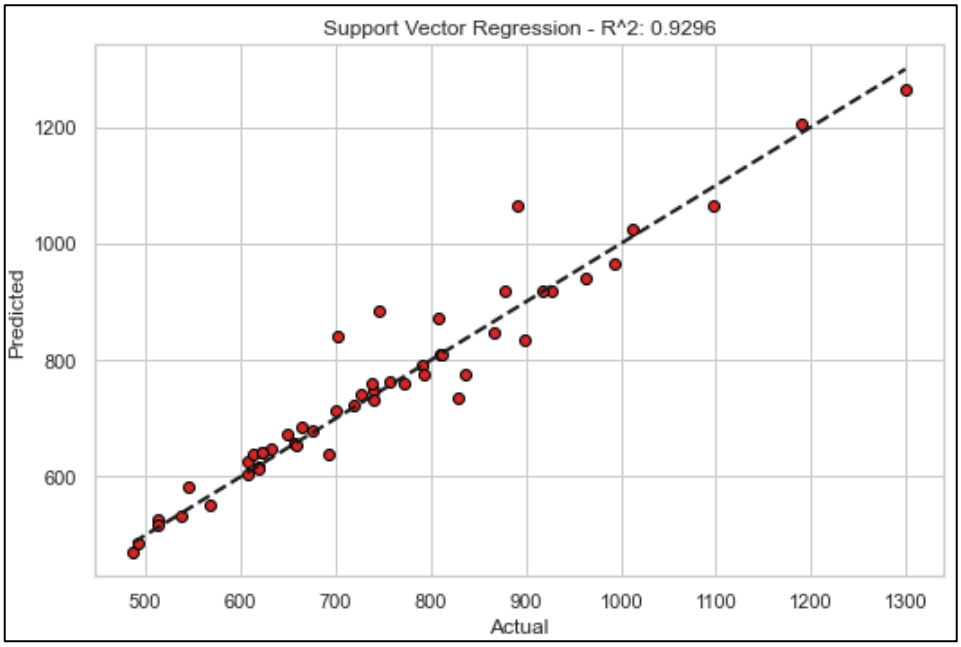
The numerical results for Hassi R'mel aquifer are presented in the table below (see table 14), where this table presents the results of the same ML models chosen in previous studies to predict TDS. The models are evaluated and compared based on the NSE and RMSE indices.

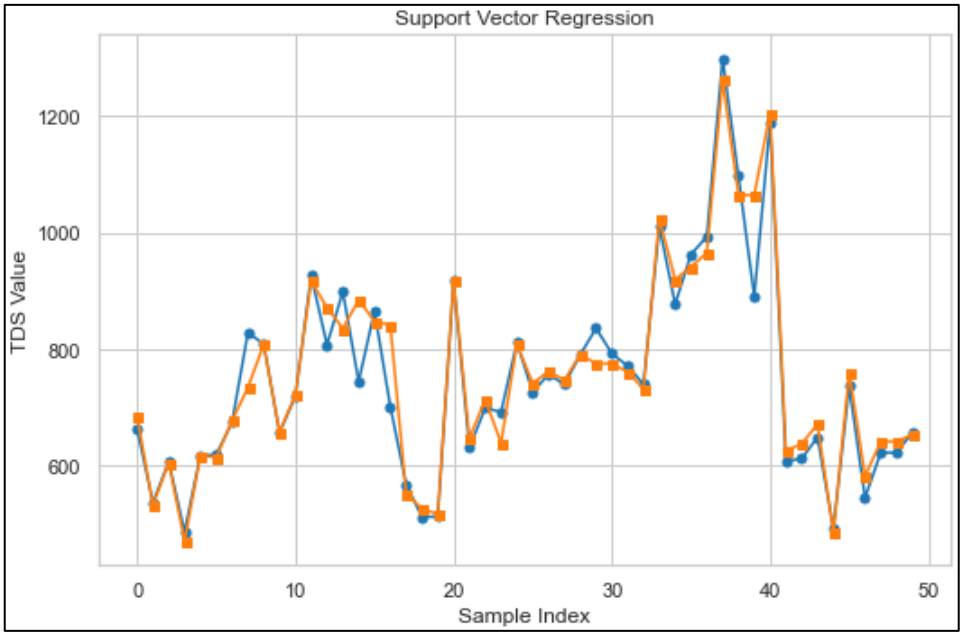
Table 14. Performance indices of the four machine learning models in the validation phase (Hassi R'mel plain)

Input Combination	Machine Learning Models							
	DT		**CatBoost**		**RF**		**SVR**	
	NS	**RMSE**	**NS**	**RMSE**	**NS**	**RMSE**	**NS**	**RMSE**
ML_{01}	0.710	0.0969	0.800	0.0796	0.794	0.0806	0.930	0.0460
ML_{02}	0.862	0.0652	0.915	0.0502	0.887	0.0579	0.949	0.0391
ML_{03}	0.820	0.0741	0.947	0.0405	0.913	0.0509	0.954	0.0375
ML_{04}	0.872	0.0628	0.949	0.0401	0.929	0.0461	0.953	0.0376
ML_{05}	0.861	0.0653	0.928	0.0463	0.939	0.0425	0.972	0.0300

The SVR model consistently outperforms other models, with the predicted TDS values ranging from 0.929 to 0.972 and the highest NS values (up to 0.972 in ML05). It also has the lowest RMSE values (0,3000 in ML05) and RMSE value between 0,4605 and 0,3000.
CatbBoost and RF also perform well, especially in ML03 and ML04 for CatbBoost model, which has NS value of 0.947 and 0.949 and RMSE value of 0.405 and 0.401 respectively.
while the RF model achieves 0.939 and 0.929 with NS values and an RMSEs of 0.425 and 0.461, respectively, from the combinations ML05 and ML04.

The DT model generally shows lower performance compared to other models, with NS values peaking at 0.872 in the combination ML04 and overall higher RMSE value of 0.0628.
Overall, SVR is the most accurate model with the combination ML05 as shown in the scatter plots (see Fig 19,20,21,22,23), which presented the observed and modeled TDS variations in the validation phase for the best combination in our models; CatBoost and RF models follow, with the DT model being less effective in predicting TDS parameter.


Support Vector Regression - R^2: 0.9296
Predicted
Actual
1200
1000
800
600
500
600
700
800
900
1000
1100
1200
1300


Support Vector Regression
TDS Value
Sample Index
1200
1000
800
600
0
10
20
30
40
50

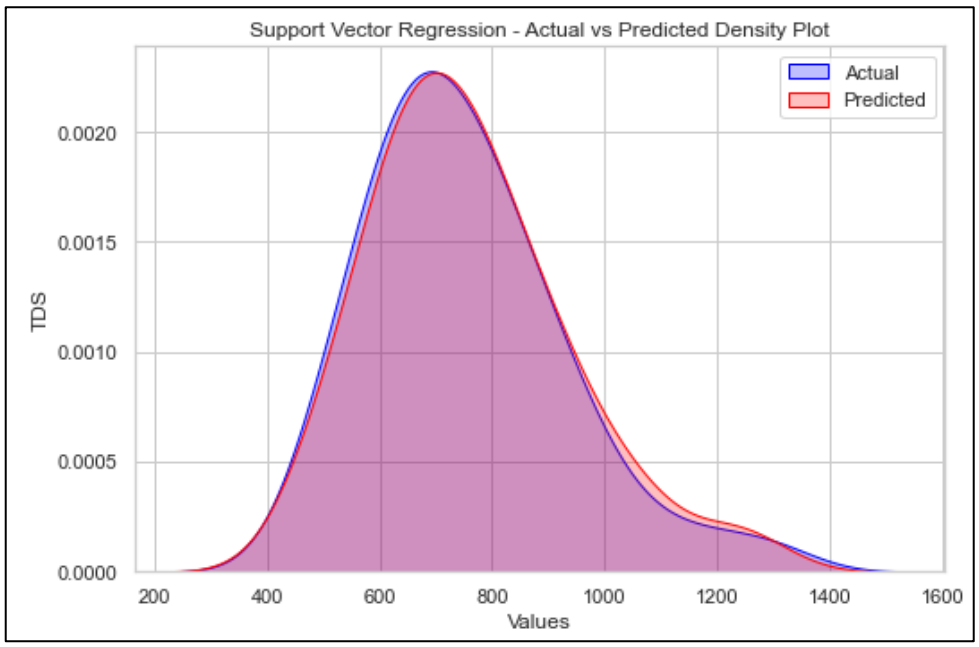
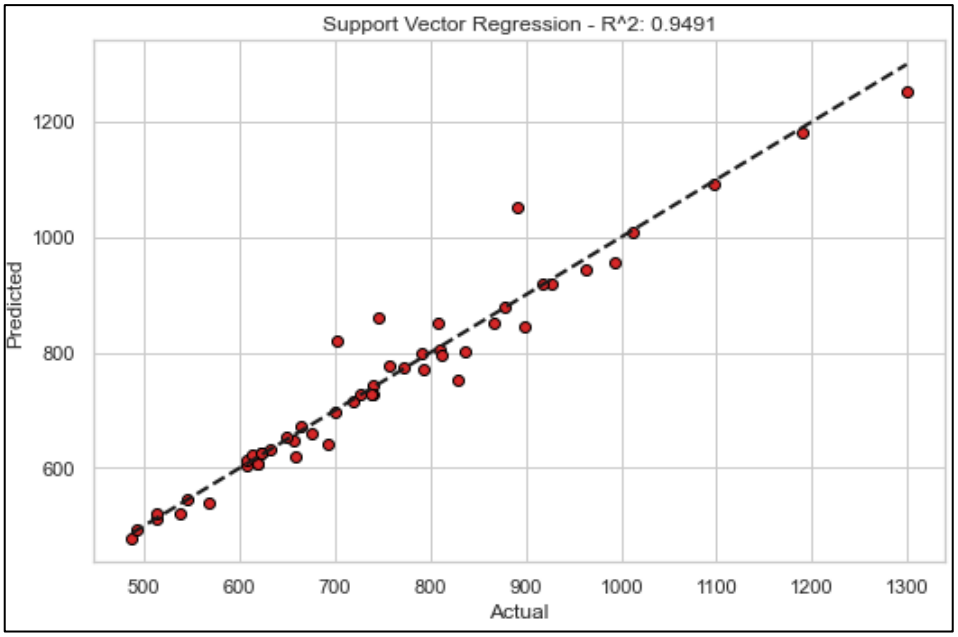
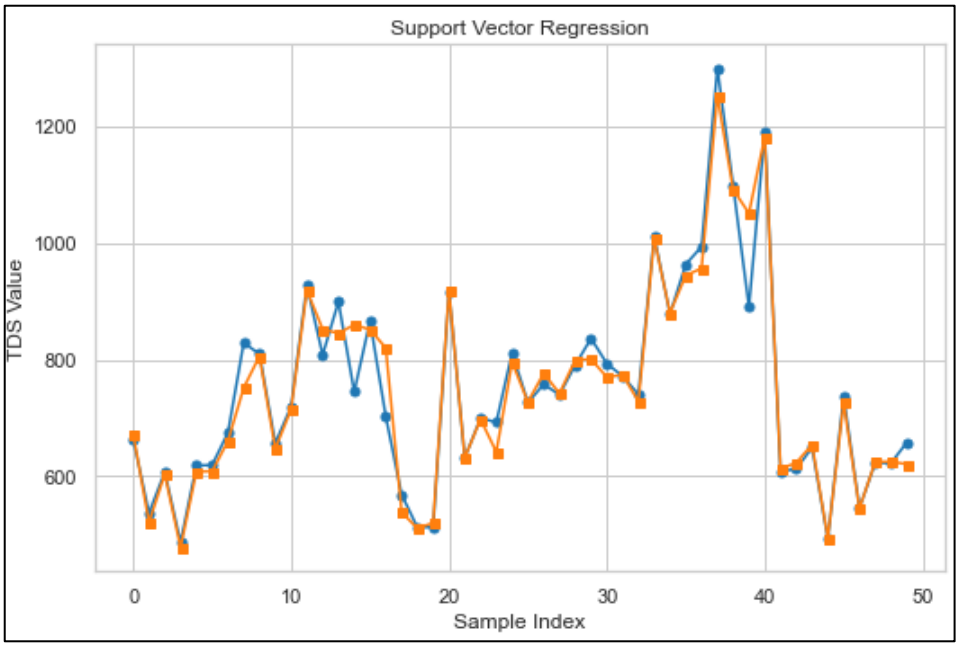


Fig 19. Observed and predicted TDS using ML01 combination: Hassi R'mel aquifer


Support Vector Regression - R^2: 0.9491
Predicted
Actual
1200
1000
800
600
500
600
700
800
900
1000
1100
1200
1300


Support Vector Regression
TDS Value
Sample Index
1200
1000
800
600
0
10
20
30
40
50

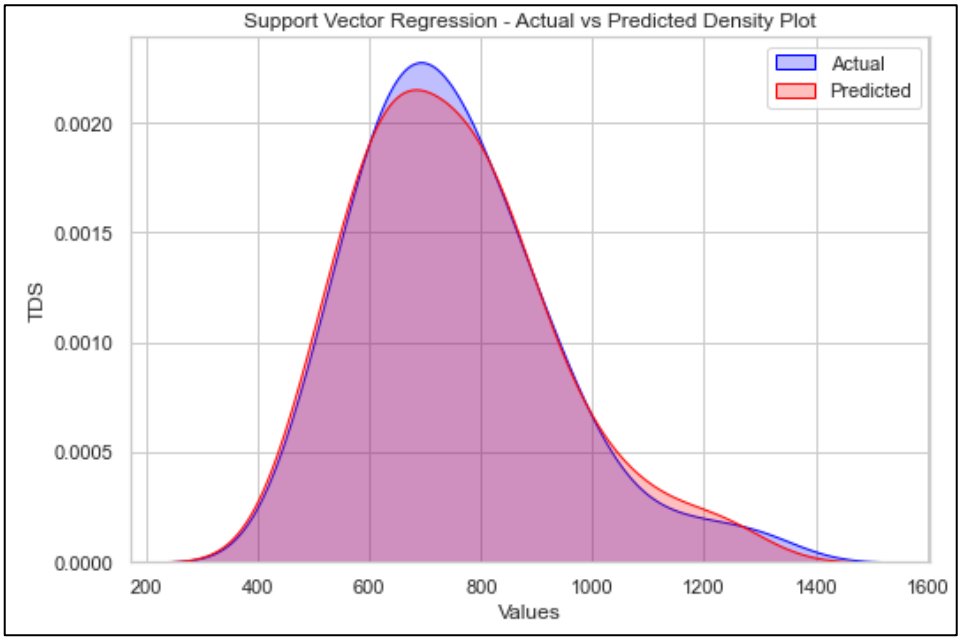
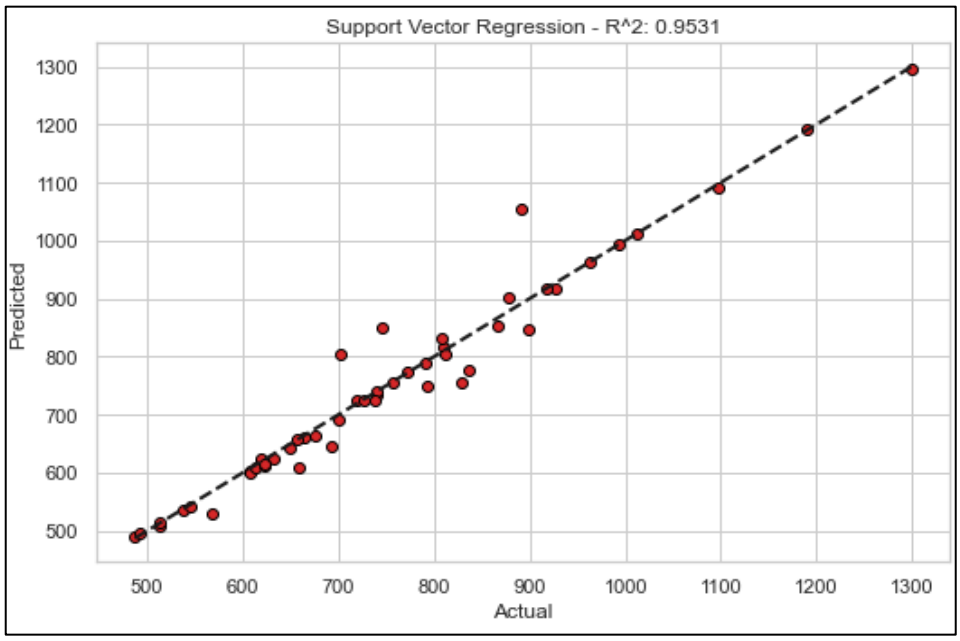
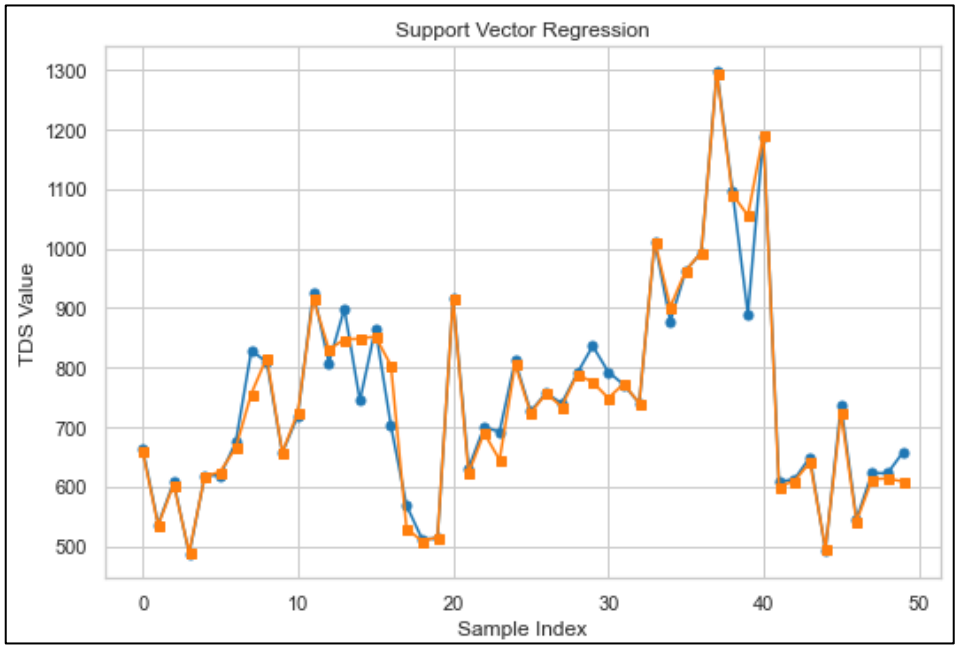


Fig 20. Observed and predicted TDS using ML02 combination: Hassi R'mel aquifer


Support Vector Regression - R^2: 0.9531
Predicted
Actual
1300
1200
1100
1000
900
800
700
600
500
500
600
700
800
900
1000
1100
1200
1300


Support Vector Regression
TDS Value
Sample Index
1300
1200
1100
1000
900
800
700
600
500
0
10
20
30
40
50

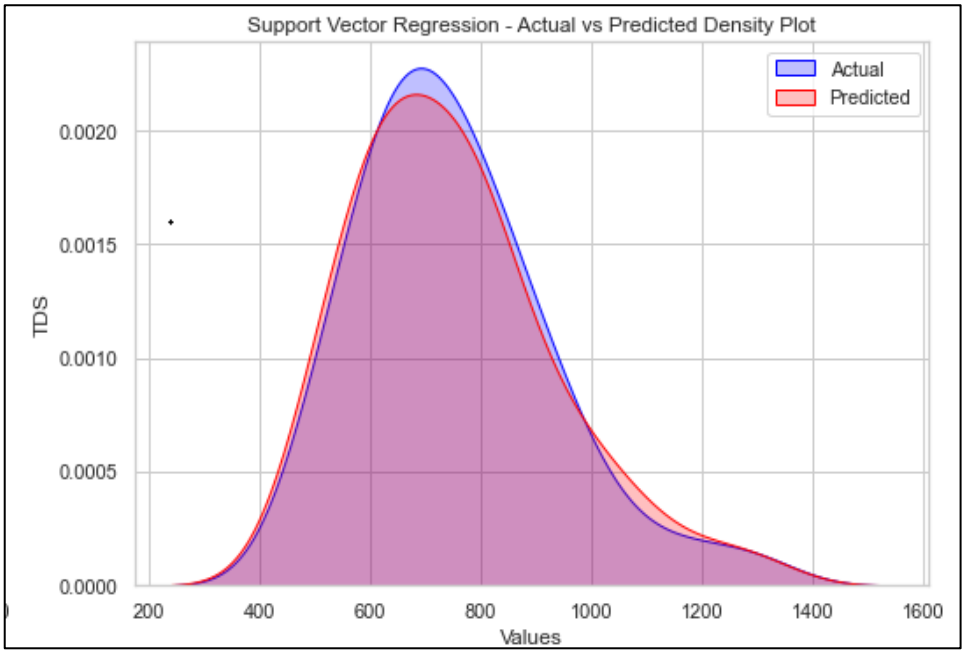
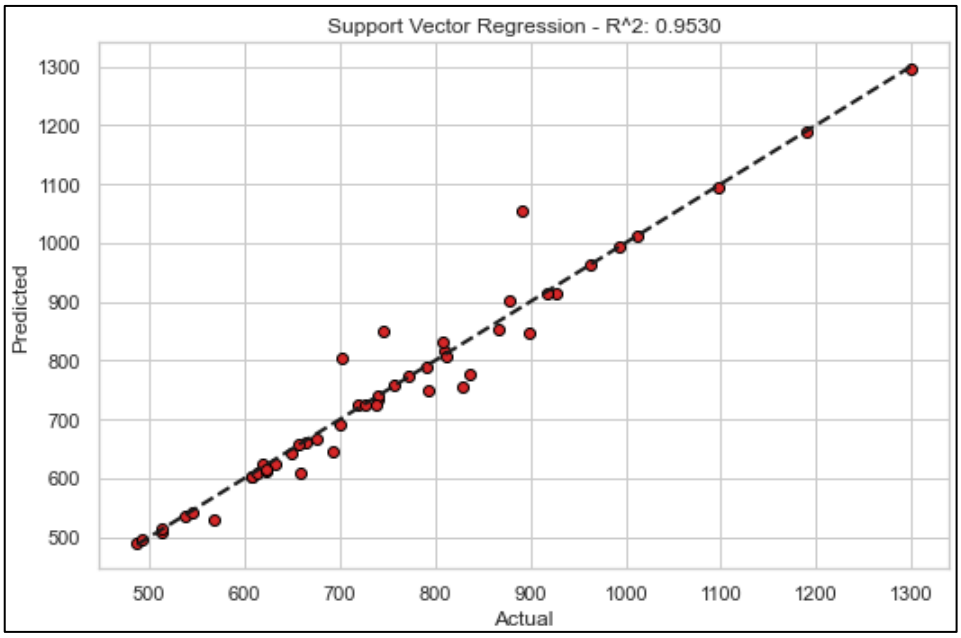
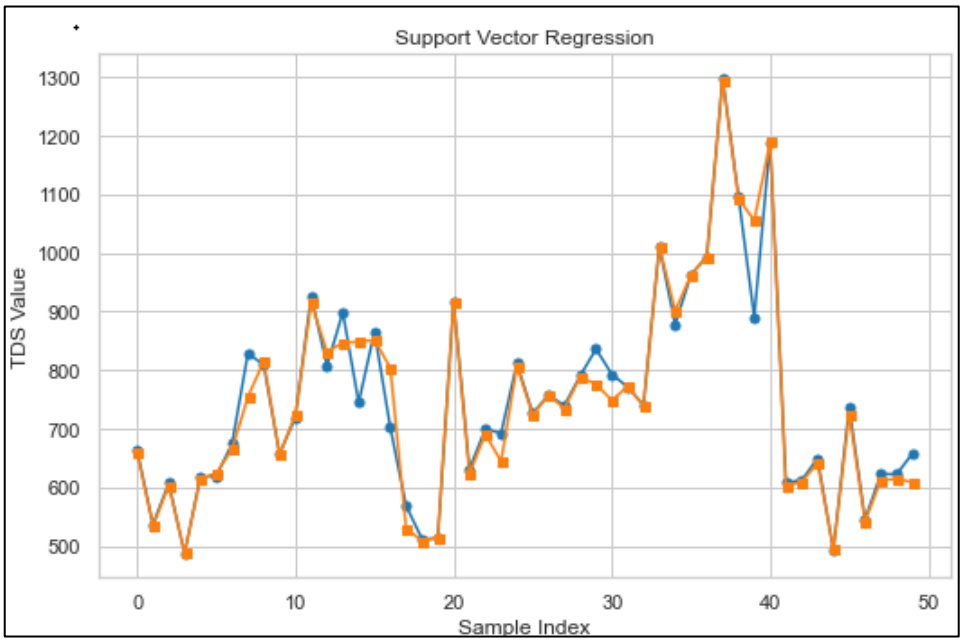


Fig 21. Observed and predicted TDS using ML03 combination: Hassi R'mel aquifer


Support Vector Regression - R^2: 0.9530
Predicted
Actual
1300
1200
1100
1000
900
800
700
600
500
500
600
700
800
900
1000
1100
1200
1300


Support Vector Regression
TDS Value
Sample Index
1300
1200
1100
1000
900
800
700
600
500
0
10
20
30
40
50

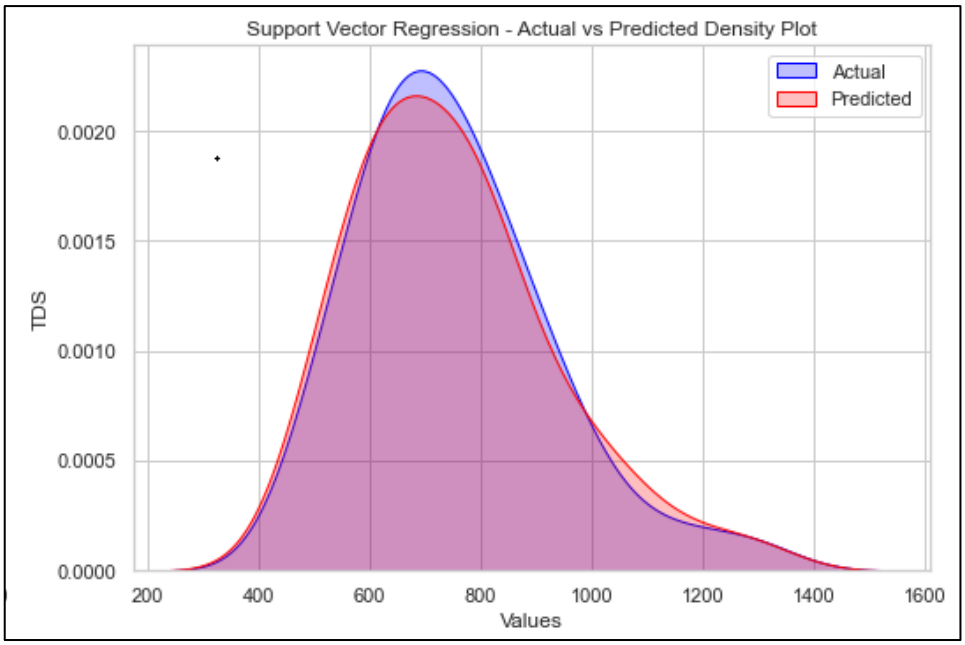
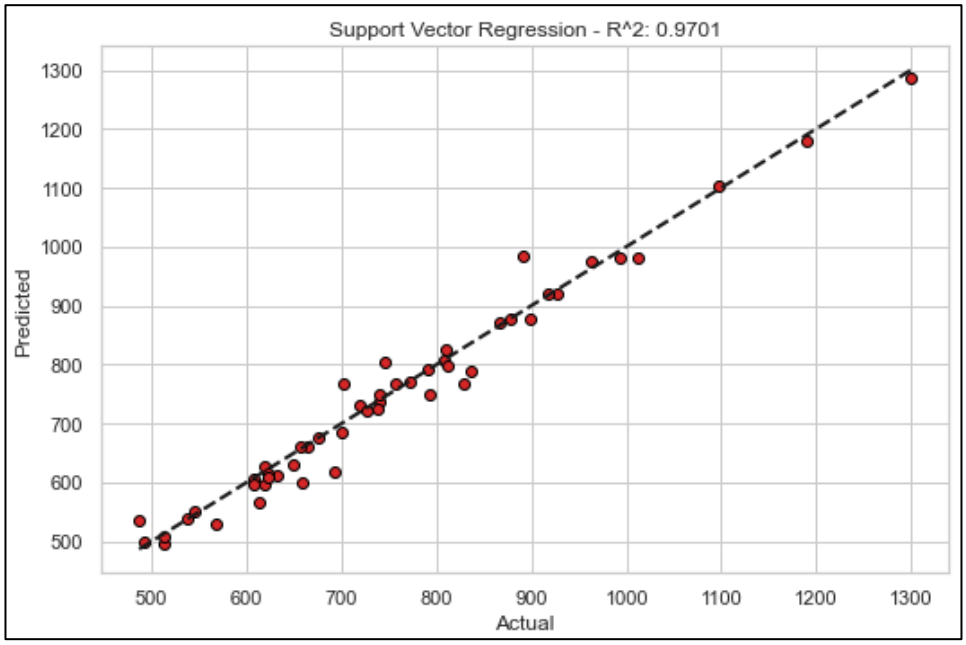
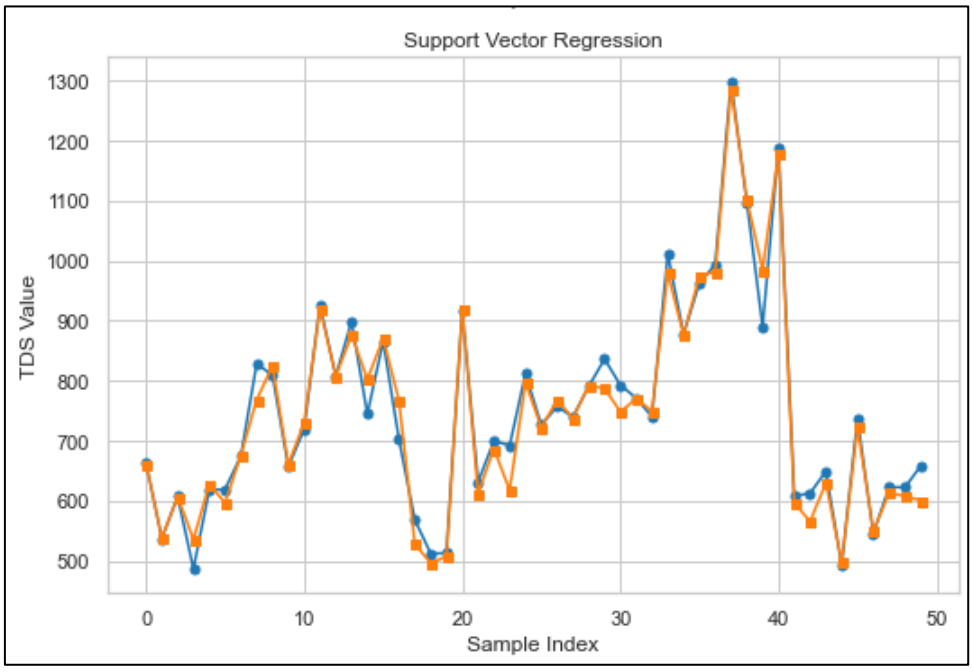


Fig 22. Observed and predicted TDS using ML04 combination: Hassi R'mel aquifer


Support Vector Regression - R^2: 0.9701
Predicted
Actual
1300
1200
1100
1000
900
800
700
600
500
500
600
700
800
900
1000
1100
1200
1300


Support Vector Regression
TDS Value
Sample Index
1300
1200
1100
1000
900
800
700
600
500
0
10
20
30
40
50

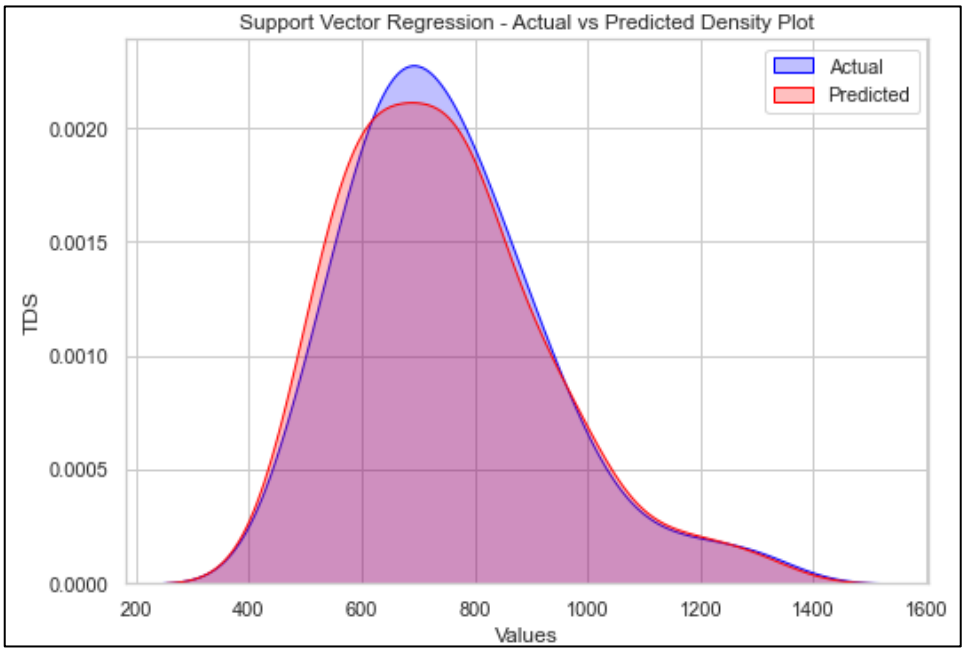
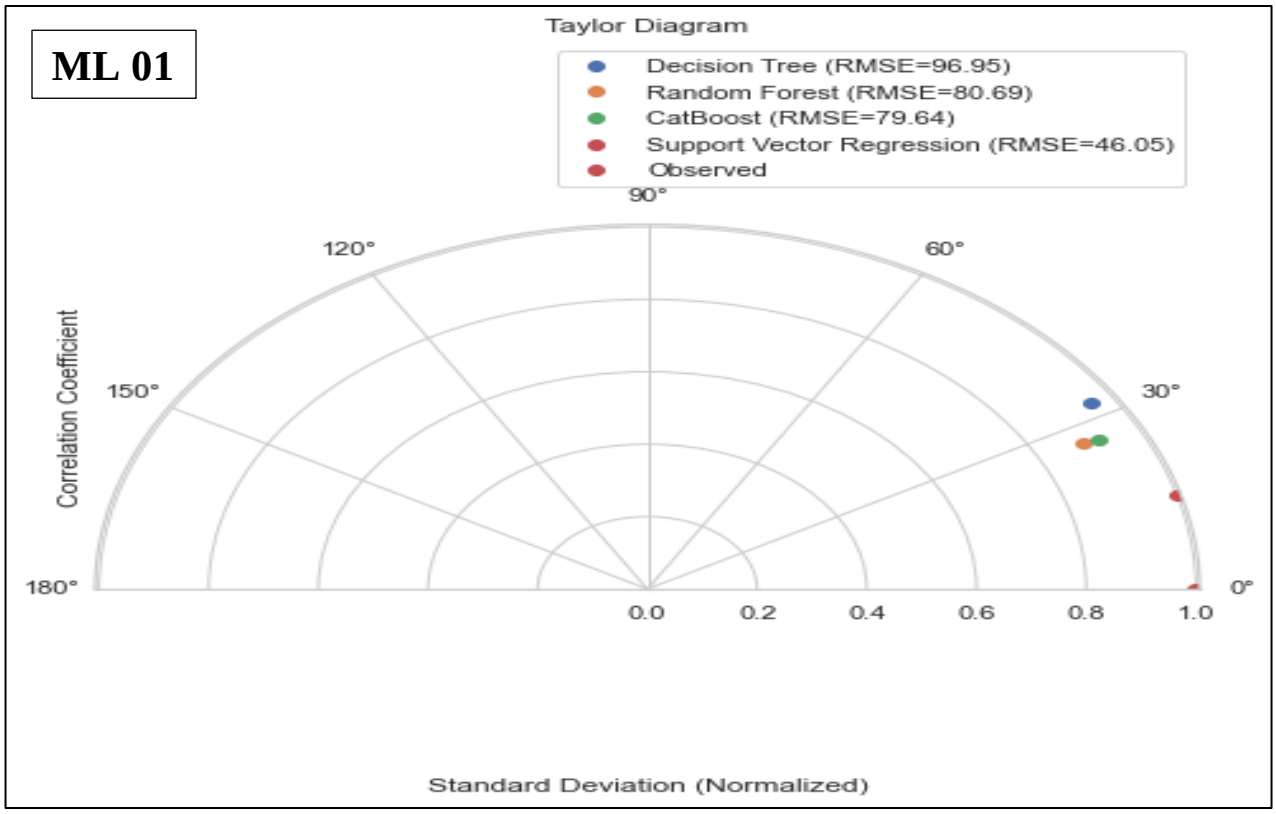
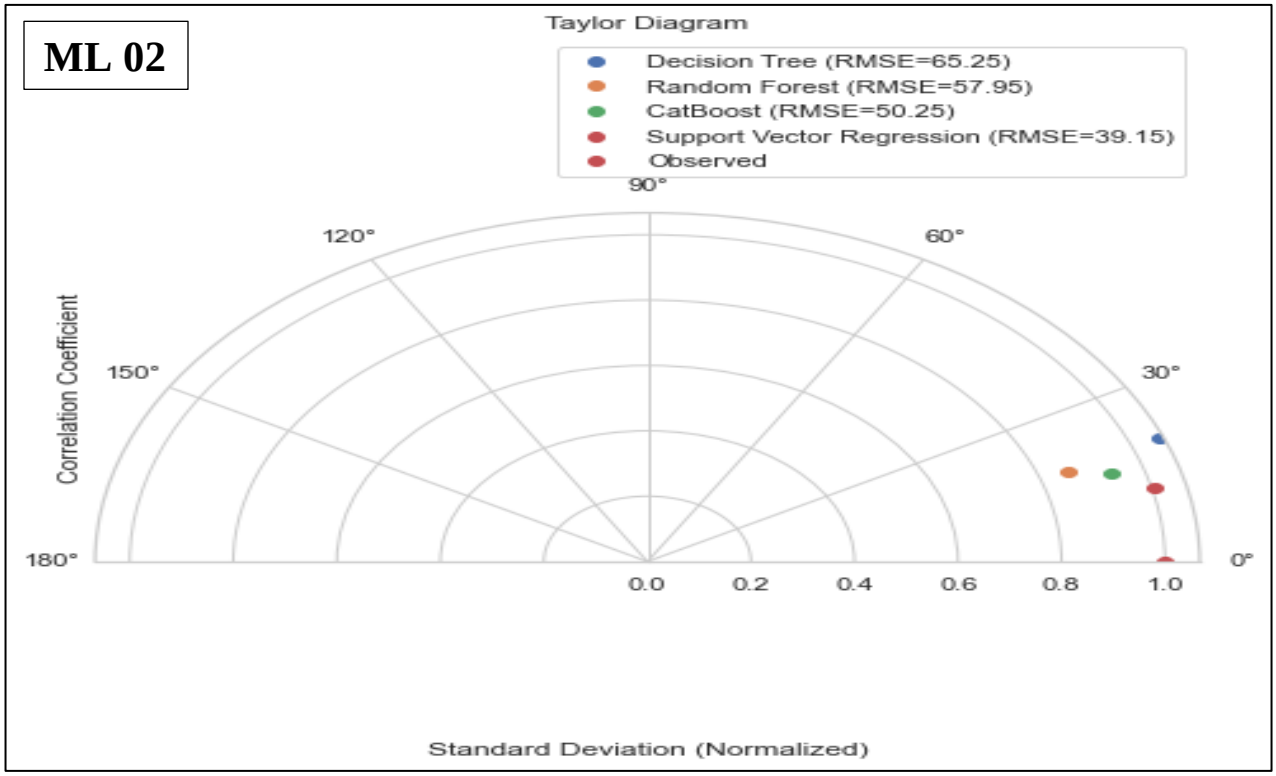
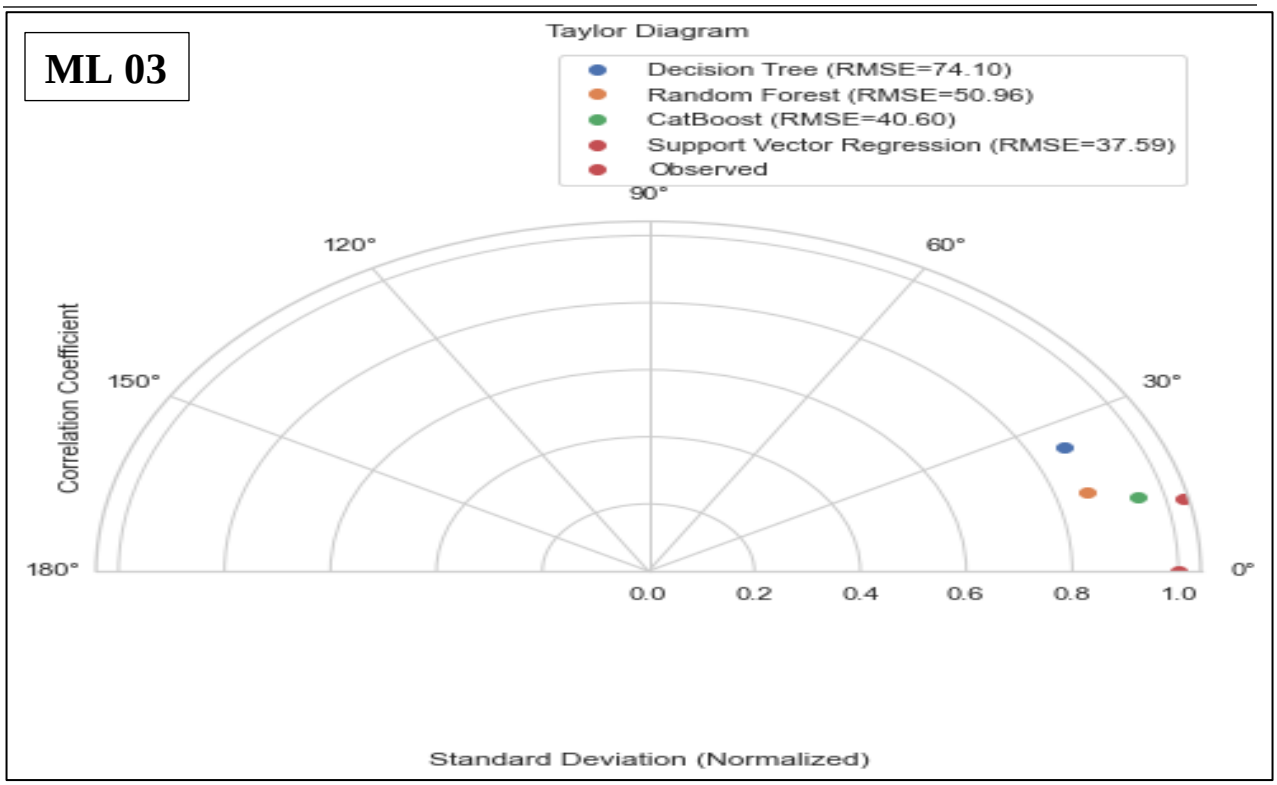


Fig 23. Observed and predicted TDS using ML05 combination:Hassi R'mel aquifer

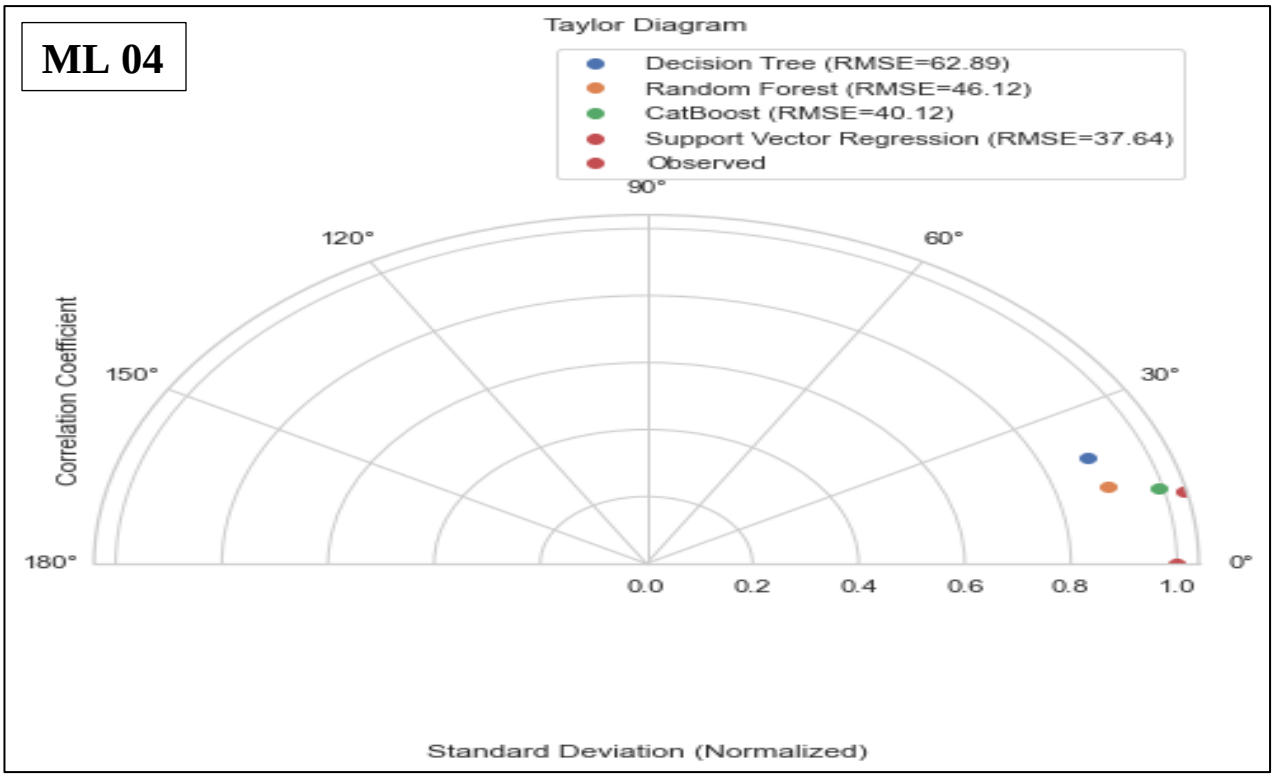

ML 01
Taylor Diagram
Decision Tree (RMSE=96.95)
Random Forest (RMSE=80.69)
CatBoost (RMSE=79.64)
Support Vector Regression (RMSE=46.05)
Observed
90°
120°
60°
150°
30°
180°
0°
0.0
0.2
0.4
0.6
0.8
1.0
Correlation Coefficient
Standard Deviation (Normalized)


ML 02
Taylor Diagram
Decision Tree (RMSE=65.25)
Random Forest (RMSE=57.95)
CatBoost (RMSE=50.25)
Support Vector Regression (RMSE=39.15)
Observed
90°
120°
60°
150°
30°
180°
0°
0.0
0.2
0.4
0.6
0.8
1.0
Correlation Coefficient
Standard Deviation (Normalized)

ML 03


Taylor Diagram
Decision Tree (RMSE=74.10)
Random Forest (RMSE=50.96)
CatBoost (RMSE=40.60)
Support Vector Regression (RMSE=37.59)
Observed
90°
120°
60°
150°
30°
180°
0°
0.0
0.2
0.4
0.6
0.8
1.0
Correlation Coefficient
Standard Deviation (Normalized)

ML 04


Taylor Diagram
Decision Tree (RMSE=62.89)
Random Forest (RMSE=46.12)
CatBoost (RMSE=40.12)
Support Vector Regression (RMSE=37.64)
Observed
90°
120°
60°
150°
30°
180°
0°
0.0
0.2
0.4
0.6
0.8
1.0
Correlation Coefficient
Standard Deviation (Normalized)

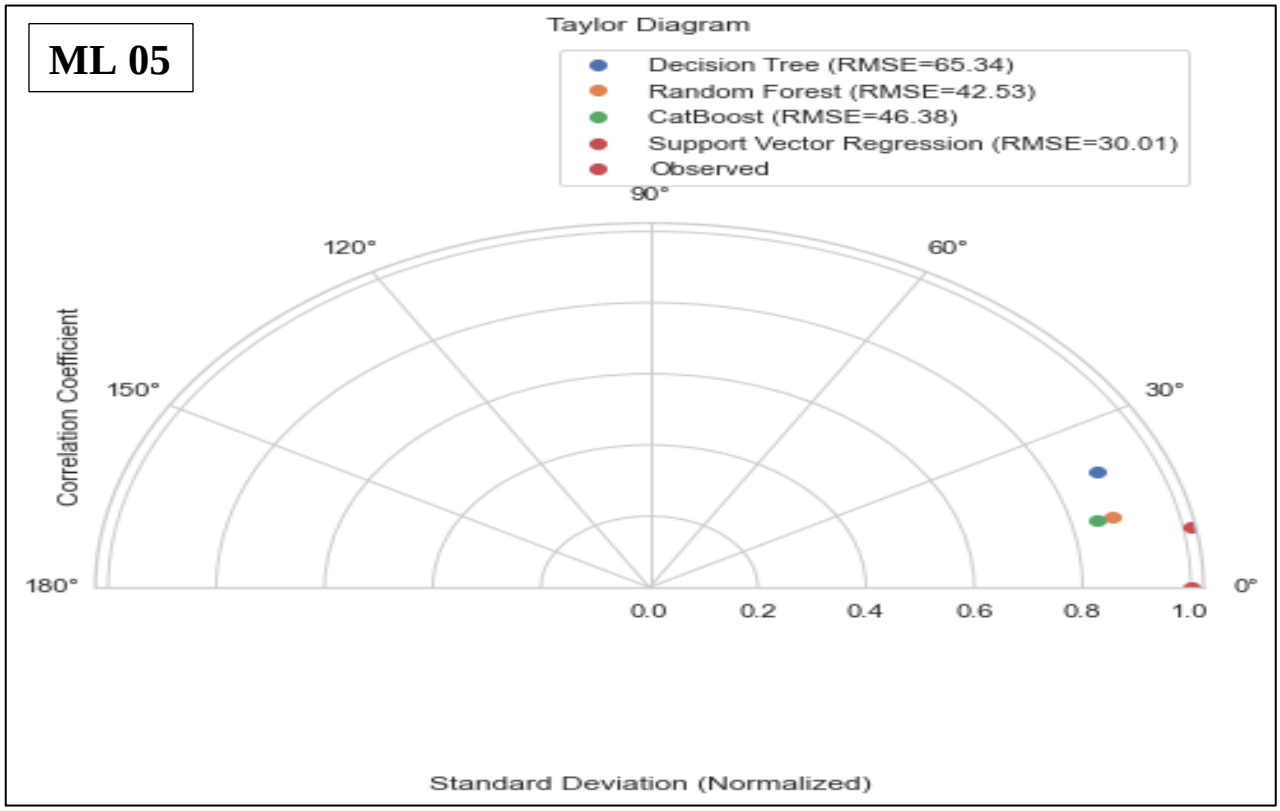


Fig 24. The Taylor diagram for the comparison between the best ML models in all combinations: Hassi R'mel aquifer

Taylor diagrams (see figure 24) prove too that the overall performance of SVR model is the highest performance in the all combinations compared the other models, followed by RF, catboost and DT models.

IV.5. Comparison of results between studied areas

To gain a better understanding of the predicted total dissolved solids and their relationship with electrical conductivity and chemical major elements using ML models (DT, RF, CatBoost, and SVR), we compared the performance and input combinations across the studied areas, using the same physicochemical analysis factors while considering the different geological structures of the three regions. In all studied areas using ML models (DT, RF, Catboost, SVR); it can be seen that during validation period the performance results of the total dissolved solids gave better results in the SVR model than the other models in all studied areas, the ML5 (EC; SO_4^{-2}, Cl^-, Na^+ and Ca^+) of the application in the Hassi R'mel aquifer gave the best performance values. In contrast, In the application of the Chellif and Mitidja aquifers, the ML4 (EC, Na^+, Cl^- and SO_4^{2-}) and ML 3 (EC, TH and Cl^-), respectively, they indicated better performance than the combination MLT for both. However, the leading performance comes from ML5 in Hassi R'mel compared to all ML combinations from the other areas.

To compare all the machine learning models, we selected the best combination for each model, where these input combinations were identified depending on the feature importance technique. This enabled us to determine the parameters most strongly associated with TDS. Therefore, we can explain that the different combinations observed in each study are due to the different nature and geological composition of the aquifers of each of the study's three regions. These geological differences influence the quality of groundwater. The impact of chemical parameters on aquifers provides the origin and history of groundwater by identifying the source of ionic and cationic distribution.

IV.6. Water quality assessment

Piper diagrams displays the chemical constituents of water from multiple boreholes. It depicts the combination of main cations and major anions in a triangular diagram of groundwater. The location of the combination of several components reveals the relative composition of groundwater and for helping to classify their type [156,157,158].

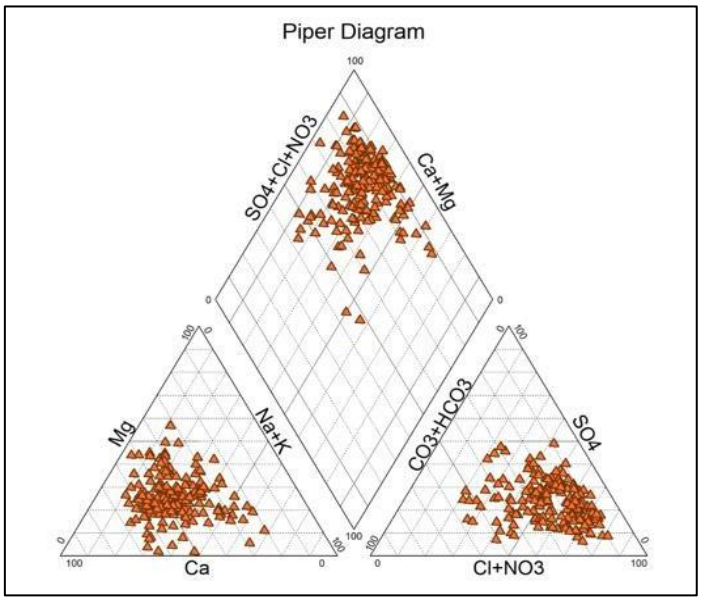


Fig 25. Piper diagram of groundwater analyses in the study region "the upper chellif"

The report on the Piper diagram of the Upper Chellif aquifer shows that the waters belong largely to the family of waters oscillating between sulfated or by the chlorinated. We note that some samples indicate that the waters are hyperchlorinated or hyper sulfated. This distribution shows the presence of mineralization; the latter is induced by sulfates or by the chlorinated.

Chlorides are present in the form of salts (NaCl, KCl, CaCl2 ...) constituting 0.05% of the lithosphere (majority in the oceans).

Sulfates and chlorides come from the leaching of land, soil erosion, dissolution of minerals including rock salt, sea spray, penetration of sea water into the soil.

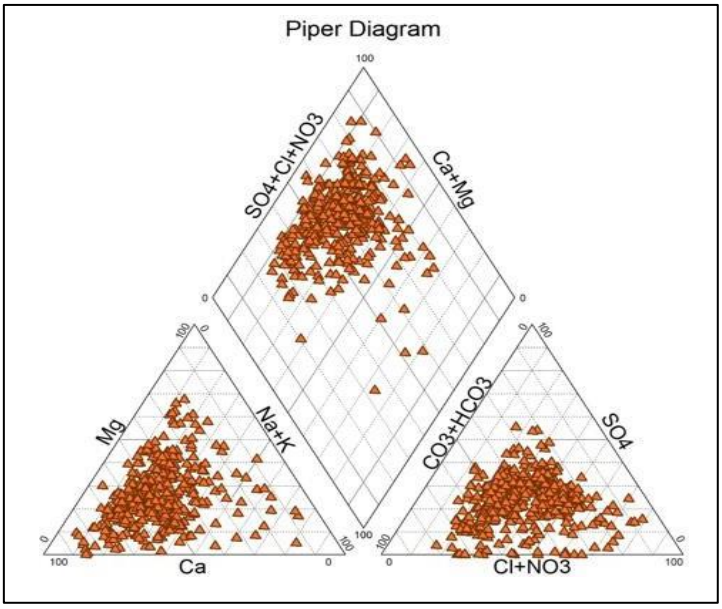


Fig 26. Piper diagram of groundwater analyses in the study region "Mitidja plain"

The Piper diagram indicates that groundwater in the research region, the "Mitidja Plain", can be classified into to two main families:

- The family of sulphated or chlorinated calcium waters,
- The family of bicarbonated calcium waters.

This distribution is linked to the geological formations encountered by the waters during their flow.

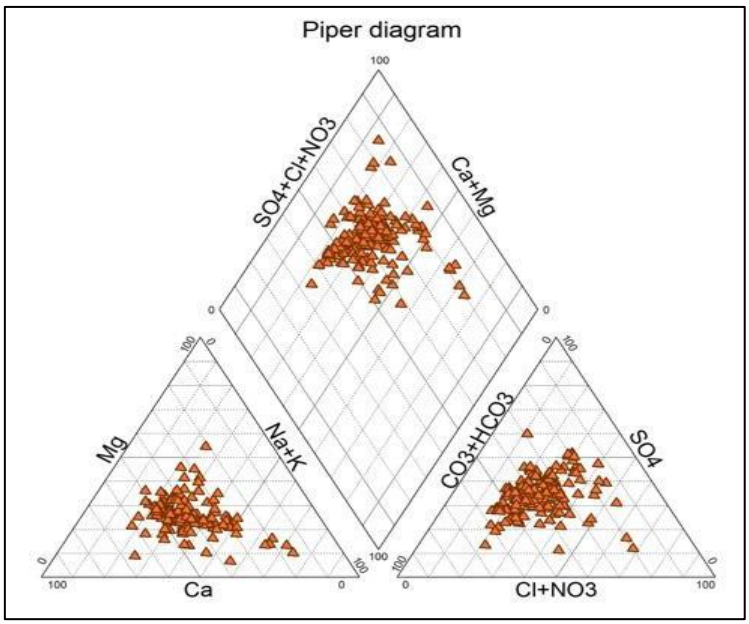


Fig 27. Piper diagram of groundwater analyses in the study region "Hassi R'mel "

The Piper diagram reveals that the groundwater in the study area, the "Hassi R'mel'" region, can be categorized into three primary families:

- the family of calcium-chlorinated sulfate waters,
- the family of sodium-chlorinated sulfate waters,
- the family of calcium-bicarbonate waters.

A mineralization of the waters is noted.

IV.7. BRGM analysis reports of TDS

On the other hand; Based on BRGM's TDS analysis report of TDS [159]; which split and categorized the principal components according to their importance via their concentration in the water into four fundamental partials groups; we obtained the following results:

-TDS 1 is the total concentrations (mg/1) of four main elements: HCO_3^-, NO_3^-, $SO_4^{-2}$and Cl^-.

-TDS2 is the total of the five major "secondary" elements: Ca^{2+}, Na^+, Mg^{2+}, K^+ and SiO_2. These elements are tasted less often than those used in the computation of TDS 1. They are also represented in mg/1.

-TDS3 is composed of F, Fe and Mn. the database expresses these three elements in mg/1,

-The TDS4 is the total of trace elements represented in pg/1 in the Observatory, including the most commonly measured: Cu, Pb, T.v, As, Cd and Al.

We also compared these results with the BRGM report, focusing on the best-performing combinations across the entire study area (ML N°04, ML N°03 and ML N°05) identified during this research. These combinations included the following elements (Na^+, SO_4^{2+}, Cl^-, Ca^{2+}, and Mg^{2+}). These elements are known to significantly contribute to TDS, with Ca^{2+}, Na^+, and Mg^{2+} classified under TDS partial 2, while $SO_4^{2-}$and Cl^- are categorized TDS partial 1. This alignment highlights the strong correlation between our findings and the BRGM report. Therefore, it can be concluded that this comparison further validates the results obtained in this study.

Conclusion

The results demonstrate that machine learning models, particularly Support Vector Regression, provide highly accurate forecasts of TDS across all study areas. Variations in model performance reflect differences in hydrogeological settings and chemical compositions. Overall, the findings confirm the value of data-driven modeling for groundwater quality assessment and support its integration into sustainable water resource management strategies.

CONCLUSION

CONCLUSION

Accurate forecasting of water quality indicators is fundamental and critical for sustainable groundwater management, as it enables and facilitates the monitoring of pollution levels, population demands and human activities. The current study explored the applicability of machine learning (ML) models, including CatBoost, Random Forest, Decision Tree and Support Victor Regression, and evaluated their effectiveness as tools for forecasting the groundwater salinity parameter (TDS).

The research was conducted on three different cases in Algeria, and the performance of the models was compared based the results, with emphasis on examining the sensitivity of these models to the output variable using various hydro chemical and physicochemical parameters as input variables.

The first case was applied on the Upper Chellif aquifer in northwestern Algeria. The four machine learning models chosen for this work were trained and tested using 191 groundwater samples.

To improve model performance, a sensitivity analysis was conducted to identify the most influential physicochemical parameters (EC, Na^+, Cl^-, SO_4^{2-}, Mg^+) on TDS. Based on these findings, five different sets of input combinations were created to construct and evaluate the performance of each model.

The second case was applied to the aquifer of the Metidja Plain, in the central region of northern Algeria. In this case, the same steps as in the first application were followed, with 363 samples of physicochemical parameters used. However, the application of the sensitivity analysis method introduced different input parameters that significantly influence TDS (EC, TH, Cl^-, Ca^{2+},Na^+); Based on these parameters, five input combinations were developed to assess the performance of the ML models.

In the last case, the study was applied to the Hassi R'mel aquifer, located in the central part of the northern Saharan Platform. A total of 168 groundwater samples were employed using the same models and steps as in the other cases. The sensitivity analyses identified five physicochemical parameters (EC, SO_4^{2-} , Cl^-, Na^+,Ca^+) as the most influential on TDS. Based on these parameters, five variable combinations were constructed as inputs to evaluate the models.

The results indicated that the use of ML models for forecasting TDS parameters produced very good and acceptable outcomes. When comparing the results across the studied aquifers, it was observed that the SVR model outperformed all other models in forecasting TDS values. However, it achieved these strong results with different input combinations in each case study, likely due to differences in the hydrogeological structure of the aquifers in the various cases. in other hand, although no statistically significant differences were observed between the performance of the various ML models, the SVR model was ultimately selected as the optimal model due to its superior validation performance. Thus, these findings highlight that prediction accuracy is influenced not only by the structure of the model but also by the selection of input variables.

In conclusion, the results demonstrate that forecasting TDS in groundwater using SVR, RF, CatBoost and DT machine learning models are highly robust, effective and based on groundwater quality parameters. These models outperform traditional approaches and are recommended for TDS forecasting. Their application has the potential to significantly enhance groundwater quality management while saving time and reducing costs.

In the future, we aim to explore different ML models and techniques to forecasting water quality parameters to enhance groundwater quality and prevent pollution in Algeria’s most vital natural source of fresh water.

REFERENCES

1. *Richey, A. S., Thomas, B. F., Lo, M. H., Reager, J. T., Famiglietti, J. S., Voss, K., ... & Rodell, M. (2015). Quantifying renewable groundwater stress with GRACE. Water resources research, 51(7), 5217-5238.https://doi:10.1002/2015WR017349*
2. *Siebert, S., Henrich, V., Frenken, K., & Burke, J. (2013). Update of the digital global map of irrigation areas to version 5. Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany and Food and Agriculture Organization of the United Nations, Rome, Italy, 10(2.1), 2660-6728.https://doi:10.13140/2.1.2660.6728*
3. *Ercin, A. E., & Hoekstra, A. Y. (2014). Water footprint scenarios for 2050: A global analysis. Environment international, 64, 71-82..https://doi.org/10.1016/j.envint.2013.11.019.*
4. *Chen, W., Pradhan, B., Li, S., Shahabi, H., Rizeei, H. M., Hou, E., & Wang, S. (2019). Novel hybrid integration approach of bagging-based fisher's linear discriminant function for groundwater potential analysis. Natural Resources Research, 28(4), 1239-1258..https://doi.org/10.1007/s11053-019-09465-w*
5. *Banadkooki, F. B., Ehteram, M., Ahmed, A. N., Fai, C. M., Afan, H. A., Ridwam, W. M., ... & El-Shafie, A. Precipitation forecasting using multilayer neural network and support vector machine optimization based on flow regime algorithm taking into account uncertainties of soft computing models. Sustainability, 2019, vol. 11, no 23, p. 6681. https://doi.org/10.3390/su11236681*
6. *Banadkooki, F. B., Ehteram, M., Panahi, F., Sammen, S. S., Othman, F. B., & El-Shafie, A. (2020). Estimation of total dissolved solids (TDS) using new hybrid machine learning models. Journal of Hydrology, 587, 124989.https://doi.10.1016/j.jhydrol.2020.124989.*
7. *Ekemen Keskin, T., Özler, E., Şander, E., Düğenci, M., & Ahmed, M. Y. (2020). Prediction of electrical conductivity using ANN and MLR: a case study from Turkey. Acta Geophysica, 68(3), 811-820.https://doi:10.1007/s11600-020-00424-1.*
8. *Drouiche, N., Khacheba, R., & Soni, R. Water policy in Algeria. 2020, Water policies in MENA countries, 19-46. https://doi.org/10.1007/978-3-030-29274-4_2*
9. *Kherbache, N. (2020). Water policy in Algeria: limits of supply model and perspectives of water demand management (WDM). Desalination and Water Treatment, 180, 141-155.. doi: 10.5004/dwt.2020.25009*
10. *Neama, M. A., Attia, M., Negm, A., & Nasr, M. (2020). Overview of water resources, quality, and management in Algeria. Water Resources in Algeria-Part I: Assessment of Surface and Groundwater Resources, 13-25. https://doi.org/10.1007/6*
11. *Banadkooki, F. B., Ehteram, M., Ahmed, A. N., Teo, F. Y., Fai, C. M., Afan, H. A., ... & El-Shafie, A. (2020). Enhancement of groundwater-level prediction using an integrated machine learning model optimized by whale algorithm. Natural resources research, 29(5), 3233-3252..https://doi.org/10.1007/s11053- 020-09634-2.*
12. *Zaman ZadGhavidel, S., &Montaseri, M. (2014). Application of different data-driven methods for the prediction of total dissolved solids in the Zarinehroud basin. Stochastic environmental research and risk assessment, 28(8), 2101-2118.https://doi.org/10.1007/s00477-014-0899-y*
13. *World Health Organization. (2004). Guidelines for drinking-water quality (Vol. 1). World health organization.*
14. *Tung, T. M., & Yaseen, Z. M. (2020). A survey on river water quality modelling using artificial intelligence models: 2000–2020. Journal of Hydrology, 585, 124670.https://doi.org/10.1016/j.jhydrol.2020.124670*
15. *Jamei, M., Ahmadianfar, I., Chu, X., &Yaseen, Z. M. (2020). Prediction of surface water total dissolved solids using hybridized wavelet-multigene genetic programming: New approach. Journal of Hydrology, 2020, vol. 589, p. 125335. https://doi.org/10.1016/j.jhydrol.2020.125335*
16. *Kisi, O., Akbari, N., Sanatipour, M., Hashemi, A., Teimourzadeh, K., & Shiri, J. (2013). Modeling of dissolved oxygen in river water using artificial intelligence techniques. Journal of Environmental Informatics, 22(2), 92-101. https://doi.org/10.3808/jei.201300248*
17. *Sorensen, D. L. (1977). Suspended and dissolved solids effects on freshwater biota: a review.*
18. *Hem, J. D. (1985). Study and interpretation of the chemical characteristics of natural water (Vol. 2254). Department of the Interior, US Geological Survey.*
19. *Liu, S., Tai, H., Ding, Q., Li, D., Xu, L., & Wei, Y. (2013). A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. Mathematical and Computer Modelling, 58(3-4), 458-465. https://doi.org/10.1016/j.mcm.2011.11.021*
20. *Melesse, A. M., Khosravi, K., Tiefenbacher, J. P., Heddam, S., Kim, S., Mosavi, A., & Pham, B. T. (2020). River water salinity prediction using hybrid machine learning models. Water, 12(10), 2951.https://doi.org/10.3390/w12102951*
21. *Chowdury, M. S. U., Emran, T. B., Ghosh, S., Pathak, A., Alam, M. M., Absar, N., ... & Hossain, M. S. (2019). IoT based real-time river water quality monitoring system. Procedia computer science, 155, 161-168.https://doi.org/10.1016/j.procs.2019.08.025*
22. *Krupková, L., Havránková, K., Krejza, J., Sedlák, P., & Marek, M. V. (2019). Impact of water scarcity on spruce and beech forests. Journal of Forestry Research, 30(3), 899-909.https://doi.org/10.1007/s11676-018-0642-5*
23. *Krupková, L., Havránková, K., Krejza, J., Sedlák, P., & Marek, M. V. (2019). Impact of water scarcity on spruce and beech forests. Journal of Forestry Research, 30(3), 899-909.https://doi.org/10.1016/j.ecolind.2018.03.079*

24. *Zhu, J. J., Yu, L. Z., Xu, T. L., Wei, X., & Yang, K. (2019). Comparison of water quality in two catchments with different forest types in the headwater region of the Hun River, Northeast China. Journal of Forestry Research, 30(2), 565-576.https://doi.org/10.1007/s11676-018-0688-4*
25. *El Bilali, A., Taleb, A., &Brouziyne, Y.(2020). Groundwater quality forecasting using machine learning algorithms for irrigation purposes. Agricultural Water Management, vol. 245, p. 106625. t, https://doi.org/10.1016/j.agwat.2020.106625*
26. *Kisi, O., Dailr, A. H., Cimen, M., & Shiri, J. (2010). Suspended sediment modeling using genetic programming and soft computing techniques. Journal of Hydrology, vol. 450, p. 48-58.https://doi.org/10.1016/j.jhydrol.2012.05.031*
27. *Yaseen, Z. M., Jaafar, O., Deo, R. C., Kisi, O., Adamowski, J., Quilty, J., & El-Shafie, A. (2020). Stream-flow forecasting using extreme learning machines: a case study in a semi-arid region in Iraq. Journal of Hydrology, 2016, vol. 542, p. 603-614.https://doi.org/10.1016/j.jhydrol.2016.09.035*
28. *Lu, H., & Ma, X.(2020). Hybrid decision tree-based machine learning models for short-term water quality prediction. Chemosphere, vol. 249, p. 126169.https://doi.org/10.1016/j.chemosphere.2020.126169*
29. *Meyers, G., Kapelan, Z., &Keedwell, E. (2017). Short-term forecasting of turbidity in trunk main networks. Water research, vol. 124, p. 67-76.https://doi.org/10.1016/j.watres.2017.07.035*
30. *Di, Z., Chang, M., &Guo, P. (2019).Water quality evaluation of the Yangtze River in China using machine learning techniques and data monitoring on different time scales. Water, vol. 11, no 2, p.339.https://doi.org/10.3390/w11020339*
31. *Rajaee, T., Ebrahimi, H., &Nourani, V. (2019). A review of the artificial intelligence methods in groundwater level modeling. Journal of hydrology, vol. 572, p. 336-351.https://doi.org/10.1016/j.jhydrol.2018.12.037*
32. *PH, G. (1996). Water resources. Encyclopedia of climate, weather, 817-823.*
33. *Remini, B. (2005). La problématique de l'eau en Algérie. Office des publications universitaires.*
34. *Rahmoune, C., & Zaimeche, S. Contribution à l'étude de l'action d'agents polluants sur des végétaux bioindicateurs.*
35. *Desjardins, R.(1997). Le traitement d'eaux, 2ème édition de l'Ecole Polytechnique de Montréal. P242 P, 6, P241.*
36. *Boeglin J.C. (2009). Propriétés des eaux naturelles, Technique de l'ingénieur, traité environnement, G1, 110p.*
37. *Degrémont, G. (1972). Mémento technique de l'eau. Ed. Techniques Ingénieur.*
38. *Molinie, L. (2009). Dispositifs rustiques d'alimentation et de traitement de l'eau potable pour des services de petites tailles en régions défavorisées. Synthèse Technique. Agro Paris Tech–Engref, 28.*
39. *Myrand D., (2008). Guide technique : captage d'eau souterraine pour des résidences isolées, Québec, P04.*
40. *Kettab, A., Mitiche, R., & Bennaçar, N. (2008). De l'eau pour un développement durable: enjeux et stratégies. Revue des sciences de l'eau, 21(2), 247-256. https://doi.org/10.7202/018469ar*
41. *Mebarki, A. (2010). La région du Maghreb face à la rareté de l'eau. L'exemple du défi algérien: mobilisation et gestion durable des ressources. In ICID+ 18 2nd International Conference: Climate, Sustainability and Development in semi-arid regions August (pp. 16-20).*
42. *Bounab S (2017). Ressources En Eau Et Développement Durable Cas De La Region Annaba-El Tarf (Nord- Est Algerien), Doctoral thesis, University of Annaba.*
43. *Benblidia, M. (2011). L'efficience d'utilisation de l'eau et approche économique. Plan Bleu, Centre d'Activités Régionales PNUE/PAM, Etude nationale, Algérie, 2011, 9-12.*
44. *Touahria, K. (2013) Evaluation De La Qualite Des Eaux De Forages Par Comparaison De Leurs Caracteristiques Physico-Chimiques (Region De Tebessa), Master's thesis, University of Souk Ahras.*
45. *Jean bontoux.M,(1993). Introduction à l'étude des eaux douces: Eaux naturelles, Eaux usées, Eaux de boisson; Qualité et santé, Editions CEBEDOC, Liège, p 79.*
46. *Cosandey, C., BIGOT, S., DACHARRY, M., GILLE, E., Laganier, R., & Salvador, P. G. (2003). Les eaux courantes: géographie et environnement. Paris: Berlin.*
47. *UNICEF., (1999). Manuel sur l'eau. N°2 , p. 42-43.*
48. *Fiambsch hambsch B, (1998). Chang from chlorine résidual distribution to no chlorine residual distribution in groundwater system. Water suppy. Vol 6 N°3/4. Germany, p.145-152.*
49. *ROUABHIA. A,(2006). Vulnerability and risk of groundwater pollution of the Miocene sand aquifer of the El Ma El Abiod plain (North-East Algeria), Doctoral thesis in science, University of Badji Mokhtar, Annaba, p 4, 5, 9, 14, 20, 43, 44, 45, 51, 54 et 55 – 73.*
50. *COLLIN. J.J,(2006). Les eaux souterraines: Connaissance et gestion, Editions BRGM et HERMANN, Paris, p3 – 5.*
51. *KETTAB.A, Traitement des eaux: Les eaux potables, Edition Office des Publications Universitaires, Alger, p 20 et 21.*
52. *Keven L, (2012) : Les eaux souterraines : captage, exploitation et gestion. Mémoire de magister, université de Kinshasa.*
53. *Degremont G., (2005). Mémento technique de l'eau, Tome 1, 10éme édition, Edit. Tec et doc, PP: 3- 38.*
54. *Cardot C., (1999). Les traitements de l'eau : procédés physico-chimiques et biologique : Cours et problèmes résolus, Ellipes Edition Marketing, paris, cedex 15, PP : 25-28.*

55. *Arjen V.D.W., (2010). Connaissances des méthodes de captage des eaux souterraines :Souterraines aux forages manuels, Un manuel d'instruction pour les équipes de forage manuel sur l'hydrogéologie appliquée, l'équipement et le développement des forages, Fondation PRACTICA, Oosteind, P10.*
56. *OSS (Observatoire du Sahara et du Sahel) (2003): Système Aquifère du Sahara Septentrional. Volume 4: Modèle Mathématique. Projet SASS; Rapport interne. Annexes. p 229*
57. *Rajendran A. and Mansiya C. (2015). Physico-chemical analysis of ground water samples of coastal areas of south Chennai in the post-Tsunami scenario. Ecotoxicology and environmental safety.*
58. *Rodier .J. (2005). L'analyse de l'eau eaux naturelles eaux résiduaires eaux de mer .9ème édition du nod Paris p 66.*
59. *HCEFLCD (2006) : Haut Commissariat Aux Eaux et Forêt et la Lutte Contre la Désertification. Etude sur la pisciculture au barrage Almassira ; CR dar CHAFAAI ; Cercle d'ELBROUGE ; Province de Settat ; 201p.*
60. *Rodier J., Legube B., Merlet N. (2009). L'analyse de l'eau, 9ème édition, Ed. Dunod, 1600 p.*
61. *REJSEK F. (2002) : Analyse de l'eau : Aspects et règlementaire et technique .Ed CRDP d'Aquitaine. France, 358 p.*
62. *Mustapha Besbes, (2010) : Hydrogéologie de l'ingénieure p23.24.26.*
63. *Reggam.A, Bouchelaghem.H , Houhamdi.M.,(2015) : Qualité Physico-Chimique des Eaux de l'Oued Seybouse (Nord-Est de l'Algérie): Caractérisation et Analyse en Composantes Principales (Physico-chemical quality of the waters of the Oued Seybouse (Northeastern Algeria): Characterization and Principal Component Analysis). J. Mater. Environ. Sci. 6 (5) (2015) 1417-1425. ISSN : 2028-2508.*
64. *Andriamiradis L., (2005). Mémento technique de l'eau, 2éme édition, Degremont. P: 8.*
65. *Jean L.C, (2002). La dégradation de la qualité de l'eau dans le réseau, Edition. Ministère de l'agriculture et de la pêche, Direction de l'espace rural et de la forêt, 22p.*
66. *WHO (2004) Inorganic tin in drinking-water. Background document for preparation of WHO Guidelines for drinking-water quality. Geneva, World Health Organization (WHO/SDE/WSH/03.04/115).*
67. *Kettab A., (1992). Traitement des eaux, Les eaux potables, Edition: Office des Publications Universitaires, Alger, PP : 111-123.*
68. *Boeglin J.C., (1998). Contrôle des eaux douces et de consommation humaine, réglementation française, décret 89.3 du 3 janvier 1998, Doc. P 4 210, 10p*
69. *Samake H. (2002).Thèse analyse physicochimique et bactériologique au L.N.S des eaux de consommation de la ville de Bamako durant la période 2000 et 2001.*
70. *PESCOD.M.B;(1985). Design; operation and maintenance of waste water stabilization ponds in treatment and use of sewage effluent for irrigation. Ed Pescodand Arar, 93-114.*
71. *Bremaude C., Claisse J .R., Leulier F., Thibault J., Ulrich E., (2006). Alimentation, santé, qualité de l'environnement et du cadre de vie en milieu rurale, Edition Educagri, Dijon, France, PP : 220-221.*
72. *Elmorhit.M, (2009).Hydrochimie, éléments traces métalliques et incidences écotoxicologiques sur les différentes composantes d'un écosysthème estuarien (Bas loukkos). Thèse de doctorat. Univ Mouhammed V. Agdal, Rabat, 232p.*
73. *Hubert E and wolkersdorfer C (2015) Establishing a conversion factor between electrical conductivity and total dissolved solids in South African mine waters. Water SA 41 (4) 490–500. https://doi.org/10.4314/wsa.v41i4.08*
74. *McNeil V H and Cox M E (2000). Relationship between conductivity and analyzed composition in a large set of natural surface-water samples, Queensland, Australia Environ. Geol. 39 1325– 1333*
75. *Balasubramanian S, pugalenthi V, anuradha K and chakradhar S (1999) Characterization of tannery effluents and the correlation between tds, bod and cod. J. Environ. Sci. Health Part A 34 (2) 461–478. https://doi.org/10.1080/10934529909376847*
76. *Chapter 5—Sampling. (2017).In NPDES Compliance Inspection Manual; U.S. Environmental Protection Agency: Washington, DC, USA. Available online: https://www.epa.gov/sites/default/files/2017-03/documents/npdesinspect-chapter-05.pdf (accessed on 25 May 2023).*
77. *Raju, N. J. (2007). A season-wise estimation of total dissolved solids from electrical conductance and silica in ground waters of upper Gunjanaeru River basin, Kadapa district, Andhra Pradesh. Current Science, 371-376.*
78. *Gustafson H and behrman AS (1939) Determination of total dissolved solids in water by electrical conductivity. Ind. Eng. Chem. Anal. Ed. 11 (7) 355–357. https://doi.org/10.1021/ac50135a001*
79. *Walton NRG (1989) Electrical conductivity and total dissolved solids—what is their precise relationship? Desalination 72 (3) 275–292. https://doi.org/10.1016/0011-9164(89)80012-8*
80. *Coury L (1999) Conductance Measurements Part 1: Theory. Curr. Sep.18 (3) 91–96.*
81. *Benamar N., Mouadih N., Benamar A., (2011). Étude de la biodiversité et de la pollution dans les canaux de l'Ouest algérien: le cas de l'oued Cheliff, Colloque international, Usages écologiques, économiques et sociaux de l'eau agricole en méditerranée: quels enjeux pour quels services ?, Université de Provence, Marseille, 20-21 janvier 2011, 6 p.*

82. *Nouayti N., Khattach D., Hilali M., (2015). Evaluation de la qualité physico-chimique des eaux souterraines des nappes du Jurassique du haut bassin de Ziz (Haut Atlas central, Maroc), Journal de Matériel et Science de l'Environnement, Vol 6, N° 4, PP : 1068-1081.*
83. *Gaujour D., (1995). La pollution des milieux aquatiques : Aide mémoire. 2éme édition,Lavoisier, P49.*
84. *ATHAMENA. M, (2006). Etude des ressources thermales de l'ensemble Sud sétifien. Algerie. Mémoire de Magistère. Univ Batna. 131p*
85. *Sedrati N., (2011). Origines et caractéristiques physico-chimiques des eaux de la wilaya de Biskra-sud est Algérien, thèse de doctorat en géologie, Hydrogéologie, faculté des sciences de la terre, département de géologie, Université Badji Mokhtar-Annaba, 252p.*
86. *Khellili R., LazaliD. (2015). Etude des propriétés physico-chimiques et bactériologiques de l'eau du barrage Harraza (Wilaya de Ain Defla) (2005).*
87. *Service de l'Eau (SEVESC). (2013). Qualité de l'eau potable en sortie de l'usine de traitement d'eau potable de Versailles et Saint Cloud, 11p.*
88. *Rodier J, (1978) : L'analyse de l'eau. Eaux naturelles, eaux résiduaires, eau de mer, 6ème édition.*
89. *BRGM, (2007). Suivi de la qualité des eaux souterraines de Martinique, compagne de saison des pluies 2006 : Résultats et interprétation.*
90. *Saoudi A, (2017) : Spécificités géologiques et hydrogéologiques de la régiondes lacs et de la plaine de Remila et son impact sur le tracé routier Oum el Bouaghi – Khenchela. Mémoire master en géologie université alarbi ben mehidi oum el bouaghi*
91. *Belkhiri.l, Mouni.L, Boudoukha.A, (2012). Geochemical evolution of groundwater in an alluvial aquifer: Case of El Eulma. aquifer, East Algeria. Journal of African Earth Sciences 66–67 (2012) 46–55.*
92. *Humbert et Pommier, (1988) in Tarik A, (2005) : Qualité physico-chimique de l'eau de boisson et la solubilité de certains médicaments utilisés chez la volaille dans certaines Région du Maroc. Thèse pour l'obtention du doctorat vétérinaire IAV Hassan II. Rabat. Maroc, (2005) 183p.*
93. *Parizot M, (2008) : Contrôle de surveillance de la qualité des masses d'eau souterraine de la Guyane conformément à la Directive Cadre Européenne sur l'Eau : saison des pluies 2008 Rapport final BRGM/RP-56890-FR.*
94. *Phogat, V., Skewes, M. A., Cox, J. W., Sanderson, G., Alam, J., & Šimůnek, J. (2014). Seasonal simulation of water, salinity and nitrate dynamics under drip irrigated mandarin (Citrus reticulata) and assessing management options for drainage and nitrate leaching. Journal of Hydrology, 513, 504-516.*
95. *Peck.H.D; (1970). Sulphur requirements and metabolism of microorganisms.*
96. *Adjovu, G. E., Stephen, H., & Ahmad, S. (2023). A machine learning approach for the estimation of total dissolved solids concentration in lake mead using electrical conductivity and temperature. Water, 15(13), 2439.*
97. *Hayashi M (2004). Temperature-electrical conductivity relation of water for environmental monitoring and geophysical data inversion Environ. Monit. Assess. 96 119–128*
98. *Jamshidzadeh, Z., & Barzi, M. T. (2018). Groundwater quality assessment using the potability water quality index (PWQI): a case in the Kashan plain, Central Iran. Environmental earth sciences, 77, 1-13.*
99. *Garcia, J., Heo, J., & Kim, C. (2024). Machine Learning Algorithms for Water Quality Management Using Total Dissolved Solids (TDS) Data Analysis. Water, 16(18), 2639.*
100. *Emenike CP, Tenebe IT, Omole DO, Ngene BU, Oniemayin BI, Maxwell O, OnokaBI (2017) Accessing safe drinking water in sub-Saharan Africa: issues and challenges in south-west Nigeria. Sustain Cities Soc 30:263–272. https ://doi.org/10.1016/j.scs.2017.01.005*
101. *Aly A (2015) Hydrochemical characteristics of Egypt western desert oases groundwater. Arab J Geosci 8:7551–7564*
102. *Balakrishnan, P., Saleem, A., & Mallikarjun, N. D. (2011). Groundwater quality mapping using geographic information system (GIS): A case study of Gulbarga City, Karnataka, India. African Journal of Environmental Science and Technology, 5, 1069–1084.*
103. *Qureshi, S. S., Channa, A., Memon, S. A., Khan, Q., Jamali, G. A., Panhwar, A., & Saleh, T. A. (2021). Assessment of physicochemical characteristics in groundwater quality parameters. Environmental Technology & Innovation, 24, 101877.*
104. *Azhar S, Ahmad ZA, Mohd KY, Mohammad FR, Hafizan J (2015) Classification of river water quality using multivariate analysis. Procedia Environ Sci 30(1):79–84. https ://doi.org/10.1016/j.proen v.2015.10.014*
105. *Cao Y, Tang C, Song X, Liu C, Zhang Y (2016) Identifying the hydro chemical characteristics of rivers and groundwater by multivariate statistical analysis in the Sanjiang Plain, China. Appl Water Sci 6:169–178*
106. *Li P, Qian H, Wu J, Zhang Y, Zhang H (2013) Major ion chemistry of shallow groundwater in the Dongsheng coalfield, Ordos Basin, China. Mine Water Environ 32(3):195–206. https ://doi.org/10.1007/s1023 0-013-0234-8*
107. *Shekhar, S., & Sarkar, A. (2013). Hydrogeological characterization and assessment of groundwater quality in shallow aquifers in vicinity of Najafgarh drain of NCT Delhi. Journal of Earth System Science, 122, 43-54.*
108. *Kothari, V., Vij, S., Sharma, S., & Gupta, N. (2021). Correlation of various water quality parameters and water quality index of districts of Uttarakhand. Environmental and Sustainability Indicators, 9, 100093.*

109. *Lozano, J.; Heo, J.; Seo, M. (2021).Historical Assessments of Inorganic Pollutants in the Sinkhole Region of Winkler County, Texas, USA. Sustainability, 13, 7513.*
110. *Hach Solids (Total & Dissolved). Available online: https://www.hach.com/parameters/solids (accessed on 8 June 2023).*
111. *Todd K (1980). Groundwater Hydrology. Published by John Wiley & Sons, New York Chichester, - 2nd Edition.*
112. *Jain, C. K., Bandyopadhyay, A., & Bhadra, A. (2010). Assessment of ground water quality for drinking purpose, District Nainital, Uttarakhand, India. Environmental monitoring and assessment, 166(1), 663-676.*
113. *Hattab.M.,(1998). Etude qualitatif des eaux du Haut Cheliff .Mémoire de fin d'etude ,Promotion 1998.CU Khemis- Miliana*
114. *SCET – ARGI (2)., (1985). Bilan des ressources hydriques, Etude du réaménagement et de l'extension du périmètre du moyen Chéliff Rap A1.1. Pub. Ministère de l'Hydraulique. PP 4- 28.*
115. *Tachi, A., Metaiche, M., Messoul, A., Bouguerra, H., & Tachi, S. E. (2023, September). Forecasting Groundwater Quality Parameters using Machine Learning Models: a Case Study of Khemismiliana Plain, Algeria. In Doklady Earth Sciences (Vol. 512, No. 1, pp. 907-914). Moscow: Pleiades Publishing.*
116. *Bensettiti, F., & Lacoste, A. (1999). Les ripisylves du nord de l'Algérie: essai de synthèse synsystématique à l'échelle de la Méditerranée occidentale. Ecologia mediterranea, 25(1), 13-39.*
117. *Derdous, O., Bouguerra, H., Tachi, S. E., & Bouamrane, A. (2020). A monitoring of the spatial and temporal evolutions of aridity in northern Algeria. Theoretical and Applied Climatology, 142, 1191-1198.*
118. *Glangeaud, L. (1952) : histoire géologique de la province d'Alger, IV congrès géologique international . monographie région d'Alger. vulnérabilité des nappes souterraines. Mem. Ingénieur. Univ. Ech Chlef .*
119. *Mehdi, M., Hakim, D. K., Amina, A., & Nourredine, G. (2023). Multivariate Statistical Analysis of Groundwater Quality of Hassi R'mel, Algeria. Journal of Ecological Engineering, 24(5).*
120. *Aricha.M. (2003). Le Complexe Hydrogéologique de Hassi R'Mel, document interne de la DP de Sonatrach.*
121. *Aït Ouali R., Delfaud J. (199)5. Les modalités d'ouverture du bassin des Ksour au Lias dans le cadre du rifting jurassique au Maghreb. Earth & planetary sciences, 320(8), 773–778.*
122. *Aït Ouali R., Nedjari A. (1996). La province triasique saharienne. 20 ans d'informations géologiques: bilan critique et réflexions. Bulletin du Service Géologique de l'Algérie, 7(2), 211–228.*
123. *Hamel A., Mania J. et Perriaux J. (1988). Etude géologique des grès triasiques du gisement pétrolier de Hassi R'Mel (Algérie). Caractérisation, extension et milieu de dépôt, Ph.D. Thesis, University of Franche-Comté, France.*
124. *Gaci.N. (2020). Caracterisation des eaux souterraines de la région Hassi R'Mel (w.Laghouat),mémoire de master de l'université de Bouira.*
125. *Freeze R.A., Cherry J.A. (1979). Groundwater. Prentice-Hall, Inc, Englewood Cliffs, NJ, 604p.*
126. *Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], 9(1), 381-386.*
127. *El Naqa, I., & Murphy, M. J. (2015). What is machine learning? (pp. 3-11). Springer International Publishing.*
128. *Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3(3), 210-229.*
129. *Mitchell TM. (1997). Machine learning. New York: McGraw-Hill.*
130. *Alpaydin E.(2014). Introduction to machine learning. 3rd ed. Cambridge, MA: The MIT Press.*
131. *Gyorfi, L., Ottucsak, G., & Walk, H. (Eds.). (2012). Machine learning for financial engineering (Vol. 8). World Scientific.*
132. *Ju, Lei et al. (2018). "An adaptive Gaussian process-based iterative ensemble smoother for data assimilation".In: Advances in water resources.*
133. *Mifdal, R. (2019). Application des techniques d'apprentissage automatique pour la prédiction de la tendance des titres financiers (Doctoral dissertation, École de technologie supérieure).*
134. *Salmani, Kimia (2013). "Multiple-instance active learning with online labeling". PhD thesis.*
135. *Thomas, T., Vijayaraghavan, A. P., & Emmanuel, S. (2020). Machine learning approaches in cyber security analytics (pp. 37-200). Singapore: Springer, 2020, pp. 37-200. https://doi.org/10.1007/978-981-15-1706-8*
Mythili, T., Mukherji, D., Padalia, N., & Naidu, A. (2013). A heart disease prediction model using SVM-decision trees-logistic regression (SDL). International Journal of Computer Applications, 68(16).https://doi.org/10.5120/11662-7250
136. *Breiman, L.(2001). Random forests. Machine learning, vol. 45, p. 5-32. https://doi.org/10.1023/A:1010933404324*
137. *Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H., & Kazakis, N. (2020). Improving prediction of water quality indices using novel hybrid machine-learning algorithms. Science of the Total Environment, 721, 137612.https://doi.org/10.1016/j.scitotenv.2020.137612*
138. *Lu, H., & Ma, X. (2020). Hybrid decision tree-based machine learning models for short-term water quality prediction. Chemosphere, 249, 126169.https://doi.org/10.1016/j.chemosphere.2020.126169*

139. *Hastie, T.; Tibshirani, R.; Friedman, J.(2009).Random forests. In The Elements of Statistical Learning; Springer: Berlin/Heidelberg, Germany; pp. 587–604.https://doi.org/10.1007/978-0-387-84858-7_15*
140. *Zuo, C., Sun, J., Li, J., Zhang, J., Asundi, A., & Chen, Q. (2017). High-resolution transport-of-intensity quantitative phase microscopy with annular illumination. Scientific reports, 7(1), 7654.https://doi.org/10.1038/s41598-017-06837-1*
141. *Biau, G. (2012). Analysis of a random forests model. The Journal of Machine Learning Research, 13(1), 1063-1095.*
142. *Wang, J., Zuo, R., & Xiong, Y. (2020). Mapping mineral prospectivity via semi-supervised random forest. Natural Resources Research, 29(1), 189-202.https://doi.org/10.1007/s11053-019-09510-8*
143. *Lu, H., & Ma, X. (2020). Hybrid decision tree-based machine learning models for short-term water quality prediction. Chemosphere, 249, 126169.https://doi.org/10.1016/j.chemosphere.2020.126169*
144. *Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. Advances in neural information processing systems, 31.https://arxiv.org/abs/1810.11363v1*
145. *Hussain, S., Mustafa, M. W., Jumani, T. A., Baloch, S. K., Alotaibi, H., Khan, I., & Khan, A. (2021). A novel feature engineered-CatBoost-based supervised machine learning framework for electricity theft detection. Energy Reports, 7, 4425-4436.https://doi.org/10.1016/j.egyr.2021.07.008*
146. *Vapnik, V. (2013). The nature of statistical learning theory. Springer science & business media.https://doi.org/10.1007/978-1-4757- 2440-0*
147. *Pham, B. T., Le, L. M., Le, T. T., Bui, K. T. T., Le, V. M., Ly, H. B., & Prakash, I. (2020). Development of advanced artificial intelligence models for daily rainfall prediction. Atmospheric Research, 237, 104845.https://doi.org/10.1016/j.atmosres.2020.104845*
148. *Samantaray, S., Tripathy, O., Sahoo, A., & Ghose, D. K. (2019, September). Rainfall forecasting through ANN and SVM in Bolangir Watershed, India. In Smart Intelligent Computing and Applications: Proceedings of the Third International Conference on Smart Computing and Informatics, Volume 1 (pp. 767-774). Singapore: Springer Singapore.https://doi.org/10.1007/978- 981-13-9282-5_74*
149. *Pan, Y., Jiang, J., Wang, R., Cao, H., & Cui, Y. (2009). Predicting the auto-ignition temperatures of organic compounds from molecular structure using support vector machine. Journal of hazardous materials, 164(2-3), 1242-1249.https://doi.org/10.1016/j.jhazmat.2008.09.031*
150. *Nouraki, A., Alavi, M., Golabi, M., & Albaji, M. (2021). Prediction of water quality parameters using machine learning models: a case study of the Karun River, Iran. Environmental Science and Pollution Research, 28(40), 57060-57072.https://doi.org/10.1007/s11356-021-14560-8*
151. *Maroufpoor, S., Fakheri-Fard, A., & Shiri, J. (2019). Study of the spatial distribution of groundwater quality using soft computing and geostatistical models. ISH Journal of Hydraulic Engineering, 2019, vol. 25, no 2, p. 232-238.https://doi.org/10.1080/09715010.2017.1408036*
Li, Z., Wang, G., Wang, X., Wan, L., Shi, Z., Wanke, H., ... & Uahengo, C. I. (2018). Groundwater quality and associated hydrogeochemical processes in Northwest Namibia. Journal of Geochemical Exploration, 186, 202-214.https://doi.org/10.1016/j.gexplo.2017.12.015
152. *Pan, C., Ng, K. T. W., Fallah, B., & Richter, A. (2019). Evaluation of the bias and precision of regression techniques and machine learning approaches in total dissolved solids modeling of an urban aquifer. Environmental Science and Pollution Research, 26(2), 1821-1833..https://doi.org/10.1007/s11356-018-3751-y*
153. *Khosravi, K., Nohani, E., Maroufinia, E., & Pourghasemi, H. R. (2016). A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. Natural hazards, 83(2), 947-987.https://doi.org/10.1007/s11069-016-2357-2*
154. *Thilagavathi R., Chidambaram S., Prasanna M.V.,Thivya C., Singaraja C. (2012). A study on groundwater geochemistry and water quality in layered aquifers systemof Pondicherry region, southeast India. Applied Water Science, 2, 253–269.*
155. *Blake S., Henry T., Murray J., Flood R., Muller M.R., Jones A.G., Rath V. (2016). Compositional multivariate statistical analysis of thermal groundwater provenance: A hydrogeochemical case study from Ireland, Applied Geochemistry, 75, 171–188.*
156. *Barkat A., Bouaicha F., Bouteraa O., Mester T., Ata B., Balla D., Rahal Z., Szabó G. (2021). Assessment of Complex Terminal Groundwater Aquifer for Different Use of Oued Souf Valley (Algeria) Using Multivariate Statistical Methods, Geostatistical Modeling, and Water Quality Index, Water, 13, 1609.*
157. *BRGM (1998)- Elaboration de tests de validation des données analytiques de l'Observation National de la Qualité des Eaux Souterraines (ONQES). Rap. BRGM R 40009, 91 p, 13 fig, 16 tabl, 9 ann.*