REPUBLIQUEALGERIENNE DEMOCRATIQUE ETPOPULAIRE
MINISTERE DEL'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

Université 20 Août 1955 Skikda

Faculté des Sciences

Département des Sciences Agronomiques

Filière : SciencesAgronomiques

Option : Aménagement Hydro-Agricole

Mémoirede fin d'études:

En vue del'obtention dudiplôme de Master en Sciences Agronomiques

Thème

Modélisation de la Dose Optimale de Coagulant dans le Traitement des Eaux à l'Aide de l'Intelligence Artificielle

Présenté par:

- **Boubellouta Rima**
- **Boukallel Ouissal**
- **Boulahia Ikram**
- **Medroua Bessma**
- **Zouied Djamila**

Membres de Jury :

Pr. : Heddam Salim	(Pr)	**Président**	Universitédu 20 Août1955-Skikda
Dr. : Daif Noureddine	(MAB)	**Examinateur**	Universitédu 20 Août1955-Skikda
Dr. : Merabet Khaled	(MCB)	**Promoteur**	Universitédu 20 Août1955-Skikda

Année universitaire : 2025-2026

Résumé

L'optimisation de la dose de coagulant constitue une étape importante dans les procédés de traitement des eaux potables, car elle influence directement l'efficacité de la coagulation-floculation ainsi que la qualité de l'eau produite. Traditionnellement, cette dose est déterminée à l'aide d'essais expérimentaux de type Jar-Test, qui nécessitent un temps d'exécution important et des ressources techniques considérables.La présente étude vise à développer des modèles de prédiction capables d'estimer la dose optimale de coagulant à la station de traitement des eaux de Taksebt (Algérie) à partir des paramètres physicochimiques de l'eau brute. Une base de données comprenant 1345 échantillons collectés entre 2018 et 2023 a été utilisée. Les variables explicatives considérées sont la conductivité électrique (SC), l'absorbance UV_{254}, l'oxygène dissous (DO), la température de l'eau (Tw), la couleur (COU), la turbidité (TU) et le pH.Trois approches de modélisation ont été développées et comparées: la Régression Linéaire Multiple (MLR), la Régression par Forêt Aléatoire (RFR) et le Réseau de Neurones à Perceptron Multicouche (MLPNN). La méthode SelectKBest basée sur le critère Mutual Information a été utilisée pour évaluer l'importance des variables d'entrée et construire six scénarios de modélisation.Les résultats montrent que les modèles d'intelligence artificielle surpassent largement la régression linéaire multiple. Le modèle MLPNN a obtenu les meilleures performances avec un coefficient de corrélation R = 0.922, un coefficient NSE = 0.849, RMSE = 2.498 mg/L et MAE = 1.702 mg/L en phase de validation. Le modèle RFR a présenté des performances très proches, tandis que le modèle MLR a enregistré les résultats les plus faibles.Cette étude démontre que les techniques d'intelligence artificielle constituent des outils efficaces et fiables pour la prédiction de la dose optimale de coagulant et peuvent contribuer à l'optimisation des procédés de traitement des eaux.

Mots-clés : Dose de coagulant, Traitement des eaux, Intelligence artificielle, MLPNN, RFR, MLR, Taksebt.

Abstract

Optimizing coagulant dosage is a critical step in drinking water treatment because it directly affects coagulation–flocculation efficiency and treated water quality. Traditionally, the optimal dosage is determined through Jar-Test experiments, which are time-consuming and require significant operational resources.This study aims to develop predictive models for estimating the optimal coagulant dosage at the Taksebt Water Treatment Plant (Algeria) using raw water physicochemical parameters. A dataset containing 1345 samples collected between 2018 and 2023 was used. The input variables included specific conductivity (SC), UV_{254} absorbance, dissolved oxygen (DO), water temperature (Tw), color (COU), turbidity (TU), and pH.Three modeling approaches were evaluated: Multiple Linear Regression (MLR), Random Forest Regression (RFR), and Multilayer Perceptron Neural Network (MLPNN). The SelectKBest feature selection method based on Mutual Information was applied to assess variable importance and generate six input scenarios.The results showed that artificial intelligence models significantly outperformed the conventional linear regression approach. The MLPNN model achieved the best predictive performance with R = 0.922, NSE = 0.849, RMSE = 2.498 mg/L, and MAE = 1.702 mg/L during the validation phase. The RFR model provided comparable results, whereas the MLR model produced the lowest accuracy.The findings demonstrate the strong potential of artificial intelligence techniques for reliable and accurate prediction of optimal coagulant dosage and for supporting decision-making in drinking water treatment plants.

Keywords: Coagulant dosage, Water treatment, Artificial intelligence, MLPNN, RFR, MLR, Taksebt Water Treatment Plant.

ملخص

تُعد عملية تحديد الجرعة المثلى للمخثر من أهم المراحل في معالجة مياه الشرب، نظراً لتأثيرها المباشر على كفاءة التخثير وجودة المياه المنتجة. وتعتمد هذه العملية تقليدياً على اختبارات (Jar-Test)، التي تتطلب وقتاً وجهداً وتكاليف تشغيلية معتبرة.تهدف هذه الدراسة إلى تطوير نماذج تنبؤية لتقدير الجرعة المثلى للمخثر بمحطة معالجة المياه لتاقسابت اعتماداً على الخصائص الفيزيائية والكيميائية للمياه الخام. وقد استُخدمت قاعدة بيانات تضم 1345 عينة جُمعت خلال الفترة الممتدة من سنة 2018 إلى سنة 2023. وشملت المتغيرات المدخلة: الناقلية الكهربائية النوعية(SC)، الامتصاصية فوق البنفسجيةUV_{254}، الأكسجين المذاب (DO)، درجة حرارة الماء (Tw)، اللون الظاهري (COU)، العكارة (TU) والأس الهيدروجيني(pH).تم تطوير ومقارنة ثلاثة نماذج هي: الانحدار الخطي المتعدد (MLR)،خوارزمية الغابات العشوائية (RFR) ، والشبكة العصبية متعددة الطبقات(MLPNN).كما استُخدمت طريقة SelectKBest لتحديد أهمية المتغيرات وإنشاء ستة سيناريوهات مختلفة للمدخلات.أظهرت النتائج تفوق نماذج الذكاء الاصطناعي على نموذج الانحدار الخطي المتعدد، حيث حقق نموذجالشبكة العصبية متعددة الطبقاتأفضل النتائج بقيم $R = 0.922$ و $NSE = 0.849$ و $RMSE = 2.498\ mg/L$و $MAE = 1.702\ mg/L$خلال مرحلة التحقق. كما حقق نموذج RFR نتائج متقاربة، في حين سجل نموذج MLRأضعف أداء.تؤكد هذه النتائج فعالية تقنيات الذكاء الاصطناعي في التنبؤ الدقيق بجرعة المخثر المثلى وإمكانية اعتمادها كأدوات مساعدة في تحسين تشغيل محطات معالجة المياه.

الكلمات المفتاحية: جرعة المخثر، معالجة المياه، الذكاء الاصطناعي، الشبكات العصبية متعددة الطبقات، الغابات العشوائية، الانحدار الخطي المتعدد، محطة تاقسابت.

Remerciements

بِسْمِ اللهِ الرَّحْمٰنِ الرَّحِيمِ
وَالصَّلَاةُ وَالسَّلَامُ عَلَى سَيِّدِنَا مُحَمَّدٍ ﷺ

Louange à Allah, Seigneur de l'univers, qui nous a accordé la réussite et nous a aidés à mener ce travail à son terme. Nous Lui demandons de le rendre utile et sincèrement consacré à Sa noble satisfaction.

Au moment de mettre la dernière touche à ce modeste travail, nous avons le plaisir d'adresser nos plus sincères remerciements, notre profonde gratitude et notre reconnaissance à Monsieur le Professeur Morabet Khaled, qui n'a ménagé aucun effort pour nous faire bénéficier de son savoir, de ses orientations judicieuses et de ses précieux conseils. Il a été pour nous un excellent guide et encadreur, et son suivi rigoureux ainsi que son encadrement constant ont grandement contribué à la réalisation de cette recherche dans les meilleures conditions. Nous lui exprimons notre profonde reconnaissance et prions Allah de le récompenser généreusement et de bénir son savoir et son œuvre.

Nous adressons également nos sincères remerciements et notre haute considération aux membres du jury pour avoir accepté d'évaluer ce travail, ainsi que pour les remarques constructives et les précieuses orientations qu'ils apporteront, contribuant ainsi à l'enrichissement de cette étude et à l'amélioration de sa qualité scientifique.

Nous exprimons aussi notre profonde gratitude à tous les enseignants qui ont contribué à notre formation scientifique et académique et qui ont joué un rôle essentiel dans l'enrichissement de nos connaissances tout au long de notre parcours universitaire.

Nous tenons également à remercier la direction et le personnel de la station d'épuration des eaux usées de Filfila pour leur coopération et pour avoir facilité l'accès aux informations et aux données nécessaires à la réalisation de ce travail.

Enfin, nous adressons notre profonde reconnaissance à toutes les personnes qui, de près ou de loin, ont contribué à l'aboutissement de ce travail et nous ont apporté leur aide, leur soutien et leurs encouragements tout au long de notre parcours d'études.

Nous prions Allah Tout-Puissant d'accorder à chacun réussite et prospérité, et de les récompenser de la meilleure des récompenses.

Dédicaces

Au nom d'Allah, le Tout Miséricordieux, le Très Miséricordieux. Louange à Allah qui a illuminé le chemin du savoir et m'a accordé la réussite pour mener à bien ce travail. Je Le remercie pour Ses innombrables bienfaits.

J'adresse mes plus sincères remerciements et ma profonde gratitude à mes chers parents, mon père et ma mère, qui ont été une source inestimable de soutien, de force et d'encouragement tout au long de mon parcours universitaire. Par leurs sacrifices, leurs conseils et leurs prières, ils ont grandement contribué à ma réussite. Je leur exprime toute mon affection, mon respect et ma reconnaissance. Qu'Allah les protège, les bénisse et les récompense généreusement.

Je tiens également à remercier ma sœur Wahida, ainsi que mes chères sœurs et mes frères, pour leur soutien constant, leur présence et leurs encouragements à chaque étape de ma vie. Mes remerciements s'adressent aussi à tous les membres de ma famille qui m'ont entourée de leur affection et de leur appui.

J'exprime ma profonde gratitude à tous mes enseignants qui ont contribué à ma formation scientifique et intellectuelle, et qui m'ont transmis leur savoir, leur expérience et leurs précieux conseils tout au long de mes études.

Je tiens à adresser une reconnaissance particulière à mon encadreur, Monsieur Khaled Merabet, pour la qualité de son encadrement, ses orientations judicieuses, sa disponibilité ainsi que ses conseils avisés qui ont été d'une grande importance dans la réalisation de ce travail.

Enfin, mes remerciements vont à toutes les personnes qui ont contribué, de près ou de loin, à la réussite de ce projet, ne serait-ce que par un mot d'encouragement ou une prière sincère.

À tous ceux qui ont participé à mon parcours et à l'aboutissement de ce travail, j'exprime ma profonde gratitude et mes sincères remerciements.

Que Dieu vous récompense tous abondamment et vous accorde bonheur et réussite.

ZOUIED Djamila

Dédicaces

﴿وَآخِرُ دَعْوَاهُمْ أَنِ الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ﴾

Louange à Dieu pour la qualité de l'introduction et la beauté de la conclusion. Je dédie ce succès à :

À celui qui a embelli mon nom avec les plus beaux titres, qui m'a soutenue sans limites et donné sans contrepartie, qui m'a appris que la vie est un combat et que son arme est le savoir et la connaissance, mon premier soutien dans mon parcours, ma force, mon appui et mon refuge après Dieu, ma fierté et mon honneur : mon père…

À celle pour qui Dieu a placé le paradis sous ses pieds, qui m'a portée dans son cœur avant ses bras, et qui a facilité mes épreuves par ses prières, le cœur tendre et la source de mon succès : ma mère…

À ma sœur de cœur, mon âme sœur "Marwa", qui a partagé avec moi les plus beaux moments de ma vie et a été un véritable soutien et aide. Merci pour ton amour sincère, ton soutien sans limite et ta présence qui a rendu le chemin plus facile et plus beau.

À mes chères sœurs "Sarra", "Ahlam" et "Ikhlas", qui m'ont apporté soutien et aide. Que Dieu vous accorde une longue vie pleine de bonheur.

À mes chers frères, merci pour votre amour et votre soutien. Que Dieu vous protège et vous accorde santé et bonheur.

À mes collègues "Basma", "DJamila", "Rima" et " Ouissal ", avec qui j'ai partagé ce travail et tous ses moments de stress, de fatigue et d'angoisse, mais aussi des moments de découverte, de joie et de rire. Merci pour votre coopération et votre soutien, aujourd'hui nous récoltons ensemble les fruits de nos efforts.

À mes amis que la distance a séparés, mais qui restent dans mon cœur : "Shaima", "Hedil" et "Rihab".

À tous mes proches, de près ou de loin, merci à vous tous.

Enfin, je remercie ma propre personne pour tous les efforts fournis, pour avoir cru en mes capacités et ne pas avoir abandonné. Je suis fière de moi et de ce que j'ai accompli.

BOULAHIA Ikram

Dédicaces

﴿ يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ ﴾

Louange à Allah, une louange abondante, pure et bénie, digne de la majesté de Son Visage et de la grandeur de Son Royaume. À Lui reviennent le mérite et la grâce, du début à la fin. C'est par Sa bénédiction que les bonnes œuvres s'accomplissent et que les objectifs se réalisent. Je Le loue et Le remercie pour la force, la patience et la détermination qu'Il m'a accordées afin de mener à bien ce parcours et d'atteindre ce moment dont j'ai longtemps rêvé.

Je dédie tout d'abord cette réussite à moi-même, à cette personne ambitieuse qui a cru en son rêve et qui a œuvré avec persévérance et assiduité jusqu'à sa concrétisation.

Avec tout mon amour et ma gratitude, je dédie le fruit de mes efforts et l'obtention de mon diplôme à mon cher père, qui a consacré de nombreuses années de sa vie à mon bien-être et qui a toujours été mon meilleur soutien. C'est grâce à lui, après Allah, que j'ai pu poursuivre mon chemin et gravir les échelons de la réussite.

À ma chère mère, qui m'a enseigné les valeurs et la morale avant les lettres, et dont les prières ont été une lumière guidant mes pas et une source de sérénité pour mon cœur. Qu'Allah la récompense de la meilleure des récompenses.

À mes frères et sœurs : Aymen, Rayane et Asmae, qui m'ont entourée de leur amour et de leur soutien, ont partagé avec moi les moments de joie comme de difficulté et ont toujours été un précieux appui tout au long de mon parcours.

J'adresse également mes sincères remerciements et ma profonde reconnaissance à mon encadreur, Monsieur Khaled Merabet, pour ses orientations précieuses, ses conseils avisés, son suivi constant, ainsi que pour sa patience et son dévouement dans l'encadrement de ce travail.

Je tiens aussi à exprimer ma gratitude à tous mes professeurs qui ont contribué à ma formation scientifique et intellectuelle et qui ont déployé de grands efforts pour nous instruire et nous guider.

À mes compagnons de route, qui ont partagé avec moi les joies et les difficultés de la vie universitaire et qui ont été d'excellents amis et soutiens, j'adresse toute mon affection et ma reconnaissance.

Enfin, je prie Allah le Très-Haut de faire de cette réussite le début d'un parcours encore plus riche en succès et en accomplissements, de me faire bénéficier du savoir qu'Il m'a accordé, d'en faire profiter les autres et d'accepter mon travail comme une œuvre sincèrement accomplie pour Lui. N

﴿ وَمَا تَوْفِيقِي إِلَّا بِاللَّهِ ۚ عَلَيْهِ تَوَكَّلْتُ وَإِلَيْهِ أُنِيبُ ﴾

MEDROUA Besma

Dédicaces

Louange à Allah et cela suffit, et que la prière et le salut soient sur le Bien-Aimé l'Élu, ainsi que sur sa famille et ceux qui lui sont restés fidèles. Ceci dit

Louange à Allah qui nous a accordé la réussite pour concrétiser cette étape de notre parcours universitaire. Ce mémoire, fruit de nos efforts et de notre réussite par Sa grâce, est dédié à :

À mes chers parents, qu'Allah les préserve, eux qui ont été une lumière éclairant mon chemin.

À toute ma chère famille qui m'a soutenue et continue du faire, frères et sœurs.

À mes compagnes de parcours qui ont partagé avec moi ses moments, qu'Allah les protège et leur accorde la réussite.

À tous ceux qui ont laissé une empreinte dans ma vie, et à tous ceux que mon cœur aime mais que ma plume a oubliés.

BOUBELLOUTA Rima

Dédicaces

À ceux dont les prières ont été le secret de ma réussite et une lumière éclairant mon chemin à chaque étape, à mes chers parents, qu'Allah les préserve et les maintienne comme une bénédiction dans ma vie.

À celle qui a partagé mes joies et mes peines et m'a toujours soutenue par ses paroles sincères et sa présence, ma sœur Maryam, ainsi qu'à mes autres sœurs et à tous les membres de ma famille.

À mes honorables enseignants qui ont éveillé en nous l'amour du savoir et de la connaissance et qui ont été pour nous les meilleurs guides et éducateurs.

À mes amies et collègues, Nermine et Iman, qui m'ont accompagnée tout au long de cette aventure universitaire et ont partagé avec moi les moments de fatigue, d'espoir et de persévérance.

Et à tous ceux qui m'ont tendu la main et ont laissé une empreinte positive dans mon parcours, je dédie ce modeste travail, en espérant qu'il soit le fruit d'un savoir utile et d'un effort sincère.

BOUKALLEL Ouissal

Table des matières

Résumé……...…… I
Abstract…….. II
ملخص……. III
Dédicaces et Remerciement……... IV
Sommaire……...…… V
Liste des tableaux……..……...…….. VIII
Liste des figures……..……..…...……...…….. X
Liste des abréviations……. XIV
Introduction Générale…….……...…… 1

Chapitre I : Généralités Sur la Pollution des Eaux Usées

I.1. Introduction 5
I.2. Généralités sur les eaux polluées 5
I.2.1. Définition de la pollution de l'eau 5
I.2.2. Sources de la pollution de l'eau 6
I.2.3. Types de pollution de l'eau 7
I.2.3.1. Pollution physique 7
I.2.3.2. Pollution chimique 7
I.2.3.3. Pollution biologique 8
I.2.3.4. Pollution thermique 8
I.2.3.5. Pollution radioactive 8
I.2.3.6. Pollution pétrolière 9
I.3. Eaux usées 9
I.3.1. Définition des eaux usées 9
I.3.2. Sources et types des eaux usées 10
I.3.2.1. Eaux usées industrielles 10
I.3.2.2. Eaux de lavage des rues et des espaces publics 10
I.3.2.3. Eaux pluviales polluées 11
I.3.2.4. Eaux usées domestiques 11
I.3.3. Caractéristiques physico-chimiques et biologiques des eaux usées 11
I.3.3.1. Température (T_w) 12

I.3.3.2. Potentiel hydrogène (pH) 12
I.3.3.3. Turbidité 12
I.3.3.4. Conductivité électrique (SC) 13
I.3.3.5. Oxygène dissous (DO) 13
I.3.3.6. Matières en suspension (MES) 13
I.3.3.7. Demande biochimique en oxygène (DBO_5) 13
I.3.3.8. Demande chimique en oxygène (DCO) 14
I.3.3.9. Ion ammonium (NH_4^+) 14
I.3.3.10. Ions nitrates (NO_3^-) 14
I.3.3.11. Ions nitrites (NO_2^-) 14
I.3.3.12. Orthophosphates (PO_4^{3-}) 15
I.3.3.13. Micro-organismes 15
I.4. Traitement des eaux usées 15
I.4.1. Définition du traitement des eaux usées 15
I.4.2. Objectifs du traitement des eaux usées 16
I.4.3. Étapes du traitement des eaux usées 16
I.4.3.1. Prétraitement 16
I.4.3.2. Traitement primaire 16
I.4.3.3. Traitement secondaire 17
I.4.3.4. Traitement tertiaire 17
I.5. Conclusion 17

Chapitre II : Présentation de la STEP de Skikda

II.1. Introduction 19
II.2. Présentation de la STEP de Filfila 19
II.3. Situation géographique et zone desservie 20
II.4. Caractéristiques techniques de la STEP 21
II.5. Filière de traitement des eaux usées 22
II.5.1. Prétraitement 23
II.5.2. Traitement biologique 24
II.5.3. Clarification secondaire 25
II.5.4. Désinfection et rejet final 25
II.6. Filière de traitement des boues 26
II.6.1. Épaississement des boues 27

II.6.2. Déshydratation mécanique 27
II.6.3. Séchage sur lits de séchage 27
II.6.4. Avantages de la filière retenue II.6.3. Séchage sur lits de séchage 28
II.7. Conclusion 28

Chapitre III : Présentation de la zone d'étude et des données

III.1. Introduction 30
III.2. Présentation de la station de traitement des eaux de Taksebt 30
III.3. Description de la base de données 32
III.4. Statistique descriptive des variables de qualité d'eau 32
III.5. Analyse de l'importance des variables basée sur l'Information Mutuelle 34
III.6. Stratégie de combinaison des variables d'entrée pour le développement des modèles 36
III.7.Critères d'évaluation des performances des modèles 38
III.8. Conclusion 38

Chapitre IV : Présentation des modèles utilisés

IV.1. Introduction 41
IV.2. Régression Linéaire Multiple (MLR) 41
IV.3.Régression par Forêt Aléatoire (RFR) 43
IV.4. Le réseau de neurones à perceptron multicouche (MLPNN) 45
IV.5. Outils logiciels utilisés 47
IV.5.1. Python, Anaconda et l'environnement Spyder 48
IV.5.2. Microsoft Excel 48
IV.6. Conclusion 49

Chapitre V : Résultats et discussion

V.1. Introduction 51
V.2. Performance du modèle MLR pour la prédiction de la dose optimale de coagulant 51
V.2.1. Analyse du scénario MLR1 52
V.2.2. Effet de la suppression progressive des variables 52
V.2.3. Dégradation des performances dans les scénarios MLR5 et MLR6 53
V.2.4. Analyse des graphiques en nuage de points (Scatter Plots) 54

V.2.5. Discussion des limitations du modèle MLR 55
V.3. Performance du modèle RFR pour la prédiction de la dose optimale de coagulant 56
V.3.1. Analyse du scénario RFR1 56
V.3.2. Influence de la réduction progressive des variables d'entrée 57
V.3.3. Dégradation des performances dans les scénarios RFR5 et RFR6 57
V.3.4. Analyse des graphiques en nuage de points (Scatter Plots) 58
V.3.5. Discussion des capacités et limitations du modèle RFR 60
V.4. Performance du modèle MLPNN pour la prédiction de la dose optimale de coagulant 61
V.4.1. Analyse du scénario MLPNN1 61
V.4.2. Influence de la réduction progressive des variables d'entrée 62
V.4.3. Dégradation des performances dans les scénarios MLPNN5 et MLPNN 63
V.4.4. Analyse des graphiques en nuage de points (Scatter Plots) 63
V.4.5. Discussion des capacités et limitations du modèle MLPNN 65
V.5. Comparaison des performances des modèles MLR, RFR et MLPNN 66
V.6. Conclusion 68
Conclusion Générale et perspectives 69
Références bibliographiques 72

Liste des tableaux

Tableau II.1	Principales caractéristiques techniques de la STEP de Filfila.	22
Tableau III.1	Statistiques descriptives des paramètres de qualité de l'eau à la station de Taksebt.	33
Tableau III.2	Combinaisons des variables d'entrée utilisées dans les différents modèles	37
Tableau V.1	Performances du modèle MLR pour la prédiction de la dose optimale de coagulant.	52
Tableau V.2	Performances du modèle RFR pour la prédiction de la dose optimale de coagulant.	56
Tableau V.3	Performances du modèle MLPNN pour la prédiction de la dose optimale de coagulant.	61

Liste des figures

Figure II.1	Localisation de la STEP de Filfila.	20
Figure II.2	Unité de dégrillage.	23
Figure II.3	dessableur-déshuileur.	24
Figure II.4	Schéma général du procédé de traitement de la STEP de Filfila.	26
Figure III.1	Situation géographique de la station de traitement des eaux de Taksebt	31
Figure III.2	Contribution relative (%) et scores des variables sélectionnées par la méthode KBest	35
Figure IV.1	Structure générale du modèle MLR	42
Figure IV.2	L'architecture du RFR	44
Figure IV.3	L'architecture du MLPNN	46
Figure IV.4	Environnement logiciel et démarche de développement des modèles de prédiction de la dose optimale de coagulant.	49
Figure V.1	Scatterplot des doses de coagulant observées et prédites à l'aide des modèles MLR1–MLR6 durant la phase de validation.	54
Figure V.2	Scatterplot des doses de coagulant observées et prédites à l'aide des modèles RFR1–RFR6 durant la phase de validation.	59
Figure V.3	Scatterplot des doses de coagulant observées et prédites à l'aide des modèles MLPNN1–MLPNN6 durant la phase de validation.	64

Liste des abréviations

ANN : Artificial Neural Network
COU: Couleur
$C_{v\ :}$ Coefficient de variation
DBO_5 : La demande biochimique en oxygène
DCO : La demande chimique en oxygène
DO: Dissolved Oxygen
DT : Temps de décantation
EH : Equivalent habitant
HU : Unité Hazen
IA : Intelligence Artificiel
IDE : Indice de Développement
MAE : Mean Absolute Error
MES : Les matières en suspension
MLPNN: Multilayer perceptron neural network
MLR: Multiple linear regression
MVS : Matières Volatiles en Suspension
NH_4^+ : L'azote ammoniacal
NO_2^- : Le nitrite
NO_3^- : Le nitrate
NSE : Le coefficient de Nash-Sutcliffe efficience
NTE : Neural Tree Ensemble
PH : Potentiel hydrogène
PO_4^{3-} : Les Orthophosphates
R: Coefficient de corrélation
R^2: Coefficient de détermination
RFR: Random forest regression
RMSE : Root Mean Square Error
SC: Conductivité Spécifique
$SO_{4\ :}$ Sulfates
STEP : Station d'Épuration des Eaux Usées
$S_{x\ :}$ Écart-type
T_{max}: Température maximale
T_{mean}: Température moyenne
T_{min}:Température minimale
TU : Turbidité
Tw: Température de l'eau
U_2: Vitesse du vent
$UV_{254:}$ Absorbance ultraviolette à 254 nm

X_{max} : Valeur maximale
X_{mean} : Valeur moyenne
X_{min} : Valeur minimale

Introduction Générale

Introduction Générale

L'eau potable constitue une ressource importante au développement humain, économique et social. Face à l'augmentation continue de la demande en eau et à la dégradation progressive de la qualité des ressources hydriques, les stations de traitement des eaux sont confrontées à des défis croissants pour garantir une eau conforme aux normes sanitaires (WHO, 2022). Parmi les différentes étapes du traitement conventionnel des eaux de surface, la coagulation-floculation représente un procédé fondamental permettant l'élimination des matières en suspension, des particules colloïdales ainsi que d'une partie importante de la matière organique présente dans l'eau brute (Duan & Gregory, 2003).

L'efficacité de ce procédé dépend fortement de la détermination de la dose optimale de coagulant. Une dose insuffisante peut entraîner une élimination incomplète des impuretés et une détérioration de la qualité de l'eau produite, tandis qu'un surdosage peut augmenter les coûts d'exploitation, produire davantage de boues et affecter les caractéristiques physicochimiques de l'eau traitée (Edzwald, 2011). Ainsi, l'optimisation de la dose de coagulant constitue un enjeu majeur pour les exploitants des stations de traitement des eaux.

Traditionnellement, la détermination de cette dose repose sur les essais de coagulation en laboratoire, communément appelés Jar-Tests. Bien que cette méthode soit largement utilisée, elle présente plusieurs limitations, notamment le temps nécessaire à sa réalisation, son coût opérationnel et son incapacité à fournir une estimation instantanée dans des conditions d'exploitation en temps réel (Fouchal et al., 2025). Ces contraintes ont conduit les chercheurs à explorer de nouvelles approches basées sur la modélisation mathématique et l'intelligence artificielle.

Au cours des dernières années, les techniques d'apprentissage automatique (Machine Learning) ont connu un développement considérable dans de nombreux domaines de l'ingénierie environnementale. Grâce à leur capacité à analyser de grandes quantités de données et à identifier des relations complexes entre les variables, ces méthodes offrent des perspectives prometteuses pour la prédiction et l'optimisation des procédés de traitement des eaux. Parmi les techniques les plus utilisées figurent les réseaux de neurones artificiels, les forêts aléatoires et les méthodes de régression statistique (Breiman, 2001 ; Haykin, 2009).

Dans ce contexte, la présente étude a pour objectif principal de développer et de comparer plusieurs modèles de prédiction de la dose optimale de coagulant au niveau de la station de traitement des eaux de Taksebt. L'étude repose sur une base de données expérimentale comprenant les principaux paramètres physicochimiques de l'eau brute, à savoir la conductivité électrique (SC), l'absorbance UV_{254}, l'oxygène dissous (DO), la température de l'eau (Tw), la couleur (COU), la turbidité (TU) et le potentiel hydrogène (pH). Ces paramètres ont été utilisés comme variables explicatives pour estimer la dose optimale de coagulant.

Trois approches de modélisation ont été développées et évaluées: la Régression Linéaire Multiple (MLR), la Régression par Forêt Aléatoire (RFR) et le Réseau de Neurones à Perceptron Multicouche (MLPNN). Afin d'identifier les variables les plus influentes et d'optimiser les performances des modèles, la méthode SelectKBest basée sur le critère Mutual Information a été appliquée. Six scénarios d'entrée ont ensuite été élaborés afin d'étudier l'effet de la réduction progressive des variables explicatives sur la qualité des prédictions.

Les performances des modèles ont été évaluées à l'aide de plusieurs indicateurs statistiques complémentaires, notamment le coefficient de corrélation (R), le coefficient d'efficacité de Nash–Sutcliffe (NSE), la racine carrée de l'erreur quadratique moyenne (RMSE) et l'erreur absolue moyenne (MAE). Des graphiques en nuage de points (Scatter Plots) ont également été utilisés afin d'analyser la concordance entre les valeurs observées et prédites.

Cette étude vise ainsi à contribuer au développement d'outils intelligents d'aide à la décision permettant d'améliorer la gestion et l'optimisation des procédés de traitement des eaux. Les résultats obtenus permettront également d'évaluer l'intérêt des techniques d'intelligence artificielle comme alternative aux méthodes conventionnelles de détermination de la dose de coagulant.

Le présent mémoire est structuré en cinq chapitres. Le premier chapitre présente une synthèse bibliographique sur la pollution des eaux usées. Le deuxième chapitre est consacré à la présentation de la STEP de Skikda. Le troisième chapitre est consacré à la présentation de la zone d'étude et des données. Le quatrième chapitre présente les modèles de prédiction développés ainsi que leurs fondements théoriques. Enfin, le cinquième chapitre expose les résultats obtenus, leur analyse et la comparaison des performances des différents modèles avant de conclure par les principales recommandations et perspectives de recherche.

Chapitre I

Généralités Sur la Pollution des Eaux Usées

I.1. Introduction

L'eau constitue une ressource naturelle indispensable à la vie et au développement socio-économique. Au cours des dernières décennies, la croissance démographique, l'urbanisation accélérée, l'amélioration du niveau de vie ainsi que l'expansion des activités agricoles, industrielles et touristiques ont entraîné une augmentation considérable de la demande en eau. Cette situation s'accompagne d'une production croissante d'eaux usées dont les rejets, lorsqu'ils ne sont pas correctement gérés, représentent une source majeure de dégradation des ressources hydriques et des écosystèmes aquatiques.

Face à la raréfaction des ressources en eau et aux exigences du développement durable, les eaux usées sont aujourd'hui considérées non seulement comme un déchet à traiter, mais également comme une ressource potentielle pouvant être valorisée après épuration dans différents domaines, notamment l'agriculture, l'industrie et l'aménagement des espaces verts.

Dans ce contexte, ce chapitre présente les notions fondamentales relatives à la pollution de l'eau et aux eaux usées. Il aborde les principales sources et formes de pollution, les caractéristiques des eaux usées, leurs origines ainsi que les procédés de traitement mis en œuvre dans les stations d'épuration.

I.2. Généralités sur les eaux polluées

I.2.1. Définition de la pollution de l'eau

Selon le Programme des Nations Unies pour l'Environnement (PNUE, 2021), un polluant est toute substance ou forme d'énergie introduite dans l'environnement susceptible de provoquer des effets nuisibles sur les écosystèmes, les ressources naturelles et la santé humaine.

La pollution de l'eau peut être définie comme toute modification des caractéristiques physiques, chimiques ou biologiques de l'eau, résultant d'activités naturelles ou anthropiques, susceptible de porter atteinte à la santé humaine, aux organismes vivants ou aux usages auxquels cette ressource est destinée (Maher, 2007).

D'après plusieurs définitions scientifiques modernes, les eaux polluées sont des eaux dont la qualité a été altérée à la suite de leur contamination par des eaux usées, des déchets ou

diverses substances polluantes, les rendant impropres à la consommation ou à d'autres usages (OMS, 2006 ; Metcalf & Eddy, 2014).

La pollution de l'eau résulte généralement de l'introduction de matières organiques, de produits chimiques toxiques, d'hydrocarbures, de métaux lourds ou de micro-organismes pathogènes dans les milieux aquatiques. Elle affecte aussi bien les eaux de surface (rivières, lacs, mers et océans) que les eaux souterraines, avec des conséquences importantes sur les écosystèmes, la santé publique et les activités économiques (Abdel Nabi Mohamed, 2019).

La dégradation de la qualité de l'eau constitue aujourd'hui l'un des principaux défis environnementaux à l'échelle mondiale. Les maladies liées à l'eau contaminée continuent de représenter une menace importante dans plusieurs régions du monde, en particulier dans les pays en développement où l'accès à l'eau potable et à l'assainissement demeure insuffisant (Shanie, 2018).

I.2.2. Sources de la pollution de l'eau

Les sources de pollution de l'eau sont nombreuses et peuvent être regroupées en plusieurs catégories principales:

- **Sources naturelles:** elles comprennent l'érosion des sols, l'altération des roches, les matières organiques issues de la décomposition des végétaux ainsi que certains minéraux naturellement présents dans l'environnement.
- **Sources agricoles:** elles résultent principalement de l'utilisation intensive des engrais chimiques, des pesticides, des herbicides et des effluents d'élevage. Le ruissellement des eaux d'irrigation contribue également au transport de ces polluants vers les milieux aquatiques.
- **Sources domestiques et urbaines:** elles regroupent les eaux usées domestiques, les rejets des réseaux d'assainissement, les eaux de ruissellement urbain et les déchets solides rejetés dans l'environnement.
- **Sources industrielles:** elles comprennent les effluents issus des procédés de fabrication, les métaux lourds, les solvants, les hydrocarbures et diverses substances chimiques susceptibles d'altérer la qualité des eaux.

- **Autres sources:** les activités minières, les décharges, les chantiers de construction, les installations pétrolières ainsi que certaines activités portuaires peuvent également contribuer à la pollution des ressources hydriques (Charabi et al., 1987).

I.2.3. Types de pollution de l'eau

I.2.3.1. Pollution physique

La pollution physique correspond à toute modification des propriétés physiques de l'eau, notamment la température, la couleur, l'odeur, la turbidité, la salinité ou la teneur en matières en suspension. Elle peut résulter de phénomènes naturels ou d'activités humaines.

L'augmentation de la salinité est souvent observée dans les régions arides où l'évaporation est importante. De même, les rejets industriels et urbains peuvent accroître la concentration en matières en suspension et dégrader la qualité visuelle des eaux (Al-Zarqa, 2018).

I.2.3.2. Pollution chimique

La pollution chimique résulte de l'introduction dans l'eau de substances chimiques susceptibles d'altérer sa qualité ou de présenter un danger pour les organismes vivants. Parmi les principaux polluants chimiques figurent les métaux lourds, les pesticides, les hydrocarbures, les détergents, les solvants industriels ainsi que les composés azotés et phosphorés.

Ce type de pollution est considéré comme particulièrement préoccupant en raison de la toxicité de certains composés, de leur persistance dans l'environnement et de leur capacité à s'accumuler dans les chaînes alimentaires (El-Roubi Mohamed, 2014).

Les principales sources de pollution chimique sont liées aux activités humaines. Elles comprennent notamment les rejets industriels contenant diverses substances toxiques, les activités agricoles caractérisées par l'utilisation intensive d'engrais et de pesticides, ainsi que les eaux de ruissellement urbain qui transportent de nombreux contaminants vers les milieux aquatiques. À cela s'ajoutent les pluies acides résultant des émissions atmosphériques polluantes, ainsi que les déchets ménagers et industriels qui contribuent à la dégradation de la qualité de l'eau et à la perturbation des écosystèmes aquatiques.

I.2.3.3. Pollution biologique

La pollution biologique est liée à la présence excessive de micro-organismes pathogènes tels que les bactéries, les virus, les protozoaires et les parasites dans les eaux. Elle résulte généralement de la contamination des ressources hydriques par les eaux usées domestiques, les déjections animales ou les rejets agricoles.

Cette forme de pollution représente un risque majeur pour la santé publique puisqu'elle peut être à l'origine de nombreuses maladies hydriques telles que le choléra, la typhoïde, les gastro-entérites et diverses infections parasitaires. Pour cette raison, les eaux contaminées doivent être soumises à des procédés de désinfection appropriés avant toute utilisation (OMS, 2006 ; Metcalf & Eddy, 2014).

I.2.3.4. Pollution thermique

La pollution thermique est provoquée par l'élévation anormale de la température des eaux naturelles à la suite du rejet d'eaux de refroidissement provenant des centrales électriques ou de certaines installations industrielles.

Cette augmentation de température réduit la solubilité de l'oxygène dans l'eau, perturbe les équilibres biologiques et peut entraîner la disparition de certaines espèces aquatiques sensibles aux variations thermiques (Ismaïl, 2003).

I.2.3.5. Pollution radioactive

La pollution radioactive résulte de la présence de substances radioactives dans les milieux aquatiques. Elle peut être liée aux activités nucléaires, aux essais atomiques, à l'exploitation des centrales nucléaires ou à la mauvaise gestion des déchets radioactifs.

Cette pollution présente des risques importants pour la santé humaine et les écosystèmes en raison du caractère persistant des radionucléides et de leur capacité à s'accumuler dans les organismes vivants (Ramade, 1992).

I.2.3.6. Pollution pétrolière

La pollution pétrolière est causée par le rejet accidentel ou volontaire d'hydrocarbures dans les milieux aquatiques. Les principales sources sont les marées noires, les accidents de pétroliers, les activités de raffinage, les plateformes pétrolières et les fuites des pipelines.

Les hydrocarbures forment une pellicule à la surface de l'eau qui limite les échanges gazeux avec l'atmosphère, réduit l'oxygénation des milieux aquatiques et affecte gravement la faune et la flore. Cette pollution peut également provoquer une dégradation importante des zones côtières et des activités touristiques (Bouziani, 2000).

I.3. Eaux usées

I.3.1. Définition des eaux usées

Les eaux usées sont des eaux dont la qualité a été altérée à la suite des différentes activités humaines, qu'elles soient domestiques, industrielles, commerciales ou agricoles. Elles contiennent des matières organiques et minérales, des micro-organismes ainsi que divers polluants dissous ou en suspension susceptibles d'affecter la qualité des milieux récepteurs et la santé publique lorsqu'elles sont rejetées sans traitement adéquat (Degremont, 2005).

D'une manière générale, les eaux usées sont constituées majoritairement d'eau, représentant environ 99.9 % de leur volume, tandis que les matières polluantes n'en représentent qu'une faible proportion. Malgré cette faible concentration, ces contaminants peuvent avoir des impacts environnementaux et sanitaires importants en raison de leur diversité et de leur charge polluante (Metcalf & Eddy, 2014).

Les eaux usées domestiques sont généralement classées en deux catégories principales. Les eaux noires proviennent des toilettes et contiennent des matières fécales, de l'urine ainsi qu'une forte concentration de micro-organismes pathogènes. Les eaux grises, quant à elles, proviennent des douches, des lavabos, des cuisines et des machines à laver. Elles sont généralement moins chargées en contaminants biologiques et présentent un potentiel de réutilisation plus important après un traitement approprié (Tchobanoglous et al., 2014).

Après un traitement adapté, les eaux usées traitées peuvent constituer une ressource alternative pour l'irrigation agricole, l'arrosage des espaces verts ainsi que certaines

applications industrielles, contribuant ainsi à une gestion durable des ressources en eau (OMS, 2006).

I.3.2. Sources et types des eaux usées

Les eaux usées proviennent de diverses activités humaines et leur composition varie en fonction de l'origine et de l'usage de l'eau. Elles renferment des polluants organiques, minéraux, chimiques et microbiologiques à des concentrations variables. En fonction de leur provenance, elles peuvent être classées en plusieurs catégories principales: les eaux usées industrielles, les eaux de lavage des espaces publics, les eaux pluviales polluées et les eaux usées domestiques.

I.3.2.1. Eaux usées industrielles

Les eaux usées industrielles résultent des différentes activités de production au sein des unités industrielles. L'eau y est utilisée à diverses fins, notamment pour le refroidissement des équipements, le nettoyage des installations, le transport des matières premières ou encore comme composant de certains procédés de fabrication. La composition de ces effluents varie considérablement selon le secteur d'activité concerné. Ils peuvent contenir des matières organiques et minérales, des métaux lourds, des hydrocarbures, ainsi que d'autres substances toxiques susceptibles de porter atteinte à l'environnement et à la santé publique. En raison de leur nature spécifique, ces eaux nécessitent généralement un traitement adapté avant leur rejet dans le milieu naturel ou dans les réseaux d'assainissement (Metcalf & Eddy, 2014).

I.3.2.2. Eaux de lavage des rues et des espaces publics

Les eaux de lavage des rues et des espaces publics proviennent des opérations de nettoyage des voies urbaines, des places publiques et des infrastructures collectives. Elles sont généralement collectées par les caniveaux puis acheminées vers les réseaux d'assainissement. Ces eaux transportent diverses matières telles que les sables, les poussières, les feuilles mortes, les déchets solides légers ainsi que différents polluants déposés sur les surfaces urbaines (Société Holding des Eaux Potables et de l'Assainissement, 2015).

I.3.2.3. Eaux pluviales polluées

Bien que l'eau de pluie soit relativement pure à l'origine, elle peut se charger en polluants lors de son passage dans l'atmosphère en absorbant des gaz et des particules issus des activités industrielles et du trafic routier. Lors de son ruissellement sur les surfaces urbaines et agricoles, elle entraîne également des matières en suspension, des hydrocarbures, des métaux lourds et divers contaminants vers les réseaux d'assainissement et les milieux récepteurs. Les eaux pluviales constituent ainsi une source importante de pollution diffuse des ressources hydriques (El Abed, 2010).

I.3.2.4. Eaux usées domestiques

Les eaux usées domestiques proviennent des différentes activités quotidiennes réalisées au sein des habitations, notamment la préparation des aliments, la lessive, le nettoyage et l'hygiène corporelle. Elles sont caractérisées par une forte teneur en matières organiques biodégradables, en matières en suspension, en nutriments et en micro-organismes.

Selon leur origine, les eaux usées domestiques peuvent être subdivisées en deux catégories (Ismaïl, 2003):

- **Les eaux ménagères (ou eaux grises)**, issues principalement des cuisines, des salles de bains et des buanderies. Elles contiennent des savons, des détergents, des graisses, des huiles ainsi que diverses matières organiques dissoutes.
- **Les eaux vannes (ou eaux noires)**, provenant des toilettes. Elles renferment une concentration élevée de matières organiques azotées, de micro-organismes pathogènes, de bactéries et de virus susceptibles de représenter un risque sanitaire important. En raison de leur forte charge polluante, elles nécessitent un traitement approprié avant toute réutilisation ou rejet dans le milieu naturel.

I.3.3. Caractéristiques physico-chimiques et biologiques des eaux usées

L'évaluation de la qualité des eaux usées repose sur l'analyse de plusieurs paramètres physiques, chimiques et biologiques. Ces indicateurs permettent d'apprécier le degré de pollution des effluents et de déterminer les procédés de traitement les plus appropriés.

I.3.3.1. Température (T_w)

La température constitue un paramètre physico-chimique fondamental influençant les propriétés de l'eau ainsi que l'activité biologique des micro-organismes. Elle agit directement sur la vitesse des réactions chimiques, la solubilité des gaz dissous, notamment l'oxygène, et le développement des populations microbiennes responsables des processus biologiques de traitement. Une augmentation excessive de la température peut perturber l'équilibre écologique des milieux aquatiques et réduire la concentration en oxygène dissous (Thomas, 1995).

I.3.3.2. Potentiel hydrogène (pH)

Le potentiel hydrogène (pH) exprime l'acidité ou l'alcalinité d'une solution aqueuse. Il représente un paramètre essentiel dans l'évaluation de la qualité des eaux usées puisqu'il influence les réactions chimiques et biologiques se produisant dans le milieu aquatique (Gaujous, 1999).

Le pH est défini par la relation suivante :

$$pH = -\log [H^+]$$

Dans les eaux usées domestiques, les valeurs de pH se situent généralement entre 6.5 et 8.5. Des valeurs en dehors de cette plage peuvent perturber l'activité des micro-organismes responsables des traitements biologiques.

I.3.3.3. Turbidité

La turbidité traduit le caractère trouble de l'eau causé par la présence de particules en suspension, telles que les matières organiques, les argiles, les microorganismes et les colloïdes. Elle constitue un indicateur important de la qualité physique de l'eau.

La turbidité est généralement mesurée à l'aide d'un turbidimètre et exprimée en unités néphélométriques de turbidité (NTU) (Rodier, 2009).

I.3.3.4. Conductivité électrique (SC)

La conductivité électrique représente la capacité de l'eau à conduire le courant électrique. Elle dépend principalement de la concentration des ions dissous tels que les chlorures, les sulfates, les nitrates et les bicarbonates.

Une conductivité élevée traduit généralement une forte minéralisation de l'eau ou la présence de rejets anthropiques provenant des activités domestiques, agricoles ou industrielles (Asfari, 2004).

I.3.3.5. Oxygène dissous (DO)

L'oxygène dissous est un indicateur essentiel de la qualité des eaux naturelles et usées. Il intervient dans les phénomènes d'autoépuration et dans le métabolisme des organismes aquatiques.

Sa concentration dépend principalement de la température, de la salinité, de la pression atmosphérique et de l'activité biologique du milieu. Une diminution importante de l'oxygène dissous traduit généralement une forte pollution organique (Asfari, 2004).

I.3.3.6. Matières en suspension (MES)

Les matières en suspension regroupent l'ensemble des particules solides présentes dans l'eau et non dissoutes. Elles comprennent notamment les sables, les matières organiques, les débris végétaux, les particules minérales et les déchets divers.

Une concentration élevée en MES peut provoquer l'envasement des cours d'eau, limiter la pénétration de la lumière et perturber les écosystèmes aquatiques. Leur élimination constitue une étape importante dans le traitement des eaux usées.

I.3.3.7. Demande biochimique en oxygène (DBO_5)

La demande biochimique en oxygène (DBO_5) correspond à la quantité d'oxygène consommée par les micro-organismes aérobies pour dégrader la matière organique biodégradable contenue dans l'eau.

Elle constitue l'un des principaux indicateurs de pollution organique. La DBO_5 est mesurée après une incubation de cinq jours à une température de 20 °C.

Des valeurs élevées de DBO_5 traduisent une forte charge organique susceptible d'entraîner une diminution de l'oxygène disponible dans le milieu récepteur (Degremont, 2005).

I.3.3.8. Demande chimique en oxygène (DCO)

La demande chimique en oxygène (DCO) représente la quantité d'oxygène nécessaire pour oxyder chimiquement l'ensemble des matières organiques et minérales oxydables présentes dans l'eau.

La DCO est généralement supérieure à la DBO_5 car elle prend en compte les substances non biodégradables. Elle constitue ainsi un indicateur rapide et fiable du niveau global de pollution des eaux usées (Metcalf & Eddy, 2014).

I.3.3.9. Ion ammonium (NH_4^+)

L'ion ammonium est une forme réduite de l'azote présente dans les eaux usées. Sa concentration constitue un indicateur important de pollution récente par des matières organiques d'origine domestique ou industrielle.

Des teneurs élevées en ammonium peuvent entraîner des effets toxiques sur les organismes aquatiques et favoriser les phénomènes d'eutrophisation (Asfari, 2004).

I.3.3.10. Ions nitrates (NO_3^-)

Les nitrates représentent la forme oxydée finale du cycle de l'azote. Ils proviennent principalement de l'oxydation biologique de l'ammonium ainsi que des activités agricoles utilisant des fertilisants azotés.

Des concentrations excessives de nitrates peuvent provoquer des phénomènes d'eutrophisation et dégrader la qualité des ressources hydriques (Secrétariat du Conseil Municipal de La Mecque, 2007).

I.3.3.11. Ions nitrites (NO_2^-)

Les nitrites constituent une forme intermédiaire du cycle de l'azote entre l'ammonium et les nitrates. Ils sont généralement présents en faibles concentrations dans les eaux naturelles.

En raison de leur forte réactivité chimique et de leur toxicité élevée, leur présence en quantités importantes peut représenter un risque pour la santé humaine et les organismes aquatiques (Journal Officiel de la République Algérienne, 2012).

I.3.3.12. Orthophosphates (PO_4^{3-})

Les orthophosphates sont des nutriments essentiels au développement des organismes vivants. Ils proviennent principalement des détergents domestiques, des effluents industriels et des engrais agricoles.

Toutefois, une concentration excessive en phosphates favorise la prolifération des algues et des plantes aquatiques, entraînant le phénomène d'eutrophisation des milieux aquatiques (El Abed, 2010).

I.3.3.13. Micro-organismes

Les eaux usées contiennent une grande diversité de micro-organismes tels que les bactéries, les virus, les champignons, les protozoaires et les algues microscopiques.

Certains de ces organismes sont pathogènes et représentent un risque sanitaire important. D'autres jouent un rôle essentiel dans les procédés biologiques de traitement, en participant à la dégradation et à la stabilisation de la matière organique (Société Holding des Eaux Potables et de l'Assainissement, 2015).

I.4. Traitement des eaux usées

I.4.1. Définition du traitement des eaux usées

Le traitement des eaux usées désigne l'ensemble des procédés physiques, chimiques et biologiques mis en œuvre afin de réduire ou d'éliminer les polluants présents dans les effluents avant leur rejet dans le milieu naturel ou leur réutilisation.

L'objectif principal est de protéger la santé publique, préserver les ressources hydriques et limiter les impacts environnementaux liés aux rejets d'eaux usées (Bachi, 2010).

I.4.2. Objectifs du traitement des eaux usées

Le traitement des eaux usées a pour objectif principal de réduire la charge polluante des effluents avant leur rejet dans le milieu naturel ou leur réutilisation. Il vise notamment à diminuer les concentrations en matières en suspension, en matières organiques et en divers polluants susceptibles de dégrader la qualité des ressources hydriques. Ce traitement permet également d'éliminer ou de réduire la présence des micro-organismes pathogènes responsables de nombreuses maladies d'origine hydrique, contribuant ainsi à la protection de la santé publique.

Par ailleurs, le traitement des eaux usées joue un rôle essentiel dans la préservation des eaux superficielles et souterraines en limitant les risques de contamination et de pollution des milieux récepteurs. Il contribue également à la protection des écosystèmes aquatiques en réduisant les impacts négatifs des rejets sur la faune et la flore. En outre, les procédés de traitement permettent d'obtenir une eau de qualité compatible avec certaines formes de réutilisation, notamment dans le domaine agricole, l'irrigation des espaces verts et certaines activités industrielles. Enfin, le traitement des eaux usées assure le respect des normes environnementales et sanitaires en vigueur, constituant ainsi un élément fondamental de la gestion durable et intégrée des ressources en eau (OMS, 2006 ; Metcalf & Eddy, 2014).

I.4.3. Étapes du traitement des eaux usées

I.4.3.1. Prétraitement

Le prétraitement correspond à la première étape du traitement des eaux usées. Il permet d'éliminer les déchets grossiers, les sables, les graisses et les huiles susceptibles de perturber le fonctionnement des installations de traitement.

Les principales opérations réalisées sont le dégrillage, le dessablage et le déshuilage (Metcalf & Eddy, 2014).

I.4.3.2. Traitement primaire

Le traitement primaire repose essentiellement sur des procédés physiques tels que la décantation. Il permet d'éliminer une partie importante des matières en suspension et des matières décantables présentes dans les eaux usées.

Cette étape réduit significativement la charge polluante avant le traitement biologique.

I.4.3.3. Traitement secondaire

Le traitement secondaire, également appelé traitement biologique, vise à éliminer la majeure partie de la pollution organique biodégradable grâce à l'action de micro-organismes spécialisés.

Les procédés les plus couramment utilisés sont les boues activées, les filtres biologiques et les lagunes de stabilisation (Metcalf & Eddy, 2014).

I.4.3.4. Traitement tertiaire

Le traitement tertiaire constitue une étape complémentaire destinée à améliorer davantage la qualité de l'eau traitée. Il permet notamment d'éliminer les nutriments (azote et phosphore), les matières résiduelles, certains micropolluants ainsi qu'une partie des microorganismes pathogènes.

Cette étape est particulièrement importante lorsque l'eau traitée est destinée à la réutilisation ou au rejet dans des milieux sensibles (Degremont, 2005).

I.5. Conclusion

La pollution des eaux constitue aujourd'hui l'un des principaux défis environnementaux à l'échelle mondiale. Les eaux usées issues des activités domestiques, agricoles et industrielles contiennent de nombreux polluants susceptibles d'altérer la qualité des ressources hydriques et de compromettre l'équilibre des écosystèmes.

La connaissance des caractéristiques physiques, chimiques et biologiques des eaux usées est indispensable pour évaluer leur degré de pollution et sélectionner les procédés de traitement les plus appropriés. Les différentes étapes de traitement mises en œuvre dans les stations d'épuration permettent de réduire efficacement la charge polluante et de produire une eau conforme aux normes de rejet ou de réutilisation.

Ainsi, le développement des infrastructures d'assainissement et l'amélioration continue des technologies de traitement constituent des éléments essentiels pour la protection de l'environnement, la préservation des ressources en eau et la promotion du développement durable.

Chapitre II

Présentation de la STEP de Skikda

II.1. Introduction

Les eaux usées constituent l'une des principales sources de pollution des milieux aquatiques lorsqu'elles sont rejetées sans traitement préalable. Elles contiennent des matières en suspension, des substances organiques biodégradables, des nutriments ainsi que des microorganismes pathogènes susceptibles d'altérer la qualité des ressources en eau et de présenter des risques pour la santé publique. Afin de limiter ces impacts, les systèmes d'assainissement assurent la collecte et l'acheminement des eaux usées vers les stations d'épuration (STEP), où elles subissent différents traitements physiques, biologiques et, dans certains cas, chimiques avant leur rejet dans le milieu naturel ou leur réutilisation (Metcalf & Eddy, 2014).

Dans ce contexte, la station d'épuration de Filfila constitue un ouvrage stratégique pour la protection du littoral de la wilaya de Skikda et l'amélioration de la qualité des eaux réceptrices.

II.2. Présentation de la STEP de Filfila

La station d'épuration des eaux usées (STEP) de Filfila, située dans la wilaya de Skikda, constitue une infrastructure environnementale essentielle destinée au traitement des eaux usées urbaines provenant principalement des zones résidentielles, touristiques et des services de la région. Sa réalisation s'inscrit dans le cadre de la politique nationale de protection de l'environnement et de préservation des ressources hydriques.

La station a été conçue pour réduire les impacts environnementaux liés au rejet direct des eaux usées dans le milieu naturel, notamment dans les zones côtières caractérisées par une forte valeur écologique et touristique. Elle contribue ainsi à la protection du littoral méditerranéen et à l'amélioration de la qualité des eaux réceptrices.

Le procédé de traitement adopté repose sur une succession d'étapes physiques et biologiques permettant l'élimination progressive des polluants contenus dans les eaux usées. Après leur collecte par le réseau d'assainissement, les effluents sont soumis à des opérations de prétraitement, de traitement biologique et de clarification avant leur désinfection et leur rejet final.

Par ailleurs, la station joue un rôle important dans l'amélioration des conditions sanitaires et environnementales de la région. Elle participe à la réduction de la pollution des plages et des eaux côtières et à la protection des écosystèmes naturels. Elle constitue également un outil essentiel pour accompagner le développement urbain et touristique de la région dans une perspective de développement durable.

II.3. Situation géographique et zone desservie

La station d'épuration de Filfila (Figure II.1) assure le traitement des eaux usées provenant de plusieurs agglomérations et zones d'extension urbaine et touristique, notamment Filfila-centre, le pôle urbain de Bouzaaroura, la cité Salah Chebel, la cité Larbi Ben M'Hidi ainsi que différentes zones touristiques relevant des communes de Filfila et de Skikda.

Le système d'assainissement desservi par la station repose sur deux types de réseaux:

- un réseau unitaire couvrant principalement les agglomérations de Filfila, Salah Chebel et Larbi Ben M'Hidi ;
- un réseau séparatif desservant Bouzaaroura et les zones d'expansion touristique.

Compte tenu de l'évolution démographique et de l'extension urbaine attendue, la station a été conçue pour desservir une population équivalente estimée à environ 153 000 équivalents-habitants à l'horizon 2040(Power Hydraulique, 2020).

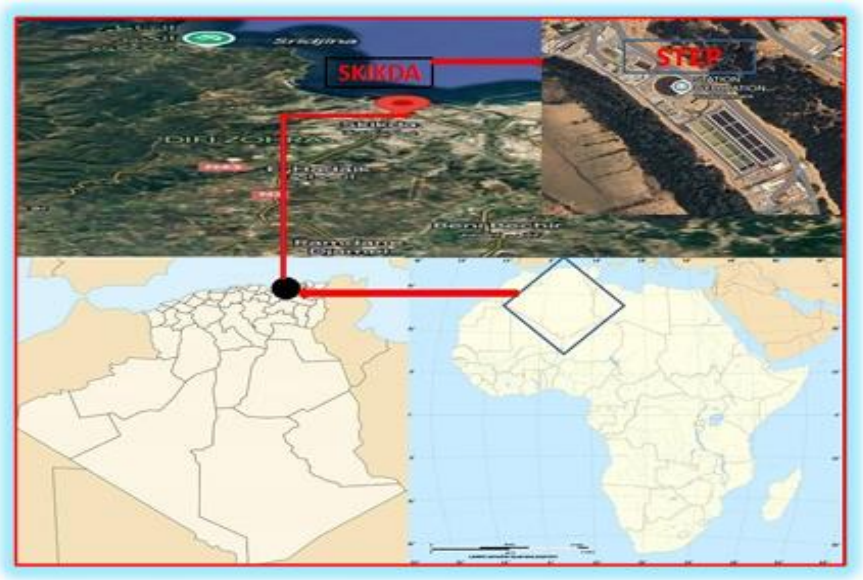


Figure II.1. Localisation de la STEP de Filfila.

II.4. Caractéristiques techniques de la STEP

La station d'épuration des eaux usées de Filfila a été conçue pour assurer le traitement des effluents urbains générés par les zones résidentielles, touristiques et les secteurs en développement de la région. Son dimensionnement repose sur les projections démographiques et les besoins futurs en matière d'assainissement, afin de garantir un niveau de traitement compatible avec les exigences environnementales et réglementaires.

La STEP adopte un procédé de traitement biologique par boues activées à faible charge, reconnu pour son efficacité dans l'élimination de la pollution organique biodégradable et des matières en suspension. Ce procédé permet de maintenir une biomasse active capable d'assurer une épuration performante tout en offrant une bonne stabilité de fonctionnement face aux variations de charge hydraulique et organique.

La capacité nominale de la station est fixée à environ 153 000 équivalents-habitants (EH), correspondant à la population actuelle et future desservie. Le débit moyen journalier de traitement est estimé à 22 000 m^3/j, valeur retenue lors des études de conception pour répondre aux besoins prévisionnels jusqu'à l'horizon 2040. La station s'étend sur une superficie d'environ 3 hectares, permettant l'implantation de l'ensemble des ouvrages de prétraitement, de traitement biologique, de clarification, de désinfection et de gestion des boues (Power Hydraulique, 2020).

Afin de garantir la qualité sanitaire des eaux traitées avant leur rejet dans le milieu récepteur, la station est équipée d'un système de désinfection par chloration. Cette étape permet de réduire la charge microbiologique résiduelle et de limiter les risques sanitaires associés aux rejets d'eaux usées traitées. L'ensemble des ouvrages a été dimensionné de manière à assurer une exploitation fiable, une continuité de service et des performances épuratoires conformes aux objectifs de protection de l'environnement et du littoral de la région de Skikda.

Les principales caractéristiques techniques de la STEP de Filfila sont présentées dans le tableau II.1.

Tableau II.1. Principales caractéristiques techniques de la STEP de Filfila.

Paramètre	Valeur
Procédé de traitement	Boues activées à faible charge
Population équivalente	153 000 EH
Débit moyen journalier	22 000 m³/j
Horizon de dimensionnement	2040
Surface approximative	3 ha
Désinfection	Chloration

La conception de la STEP de Filfila s'inscrit dans une démarche de développement durable visant à réduire les impacts environnementaux des rejets d'eaux usées, à préserver les ressources hydriques et à accompagner l'expansion urbaine et touristique de la région. Grâce à ses différentes unités de traitement, la station contribue à l'amélioration de la qualité des eaux réceptrices et à la protection des écosystèmes côtiers méditerranéens.

II.5. Filière de traitement des eaux usées

La station d'épuration des eaux usées de Filfila est conçue selon le procédé des boues activées à faible charge, l'une des technologies biologiques les plus répandues pour le traitement des eaux usées urbaines. Ce procédé se distingue par des temps de séjour relativement élevés et une faible charge appliquée à la biomasse, favorisant ainsi une dégradation efficace de la pollution organique et une bonne stabilité de fonctionnement. Selon la note de calcul du projet, la station fonctionne avec une charge massique d'environ 0.07 kg DBO_5/j/kg MVS et une charge volumique voisine de 0.15 kg DBO_5/j/m³, valeurs caractéristiques d'un système à faible charge permettant d'obtenir des rendements épuratoires élevés (Power Hydraulique, 2020).

La filière de traitement a été dimensionnée afin d'assurer l'élimination progressive des différentes formes de pollution présentes dans les eaux usées domestiques. Elle comprend plusieurs ouvrages fonctionnant de manière complémentaire, depuis la réception des effluents jusqu'au rejet des eaux traitées dans le milieu récepteur.

Les principales unités constituant la filière de traitement sont :

- le déversoir d'orage ;
- le dégrillage mécanisé ;

- le dessablage-déshuilage aéré ;
- le bassin d'aération à boues activées ;
- les clarificateurs secondaires ;
- le bassin de désinfection par chloration.

II.5.1. Prétraitement

Le prétraitement constitue la première étape de la filière d'épuration. Son rôle est de protéger les équipements situés en aval et d'améliorer les conditions de fonctionnement du traitement biologique.

Le déversoir d'orage permet de gérer les surcharges hydrauliques pouvant survenir lors des épisodes pluvieux. Il limite l'arrivée de débits excessifs dans la station et contribue à la protection des ouvrages de traitement.

Les eaux usées transitent ensuite par l'unité de dégrillage (Figure II.2) où sont retenus les déchets grossiers tels que les plastiques, papiers, chiffons, bois et autres matières volumineuses susceptibles d'endommager les équipements électromécaniques.



Figure II.2. Unité de dégrillage

Après le dégrillage, les effluents sont dirigés vers le dessableur-déshuileur aéré(Figure II.3). Cette unité assure l'élimination des particules minérales lourdes (sables, graviers, fragments minéraux) par décantation ainsi que la séparation des huiles, graisses et matières flottantes. L'extraction de ces éléments permet de réduire l'usure des équipements et d'éviter leur accumulation dans les bassins biologiques.



Figure II.3. dessableur-déshuileur.

II.5.2. Traitement biologique

Le traitement biologique constitue le cœur du procédé d'épuration. Il est réalisé dans un bassin d'aération fonctionnant selon le principe des boues activées à faible charge.

Dans ce bassin, les eaux prétraitées sont mises en contact avec une biomasse constituée principalement de bactéries, protozoaires et autres microorganismes capables de dégrader les matières organiques biodégradables présentes dans les effluents. L'apport continu d'oxygène par le système d'aération crée des conditions favorables au développement de cette biomasse et à l'oxydation de la pollution carbonée.

La matière organique dissoute est ainsi transformée en biomasse, en dioxyde de carbone, en eau et en composés minéraux. Cette étape permet de réduire significativement la demande biochimique en oxygène (DBO_5), la demande chimique en oxygène (DCO) ainsi qu'une partie des composés azotés présents dans les eaux usées.

II.5.3. Clarification secondaire

À la sortie du bassin d'aération, le mélange liquide-boues, appelé liqueur mixte, est acheminé vers les clarificateurs secondaires. Ces ouvrages assurent la séparation gravitaire entre les eaux épurées et les flocs biologiques produits au cours du traitement.

Sous l'effet de la décantation, les boues biologiques se déposent au fond du clarificateur tandis que l'eau clarifiée est récupérée en surface. Une partie des boues décantées est recyclée vers le bassin d'aération afin de maintenir une concentration suffisante de biomasse active. L'excédent de boues est quant à lui dirigé vers la filière de traitement des boues.

Cette étape joue un rôle essentiel dans la qualité finale de l'eau traitée puisqu'elle conditionne l'élimination des matières en suspension résiduelles.

II.5.4. Désinfection et rejet final

Après clarification, les eaux épurées subissent une étape de désinfection par chloration destinée à réduire la concentration des microorganismes pathogènes encore présents dans l'effluent traité.

L'injection contrôlée du chlore permet d'améliorer la qualité microbiologique des eaux avant leur rejet dans le milieu récepteur. Cette étape constitue une mesure de sécurité sanitaire particulièrement importante dans les zones côtières où les eaux rejetées peuvent avoir un impact sur les activités touristiques, les écosystèmes marins et la qualité des eaux de baignade.

Les eaux ainsi traitées sont finalement évacuées vers le milieu récepteur conformément aux exigences réglementaires relatives à la protection de l'environnement et à la préservation des ressources hydriques.

La figure II.4 présente le schéma général de la filière de traitement adoptée au niveau de la STEP de Filfila.



Figure II.4. Schéma général du procédé de traitement de la STEP de Filfila.

II.6. Filière de traitement des boues

Au cours du processus d'épuration des eaux usées, les différentes étapes de traitement génèrent des quantités importantes de boues constituées principalement de matières organiques, de microorganismes et de matières minérales. Ces boues doivent être traitées de manière appropriée afin de réduire leur volume, leur teneur en eau et leurs nuisances potentielles avant leur évacuation ou leur valorisation éventuelle.

La STEP de Filfila est équipée d'une filière spécifique de traitement des boues permettant d'assurer leur concentration et leur déshydratation. Cette filière a été conçue pour limiter les volumes à gérer tout en réduisant les coûts d'exploitation et les contraintes liées à leur élimination.

Les principales unités de la filière boues sont :

- un épaississeur ;
- une unité de déshydratation mécanique ;
- des lits de séchage.

II.6.1. Épaississement des boues

L'épaississement constitue la première étape du traitement des boues. Son objectif est d'augmenter la concentration en matières sèches en éliminant une partie de l'eau libre contenue dans les boues biologiques extraites des clarificateurs secondaires.

Cette opération permet de réduire significativement le volume des boues à traiter dans les étapes suivantes et d'améliorer l'efficacité des procédés de déshydratation. Les eaux séparées lors de cette phase sont généralement renvoyées en tête de station pour être à nouveau traitées.

II.6.2. Déshydratation mécanique

Après épaississement, les boues sont dirigées vers une unité de déshydratation mécanique destinée à extraire une quantité supplémentaire d'eau. Cette opération permet d'augmenter la teneur en matières sèches et de réduire davantage les volumes à transporter ou à stocker.

La déshydratation mécanique améliore également la maniabilité des boues et facilite leur gestion ultérieure. Selon les conditions d'exploitation, cette étape peut être complétée par un séchage naturel afin d'obtenir un produit final plus stable.

II.6.3. Séchage sur lits de séchage

Les boues déshydratées peuvent être dirigées vers des lits de séchage où elles subissent un processus de déshydratation naturelle sous l'effet combiné du drainage et de l'évaporation.

Les lits de séchage sont généralement constitués d'un système multicouche comprenant:

- une couche superficielle de sable assurant la filtration ;
- une couche drainante de gravier ;
- un réseau de drainage permettant l'évacuation des eaux percolées.

L'eau interstitielle contenue dans les boues traverse progressivement ces couches drainantes puis est collectée et évacuée. Sous l'effet des conditions climatiques, notamment du rayonnement solaire, de la température et du vent, la teneur en eau des boues diminue progressivement jusqu'à l'obtention d'un produit plus sec et plus facile à manipuler.

II.6.4. Avantages de la filière retenue

La filière de traitement des boues adoptée au niveau de la STEP de Filfila présente plusieurs avantages techniques et économiques. Le procédé des boues activées à faible charge mis en œuvre dans la station ne nécessite ni décanteur primaire ni digesteur de boues, ce qui simplifie l'exploitation des ouvrages et réduit les investissements nécessaires à la construction de la station.

Cette configuration permet également de limiter les coûts d'exploitation et de maintenance tout en garantissant une gestion efficace des boues produites. L'association de l'épaississement, de la déshydratation mécanique et du séchage naturel constitue ainsi une solution adaptée aux conditions locales et aux besoins de la station (Power Hydraulique, 2020).

II.7. Conclusion

La station d'épuration de Filfila constitue une infrastructure essentielle pour la protection de l'environnement et la préservation des ressources hydriques dans la région de Skikda. Grâce à sa capacité de traitement et à son procédé à boues activées à faible charge, elle assure une réduction significative de la pollution organique et des matières en suspension contenues dans les eaux usées urbaines.

L'étude de la station a permis de présenter sa situation géographique, la zone desservie ainsi que ses principales caractéristiques techniques. Les différentes unités composant la filière de traitement des eaux usées ont également été décrites, depuis les opérations de prétraitement jusqu'à la désinfection finale, en passant par le traitement biologique et la clarification secondaire. Par ailleurs, la filière de traitement des boues a été présentée afin de mettre en évidence les opérations d'épaississement, de déshydratation et de séchage mises en œuvre pour assurer une gestion adéquate des sous-produits issus de l'épuration.

L'ensemble de ces ouvrages fonctionne de manière complémentaire afin de garantir des performances épuratoires conformes aux exigences environnementales et de limiter l'impact des rejets sur le milieu récepteur. La STEP de Filfila contribue ainsi à l'amélioration de la qualité des eaux côtières, à la protection des écosystèmes naturels et à l'accompagnement du développement urbain et touristique de la région.

Chapitre III

Présentation de la zone d'étude et des données

III.1. Introduction

L'optimisation de la dose de coagulant est une étape essentielle dans le traitement des eaux potables, car elle influence directement l'efficacité de la coagulation-floculation et la qualité de l'eau produite. Traditionnellement, cette dose est déterminée par des essais de Jar-Test, une méthode fiable mais coûteuse en temps et en ressources. Les techniques d'intelligence artificielle offrent aujourd'hui une alternative performante permettant de prédire la dose optimale à partir des caractéristiques physicochimiques de l'eau brute.

Bien que le stage pratique ait été effectué au niveau de la STEP de Filfila, cette station ne dispose pas d'une base de données suffisamment complète pour le développement de modèles prédictifs. Par conséquent, cette étude s'appuie sur les données de la station de traitement des eaux potables de Taksebt, choisie pour la disponibilité d'un historique de données couvrant plusieurs années d'exploitation.

Ce chapitre présente la station de Taksebt, la base de données utilisée ainsi que la méthodologie adoptée pour développer les modèles de régression linéaire multiple (MLR), de réseau de neurones multicouches (MLPNN) et de forêt aléatoire (RFR) destinés à la prédiction de la dose optimale de coagulant.

III.2. Présentation de la station de traitement des eaux de Taksebt

La station de traitement des eaux potables de Taksebt, mise en service en juillet 2008, constitue l'une des infrastructures hydrauliques les plus importantes en Algérie. Elle est située à environ 8 km du barrage de Taksebt, dans la wilaya de Tizi-Ouzou, et s'étend sur une superficie d'environ 34 hectares (Figure III.1). Cette station possède une vocation régionale stratégique puisqu'elle assure l'alimentation en eau potable de 25 communes ainsi que des grandes agglomérations telles qu'Alger, Boumerdès et Tizi-Ouzou (Fouchal et al., 2025).

Le barrage de Taksebt, d'une capacité de stockage estimée à 181 millions de m^3, représente la principale source d'alimentation de la station. L'acheminement de l'eau traitée s'effectue gravitairement depuis la station vers le réservoir de Draâ Ben Khedda, puis vers le réservoir de Boudouaou à travers un réseau de conduites en fonte et de tunnels hydrauliques. Avant son transfert vers les stations de pompage et les réseaux de distribution, l'eau brute issue du barrage subit plusieurs étapes de traitement afin de satisfaire les normes de qualité de l'eau potable.

La station fonctionne avec une capacité de production journalière d'environ 607000 m³/jour, ce qui en fait la plus grande station de traitement d'eau potable en Algérie en termes de capacité (Fouchal et al., 2025).

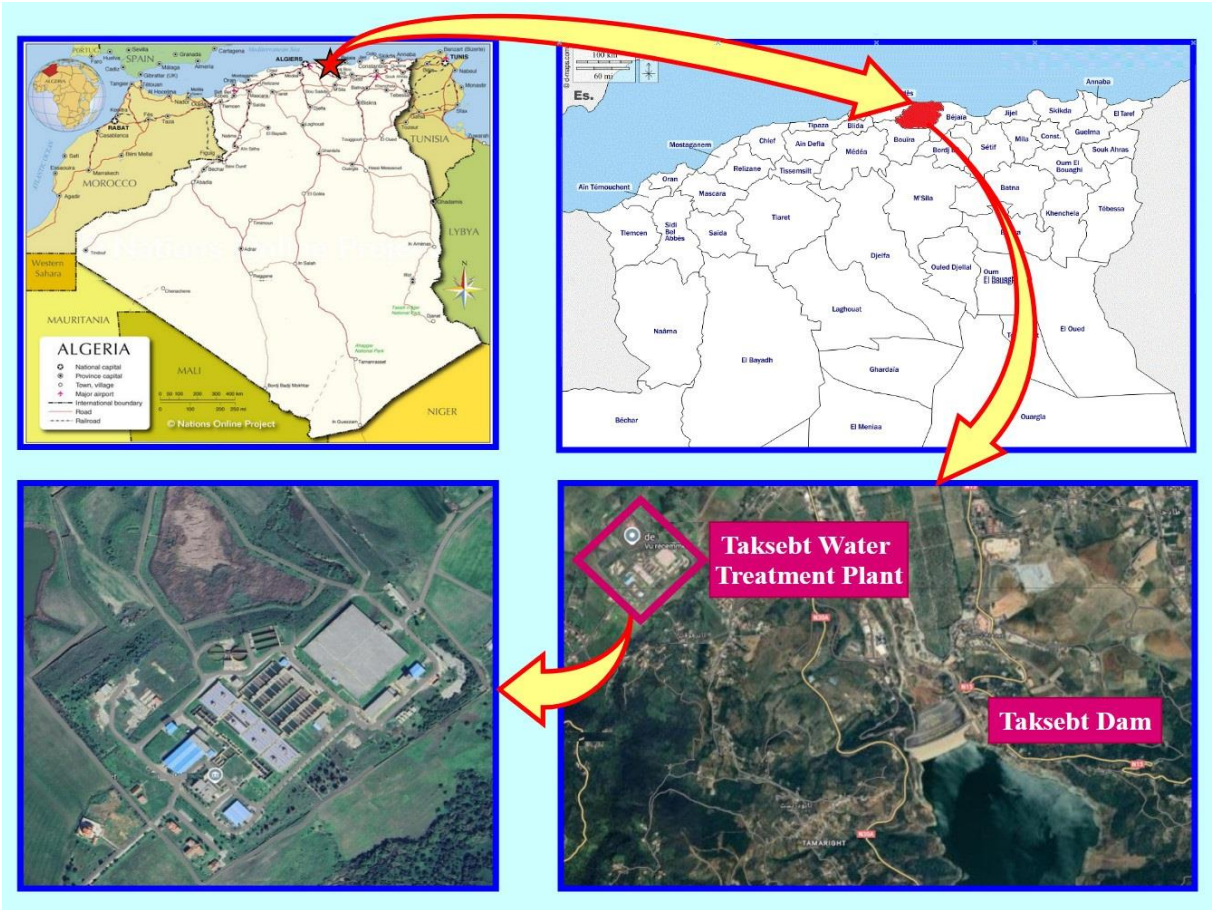


Figure III.1. Situation géographique de la station de traitement des eaux de Taksebt (Fouchal et al., 2025)

Le processus de coagulation est assuré à l'aide du sulfate d'aluminium hydraté ($Al_2(SO_4)_3.18H_2O$), utilisé comme coagulant principal avec une dose moyenne d'environ 10.059 mg/L. Après l'étape de coagulation-floculation, les particules colloïdales et les matières en suspension sont éliminées au cours de la décantation et de la filtration. La désinfection constitue ensuite une étape indispensable afin de garantir la qualité microbiologique de l'eau distribuée. Enfin, l'eau traitée est stockée dans deux réservoirs d'une capacité totale de 28000 m³ avant sa distribution vers les consommateurs.

Dans cette étude, la station de Taksebt a été choisie comme zone d'étude en raison de son importance stratégique, de la variabilité des caractéristiques physicochimiques de l'eau brute

ainsi que de la disponibilité d'une base de données historique importante permettant l'application des techniques d'intelligence artificielle pour la modélisation de la dose de coagulant.

III.3. Description de la base de données

La présente étude repose sur une base de données collectée au niveau de la station de traitement des eaux potable de Taksebt durant la période allant du 01 février 2018 au 02 août 2023. Les données ont été obtenues à partir des analyses de contrôle réalisées quotidiennement au laboratoire de la station. L'ensemble de la base comprend 1345 échantillons représentant les caractéristiques physicochimiques de l'eau brute ainsi que les doses de coagulant appliquées au cours du traitement.

Les variables explicatives retenues dans cette étude sont: (*i*) la température de l'eau (Tw) en °C ; (*ii*) le potentiel hydrogène pH ; (*iii*) la conductivité électrique (SC) en µS/cm ; (*iv*) la turbidité (TU) en NTU ; (*v*) l'oxygène dissous (DO) en mg/L ; (*vi*) l'absorbance UV_{254} en cm^{-1} ; et (*vii*) la couleur apparente (COU) en HU.

La variable cible correspond à la dose de sulfate d'aluminium utilisée dans le procédé de coagulation. Cette dose a été déterminée expérimentalement à l'aide de la méthode conventionnelle du Jar-Test, qui constitue la méthode de référence pour l'optimisation des doses de coagulant dans les stations de traitement des eaux.

Afin de développer et valider les modèles d'intelligence artificielle, la base de données a été divisée en deux parties: (*i*) 70 % des données (942 échantillons) pour la phase d'apprentissage (Training) ; (*ii*) 30 % des données (403 échantillons) pour la phase de validation (Validation).

III.4. Statistique descriptive des variables de qualité d'eau

Le tableau III.1 présente les statistiques descriptives des différentes variables de qualité de l'eau utilisées dans cette étude. Les paramètres statistiques calculés comprennent la moyenne (X_{mean}), la valeur maximale (X_{max}), la valeur minimale (X_{min}), l'écart-type (S_x), le coefficient de variation (C_v) ainsi que le coefficient de corrélation avec la dose de coagulant (R).

L'analyse statistique montre que certaines variables présentent une faible variabilité tandis que d'autres sont caractérisées par des fluctuations importantes.

Table III.1. Statistiques descriptives des paramètres de qualité de l'eau à la station de Taksebt.

Variables	**Subset**	**Unit**	X_{mean}	X_{max}	X_{min}	S_x	C_v	R
SC	Training	*µs/cm*	488.803	655.500	419.500	43.661	0.089	0.361
	Validation	*µs/cm*	489.943	620.500	386.500	41.325	0.084	0.267
	All data	*µs/cm*	489.144	655.500	386.500	42.977	0.088	0.334
UV_{254}	Training	cm^{-1}	4.800	46.000	0.041	4.402	0.917	0.619
	Validation	cm^{-1}	4.766	55.500	0.041	5.375	1.128	0.708
	All data	cm^{-1}	4.789	55.500	0.041	4.715	0.984	0.645
DO	Training	*mg/L*	4.544	10.475	0.380	2.311	0.509	0.269
	Validation	*mg/L*	4.469	9.755	0.365	2.335	0.522	0.240
	All data	*mg/L*	4.522	10.475	0.365	2.319	0.513	0.261
Tw	Training	*°C*	17.654	29.850	8.250	3.654	0.207	0.206
	Validation	*°C*	17.349	28.700	11.200	3.244	0.187	0.159
	All data	*°C*	17.563	29.850	8.250	3.539	0.201	0.163
COU	Training	*HU*	18.855	552.000	0.000	30.976	1.643	0.343
	Validation	*HU*	22.089	727.500	0.000	49.313	2.232	0.480
	All data	*HU*	19.824	727.500	0.000	37.454	1.889	0.384
TU	Training	*NTU*	3.095	71.300	0.115	4.907	1.586	0.381
	Validation	*NTU*	3.362	91.650	0.170	7.091	2.109	0.483
	All data	*NTU*	3.175	91.650	0.115	5.652	1.780	0.410
pH	Training	/	7.653	8.290	7.105	0.230	0.030	0.290
	Validation	/	7.634	8.350	7.090	0.224	0.029	0.188
	All data	/	7.648	8.350	7.090	0.228	0.030	0.262
Dose	Training	*mg/L*	10.221	46.000	0.000	7.830	0.766	1.000
	Validation	*mg/L*	9.682	46.000	0.000	7.570	0.782	1.000
	All data	*mg/L*	10.059	46.000	0.000	7.757	0.771	1.000

La conductivité électrique (SC) présente un faible coefficient de variation ($C_V \approx 0.088$), traduisant une relative stabilité de la composition ionique de l'eau brute durant toute la période d'étude. Cette stabilité peut être expliquée par la nature géologique relativement constante du bassin versant alimentant le barrage de Taksebt. De manière similaire, le pH affiche une très faible variabilité ($C_V \approx 0.03$), indiquant un système aquatique bien tamponné et chimiquement stable. Ces paramètres stables constituent généralement des variables fiables pour l'estimation de la dose optimale de coagulant.

En revanche, plusieurs paramètres présentent une forte variabilité temporelle. La turbidité (TU), la couleur (COU) ainsi que l'absorbance UV_{254} montrent des coefficients de variation

élevés, témoignant d'importantes fluctuations de la qualité de l'eau brute. La turbidité possède un coefficient de variation supérieur à 1 dans les deux sous-ensembles, ce qui traduit des variations significatives des matières en suspension et des particules colloïdales dans l'eau. Ces variations sont souvent liées aux épisodes pluvieux, au ruissellement de surface ainsi qu'aux apports saisonniers en sédiments.

La couleur de l'eau (COU) présente également une dispersion importante avec des valeurs maximales très élevées atteignant 727.5 HU. Cette forte variabilité indique la présence fluctuante de matières organiques naturelles dissoutes dans l'eau brute. L'absorbance UV254, indicateur de la matière organique dissoute aromatique, montre aussi des variations importantes, confirmant la complexité de la composition organique de l'eau à traiter.

La température de l'eau (Tw) présente une variabilité modérée avec un coefficient de variation avoisinant 0.20. Cette variation est principalement liée aux changements saisonniers. La température joue un rôle indirect mais important dans les mécanismes de coagulation-floculation puisqu'elle influence la viscosité de l'eau ainsi que la cinétique des réactions chimiques.

L'oxygène dissous (DO) montre également une variabilité moyenne, reflétant les changements biologiques et physicochimiques du milieu aquatique au cours du temps.

Concernant la variable cible, à savoir la dose de coagulant (Dose), les résultats montrent une variabilité modérée avec un coefficient de variation d'environ 0.77. Cette variabilité confirme que la détermination de la dose optimale de coagulant dépend simultanément de paramètres stables et d'autres fortement fluctuants. Par conséquent, la relation entre les variables d'entrée et la dose de coagulant peut être considérée comme complexe et potentiellement non linéaire.

Enfin, la similarité des statistiques descriptives entre les ensembles d'apprentissage et de validation indique une bonne homogénéité de la base de données et l'absence de changements significatifs de distribution. Cela garantit la fiabilité des modèles développés ainsi qu'une bonne capacité de généralisation des techniques de prédiction utilisées dans cette étude.

III.5. Analyse de l'importance des variables basée sur l'Information Mutuelle

La Figure III.2 présente le classement des variables explicatives obtenu à l'aide de la méthode de sélection KBest basée sur le critère de l'information mutuelle (Mutual Information). Cette

approche statistique permet d'évaluer le degré de dépendance entre chaque variable d'entrée et la variable cible, à savoir la dose de coagulant. Contrairement aux méthodes classiques de corrélation linéaire, l'information mutuelle est capable de détecter aussi bien les relations linéaires que non linéaires, ce qui la rend particulièrement adaptée aux systèmes complexes tels que les procédés de traitement des eaux.

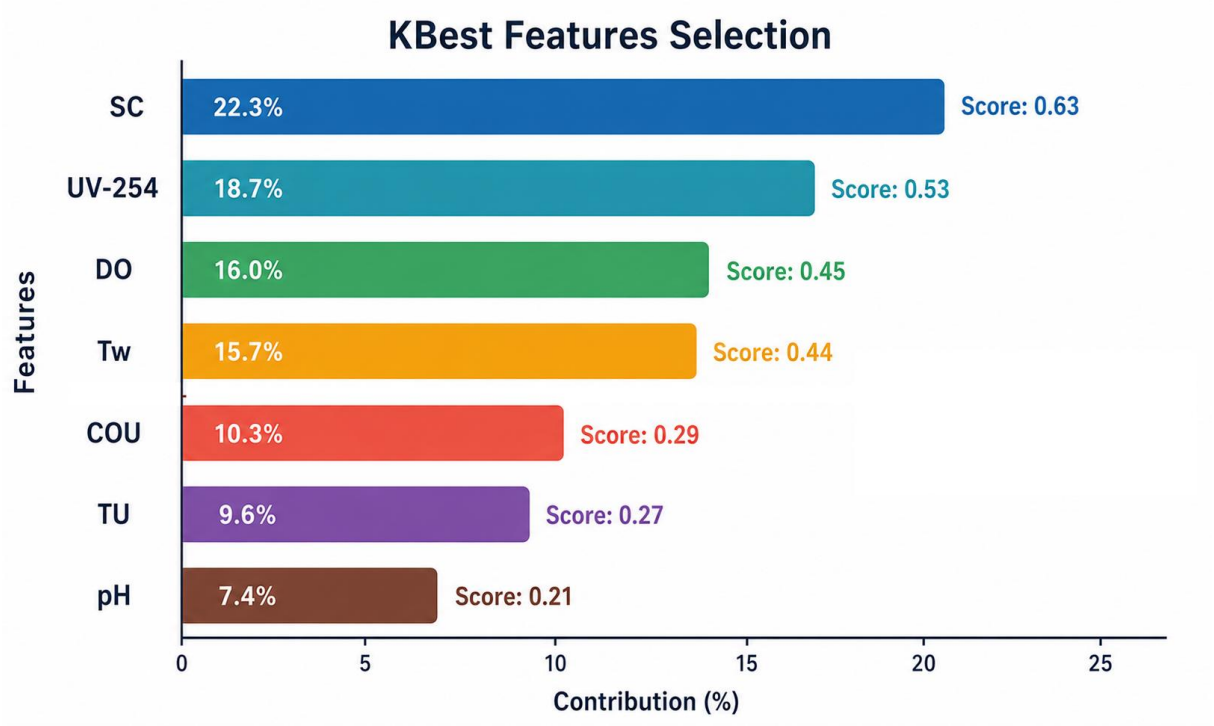


Figure III.2. Contribution relative (%) et scores des variables sélectionnées par la méthode KBest

Les résultats obtenus montrent que la conductivité électrique (SC) représente la variable la plus influente dans la prédiction de la dose de coagulant. Elle contribue à hauteur de 22.3 % du score maximal d'information mutuelle, avec une valeur de 0.63. Cette forte contribution traduit l'existence d'une relation importante entre la composition ionique de l'eau brute et les besoins en coagulant. En effet, la variation de la concentration des ions dissous peut influencer directement les mécanismes de coagulation-floculation ainsi que la stabilité des particules colloïdales présentes dans l'eau.

L'absorbance UV_{254} occupe la deuxième position avec une contribution de 18.7 % et un score de 0.53. Ce résultat met en évidence le rôle majeur de la matière organique dissoute dans le procédé de coagulation. Les composés organiques absorbant dans l'ultraviolet nécessitent

généralement des doses plus élevées de coagulant afin d'assurer leur élimination efficace et de réduire les risques de formation de sous-produits de désinfection.

L'oxygène dissous (DO) et la température de l'eau (Tw) présentent également des contributions importantes, respectivement de 16.0 % et 15.7 %. Bien que leur influence soit indirecte, ces paramètres jouent un rôle significatif dans les mécanismes physicochimiques du traitement. La température affecte notamment la viscosité de l'eau ainsi que la vitesse des réactions chimiques, tandis que l'oxygène dissous peut être associé à l'activité biologique et à l'évolution de la matière organique dans le milieu aquatique.

En revanche, certaines variables telles que la couleur (COU), la turbidité (TU) et le pH présentent des scores d'information mutuelle relativement faibles. Le pH affiche la contribution la plus faible avec seulement 7.4 % et un score de 0.21, indiquant une relation moins marquée avec la dose de coagulant comparativement aux autres paramètres étudiés. Cette faible influence peut s'expliquer par la stabilité relative du pH observée dans la base de données, ce qui réduit son impact sur les variations de dosage.

Par ailleurs, le seuil moyen de contribution, estimé à 14.3 %, montre que les variables SC, UV_{254}, DO et Tw constituent les prédicteurs les plus pertinents pour la modélisation de la dose optimale de coagulant. Ce classement confirme la pertinence de la stratégie progressive de réduction des variables d'entrée adoptée dans cette étude et présentée dans le Tableau III.2, et facilite également l'analyse comparative des performances des différents modèles d'intelligence artificielle utilisés pour la prédiction de la dose de coagulant.

Les résultats de cette analyse démontrent que la relation entre les paramètres de qualité d'eau et la dose de coagulant est complexe et non linéaire, ce qui justifie davantage l'utilisation de techniques avancées d'apprentissage automatique capables de capturer ces interactions complexes avec une meilleure précision.

III.6. Stratégie de combinaison des variables d'entrée pour le développement des modèles

Le Tableau III.2 présente les différentes configurations des variables d'entrée utilisées dans cette étude pour le développement des modèles de prédiction de la dose de coagulant. Six scénarios hiérarchiques ont été appliqués de manière uniforme à l'ensemble des méthodes de modélisation.

L'objectif principal de cette stratégie est d'évaluer l'influence progressive des variables explicatives sur les performances des différents modèles d'intelligence artificielle et d'identifier les paramètres les plus pertinents pour la prédiction de la dose optimale de coagulant.

La démarche adoptée repose sur une stratégie de réduction progressive des variables. Dans le premier scénario, l'ensemble des sept variables explicatives est utilisé.

À partir du deuxième scénario jusqu'au sixième scénario, les variables sont supprimées progressivement selon leur importance déterminée par l'analyse basée sur l'information mutuelle. Cette réduction hiérarchique permet d'étudier la sensibilité des modèles face à la diminution du nombre de variables d'entrée.

Le scénario 2 exclut la variable pH, considérée comme la moins influente. Le scénario 3 retire ensuite la turbidité (TU), suivi du scénario 4 où la couleur (COU) est également supprimée. Dans le scénario 5, seule une combinaison de trois variables importantes (SC, UV_{254} et DO) est conservée. Enfin, le scénario 6 représente la configuration minimale utilisant uniquement les deux variables les plus influentes, à savoir SC et UV_{254}.

L'utilisation des mêmes scénarios pour l'ensemble des trois modèles permet de garantir que les différences de performance observées proviennent principalement des capacités des algorithmes eux-mêmes et non d'un déséquilibre des données d'entrée.

Par ailleurs, cette réduction progressive des variables, passant de sept à deux paramètres seulement, permet également d'analyser la capacité des modèles à maintenir des performances élevées même lorsque l'information disponible devient limitée

Tableau III.2. Combinaisons des variables d'entrée utilisées dans les différents modèles

Models			**Inputs**	**Output**
MLR1	MLPNN1	RFR1	SC, UV_{254}, DO, Tw, COU, TU, pH	*Dose*
MLR2	MLPNN2	RFR2	SC, UV_{254}, DO, Tw, COU, TU	*Dose*
MLR3	MLPNN3	RFR3	SC, UV_{254}, DO, Tw, COU	*Dose*
MLR4	MLPNN4	RFR4	SC, UV_{254}, DO, Tw,	*Dose*
MLR5	MLPNN5	RFR5	SC, UV_{254}, DO	*Dose*
MLR6	MLPNN6	RFR6	SC, UV_{254}	*Dose*

III.6.Critères d'évaluation des performances des modèles

Dans cette étude, les performances des différents modèles développés pour la prédiction de la dose de coagulant ont été évaluées à l'aide de quatre indicateurs statistiques largement utilisés dans les études de modélisation environnementale et hydrologique, à savoir l'erreur absolue moyenne (MAE), la racine carrée de l'erreur quadratique moyenne (RMSE), le coefficient d'efficacité de Nash–Sutcliffe (NSE) et le coefficient de corrélation (R). Le MAE permet d'estimer l'erreur moyenne absolue entre les valeurs observées et prédites, tandis que le RMSE mesure l'importance des écarts en accordant une plus grande sensibilité aux erreurs élevées. Le coefficient NSE évalue la capacité du modèle à reproduire les valeurs observées par rapport à la moyenne des données mesurées, alors que le coefficient de corrélation R mesure le degré de la relation linéaire entre les valeurs observées et simulées. Un modèle performant doit généralement présenter de faibles valeurs de MAE et RMSE ainsi que des valeurs élevées de NSE et R, proches de 1.

$$MAE = \frac{\sum_{i=1}^{N} |Dose_{pre,i} - Dose_{obs,i}|}{N} \quad \text{(III.1)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Dose_{obs,i} - Dose_{pre,i})^2}{N}} \quad \text{(III.2)}$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^{N} (Dose_{obs,i} - Dose_{pre,i})^2}{\sum_{i=1}^{N} (Dose_{obs,i} - \overline{Dose_{obs}})^2}\right] \quad \text{(III.3)}$$

$$R = \left(\frac{\sum_{i=1}^{N} \left(Dose_{obs,i} - \overline{Dose_{obs}}\right)\left(Dose_{pre,i} - \overline{Dose_{pre}}\right)}{\sqrt{\sum_{i=1}^{N} (Dose_{obs,i} - \overline{Dose_{obs}})^2 \sum_{i=1}^{N} (Dose_{pre,i} - \overline{Dose_{pre}})^2}}\right) \quad \text{(III.4)}$$

Dans ces expressions, $Dose_{obs,i}$ et $Dose_{pre,i}$ représentent respectivement les valeurs observées et prédites de la dose de coagulant au pas de temps (i), tandis que (N) désigne le nombre total d'observations. Les termes .$\overline{Dose_{obs}}$, et $\overline{Dose_{pre}}$ correspondent respectivement aux valeurs moyennes des doses de coagulant observées et prédites.

III.7. Conclusion

Ce chapitre a présenté les principaux éléments de la base de données utilisée ainsi que les méthodes d'analyse adoptées. Dans un premier temps, une description des caractéristiques de qualité de l'eau à la station de Taksebt a été réalisée à travers des statistiques descriptives

permettant de mieux comprendre la distribution et la variabilité des différentes variables. Ensuite, une analyse exploratoire des données a été effectuée afin d'identifier les relations potentielles entre les paramètres étudiés. Enfin, la méthode de sélection des caractéristiques (SelectKBest) a permis de déterminer les variables les plus influentes pour la suite de l'étude. Ces étapes constituent une base essentielle pour le développement et l'évaluation des modèles prédictifs dans les chapitres suivants.

Chapitre IV

Présentation des modèles utilisés

IV.1. Introduction

Ce chapitre présente les différents modèles utilisés dans le cadre de cette étude pour la modélisation et la prédiction de la dose optimale de coagulant dans le traitement des eaux. Trois types de modèles ont été retenus. La régression linéaire multiple (Multiple Linear Regression –MLR) est utilisée comme méthode de référence en raison de sa simplicité et de sa facilité d'interprétation. Ensuite, la régression par forêt aléatoire (Random Forest Regression – RFR), basée sur les techniques d'ensemble (ensemble learning), permet de prendre en compte les relations non linéaires et les interactions complexes entre les variables. Enfin, le réseau de neurones à perceptron multicouche (Multilayer Perceptron Neural Network–MLPNN) est introduit comme une approche d'apprentissage automatique(machine learning) capable de modéliser des relations fortement non linéaires.

La présentation de ces modèles inclut leurs principes théoriques, leurs formulations mathématiques ainsi que leurs avantages et limites dans le contexte de la prédiction de la qualité de l'eau. Ce chapitre constitue ainsi une étape essentielle avant la phase d'application et de comparaison des performances des modèles.

IV.2. Régression Linéaire Multiple (MLR)

La régression linéaire multiple (Figure IV.1) est une méthode statistique largement utilisée pour modéliser la relation entre une variable dépendante et plusieurs variables explicatives. Elle représente une extension de la régression linéaire simple en intégrant simultanément plusieurs prédicteurs afin d'expliquer et de prédire le comportement de la variable cible. Cette approche permet d'évaluer l'influence combinée de plusieurs paramètres sur un phénomène donné et constitue l'une des techniques classiques les plus utilisées dans les études de modélisation environnementale et hydrologique (Nimon and Oswald, 2013).

Dans le cadre de cette étude, la méthode MLR a été utilisée pour établir une relation mathématique entre la dose de coagulant et les différents paramètres physicochimiques de la qualité de l'eau brute. L'objectif principal est de déterminer dans quelle mesure chaque variable d'entrée contribue à la variation de la dose optimale de coagulant.

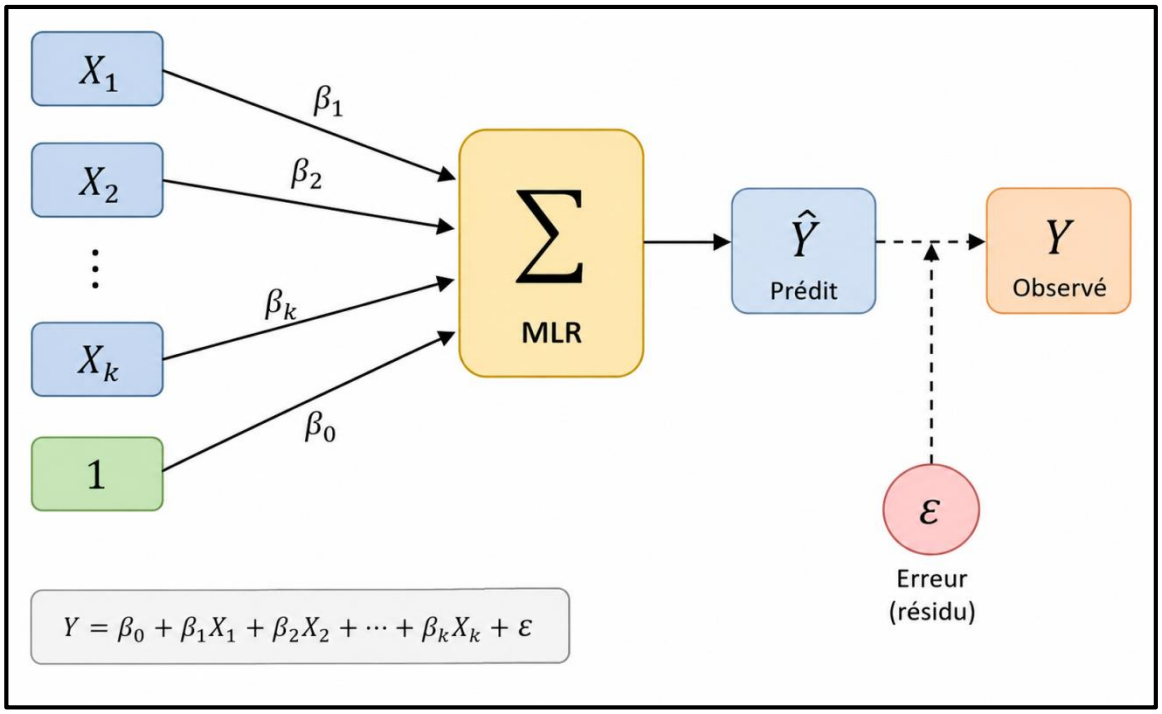


Figure IV.1.Structure générale du modèle MLR

L'un des principaux avantages de la régression linéaire multiple réside dans sa simplicité d'interprétation. En effet, chaque variable explicative est associée à un coefficient de régression permettant d'indiquer la direction de l'influence (positive ou négative) ainsi que l'importance de l'effet exercé sur la variable cible. Cette propriété facilite l'analyse des relations entre les variables et permet de mieux comprendre les mécanismes influençant le processus de coagulation.

Cependant, malgré sa simplicité et son efficacité dans certains cas, la méthode MLR repose sur plusieurs hypothèses statistiques importantes. Elle suppose principalement que la relation entre les variables explicatives et la variable cible est linéaire. Par conséquent, cette méthode peut présenter des limitations lorsqu'il existe des relations complexes ou fortement non linéaires entre les variables, comme c'est souvent le cas dans les procédés de traitement des eaux. De plus, la présence de corrélations élevées entre les variables explicatives (multicolinéarité) peut affecter la stabilité et la précision des coefficients estimés (Etemadi and Khashei, 2021 ; Shams et al., 2021).

Le principe fondamental de la régression linéaire multiple consiste à exprimer la variable cible comme une combinaison linéaire des variables explicatives. Les coefficients du modèle sont estimés à l'aide de la méthode des moindres carrés, qui cherche à minimiser la somme

des carrés des erreurs entre les valeurs observées et les valeurs prédites. Ainsi, la relation mathématique générale de la méthode MLR peut être exprimée par l'équation suivante:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad \text{(IV.1)}$$

où :

Y représente la variable dépendante ;(X_1, X_2, ... , X_k) représentent les variables explicatives ;β_0 correspond à la constante du modèle ;β_1, β_2, ... β_k représentent les coefficients de régression associés aux variables explicatives ; ε représente l'erreur résiduelle.

La méthode MLR a été largement appliquée dans les domaines de la qualité des eaux et de la prédiction des doses de coagulant. Plusieurs travaux antérieurs ont démontré son utilité pour l'analyse des paramètres influençant les procédés de traitement des eaux, notamment les études réalisées par Agirre-Basurko et al. (2006), Chakraborty and Goswami (2017) et Achite et al. (2024). Malgré certaines limites liées à sa nature linéaire, cette méthode demeure une référence importante pour comparer les performances des modèles avancés d'intelligence artificielle utilisés dans cette étude.

IV.3.Régression par Forêt Aléatoire (RFR)

La régression par forêt aléatoire (Figur IV.2) est une méthode d'apprentissage automatique appartenant à la famille des techniques d'ensemble (Ensemble Learning). Cette approche repose sur la construction d'un grand nombre d'arbres de décision durant la phase d'apprentissage, puis sur la combinaison de leurs résultats afin d'améliorer la précision des prédictions et la capacité de généralisation du modèle (Breiman, 2001). Grâce à cette stratégie collective, le modèle exploite les connaissances obtenues à partir de plusieurs architectures d'arbres pour produire des estimations plus robustes et plus fiables (Achite et al., 2023 ; Zhou et al., 2023).

Le modèle RFR est particulièrement adapté aux problèmes complexes caractérisés par des relations non linéaires entre les variables d'entrée et la variable cible. Il possède également une grande capacité à traiter des bases de données volumineuses et multidimensionnelles tout en maintenant de bonnes performances computationnelles (Ahmad et al., 2018 ; Zhang et al., 2025). Dans le cadre de cette étude, cette méthode a été utilisée pour modéliser la relation entre les paramètres physicochimiques de la qualité de l'eau et la dose optimale de coagulant.

L'élément fondamental du modèle RFR est l'arbre de décision (DT). Toutefois, contrairement à un arbre de décision classique, la forêt aléatoire introduit des mécanismes probabilistes lors du processus de construction des arbres, notamment dans la sélection des variables explicatives et dans le choix des points de séparation des données (Breiman, 2001 ; Feng et al., 2025). Cette procédure aléatoire permet de réduire la dépendance entre les arbres individuels et d'améliorer la capacité du modèle à capturer les interactions complexes entre les variables.

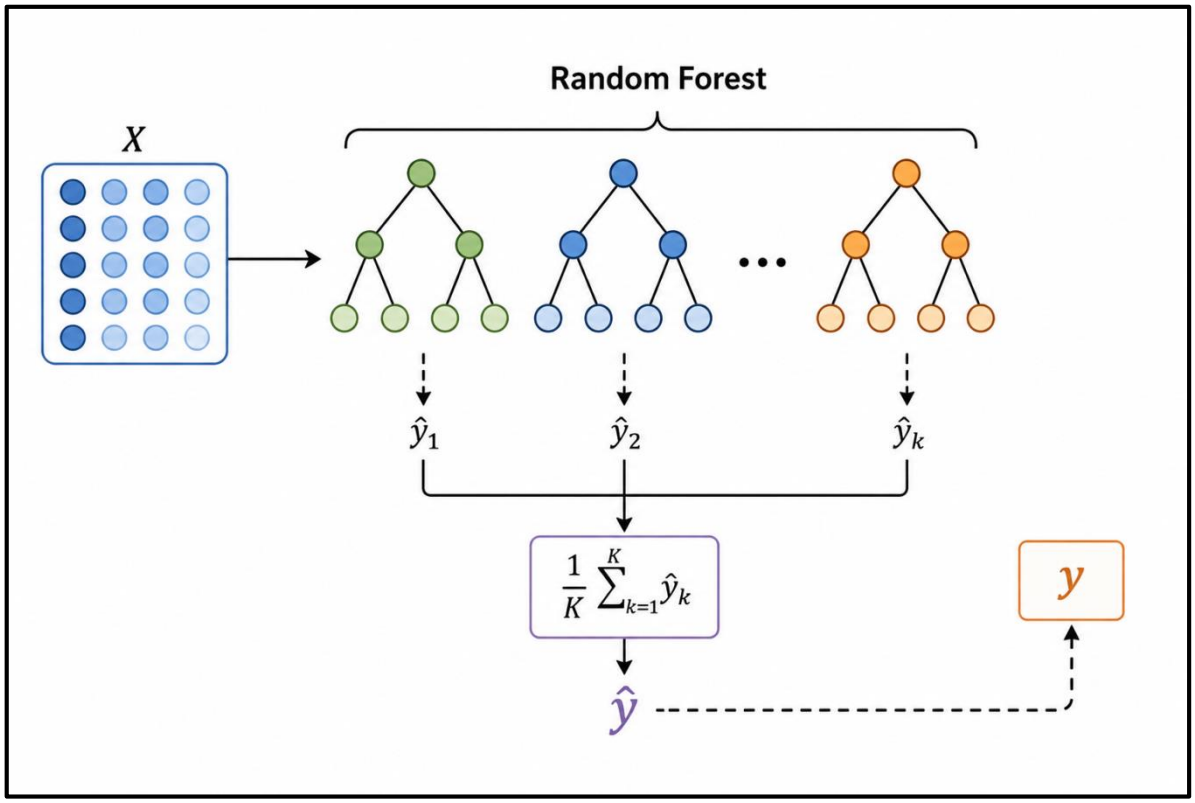


Figure IV.2. L'architecture du RFR

Par ailleurs, la diversité générée par les arbres construits de manière aléatoire contribue à limiter le phénomène de surapprentissage (Overfitting) et améliore la stabilité du modèle lors de l'application à de nouvelles données (Zhou et al., 2023 ; Xu et al., 2025).

Le modèle RFR utilise la technique du bootstrap aggregation (bagging), qui consiste à générer plusieurs sous-échantillons aléatoires à partir de la base de données d'apprentissage. Chaque arbre de décision est ensuite entraîné indépendamment sur un sous-échantillon différent. Les prédictions finales sont obtenues en calculant la moyenne des prédictions produites par l'ensemble des arbres. Cette stratégie améliore la flexibilité du modèle et réduit la variabilité des prédictions (Zhou et al., 2023).

La formulation mathématique du modèle RFR peut être exprimée par l'équation suivante:

$$\hat{f}_{rf}^{K}(X) = \frac{1}{K}\sum_{k=1}^{K} T_k(X) \quad \text{(IV.2)}$$

où :

$\hat{f}_{rf}^{K}(X)$représente la prédiction du modèle RFR utilisant (K) arbres pour une entrée (X) ; $T_k(X)$ représente la prédiction du k^{ème} arbre de décision ; $\frac{1}{K}\sum_{k=1}^{K} T_k(X)$correspond à la moyenne des prédictions générées par l'ensemble des arbres.

L'un des principaux avantages du modèle RFR est sa robustesse face au bruit présent dans les données. En effet, grâce à l'utilisation simultanée de plusieurs arbres de décision indépendants, le modèle réduit efficacement les erreurs de généralisation et améliore la précision des prédictions. Cette caractéristique est particulièrement importante dans les applications environnementales où les données présentent souvent des fluctuations importantes et des relations complexes entre les variables.

Cependant, malgré ses nombreux avantages, le modèle RFR possède certaines limites. Étant basé sur des arbres de décision, il réalise les prédictions à partir des valeurs moyennes des feuilles terminales des arbres. Par conséquent, il présente des difficultés à prédire des valeurs situées en dehors de l'intervalle des données utilisées lors de l'apprentissage (Zhou et al., 2023 ; Xu et al., 2025). Cette limitation peut affecter les performances du modèle dans le cas de conditions extrêmes ou de situations rarement observées dans la base de données.

La méthode RFR a été largement appliquée dans les domaines de l'hydrologie, de la qualité des eaux et de la modélisation environnementale. Plusieurs études récentes ont démontré son efficacité pour la prédiction des doses de coagulant et l'analyse des procédés de traitement des eaux, notamment les travaux réalisés par Ahmad et al. (2018), Feng et al. (2025) et Xu et al. (2025)

IV.4. Le réseau de neurones à perceptron multicouche (MLPNN)

Les réseaux de neurones artificiels (ANN) sont des modèles mathématiques inspirés du fonctionnement du cerveau humain. Ce sont des modèles non linéaires et appartiennent à la catégorie des techniques d'intelligence artificielle (IA), généralement appelées modèles boîte noire. L'une des architectures ANN les plus rapportées dans la littérature est le réseau de neurones à perceptron multicouche (MLPNN: Multilayer Perceptron Neural Network) (Rumelhart et al., 1986a). En raison de sa forte capacité de généralisation, il est principalement utilisé pour les problèmes de prédiction et d'approximation; et considéré

comme un approximateur universel (Hornik et al., 1989; Hornik, 1991). Le MLPNN se compose de trois types de couches: une couche d'entrée qui contient les variables explicatives, une couche de sortie qui contient la variable à modéliser, et une ou plusieurs couches cachées (Figure IV.3). Les poids et les biais sont les paramètres du modèle MLPNN. L'apprentissage du MLPNN se fait avec des paramètres initiaux aléatoires. Le processus d'apprentissage est répété plusieurs fois, afin de déterminer le meilleur modèle ayant l'erreur quadratique moyenne (MSE) la plus faible entre la variable à modéliser mesuré et calculé (Merabet & Heddam, 2023).

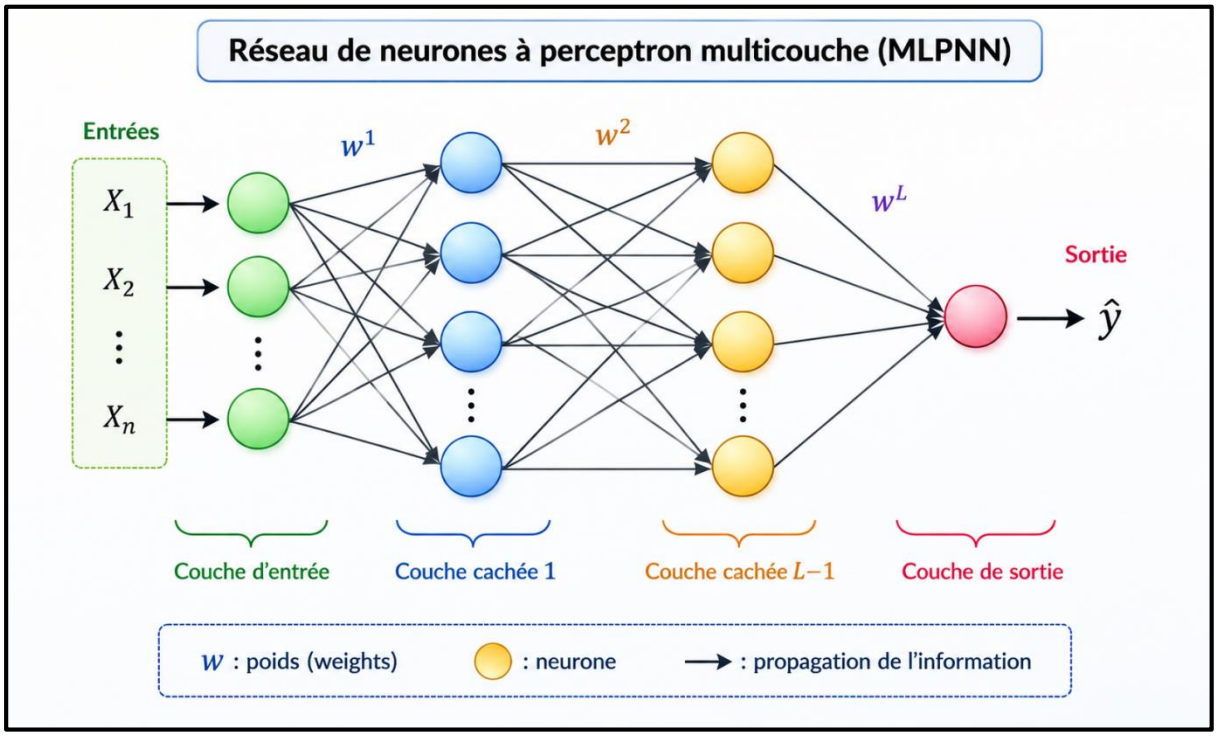


Figure IV.3. L'architecture du MLPNN

Le MLPNN utilisé dans cette étude est un réseau de neurones avec une fonction d'activation sigmoïde dans la couche cachée et une fonction d'activation linéaire dans la couche de sortie.
La formulation mathématique du MLPNN est illustrée comme suit: De la couche d'entrée à la couche cachée, chaque neurone caché (I) calcule la somme pondérée de toutes les variables d'entrée (x_i) plus un biais (δ_j):

$$I_j = \delta_j + \sum_{j=1}^{n} x_i w_{ij} \qquad \text{(IV.3)}$$

Où x_i est la variable d'entrée, w_{ij}est le poids entre l'entrée i et le neurone caché j, I_j est le niveau d'activité interne du neurone j dans la couche cachée et δ_j est le biais associé au neurone caché j.

Une fonction d'activation (f) appartenant à la couche cachée est appliquée à la somme des résultats :

$$E_j = f_1(I_j) \quad \text{(IV.4)}$$

E_j est la sortie du neurone j dans la couche cachée et f_1 la fonction d'activation, généralement la fonction sigmoïde, représentée par l'équation IV.5:

$$f_1(x) = \frac{1}{1+e^{-x}} \quad \text{(IV.5)}$$

Les sorties des neurones cachés sont utilisées comme entrées dans la couche de sortie:

$$O = \delta_0 + \sum_{j=1}^{n} E_j w_{jk} \quad \text{(IV.6)}$$

w_{jk} est le poids de connexion du neurone j dans la couche cachée au neurone unique k dans la couche de sortie; O est l'entrée du neurone de sortie k et δ_0 est le biais du neurone de sortie k. Enfin, la sortie du neurone k dans la couche de sortie est calculée à l'aide d'une fonction d'activation linéaire f_0 :

$$Y = f_2(O) \quad \text{(IV.7)}$$

La sortie du modèle MLPNN a été calculée selon Jin et al. (2016):

$$Y = f^2[W^2 f^1(W^1 X + b^1) + b^2] \quad \text{(IV.8)}$$

où Y est le vecteur de sortie (Dose de coagulant); X est le vecteur d'entrée (Sept variables physicochimiques de l'eau brute) ; f^1 et f^2 sont des fonctions de transfert des couches cachée et de sortie, sigmoïde et linéaire, respectivement; b^1 et b^2 sont respectivement les biais des couches cachée et de sortie; et W^1 et W^2 sont respectivement les poids des couches d'entrée et cachée (Jin et al., 2016).

L'apprentissage du MLPNN s'effectue à l'aide d'un algorithme d'apprentissage supervisé, le plus connu est l'algorithme de rétropropagation du gradient (LeCun, 1985; Rumelhart et al., 1986a; 1986b), qui nécessite un ensemble d'entrées et de sortie désirée. L'objectif de l'algorithme d'apprentissage est de trouver les valeurs optimal de poids et de biais, qui fournit l'erreur minimale entre le résultat désirée et la sortie actuelle du modèle (Haykin, 2009).

IV.5. Outils logiciels utilisés

Dans cette étude, plusieurs outils logiciels ont été utilisés pour le traitement des données, le développement des modèles de prédiction et l'analyse des résultats.

IV.5.1. Python, Anaconda et l'environnement Spyder

Le langage de programmation Python a été utilisé comme principal outil de modélisation et d'analyse des données dans cette étude. Les différents programmes ont été développés et exécutés à l'aide de la plateforme Anaconda, une distribution largement utilisée en science des données qui facilite la gestion des bibliothèques scientifiques et des environnements de travail. L'environnement de développement intégré (IDE) Spyder, inclus dans Anaconda, a été utilisé pour l'écriture, l'exécution et le débogage des scripts Python.

Python a été employé pour l'implémentation des modèles d'apprentissage automatique développés dans cette étude, notamment le réseau de neurones à perceptron multicouche (MLPNN) et la régression par forêt aléatoire (RFR). Il a également permis la préparation des données, l'apprentissage des modèles, la validation des résultats et l'évaluation des performances prédictives.

Parmi les bibliothèques utilisées, Scikit-learn (Sklearn) a constitué l'outil principal pour le développement des modèles de régression et l'évaluation de leurs performances. La bibliothèque Pandas a été utilisée pour la gestion et le traitement des données, NumPy pour les calculs numériques et les opérations matricielles, tandis que Matplotlib a servi à la visualisation graphique des résultats et à la comparaison entre les valeurs observées et prédites.

L'utilisation conjointe de Python, Anaconda et Spyder a permis de disposer d'un environnement de travail complet, performant et adapté aux exigences de la modélisation et de l'apprentissage automatique appliqués à la prédiction de la dose optimale de coagulant.

IV.5.2. Microsoft Excel

Le logiciel Microsoft Excel a été utilisé pour les étapes préliminaires de traitement et d'analyse des données ainsi que pour le développement du modèle de régression linéaire multiple (MLR). Grâce à ses outils statistiques intégrés, notamment le module « Analyse de données » (Data Analysis ToolPak), Excel a permis de réaliser les analyses descriptives, les matrices de corrélation et l'estimation des coefficients de la régression linéaire multiple.

Excel a également été utilisé pour l'organisation de la base de données, la vérification de la qualité des données et la réalisation de certains graphiques exploratoires. Son interface conviviale et ses fonctionnalités de calcul statistique en font un outil largement utilisé dans les études de modélisation environnementale et hydrologique.

L'association de Python et de Microsoft Excel a ainsi permis de combiner la puissance des techniques modernes d'apprentissage automatique avec la simplicité des méthodes statistiques classiques, offrant un cadre efficace pour le développement et l'évaluation des modèles de prédiction de la dose optimale de coagulant.

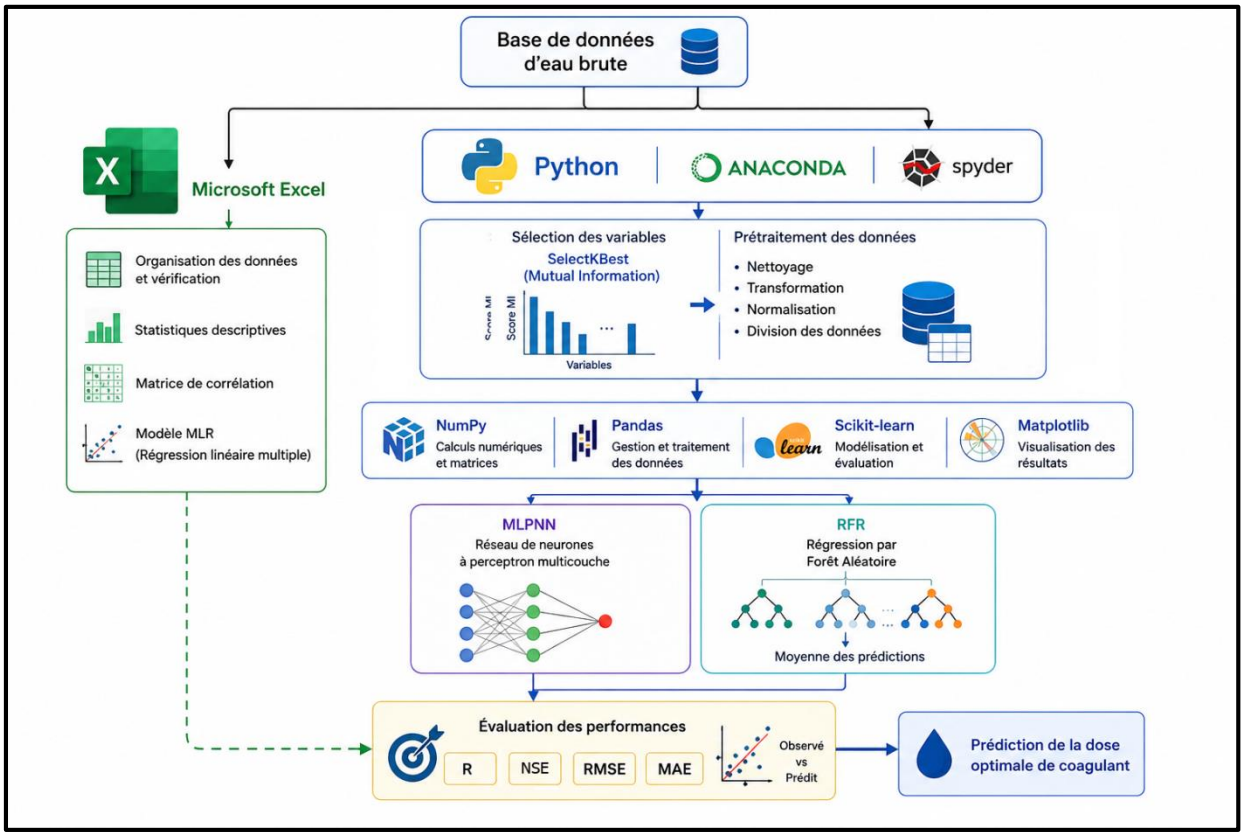


Figure IV.4. Environnement logiciel et démarche de développement des modèles de prédiction de la dose optimale de coagulant.

IV.6. Conclusion

Dans ce chapitre, les principaux modèles de prédiction utilisés dans cette étude ont été présentés, à savoir la régression linéaire multiple (MLR), la régression par forêt aléatoire (RFR) et le réseau de neurones à perceptron multicouche (MLPNN). Les fondements théoriques, les formulations mathématiques ainsi que les avantages et les limites de chaque approche ont été décrits. Par ailleurs, les outils logiciels mobilisés pour la mise en œuvre des modèles, notamment Python/Spyder et Microsoft Excel, ont été présentés. Ces éléments constituent la base méthodologique nécessaire à l'application, à l'entraînement et à la comparaison des performances des différents modèles dans le chapitre suivant.

Chapitre V:

Résultats et discussion

V.1. Introduction

Ce chapitre présente et discute les résultats obtenus à partir des trois modèles de prédiction développés dans cette étude, à savoir la régression linéaire multiple (MLR), le réseau de neurones à perceptron multicouche (MLPNN) et la régression par forêt aléatoire (RFR). L'objectif principal est d'estimer la dose optimale de coagulant au niveau de la station de traitement des eaux de Taksebt à partir des principaux paramètres physicochimiques caractérisant la qualité de l'eau brute.

Afin d'optimiser la sélection des variables explicatives, une analyse de l'importance des paramètres d'entrée a été réalisée à l'aide de la méthode SelectKBest basée sur le critère Mutual Information. Sur cette base, six scénarios d'entrée ont été élaborés afin d'étudier l'influence progressive des variables physicochimiques sur les performances des modèles de prédiction et d'identifier les combinaisons les plus pertinentes pour l'estimation de la dose optimale de coagulant.

L'analyse et l'évaluation des performances des modèles ont été réalisées à l'aide de graphiques de dispersion (Scatter Plots) et de plusieurs indicateurs statistiques complémentaires, notamment le coefficient de corrélation (R), le coefficient d'efficacité de Nash–Sutcliffe (NSE), la racine carrée de l'erreur quadratique moyenne (RMSE) et l'erreur absolue moyenne (MAE). Les Scatter Plots ont permis d'examiner les relations entre les variables explicatives et la variable cible, ainsi que de comparer les valeurs observées et prédites. Quant aux indicateurs statistiques, ils ont servi à évaluer et à comparer les performances des modèles MLR, MLPNN et RFR afin d'identifier l'approche la plus efficace pour la prédiction de la dose optimale de coagulant au niveau de la station de traitement des eaux de Taksebt.

V.2. Performance du modèle MLR pour la prédiction de la dose optimale de coagulant

Les résultats obtenus durant les phases d'apprentissage et de validation sont présentés dans le Tableau V.1. Globalement, les performances du modèle MLR varient selon le nombre et la nature des variables d'entrée utilisées. Une diminution progressive des performances est observée à mesure que le nombre de variables explicatives est réduit, ce qui met en évidence l'importance de l'information apportée par chaque paramètre physicochimique dans l'estimation de la dose de coagulant.

Tableau V.1. Performances du modèle MLR pour la prédiction de la dose optimale de coagulant.

Models	**Training**				**Validation**			
	R	NSE	RMSE (mg/L)	MAE (mg/L)	R	NSE	RMSE (mg/L)	MAE (mg/L)
MLR1	0.715	0.511	4.521	3.630	0.713	0.507	4.510	3.576
MLR2	0.715	0.511	4.522	3.634	0.712	0.506	4.514	3.584
MLR3	0.714	0.510	4.523	3.639	0.711	0.504	4.520	3.587
MLR4	0.709	0.503	4.556	3.657	0.709	0.500	4.541	3.610
MLR5	0.624	0.389	5.053	4.048	0.595	0.349	5.180	4.062
MLR6	0.556	0.310	5.370	4.142	0.555	0.299	5.373	4.070

V.2.1. Analyse du scénario MLR1

Le scénario MLR1, qui utilise l'ensemble des variables d'entrée, présente les meilleures performances parmi tous les scénarios étudiés. Durant la phase d'apprentissage, le modèle atteint un coefficient de corrélation de R = 0.715 et un coefficient NSE = 0.511, avec des valeurs RMSE et MAE respectivement égales à 4.521 mg/L et 3.630 mg/L. Des performances très proches sont observées en phase de validation avec R = 0.713, NSE = 0.507, RMSE = 4.510 mg/L et MAE = 3.576 mg/L.

Ces résultats indiquent que le modèle possède une capacité relativement satisfaisante à reproduire les variations de la dose de coagulant. La faible différence entre les phases d'apprentissage et de validation témoigne d'une bonne capacité de généralisation.

Les bonnes performances obtenues par MLR1 peuvent être attribuées à l'utilisation simultanée de plusieurs paramètres influençant directement le processus de coagulation-floculation. La turbidité (TU) et la couleur (COU) sont associées à la présence de particules en suspension et de matières organiques, tandis que l'absorbance UV_{254} constitue un indicateur pertinent de la matière organique dissoute. Le pH influence directement l'efficacité des réactions de coagulation, alors que les solides en suspension (SC), l'oxygène dissous (DO) et la température de l'eau (Tw) contribuent à caractériser l'état physicochimique global de l'eau brute.

V.2.2. Effet de la suppression progressive des variables

Les scénarios MLR2 et MLR3 présentent des performances très proches de celles obtenues avec MLR1. La suppression du pH dans le scénario MLR2 entraîne seulement une légère

diminution des indicateurs statistiques. En phase de validation, le coefficient R passe de 0.713 à 0.712 et le NSE de 0.507 à 0.506. Cette faible variation suggère que l'information apportée par le pH est partiellement compensée par les autres variables explicatives.

Dans le scénario MLR3, la suppression supplémentaire de la turbidité entraîne une baisse progressive des performances avec R = 0.711 et NSE = 0.504 en validation. Cette évolution confirme l'importance de la turbidité dans la détermination de la dose optimale de coagulant, puisqu'elle reflète directement la concentration des particules colloïdales à éliminer au cours du traitement.

Le scénario MLR4, basé uniquement sur les variables SC, UV_{254}, DO et Tw, présente une diminution plus marquée des performances. Le coefficient NSE diminue à 0.500 en validation tandis que le RMSE augmente à 4.541 mg/L. Cette tendance montre que la suppression de la couleur réduit la capacité du modèle à décrire correctement les variations de la dose de coagulant. La couleur étant souvent liée à la présence de matières organiques naturelles, son influence sur la consommation en coagulant demeure importante.

V.2.3. Dégradation des performances dans les scénarios MLR5 et MLR6

Les scénarios MLR5 et MLR6 enregistrent les performances les plus faibles parmi l'ensemble des configurations étudiées. Dans le scénario MLR5, où seules les variables SC, UV_{254} et DO sont conservées, le coefficient de corrélation diminue à R = 0.595 en validation et le coefficient NSE à 0.349. Parallèlement, les erreurs augmentent avec RMSE = 5.180 mg/L et MAE = 4.062 mg/L.

La dégradation devient encore plus prononcée dans le scénario MLR6 utilisant uniquement les variables SC et UV_{254}. Le coefficient de corrélation atteint seulement R = 0.555 et le coefficient NSE chute à 0.299, tandis que le RMSE et le MAE atteignent respectivement 5.373 mg/L et 4.070 mg/L.

La comparaison entre les scénarios MLR1 et MLR6 met clairement en évidence l'impact de la réduction des variables d'entrée. Le coefficient NSE passe de 0.507 à 0.299 tandis que le RMSE augmente de 4.510 à 5.373 mg/L. Ces résultats montrent que l'utilisation d'un nombre limité de variables ne permet pas de représenter correctement la complexité du processus de coagulation, qui dépend simultanément de plusieurs facteurs physiques, chimiques et organiques interconnectés.

V.2.4. Analyse des graphiques en nuage de points (Scatter Plots)

La Figure V.1 présente les diagrammes de dispersion des doses de coagulant observées et prédites obtenues à partir des six configurations du modèle MLR. Les scénarios MLR1 à MLR6 correspondent aux différentes combinaisons de variables d'entrée considérées dans cette étude.

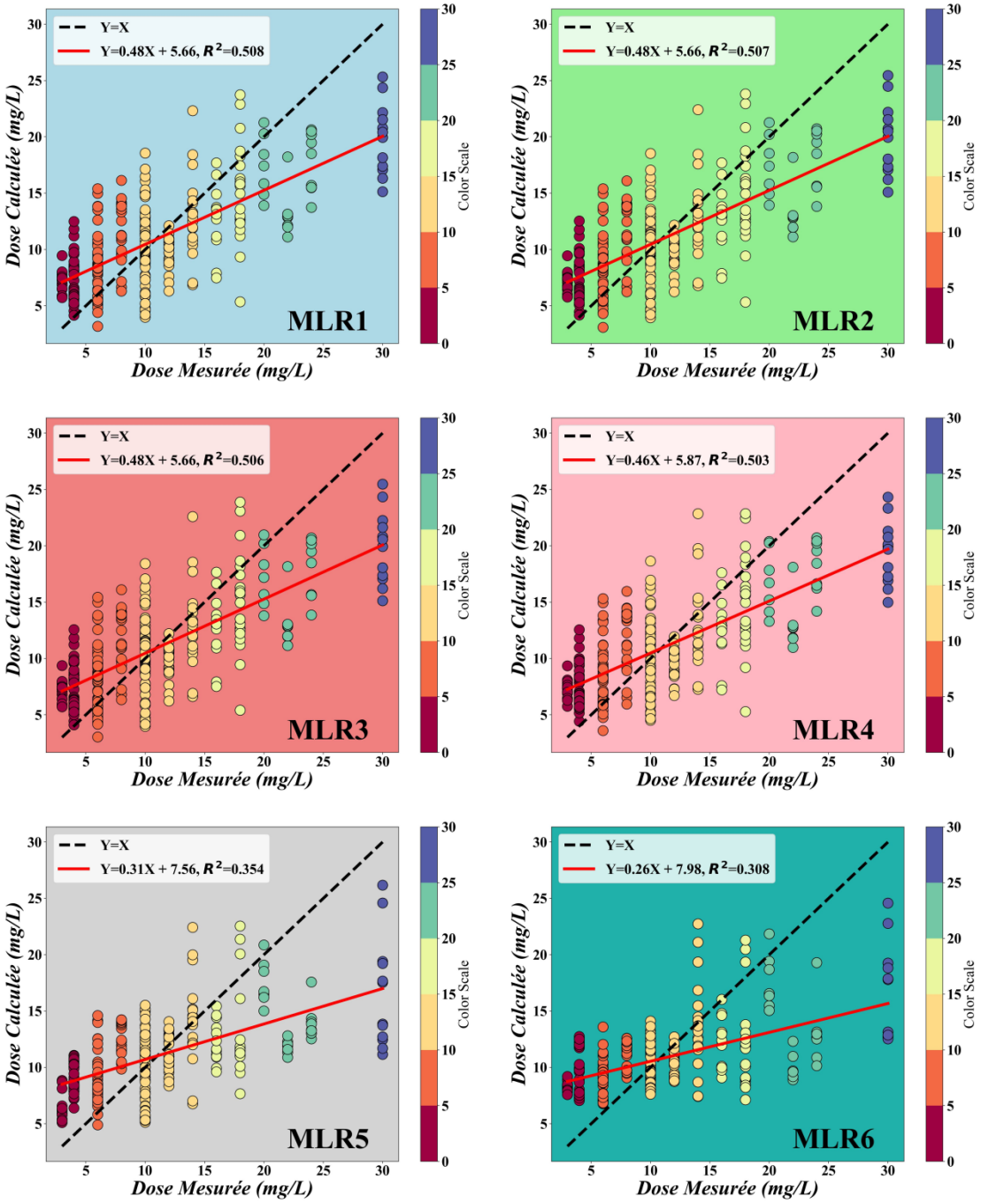


Figure V.1. Scatterplot des doses de coagulant observées et prédites à l'aide des modèles MLR1–MLR6 durant la phase de validation.

La ligne noire discontinue (Y = X) représente la situation idéale dans laquelle les valeurs prédites sont parfaitement égales aux valeurs observées. La ligne rouge correspond à la droite de régression entre les doses observées et calculées. Globalement, les graphiques montrent une dispersion relativement importante des points autour de la ligne idéale, traduisant les limites du modèle MLR dans la reproduction précise de la dose optimale de coagulant.

Les coefficients de détermination (R^2) obtenus varient entre 0.308 et 0.508. Le modèle MLR1 présente les meilleures performances avec $R^2 = 0.508$, tandis que le modèle MLR6 enregistre la valeur la plus faible avec $R^2 = 0.308$. Les nuages de points montrent également une dispersion croissante autour de la ligne idéale lorsque le nombre de variables d'entrée diminue, traduisant une perte progressive de précision des modèles.

Par ailleurs, les figures révèlent une tendance à la sous-estimation des fortes doses de coagulant, comme l'indique l'éloignement progressif des points par rapport à la ligne idéale pour les valeurs élevées. Ce comportement met en évidence les limites de la régression linéaire multiple pour représenter les relations complexes et non linéaires existant entre les paramètres de qualité de l'eau et la dose optimale de coagulant.

V.2.5. Discussion des limitations du modèle MLR

Bien que le modèle MLR fournisse des résultats relativement satisfaisants, plusieurs limitations doivent être soulignées. La régression linéaire multiple repose sur l'hypothèse d'une relation linéaire entre les variables explicatives et la variable cible. Or, le processus de coagulation-floculation est gouverné par des mécanismes physicochimiques complexes impliquant des interactions non linéaires entre les différents paramètres de qualité de l'eau.

De plus, certains paramètres peuvent interagir entre eux de manière difficilement représentable par un modèle strictement linéaire. Par exemple, l'effet du pH sur l'efficacité de coagulation dépend souvent simultanément de la turbidité, de la matière organique et de la température de l'eau. Ces interactions complexes peuvent expliquer les valeurs relativement élevées des erreurs RMSE et MAE observées dans les différents scénarios.

Par conséquent, bien que le modèle MLR constitue une première approche pertinente pour l'estimation de la dose optimale de coagulant, il est raisonnable de s'attendre à ce que les modèles non linéaires d'apprentissage automatique, tels que le réseau de neurones multicouche (MLPNN) et la forêt aléatoire (RFR), offrent des performances supérieures grâce à leur capacité à capturer des relations complexes entre les variables d'entrée et la variable cible.

V.3. Performance du modèle RFR pour la prédiction de la dose optimale de coagulant

Les résultats obtenus durant les phases d'apprentissage et de validation sont présentés dans le Tableau V.2. Globalement, le modèle RFR présente des performances nettement supérieures à celles obtenues avec le modèle MLR dans l'ensemble des scénarios étudiés. Les valeurs élevées de R et NSE, associées à des erreurs RMSE et MAE relativement faibles, confirment la capacité du modèle à reproduire avec précision les variations de la dose optimale de coagulant.

Tableau V.2. Performances du modèle RFR pour la prédiction de la dose optimale de coagulant.

Models	**Training**				**Validation**			
	R	NSE	RMSE (mg/L)	MAE (mg/L)	R	NSE	RMSE (mg/L)	MAE (mg/L)
RFR1	0.980	0.958	1.327	0.82	0.915	0.835	2.604	1.738
RFR2	0.980	0.956	1.349	0.839	0.915	0.835	2.612	1.754
RFR3	0.980	0.957	1.336	0.813	0.911	0.829	2.655	1.741
RFR4	0.977	0.953	1.399	0.851	0.906	0.820	2.725	1.796
RFR5	0.964	0.925	1.765	1.103	0.857	0.734	3.313	2.200
RFR6	0.925	0.850	2.500	1.710	0.762	0.574	4.191	3.012

V.3.1. Analyse du scénario RFR1

Le scénario RFR1, utilisant l'ensemble des variables d'entrée, présente les meilleures performances parmi tous les scénarios étudiés. Durant la phase d'apprentissage, le modèle atteint des valeurs très élevées avec R = 0.980 et NSE = 0.958. Les erreurs associées sont particulièrement faibles avec RMSE = 1.327 mg/L et MAE = 0.820 mg/L.

En phase de validation, le modèle conserve d'excellentes performances avec R = 0.915 et NSE = 0.835. Les erreurs demeurent relativement faibles avec RMSE = 2.604 mg/L et MAE = 1.738 mg/L. Ces résultats démontrent une forte capacité de généralisation du modèle malgré une légère diminution des performances entre l'apprentissage et la validation.

Les excellentes performances obtenues par RFR1 montrent que l'algorithme Random Forest est capable de représenter efficacement les interactions complexes entre les paramètres physicochimiques et la dose de coagulant. Contrairement à la régression linéaire multiple,

cette approche ne repose pas sur l'hypothèse de linéarité et permet ainsi de capturer les effets non linéaires ainsi que les interactions multiples entre les variables explicatives.

V.3.2. Influence de la réduction progressive des variables d'entrée

Les scénarios RFR2 et RFR3 présentent des performances très proches de celles observées pour RFR1. Dans le scénario RFR2, la suppression du pH entraîne une variation presque négligeable des performances. En validation, les indicateurs restent pratiquement inchangés avec R = 0.915 et NSE = 0.835. Cette stabilité montre que l'algorithme Random Forest est capable de compenser partiellement l'absence de certaines variables grâce aux informations fournies par les autres paramètres.

Le scénario RFR3 présente une légère diminution des performances avec R = 0.911 et NSE = 0.829 en validation. Bien que cette baisse reste limitée, elle confirme l'importance de la turbidité dans l'estimation de la dose optimale de coagulant, ce paramètre étant directement lié à la charge particulaire de l'eau brute.

Dans le scénario RFR4, utilisant uniquement les variables SC, UV_{254}, DO et Tw, les performances diminuent progressivement avec R = 0.906 et NSE = 0.820 en validation. Les erreurs augmentent légèrement pour atteindre RMSE = 2.725 mg/L et MAE = 1.796 mg/L. Cette évolution suggère que la suppression de la couleur réduit partiellement la capacité du modèle à représenter les variations associées à la matière organique présente dans l'eau.

V.3.3. Dégradation des performances dans les scénarios RFR5 et RFR6

Les scénarios RFR5 et RFR6 enregistrent les performances les plus faibles parmi les modèles RFR en raison de la réduction importante du nombre de variables d'entrée.

Dans le scénario RFR5, basé uniquement sur les variables SC, UV_{254} et DO, le coefficient de corrélation atteint R = 0.857 en validation tandis que le coefficient NSE diminue à 0.734. Les erreurs augmentent à RMSE = 3.313 mg/L et MAE = 2.200 mg/L. Malgré cette diminution, les performances demeurent relativement satisfaisantes.

La dégradation devient plus marquée dans le scénario RFR6 utilisant uniquement les variables SC et UV_{254}. En validation, le coefficient de corrélation diminue à R = 0.762 et le coefficient

NSE à 0.574. Les erreurs RMSE et MAE augmentent respectivement à 4.191 mg/L et 3.012 mg/L.

La comparaison entre les scénarios RFR1 et RFR6 met clairement en évidence l'importance des variables explicatives. Le coefficient NSE diminue de 0.835 à 0.574 tandis que le RMSE augmente de 2.604 à 4.191 mg/L. Ces résultats montrent que, même si Random Forest possède une forte capacité de modélisation non linéaire, la disponibilité d'informations suffisantes demeure essentielle pour garantir des prédictions précises.

V.3.4. Analyse des graphiques en nuage de points (Scatter Plots)

La Figure V.2 présente les diagrammes de dispersion des doses de coagulant observées et prédites obtenues à partir des six configurations du modèle Random Forest Regression (RFR). Les scénarios RFR1 à RFR6 correspondent aux différentes combinaisons de variables d'entrée considérées dans cette étude.

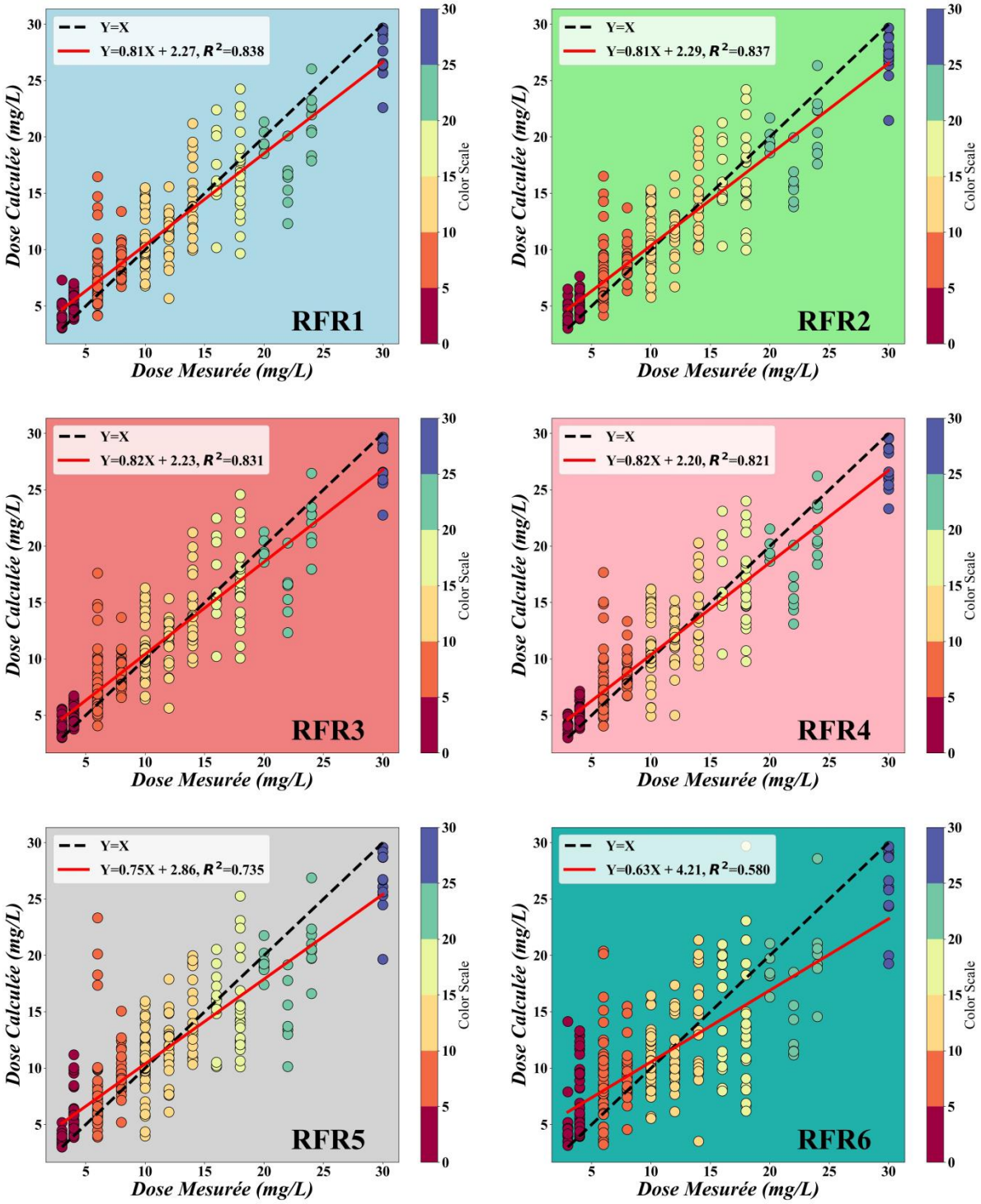


Figure V.2. Scatterplot des doses de coagulant observées et prédites à l'aide des modèles RFR1–RFR6 durant la phase de validation.

Comparativement aux modèles MLR, les modèles RFR présentent une dispersion nettement plus faible des points autour de la ligne idéale, traduisant une amélioration significative de la précision des prédictions.

Les coefficients de détermination (R^2) obtenus varient entre 0.580 et 0.838. Le modèle RFR1 présente la meilleure performance avec $R^2 = 0.838$, indiquant une forte capacité à reproduire les variations de la dose de coagulant. Les modèles RFR2, RFR3 et RFR4 conservent également des performances élevées avec des valeurs de R^2 supérieures à 0.82, ce qui démontre la robustesse du modèle malgré la réduction progressive du nombre de variables.

En revanche, une diminution plus marquée des performances est observée pour les modèles RFR5 et RFR6, dont les coefficients de détermination atteignent respectivement 0.735 et 0.580. Les nuages de points montrent une dispersion croissante autour de la ligne idéale à mesure que le nombre de variables d'entrée diminue, traduisant une perte progressive de précision.

Globalement, les résultats obtenus confirment que la méthode Random Forest fournit des performances largement supérieures à celles de la régression linéaire multiple. La faible dispersion des points et les valeurs élevées de R^2 démontrent la capacité de cette approche à représenter efficacement les relations non linéaires et les interactions complexes qui caractérisent le processus de coagulation.

V.3.5. Discussion des capacités et limitations du modèle RFR

Le modèle Random Forest présente plusieurs avantages majeurs dans le contexte de la modélisation des procédés de traitement des eaux. Il possède une excellente capacité à représenter les relations non linéaires, une forte robustesse vis-à-vis du bruit et des valeurs aberrantes, ainsi qu'une bonne aptitude à gérer les interactions complexes entre les variables explicatives. De plus, le principe d'agrégation de plusieurs arbres de décision contribue à réduire le risque de surapprentissage et à améliorer la stabilité des prédictions.

Toutefois, certaines limitations peuvent être mentionnées. Le modèle nécessite généralement un temps de calcul plus important que les approches statistiques classiques. Par ailleurs, l'interprétation physique des relations entre les variables reste plus difficile en raison du caractère complexe de l'algorithme et du grand nombre de décisions internes impliquées dans le processus de prédiction.

Malgré ces limites, les résultats obtenus montrent clairement que le modèle Random Forest constitue une approche particulièrement performante pour l'estimation de la dose optimale de coagulant au niveau de la station de traitement des eaux de Taksebt. Les performances élevées observées confirment le potentiel des techniques d'intelligence artificielle avancées pour l'optimisation des procédés de traitement des eaux.

V.4. Performance du modèle MLPNN pour la prédiction de la dose optimale de coagulant

Le modèle MLPNN a été développé afin d'estimer la dose optimale de coagulant au niveau de la station de traitement des eaux de Taksebt à partir des paramètres physicochimiques de la qualité de l'eau. Comme pour les modèles MLR et RFR, six scénarios d'entrée ont été considérés afin d'évaluer l'influence progressive des variables explicatives sur les performances prédictives du modèle.

Les résultats obtenus durant les phases d'apprentissage et de validation (Tableau V.3) montrent que le modèle MLPNN présente globalement les meilleures performances parmi tous les modèles étudiés. Les valeurs très élevées de R et NSE ainsi que les faibles valeurs de RMSE et MAE confirment la grande capacité du réseau neuronal à reproduire les variations de la dose optimale de coagulant.

Tableau V.3. Performances du modèle MLPNN pour la prédiction de la dose optimale de coagulant.

Models	**Training**				**Validation**			
	R	NSE	RMSE (mg/L)	MAE (mg/L)	R	NSE	RMSE (mg/L)	MAE (mg/L)
MLPNN1	0.999	0.999	0.208	0.161	0.922	0.849	2.498	1.702
MLPNN2	0.999	0.999	0.229	0.180	0.918	0.842	2.550	1.752
MLPNN3	0.999	0.998	0.302	0.238	0.920	0.845	2.526	1.742
MLPNN4	0.998	0.996	0.397	0.303	0.913	0.834	2.613	1.754
MLPNN5	0.999	0.999	0.100	0.077	0.845	0.708	3.471	2.380
MLPNN6	0.971	0.939	1.593	1.078	0.756	0.559	4.263	3.060

V.4.1. Analyse du scénario MLPNN1

Le scénario MLPNN1, utilisant l'ensemble des variables disponibles (SC, UV_{254}, DO, Tw, COU, TU et pH), présente les meilleures performances globales de toute l'étude. Durant la

phase d'apprentissage, le modèle atteint des résultats exceptionnels avec R = 0.999 et NSE = 0.999. Les erreurs de prédiction sont extrêmement faibles avec RMSE = 0.208 mg/L et MAE = 0.161 mg/L.

En phase de validation, le modèle conserve une excellente capacité prédictive avec R = 0.922 et NSE = 0.849. Les erreurs restent relativement faibles avec RMSE = 2.498 mg/L et MAE = 1.702 mg/L.

Ces résultats démontrent la forte capacité du réseau neuronal à apprendre les relations complexes existant entre les paramètres physicochimiques de l'eau et la dose optimale de coagulant. Les performances élevées observées en validation confirment également une bonne capacité de généralisation du modèle.

La supériorité du scénario MLPNN1 s'explique par l'utilisation simultanée de l'ensemble des variables influençant le processus de coagulation-floculation. Les paramètres TU, COU et UV254 constituent des indicateurs importants de la matière particulaire et organique présente dans l'eau brute, tandis que le pH, la température, la conductivité et l'oxygène dissous influencent directement les mécanismes physicochimiques de coagulation.

V.4.2. Influence de la réduction progressive des variables d'entrée

Les scénarios MLPNN2 et MLPNN3 présentent des performances très proches de celles obtenues avec MLPNN1. La suppression du pH dans MLPNN2 entraîne une légère diminution des performances avec R = 0.918 et NSE = 0.842 en validation. Les erreurs augmentent faiblement avec RMSE = 2.550 mg/L et MAE = 1.752 mg/L.

Dans le scénario MLPNN3, la suppression supplémentaire de la turbidité entraîne des performances très similaires avec R = 0.920 et NSE = 0.845. Les valeurs RMSE et MAE demeurent faibles, atteignant respectivement 2.526 mg/L et 1.742 mg/L. Cette stabilité met en évidence la robustesse du modèle MLPNN face à la réduction du nombre de variables d'entrée.

Le scénario MLPNN4, utilisant uniquement SC, UV_{254}, DO et Tw, présente une légère baisse des performances avec R = 0.913 et NSE = 0.834 en validation. Les erreurs augmentent modérément avec RMSE = 2.613 mg/L et MAE = 1.754 mg/L. Malgré cette diminution, les

performances restent très satisfaisantes et demeurent supérieures à celles obtenues avec le modèle MLR.

V.4.3. Dégradation des performances dans les scénarios MLPNN5 et MLPNN6

Les scénarios MLPNN5 et MLPNN6 enregistrent une diminution plus marquée des performances suite à la réduction importante du nombre de variables explicatives.

Le scénario MLPNN5, basé uniquement sur les variables SC, UV_{254} et DO, atteint en validation un coefficient de corrélation de R = 0.845 et un coefficient NSE de 0.708. Les erreurs augmentent à RMSE = 3.471 mg/L et MAE = 2.380 mg/L.

Le scénario MLPNN6, utilisant uniquement SC et UV_{254}, présente les performances les plus faibles parmi les modèles MLPNN. En validation, le coefficient R diminue à 0.756 tandis que le NSE atteint 0.559. Les erreurs augmentent à RMSE = 4.263 mg/L et MAE = 3.060 mg/L.

Ces résultats montrent que la réduction excessive du nombre de variables limite la capacité du réseau neuronal à représenter correctement la complexité du processus de coagulation. Bien que le modèle conserve des performances acceptables, l'absence de certaines variables importantes réduit significativement la précision des prédictions.

V.4.4. Analyse des graphiques en nuage de points(Scatter Plots)

La Figure V.3 présente les diagrammes de dispersion des doses de coagulant observées et prédites obtenues à partir des six configurations du modèle MLPNN. Les scénarios MLPNN1 à MLPNN6 correspondent aux différentes combinaisons de variables d'entrée considérées dans cette étude.

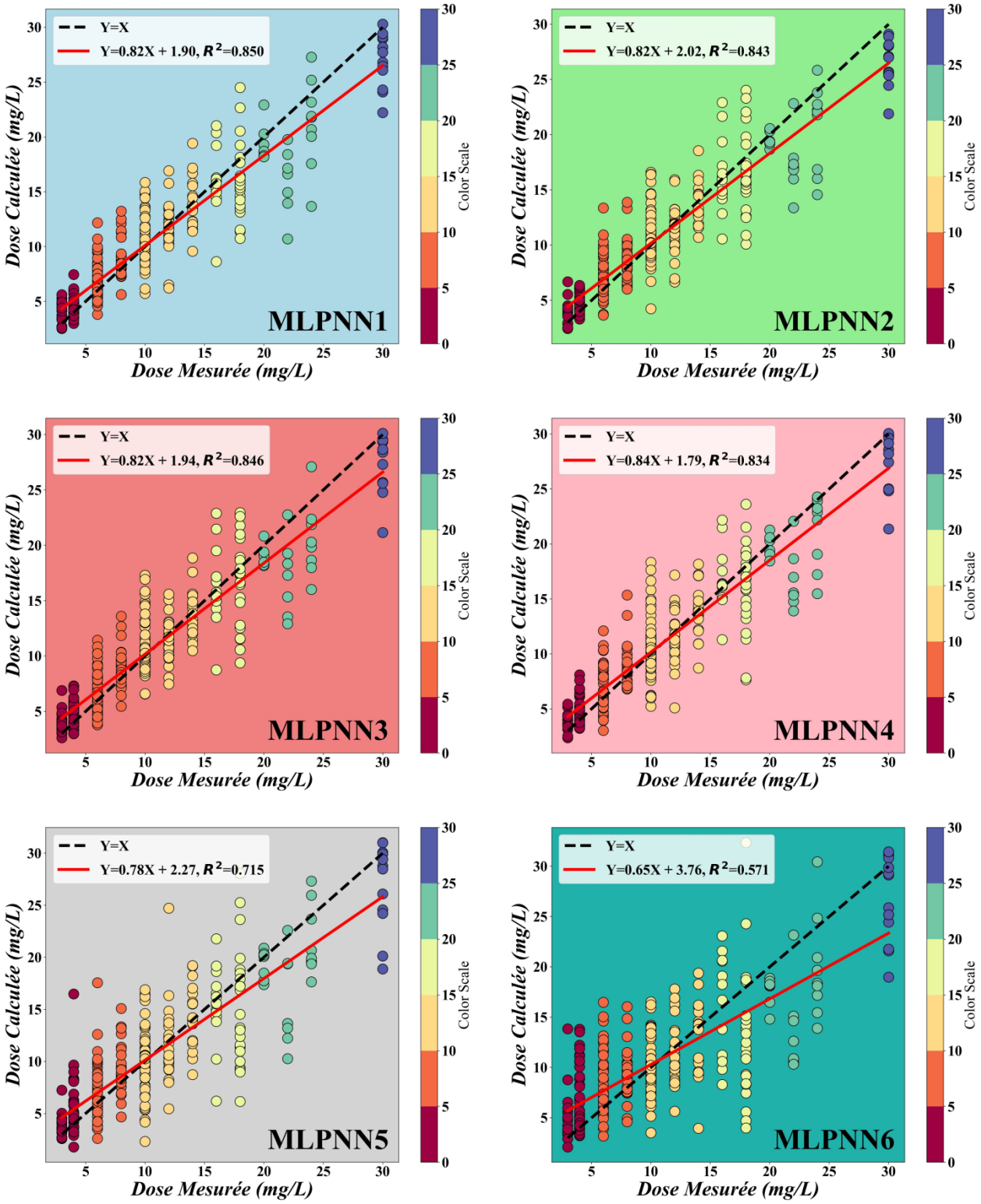


Figure V.3. Scatterplot des doses de coagulant observées et prédites à l'aide des modèles MLPNN1–MLPNN6 durant la phase de validation.

Les résultats montrent une forte concentration des points autour de la ligne idéale pour les quatre premiers scénarios, traduisant une excellente capacité du réseau neuronal à reproduire

les variations de la dose de coagulant. Les coefficients de détermination (R^2) varient entre 0.571 et 0.850. Le modèle MLPNN1 présente la meilleure performance avec $R^2 = 0.850$, suivi par MLPNN2, MLPNN3 et MLPNN4 dont les valeurs de R^2 restent supérieures à 0.83.

Une diminution progressive des performances est observée pour les modèles MLPNN5 et MLPNN6, dont les coefficients de détermination atteignent respectivement 0.715 et 0.571. Les nuages de points montrent une dispersion croissante autour de la ligne idéale lorsque le nombre de variables d'entrée diminue, indiquant une perte progressive de précision.

Globalement, les diagrammes de dispersion confirment les résultats statistiques obtenus. Le modèle MLPNN présente les meilleures performances de l'étude et reproduit avec une grande précision les variations de la dose optimale de coagulant.

V.4.5. Discussion des capacités et limitations du modèle MLPNN

Le modèle MLPNN présente plusieurs avantages majeurs dans le contexte de la prédiction de la dose optimale de coagulant:

- Excellente capacité d'apprentissage des relations non linéaires.
- Très haute précision prédictive.
- Capacité à modéliser des interactions complexes entre les variables.
- Bonne capacité de généralisation lorsque les données sont suffisantes.

Cependant, certaines limitations doivent également être mentionnées. Les réseaux neuronaux nécessitent généralement un temps d'apprentissage plus important ainsi qu'un réglage adéquat de plusieurs hyperparamètres tels que le nombre de neurones, le nombre de couches cachées et les fonctions d'activation.

Par ailleurs, les performances extrêmement élevées observées durant la phase d'apprentissage peuvent parfois traduire un risque de surapprentissage. Néanmoins, dans cette étude, les performances obtenues en validation restent suffisamment élevées pour confirmer la robustesse et la fiabilité du modèle.

Enfin, contrairement au modèle MLR, l'interprétation physique des relations internes du réseau neuronal demeure relativement difficile en raison de son caractère « boîte noire ».

Malgré ces limitations, les résultats obtenus montrent clairement que le modèle MLPNN constitue l'approche la plus performante pour la prédiction de la dose optimale de coagulant au niveau de la station de traitement des eaux de Taksebt. Les performances élevées observées confirment le potentiel considérable des réseaux de neurones artificiels dans l'optimisation des procédés de traitement des eaux.

V.5. Comparaison des performances des modèles MLR, RFR et MLPNN

Afin d'identifier l'approche la plus performante pour la prédiction de la dose optimale de coagulant à la station de traitement des eaux de Taksebt, une comparaison a été réalisée entre les trois modèles développés dans cette étude : la Régression Linéaire Multiple (MLR), la Régression par Forêt Aléatoire (RFR) et le Réseau de Neurones à Perceptron Multicouche (MLPNN). Cette comparaison repose sur les indicateurs statistiques R, NSE, RMSE et MAE obtenus lors de la phase de validation.

Les résultats montrent clairement une hiérarchie des performances entre les trois approches. Le modèle MLPNN présente les meilleurs résultats globaux, suivi du modèle RFR, tandis que le modèle MLR affiche les performances les plus faibles. Cette tendance est observée dans l'ensemble des scénarios étudiés et devient particulièrement évidente lorsque toutes les variables d'entrée sont utilisées.

Pour le scénario complet utilisant les sept paramètres physicochimiques (SC, UV_{254}, DO, Tw, COU, TU et pH), le modèle MLR1 atteint un coefficient de corrélation de R = 0.713 et un coefficient NSE de 0.507, avec RMSE = 4.510 mg/L et MAE = 3.576 mg/L. En comparaison, le modèle RFR1 améliore considérablement la qualité des prédictions avec R = 0.915 et NSE = 0.835, tandis que les erreurs diminuent à RMSE = 2.604 mg/L et MAE = 1.738 mg/L. Le modèle MLPNN1 fournit les meilleures performances avec R = 0.922 et NSE = 0.849, associés à RMSE = 2.498 mg/L et MAE = 1.702 mg/L.

Cette amélioration significative des modèles RFR et MLPNN par rapport au modèle MLR montre que la relation entre les paramètres de qualité de l'eau et la dose optimale de coagulant n'est pas purement linéaire. Le procédé de coagulation-floculation dépend d'interactions physicochimiques complexes impliquant simultanément la turbidité, la couleur, la matière organique dissoute, le pH, la température et d'autres caractéristiques de l'eau brute. Les

modèles d'intelligence artificielle sont capables de représenter ces relations non linéaires avec une précision nettement supérieure à celle de la régression linéaire multiple.

L'analyse des différents scénarios met également en évidence une diminution progressive des performances lorsque le nombre de variables d'entrée est réduit. Toutefois, cette dégradation reste beaucoup plus limitée pour les modèles RFR et MLPNN que pour le modèle MLR. Par exemple, dans le scénario le plus réduit (deux variables d'entrée), les modèles RFR6 et MLPNN6 conservent des coefficients NSE respectifs de 0.574 et 0.559, alors que le modèle MLR6 n'atteint qu'un NSE de 0.299. Cette observation confirme la capacité des modèles d'apprentissage automatique à extraire davantage d'informations à partir d'un nombre limité de variables.

Les dscatterplots confirment également cette tendance. Les modèles MLPNN et RFR présentent une concentration plus importante des points autour de la ligne idéale (Y = X), traduisant une meilleure concordance entre les valeurs observées et prédites. À l'inverse, les modèles MLR montrent une dispersion plus importante ainsi qu'une tendance à sous-estimer les fortes doses de coagulant.

Entre les deux modèles d'intelligence artificielle, le réseau de neurones MLPNN présente globalement les meilleures performances. Cette supériorité peut être attribuée à sa grande capacité d'apprentissage et à son aptitude à approximer des fonctions non linéaires très complexes. Toutefois, les écarts entre MLPNN et RFR demeurent relativement faibles pour la majorité des scénarios. Le modèle RFR présente quant à lui l'avantage d'être plus robuste vis-à-vis du bruit et généralement plus facile à paramétrer.

Dans l'ensemble, les résultats obtenus démontrent que les techniques d'intelligence artificielle constituent des outils particulièrement efficaces pour la prédiction de la dose optimale de coagulant. Le modèle MLPNN apparaît comme l'approche la plus performante dans le contexte de cette étude, suivi de très près par le modèle RFR. La régression linéaire multiple, bien qu'elle fournisse des résultats acceptables et facilement interprétables, demeure limitée par son incapacité à représenter pleinement les relations non linéaires complexes caractérisant les procédés de traitement des eaux.

V.6. Conclusion

Dans ce chapitre, les performances de trois approches de modélisation, à savoir la Régression Linéaire Multiple (MLR), la Régression par Forêt Aléatoire (RFR) et le Réseau de Neurones à Perceptron Multicouche (MLPNN), ont été évaluées pour la prédiction de la dose optimale de coagulant à la station de traitement des eaux de Taksebt à partir des paramètres physicochimiques de la qualité de l'eau (SC, UV_{254}, DO, Tw, COU, TU et pH).

La comparaison des modèles a révélé une nette supériorité des approches d'intelligence artificielle par rapport à la régression linéaire multiple. Le modèle MLR a fourni des performances acceptables, mais limitées par son caractère linéaire. À l'inverse, les modèles RFR et MLPNN ont montré une meilleure capacité à représenter les relations complexes et non linéaires entre les paramètres de qualité de l'eau et la dose optimale de coagulant.

Parmi les modèles étudiés, leMLPNN a obtenu les meilleures performances globales, avec les valeurs les plus élevées de R et NSE ainsi que les plus faibles valeurs de RMSE et MAE. Le modèle RFR a également présenté des résultats très satisfaisants et proches de ceux du MLPNN, confirmant l'efficacité des méthodes d'apprentissage automatique pour la modélisation des procédés de traitement des eaux. Les analyses graphiques basées sur les Scatter Plots ont confirmé les résultats statistiques en montrant une meilleure concordance entre les valeurs observées et prédites pour les modèles MLPNN et RFR comparativement au modèle MLR.

Ainsi, les résultats obtenus démontrent que les techniques d'intelligence artificielle, et en particulier le modèle MLPNN, constituent des outils performants et fiables pour la prédiction de la dose optimale de coagulant. Leur utilisation peut contribuer à améliorer la gestion et l'optimisation des procédés de coagulation-floculation, tout en réduisant le recours aux essais expérimentaux traditionnels tels que les Jar-Tests.

Conclusion Générale

Conclusion Générale

L'objectif principal de cette étude était de développer des modèles de prédiction capables d'estimer la dose optimale de coagulant à partir des paramètres physicochimiques de l'eau brute de la station de traitement des eaux de Taksebt. Cette problématique revêt une importance particulière dans le domaine du traitement des eaux, puisque l'optimisation de la coagulation-floculation influence directement la qualité de l'eau produite, les coûts d'exploitation et l'efficacité globale du procédé.

L'analyse exploratoire des données a permis de mettre en évidence l'existence de relations significatives entre la dose de coagulant et les paramètres de qualité de l'eau étudiés. La méthode SelectKBest basée sur le critère Mutual Information a montré que certains paramètres, notamment SC, UV_{254}, DO et Tw, jouent un rôle déterminant dans la prédiction de la dose optimale de coagulant.

Trois approches de modélisation ont été développées et comparées: la Régression Linéaire Multiple (MLR), la Régression par Forêt Aléatoire (RFR) et le Réseau de Neurones à Perceptron Multicouche (MLPNN). Afin d'évaluer l'impact des variables d'entrée sur les performances prédictives, six scénarios différents ont été construits à partir de combinaisons progressives des paramètres physicochimiques disponibles.

Les résultats obtenus ont montré que les performances des modèles diminuent progressivement lorsque le nombre de variables d'entrée est réduit, ce qui confirme l'importance d'une représentation complète des caractéristiques de l'eau brute. Le scénario utilisant l'ensemble des variables disponibles a systématiquement fourni les meilleurs résultats pour les trois approches étudiées.

La comparaison des modèles a mis en évidence la supériorité des techniques d'intelligence artificielle par rapport à la méthode statistique classique MLR. Bien que la régression linéaire multiple ait fourni des résultats acceptables, ses performances sont restées limitées par l'hypothèse de linéarité entre les variables explicatives et la variable cible.

Le modèle Random Forest Regression a présenté des performances nettement supérieures grâce à sa capacité à représenter les interactions complexes et les relations non linéaires entre les paramètres physicochimiques de l'eau. Cette approche s'est révélée robuste et stable, même lorsque certaines variables d'entrée étaient supprimées.

Le MLPNN a obtenu les meilleures performances globales. Les valeurs élevées des coefficients de corrélation et d'efficacité ainsi que les faibles erreurs de prédiction démontrent

sa capacité à reproduire avec précision les variations de la dose optimale de coagulant. Ces résultats confirment que les phénomènes régissant le processus de coagulation-floculation sont fortement non linéaires et que les réseaux neuronaux constituent des outils particulièrement adaptés à leur modélisation.

Les Scatter Plots entre les valeurs observées et prédites ont confirmé les résultats statistiques en montrant une meilleure concordance pour les modèles MLPNN et RFR comparativement au modèle MLR. L'ensemble des résultats obtenus démontre ainsi le potentiel considérable des techniques d'intelligence artificielle pour l'optimisation des procédés de traitement des eaux.

En conclusion, cette étude a montré que les modèles d'apprentissage automatique peuvent constituer une alternative fiable et performante aux méthodes conventionnelles de détermination de la dose de coagulant. Leur utilisation pourrait contribuer à améliorer la prise de décision opérationnelle au sein des stations de traitement des eaux et à réduire le recours systématique aux essais expérimentaux.

Perspectives

Les résultats obtenus ouvrent plusieurs perspectives de recherche intéressantes :

- Tester d'autres algorithmes avancés d'apprentissage automatique tels que XGBoost, LightGBM, ou les réseaux neuronaux profonds (Deep Learning).
- Intégrer des paramètres hydrologiques et météorologiques supplémentaires afin d'améliorer davantage les performances prédictives.
- Développer des systèmes de prédiction en temps réel couplés aux capteurs de surveillance de la qualité de l'eau.
- Étendre l'application des modèles à d'autres stations de traitement afin d'évaluer leur capacité de généralisation dans différents contextes environnementaux.
- Développer des plateformes d'aide à la décision permettant aux exploitants de déterminer instantanément la dose optimale de coagulant.
- Étudier l'utilisation de techniques hybrides combinant plusieurs modèles d'intelligence artificielle afin d'obtenir des performances encore plus élevées.

Ces perspectives constituent des pistes prometteuses pour le développement de systèmes intelligents destinés à améliorer durablement la gestion et l'optimisation des procédés de traitement des eaux potables.

Références bibliographiques

Références bibliographiques

Abdel Nabi Mohamed, S. (2019). *La pollution environnementale : une épidémie à l'ère de la mondialisation*. Égypte : Agence de presse arabe.

Achite, M., Farzin, S., Elshaboury, N., Valikhan Anaraki, M., Amamra, M., Toubal, A.K. (2024). Modeling the optimal dosage of coagulants in water treatment plants using various machine learning models. Environment, Development and Sustainability, 26(2), 3395-3421.https://doi.org/10.1007/s10668-022-02835-0

Agirre-Basurko, E., Ibarra-Berastegi, G., Madariaga, I. (2006). Regression and multilayer perceptron-based models to forecast hourly O3 and NO2 levels in the Bilbao area. Environmental Modelling and Software, 21(4), 430-446.https://doi.org/10.1016/j.envsoft.2004.07.008

Ahmad, M.W., Reynolds, J., Rezgui, Y. (2018). Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. Journal of Cleaner Production, 203, 810-821.https://doi.org/10.1016/j.jclepro.2018.08.207

Al Zarqa, M. A. N. A. (2018). *Destruction des polluants de l'eau dans les gouvernorats du Nord et du Centre et leurs effets sur la santé humaine* (Mémoire de master). Université Islamique de Gaza.

Asfari, A. F. (2004). *Directives pour la conception, l'exploitation et la maintenance du traitement des eaux usées*. Jordanie.

Bachi, O. E. K. (2010). *Diagnostic sur la valorisation de quelques plantes du gardien d'épuration de la station du Vieux Ksar de Témacine* (Mémoire de magister). Université Kasdi Merbah, Ouargla, Algérie.

Bouziani. (2000). *De la pénurie d'eau aux maladies*. Alger : Éditions Ibn Khaldoun.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.https://doi.org/10.1023/A:1010933404324

Chakraborty, A., Goswami, D. (2017). Prediction of slope stability using multiple linear regression (MLR) and artificial neural network (ANN). Arabian Journal of Geosciences, 10(17), 385.https://doi.org/10.1007/s12517-017-3167-x

Charabi, N. E., Minour, H., & Ziad, A. L. (1987). *Fondamentaux de microbiologie : Partie pratique*. Damas, Syrie : Al-Matloua Al-Jadida.

Degremont. (2005). *Mémento technique de l'eau* (10e éd.). Paris, France : Lavoisier.

Duan, J., & Gregory, J. (2003). Coagulation by hydrolysing metal salts. *Advances in colloid and interface science*, *100*, 475-502.https://doi.org/10.1016/S0001-8686(02)00067-2

Edzwald, J. K. (2011). *Water Quality & Treatment*. AWWA.

El Abed, I. (2010). *Traitement des eaux usées de la région de Touggourt par des plantes épuratrices locales* (Thèse de doctorat). Université Kasdi Merbah, Ouargla, Algérie.

El-Roubi Mohamed, M. M. (2014). *Le contrôle administratif et son rôle dans la protection de l'environnement*. p. 295.

Etemadi, S., Khashei, M. (2021). Etemadi multiple linear regression. Measurement, 186, 110080.https://doi.org/10.1016/j.measurement.2021.110080

Feng, L., Zhang, Y., Wei, X., Wang, M., Niu, Z., Wang, C. (2025). Interpretable prediction of coagulant dosage in drinking water treatment plant based on automated machine learning and SHAP method. Journal of Water Process Engineering, 75, 107925.https://doi.org/10.1016/j.jwpe.2025.107925

Fouchal, A., Merabet, K., Kisi, O., Di Nunno, F., Granata, F., Kim, S., ... & Yaseen, Z. M. (2025). Modeling Coagulant Dosage in Water Treatment Plant Using Explainable Deep Learning Model–Based SHapley Additive exPlanations (SHAP). *Water Conservation Science and Engineering*, *10*(3), 109. https://doi.org/10.1007/s41101-025-00434-4

Gaujous, D. (1999). *La pollution des milieux aquatiques : Aide-mémoire* (2e éd.). Paris, France : TEC & DOC – Lavoisier.

Haykin, S. (2009). *Neural Networks and Learning Machines* (3rd ed.).

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. Neural Netw 4(2):251–257. https://doi.org/10.1016/0893-6080(91)90009-T

Hornik, K., Stinchcombe, M., White, H. (1989). Multilayer feedforward networks are universal approximators. Neural Netw 2:359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Ismaïl, M. (2003). *Traitement des eaux*. Le Caire, Égypte : Dar El-Kotob El-Ilmiyah.

Jin, X., Du, J., Liu, H., Wang, Z., Song, K. (2016). Remote estimation of soil organic matter content in the Sanjiang Plain, Northest China: the optimal band algorithm versus the GRA-ANN model. Agric For Meteorol 218–219:250–260. https://doi.org/10.1016/j.agrformet.2015.12.062

Journal Officiel de la République Algérienne. (2012). *Arrêté interministériel du 2 janvier 2012 fixant les caractéristiques des eaux usées épurées utilisées à des fins d'irrigation*. Journal officiel de la République Algérienne, 41.

LeCun, Y. (1985). Une procédure d'apprentissage pour réseau à seuil asymétrique Cognitiva, Vol. 85, pp 599-604.

Maher, G. N. (2007). *Analyse et évaluation de la qualité de l'eau*. Égypte : Dar El Maârif.

Merabet, K., & Heddam, S. (2023). Improving the accuracy of air relative humidity prediction using hybrid machine learning based on empirical mode decomposition: a comparative study. *Environmental Science and Pollution Research*, *30*(21), 60868-60889. https://doi.org/10.1007/s11356-023-26779-8

Metcalf & Eddy, Inc. (2014). Wastewater engineering: Treatment and resource recovery (5th ed.). New York, NY: McGraw-Hill Education.

Metcalf & Eddy, Inc. (2014). *Wastewater Engineering: Treatment and Resource Recovery* (5th ed.). McGraw-Hill Education.

Nimon, K.F., Oswald, F.L. (2013). Understanding the results of multiple linear regression: Beyond standardized regression coefficients. Organizational Research Methods, 16(4), 650-674.https://doi.org/10.1177/1094428113493929

Organisation mondiale de la Santé. (2006). Directives pour l'utilisation sans risque des eaux usées, des excreta et des eaux ménagères. Genève, Suisse : OMS.

Power Hydraulique. (2020). *Note de calcul process de la station d'épuration du groupement urbain de Filfila, Wilaya de Skikda (DRE-SK-EX-PE-TR-002(01))*

Programme des Nations Unies pour l'Environnement. (2021). *Making peace with nature: A scientific blueprint to tackle the climate, biodiversity and pollution emergencies*. United Nations Environment Programme.

Ramade, F. (1992). *Éléments d'écologie appliquée*. Paris, France : McGraw-Hill.

Rodier, J. (2009). *L'analyse de l'eau* (9e éd.). Paris, France : Dunod.

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986b). Learning internal representations by back-propagating errors. Nature, 323:533-536.

Rumelhart, D.E., Hinton, G.E., Williams, R.J. (1986a).Learning internal representations by error propagation. In: Rumelhart DE, McClelland PDP, Research Group (eds) Parallel distributed processing: explorations in the microstructure of cognition. Foundations, vol I. Cambridge, MA, MIT Press, p 318–62

Secrétariat du Conseil Municipal de La Mecque. (2007). *Traitement des eaux usées à la station d'assainissement d'Al-Akashiya à La Mecque* (pp. 15–21). Centre Fakieh de Recherche et Développement.

Shams, S.R., Jahani, A., Kalantary, S., Moeinaddini, M., Khorasani, N. (2021). The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models for predicting SO2 concentration. Urban Climate, 37, 100837.https://doi.org/10.1016/j.uclim.2021.100837

Shanie, A. A. A. (2018). *Dégradation des polluants organiques de l'eau* (Mémoire de licence en chimie). Université Al-Qadisiyah, Irak.

Société Holding des Eaux Potables et de l'Assainissement. (2015). *Guide du stagiaire technicien d'exploitation de l'assainissement : Spécifications et caractéristiques des eaux usées*. Programme de formation professionnelle des travailleurs du secteur de l'eau potable et de l'assainissement, Égypte.

Tchobanoglous, G., Stensel, H. D., Tsuchihashi, R., Burton, F. L., Abu-Orf, M., Bowden, G., & Pfrang, W. (2014). Wastewater engineering: Treatment and resource recovery (5th ed.). New York, NY: McGraw-Hill Education.

Thomas, O. (1995). *Métrologie des eaux résiduaires*. Belgique : Cebedoc.

WHO. (2022). *Guidelines for Drinking-water Quality*. World Health Organization

Xu, S., Wang, S., Yang, Y., Zheng, L., & Xiao, F. (2025). Importance analysis of key parameters for coagulation dosage prediction based on machine learning models. *Process Safety and Environmental Protection, 205*, 108152. https://doi.org/10.1016/j.psep.2025.108152

Zhang, J., Tylor, N., Zhao, W., Rui, R., He, C. (2025). Enhancing coagulation prediction in water treatment using a similarity score–based piecewise machine learning model.

Journal of Environmental Engineering, 151(6), 04025023.https://doi.org/10.1061/JOEEDU.EEENG-7969

Zhou, Z., Qiu, C., Zhang, Y. (2023). A comparative analysis of linear regression, neural networks and random forest regression for predicting air ozone employing soft sensor models. Scientific Reports, 13(1), 22420. https://doi.org/10.1038/s41598-023-49899-0

الجمـــهوريـة الجزائـــريـة الديمقراطية الشعبيـــة.
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement
Supérieur et de la Recherche Scientifique
Université 20 Août 1955 - SKIKDA -
Faculté des Sciences
Dpartement D' Agronomie
Réf : / D S AGRO/2026

وزارة التعليم العالي و البحث العلمي
جامعة 20 أوت 1955- سكيكدة
كلية العلـوم
قسم علوم الفلاحة
رقم : / ق ع ف /2026

Skikda le : .....
30 JUIN 2026

جامعة 20 أوت 1955 سكيكدة
قسم العلوم الفلاحية
كلية العلوم

Autorisation de Dépôt de Mémoire de Master

Je soussigné : **Pr. Heddam Salim** Président de Jury

Certifie que :

L'étudiant(e) : **Medroua Bessma**

Spécialité : **Aménagement Hydro-Agricole**

Intitulé : **Modélisation de la Dose Optimale de Coagulant dans le Traitement des Eaux à l'Aide de l'Intelligence Artificielle.**

A apporté les corrections relatives à son travail de mémoire.

Encadrant

Président de Jury

الجمـــهوريـة الجزائـــريـة الديمقراطية الـشعبيـــة.

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement
Supérieur et de la Recherche Scientifique
Université 20 Août 1955 - SKIKDA -
Faculté des Sciences
Dpartement D' Agronomie
Réf : / D S AGRO/2026

وزارة التعليم العالي و البحث العلمي
جامعة 20 أوت 1955- سكيكدة
كلية العلـوم
قسم علوم الفلاحة
رقم : / ق ع ف /2026

Skikda le : ..... 30 JUIN 2026

Autorisation de Dépôt de Mémoire de Master

Je soussigné : **Pr. Heddam Salim** Président de Jury

Certifie que :

L'étudiant(e) : **Boubellouta Rima**

Spécialité : **Aménagement Hydro-Agricole**

Intitulé : **Modélisation de la Dose Optimale de Coagulant dans le Traitement des Eaux à l'Aide de l'Intelligence Artificielle.**

A apporté les corrections relatives à son travail de mémoire.

Encadrant

Président de Jury

الجمـــهوريـة الجزائـــريـة الديمقراطية الـشعبيـــة.
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement
Supérieur et de la Recherche Scientifique
Université 20 Août 1955 - SKIKDA -
Faculté des Sciences
Dpartement D' Agronomie
Réf : / D S AGRO/2026

وزارة التعليم العالي و البحث العلمي
جامعة 20 أوت 1955- سكيكدة
كلية العلـوم
قسم علوم الفلاحة
رقم : / ق ع ف /2026

Skikda le : .....
30 JUIN 2026

جامعة 20 أوت 1955 سكيكدة
قسم العلوم الفلاحية
كلية العلوم

Autorisation de Dépôt de Mémoire de Master

Je soussigné : **Pr. Heddam Salim** Président de Jury

Certifie que :

L'étudiant(e) : **Boukallel Ouissal**

Spécialité : **Aménagement Hydro-Agricole**

Intitulé : **Modélisation de la Dose Optimale de Coagulant dans le Traitement des Eaux à l'Aide de l'Intelligence Artificielle.**

A apporté les corrections relatives à son travail de mémoire.

Encadrant

Président de Jury

الجمـــهوريـة الجزائـــريـة الديمقراطية الـشعبيـــة.

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement
Supérieur et de la Recherche Scientifique
Université 20 Août 1955 - SKIKDA -
Faculté des Sciences
Dpartement D' Agronomie
Réf : / D S AGRO/2026

وزارة التعليم العالي و البحث العلمي
جامعة 20 أوت 1955- سكيكدة
كلية العلـوم
قسم علوم الفلاحة
رقم : / ق ع ف /2026

Skikda le : ……
30 JUIN 2026

Autorisation de Dépôt de Mémoire de Master

Je soussigné : **Pr. Heddam Salim** Président de Jury

Certifie que :

L'étudiant(e) : **Boulahia Ikram**

Spécialité : **Aménagement Hydro-Agricole**

Intitulé : **Modélisation de la Dose Optimale de Coagulant dans le Traitement des Eaux à l'Aide de l'Intelligence Artificielle.**

A apporté les corrections relatives à son travail de mémoire.

Encadrant

Président de Jury

الجمـــهورية الجزائـــرية الديمقراطية الشعبيـــة.
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement
Supérieur et de la Recherche Scientifique
Université 20 Août 1955 - SKIKDA -
Faculté des Sciences
Dpartement D' Agronomie
Réf : / D S AGRO/2026

وزارة التعليم العالي و البحث العلمي
جامعة 20 أوت 1955- سكيكدة
كلية العلوم
قسم علوم الفلاحة
رقم : / ق ع ف /2026

Skikda le : 30 JUIN 2026

جامعة 20 أوت 1955 سكيكدة
قسم
العلوم الفلاحية

Autorisation de Dépôt de Mémoire de Master

Je soussigné : **Pr. Heddam Salim** Président de Jury

Certifie que :

L'étudiant(e) : **Zouied Djamila**

Spécialité : **Aménagement Hydro-Agricole**

Intitulé : **Modélisation de la Dose Optimale de Coagulant dans le Traitement des Eaux à l'Aide de l'Intelligence Artificielle.**

A apporté les corrections relatives à son travail de mémoire.

Encadrant

Président de Jury