REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE YAHIA FARES DE MEDEA

Département de Génie des Procédés et Environnement

Laboratoire de Biomatériaux et Phénomènes de Transport (LBMPT)

THÈSE

EN VUE DE L'OBTENTION DU DIPLOME DE DOCTORAT

Faculté : Technologie

Filière : Génie des Procédés

Spécialité : Génie des Procédés

Présentée par

TAHRAOUI HICHEM

Licence en Chimie biologie : Université de Médéa, Algérie
Master en Pollution chimique et gestion de l'environnement : Université de Médéa, Algérie

En vue de l'obtention du titre de Doctorat en Génie des Procédés

Thème

Optimisation de l'élimination des micropolluants organiques

Soutenue le : 08/04/2021 **Devant le Jury composé de :**

Nom et Prénom	**Grade**	
Mr Benkortbi Othmane	Professeur, Université de Médéa	Président
Mr Allouaoui Redha	Maitre de Conférences A, Université de Médéa	Examinateur
Mme Seddari Soumia	Maitre de Conférences A, Université de Médéa	Examinatrice
Mr Kerchich Yacine	Maitre de Conférences A, école polytechnique, El-Harrach	Examinateur
Mr Belhadj Abd-Elmouneïm	Maitre de Conférences A, Université de Médéa	Directeur de thèse
Mr Hamitouche Adhya edinne	Directeur de recherche, CRAPC	Co-Directeur
Mme Boucherit Nabilla	Maitre de Conférences A, Université de Médéa	Invité

Année Universitaire : 2020/2021

Dédicace

Au-delà des personnes, des lieux et des temps, je dédie ce travail avec ma profonde conviction, à tous ceux qui ont toujours cru et sans doute à la Science. Cette lumière qui éclaire les esprits et leurs permet de transcender les limites installées par les sociétés et les cultures dans leurs périodes de dégénérescence.

À ceux qui savent concrétiser leurs idées et savent réussir.

À ceux qui ne portent que du bien pour les autres.

À ceux qui ne vivent que dans et par la vérité.

Je dédie ce travail à mes chers parents SLIMANI FATMA et l'âme de mon père TAHRAOUI OMAR pour leur soutien et leur tendresse quotidienne depuis toujours. Je leur dois l'épanouissement de ma vocation.

À mes frères (SAMIR, FAYÇAL et REDHA) et mes belles sœurs (FARIDA, SIHAM et SABRINA).

À mes sœurs (AMINA et ILHEM) et mes beaux-frères (AMINE et KAMEL), en particulier AMINA et AMINE. Qu'ils soient remerciés pour ses aides et pour leurs soutiens et leurs tendresses quotidiennes depuis toujours.

À Selma toumi. Qu'elle soit remerciée pour son amitié et qu'elle trouve ici le témoignage de notre immense estime.

À mes amis de l'Université Dr. Yahia FARES de Médéa.

HICHEM TAHRAOUI

Remerciements et gratitude

Au terme de ce modeste travail, je voudrais commencer par remercier ALLAH Tout-Puissant qui m'a donné le courage, la volonté et la force de mener à bien ce travail.

Je tiens à exprimer ma gratitude et mes sincères remerciements à mon directeur de thèse Dr Abd-Elmouneïm Belhadj qui m'a aidé, guidé et suivi attentivement afin de mener à bien ce travail, pour ses précieux conseils, sa générosité et sa patience, pour son Support. Ces exigences m'ont permis de progresser à la fois professionnellement et personnellement. Sans oublier qu'Il était frère avant d'être mon directeur de thèse. Merci beaucoup.

Je tiens à exprimer ma gratitude et mes sincères remerciements à mon co-directeur de thèse Dr Adhya eddine Hamitouche pour m'avoir donné accès au laboratoire de CRAPC au bou ismail, pour leur disponibilité, leur qualité humaine et leur réel soutien. Merci beaucoup.

Je suis profondément reconnaissant au Pr Abdlatif Amrane de l'Université de Rennes d'avoir suivi et guidé mon travail avec gentillesse. Je tiens à lui exprimer ma gratitude pour son aide et ses qualités scientifiques et humaines. Son accompagnement et son suivi durant les trois années de préparation de cette thèse, ainsi que les nombreuses discussions qui m'ont permis d'avancer et de compléter cette thèse. Je dis merci de me donner l'opportunité de travailler sous votre qualité scientifique. Merci beaucoup.

Je tiens à exprimer mes sincères remerciements à M. Benkortbi Othmane Professeur de l'Université de Médéa pour avoir accepté de présider le jury et de juger mon travail.

Je tiens à exprimer mes sincères remercie

ments aux membres du jury: M. Allouaoui Redha `` Maitre de Conférence A " de l'Université de Médéa, Mme Seddari Soumia `` Maitre de Conférence A " de l'Université de Médéa, M. Kerchich Yacine "Maître de conférences A" de l'Ecole polytechnique d'El-Harrach et Mme Boucherit

Nabilla "Maître de conférences A" de l'Université de Médéa, pour les avoir acceptés d'examiner mon modeste travail.

Je tiens également à remercier le professeur Pierre Cardol de l'Université de Liège pour m'avoir accueilli dans son équipe `` génétique et physiologie des microalgues '', je tiens à lui exprimer ma profonde gratitude pour m'avoir suivi et orienté mon travail avec gentillesse, je voudrais également remercier son équipe du laboratoire de génétique et physiologie des microalgues, en particulier M. Fabrizio Iacono, Mme Gain Gwenaelle et Mme Rebecca Houet.

Ce serait de l'ingratitude si j'ose oublier l'aide précieuse et désintéressée du Docteur Nassim Moula de l'université de liège. Il n'a cessé de s'intéresser à mon projet malgré le fait que le thème du travail ne coïncidait pas avec son profil. Je ne saurais oublier son aide, ses conseils, pour son disponibilité, je ne peux que le remercier du fond du cœur.

Je remercie sincèrement M. Salam abudura, Mme Sonia Lefnaoui, Mme Amina Benkiar, Mme Saliha Bourenane, Mme Selma Toumi, M. Mohamed Rahal, M. Monir Bouhheda et M. Hantebli Mohamed. Qu'ils m'ont aidé à mettre en valeur ce travail, j'exprime ma profonde gratitude pour leurs disponibilités, leurs qualités humaines et leur réel soutien, à travers leurs implications maximales.

Je tiens particulièrement à remercier toute l'équipe de laboratoire de l'algérienne de l'eau (ADE), en particulier Madame la directrice du laboratoire, DJEZZAR, pour m'avoir donné accès au laboratoire et aussi pour les nombreuses discussions qui m'ont permis d'avancer et compléter cette thèse, ainsi que chef de service Mme Mounia.

Je voudrais particulièrement remercier toute l'équipe du Centre d'analyse physique et chimique (CRAPC) de Bou issmail, en particulier M. Mohamed Abbaci, M. Mohamed Hadjadj, M. Yacine Laichi, M.Abou moustapha Mohamed, Mme Souad Touati et Mme Boukhelata Nacera.

Je tiens en particulier à remercier toute l'équipe du groupe Saidal, en particulier Madame la Directrice, Sonia Keskes, pour m'avoir donné accès au laboratoire, ainsi que la chef de département, Mme Bakhti.

Je voudrais en particulier remercier toute l'équipe du laboratoire Hall de la Faculté de technologie, y compris M. Younes Chiba, M. Sofiane Mouloud, Mme Radia Fargani et Mme Meriem Benhaoua.

Je tiens particulièrement à remercier toute l'équipe du laboratoire de la Faculté des sciences, en particulier M. Doyen Hadidi Noureddine pour m'avoir donné accès au laboratoire, Mme Amina Tifoura, Mme Faiza Masaad et Mme Fairouz.

Je voudrais particulièrement remercier les enseignants de la formation doctorale: M. Abd-Elmouneïm Belhadj, M. Benkortbi Othmane, M. Laidi Maamar, M. Zakaria Triki, M. Younes Chiba, Mme Ferhat Samira, Mme Amina Benkiar et Mme Naila Terkman, je les dis merci pour les qualités pédagogiques et les précieux conseils.

A tous les enseignants du département Génie des Procédés et de l'Environnement, je dis merci pour la qualité de votre enseignement qui restera à jamais gravé dans ma thèse.

Merci à tous ceux qui m'ont aidé directement ou indirectement au cours de ces années de recherche et d'étude. Je ne peux conclure sans un mot de remerciement à tous nos collègues.

HICHEM TAHRAOUI

ملخص: تحسين التخلص من الملوثات العضوية الدقيقة

الهدف من هذه الدراسة هو تحسين التخلص من الملوثات العضوية الدقيقة.

ركزت مساهمتنا الأولى على تنبؤ جرعة المخثر للتخلص من التعكر (المواد العضوية) لتجنب إضاعة الوقت وكذلك المواد الكيميائية. لهذا الغرض ، تم النظر في أربعة نماذج رياضية مختلفة في هذه الدراسة ، وهي: منهجية سطح الاستجابة (RSM) ، والشبكات العصبية الاصطناعية (RNA) ، وآلة ناقلات الدعم (SVM) والنظام التكيفي للاستدلال العصبي الضبابي (ANFIS) . أظهرت النتائج أن جميع النماذج تناسب البيانات التجريبية بدقة ، على الرغم من أن نموذج RNA كان أعلى قليلاً من النماذج الأخرى. ومع ذلك ، أدى نموذج SVM إلى نتائج مشابهة تقريبًا لتلك الخاصة بنموذج RNA ؛ كان الاختلاف الوحيد في مرحلة التحقق ، لأن معامل الارتباط كان مرتفعًا جدًا والمؤشرات الإحصائية صغيرة جدًا لنموذج RNA مقارنة بنموذج SVM. ومع ذلك ، من وجهة نظر اقتصادية ، كان نموذج SVM أكثر ملاءمة من نموذج RNA ، نظرًا لأن عدد المعلمات كان 22 ، وهو ما يقرب من نصف عدد معلمات نموذج RNA ، بينما كانت النتائج تقريبًا مماثلة في جميع أنحاء مرحلة البيانات. لزيادة خفض التكاليف الاقتصادية ، يمكن أيضًا استخدام نموذج RSM الذي ظل مفيدًا للغاية بسبب معاملاته العالية إذا كان مرتبطًا بعدد المعلمات ، فقط 13. بالإضافة إلى ذلك ، ظلت المؤشرات الإحصائية لنموذج RSM مقبولة .. ركزت مساهمتنا الثانية على إيجاد مصدر للمياه الخام شديدة التلوث من أجل التحقق من صحة النموذج الرياضي واختبار فعالية المعالجة في إزالة المواد العضوية وغيرها من المعايير الفيزيائية والكيميائية للمياه من أجل جعل المياه النظيفة للاستهلاك. أظهرت النتائج الأولية أن وادي شفة كان ملوثاً جداً (فترة الكوليرا). أظهرت نتائج التحليلات الفيزيائية والكيميائية للمياه المعالجة كفاءة عالية جدًا في المعالجة لإزالة المادة العضوية وجعل الماء صالحًا للشرب ، مما يدل على كفاءة نموذجنا. ركزت مساهمتنا الثالثة على إيجاد نموذج رياضي للكبريتات والبيكربونات كدالة للمعلمات الفيزيائية والكيميائية. بالنسبة للكبريتات ، تم استخدام مجموعة من 84 نقطة بيانات. بالنسبة لـ RNA ، تم استخدام 18 خلية عصبية في طبقة الإدخال ، و 8 خلايا عصبية في الطبقة المخفية و 1 في طبقة الإخراج. تم استخدام خوارزمية التعلم Levenberg Marquardt مع وظيفة النقل السيني المماس الزائدية اللوغاريتمية في الطبقات المخفية والمخرجة. أظهرت مقارنة النتائج التي تم الحصول عليها من حيث متوسط الخطأ التربيعي (RMSE) ومعامل الارتباط (R) باستخدام نموذجي RNA و MLR تفوق النموذج (RNA) في التنبؤ بمحتوى الكبريتات الذائبة في مياه الشرب. بالنسبة للبيكربونات ، تم تقسيم مجموعة البيانات الناتجة إلى جزأين لتدريب الشبكة العصبية الاصطناعية ، حيث تم استخدام 70٪ من مجموعة البيانات للتدريب وتم تقسيم الـ 30٪ المتبقية أيضًا إلى جزأين. أخرى للتحقق من صحة النموذج. تم تقييم النموذج الذكي الناتج بناءً على معامل الارتباط الأقرب إلى 1 وأدنى خطأ مربع متوسط (RMSE). تم استخدام مجموعة من 84 نقطة بيانات في هذه الدراسة. لهذا الغرض ، يتم استخدام 18 معلمة في طبقة الإدخال وخمسة خلايا عصبية في الطبقة المخفية ومعلمة واحدة مستخدمة في طبقة الإخراج لنمذجة RNA. تم استخدام خوارزمية التعلم Levenberg Marquard ، السيني اللوغاريتمي ووظيفة النقل الخطي ، على التوالي للطبقة المخفية والمخرجة. أظهرت النتائج التي تم الحصول عليها خلال هذه الدراسة معامل ارتباط R = 0.99276 مع جذر متوسط الخطأ التربيعي RMSE = 11.52613 مجم / لتر.

Résumé: Optimiser l'élimination des micropolluants organiques

L'objectif de cette étude est d'optimiser l'élimination des micropolluants organiques.

Notre première contribution a porté sur la prédilection de la dose de coagulant pour l'élimination de la turbidité (matière organique) pour éviter de perdre du temps et aussi les produits chimiques. Pour cela, Quatre modèles mathématiques différents ont été considérés dans cette étude, ce sont: la méthodologie de surface de réponse (RSM), les réseaux neuronaux artificiels (RNA), la machine à vecteur de support (SVM) et le système adaptatif d'inférence neuro-floue (ANFIS). Les résultats ont montré que tous les modèles ajustaient avec précision les données expérimentales, même si le modèle RNA était légèrement au-dessus des autres modèles. Cependant, le modèle SVM a conduit à des résultats presque similaires à ceux du modèle RNA; la seule différence résidait dans la phase de validation, car le coefficient de corrélation était très élevé et les indicateurs statistiques très petits pour le modèle RNA par rapport au modèle SVM. Cependant, d'un point de vue économique, le modèle SVM était plus approprié que le modèle RNA, puisque son nombre de paramètres était de 22, soit près de la moitié du nombre de paramètres du modèle RNA (43 paramètres), alors que les résultats étaient presque similaires dans toute la phase de données. Pour réduire encore les coûts économiques, le modèle RSM peut également être utilisé qui est resté très utile en raison de ses coefficients élevés s'il est lié au nombre de paramètres, seulement 13. De plus, les indicateurs statistiques du modèle RSM sont restés acceptables. Notre deuxième contribution a porté de trouver une source d'eau brute très pollué afin de valider le modèle mathématique et testé l'efficacité de traitement sur l'élimination de les matières organique et les autres paramètres physicochimique de l'eau afin de rendre l'eau propre a le consommation à les humines. Les résultats première a été montré que oued cheffa a été très pollué (période de choléra). Les résultats des analyses physicochimiques de l'eau traitée ont montrés une efficacité très élevées de traitement à l'élimination des matières organiques et rendre l'eau potable, ce que signifié l'efficacité de notre modèle. Notre troisième contribution a porté de trouver un modèle mathématique pour sulfate et bicarbonate en fonction les paramètres physicochimique. Pour le sulfate, un ensemble de 84 points de données a été utilisé. Pour le RNA, 18 neurones ont été utilisés dans la couche d'entrée, 8 neurones dans la couche cachée et 1 dans la couche de sortie. L'algorithme d'apprentissage de Levenberg Marquardt (LM) avec fonction de transfert sigmoïde tangente hyperbolique logarithmique a été utilisé au niveau des couches cachées et de sortie. La comparaison des résultats obtenus en termes d'erreur quadratique moyenne (RMSE) et de coefficient de corrélation (R) à l'aide des modèles RNA et MLR a révélé la supériorité du modèle (RNA) dans la prédiction de la teneur en sulfate soluble dans l'eau potable. Pour bicarbonate, l'ensemble de données obtenu a été divisé en deux parties pour former le réseau de neurones artificiels, où 70% de l'ensemble de données a été utilisé pour la formation et les 30% restants ont également été divisés en deux parties égales: une pour les tests et l'autre pour validation du modèle. Le modèle intelligent obtenu a été évalué en fonction du coefficient de corrélation le plus proche de 1 et de l'erreur quadratique moyenne (RMSE) la plus faible. Un ensemble de 84 points de données a été utilisé dans cette étude. Pour cela, 18 paramètres dans la couche d'entrée, cinq neurones dans la couche cachée et un paramètre utilisé dans la couche de sortie sont utilisés pour la modélisation RNA. L'algorithme d'apprentissage de Levenberg Marquardt (LM), le sigmoïde logarithmique et la fonction de transfert linéaire ont été utilisés, respectivement pour la couche cachée et de sortie. Les résultats obtenus au cours de la présente étude ont montré un coefficient de corrélation de R = 0,99276 avec l'erreur quadratique moyenne RMSE = 11,52613 mg/L.

Abstract: optimize the elimination of organic micro pollutants.

Our first contribution focused on the predilection of the dose of coagulant for the elimination of turbidity (organic matter) to avoid wasting time and also chemicals. For this, Four different mathematical models were considered in this study, they are: the response surface methodology (RSM), artificial neural networks (ANN), the support vector machine (SVM) and the adaptive system of neuro-fuzzy inference (ANFIS). The results showed that all models accurately fitted the experimental data, even though the ANN model was slightly above the other models. However, the SVM model led to results almost similar to those of the ANN model; the only difference was in the validation phase, because the correlation coefficient was very high and the statistical indicators very small for the ANN model compared to the SVM model. However, from an economic point of view, the SVM model was more suitable than the ANN model, since its number of parameters was 22, which is almost half the number of parameters of the ANN model (43 parameters), while the results were almost similar throughout the data phase. To further reduce the economic costs, the RSM model can also be used which has remained very useful due to its high coefficients if it is related to the number of parameters, only 13. In addition, the statistical indicators of the RSM model have remained acceptable. . Our second contribution focused on finding a source of highly polluted raw water in order to validate the mathematical model and tested the effectiveness of treatment on the removal of organic matter and other physicochemical parameters of water in order to make the clean water for consumption in humins. The first results were shown that oued cheffa was very polluted (period of cholera). The results of the physicochemical analyzes of the treated water showed a very high efficiency of treatment to remove organic matter and make the water drinkable, which signified the efficiency of our model. Our third contribution focused on finding a mathematical model for sulfate and bicarbonate as a function of the physicochemical parameters. For sulfate, a set of 84 data points was used. For ANN, 18 neurons were used in the input layer, 8 neurons in the hidden layer and 1 in the output layer. The Levenberg Marquardt (LM) learning algorithm with logarithmic hyperbolic tangent sigmoid transfer function was used at the hidden and output layers. Comparison of the results obtained in terms of mean squared error (RMSE) and correlation coefficient (R) using the ANN and MLR models revealed the superiority of the model (ANN) in the prediction of soluble sulfate content. in drinking water. For bicarbonate, the resulting data set was split into two parts to train the artificial neural network, where 70% of the data set was used for training and the remaining 30% was also split into two parts. equal: one for the tests and the other for validation of the model. The resulting intelligent model was evaluated based on the correlation coefficient closest to 1 and the lowest mean square error (RMSE). A set of 84 data points was used in this study. For this, 18 parameters in the input layer, five neurons in the hidden layer and one parameter used in the output layer are used for ANN modeling. The Levenberg Marquardt (LM) learning algorithm, logarithmic sigmoid and linear transfer function were used, respectively for the hidden and output layer. The results obtained during the present study showed a correlation coefficient of R = 0.99276 with the root mean square error RMSE = 11.52613 mg / L., 18 neurons were used in the input layer, 8 neurons in the hidden layer and 1 in the output layer. The Levenberg Marquardt (LM) learning algorithm with logarithmic hyperbolic tangent sigmoid transfer function was used at the hidden and output layers. Comparison of the results obtained in terms of mean squared error (RMSE) and correlation coefficient (R) using the ANN and MLR models revealed the superiority of the model (ANN) in the prediction of soluble sulfate content. in drinking water. For bicarbonate, the resulting data set was split into two parts to train the artificial neural network, where 70% of the data set was used for training and the remaining 30% was also split into two parts.

equal: one for the tests and the other for validation of the model. The resulting intelligent model was evaluated based on the correlation coefficient closest to 1 and the lowest mean square error (RMSE). A set of 84 data points was used in this study. For this, 18 parameters in the input layer, five neurons in the hidden layer and one parameter used in the output layer are used for ANN modeling. The Levenberg Marquardt (LM) learning algorithm, logarithmic sigmoid and linear transfer function were used, respectively for the hidden and output layer. The results obtained during the present study showed a correlation coefficient of R = 0.99276 with the root mean square error RMSE = 11.52613 mg / L.

Table de matière

Dédicaces

Remerciements

Résumé

Table de matière

Liste de figures

Liste des tableaux

Liste des abréviations

Introduction générale **01**

SYNTHESE BIBLIOGRAPHE

PARTIE I : MATIERE ORGANIQUE

I.1. Introduction **3**

I.2. La matière organique dans les eaux de surface **3**

I.2.1. Origines des matières organiques dans les eaux de surface **5**

I.2.2. Formes des matières organiques dans les eaux de surface **7**

Solubles et colloïdales (carbone organique dissous COD) **7**

Particulaire (carbone organique particulaire COP) **7**

I.2.3. Catégories de la matière organique naturelle **8**

I.3. Caractérisation de la matière organique naturelle **8**

I.4. Interactions de la matière organique naturelle avec les réactifs de traitement des eaux **10**

I.4.1. Interaction de la matière organique naturelle avec les métaux dans la nature **10**

I.4.2. Interaction de la matière organique naturelle avec les composés organiques **12**

I.5. Interactions de la matière organique naturelle avec les réactifs de traitement des eaux **14**

- **I.5.1. Interaction de la matière organique naturelle avec les métaux de coagulants** **14**
- **I.5.2. Interactions de la matière organique naturelle avec les oxydants** **19**
- **I.6. Conclusion** **21**

Table de matière

PARTIE II : COAGULATION ET FLOCCULATION

II-1-Introduction 22

II-2-But de la coagulation floculation 22

II-3-La coagulation 24

II-3-1-Principe de la coagulation 25

II-3-2-Réactifs coagulants 27

II-3-3-Principaux réactifs coagulants 29

II-3-3-1-Les coagulants minéraux 29

II-3-3-2-Les coagulants organiques de synthèse 29

II-3-4-Avantages et inconvénients de divers coagulants chimiques 30

II-4-La floculation 30

II-4-1-Types de Floculation 31

II-4-1-1-Floculation péricinétique 31

II-4-1-2-Floculation orthocinétique 31

II-4-2-Les adjuvants de floculation ou floculants 32

II-4-Effet de la coagulation sur les paramètres de l'eau 32

1)-La turbidité 32

2)-La couleur 33

3)-La résistivité 33

4)-Le PH 33

5)-La silice, les phosphates et le fluor 33

II-5-Facteurs intervenant dans la coagulation des particules en suspension 33

A/-La valence des colloïdes 33

B/-Le potentiel Zêta 33

C/-Grosseurs et concentration des particules colloïdales 34

D/-L'état d'hydratation 34

E/-Electrisation des particules 34

F/-Concentration du coagulant 35

G/-Vitesse et temps d'agitation 35

H/-La température 35

a) I/-Le potentiel d'hydrogène pH 35

J/-Rôle des bicarbonates dans l'eau 36

II-5-Paramètres influençant la coagulation floculation 36

Table de matière

II-5-1-Facteurs liés à la qualité de l'eau brute 36

II-5-1-1-La nature de la couleur et de la turbidité et leur proportion relative dans l'eau brute 36

II-5-1-2-La température 37

II-5-1-3-L'influence du pH 39

II-5-1-4-Effets de l'alcalinité 41

II-5-1-5-Effets des sels dissous 41

II-5-1-6-L'influence de la minéralisation 41

II-5-2-Facteurs liés au traitement 42

II-5-2-1-La nature et le dosage du coagulant 42

II-5-2-2-Les conditions d'agitation 42

II-5-2-3-La nature et le dosage de l'aide coagulant 43

II-6-Conclusion 43

MATERIELS ET METHODES

- **Introduction 44**
- **PARTIE I : COAGULATION**
- **I.1. Travail au laboratoire 45**
- **I.1.1. Principe 45**
- **I.1.2. Réactifs 45**

I.1.3. Appareillage 45

I.1.4. Préparation et conservation des échantillons 46

I.1.5. Mode opératoire 46

I.2. Méthodes de prédiction 47

I.2.1. Plan d'expérience (surface de réponse) 49

I.2.2. réseau neurone 50

I.2.3. Support vector machine 51

I.2.4. Système d'inférence neuro floue adaptatif 52

II. PARTIE II: CARACTERISATION DES EAUX BRUTES

Caractérisation des eaux brutes et traitées 53

II.1. Mode de prélèvement 53

***II*.2. Matières 53**

Table de matière

II.3. Matériels utilisés **54**

I.4. Méthodes d'analyse physico-chimique **54**

II.4.1. Analyse physique **54**

A. a) Température, salinité, Total solides dissous, et conductivité **54**

B. b) Turbidité **54**

C. c) PH **54**

II.4.2. Analyse chimique **55**

a) Dureté Totale **55**

b) Dosage des chlorures par la méthode de « MOHR » **55**

c) Dosage des nitrites (Méthode au réactif de ZEMBELLI) **55**

d) Dosage des nitrates (Méthode au salicylate de sodium) **55**

e) Dosage des phosphates **56**

f) Dosage du Sodium, Potassium et Calcium (Méthode par spectrométrie d'émission de flamme) **56**

g) Dosage de la matière organique (Méthode à chaud) **56**

h) Résidu sec **57**

i) Dosage de L'ammonium(Méthode de Nessler) **57**

j) dosage de calcium (Méthode titremetrique à l'EDTA) **57**

k) Détermination de sulfate **57**

l) Détermination d'Aluminium (par spectromètrie). **57**

m) Détermination du Manganèse (par spectrométrie) **58**

n) Détermination de l'Alcalinité **58**

o) Dosage du fer (par spectrométrie) **58**

I.5. Analyses physico-chimiques **58**

II.5.1. Mode opératoire **59**

II.5.2. Turbidité **59**

II.5.3. Dosage de Calcium Ca^{2+} **59**

II.5.4. Dureté Totale (TH) **60**

II.5.5. Dosage de Chlorure ($Cl^{-)}$ **61**

II.5.6. Détermination de Magnésium (Mg2+) **61**

II.5.7. Détermination de Titre Alcalimétrique Complet (TAC) et l'acidité (HCO_3) **61**

II.5.8. Détermination de Résidu Sec **62**

Table de matière

II.5.9. Détermination d'ammoniacal (NH^{+}_{4}) **62**

II.5.10. Détermination de nitrites NO^{-}_{2} **62**

II.5.11. Détermination du phosphate PO_4^{3-} **63**

II.5.12. Détermination de sulfates SO_4^{2-} **63**

II.5.13. Détermination de nitrate NO^{-}_{3} **63**

II.5.14. Détermination de fer **64**

II.5.15. Détermination du manganèse **64**

II.5.16. Détermination de sodium et de potassium **65**

II.5.17. Détermination de l'oxydation au permanganate de potassium **65**

II.5.18. Détermination d'Aluminium Al^{3+} **66**

PARTIE III : MODELISATION

III.1. Modélisation **68**

III.1.1. Sulfate **68**

III.1.1.1. Base de données **68**

III.1.1.2. Méthodes de prédiction **68**

III.1.1.2.1. Réseau de neurone **68**

III.1.1.2.2. Régressions linéaires multiples **69**

III.1.1.3. critères d'évaluation statistique **69**

III.1.2. Bicarbonate **70**

III.1.2.1. Base de données **70**

III.1.2.2. critères d'évaluation statistique **70**

RESULTATS ET DISSCUSSION

PARTIE I : ELIMINATION DES MICROPOLLUANTS ORGANIQUES PAR COAGULATION

I.Prédiction des doses de coagulation **72**

I.1. RSM **72**

I.2. ANN **75**

I.3. SVM **77**

I.4. ANFIS **80**

I.5. Comparaison des modèles **82**

I.6. Residus **84**

PARTIE II : RESULTATS DES ANALYSES PHYSICO-CHIMIQUES DE L'EAU BRUTE ET TRAITEES

Table de matière

II.1.1. Paramètres physiques **85**

II.1.1.1 Potentielle d'hydrogène (pH) **85**

II.1.1.2 Conductivité électrique **86**

- **II.1.1.3. Turbidité** **87**
- **II.1.2 Paramètres de pollution** **88**
- **II.1.2.1 Matière organique** **88**

II.1.2.2 Ammonium **89**

II.1.2.3. Phosphate **90**

II.1.2.4 Nitrate **90**

II.1.2.5 Nitrite **91**

II.1.3. Minéralisation globale **92**

II.1.3.1 Calcium, Magnésium et TH **92**

II.1.3.2Chlorure **95**

II.1.3.3 Bicarbonate **96**

II.1.3.4 Sulfate **97**

II.1.4.Paramètres indésirables **98**

II.1.4.1. Fer **98**

II.1.4.2. Manganèse **99**

II.1.4.3 Aluminium **100**

II.1.5.Paramètres des alcalins **101**

II.1.5.1. Potassium **101**

II.1.5.2. Sodium **101**

II.1.5.2. Titre alcalimétrique complet (TAC) **102**

II.1.6. Résidu sec **104**

Partie III : Résultats de Modélisation

III.1. Modélisation de sulfate **106**

III.1.1. Multiple Linaire Régression **106**

III.1.2. Modélisation RNA **109**

III.1.3. Performances de prédiction **111**

III.1.4. Performances d'interpolation **113**

III.1.5. Performances d'extrapolation **114**

III.1.6. Comparaison de MLR et ANN pour prédire les concentrations de **115**

Table de matière

sulfate soluble

III.1.7 Etude des résidus **118**

III.2. Modélisation de bicarbonate **120**

III.2.*1* Performances de prédiction **123**

III.2.2 Etude des résidus **125**

CONCLUSION GENERALE **127**

REFERENCE BIBLIOGRAPHIQUE

Listes des figures

CHAPITRE I : SYNTHESE BIBLIOGRAPHE

PARTIE II : MATIERE ORGANIQUE

Figure I.1. Concentration approximatives en carbone organique dissous et particulaire des eaux naturelles 4

Figure I.2. Schéma du cycle de la matière organique et des transferts de carbone 5

Figure I.3. Granulométrie des principaux constituants de la matière organique contenue dans les eaux naturelles. Le terme CS se rapporte aux composés simples comme les carbohydrates, les glucides et les lipides 7

Figure I.4. Influence des acides humiques et fulviques sur l'adsorption de l'atrazine (5 mg/l) sur CAP (20 mg/l) 14

Figure I.5. Diagramme schématique montrant la reconformation des macromolécules humiques avec l'addition de polycations Al13 16

Figure I.6. Influence de la dose de coagulant sur l'élimination d'acides organiques (5mg/l) 18

PARTIE II : COAGULATION ET FLOCCULATION

Figure II.1. Temps de décantation des particules 22

Figure II.2. La coagulation floculation 22

Figure II.3. La barrière énergétique. 25

Figure II.4. Potentiels électrostatiques 26

Figure II.5. Solubilité d'Al en fonction du pH et du TAC 40

CHAPITR II : MATERIELS ET METHODES

I. 1ERE PARTIE : COAGULATION

Figure I.1. Jar test 46

Figure I.2. **Organigramme pour le développement et l'optimisation de l'architecture RNA** 51

Figure I.3. **Organigramme pour le développement et l'optimisation du modèle SVM** 52

II. 2 EME PARTIE : CARACTERISATION DES EAUX BRUTES

Listes des figures

Figure II.1. Schéma du principe d'un photomètre de flamme **56**

Figure II.2. Image prise d'un multi-paramètre utilisé dans la présente étude. **58**

Figure II.3. Image prise d'un PH-mètre utilisé dans la présente étude. **59**

Figure II.4. Image prise du dosage de calcium par l'EDTA. **60**

Figure II.5. Image prise du résultat de dosage de la dureté par l'EDTA. **60**

Figure II.6. Image prise du dosage des ins chlore par l'EDTA **61**

Figure II.7. Image prise d'un protocole de dosage de TAC de la présente étude. **62**

Figure II.8. Image prise lors de l'expérience de la détermination de nitrate. **64**

Figure II.9. Image prise pendant la détermination de la matière organique. **66**

Figure II.10. Image prise de sulfate ferreux ammoniacal utilisé dans l'expérience. **66**

CHAPITRE III : RESULTATS ET DISSCUSSION

PARTIE I : ELIMINATION DES MICROPOLLUANTS ORGANIQUES PAR COAGULATION

Figure I.1. Effets des facteurs de fonctionnement et de leurs interactions. **74**

Figure I.2. Relation entre la dose observée de coagulant et celles estimées par le modèle RSM **75**

Figure I.3. Relation entre la dose observée de coagulant et celles estimées par le modèle ANN **77**

Figure I.4. Relation entre la dose observée de coagulant et celles estimées par le modèle SVM **79**

Figure I.5. Relation entre la dose de coagulant observée et celles estimées par le modèle ANFIS **81**

Figure I.6. Relation entre les données expérimentales et les données prédites à l'aide de la modélisation RSM, RNA, SVM et ANFIS **84**

Figure I.7. Distribution des résidus pour les modèles RSM (a), ANN (b), SVM (c) et ANFIS (d) pour toutes les données. **85**

PARTIE II : RESULTATS DES ANALYSES PHYSICO-CHIMIQUES DE L'EAU BRUTE ET TRAITEES

Figure II.1. Variation des moyennes mensuelles du pH en fonction du temps (mois) des eaux brute **85**

Listes des figures

Figure II.2. Variation des moyennes mensuelles de la conductivité (µs/cm) en fonction du temps (mois) des eaux brutes (A) et traitées (B) **86**

Figure II.3. Variation des moyennes de la turbidité (NTU) en fonctions du temps (mois) des eaux brutes (A) et traitées (B) **87**

Figure II.4. Variation des moyennes mensuelles de matière organique (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B) **88**

Figure II.5. Variation des moyennes mensuelles d'ammonium (mg/l) en fonction du temps (mois) des eaux brutes (A) **89**

Figure II.6. Variation des moyennes mensuelles de phosphate (mg/l) en fonction du temps (mois) des eaux brutes (A) **90**

Figure II.7. Variation des moyennes mensuelles de nitrate (mg/l) en fonction de temps (mois) des eaux brutes (A) **90**

Figure II.8. Variation des moyennes mensuelles de nitrite (mg/l) en fonction de temps (mois) des eaux brutes (A) **91**

Figure II.9. Variation des moyennes mensuelles du calcium (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B). **92**

Figure II.10. Variation des moyennes mensuelles du Magnésium (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B). **93**

Figure II.11. Variation des moyennes mensuelles du la dureté (F°) en fonction du temps (Mois) des eaux brutes (A) et traitées (B). **94**

Figure II.12. Variation des moyennes mensuelles du chlorure (mg/l) en fonction de temps (mois) des eaux brutes (A) et traitées (B). **95**

Figure II.13 variation des moyennes mensuelles du bicarbonate (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B). **96**

Figure II.14. Variation des moyennes mensuelles du sulfate (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B) **97**

Figure II.15 variation des moyennes mensuelles du fer (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B). **98**

Figure II.16. Variation des moyennes mensuelles du manganèse (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B). **99**

Figure II.17. Variation des moyennes mensuelles du l'aluminium en fonction du temps (mois) des eaux traitées (B) **100**

Listes des figures

Figure II.18. Variation des moyennes mensuelles du potassium (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B). 101

Figure II.19. Variation des moyennes mensuelles du sodium (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B). 102

Figure II.20. Variation des moyennes mensuelles du TAC (mg/l de CaCO3) en fonction du temps (mois) des eaux brutes (A) et traitées (B). 103

Figure II.21. Variation des moyennes mensuelles du résidu sec (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B). 104

Partie III : Résultats de Modélisation

Figure III.1. Comparaison entre les valeurs expérimentales et calculées 108

Figure III.2. Evolution du MSE correspondant aux trois phases d'apprentissage 110

Figure III.3. Comparaison entre les valeurs expérimentales normalisées et calculées 111

Figure III.4. Comparaison entre les valeurs expérimentales normalisées et calculées 112

Figure III.5. Comparaison entre les valeurs expérimentales et calculées 113

Figure III.6. Comparaison entre les valeurs expérimentales et calculées 115

Figure III.7. Distribution des résidus pour les modèles ANN et MLR. 118

Figure III.8. Résidus relatifs aux modèles établis par régression linéaire multiple et réseau de neurones artificiels en fonction des valeurs estimées. 119

Figure III.9. Représentation schématique du modèle ANN 120

Figure III.10. Représentation schématique du modèle neuronal ANN 121

Figure III.11. Évolution du MSE 121

Figure III.12. Variations de gradient d'erreur, de taux d'apprentissage et d'erreur de validation en fonction du nombre d'itérations 122

Figure III.13. Relation entre les niveaux de bicarbonate observés et ceux estimés par le modèle ANN 123

Figure III.14. Comparaison entre les données expérimentales et prédites estimées par le modèle ANN 124

Figure III.15. Relation entre les données expérimentales et les données prédites à l'aide de la modélisation ANN: (1) formation, (2) test, (3) validation et (4) toutes les données 125

Figure III.16. Résidus relatifs aux modèles établis par le réseau de neurones artificiels en fonction de valeurs estimées 126

Liste des tableaux

SYNTHESE BIBLIOGRAPHIQUE

PARTIE I : MATIERE ORGANIQUE

Tableau I.1. Origine des polymères présentent dans les ressources en eaux 8

Tableau I.2. Outils analytique utilisés pour caractériser la MON 9

Tableau I.3. Structures des composées organiques étudiées 17

PARTIE II : COAGULATION ET FLOCCULATION

Tableau II.1. Interaction entre deux particules en fonction de la distance les séparant. 26

Tableau II.2. Perte d'alcalinité et production de CO2 de quelques coagulants. 29

Tableau II.3. Avantages et inconvénients de divers coagulants chimiques 30

Figure II.4. Potentiels électrostatiques 39

MATERIELS ET METHODES

I. 1ERE PARTIE : COAGULATION

Tableau I.1. Codage et niveaux réels pour le modèle CCD. 49

II. 2 EME PARTIE : CARACTERISATION DES EAUX BRUTES

Tableau II.1.Types de solution et produite utilisés pour les différents analyses. 53

Tableau II.2. Les différents matériels utilisés pour les différentes analyses. 54

RESULTATS ET DISSCUSSION

PARTIE I : ELIMINATION DES MICROPOLLUANTS ORGANIQUES PAR COAGULATION

Tableau I.1. Valeurs de B, xi i caractérisation des variables indépendantes du RSM. 72

Tableau I.2. RSM performances 75

Tableau I.3. Performances des différentes architectures ANN testées 76

Tableau I.4. Performances de la prédiction utilisant ANN 77

Tableau I.5. Performances des différentes fonctions du noyau SVM testées 78

Liste des tableaux

Tableau I.6. Performances de la prédiction à l'aide de SVM **80**

Tableau I.7. Performances des différentes architectures ANFIS testées **80**

Tableau I.8. Performances de la prédiction à l'aide de SVM **82**

Tableau I.9. Comparaison des performances prédictives des modèles RSM, ANN, SVM et ANFIS pour toutes les données **82**

PARTIE III : RESULTATS DE MODELISATION

Tableau III.1. Valeurs de, xi et caractérisation des variables indépendantes du RLM. **106**

Tableau III.2. Performances de la prédiction utilisant MLR **109**

Tableau III.3. Performances des différentes architectures ANN testées **109**

Tableau III.4. Performances de la prédiction utilisant ANN **112**

Tableau III.5. Performances de l'interpolation **114**

Tableau III.6. Performances de l'extrapolation **115**

Tableau III.7. Comparaison des performances entre les modèles de prédiction MLR et ANN **116**

Tableau III.8. Structure des réseaux de neurones **120**

Tableau III.9. Résultats des performances de prédiction en termes de toutes les erreurs et vecteur d'accord [α, β, R] **124**

Liste des abréviations

ADE: Algérienne Des Eaux.

AEP: Alimentation en Eau Potable.

ANFIS : Adaptive neuro fuzzy inference system

ANN : Artificielle neural network

b: Biais

Ca^{2+}: Ion Calcium.

Cl^{-}: Ion Chlorure.

cm : Centimètre.

cm^2: Centimètre carré.

DBO_5: Demande Biochimique en Oxygène.

DCO: Demande Chimique en Oxygène.

DO: Densité Optique

EDTA: Acide Ethylène Diamine Tétra acétique.

EPM: Erreur de Prédictiondu Modèle

ESP: Erreur Standard de Prédiction.

FEP: photométrie d'émission atomique de flamme.

GAC: Charbon actif granulaire.

IR: Infrarouge.

MAE : l'erreur absolue moyenne.

MES: Matières En Suspension.

mg/l: Milligramme par litre.

Liste des abréviations

Mg^{2+}: Ion Magnésium.

MLR : Multiple linaire régression

MO: Matière Organique.

MSE : erreur quadratique moyenne.

NA: Normes Algériennes.

Na^{+}: Ion Sodium.

NET: Noir d'ériochrome.

OMS: Organisation Mondiale de la Santé.

PH: Potentiel d'Hydrogéné.

pH: potentiel hydrogène.

R: Coefficient de Régression.

REQM: Racine de l'Erreur Quadratique Moyenne.

RMN: résonance magnétique nucléaire.

RNA : Réseau de Neurones Artificielle

RNA: Réseau de Neurones Artificiel

RNFF: Réseau de Neurones type de Feed-Forward.

RS: Résidu Sec.

RS: Résidu Sec.

RSM : Réponse surface méthodologie

SPC: plusieurs sous-produits de chloration.

SVM : Support vecteur machine

T: Température.

Liste des abréviations

TA: titre alcalimétrique.

TAC: Titre Alcalimétrique Complet.

TDS: Total des Solides Dissous.

TH: Titre Hydrotimétrique.

THM: trihalométhanes.

UCV: Unités de Couleur Vraie.

UJT: unité Jackson de turbidité.

UTN: Unités de Turbidité Néphalométriques.

UV: Ultra Violet.

w: Poids.

µs / cm: Micro-siémens par centimètre.

%: Pourcentage.

°C: Degré Celsius.

°F: degré fahrenheit.

°F: degré français.

Introduction générale

Introduction générale

L’eau est à l’origine de la vie sur terre. Elle est indispensable à la survie des êtres vivant et les plantes dans la nature. Depuis longtemps, notre environnement et directement influencé le mode de vie des êtres humains.

Les trois quarts de la surface de notre planète sont recouvert par de l'eau. Aussi, elle se trouve sous multiples formes : pluies, mers, océans, lacs, nappes souterraines,…etc.

L’Algérie comme d'autres pays sont confrontés à la problématique de l’eau potable. La surcharge du littoral, les disparités entre zones rurales et urbaines, les périodes de sécheresse et l’accroissement de la pollution, sont autant de facteurs qui déstabilisent l’équilibre déjà précaire de l’environnement. Notre pays dispose de ressources en eau très limité est tout au moins au nord ou vivent les trois quarts de la population. En plus de cette rareté d’eau et la grande disparité entre le nord qui est arrosé et le sud qui est aride, s’ajoute le problème de la dureté des eaux. En effet, les eaux du sud algériens ont fortement chargées en sels minéraux, ce qui conduit à une dureté très élevée [1].

L'eau potable est un bien des plus précieux que nous devons nous servir quotidiennement dans le présent à l'aide d'un robinet. Plusieurs communes de la wilaya de Médéa sont confrontées ces dernières années à de grandes perturbations dans l’alimentation en eau potable. Cela est dû à une croissance assez importante de la population. Afin de subvenir aux besoins de la population et mettre fin à ce manque d’eau. Une autre source de production en eau potable est mise en service ces derniers temps sis au barrage de Koudiet-Acerdoune de Bouira. Cependant, pour protéger et répondre aux besoins des consommateurs d'eau potable pour les personnes et aux échelles industrielles, il est fortement demandé d'exploiter la purification des eaux ainsi que les rendre buvables. Pour cette raison, dans ce travail nous allons se focaliser à étudier la qualité physico-chimique et d'autres paramètres via le service du laboratoire de l'Algérienne des eaux "ADE" de différentes sources dans les régions de Médéa.

D’autre part, la matière organique pose actuellement un problème aux stations de traitement des eaux potables. En effet, les principaux problèmes engendrés sont globalement la détérioration de la qualité organoleptique, un développement bactérien dans les conduites du réseau de distribution et par suite aggravation de la corrosion, et aussi une consommation importante du chlore lors de la désinfection [2].

L'objectif général de ce travail est d'éliminer la matière organique naturelle de l'eau de surface lors de son traitement par les méthodes conventionnelles d'eau potable. L'étude examinera l'efficacité du traitement de coagulation-floculation pour éliminer la matière organique et

rendre l'eau potable pour la consommation humaine. Le coagulant utilisé est le sulfate d'aluminium. Les tests porteront sur des échantillons d'eau brute les plus pollués de chafa. Pour cela, une étude préliminaire a été lancée afin de voir les interactions des doses de coagulant avec certains paramètres physico-chimiques et de créer un modèle mathématique afin de réduire le temps d'optimisation des doses de coagulant et de gagner du temps. Ce modèle mathématique a été appliqué à l'eau brute d'une station de traitement chafa afin d'évaluer l'efficacité de notre modèle par les analyses physico-chimiques de l'eau traitée, notamment pour l'analyse des paramètres de la matière organique.

L'étude s'organise en trois chapitres : **un chapitre**, regroupe une synthèse bibliographique relative à la généralité sur l'eau, matière organique, coagulation-floculation, intelligence artificiel. Dans **le second chapitre**, on aborde les méthodes et matériel utilisé dans notre travail tel que : les protocoles de l'analyse des paramètres physicochimique, la modélisation … . Le **troisième chapitre**, concerne le résultat. L'étude se termine par une conclusion générale qui rassemble les principaux résultats obtenus durant cette étude.

Revue bibliographique

Partie I : Matière organique

I.1. Introduction

Une eau de surface est un milieu très dilué et complexe, contenant des matières minérales et organiques diverses en suspension ou en solution. Les matières dissoutes et colloïdales constituent à elles seules 60 à 80% de la charge organique d'une eau : à côté des argiles et des hydroxydes métalliques, on trouve des acides humiques, fulviques, lignosulfoniques[1]. Cette composition de l'eau que nous venons de décrire brièvement n'est pas stable mais varie qualitativement et quantitativement dans l'espace et dans le temps.

Dans le milieu naturel, la matière organique aquatique est composée principalement de macromolécules (acides fulviques et humiques), associée à de petites molécules organiques. Ces dernières peuvent interagir avec les substances humiques par différentes interactions (Forces Van Der Waals, compléxation, …) et modifier leur comportement. De même, toutes ces substances organiques peuvent complexer diverses substances de nature minérale et notamment métallique.

Au cours de ce chapitre, nous décrirons quelques notions sur la matière organique naturelle. Nous présentons également son origine, formes, caractéristiques et sa présence dans les eaux de surface. Nous nous sommes aussi donc intéressés à l'interaction de la matière organique naturelle avec les constituants minéraux et organiques de l'eau, ainsi qu'à leurs réactivités avec les réactifs de traitement des eaux, et l'influence de la structure de ces composés sur les mécanismes de la réactivité.

I.2. La matière organique dans les eaux de surface

La matière organique naturelle est un ensemble de macromolécules provenant de la dégradation de débris animaux et végétaux ainsi que de la production primaire du phytoplancton dans les eaux naturelles.

Les eaux de surface contiennent généralement de 2 à 10 mg/l de carbone organique total (COT) mais certaines valeurs plus élevées (jusqu'à 30 mg/l de COT) peuvent être rencontrées dans les eaux retenues (barrages, lacs)[2].

En moyenne, le COT inclut moins de 10% de 90% de carbone organique particulaire (COP).

Le carbone organique dissous (COD) est essentiellement issu des processus de dégradation des végétaux et des détritus animaux[3]. Certains micropolluants provenant de l'activité agricole, des rejets industriels et urbains y sont également présents, mais à des concentrations faibles, de l'ordre du µg/l, voire du ng/l [2]. Le carbone organique particulaire est principalement composé de sédiments, zooplancton, phytoplancton et bactéries. La figure I.1 représente l'ordre de grandeur des pourcentages en ces deux types de matière organique dans différents types d'eaux.

Une eau de surface est un milieu très dilué et complexe, contenant des matières minérales et organiques diverses en suspension ou en solution. Les matières dissoutes et colloïdales constituent à elles seules 60 à 80% de la charge organique d'une eau : à côté des argiles et des hydroxydes métalliques, on trouve des acides humiques, fulviques, lignosulfoniques[4]. Cette composition de l'eau que nous venons de décrire brièvement n'est pas stable mais varie qualitativement et quantitativement dans l'espace et dans le temps.

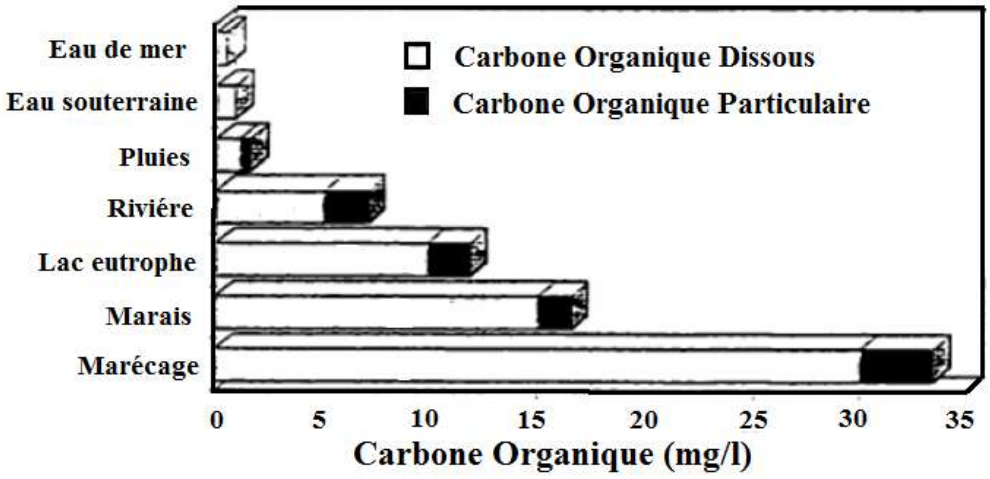


Figure I.1. Concentration approximatives en carbone organique dissous et particulaire des eaux naturelles[3].

Selon Jadas-Hécart (1989), on peut considérer que la matière organique est constituée par[5]

- Des substances humiques (30 à 50% du COD).
- Des acides hydrophiles (environ 30% du COD).
- Des acides carboxyliques (5 à 8% du COD).
- Des acides aminés libres et combinés (2 à 3% du COD).
- Des hydrates de carbone (5 à 10% du COD).
- Un ensemble complexe de composés à l'état de traces.

Les molécules simples (acides aminés, sucres…) représentent environ seulement 20 % de la matière organique aquatique et que les 80 % restants sont des macromolécules complexes difficilement analysables. La MOD présente dans les eaux douces est constituée en moyenne de 80 % d'acides hydrophobes ou hydrophiles, le reste étant des molécules neutres ou basiques [6].

La majeure partie des acides hydrophobes, en tout cas pour les eaux douces, est composée d'acides humiques (insolubles à pH < 2, minoritaires) et d'acides fulviques (solubles à tous les pH, majoritaires), regroupés sous le terme de substances humiques. Ces composés sont des

macromolécules amphiphiles de poids moléculaire supérieur à 500 Da (1 Dalton (Da) = 1 g.mol-1) présents sous forme de mélange hétérogène.

Les substances humiques sont souvent considérées comme la portion de la MOD la plus influente sur le transport et le comportement des polluants organiques et inorganiques [7], et représentent dont la fraction la plus importante de la matière organique, elles peuvent constituer jusqu'à 90% du COD dans certaines eaux très colorées[2,3].

I.2.1. Origines des matières organiques dans les eaux de surface

L'état dynamique de la matière organique rend son étude difficile du fait qu'elle subit sans cesse des cycles de synthèse biologique, métabolisation et décomposition. Dans l'environnement aquatique, la matière organique est majoritairement synthétisée par le phytoplancton puis utilisée comme nutriments par d'autres espèces qui peuvent la métaboliser puis l'excréter sous une autre forme (Figure I.2). Dans les eaux profondes ou plus généralement dans les zones aphotiques, la production par les bactéries devient majoritaire [8]. Une troisième zone de production de la matière organique dans les eaux naturelles est l'interface eau-sédiment où elle subit des transformations biologiques ou chimiques lentes[8]. La matière organique est également émise par dégradation des organismes animaux ou végétaux, que ce soit dans les milieux aquatiques ou terrestres. La matière organique fraîchement produite par la matière vivante (phytoplancton, bactéries, micro-organismes…) est généralement appelée matière organique récente en opposition à la matière organique plus humifiée (ancienne) issue de multiples étapes de dégradation de cette matière organique.

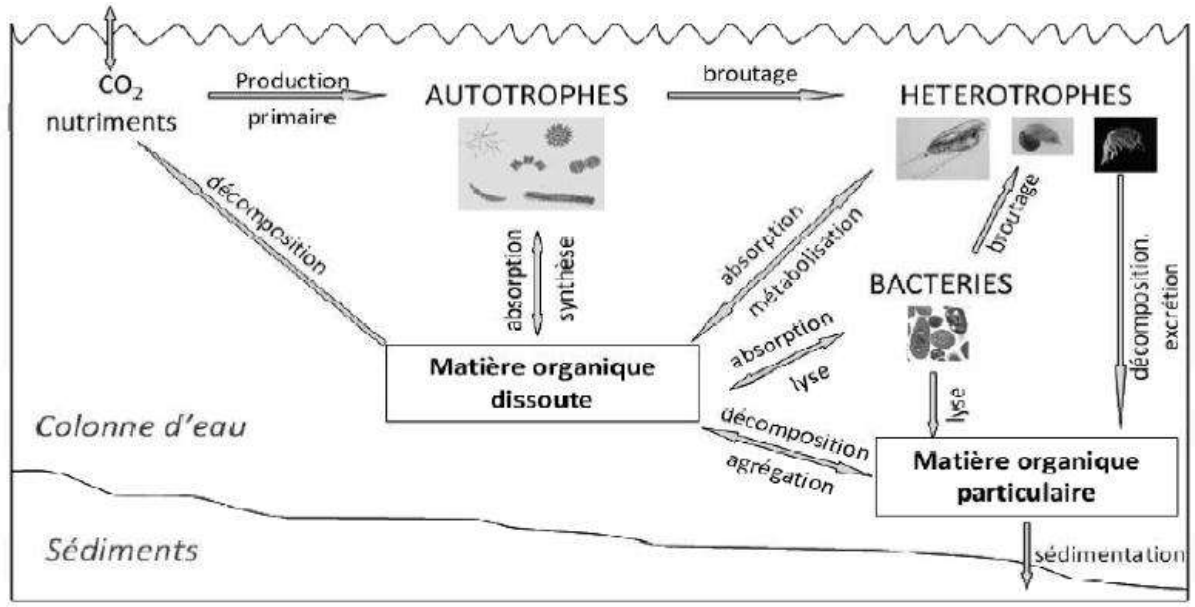


Figure I.2. Schéma du cycle de la matière organique et des transferts de carbone[9].

On peut regrouper les sources de matières organiques en grands ensembles suivants qu'elles soient d'origine diffuse ou ponctuelle ou qu'elles soient d'origines endogène ou exogène au milieu aquatique.

La partie importante de la matière organique de l'eau générée dans le milieu aquatique est dite endogène par opposition à celle qui vient des versants qui est dite exogène. Participent à la matière organique endogène les cellules et organismes vivants (macro et micro algues, autres microorganismes aquatiques) ainsi que les molécules issues directement de leurs métabolismes, mais aussi la matière organique « morte » provenant de la dégradation des végétaux par les organismes aquatiques vivants[10].

Les quantités de matière organique endogène produites dans les retenues d'eau eutrophies peuvent être extrêmement importantes.

La matière organique exogène provient des bassins versants, de manière diffuse ou ponctuelle. Les sources diffuses sont associées aux sols. Ceux-ci stockant une quantité très importante de matière organique, les formes de MO présentes dans les sols peuvent se retrouver dans les eaux. On distingue particulièrement quatre groupes:

- La matière organique vivante, végétale et animale, qui englobe la totalité de la biomasse en activité.
- Les débris d'origine végétale (résidus végétaux, exsudats) et d'origine animale (déjections et cadavres) qui sont regroupés sous le terme de matière organique fraîche.
- La nature de cette matière organique fraîche est directement liée aux activités de surface et en premier lieu au couvert végétal (foret, prairie, cultures). Elle est donc en majorité composée de MO végétale (cellulose, lignines, pectines, algines, etc.).
- Des composés organiques intermédiaires, encore appelés produits transitoires ou
- matière organique transitoire, qui est les matières organiques en cours d'évolution
- entre la MO fraîche et les composés finaux : les matières humiques (humus).
- Les composés organiques stabilisés ; les matières humiques, avec une fraction extractible par les solvants alcalins (acide fulviques et acides humiques) et une fraction non extractible par ces solvants (l'humine).
- Une autre source de MO des versants peut potentiellement avoir une origine anthropique correspond à la matière organique apportée sur les sols sous la forme d'amendements organique de type fumier et lisier. Une partie de cette MO est incorporée dans les sols et se décompose suivant ce qui est présenté ci-dessus. Une

autre partie peut potentiellement être transportée, par ruissellement de surface et après percolation dans les sols, et rejoindre de façon diffuse le milieu aquatique[10].

I.2.2. Formes des matières organiques dans les eaux de surface

A la multiplicité des sources, correspond une multiplicité des formes que prend la matière organique dans les eaux brutes (Figure I.3). Pour des raisons analytiques, on distingue les formes dissoutes et particulaires [10]:

A. Solubles et colloïdales (carbone organique dissous COD)

- Les acides humiques et fulviques.
- La cellulose.
- Les hydrocarbures.
- Les acides aminés, lipides et sucres (hydrate de carbone).
- Les sécrétions animales ou végétales, autolyse des organismes aquatiques, virus.

B. Particulaire (carbone organique particulaire COP)

- Les micro-organismes tels que les algues (phytoplancton), le zooplancton, les champignons et les bactéries ;
- Les substances humiques fixées aux argiles (complexe argilo-humique) ;
- Les débris divers issus de la fragmentation (feuilles mortes à l'automne, résidus de
- fauche sur les rives).

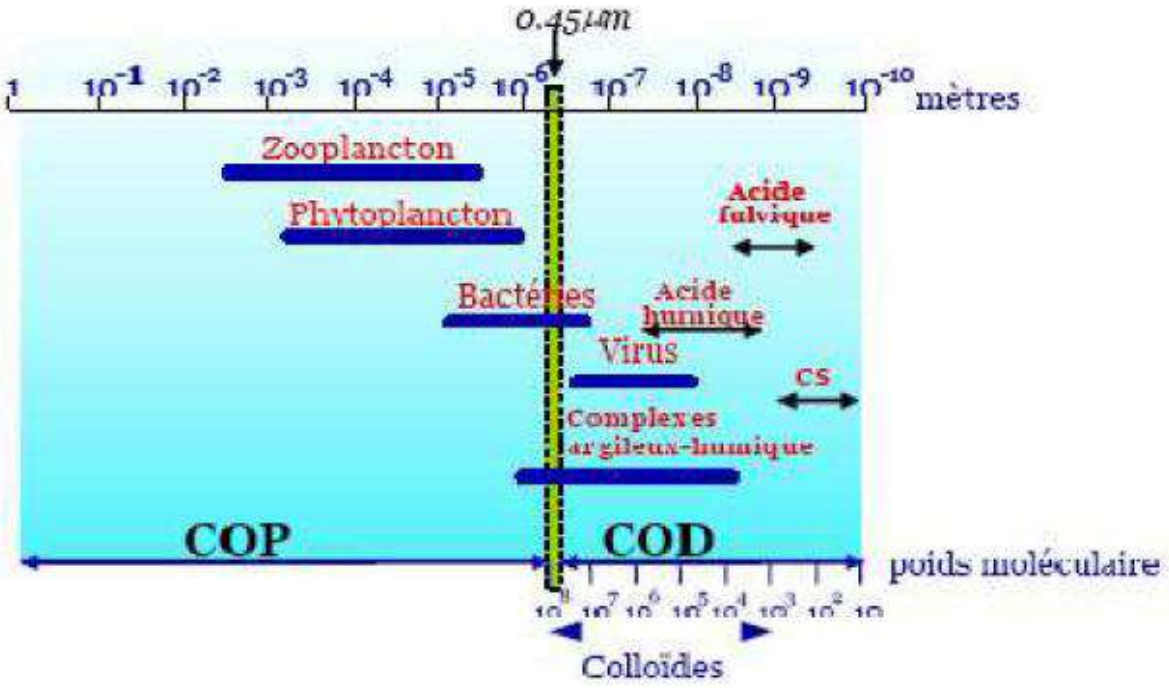


Figure I.3. Granulométrie des principaux constituants de la matière organique contenue dans les eaux naturelles.

Le terme CS se rapporte aux composés simples comme les carbohydrates, les glucides et les lipides[11].

I.2.3. Catégories de la matière organique naturelle

Les matières organiques naturelles constituent un mélange très complexe de biopolymères et de monomères générés par l'activité biologique sur un bassin versant et dans les ressources elles-mêmes. Les polymères les plus abondants sont les acides humiques, les polysaccharides et les protéines, alors que les monomères, moins abondants car facilement biodégradables, sont principalement représentés par les sucres aminés et les acides aminés. L'origine de ces biopolymères est présentée dans le tableau I.1 [12].

Tableau I.1. Origine des polymères présentent dans les ressources en eaux[12,13]

Types de biopolyméres	Origines
Polysaccharides (stockés sous forme d'amidon, ou structurel sous forme de cellulose)	Aquatique (algues et bactéries) Sol (résidus de plantes)
Protéines	Aquatique (algues et phytoplancton)
Sucre aminés	Parois cellulaires des bactéries et champignons)
Polyhydroxyaromatiques (PHA)	Plantes (lignine) aquatiques (algues, diatomées, animaux)

I.3. Caractérisation de la matière organique naturelle

Une classification simple des composants organiques aquatiques naturels n'est pas aisée. Toutefois, une distinction peut être faite entre différentes classes de composants par des méthodes d'isolation[14]. La matière organique naturelle a un caractère hétérogène, et se compose de carbone dissous (COD), colloïdal (COC) et de carbone organique sous forme particulaire (COP). Les fractions de COD sont généralement plus importantes que les fractions de COP. La détermination de la fraction de carbone organique particulaire se fait par filtration dans un filtre de 0,45 µm. Quant aux particules de dimension supérieure à 0,45 µm, elles peuvent généralement être classifiées en deux groupes, c'est-à-dire en substances non humiques et humiques. Les substances non humiques de ce groupe sont non solubles en milieu acide, c'est-à-dire de pH<2 et le deviennent à des valeurs de pH plus élevées. Les substances humiques restent non solubles, quelles que soient les valeurs de pH [15].

La fraction colloïdale est constituée de solides en suspension et se compose d'organismes vivants, d'exsudats cellulaires et de matières dégradées pouvant tous être associés à des phases minérales. La fraction COD répond à différentes dénominations telles que substances humiques, acides humiques et acides fulviques. Elle est généralement déterminée analytiquement en tant que carbone organique dissous.

Owen et al. (1993) suggèrent le protocole analytique suivant pour la caractérisation de la MON [16]:

- MON « Bulk » et diverses fractions de MON.
- En fonction de la taille; matière organique dissoute (MOD), colloïdale (MOC) ou particulaire (MOP) par utilisation d'une filtration membranaire séquentielle.

Une caractérisation ultérieure de la MOD par fractionnement est également possible via le poids moléculaire apparent en utilisant l'ultrafiltration et la chromatographie à perméation de gel et, selon le caractère humique et non humique, en utilisant la chromatographie d'adsorption par isolation de la résine XAD-8. La caractérisation de la fraction humique requiert également que l'on établisse la teneur en groupe fonctionnel acide (carboxylique et phénolique) par titration potentiométrique, la polarité relative par chromatographie liquide à haute pression en phase inversée et la capacité de liaison aluminium par extinction de fluorescence[16].

Le Centre de Recherche Coopératif sur la Qualité et le Traitement de l'Eau a réalisé une étude approfondie sur la MON dans le but de comprendre et de limiter l'impact sur la qualité et les procédés de traitement de l'eau en Australie[17]. L'étude a suggéré différents outils analytiques destinés à caractériser différents paramètres, comme il ressort du (Tableau I.2).

Tableau I.2. Outils analytique utilisés pour caractériser la MON

Paramètres	**Outils analytique**
Couleur	Spectrophotométrie visible
	Comparateurs visuels
Aromaticité (absorbance UV)	Spectrophotométrie UV
Carbone organique Total	Analyseur (COA)
Carbone organique Dissous (COD)	
Carbone organique biodégradable (COBD)	
Carbone organique assimilable (COA)	Potentiel de Recroissance Bactérienne (PRB)
Groupes fonctionnels (Aliphatique,	Chromatographie gazeuse- spectroscopie de

Aromatique, Contenant de l'azote)	Masse (CG-SM)
	Spectroscopie Infrarouge (FT-IR)
	Résonnance Magnétique Nucléaire (RMN)
DBO	DBO mètre
DCO	Kits Dr Lange LCK 414 de 5 à 60 mg / l O_2
COT	Kits Dr Lange LCK 385 de 3 à 30 mg / l
Absorbance en UV	Spectrophotomètre Hach

I.4. Interactions de la matière organique naturelle avec les réactifs de traitement des eaux

I.4.1. Interaction de la matière organique naturelle avec les métaux dans la nature

La charge des substances humiques est toujours négative ou nulle, d'intensité variable suivant le pH du milieu et provient de la dissolution des groupements fonctionnels. La nature de ces groupements fonctionnels est relativement bien connue car elle a fait l'objet de nombreuses études en raison de leur implication dans la complexation des cations métalliques. Les substances humiques possèdent un grand nombre de groupements fonctionnels hydrophiles : carboxyles (COOH), phénolique et/ou alcooliques (OH), carbonyles (C=O) ainsi que des groupements amines (NH_2). Les groupes acides COOH et OH contribuent en majorité à la nature acide des substances humiques, avec les groupements COOH dont la proportion est le plus souvent supérieure à 50%. La concentration en groupements fonctionnels acides dans les substances humiques exprimée en moles de groupements titrables rapportés au carbone organique, est généralement comprise entre 4,1 et 10 μmol H+ / mg COT (carbone organique total) [18].

Des résultats en spectroscopie infra-rouge ont confirmé que les groupes COOH ou plus précisément carboxylates (COO^-) jouent un rôle prédominant dans la complexation des métaux par les substances humiques. Les AF du fait de leur charge importante sont les plus réactives vis-à-vis des métaux. L'affinité des métaux pour les substances humiques suit dans la plupart des travaux, l'ordre décroissant suivant : $Cu^{2+} > Pb^{2+} > Zn^{2+} > Cd^{2+}$ [19].

Les substances humiques peuvent former avec les métaux différents types d'association suivant la concentration en métal et le métal impliqué[19–21]:

1. Réaction d'un COOH avec le métal pour former un complexe monodenté.
2. Réaction entre deux groupes COOH voisins pour former des complexes types phatale.

3. Réaction entre un groupe COOH et un groupe OH voisin pour former des complexes type salicylate.

4. Attraction électrostatique et formation d'une liaison hydrogène entre l'ion métallique hydraté et un groupe OH.

Les différents complexes peuvent se former entre groupes appartenant au même polymère ou entre groupes appartenant à des polymères différents pour former des liaisons intermoléculaires qui peuvent conduire à l'agrégation des polycondensats [22]. Ces réactions peuvent avoir lieu simultanément. La réaction (I.4) conduit à la formation de complexes de sphère externe. Les réactions (I.2) et (I.3) décrivent la formation de complexes de sphère interne qui sont plus stables que les complexes de sphère externe.

De nombreux travaux ont montré que la sélectivité et la force de liaison des métaux par les substances humiques dépendent en partie de la quantité de métal disponible pour la liaison et de la quantité de métal complexé. Par exemple, différentes études ont montré que la force de liaison des métaux par les groupes fonctionnels des substances humiques diminuait lorsque la concentration en métal complexé augmentait[23,24].

Les études utilisant la spectrophotométrie ESR ont permis de montrer l'existence de deux niveaux de liaison des métaux. Lorsque la teneur en métal est faible, les liaisons covalentes sont préférentiellement former pour donner des complexes bidentés ou monodentés alors qu'à des concentrations plus élevées, les liaisons deviennent de plus en plus ioniques[25]. Par ailleurs, le type de complexes formés et/ou la force de liaison avec laquelle les métaux vont être complexés par les groupes fonctionnels des substances humiques dépendent de l'état électronique des sites de complexation, des chaines aliphatiques associées, et de la présence éventuelle d'un noyau aromatique et, de la géométrie du site fonctionnel. Le pH et la force ionique du milieu vont aussi affecter les processus de complexation.

Les substances humiques jouent un rôle important dans la biodisponibilité et le transport de l'aluminium sous forme de composés stables et plus ou moins solubles. De nombreuses études ont été menées sur la complexation de l'aluminium par les substances humiques[26,27]. Plus particulièrement avec l'acide fulvique de masse molaire plus faible et de structure un peu moins complexe que l'acide humique.

La disponibilité de l'ion Al^{3+} et de ses monomère hydratés ($Al(OH)^{2+}$) , pour les plantes diminue lorsque les ions sont complexés par les substances humiques. Cela est d'autant plus vrai que le complexe métallo-organique a une faible solubilité en milieu aqueux.

Malgré la rétention de Al^{3+} par les acides humiques, une certaine quantité d'aluminium reste toujours disponibles dans le sol et est susceptible de former, avec d'autres composés

organiques et inorganiques de petite tailles et par conséquent plus mobiles[28], des composés solubles et donc transportable vers d'autres environnement.

Les différents sites de complexation que l'on peut rencontrer dans les acides humiques n'ont pas le même pouvoir complexant vis-à-vis de l'aluminium. L'affinité de ces ligands pour l'ion métallique augmente avec la basicité du groupe donneur : d'une manière générale les groupements ayant le plus grand pKa auront donc la plus grande affinité pour l'ion Al^{3+}.

Parmi les principaux groupements chélateurs de l'acide humique, on retrouve la fonction acide carboxylique, le groupement catéchol et le groupement hydroxy_carbonyle. Chacun de ces sites de chélation piège le métal en formant une pince autour de lui. Ces pinces sont plus ou moins grandes suivant le type de site et donc ne peuvent pas chélater tous les métaux, tout dépendra de leur taille et de leur environnement électronique. Par exemple, l'aluminium qui est de petite taille peut se fixer sur tous les sites.

Le pouvoir complexant de l'acide humique est très important, il peut non seulement facilement fixer les ions métalliques mais aussi complexer de petites molécules. Si l'acide humique est un bon complexant pour les métaux nécessaires au développement des plantes, il l'est aussi pour certains métaux issus de l'activité humaine (aluminium, plomb, zinc, cadmium, mercure…).

I.4.2. Interaction de la matière organique naturelle avec les composés organiques

Les substances humiques (SH), constituant 25 à 50 % du carbone organique dissous (COD) des eaux de surface[3], sont aussi présentes dans les sédiments. Ces acides polyélectrolytiques, qualifiés d'humiques ou fulviques selon leur solubilité aqueuse à pH inférieur à 2, sont produits par la décomposition des molécules organiques d'origine animale et végétale contenue dans les eaux. Leur structure macromoléculaire polyfonctionnelle varie selon leur origine biologique et leur processus de formation mais présente toujours des capacités complexantes à l'égard des métaux[29] et des propriétés amphiphiles leur conférant un rôle important dans le transport des micropolluants aromatiques[30,31] La matière humique présente en effet des propriétés hydrophobes rendant possible les interactions soluté avec des composés organiques non ioniques[32].

Les hydrocarbures aromatiques polycycliques (HAP) interagissent avec les substances humiques dissoutes ou fixées sur les colloïdes et sédiments, les substances humiques amphiphiles aussi peuvent recouvrir les colloïdes minéraux (argileux ou siliceux) avec une affinité dépendante de leur poids moléculaires apparent[33], de la nature du minéral et du pH de la solution[34]. La nature et la quantité de matériel humique fixé au minéral peuvent alors perturber l'adsorption des hydrocarbures aromatiques polycycliques (HAP) sur le solide. Les

différentes interactions entre les solutés (HAP_d et SH_d) et le solide peuvent se résumer par les équilibres suivants :

$(HAP)_d + (SH)_d \longleftrightarrow (HAP\text{-}SH)_d$ **(I.1)**

$(SH)_d$ + solide $\longleftrightarrow$ (H)-solide **(I.2)**

$(HAP)_d$ + solide $\longleftrightarrow$ HAP-solide **(I.3)**

$(HAP)_d + (SH)_d$ + solide $\longleftrightarrow (HAP\text{-}SH)_d$ + (SH)-solide + HAP-solide-SH **(I.4)**

L'équilibre (04) représente globalement les interactions des solutés humiques et aromatiques polycycliques en phase liquide et avec solide. La nature et la concentration de composés humiques en solution peuvent ainsi influencer la thermodynamique ou la cinétique de désorption des hydrocarbures aromatiques polycycliques du solide.

Les colloïdes naturels sont dans la majorité des cas revêtus d'un film de substances humiques [34]. Cette gangue organique est alors susceptible de modifier les phénomènes de sorption des HAP sur le solide.

Des essais de sorption du pyrène ont été menés à pH et force ionique identique, en solution aqueuse dépourvue d'AH, sur de la kaolinite préalablement mise en contact avec une solution aqueuse d'acide humique; lorsque la surface de la kaolinite est recouverte de composés humiques, sa capacité à fixer le pyrène est fortement modifiée. La quantité de pyrène adsorbé différent de celle fixée sur la kaolinite pure. Les molécules d'acide humique à la surface de l'argile semblent alors perturber la formation de microcristaux de pyrène. Les pesticides, provenant de l'activité agricole et des rejets industriels et urbains recouvrent un ensemble de substances chimiques (naturelle ou synthétiques) utilisées pour détruire diverses espèces animales ou végétales considérées comme nuisibles. Ces composées étant peu solubles dans les eaux, les risques de pollution par ruissellement ou drainage sont donc limités. Leurs associations avec les substances humiques[35] peut cependant modifier leur activité entrainant notamment une augmentation de leur solubilité et donc de la pollution des eaux[36].

Toutefois ces associations ne dépendant pas seulement de la nature du pesticide, mais sont fortement influencées par l'origine même des substances humiques [36].

Ayele et al., 1989, ont montré l'influence des substances humiques sur l'adsorption de l'atrazine sur CAP ; les substances humiques sont utilisées à des concentrations de 0 à 30 mg/l en mélange avec l'atrazine (5mg/l) sur CAP (20 mg/l) pour des pH de 5,6 ; 7,0 et 8,8. La vitesse d'adsorption de l'atrazine à pH acide est plus importante en présence de substances humiques[36]. La figure I.4 montre l'influence des acides humique et des acides fulviques sur l'adsorption de l'atrazine.

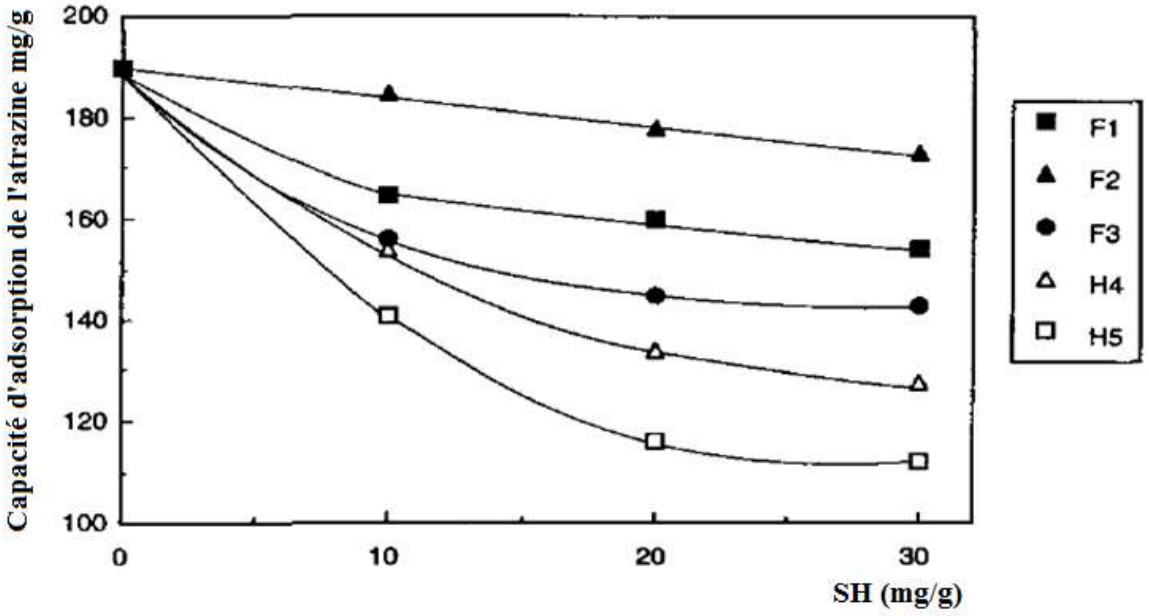


Figure I.4. Influence des acides humiques et fulviques sur l'adsorption de l'atrazine (5 mg/l) sur CAP (20 mg/l)[36]

I.5. Interactions de la matière organique naturelle avec les réactifs de traitement des eaux

I.5.1. Interaction de la matière organique naturelle avec les métaux de coagulants

Le développement du thème agrégation s'est effectué initialement autour du procédé de coagulation-floculation utilisé dans la production d'eau potable. Au cours de cette opération, une quantité suffisante d'un sel métallique hydrolysable est ajoutée à une eau de surface dans un réacteur sous agitation contrôlée afin de rassemblé les matières colloïdales organique et minérales qu'elle contient sous la forme d'agrégats sédimentable. Les phénomènes mise en jeu sont particulièrement complexes et sont à la fois d'origine physico-chimique (hydrolyse du coagulant, interaction avec les surfaces des colloïdes naturels, structuration des agrégats …) et d'origine hydrodynamique (dispersion du coagulant, croissance/rupture des agrégats en fonction des gradients de vitesse locaux…).

Même lorsqu'elle s'effectue dans des conditions contrôlées en laboratoire, l'hydrolyse des sels d'aluminium et de fer conduit à la formation d'espèces hydrolysées dont les précurseurs ne sont pas toujours identifiés de manière certaine. C'est le cas du polycation Al^{13}, formé d'un aluminium central en position tétraédrique et entouré de douze aluminiums en coordinence octaédrique, pour lequel les trimères d'aluminium proposés pour rendre compte de sa formation, n'ont pas encore été expérimentalement isolés.

Un cas particulièrement intéressant concerne la matière organique naturelle dont les rendements d'élimination dépassent rarement les 50 – 60% en milieu industriel. Dans la filière de potabilisation, cette matière organique est à l'origine de la production de produits potentiellement cancérigènes lors de l'étape de désinfection, et contribue à la prolifération bactérienne dans les réseaux de distribution. En utilisant un modèle de substance humique synthétisé à partir de l'oxydation d'un mélange catéchol-glycine, Bihannic et al., (2003) [37]ont pu montrer en croisant les résultats obtenu par RMN de 27Al et RMN-PC de ^{13}C, que le mécanisme de déstabilisation de l'acide humique synthétisé ne dépond pas du pH et de la concentration en coagulant (polycations Al^{13}). La seule modification majeure observée en fonction du pH correspond à l'ordre d'interaction des espèces coagulantes avec les groupements fonctionnels ionisés de l'acide humique modèle. Ainsi, à pH faiblement acide, ce sont les groupements carboxyliques qui fixent les polycations Al^{13} ; cependant, aux concentrations en coagulant plus élevés, il semble que la présence des polycations induise la déprotonation des sites phénoliques qui s'associent alors à leur tour avec les espèces coagulantes. A pH 8, l'ordre d'interaction est inverse : ce sont les groupements phénoliques qui s'associent en premier avec les polycations, les groupements carboxyliques participant plus tardivement à l'agrégation.

Un autre aspect de l'interaction espèce coagulant-matière organique concerne la reconformation des macromolécules humiques. En effet, comme les espèces coagulantes sont fortement chargées positivement et que la matière organique humique peut être assimilée à des polymères anioniques, une telle reconformation est susceptible de se produire au cours de l'agrégation. La spectroscopie de fluorescence utilisant la molécule de pyrène comme sonde a effectivement permis de montrer que les chaines de macromolécules humiques se réarrange au cours de l'agrégation en présence de polycations Al^{13}. A faible concentration en coagulant, des pseudo-micelles intra ou intermoléculaire apparaissent. Au-delà d'une certaine concentration, la tendance s'inverse et la structure initiale semble restaurée dans une certaine mesure (figure I.5).

La mise en évidence de ce phénomène de reconformation permet alors de rediscuter du problème de rendement d'élimination de la matière organique naturelle lors de la coagulation floculation des eaux de surface. En effet, les jar-tests montrent qu'à l'optimum de coagulation, les espèces coagulantes neutralisent la totalité des groupements carboxyliques de la matière organique à pH 6, alors qu'à pH 8, la charge des macromolécules n'est que partiellement compensée. Cette équivalence des charges cationiques et anioniques à pH 6 s'explique aisément si l'on considère qu'une reconformation des polymères d'acide humique

se produit au fur et à mesure de la fixation des polycations sur les macromolécules, ce qui facilite alors l'accès des sites ionisés aux espèces coagulantes. En revanche, à pH 8, il est probable que les clusters d'Al^{13} moins chargés, déterminent un réarrangement plus lent des polymères d'acide humique, ce qui conduit à une neutralisation partielle des macromolécules.
Ce même raisonnement étendu à une distribution en taille des macromolécules suggère que les polymères de grande taille se réarrangeant lentement sont neutralisés avant les polymères de plus petite taille se reconformant plus rapidement. Un tel phénomène pourrait expliquer que les acides humiques soient plus facilement éliminés que les acides fulviques dans les eaux de surface.

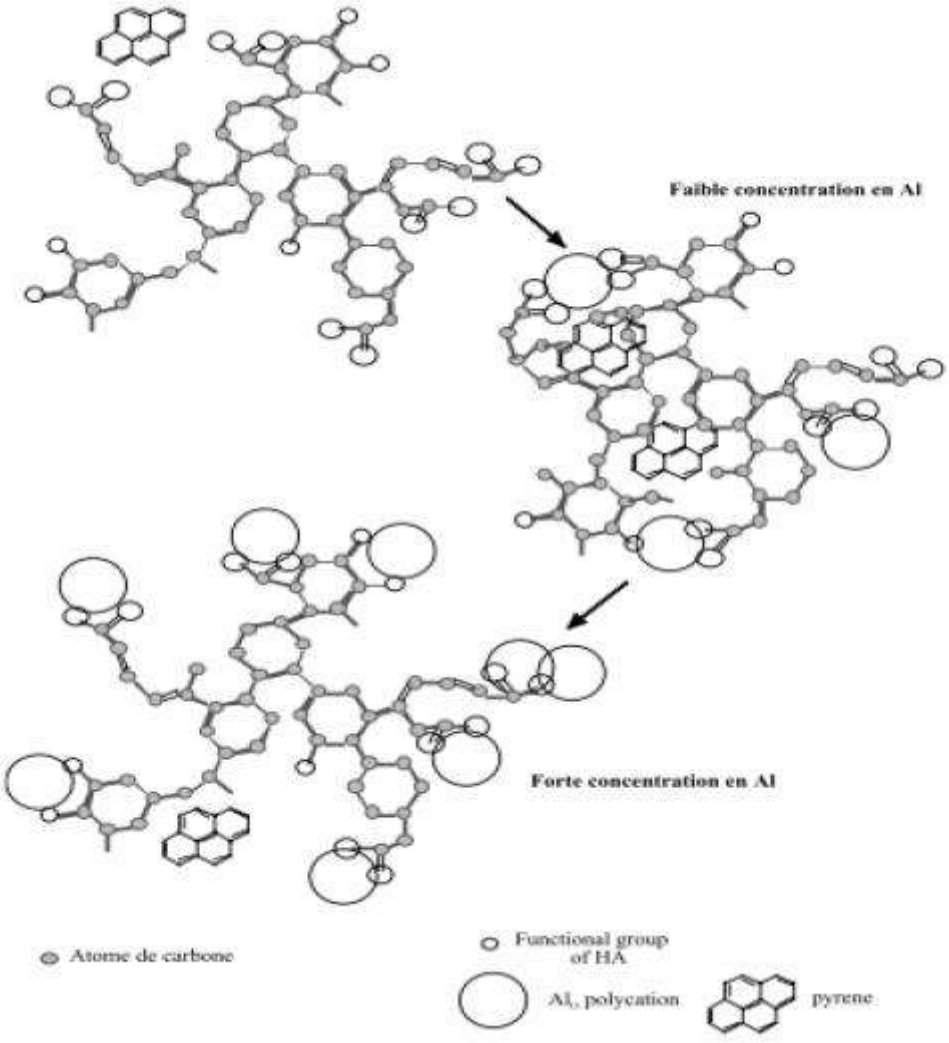


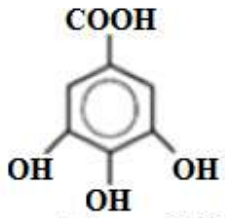
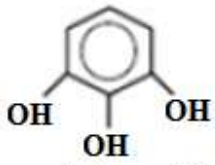
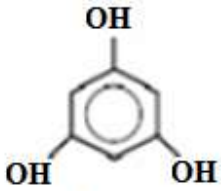
Figure I.5. Diagramme schématique montrant la reconformation des macromolécules humiques avec l'addition de polycations Al^{13} [37] ;

Les essais de coagulation-floculation sont conduits sur des solutions synthétiques de composés organiques dissous à raison de 5mg/l en eau distillée (phloroglucinol) (tableau I.3). Des doses croissantes de sulfate d'aluminium sont introduites dans les différentes solutions.

Les résultats présentés sur la figure I.6 montrent que les rendements d'élimination des molécules organiques évoluent différemment selon la structure chimique des composés[38]. Nous pouvons ainsi distinguer que lepourcentage d'abattement du phloroglucinol est faible. Ce composé présente un léger optimum pour 10 et 30 mg/l en sulfate d'aluminium. Par ailleurs, la catégorie des composées de type acide gallique et pyrogallol sont bien éliminés. A l'optimum, l'élimination de l'acide gallique atteint 72,58% alors que celle du pyrogallol atteint 53,58%. Ces rendements sont observés respectivement pour une dose de 90 mg/l et 5 mg/l en sulfate d'aluminium. De plus il faut remarquer que le pourcentage d'abattement de l'acide gallique se stabilise au-delà de l'optimum.

Tableau I.3. Structures des composés organiques étudiées [38]

En se référant aux données bibliographique[2,39],

acide gallique	Pyrogallol	Phloroglucinol
 Masse molaire = 172.12g	 Masse molaire = 126.11g	 Masse molaire = 162.14g

En se référant aux données bibliographique[2,39], les composés aromatiques présentent une seule fonction acide ou deux ou trois fonction non contigües (fonction en position méta ou para) ne sont pas bien éliminés. C'est le cas du phloroglucinol[38] Par ailleurs, l'élimination importante obtenus dans le cas de l'acide gallique et du pyrogallol confirment que la présence de deux fois deux groupements acides en contigües (COOH ou OH) facilite l'élimination des molécules organiques par coagulation floculation.

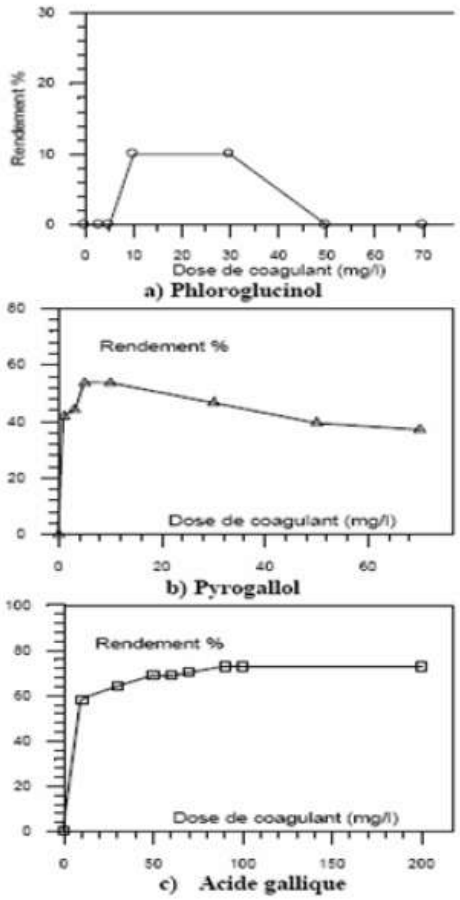


Figure I.6. Influence de la dose de coagulant sur l'élimination d'acides organiques (5mg/l)[38].

Deux grands types de mécanismes se distinguent[2,40] pour expliquer les interactions entre la matière organique et les oxo-hydroxy-métalliques solides. Il s'agit pour ce dernier mécanisme du phénomène d'adsorption qui pourra s'établir par différents types de réactions.

Le second mécanisme envisage des réactions entre la matière organique et les formes solubles hydrolysées métalliques.

La coagulation du phénol et du catéchol (20 mg/l) en eau distillée par les doses de sulfate d'aluminium introduites correspondant à l'optimum de leur élimination a été réalisée pour une gamme de pH entre de 2 et 11[41]. L'ajustement du pH a été réalisé durant la phase rapide de l'agitation. Il faut remarquer qu'à pH optimal, l'amélioration rendements est appréciable, surtout pour le catéchol. Ainsi les valeurs des rendements optima passent de 13,54% et 26,70% à pH non ajusté à 14,35% et 39,56% à pH = 7 et 9 respectivement pour le phénol et le catéchol. L'étude du paramètre pH s'avère importante du fait qu'il conditionne d'une part la dissociation des groupements fonctionnels existant dans la structure des composés organiques en fonction de leur pKa et d'autre part la spéciation du coagulant (forme solubles ou insolubles, chargées ou natures). Les résultats obtenus confirment les données

bibliographiques[38,42,43] concernant l'élimination de molécules organiques pour des pH se situent globalement entre 5 et 7 selon la structure du composé ainsi que la nature du coagulant.

I.5.2. Interactions de la matière organique naturelle avec les oxydants

L'emploi de l'oxydant est fondamental aussi bien dans le traitement des eaux souterraines que des eaux de surface. Les critères bactériologiques constituent encore un des objectifs de la chaine de traitement d'une eau destinée à la consommation. Cependant, certains oxydants chimiques et notamment le chlore peuvent intervenir sur les substances minérales et organiques des eaux[44].

Depuis une trentaine d'années, il a été établi par certains auteurs qu'une faible fraction de chlore introduit dans une eau de surface en vue de leur préoxydation ou de leur désinfection conduit à la formation des composés organohalogénès volatils et non volatils [44,45].

L'évolution globale de la matière organique en cours de la chloration présente un abattement très faible en carbone organique total (COT) inférieur à (5 à 10%), car peu de molécules subissent une oxydation totale jusqu'au stade de CO2. (MERLET, 1986 ; CROUE, 1987) En solution aqueuse, le chlore réagit sur de nombreux composés organiques, suivant des réactions d'addition, de substitution et d'oxydation pour conduire à des produits de structures très variables (DORE, 1989 ; ACHOUR, 2001 ; BACHA et al, 2006). Avec les composés aromatiques, il peut conduire à des réactions de substitution électrophiles, d'hydroxylation, de décarboxylation, de polycondensation, d'addition sur les doubles liaisons (DORE, 1989). Ces réactions conduisent à la formation des produits organohalogénès de structures très variables et en particulier les trihalométhanes (THM)[46]. La formation de THM dépend de la quantité de matières organiques présentes dans les eaux et les teneurs de ces différents composés sont très variables et dépendent essentiellement de la nature des eaux et des conditions de chloration : pH, taux de chloration, temps de contact et la température[45,46]. En solutions synthétiques et en absence de tout composé réducteur, les substances humiques présentent des consommations en chlore suffisamment importantes, de l'ordre de 0,6 à 0,8 mg Cl_2/mgSH [47].

Les consommations en chlore mesurées sur des solutions synthétiques de composés organiques simples à pH = 7, indiquent que la réactivité des composés aliphatiques, même avec des fonctions oxygénées (alcools, acides carboxyliques, cétones,…) reste limitée. Des consommations inférieures à 0,1 mole de chlore par mole de produit organique sont observées[48,49].

Cependant, certaines structures ayant des groupements fonctionnels azotés (amines, acides aminés) peuvent réagir rapidement avec le chlore en formant des chloramines organiques plus ou moins stables selon le composé et le pH du milieu[50,51]. Ainsi, les demandes en chlore indiquent des valeurs de 2 moles par fonction NH_2 présente dans la structure des acides aminés aliphatiques (alanine, aspargine)[44].

Les plus fortes demandes sont obtenues dans le cas de la tyrosine qui présente un cycle aromatique activé par un groupement donneur hydroxylé [49]. En effet, de nombreuses recherches effectuées ces dernières années confirment l'influence des substituants sur la réactivité des dérivés aromatiques [45,48,49]. Les plus forts consommateurs de chlore sont ainsi des composés aromatiques possédant des groupements actives (OH, NH_2, NR_2,…) tels que le phénol, l'aniline et leurs dérivés (5 à 12 moles de Cl_2 par mole de composé organique)[47,49].

Des travaux de coagulation-floculation sur des solutions synthétiques pour l'élimination des substances humiques en présence d'une étape préalable de chloration[52] montrent l'action oxydative du chlore sur les substances humiques produit ainsi des dérivés chlorés ou non chlorés tels les trihalométhanes (THM), des acides, des acides aromatiques ou aliphatiques. Les forts taux de chloration et le temps de réaction plus ou moins longs produisent en effet un clivage des polymères organiques en molécules plus petites ; les monomères formés peuvent être ensuite oxydés en acides mono ou dicarboxyliques. La réduction de l'efficacité de la floculation des substances humiques pourrait alors s'expliquer par la difficulté d'une partie des ions aluminium du coagulant à former des hydroxydes ou des humâtes d'aluminium insolubles. Il y'aurait donc une augmentation de la formation de complexes organoaluminiques solubles. La polarité des produits de la chloration des substances humiques pourrait également expliquer le fait que ces substances soient plus difficilement éliminables que les substances humiques dont ils dérivent. Par ailleurs, les dosages plus importants en coagulant après 24 heures de préchloration s'expliqueraient par une augmentation de la teneur en groupements carboxyliques.

Des travaux (VAN BREEMEN et al, 1979 ; LEVEBVRE et LEGUBE, 1990) [53] avaient montré en effet le rôle des groupements COOH dans la stoechiométrie de la réaction de coagulation des substances humiques par le fer ferrique ou l'aluminium.

En fait l'action du chlore ou de l'ozone conduit à une diminution des masses moléculaires de la matière organique aquatique [54].

I.6. Conclusion

La matière organique contenue dans les eaux de surface regroupe un nombre important de composés appartenant à des familles chimiques très variées. Les substances humiques représentent la fraction organique la plus importante. Ces substances sont définies comme un mélange complexe de matières organiques naturelles qui incluent dans leurs structures des chaînes aliphatiques et aromatiques portant des substituants oxygénés (COOH, OH). La présence des matières organiques d'origine naturelle est la caractéristique essentielle des eaux de surface. Les substances humiques (acide humiques et fulviques) constituent généralement la catégorie la plus importante dans ces eaux et à des concentrations supérieures à 1 mg/l. Les substances humiques représentent en moyenne 30 à 50 % du carbone organique dissous (COD) et parfois 90% dans certaines eaux colorées. Leur présence dans les eaux est gênante. Elles peuvent détériorer la qualité organoleptique de l'eau et conduisant par chloration à des composés organohalogénés souvent toxique.

La matière organique des eaux de surface constitue un milieu très hétérogène, comprenant des molécules à structure très complexe de masse moléculaire élevée mais aussi des composées organique simples.

La matière organique qui se trouve dans les eaux de surface est principalement constituée de macromolécules (acides humiques et fulviques) complexes possédant de nombreuses fonctions carboxyles (COOH), phénoliques et ou alcooliques (OH), carbonyles (C=O) ainsi que des groupements amines (NH_2), pouvant complexer les éléments minéraux.

Les processus de coagulation-floculation a fait apparaitre divers mécanismes, souvent complexes et dépendant fortement du pH de l'eau à traiter, les réactions intervenant entre des molécules organiques, micropolluant organique ou minérale et des cations utilisés en coagulation floculation, entrainent la formation de complexes moléculaires, ces complexes, une fois formés, auront un comportement différent de celui de la molécule seule.

Partie II : Coagulation et floculation

II.1. Introduction:

Le mot coagulation vient du Latin "coagulare" qui signifie "agglomérer". La couleur et la turbidité d'une eau de surface sont dues à la présence de particules de très faible diamètre (les colloïdes). Leur élimination ne peut se baser sur la simple décantation. En effet, leur vitesse de sédimentation est extrêmement faible (figure II.1). Le temps nécessaire pour parcourir 1 m en chute libre peut être de plusieurs années.

La coagulation et la floculation sont les processus qui permettent l'élimination des colloïdes. La coagulation consiste à les déstabiliser. Il s'agit de neutraliser leurs charges électrostatiques de répulsion pour permettre leur rencontre. La floculation rend compte de leur agglomération en agrégats éliminés par décantation et/ou filtration[55].

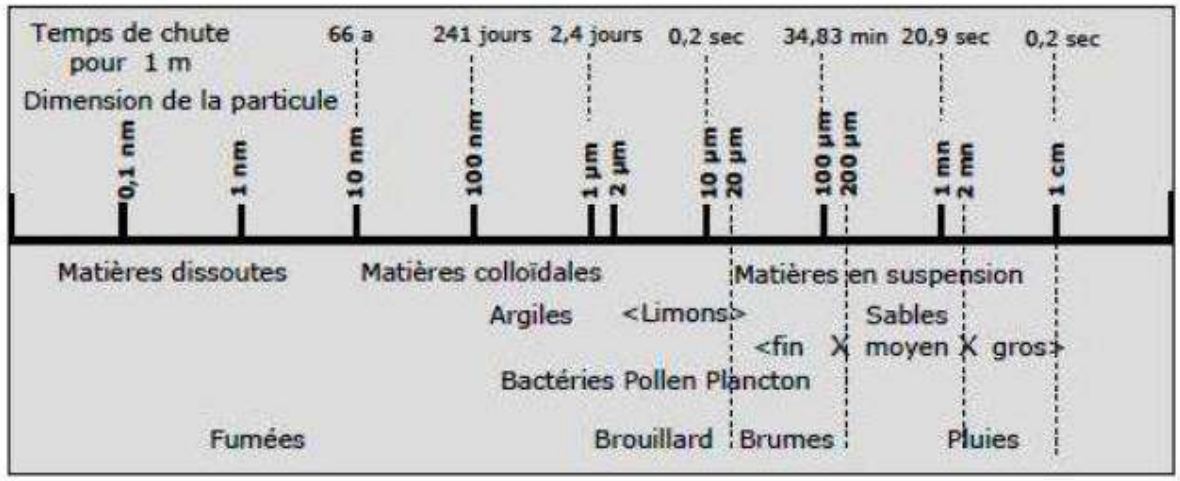


Figure II.1. Temps de décantation des particules.

II.2. But de la coagulation floculation:

L'opération de coagulation floculation a pour but la croissance des particules (qui sont essentiellement colloïdales) par déstabilisation des particules en suspension puis formation de flocs qui seront décantés et filtrés par la suite. Le principe de la coagulation est présenté dans la figure suivante[56].

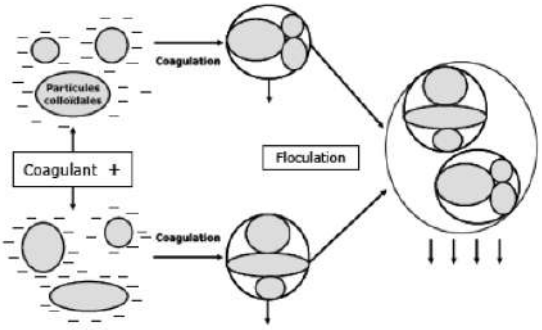


Figure II.2. La coagulation floculation.

Les principaux objectifs de la coagulation sont[57]:

- Obtention d'une turbidité minimale: Les normes OMS et Algériennes fixent la valeur maximale à 5 NTU pour les eaux désinfectées;
- Obtention d'une couleur minimale: La couleur ne peut être complètement éliminée par coagulation floculation. La valeur maximale fixée par l'OMS et Algérienne est de 15 UH (unité Hazen);
- Désinfection finale efficace: Elle est liée à l'enlèvement de la turbidité et à la valeur du pH de l'eau;
- Obtention d'un floc de bonne qualité: Le floc doit être résistant et facilement décantable. Le dosage doit être tel que la quantité de floc soit limitée de façon à éviter une surproduction de boues et un encrassage rapide du filtre;
- Obtention d'un pH final optimal: Le pH final doit être proche de la valeur d'équilibre où l'eau n'est ni agressive ni corrosive de manière à éliminer l'ajout de correctif;
- Teneur en carbone organique total (COT) minimale et élimination des précurseurs de THM (trihalométhane): Etant donné la baisse constante de la qualité des eaux brutes qui contiennent de plus en plus de micropolluants organiques, le traitement doit être aussi efficace que possible pour les éliminer et éviter, lors de la chloration, la formation de trihalométhane. Or, les substances humiques sont reconnues pour être des précurseurs de THM. La couleur ne constitue pas un bon indice de cette tendance. Par contre, la mesure du COT et l'absorption en UV reflètent la présence des acides humiques. Le COT dissous est plus difficile à éliminer. Par contre une coagulation préalable au sulfate d'aluminium améliore sensiblement l'absorption des matières organiques en solution sur charbon actif;
- Efficacité à basse température: Une simple diminution de température de 0.5 °C (entre 4.5 et 4 °C) entraîne aussi bien avec l'alun qu'avec le chlorure ferrique une augmentation notable de la turbidité de l'effluent. Si l'on augmente la dose de coagulant, une sursaturation de l'hydroxyde de métal importante, entraîne une augmentation du métal résiduel. Ceci est due à la quantité de particules de taille inférieure qui augmente à basse température, et exige un temps de complète du sel de métal;
- Autres applications: Nous avons insisté surtout sur la coagulation appliquée à la préparation de l'eau potable. Rappelons que les dérivés de l'Al et du fer sont également employés dans le traitement physico-chimique des eaux usées où les mécanismes sont

similaires avec des problèmes moins liés à la turbidité ou l'alcalinité disponibles qu'à la qualité et au volume des boues produites. Certains procédés industriels nécessitent également une coagulation efficace, en particulier dans le domaine des pâtes et papiers.

- Teneur en aluminium ou en fer résiduel minimale: Dans la grande majorité des usines, le traitement s'accompagne d'une hausse de la teneur en aluminium total qui peut être de 6 à 1500 µg/l. Or, certaines études relieraient l'absorption d'une quantité importante d'aluminium dans l'organisme humain à l'apparition de désordres neurologiques tel que la maladie d'Alzheimer, dialysis encéphalopathie, insuffisance rénale. La quantité d'aluminium ingérée chaque jour par l'homme est de l'ordre de 20 mg et la contribution de l'eau potable est inférieure à 5 %. L'aluminium résiduel doit être maintenu à une valeur minimale, évitant du même coup les post-précipitations éventuelles.
- Le pH est le facteur principal qui influence l'aluminium résiduel: La quantité d'aluminium varie 100 fois pour une variation d'une unité de pH. La solubilité d'aluminium à pH (6.5-7) est de (0.025-0.075 mg/l) comparée à la solubilité d'aluminium à pH 8 qui est de 0.5 mg/l;
- La température: Une basse température peut favoriser la solubilité de l'aluminium en ralentissant les cinétiques des réactions;
- La présence de ligand (F, SO_4^{2-}, PO_4^{2-}, substances humiques) favorise la formation de complexes d'aluminium solubles non éliminables par filtration;
- Le fer n'est pas un élément nuisible pour la santé. Au contraire c'est même un élément nutritif essentiel. On doit cependant en contrôler les concentrations, et ce pour des raisons d'esthétique. En effet la présence d'une certaine quantité de fer dans l'eau peut entraîner certains inconvénients. Cela peut:
- Donner à l'eau un goût métallique désagréable;
- Tacher les tissus lors de la lessive (taches qui vont du jaune au brun);
- Tacher les équipements sanitaires (baignoires…);
- Former un précipité dans les conduites, ce qui favorise la croissance des bactéries ferrugineuses.

II.3. La coagulation:

La turbidité et la couleur d'une eau sont principalement causées par les particules colloïdales. Ces particules, qui peuvent rester en suspension dans l'eau durant de très longues périodes,

peuvent même traverser un filtre très fin[58]. Par ailleurs, du fait de leur grande stabilité (portant des charges négatives sur leur surface, elles ont tendance à se repousser). Pour éliminer ces particules, on a recours aux procédés de coagulation et de floculation[58].

La coagulation a pour but principal de déstabiliser les particules en suspension, c'est-à-dire faciliter leur agglomération, donc déstabiliser les colloïdes en diminuant les forces de répulsion électrostatiques[58]: En pratique ce procédé est caractérisé par la neutralisation des charges négatives en ajoutant un réactif de charge positive ou coagulant.

C'est aussi l'agglomération des particules fines colloïdales en suspension et aussi des précipités obtenus par la réaction chimique des ions ferriques Fe^{3+} avec les substances dissoutes dans l'eau. Les principales réactions chimiques en causes sont les suivantes[58]:µ

$$Fe_2(SO_4)_3 + 6H_2O \longrightarrow 2Fe(OH) + 3H_2SO_4 \quad \textbf{(II.1)}$$

$$3Ca(HCO_3) + 3H_2SO_4 \longrightarrow 3CaSO_4 + 6CO_2 + 6H_2O \quad \textbf{(II.2)}$$

$$2Fe_2(SO4)_3 + 6Ca(HCO_3)_2 \longrightarrow 4Fe(OH)_3 + 6CaSO_4 + 12CO_2 \quad \textbf{(II.3)}$$

II.3.1. Principe de la coagulation:

Considérons deux particules colloïdales A et B. Entre ces deux particules, il existe des forces de répulsion électrostatiques, dues à la charge de surface, et des forces d'attraction intermoléculaires, dues aux interactions de Van der Walls. L'énergie potentielle de A et B varie donc en fonction de la distance les séparant [59–61].

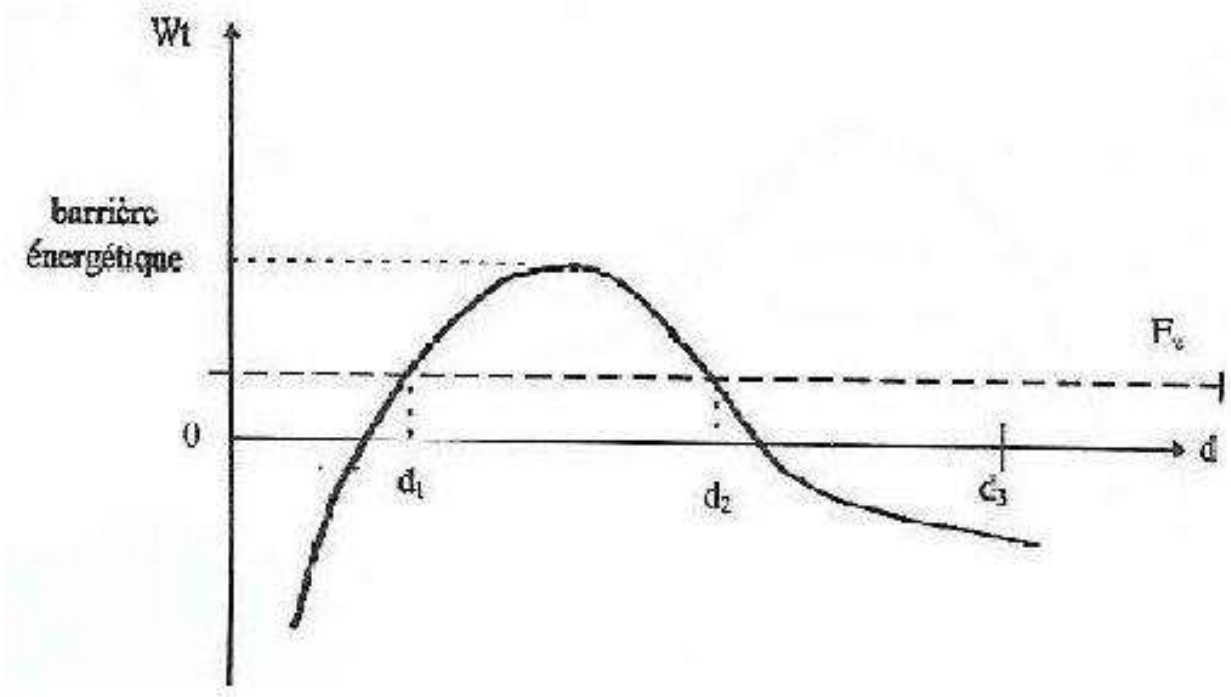


Figure II.3. La barrière énergétique.

Tableau II.1. Interaction entre deux particules en fonction de la distance les séparant.

Distance interparticulaire

Distance interpaticulaire	Interprétation
$d>d_3$	Aucune interaction
$d_3<d<d_2$	Attraction faible
$d_2<d<d_1$	Répulsion
$d<d_1$	Attraction forte, adhésion

Dans un premier temps, la coagulation, par un ajout de sels métalliques (généralement de fer ou d'aluminium), permet de supprimer les répulsions intercolloïdales: les cations métalliques (Al^{3+} et Fe^{3+}) se lient aux colloïdes et les neutralisent. Les particules colloïdales peuvent désormais se rencontrer.[61]

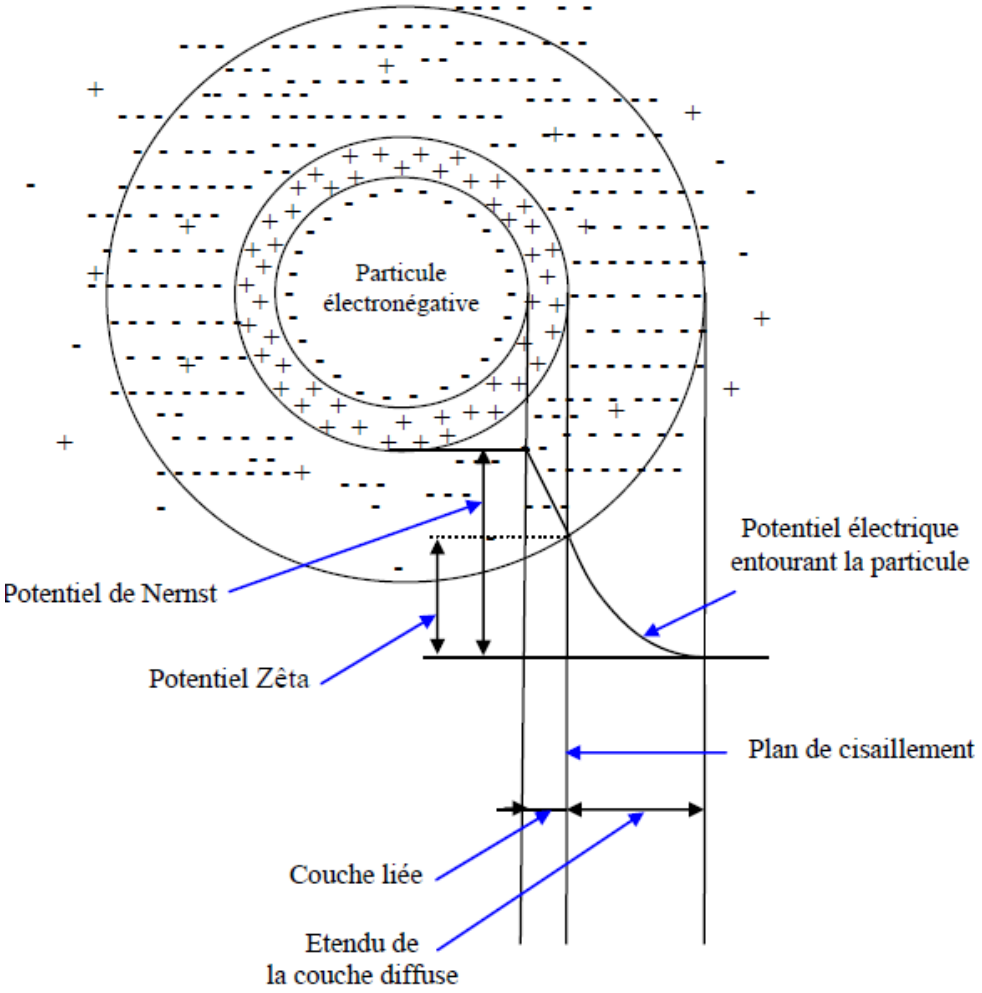


Figure II.4. Potentiels électrostatiques[52].

Il existe entre ces deux couches un potentiel électrostatique ou de NERNST, qui varie en fonction de la distance par rapport à la surface du colloïde.

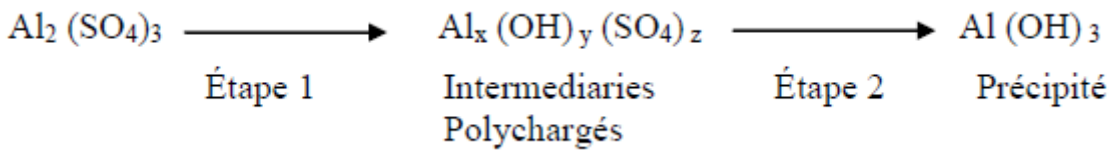
Dans la couche liée, le potentiel de NERNST décroît linéairement car les cations constitutifs sont empilés uniformément. En revanche, dans la couche de GOUY, le potentiel électrostatique varie de manière non linéaire, étant donné que la répartition ionique résulte d'un mélange aléatoire de cations et d'anions (la valeur du potentiel à la surface de la couche de NERNST est appelée potentiel zêta). Les colloïdes étant chargés négativement, ce potentiel est négatif. Dans les eaux naturelles, sa valeur varie de -30 à -35 mV[57].

On en déduit donc que l'élimination des colloïdes passe par l'annulation du potentiel zêta afin d'annuler les forces de répulsion.[62]

II.3.2. Réactifs coagulants:

L'efficacité des coagulants est directement liée à la valence des cations utilisés. Les ions monovalents sont moins efficaces que les bi ou trivalents. Il est évident que les doses de coagulant à mettre en œuvre sont fonction de leur efficacité relative[63]. Dans le cas des ions monovalents la quantité nécessaire est de 100 pour une dose de 1 pour les ions trivalents. En conséquence, les coagulants utilisés sont des sels cationiques trivalents[63].

La mise en solution d'un coagulant se déroule en deux étapes. Le cas de sulfate d'alumine est très significatif. Les réactions peuvent être représentées de la façon suivante[62]



L'étape 1: est une phase d'hydrolyse. Des intermédiaires poly chargés positifs se forment. Ces composés assez fugaces présentent un atome d'aluminium dont le nombre d'oxydation est très Conformément à la règle de SCHULZE-HARDY, ces intermédiaires poly chargés positifs sont très efficaces pour neutraliser la charge primaire négative des colloïdes. Il s'agit de la véritable forme coagulante qui déstabilise les particules chargées négativement.

L'étape1 dépend de la température et nécessite un PH compatible avec l'existence de ces intermédiaires poly chargés. Le temps de formation de ces composés est de l'ordre de 0.1 seconde.

L'étape 2 permet la formation de précipités $Al(OH)_3$. cette réaction dépend de l'agitation du milieu.

Ce précipité est l'élément qui assure le pontage et la coalescence entre les colloïdes déstabilisés: c'est la forme floculant. Tout coagulant présente successivement les deux formes

actives coagulantes et floculant. Le maintien de cette dernière dépend du PH du milieu. Cette notion de PH permet de définir les zones optimales de coagulation floculation[63].

La gamme optimale de PH pour l'utilisation du sulfate d'alumine est conditionnée par l'obtention et la persistance de la forme précipitée $Al(OH)_3$. il s'agit ici de la fourchette 5.8 à 7.2, entre ces deux valeurs le précipité est prédominent. De plus, la forme dissoute Al^{3+} est inexistante. Toute fuite de cet ion au niveau d'une filière de traitement est problématique. En effet, l'ajout d'un adjuvant de floculation pour obtenir un floc cohérent est recommandé dans le cas de fuite chronique après vérification de PH[56].

Le chlorure ferrique possède une fourchette plus large comprise entre 5.5 et 8.3 toute fuite de Fe^{3+} peut entraîner une coloration rouge de l'eau et une corrosion sur le réseau de distribution de fait de l'existence des Ferro bactéries et des phénomènes d'oxydoréduction[64].

L'utilisation de formes polymérisées de type WAC permet de s'affranchir de la première étape. Leur formule chimique est identique aux intermédiaires poly chargés. Leur prix est évidemment beaucoup plus élevé. En contrepartie, les difficultés de coagulation en hiver sont résolues.

Une autre particularité de l'addition de coagulant dans l'eau est à examiner. Tout coagulant mis en solution consomme de l'alcalinité. Cette perte de TAC est fonction de la nature et de la concentration du composé mis en jeu. L'équation d'hydrolyse du sulfate d'alumine permet d'expliquer ce phénomène[63].

$$\mathbf{Al_2(SO_4)3, 18\,H_2O + 3Ca(HCO_3)_2 \longrightarrow 18\,H_2O + 3Ca\,SO_4 + 2Al(OH)_3 + 6CO_2} \quad \mathbf{(II.4)}$$

La consommation du TAC par le coagulant et la production de CO_2 sont les éléments importants de cette réaction. Cette perte d'alcalinité se traduit par une chute de PH, donc par une acidification de l'eau, sauf si en même temps, on a un dégazage de CO_2. Le précipité d'hydroxy d'alumine se forme aux dépens de la dureté temporaire ($Ca(HCO_3)_2$) qui se transforme en dureté permanente sous forme de sulfate de calcium. Ce composé peut précipiter et donner le gypse[65].

L'acidification de l'eau qui en découle peut perturber la réaction d'hydrolyse du coagulant. La correction de PH par addition de soude, de chaux ou de calcaire dans le cas d'eaux acides et d'acide sulfurique ou de CO_2 pour les eaux incrustantes est quelquefois nécessaire. L'acidité est à vérifier avant et après la coagulation. Il s'agit de problème de l'équilibre cal co-carbonique[62].

Tableau II.2. Perte d'alcalinité et production de CO_2 de quelques coagulants. Les valeurs sont données pour 100g de produit commercial et dépendent de la qualité de l'eau brute. La perte d'alcalinité mesurée en °F[44].

Nom commercial	Perte d'alcalinité En g de CaO	Perte de TAC en °F	Production de CO_2 en g
Sulfate d'aluminium	25	4.5	40
WAC	5.4	1	8.6
Aqualence	5	9	8
Chlorure ferrique	21.1	3.8	33.8
Clairtan	20.1	3.6	31.6

II.3.3. Principaux réactifs coagulants:

Le sulfate d'aluminium, le chlorure ferrique et le polychlorure d'aluminium sont les principaux produits chimiques utilisés pour le traitement de l'eau potable. Ces produits agissent en neutralisant la charge négative des composés dissous et des particules en suspension dans l'eau au moyen de la charge positive de l'aluminium ou du fer[66].

II.3.3.1. Les coagulants minéraux:

Les sels d'aluminium III et de fer III sont largement utilisés. Les plus courants sont [62]:

- Les sulfates d'aluminium et de fer II : $Al_2(SO_4)_3$ et $Fe_2(SO_4)_3$.
- Les chlorures d'aluminium et de fer II : $AlCl_3$ et $FeCl_3$.
- Les mélanges de sulfates et de chaux : $M2(SO_4)_3 + Ca(OH)_2$.
- Les produits mixtes Al^{3+}/Fe^{3+}

La réaction de base lors de la mise en solution des coagulants minéraux est une hydrolyse qui modifie les caractéristiques physico-chimiques de l'eau à traiter (pH, conductivité)[63].

$M_3^+ + 3H_2O <=> M(OH)_3 Precip + 3H^+$

Le problème posé par les coagulants minéraux vient de la formation d'hydroxydes précipités qui entraîne la production d'un volume de boue important. Ces boues doivent être éliminées ultérieurement dans le processus de séparation liquide-solide[67].

II.3.3.2. Les coagulants organiques de synthèse:

Ce sont des molécules organiques de synthèse à caractère cationique. Ces produits neutralisent les colloïdes négatifs et sont utilisables directement, en remplacement total ou partiel d'un coagulant minéral[62].

Ils présentent l'avantage de réduire considérablement le volume de boue produite. Mais, dans certains cas, ils sont moins efficaces. Donc, l'utilisation conjointe des deux types de coagulant

permet une très nette diminution de la quantité de coagulant minéral nécessaire (40 à 80 %) tout en assurant un moindre volume de boue produite[68].

II.3.4. Avantages et inconvénients de divers coagulants chimiques:

Tableau II.3. Avantages et inconvénients de divers coagulants chimiques[69].

Produits	Avantages	Inconvénients
Chlorure ferrique ou sulfate ferrique	- Le meilleur pour éliminer les matières organiques. - Bas prix. - N'ajoute pas d'aluminium à l'eau. - Les concentrations élevées de résidus de fer sont faciles à détecter.	- Un mauvais dosage produits une coloration rougeâtre et une concentration élevée de résidus de fer. - La forte concentration de résidus de fer risque de boucher les filtres et de tacher la lessive. - Très acide (la porte d'un équipement - de protection complet est de rigueur) - boue inesthétique (brune)
Sulfate d'aluminum	- relativement peu couteux - moins acide (il faut quand même porter un équipement de protection)	-Elimination du COD réduite. -les erreurs de dosage produisent une grande quantité de résidus d'aluminium. posant un risque potentiel pour la santé.
Polychlorure d'aluminium	- Taux de résidus plus faible que le sulfate d'aluminium. - Moins d'impact sur le pH et l'alcalinité.	- Coute trois fois plus cher que les autres coagulants.

II.4. La floculation:

Après avoir été déstabilisées, les particules colloïdales ont tendance à s'agglomérer lorsqu'elles entrent en contact les unes avec les autres. Le taux d'agglomération des particules dépend de la probabilité des contacts et de l'efficacité de ces derniers. La floculation a justement pour but d'augmenter la probabilité de rencontre entre les particules grâce à l'agitation lente du fluide (accroître le volume, le poids et la cohésion du floc)[70].

La floculation consiste à alourdir les flocs déjà formés par coagulation qu'on pourra facilement les éliminer par décantation. Le réactif, floculant, ou adjuvant de floculation,

injecté pour former des " ponts" entre les flocs. Le floculant est en général un polymère d'origine naturelle, minérale (silice activée) ou organique (alginates ou amidon). Cette étape qui se place donc après la coagulation, nécessite un milieu où l'agitation est très faible pour ne pas casser les flocs[71].

II.4.1. Types de Floculation:

II.4.1.1. Floculation péricinétique:

Durant ce type de floculation, le mouvement brownien des particules est fonction de la température de I'eau. Lorsqu'elles entrent en contact, les particules ont tendance à s'agglutiner les unes aux autres. Le nombre total de particules varie donc selon l'équation suivante[72]:

$$\frac{dN}{dt} = -4\ \frac{\Gamma k\theta (N)^2}{3\eta} \qquad \textbf{(II.5)}$$

Où:

N: nombre de particules par m^3.

Γ: facteur tenant en compte l'efficacité des collisions.

κ: constante de Boltzmann (1.38x10-23 J*k-1)

θ: température (k).

η: viscosité dynamique de l'eau (Pa*s).

Le taux de variation du nombre de particules, lors de la floculation péricinétique est important lorsque la taille des particules est inférieure à 1 micron. Au-delà de cette taille, d'autres mécanismes doivent être utilisés afin d'assurer un statut d'agrégation des particules raisonnable.

II. 4.1.2.Floculation orthocinétique:

Ce type de floculation est fonction de l'énergie dissipée dans le fluide. Elle est donc causée par le brassage de l'eau qui permet d'augmenter les chances de contacts entre les particules. La vitesse du fluide varie dans l'espace et dans le temps et elle est caractérisée par le gradient de vitesse G. L'efficacité du processus est exprimée selon l'équation suivante lorsque le régime est turbulent [73]:

$$\frac{dN}{dt} = -\frac{2}{3}\ \Gamma G d^3 N^2 \qquad \textbf{(II.6)}$$

Où:

G: gradient de vitesse (s^{-1}).

d: diamètre des particules (m).

Le gradient de vitesse est défini selon l'équation [74]:

$$G = \sqrt{\frac{P}{\mu V}} \quad \textbf{(II.7)}$$

Où:

P: puissance réellement dissipée (w).

V: volume du bassin (m^3).

μ: viscosité dynamique de l'eau (kg/m*s).

Selon le degré de floculation désiré, une agitation plus ou moins énergétique doit être effectuée[60].

II.4.2. Les adjuvants de floculation ou floculants:

Le floc qui se forme par l'agglomération de plusieurs colloïdes peut ne pas être suffisamment large pour se décanter ou pour se déshydrater à la vitesse souhaitée. L'utilisation d'un floculant est alors nécessaire[75]. Il rassemble toutes les particules de floc dans un filet, construisant un pont d'une surface à l'autre et liant chaque particule pour former de vastes agglomérats.

Le temps à respecter entre les ajouts de coagulant et de floculant est primordial. En effet, un floculant n'est en général efficace que lorsque la phase de coagulation est achevée. Pour choisir les floculants, il faut considérer la taille, la cohésion et la vitesse de décantation du floc[76]. Comme pour les coagulants, le taux de traitement à mettre en œuvre est donné par un essai de floculation.

On utilise[77]:

- Des polymères minéraux tels que la silice activée (SiO_2), généralement associée au sulfate d'aluminium en eau froide;
- Des polymères naturels extraits de substances animales ou végétales: amidons, alginates (obtenus à partir d'algues marines);
- Des polymères de synthèse apparus plus récemment qui ont fait évoluer considérablement les performances de la floculation. Ils conduisent souvent à un volume de boue très inférieur.

II.4. Effet de la coagulation sur les paramètres de l'eau:

1) La turbidité:

L'élimination de la turbidité est très importante. L'élimination de gros colloïdes minéraux, ceux-là même qui sont responsables de la turbidité est le domaine de prédilection de la coagulation[76].

2) La couleur:

La coagulation est aussi l'élimination de la couleur, c'est-à-dire des matières organiques à l'état colloïdal et dissous et les micropolluants organiques[78].

3) La résistivité:

Elle baisse légèrement à cause des ions sulfatés ajoutés avec l'aluminium dans le cas du traitement avec le sulfate d'aluminium et des polymères d'aluminium (sulfates)[70].

4) Le pH:

Le pH est plus ou moins affecté selon le type de coagulant utilisé et sa concentration. Quand le sulfate d'aluminium $Al_2(SO_4)_3$ est utilisé comme coagulant et est additionné à l'eau, il se déclenche une hydrolyse qui entraîne une baisse sensible du pH. Généralement, chaque Al3+ nécessite pour passer à l'état de l'hydroxyde d'aluminium ($Al(OH)_3$), $3OH^-$ provenant des bicarbonates et de l'eau elle-même. Les protons H^+ sont ainsi libérés[44]. Le T.A.C baisse donc également. Par contre, l'utilisation des polymères basiques d'aluminium comme coagulant n'entraîne qu'une très faible variation du pH qui n'affecte en rien la déstabilisation[62].

5) La silice, les phosphates et le fluor:

La coagulation élimine, en partie, la silice par suite de la formation de silicates due à l'affinité de la silice pour l'aluminium. Les phosphates et le fluor sont considérablement diminués par suite de la forte attirance qu'ils ont pour l'aluminium[72].

II.5. Facteurs intervenant dans la coagulation des particules en suspension:

A/ La valence des colloïdes:

Selon la nature des particules colloïdales et leur interaction vers le milieu aqueux, c'est-à-dire leur faculté d'échange des cations périphériques selon qu'elle s'avère forte ou faible, elle favorise plus ou moins le processus de coagulation[76].

B/ Le potentiel Zêta:

Certaines forces attractives existent entre tous les types de particules colloïdales quelles que soit leur nature chimique. Ces forces attractives (forces de Van der Waals), sont responsables de l'agrégation de beaucoup de systèmes colloïdaux. Leur ampleur dépend de la nature des atomes qui forment les particules colloïdales et de leur densité[44]. Ces forces attractives diminuent avec l'augmentation de la distance séparant les particules, ainsi que l'énergie potentielle attractive qui en résulte. Les effets de charge sur la stabilité apparente des colloïdes irréversibles doivent être déterminés par la somme des énergies attractives et répulsives de l'interaction. Cette nette répulsion peut être considérée comme une barrière d'énergie d'activation qui doit être surmontée pour que l'agrégation ait lieu[62]. Ses amplitudes

dépendant de la charge sur les particules et de la composition ionique de la solution. A des forces ioniques élevées, cette barrière d'énergie peut disparaître. Le taux de coagulation dépend de la grandeur de la barrière d'énergie et de l'énergie cinétique qui anime les particules colloïdales. Les systèmes colloïdaux qui ont une haute énergie d'activation et une faible énergie cinétique coagulent très lentement (colloïdes persistants) par contre les colloïdes avec une faibles énergie d'activation et/ou une haute énergie cinétique coagulent rapidement (systèmes caducs)[76].

C/ Grosseurs et concentration des particules colloïdales:

Il existe un seul propice, pour la coagulation, du point de vue dimensionnel. Pour les particules dont les dimensions sont inférieures au micron, les forces de surface deviennent prépondérantes vis-à-vis des forces de masse. Dans ces conditions, on constate un état de dispersion stable dû aux effets conjugués du mouvement brownien[62]. De même, il s'avère que la concentration des particules, c'est-à-dire le nombre de charges électriques à mettre en jeu pour rompre le potentiel Zêta à une grande influence sur la coagulation. Cette coagulation est très difficile quand la concentration colloïdale est basse, du moment que le taux de contact interparticulaire se trouve probablement baissé[71].

D/ L'état d'hydratation:

L'hydratation est un phénomène complexe dont les effets se répercutent sur le taux d'agrégation des colloïdes[44]. Dans une solution aqueuse, les particules microscopiques s'hydratent et s'enveloppent de liquide inter-micellaire. Cette monocouche d'eau absorbée chimiquement sur les surfaces des particules (ex: les oxydes métalliques, les argiles et le quartz) produit l'effet d'une gaine isolante protectrice, s'opposant à l'agrégation des particules, par réduction de la tension superficielle à l'interface particule-liquide. Par conséquent, la stabilité des colloïdes se trouve renforcée[62]. Ce phénomène peut être contrarié par l'augmentation de la quantité de coagulant et le prolongement de l'opération[63].

E/ Electrisation des particules:

Comme il a été souligné précédemment, toutes les particules sont ionisées. Cette électrisation est due soit aux constituants superficiels par adsorption préférentielle soit par combinaison de la particule avec le solvant suivie d'une dissociation de cette combinaison. La charge électrique agit par modification de la tension superficielle. Si cette tension est positive, les particules se dispersent[62]. En réalité, l'électrisation n'est pas la seule cause de dispersion et ne fait que superposer son effet à ceux des autres agents. Mais elle s'avère un facteur prépondérant du moment que son abaissement au-dessous d'un certain point critique provoque ou du moins accélère la coagulation[71].

F/ Concentration du coagulant:

La détermination de la concentration du coagulant est un paramètre essentiel pour la déstabilisation des colloïdes. Une surdose du polymère produit inévitablement une restabilisation des particules colloïdales qui sont susceptibles de changer de charge de surface. Aussi les sites d'adsorption des particules colloïdales ne seraient plus disponibles pour la formation des ponts interparticulaires[69]. Un traitement de coagulation mal ajusté peut donc conduire à une dégradation rapide de la qualité de l'eau et à des dépenses d'exploitation non justifiées. Bien souvent, la dose de coagulant conditionne le fonctionnement des ouvrages de séparation et il se trouve tout simplement impossible de réaliser la clarification si cette dose est mal ajustée[72].

G/-Vitesse et temps d'agitation:

Lors de l'introduction du coagulant dans l'eau, on procède à une agitation rapide pour la dispersion et l'homogénéisation de la solution. Mais une agitation intense empêche l'agrégation des particules, tandis qu'une agitation prolongée conduit systématiquement à la restabilisation du système agrégé, l'effraction des liaisons entre coagulant et la surface des particules, et le rabattement ultérieur des segments étendus sur la surface des particules[79].

H/ La température:

La température est un facteur physique important, vu ses effets sur la coagulation et son rôle sur la floculation. Toute diminution de la température suivie d'une augmentation de la viscosité et par conséquent le taux de formation des flocs se trouve réduit. Aussi il est bien connu qu'en hiver le floc est plus fragile et moins dense[79].

De plus, la température joue probablement un rôle dans la diminution du taux de réaction chimique. De même, la valeur du pH décroît avec la diminution de la température et ce changement devient plus important avec les petites doses de coagulant. Ces difficultés dues aux basses températures peuvent être généralement vaincues par une conduite de la coagulation à un pH voisin, autant que possible du pH optimum. Ceci peut être obtenu par l'augmentation de la dose de coagulant, qui non seulement accroît la probabilité de collision des particules mais réduit l'effet du changement du pH optimum[80].

I/ Le potentiel d'hydrogène pH:

Le pH est l'un des paramètres les plus importants qu'il faut noter du fait qu'il joue un rôle déterminant dans la coagulation. Il influence le signe de l'électrisation des colloïdes suivant qu'il est acide ou basique. Les flocons d'hydroxyde de fer se forment dans des limites de pH assez larges entre 4 et 11. Pour les hydroxydes d'aluminium, les limites sont plus étroites.

En pratique, le pH optimum dépend en premier lieu des colloïdes à éliminer. Les matières organiques acides sont mieux éliminées à un pH bas, tandis que les substances basiques sont mieux coagulées à un pH élevé[79,80].

J/ Rôle des bicarbonates dans l'eau:

Pendant la coagulation, il est important que les bicarbonates soient en proportion suffisante dans l'eau pour que les sels d'aluminium s'exercent de façon optimum en formant des hydrates d'aluminium sous forme colloïdale[62,81]. La formation d'ions complexes mobilise un certain nombre d'ions hydroxydes OH^- et fait entrer en jeu l'effet tampon des bicarbonates présents. L'ydrolyse conduit donc à un abaissement de l'alcalinité. L'effet tampon ne peut par nature même neutraliser entièrement le phénomène qui la déclenche, il y a également un abaissement du pH et une augmentation de la teneur en gaz carbonique dissous[62,81].

$$H_2O \rightleftharpoons H^+ + OH^-$$
$$CO_3H^- + 2H^+ \rightleftharpoons CO_2 + H_2O \qquad \textbf{(II.8)}$$

L'alcalinité bicarbonatée peut être corrigée par neutralisation et addition de CaO mais il faut se garder d'un excès de CaO qui provoquerait ou renforcerait les charges négatives[65].

II.5. Paramètres influençant la coagulation floculation:

Le processus de coagulation est complexe et peut être influencé par plusieurs facteurs. Certains sont liés à la qualité de l'eau brute et d'autres au traitement lui-même.

II.5.1. Facteurs liés à la qualité de l'eau brute:

II.5.1.1. La nature de la couleur et de la turbidité et leur proportion relative dans l'eau brute:

Rappelons qu'il n'y a pas de solution de continuité entre les mesures de couleur et de turbidité des eaux, et que seules les tailles des particules et dans une certaine mesure leur nature, permettent d'établir une classification arbitraire[82].

Les particules de grande taille et de nature généralement minérale (argiles) constituent la turbidité. Les petites particules et certains composés solubles de nature généralement organique sont responsables de la couleur[81].

La floculation des acides humiques exige des doses élevées de coagulant approximativement proportionnelles à leur concentration dans l'eau. Au contraire, les colloïdes minéraux sont facilement floculés par de faibles doses de coagulant, sans rapport avec l'intensité de la turbidité[82,83].

Des études réalisées sur la coagulation floculation de suspensions colloïdales affirment que[84]:

- La nature de la suspension influe sur la turbidité résiduelle et le taux de sulfate d'aluminium à l'optimum de floculation, aussi il existe une relation stoechiométrique entre la quantité des colloïdes organiques et la dose du coagulant consommé;
- L'ajout d'une faible quantité d'argile permet d'améliorer le rendement d'élimination des substances humiques;
- La structure minéralogique et les propriétés surfaciques des particules affectent la dose du coagulant;
- L'élimination de la matière organique est liée directement à son poids moléculaire, aux nombre et degré d'ionisation du groupement fonctionnel;
- Dans le cas de la déstabilisation par neutralisation de charge, la dose du coagulant est liée à:
- La solubilité de l'hydroxyde de l'aluminium;
- La densité de charge des particules primaires;
- La surface spécifique des particules primaires.

II.5.1.2. La température:

Des études réalisées sur l'influence de la température sur le processus de coagulation floculation montrent que l'abaissement de la température entraîne[85]:

- Une augmentation de la concentration en sulfate d'aluminium à l'optimum de floculation;
- Une décroissance de la taille des flocs;
- Un déplacement du pH optimum de floculation vers les pH basiques;
- Une inhibition des mécanismes de floculation pour des températures inférieures à 10°C, en particulier, pour les eaux peu turbides.

DESJARDINS (1996), montre que la température affecte de manière significative l'abattement de la matière organique par le sulfate d'aluminium[83].

Comme toute réaction chimique, la floculation est accélérée par l'élévation de la température. La floculation des eaux froides est donc plus lente et moins complète que celle des eaux chaudes.

D'autre part, la rapide augmentation de la viscosité de l'eau lorsque la température s'abaisse (la valeur de la viscosité passe de 1 à 2 entre 25 et 0 °C), joue contre la fréquence des rencontres des particules dans la phase de floculation orthocinétique et diminue le nombre de chocs efficaces. Enfin, cette viscosité entrave le phénomène de décantation[83,86]. Nous savons qu'une baisse de la température ralentit la cinétique des réactions chimiques. Les premières

études sur la coagulation floculation en eau froide remontent à plusieurs années. En effet, Camp et al (1940) expliquent le problème en eau froide de deux façons[87] :

1) La solubilité des coagulants est plus faible;

2) L'augmentation de la viscosité de l'eau freine la décantation des flocs.

Des études ont été faites afin de déterminer le coagulant qui est le moins affecté par une baisse de la température. Morris et Knocke (1984) comparent l'effet de la température sur les performances du chlorure ferrique $FeCl_3$ et du chlorure d'aluminium $AlCl_3$[88].

L'évaluation des produits est basée sur la cinétique des réactions, les caractéristiques des flocs formés et l'enlèvement de la turbidité. Ils concluent que les cinétiques de réaction sont plus faibles en eaux froides et que leur meilleur enlèvement de la turbidité est réalisé par le $FeCl_3$ comparé au $AlCl_3$. De plus, ils remarquent que les flocs formés à partir du chlorure d'aluminium sont plus petits et plus difficilement décantables[88]. Par contre, le taux de précipitation n'est pas affecté par la température. Ils ont donc proposé d'utiliser le chlorure ferrique $FeCl_3$ comme coagulant primaire ou aide-coagulant. Il est à noter que les conditions des tests au laboratoire simulent une coagulation floculation classique avec un mélange rapide de 1 minute suivie d'un mélange lent de 20 minutes et d'une période de décantation d'une heure[89].

Une étude approfondie de l'effet de la température sur la floculation a été faite par Mohtadi et Rao (1973). Les variations du potentiel Zêta sont analysées de même que les taux de formation de flocs, les doses des produits chimiques ainsi que les pH optimaux[90].

Dans cette étude, les coagulants utilisés sont l'alun et des polyélectrolytes cationiques (Purifloc 32-C), non ioniques (Magnifloc 900N) et anioniques (Magnifloc 870A). Les résultats montrent que dans les conditions des essais (glaises kaolines et bentonite à des températures entre 1°C et 20°C) les potentiels zêta ne varient pas avec la température[90].

Pour l'alun, l'impact d'une baisse de la température est compensé par une augmentation de la dose d'alun ou par une augmentation du pH (nouvelle zone optimale de pH 6,6 à 6,8). Les résultats montrent également que le taux de formation des flocs est dépendant de la température dans le cas de l'alun, par contre lors de l'utilisation d'un polyélectrolyte cationique, la température n'a pas d'effet sur le taux de formation des flocs. Les auteurs montrent aussi que les polyélectrolytes anioniques et nonioniques ne réussissent pas à provoquer une floculation des particules chargées négativement. Ce résultat va à l'encontre de ce qu'ont démontré Gan et al. (1991)[91].

Une étude (Kank et al, 1995) tente de déterminer les effets de la température sur la coagulation et la floculation d'une glaise kaoline par le nitrate ferrique. Il a d'abord été montré

que l'augmentation du pH de 6 à 8 fait augmenter le taux de floculation. Ce taux peut aussi être amélioré en augmentant l'intensité du mélange lent. L'effet de la température sur le pH optimal est bien illustré. Ce dernier passe de 6,8 à 23°C à 7,5 à 5°C afin que le nombre d'ions OH- soit maintenu[58].

Le gradient de vitesse, donné par la relation ci-après, est l'un des paramètres agissant sur la probabilité de rencontre des particules dans le processus de floculation. Il dépend de la viscosité dynamique de l'eau, et donc de sa température[58].

$$G = \sqrt{\frac{P}{VX\eta}} \qquad \textbf{(II.9)}$$

G: Gradient de vitesse (s^{-1}).

P: Puissance dissipée (Watt).

V: Volume du bassin (m3).

η: Viscosité dynamique (Pa.s).

II.5.1.3. L'influence du pH:

Le rôle du pH est particulièrement important dans la mesure où il détermine à la fois la charge des colloïdes et la nature des espèces qui naissent de sels minéraux d'Al et de Fe. Pour chaque eau il existe une plage de pH pour laquelle la coagulation a lieu rapidement, plage qui est fonction du coagulant utilisé, de sa concentration et de la composition de l'eau à traiter[83]. Les coagulants inorganiques changent les propriétés physico-chimiques (pH, conductivité,..) de l'eau à traiter en raison de leur hydrolyse.

$$\mathbf{M^{3+} + 3H_2O <=> M(OH)_3 + 3H^+} \qquad \textbf{(II.10)}$$

Le pH nécessaire pour la coagulation peut être ajusté en ajoutant une base ou un acide.

Tableau II.4. pH optimum pour la coagulation floculation[62].

Cation	pH optimum pour la coagulation folculation
AL^{3+}	6.0 -7.4
Fe^{2+}	>5

Le pH est l'un des paramètres qui affecte le plus la coagulation floculation. Son contrôle est donc suggéré. Selon les caractéristiques d'une eau, il existe un pH optimal permettant la meilleure coagulation (Julien, 1983). Ce pH se situe souvent dans la plage où la solubilité du coagulant utilisé est minimale, ce qui permet une meilleure précipitation. Le contrôle du pH permet aussi d'améliorer la coagulation lorsque la température de I'eau est faible[61]. Dans la plupart des usines au Québec on utilise plutôt un sel d'aluminium (l'alun). À pH élevé, les concentrations d'aluminium dissout augmentent considérablement dans l'eau du robinet, ce

qui est incompatible avec de l'eau potable. Le sulfate ferrique travaille très bien à des pH variant de 5,5 à 11,5. Le sel de fer offre définitivement plus de flexibilité. Il réagit beaucoup mieux et l'hydroxyde de fer produit est quasi insensible aux variations de pH dans la plage de pH utilisée. Même en ajoutant 5 à 7 mg/l de fer en provenance du sulfate ferrique, on obtient une eau traitée ayant des teneurs inférieures à 0,1 mg/l de fer, la norme esthétique maximale étant plus petite ou égale à 0,3 mg/l[79].

Aux valeurs habituelles de pH rencontrées dans les eaux naturelles, les ions OH^- libres n'existant pas, ils résulteront de la décomposition des bicarbonates (HCO_3^-) par le floculant (et formation de gaz carbonique CO_2)[81]:

$$HCO_3^- \longrightarrow CO_2 + OH^- \qquad \textbf{(II. 11)}$$

Réaction à laquelle se superpose celle de l'équilibre du bicarbonate de calcium avec ses produits de dissociation. L'équation globale type de coagulation peut donc s'écrire [81]:

$$X^{3+} + 3HCO_3^- \longrightarrow X(OH)_3 + 3CO_2 \qquad \textbf{(II.12)}$$

Donc, précipitation d'une hydroxyde métallique $X(OH)_3$ et formation de CO_2[65].

Rappelons que

$$CO_2 + H_2O \longleftrightarrow H_2CO_3 \text{ (acide carbonique)} \qquad \textbf{(II.13)}$$

L'apparition d'acide carbonique et la diminution subséquente de l'alcalinité bicarbonatée conduit à un abaissement du pH du milieu, dont l'ampleur dépend du pouvoir tampon (alcalinité initiale et force ionique) et de la dose de coagulant introduite. C'est pour cette raison qu'il est essentiel de tenir compte du pH, de l'alcalinité et de la minéralisation de l'eau brute pour la définition du procédé de coagulation floculation[65].

Et voici un exemple avec l'utilisation de l'aluminium comme coagulant[62]:

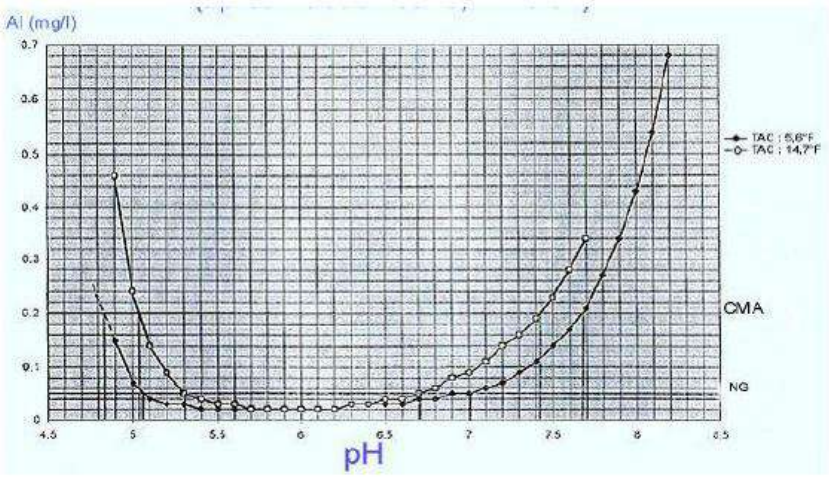


Figure II.5. Solubilité d'Al en fonction du pH et du TAC (après filtration sur 0.2 micron).

Il peut être nécessaire, pour obtenir une floculation correcte, d'ajuster le pH de l'eau en y ajoutant soit un acide, soit une base (soude, chaux, carbonate de sodium).

II.5.1.4. Effets de l'alcalinité:

La capacité de l'eau à neutraliser des acides. Cette propriété dépend entre autres de la concentration en carbonate, bicarbonate et hydroxydes de l'eau. L'alcalinité se mesure en milligramme par litre équivalent carbonate de calcium (mg/l $CaCO_3$)[89].

L'alcalinité naturelle de I'eau réagit avec l'alun pour former le précipité d'hydroxyde d'aluminium. Il faut donc un minimum d'alcalinité pour que l'alun agisse correctement.

Ce minimum permet aussi de stabiliser le pH dans la plage optimale au cours de la coagulation (Julien, 1983). Dans les eaux de faibles alcalinités, des coagulants préhydrolisés peuvent s'avérer avantageux car ils en consomment moins[92].

II.5.1.5. Effets des sels dissous:

Les sels dissous dans une eau exercent les influences suivantes sur la coagulation floculation[84] :

- Modification de la plage de pH optimal;
- Modification du temps requis pour la floculation;
- Modification de la quantité de coagulant requis;
- Modification de la quantité résiduelle de coagulant dans l'effluent.

La coagulation à l'aide des sels d'aluminium ou de fer est davantage affectée par les anions que par les cations. Exemple: SO_4^{2-}, PO_4^{2-}. Les cations Na^+, Mg^{2+}, Ca^{2+} exercent une influence minime sur la coagulation.

II.5.1.6. L'influence de la minéralisation:

L'influence de la minéralisation totale ou de teneur en certains cations ou anions a été abordée par différents auteurs, il en résulte que[81]:

Les anions et cations ont une double action[81]:

- Sur l'environnement des colloïdes: L'augmentation de la teneur en ions entraîne une diminution des forces de répulsion interparticulaire par compression de la double couche;
- Sur la formation des produits d'hydrolyse des sels de fer ou d'aluminium.
- L'addition des ions sulfates (SO_4^{2-}) élargissait la plage de pH optimal vers l'acidité et favorise la formation de composés neutres ou chargés négativement, alors que l'addition d'ions phosphates (PO_4^{3-}) déplaçait sans élargir la plage de pH optimal et

que les ions bicarbonates avaient un effet sur la précipitation des hydroxydes d'aluminium.

ZOUGRANA (1993), montre que l'alcalinité de l'eau brute joue un rôle important dans le processus de coagulation floculation et l'enlèvement du carbone organique total (COT)[11].

PENG (1994), a mis en évidence que les sels de calcium et ceux d'aluminium affectent la floculation de suspension de kaolin avec du polymère anionique[83].

II.5.2. Facteurs liés au traitement:

II.5.2.1. La nature et le dosage du coagulant:

Même si l'alun est le coagulant le plus utilisé, il peut être avantageux de le remplacer par un sel de fer ou par des P.A.C ou P.A.S selon les caractéristiques de l'eau à traiter. On ne peut choisir un coagulant et en déterminer la concentration optimale qu'après essai au laboratoire vu la complexité du phénomène[93].

II.5.2.2. Les conditions d'agitation:

Le processus de coagulation floculation se déroule généralement en deux étapes:

- Une étape de mise en contact entre l'eau à traiter et le coagulant. Cette étape se déroule généralement sous une agitation rapide. Son but est également de permettre la neutralisation de la charge des particules et d'amorcer le processus de floculation par la formation de microflocs. LETTERMAN (1973), montre que la turbidité résiduelle dépend du temps de mélange rapide et pour chaque combinaison de gradient de vitesse et dose du coagulant injecté; il existe un temps de mélange rapide associé au minimum de turbidité[94]. FRANCESHI (1991), indique qu'il existe une valeur optimale de gradient de vitesse lors de la phase d'agitation rapide pour laquelle la turbidité résiduelle est minimale;
- Une étape d'agitation lente qui assure la mise en contact et la croissance des flocs, CAMP et STEIN proposent que le critère GT (dit critère de Camp) soit retenu pour le bon contrôle du bon fonctionnement d'un floculateur avec $20<G<40\ s^{-1}$, MLYANAMI (1982), confirme ce résultat dans le cas d'agitation par rotation et par vibration. KELLIL (1989), a montré qu'il existe un maximum pour la taille de floc de l'ordre de 500 μm qui correspond à une valeur de GT=104:
- Pour les faibles valeurs de GT, le mécanisme de croissance prédomine;
- Pour GT>104 le mécanisme de cassage de flocs est plus important;
- Plus la valeur de GT est grande plus le cassage du floc est important.

L'injection du coagulant doit être extrêmement énergétique pour qu'il se disperse rapidement et uniformément dans toute la masse d'eau. Ceci est dû à:

-L'importance du contact entre les particules et le coagulant évitant la déperdition de ce dernier dans l'eau.

-L'instabilité de certaines espèces (en particulier les espèces polycationiques). Les usines de traitement doivent être dotées de véritables mélangeurs rapides.

II.5.2.3. La nature et le dosage de l'aide coagulant:

L'introduction des adjuvants pour améliorer la coagulation se fait surtout dans les cas suivants[55,95]:

- Eau de faible turbidité.
- Eau de faible alcalinité.
- Eau froide.

Ces adjuvants doivent être injectés sous mélange énergétique avant ou après le coagulant. On peut citer: l'aluminate de sodium (pour une eau de faible alcalinité), la silice activée et les polyélectrolytes synthétiques.

II.6. Conclusion:

La coagulation est une opération essentielle dans l'étape de clarification des eaux de surface. Une méthode de traitement de l'eau sûre et efficace, qui améliore la qualité de l'eau en abaissant la teneur en composés organiques, en phosphore dissous, en fer et en particules en suspension et en diminuant la coloration de l'eau. Elle est aussi très complexe car elle dépend de nombreux facteurs afin d'obtenir une eau prête à être distribuée au consommateur. Pour cela les opérateurs d'usines doivent en particulier:

- Surveiller en continu les évolutions des caractéristiques physico-chimiques de l'eau brute, et donc posséder des équipements de mesure et d'analyse des différents paramètres influençant sur la coagulation;
- Conduire des tests (Jar Test) suivant une procédure rigoureuse et maîtrisée, en particulier au niveau du maintien de la température à la valeur de celle de l'eau brute à tester, de l'élimination des risques de contamination des échantillons par contact avec des accessoires en aluminium, de la conservation des échantillons … ;
- Disposer des installations de préparation et d'injection de réactifs souples permettant de remplacer le sulfate d'aluminium par un autre coagulant, tout au moins durant certaines périodes critiques, le sulfate d'aluminium restant le réactif le plus économique d'utilisation.

Méthodes et matériels

Méthode et matériels

Introduction:

Le présent chapitre est consacré à la description des méthodes et des moyens expérimentaux utilisés durant notre travail.

La 1éme partie a été consisté à étudier l'élimination des micropolluants organiques par coagulation basant sur turbidité et étudié les interactions physicochimiques avec la dose de coagulant. Pour cela, nous avons utilisées la méthode de plan d'expérience pour étudier les interactions physicochimiques avec les doses de coagulant et aussi trouver un modèle mathématique afin de faciliter notre travail. Alors, nous avons optimisé trois méthodes (RNA, SVM, ANFIS) à base des données construites par réponse de surface méthodologie (RSM).

La 2éme partie consistait en les caractérisations des propriétés physico-chimiques et autres éléments principaux de l'eau brute après une première caractérisation pour trouver un site très pollué des différentes sources de Médéa. Pour cela, les eaux brutes ont été caractérisées par différentes méthodes d'analyse. Après, nous avons appliqué le modèle mathématique de la partie 1 dans la station ADE afin de fixer les doses de coagulant. A la fin, nous avons caractérisé l'eau traitée afin de la comparer avec l'eau brute et voir l'efficacité de notre modélisation

Dans la 3éme partie a été consisté aux caractérisations des propriétés physico-chimiques et autres éléments principaux dans les eaux traitées des différentes sources de Médéa et aussi créer un modèle mathématique pour sulfate et bicarbonate. Vu que ces deux paramètres influent sur coagulation et ils ne sont pas stable à leurs analyses.

Partie I : Coagulation

I. Coagulation

I.1. Travail au laboratoire:

I.1.1. Principe:

A des volumes identiques d'eau à traiter, on injecte des quantités croissantes de coagulant de façon à déterminer la dose permettant d'obtenir une eau traitée de meilleure qualité possible. La même opération est répétée pour différents types de coagulant. Les résultats permettent de choisir la nature et la dose de coagulant[62].

I.1.2. Réactifs:

- Solution mère de coagulant (100 g/l de sel d'aluminium exprimée en produit commercial.
- Solution fille (10 g/l) préparée le jour de l'essai.

I.1.3. Appareillage:

- Un chronomètre;
- Une éprouvette de un litre;
- Un floculateur de laboratoire avec six béchers de floculation;
- Un siphon (d $\approx$0,5 mm) recourbé vers le haut pour éviter l'aspiration des boues;
- 7 vases de 0,5 litre pour récupérer les échantillons;
- 1 pipette graduée de 10 ml;
- 1 pH mètre;
- 1 turbidimètre;

-Le jar test: Équipement de laboratoire pour tester des différents produits chimiques à utiliser ultérieurement dans le traitement physico-chimique. Cet équipement est composé de 6 postes avec régulation de vitesse d'agitation et programmation du temps d'essai.

L'origine du jar test est difficile à retracer, son utilisation remonte à plusieurs décennies. Les avantages du jar test sont importants, en voici quelques-uns:

- Déterminer les doses optimales de produits chimiques (coagulants et aides-coagulants);
- Déterminer le pH optimal de coagulation;
- Déterminer la solidité des flocs formés;
- Déterminer les intensités de mélange optimal.



Figure I.1. Jar test

I.1.4. Préparation et conservation des échantillons:

- Le sel d'aluminium a été utilisé dans cette étude comme coagulant. Les tests en pots ont été réalisés à l'aide de béchers remplis d'1 L d'eau, à savoir des eaux brutes collectées dans différentes localités de la région de Médéa, avec différents niveaux de turbidité ajustés par l'acide humique. Parce que les eaux acides et oligotrophes ont souvent une teinte "thé" brune à jaune, notamment due aux acides humiques (les acides humiques sont l'une des fractions les plus importantes de l'humus. ... Grâce à leurs fonctions hydrophiles de type acide carboxylique), ils peuvent retenir une quinzaine de fois leur poids en eau et jouer ainsi un rôle fondamental dans la rétention d'eau et la réserve d'eau utile d'un sol. Ainsi, après avoir créé de la turbidité par l'acide humique, l'eau prend une teinte "thé", et sans oublier que la turbidité contient diverses matières minérales et organiques en suspension ou en solution. Les matières dissoutes et colloïdales constituent à elles seules 60 à 80% de la charge organique (acides humiques). Si nous parvenons à éliminer la turbidité, nous pouvons dire que nous avons éliminé les matières organiques et minérales (par complexation). En raison que, l'élimination de la couleur, c'est-à-dire l'élimination des matières organiques à l'état colloïdal et dissous et les micropolluants organiques[78]

I.1.5. Mode opératoire[96]:

- Placer les béchers sur le floculateur et mettre les pales d'agitation en place.

- Démarrer l'agitation à la vitesse de 180 tours/min. Les pales et leur vitesse de rotation doivent être identiques dans tous les béchers.
- Noter la température de l'eau.
- A l'aide d'une pipette graduée, ajouté le coagulant le plus rapidement possible et de manière reproductible dans chaque bécher. - Les doses appliquées dans chaque bécher sont croissantes de façon à "encadrer" la dose optimale présumée.
- Déclencher le chronomètre (t = 0).
- Après 2 min (t = 2 min) de mélange rapide (dispersion du réactif, coagulation), réduire la vitesse de rotation des pales à 40 tours/min (agitation lente, floculation).
- Noter le temps d'apparition du floc après introduction du coagulant.
- Au temps t = 10 min et t = 20 min, noter l'aspect de la floculation exprimée par une note.
- Au temps t = 20 min, relever (doucement en prenant soin de ne pas casser le floc formé) les pales des agitateurs en respectant un décalage de 15 à 30 s entre chaque relevé.
- Attendre 45 min de décantation puis prélevé, à l'aide d'un siphon, 0,5 litre dans chaque bécher suivant le schéma suivant en veillant à ne pas siphonner la surface de l'eau:
- Après 45 min de décantation, le surnageant de chaque pot a été retiré de l'échantillon de port et analysé pour le pH, la température, la turbidité, la conductivité et le titre métrique total alcalin (TAC). Le pH, la conductivité, la température et la turbidité ont été mesurés au moyen d'un pH-mètre, d'un conductimètre, d'un thermomètre et d'un turbidimètre, respectivement. Le TAC a été analysé immédiatement suivant la méthode de dosage décrite dans partie 1 [29].

I.2. Méthodes de prédiction

Plusieurs méthodes ont été appliquées pour résoudre des problèmes liés à la prédiction et à la modélisation de systèmes non linéaires complexes. Ces méthodes sont particulièrement utiles lorsque de tels systèmes sont difficiles à modéliser à l'aide de méthodes conventionnelles[97]. Dans cette étude, quatre méthodes ont été testées pour prédire les doses de coagulant à partir des paramètres physico-chimiques de l'eau, à savoir RSM, RNA, SVM et ANFIS.

Après avoir construit la base de données selon RSM, la base de données a été normalisée dans l'intervalle [-1, + 1] selon la méthode RSM en utilisant la fonction mapminmax du logiciel Matlab[98]; il a également été divisé en deux parties pour les trois autres modèles (RNA, SVM, ANFIS): 70% du jeu de données pour la formation. Les 30% restants des échantillons, qui n'ont pas participé à l'apprentissage du modèle, ont été utilisés pour les performances de validation et de prédiction des modèles [99,100]

Où xN est la valeur des données après normalisation, x max et x min désignent respectivement le maximum et le minimum des données; y max et y min sont pris comme -1 et 1; x les valeurs d'origine.

$$x_N = (y_{\max} - y_{\min})\left(\frac{x - x_{\min}}{x_{\max} - x_{\min}}\right) + y_{\min}\,. \qquad \textbf{(I.1)}$$

Le coefficient de corrélation (R), le coefficient de détermination (R^2), le coefficient ajusté (R^2adj), l'erreur quadratique moyenne (RMSE), l'erreur quadratique moyenne (MSE), l'erreur absolue moyenne (MAE), la prédiction standard d'erreur (ESP), la prédiction d'erreur Des modèles (EPM) ont été utilisés pour estimer les performances de chaque modèle. Les valeurs correspondantes ont été calculées à l'aide des équations suivantes [100–102]

$$R = \frac{\sum_{i=1}^{N}\left(y_{\exp} - \overline{y_{\exp}}\right)\left(y_{\mathrm{p}red} - \overline{y_{pred}}\right)}{\sqrt{\sum_{i=1}^{N}\left(y_{\exp} - \overline{y_{\exp}}\right)^2 \sum_{i=1}^{N}\left(y_{pred} - \overline{y_{pred}}\right)^2}} \qquad \textbf{(I.2)}$$

$$R^2_{adj} = 1 - \frac{\left(1 - R^2\right)\left(N - 1\right)}{N - K - 1} \qquad \textbf{(I.3)}$$

$$RMSE = \sqrt{\left(\frac{1}{N}\right)\left(\sum_{i=1}^{N}\left(y_{\exp} - y_{pred}\right)^2\right)} \qquad \textbf{(I.4)}$$

$$MSE = \left(\frac{1}{N}\right)\left(\sum_{i=1}^{N}\left(y_{\exp} - y_{pred}\right)^2\right) \qquad \textbf{(I.5)}$$

$$MAE = \left(\frac{1}{N}\right)\sum_{i=1}^{N}\left|y_{\exp} - y_{pred}\right| \qquad \textbf{(I.6)}$$

$$ESP\left(\%\right) = \frac{RMSE}{\overline{y_{\exp}}} \times 100 \qquad \textbf{(I.7)}$$

$$EPM\left(\%\right) = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{\left(y_{\exp} - y_{pred}\right)}{y_{\exp}}\right| \qquad \textbf{(I.8)}$$

Où N est le nombre de données; K est le nombre de variables (entrées); yexp et ypred sont respectivement les valeurs expérimentales et prédites, $\overline{y_{\exp}}$ et $\overline{y_{pred}}$ sont respectivement les valeurs moyennes des valeurs expérimentales et prédites[103–105].

I.2.1. Plan d'expérience (surface de réponse)

Le Centre Compose Design a été sélectionné pour être utilisé dans cette étude [106]. Deux niveaux, une valeur de a = 1 et 5 répétitions au point central ont été considérés. Enfin, les variables d'entrée ont été codées par intervalles [-1, +1] (tableau 1).

Tableau I.1. Codage et niveaux réels pour le modèle CCD.

Variables	Symbole	Unité	Niveaux de facteur de code		
			-1	0	+1
pH	X1	-	7.25	8.02	8.78
Température	X2	°C	10.65	19.48	28.30
Conductivité	X3	Us/cm	106	988	1871
Turbidité	X4	NTU	0.86	519.93	1039
Tac	X5	°F	12.00	18.20	24.4

Le nombre d'expériences est déterminé selon l'équation. (I.9) [107].

$$N = 2^k + 2k + c \quad \textbf{(I.9)}$$

Nous avons appliqué la méthode des plans d'expériences (surface de réponse) sur les résultats obtenus de la station de traitement de Médéa. Le choix du modèle est porté sur un modèle polynomial de degré 2 de la forme [107]:

$$Y = \mathrm{B}_0 + \sum_i^k \mathrm{B}_i X_i + \sum_i^k \mathrm{B}_{ii} X^2 + \sum_{ij}^k \mathrm{B}_{ij} X_i X_j + E \quad \mathrm{Y} = \mathrm{a0} + \Sigma \mathrm{ai\ xi} + \Sigma \mathrm{aij\ xi\ xj} + \Sigma \mathrm{aii\ xi2} \quad \textbf{(I.10)}$$

Rappelons que les facteurs retenus dans cette modélisation sont les suivants :

- La température (T).
- Le pH (P).
- La conductivité de l'eau (C).
- La turbidité (TU).
- Le titre alcalimétrique (TA).

La présentation des plans d'expériences s'est fait manuellement. Les résultats de cette présentation sont donnés dans le tableau suivant.

Le nombre d'expériences est déterminé selon l'équation. (I.9) [107].

Où k est le nombre de variables indépendantes, 2^k est le facteur de points, 2k est les points a xi aux et c est le nombre d'expériences au centre[108]. Le modèle utilisé pour prédire les

réponses et décrit la relation entre les variables indépendantes était l'équation polynomiale du second ordre [107].

$$Y = \mathrm{B}_0 + \sum_{i}^{k} \mathrm{B}_i X_i + \sum_{i}^{k} \mathrm{B}_{ii} X^2 + \sum_{ij}^{k} \mathrm{B}_{ij} X_i X_j + E \qquad \textbf{(I.11)}$$

I.2.2. réseau neurone

Afin de relier les doses de coagulation avec le pH, température, conductivité et turbidité, on a construit un modèle prédictif à base de RN type feed-forward (RNFF). Le RNFF proposé renferme trois couches: une couche d'entrée avec quatre nœuds représentant les facteurs étudiés, une seule couche cachée (une seule couche cachée est jugée suffisante vue le nombre limité des points expérimentaux) et une couche de sortie avec un neurone générant la valeur estimée des doses de coagulant[109].

Afin de rendre l'apprentissage du RNFF plus efficace en empêchant la fonction de transfert de devenir saturé et rendre l'apprentissage très rapide, toutes les variables d'entrée et de sortie ont été normalisées dans l'intervalle (IN) -1 à 1 en utilisant l'équation (I.1) de normalisation.

La fonction de transfert linéaire "logsig et tansig" ont été attribuée à la couche de l'entrée et "purelin et tansig" ont été attribuée à la couche de sortie[100].

L'algorithme LM est utilisé dans l'apprentissage du réseau de neurones. Cette méthode d'apprentissage recherche le nombre de neurones dans la couche cachée qui produisent de bonnes qualités de prédictions par optimisation des poids et des biais L'algorithme de LM (fait partie des méthodes[100]. Le nombre d'itérations est nettement plus faible et les temps de calcul réduits d'autant. Vue le nombre limité des points expérimentaux, la base de données a été aléatoirement répartie en deux sous-ensembles, en réservant 70 % des données pour la phase d'apprentissage et 30 % pour la phase de validation. Afin d'optimiser l'architecture du RN, une procédure d'essai et erreur a été exécutée. Le nombre de neurones (N) dans la couche cachée a été varié de 3 à 10[100].

Dans ce travail, l'optimisation de l'architecture RNA à l'aide du logiciel MATLAB R2013a est décrite par un organigramme représenté sur la figure I.2 [100];

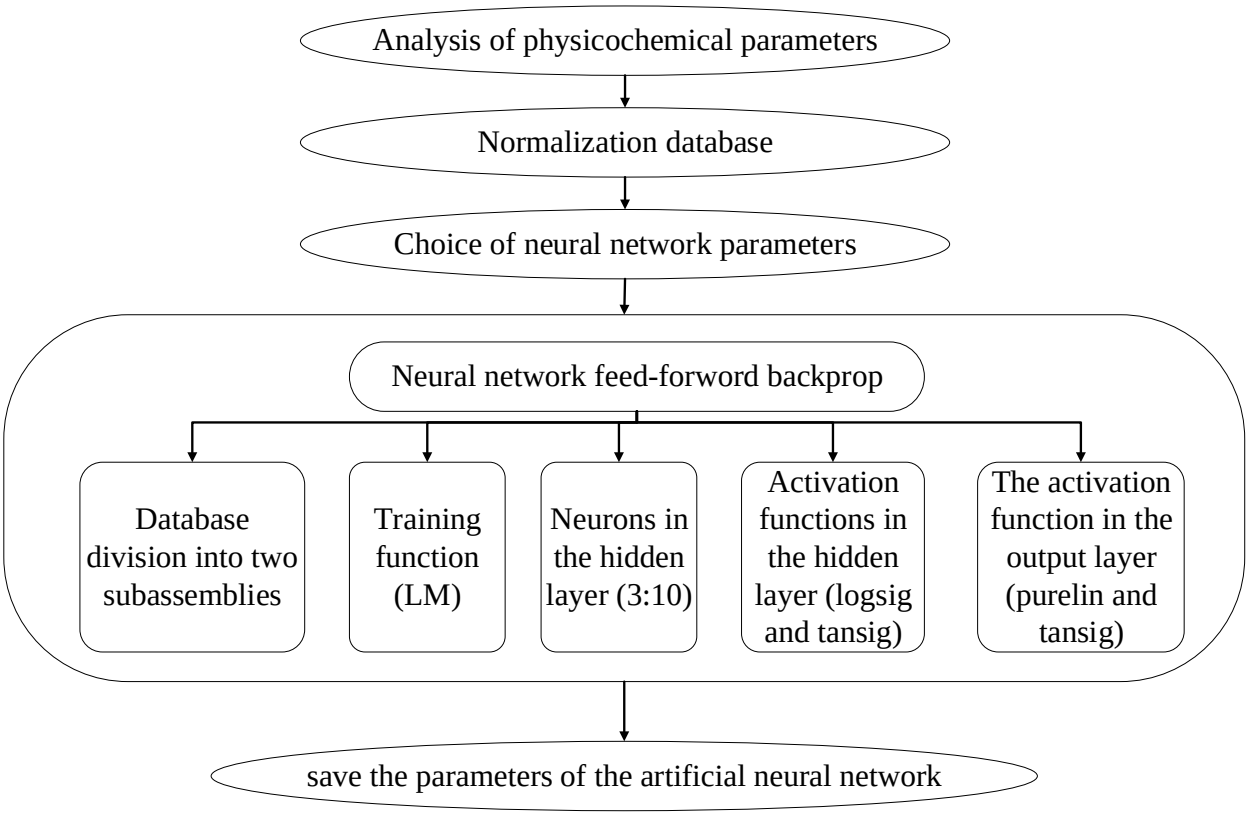


Figure I.2. Organigramme pour le développement et l'optimisation de l'architecture RNA

I.2.3. Support vector machine

Dans ce travail, SVM est utilisé pour la modélisation non linéaire utilisant les différentes fonctions du noyau, car il n'y a aucune garantie qu'un noyau donné apparaîtra plus efficace qu'un autre concent des données spécifiques. A partir de là, il est nécessaire d'optimiser les différentes fonctions du noyau et de tester leurs performances. Ces fonctions du noyau se trouvent dans la boîte à outils Matlab R 2018a*[110]*.

Les fonctions du noyau qui ont été sélectionnées dans ce travail sont les suivantes *[110]*:

- Poly (polynomial)

$$k\left(X_i, X_j\right) = (ax^T + c)^d \quad \textbf{(I.12)}$$

- Linear

$$k\left(X_i, X_j\right) = x^T y + c \quad \textbf{(I.13)}$$

- Gaussian

$$k\left(X_i, X_j\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-X_i - X_j^2}{2\sigma^2}\right\} \quad \textbf{(I.14)}$$

Avec C= BOxConstraint, ε =epsilon, d= PolynomialOrder et σ^2 = sigma

Où d, c, et σ^2 sont des paramètres du noyau définis par l'utilisateur

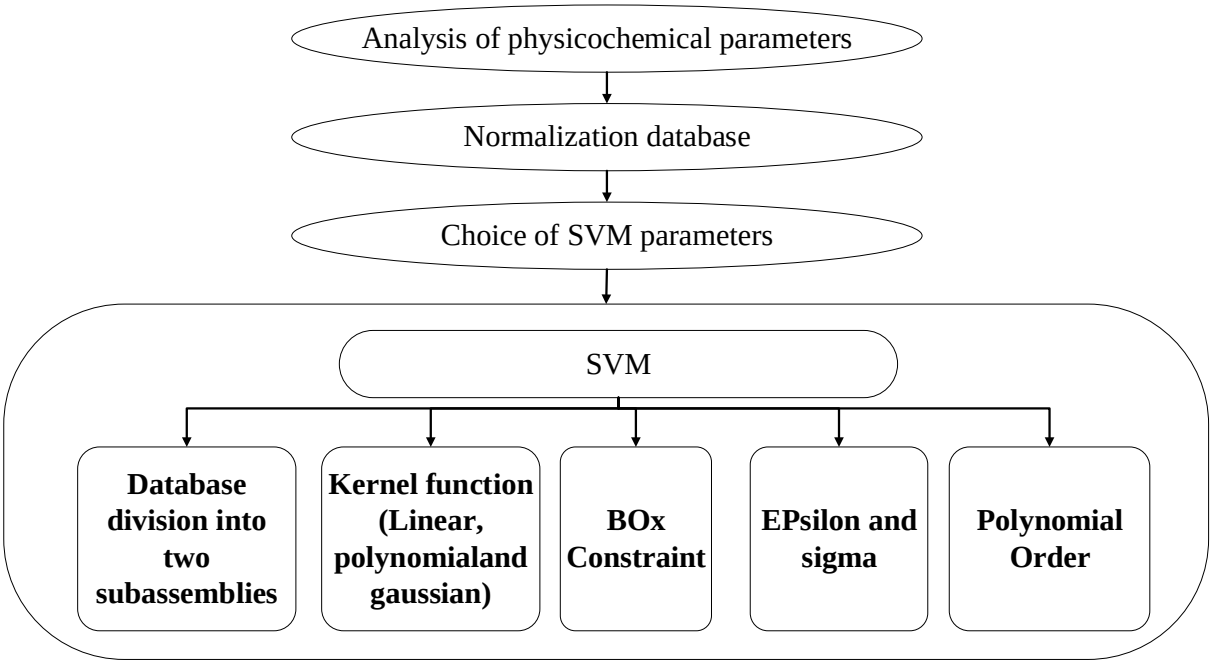


Figure I.3. Organigramme pour le développement et l'optimisation du modèle SVM

I.2.4. Système d'inférence neuro floue adaptatif

Dans ce travail, nous avons optimisé les algorithmes de pondération, les fonctions (trimf, trapmf, gbellmf, gaussmf, gauss2mf, pimf, dsigmf et psigmf) pour l'entrée et (constante et linéaire) pour la sortie et aussi de nombreux nœuds pour chacun entrée pour obtenir le résultat le plus précis; il est à noter que ces algorithmes et fonctions se trouvent dans la boîte à outils de Matlab R 2013[111].

Comme décrit ci-dessus, deux étapes ont été suivies pour la modélisation ANFIS, où deux ensembles de données ont été utilisés, respectivement les données d'apprentissage et de vérification. Tout d'abord, la formation a été effectuée à l'aide de l'ensemble de données de formation. Deuxièmement, la vérification de l'ensemble de données a été envisagée pour vérifier l'exactitude et l'efficacité du modèle ANFIS[111].

Deux critères ont été retenus pour optimiser le modèle ANFIS: le nombre de fonctions d'appartenance affectées aux entrées et sorties et la valeur RMSE. En outre, les quatre aspects suivants ont été pris en compte pour la phase d'apprentissage de l'ANFIS, à savoir le nombre de fonctions d'appartenance utilisées, le type de fonctions d'appartenance, le surajustement et les options de formation.

Partie II : Caractérisation des eaux brutes et traitées

II. Caractérisation des eaux brutes et traitées

Dans notre travail, nous nous sommes basés sur le site le plus pollué afin d'appliquer et de tester l'efficacité du coagulant par le modèle mathématique qui a été créé dans la 1ère partie sur l'élimination de la matière organique et d'autres paramètres physico-chimiques afin de rendre l'eau potable

La 2^{éme} partie consistait en les caractérisations des propriétés physico-chimiques et autres éléments principaux de l'eau brute après une première étude pour trouver un site très pollué dans la région de Médéa. Pour cela, les eaux brutes ont été caractérisées par différentes méthodes d'analyse. Après, nous avons appliqué le modèle mathématique de la partie 1 dans la station ADE afin de fixer les doses de coagulant. A la fin, nous avons caractérisé l'eau traitée afin de la comparer avec l'eau brute et voir l'efficacité de notre modélisation

II.1. Mode de prélèvement

Les échantillons d'eau sont recueillis dans des flacons en verre bien lavés. Les prélèvements se font à une profondeur de 15 à 30 cm de la surface de l'eau, en évitant la pénétration de l'air. Le transport des échantillons depuis le point de prélèvement jusqu'au laboratoire a été effectué dans une glacière à 4 °C. Avant de procéder aux opérations analytiques, il est essentiel que toutes les dispositions soient prises, telles que l'homogénéisation au moment du dosage.

II.2. Matières utilisées

Le tableau ci-dessous nous présentons les différents produits utilisés lors de notre étude.

Tableau II.1. Types de solution et produite utilisés pour les différents analyses.

Paramètres mesurés	Types de solutions et produits
Chlorures	Chromate de potassium, Nitrates d'argent
TA et TAC	Phénolphtaléine et méthyle d'orange HCl
Ca^{2+}et Dureté Totale	NaOH, Murexide, solution Tampon pH=10, Noir eriochrome(NET) et EDTA
Matière Organique	Acide sulfurique, Permanganate de potassium et sulfate ferreux ammoniacal

II.3. Matériels utilisés

Le tableau ci-dessous nous présentons les différents appareils utilisés lors de notre étude.

Tableau II.2. Les différents matériels utilisés pour les différentes analyses.

Paramètre mesurés	Types d'appareils
Température	Multi-paramètre WTW
TDS	
Salinité (%)	
Conductivité (µs / cm)	
Turbidité	Turbidimètre de laboratoire modèle 2100N HACH à 4000NTU
PH	PH-mètre
Ca^{2+}, K^{+}, Na^{+}	La photométrie d'émission atomique
La densité optique	Spectromètre

I.4. Méthodes d'analyse physico-chimique

Les prélèvements des échantillons pour l'analyse physico-chimique ont été mis dans des flacons en verre puis acheminés au laboratoire pour l'analyse. Les analyses physico-chimiques ont concerné les paramètres suivants : T°C, PH, Conductivité électrique, Dureté totale, Chlorures.....etc.

Les méthodes analytiques utilisées sont décrites par[5] :

II.4.1. Analyse physique

a) Température, salinité, Total solides dissous, et conductivité

Ces paramètres sont mesurés à l'aide d'un Multi-paramètre WTW.

b) Turbidité

La turbidité est mesurée au laboratoire à l'aide d'un turbidimètre « Hach » à cellule photo-électrique, par la méthode néphélémétrique. Les résultats sont exprimés en unité de turbidité néphélémétrique (NTU).

c) PH

La mesure de pH est effectuée couramment par la méthode potentiométrique avec électrode de verre plus.

- **Principe**

La différence de potentiel existant entre une électrode de verre et une électrode de référence (calomel-KCl saturé) plongeant dans une même solution, est une fonction linéaire du pH de celle-ci. Selon l'expression de NERNST, le potentiel de l'électrode est lié à l'activité des ions H^+.

II.4.2. Analyse chimique

a) Dureté Totale

- **Principe**

Dureté totale par méthode titrimétrie à l'EDTA permet de doser rapidement les ions calcium et magnésium, avec certaines précautions, elle est appropriée à la plupart des types d'eaux.

La dureté totale détermine la concentration en calcium et du magnésium dissous. Les alcalino-terreux présents dans l'eau sont amenés à former un complexe de type chélate par le sel di sodique de l'Acide Éthylène Diamine tétra acétique (EDTA).

b) Dosage des chlorures par la méthode de « MOHR »

- **Principe de la méthode**

Réaction des ions chlorures avec des ions argent pour former du chlorure d'argent insoluble qui est précipité quantitativement .Addition d'un petit excès d'ions argent et formation du chromate d'argent brun-rouge avec des ions chromates qui ont été ajoutés comme indicateur. Cette réaction est utilisée pour l'indication du virage. Durant le titrage, le pH est maintenu entre 5 et 9.5 afin de permettre la précipitation.

c) Dosage des nitrites (Méthode au réactif de ZEMBELLI)

- **Principe de la méthode**

Réaction des ions nitrites présents dans une prise d'essai avec le réactif amino-4 benzène sulfonamide en présence d'aide ortho phosphorique pour former un sel diazoïque qui forme un complexe de coloration rose avec le dichlor-hydrate de N-(naphtyl-1)diamino-1,2 éthane (ajouter avec le réactif amino-4 benzène sulfonamide) .

d) Dosage des nitrates (Méthode au salicylate de sodium)

- **Principe de la méthode**

En présence de salicylate de sodium, les nitrates donnent du paranitrosalisylate de sodium, coloré en jaune susceptible d'un dosage Calorimétrique.

e) Dosage des phosphates

- **Principe de la méthode**

Les ions orthophosphates peuvent être dosés avec précision par colorimétrie. Ils réagissent avec le molybdate d'ammonium pour donner de l'acide molybdophosphorique. Cet acide donne lieu à diverses réactions colorées. La réaction au chlorure stanneux ou à l'acide ascorbique donne du molybdène qui développe une coloration bleu susceptible d'un dosage spectrophotométrie à une longueur d'onde de 825 nm.

f) Dosage du Sodium, Potassium et Calcium (Méthode par spectrométrie d'émission de flamme)

- **Principe de la méthode**

Le photomètre de flamme est principalement utilisé pour le dosage du sodium, du potassium, du lithium et du calcium. En raison de la facilité et de la rapidité de son utilisation ainsi que du peu d'interférences. Cette méthode est largement utilisée pour ces éléments qui sont difficiles à doser autrement. Lorsque le cation, préalablement nébulisé, arrive dans la flamme, il est alors excité. Il se désexcite alors en émettant des photons, c'est-à-dire une émission lumineuse. Cette intensité émise est proportionnelle à la concentration du composé.

La mise au point d'un dosage par émission de flamme exige un étalonnage de l'appareil.

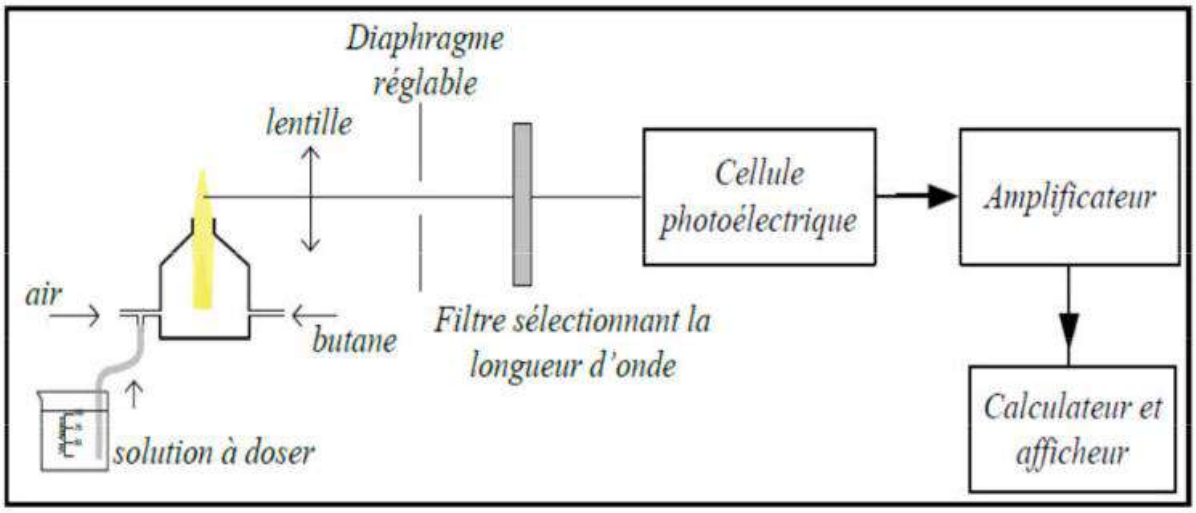


Figure II.1. Schéma du principe d'un photomètre de flamme [53].

g) Dosage de la matière organique (Méthode à chaud)

- **Principe**

L'indice de permanganate d'une eau est la concentration en masse d'oxygène équivalente à la quantité d'ions permanganates consommée quand un échantillon d'eau est traité par le permanganate dans les conditions définies. L'échantillon d'eau est en présence d'une quantité connue de permanganate de potassium et d'acide sulfurique pendant une période donnée (10

minutes). Une partie du permanganate est réduite par les matières oxydables de l'échantillon. L'excès de permanganate est déterminé par addition d'un excès d'oxalate, suivie par un titrage de l'oxalate en excès par du permanganate (principe du titrage en retour).

h) Résidu sec

- **Principe**

Une certaine quantité d'eau bien mélangée est évaporée dans une capsule tarée. Le résidu desséché est ensuite pesé.

i) Dosage de L'ammonium (Méthode de Nessler)

- **principe**

Mesurage spectrométrique de la solution de tartrate de potassium et de citrate de sodium ajoutée à l'échantillon afin d'éviter la précipitation subséquente d'hydroxydes de calcium et de magnésium.

Par la suite, l'azote ammoniacal réagit en présence de salicylate, de nitroprussiate de sodium et de chlore actif pour former un complexe vert dont l'absorbance à 660 nm est proportionnelle à la concentration d'azote ammoniacal.

j) dosage de calcium (Méthode titremetrique à l'EDTA)

- **Principe**

Titrage des ions calcium avec une solution aqueuse de l'EDTA à un pH compris entre 12 et 13. L'indicateur utilisé est le murexide, qui forme un complexe rose avec le calcium. Lors du titrage, l'EDTA réagit avec les ions calcium, l'indicateur vire alors de la couleur rose à la couleur violet.

k) Détermination de sulfate

- **Principe**

Les ions sulfates sont précipités et dosés à l'état de sulfate de baryum suivant la réaction:

$$BaCl_2 + SO_4^{-2} \longrightarrow BaSO_4 + 2Cl^- \qquad \textbf{(II.1)}$$

l) Détermination d'Aluminium (par spectromètrie)

- **Principe**

Réaction de l'aluminium avec l'ériochrome cyanine à un pH de 5.9 en présence de l'acétate d'ammonium.

m) Détermination du Manganèse (par spectrométrie)

- **Principe**
- C'est une réaction des ions manganèses avec une solution d'acide ascorbique, forme un complexe rouge orangé mesurable à la longueur d'onde de 450 nm dans un domaine de pH situé entre 9.5 et 10.5.

n) Détermination de l'Alcalinité

- **principe**

Détermination des volumes successifs d'acide fort en solution diluée nécessaires pour neutraliser, aux niveaux de pH 8.3 et 4.3, le volume d'eau à analyser. la première détermination sert à calculer le titre alcalimétrique (TA), La seconde à calculer le titre alcalimétrique complet (TAC).

o) Dosage du fer (par spectrométrie)

- **principe**

Addition d'une solution de phénantroline-1.10 a une prise d'essai et mesurage photométrique du complexe rouge-orange à une longueur d'onde de 510 nm. Le complexe fer (II-phénantroline-1.10 est stable dans l'intervalle de pH de 2.5 à 9 et l'intensité de la coloration est proportionnelle à la quantité du fer.

I.5. Analyses physico-chimiques

Mesure de Totale des solides dissous (TDS), Potentiel d'hydrogéne (PH), Température (T°), Oxygène dissous (O_2dissous), Conductivité électrique et Salinité à l'aide d'un multi-paramètre[5].



Figure II.2. Image prise d'un multi-paramètre utilisé dans la présente étude.

On peut mesurer le PH à l'aide d'un pH-mètre.



Figure II.3. Image prise d'un PH-mètre utilisé dans la présente étude.

II.5.1. Mode opératoire

➢**Dosage de l'échantillon**

- Prendre environ 100 ml d'eau à analyser
- Mettre un agitateur avec une bécher
- Laisser stabilisé un moment avec une faible vitesse d'agitateur
- Puis noter le pH, Tº, l'oxygène dissous, conductivité électrique et salinité

Les résultats donnés directement en % et mg/l (oxygène dissous), µs/cm (conductivité) et en g/l (salinité).

II.5.2. Turbidité

II.5.2.1. Mode opératoire

Remplir une cuvette de mesure propre et bien essuyer avec du papier hygiénique avec l'échantillon à analyser bien homogénéisé et effectuer rapidement la mesure, il est nécessaire de vérifier l'absence de bulle d'air avant la mesure[5].

La mesure est obtenue directement en NTU.

II.5.3. Dosage de Calcium Ca^{2+}

II.5.3.1. Mode opératoire

- Prélever une prise d'essai de 50ml de l'échantillon, ajouter 2ml de la solution d'hydroxyde NaOH (2N) et une pincée d'indicateur (Murexide), la solution doit se colorer en rose bien mélanger le tout ;
- Titrer avec la solution d'EDTA, en versant lentement le virage est atteint lorsque la couleur devient nettement violette ;



Figure II.4. Image prise du dosage de calcium par l'EDTA.

- La couleur ne doit plus changer avec l'ajout d'une goutte supplémentaire de la solution d'EDTA.

$$Ca = (C_1 * V_1 * A) / V_0 \left\{ F_c * 1000 * F \right\} \quad \textbf{(II.2)}$$

Ou :

C_1 : concentration exprimée en mole par litre de la solution EDTA soit 0,01.

V_0 : est le volume en millilitre de la prise d'essai soit 50ml.

V_1 : est le volume en millilitre de la solution d'EDTA utilisé pour le dosage.

A : est la masse atomique du calcium (40,08g).

F_c : facteur de correction du titre.

F : facteur de dilution.

II.5.4. Dureté Totale (TH)

II.5.4.1. Mode opératoire

- Placer 50 ml d'eau à analyser dans un erlenmeyer;
- Ajouter 4 ml de solution Tampon et 3 goutte de solution de noir ériochrome;
- Titrer la solution de l'EDTA rapidement en début puis goutte à goutte jusqu'à virage bleu;

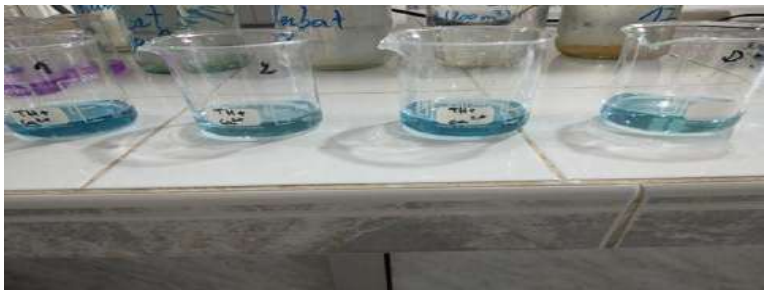


Figure II.5. Image prise du résultat de dosage de la dureté par l'EDTA.

▪Noter $V_{éq}$ de Solution l'EDTA.

TH = 1000 * [(C*V1) / V2] **(II.3)**

TH : C'est le titre hydrométrique en mg/l

C : Concentration en milliéquivalent par litre d'EDTA

V1: Volume ml de solution d'EDTA utilisé pour le titrage

V2: Volume d'échantillon.

II.5.5. Dosage de Chlorure (Cl^{-})

II.5.5.1. Mode opératoire

- Introduire 100 ml de l'échantillon dans une fiole;
- placer sur un fond blanc et ajouter 1 ml d'indicateur de chromate de potassium(b)
- titrer la solution par addition goutte à goutte de solution de nitrate d'Argent jusqu'à ce que la solution prenne une couleur rougeâtre;



Figure II.6. Image prise du dosage des ins chlore par l'EDTA.

- Après addition d'une goutte de solution de Chlorure de Sodium cette coloration doit disparaître.

II.5.6. Détermination de Magnésium (Mg^{2+})

La différence entre la dureté totale et la dureté calcique donne directement la dureté magnésienne de l'eau analysée.

II.5.7. Détermination de Titre Alcalimétrique Complet (TAC) et l'acidité (HCO_3)

II.5.7.1. Mode opératoire

- Introduire 100 ml de l'échantillon dans un bécher et noter le pH ;

- Amener 100 ml d'eau à analyser pH au 4.3 en versant lentement l'acide chlorhydrique (HCl) pour obtenir cette valeur à l'aide d'un disimal ;



Figure II.7. Image prise d'un protocole de dosage de TAC de la présente étude.

- Noter le volume de HCl.

$TAC=V_{HCl}*50=mg/L$ **(II.4)**

$HCO_3=V_{HCl}*61=mg/l$ **(II.5)**

II.5.8. Détermination de Résidu Sec

II.5.8.1. Mode opératoire

▪Tracer une capsule préalablement lavée, rincer à l'eau distillée et desséchée;

▪Prélever 100 ml d'eau à analyser dans une fiole jaugée et déverser la dans la capsule;

▪Porter cette dernière à l'étuve à 105 C° pendant 24 heures;

▪Laisser refroidir pendant ¼ d'heure au dessiccateur ;

▪Peser immédiatement et rapidement [5].

$R.S= (V_2-V_1)\ x2x104$ **(II.6)**

II.5.9. Détermination d'ammoniacal (NH^+_4)

II.5.9.1. Mode opératoire

- Prendre 40ml d'eau à analyser ;
- Ajouter 2ml du réactif I(1g Acide dichloroisocyanurique + 16g Hydroxyde de sodium (NaOH) + eau distillé) ;
- Ajouter 2ml du réactif II (65g Tricitirate de sodium + 16g Salicylate de sodium + 0.485g Nitroprussaite de sodium) et ajuster à 50ml avec H_2O distillée et attendre 1heure ;
- L'apparition de la coloration verdâtre indique la présence de **NH^+_4;**
- faire une lecture spectrométrique à environ 655 nm.

II.5.10. Détermination de nitrites NO^-_2

II.5.10.1. Mode opératoire

- Prendre 50ml d'eau à analyser ;
- Ajouter 1ml du réactif mixte (40g Sulfanilamide+ 100 ml Acide phosphorique+ N-1 Naphtyle éthylène diamine);

- Attendre 10mn ;
- L'apparition de la coloration rose indique la présence des NO^-_2;
- faire une lecture spectrométrique à environ 543 nm.

II.5.11. Détermination du phosphate PO_4^{3-}

II.5.11.1. Mode opératoire

- Introduire 40 ml d'échantillon;
- ajouter 1 ml d'acide ascorbique + 2mL de réactif mélange dans une fiole jaugé de 50 ml;
- Compléter le volume;
- porter à l'étuve 80°C pendant 20 min;
- faire une lecture spectrométrique à 825 nm.

II.5.12. Détermination de sulfates SO_4^{2-}

II.5.12.1. Mode opératoire

- Introduire 40 ml d'échantillon;
- ajouter 5ml solution stabilisante + 2mL de solution de chlorure de baryum dans une fiole jaugé de 50 ml;
- Compléter le volume;
- faire une lecture spectrométrique à 420 nm.

II.5.13. Détermination de nitrate NO^-_3

II.5.13.1. Mode opératoire

- Introduire 10ml d'eau à analyser ;
- Ajouter 3 gouttes de la solution d'hydroxyde de Sodium à 30% **(b)** ;
- Ajouter 1 ml de solution de Salicylate de sodium à 0.5 % **(a)** ;
- Evaporer à sec au bain marie ou à l'étuve 75 - 88°C ;
- Prendre le résidu avec 2 ml d'acide sulfurique concentré **(d)** ;
- Laisser reposer 10 minutes ;
- Ajouter 15 ml d'eau distillée ;
- Ajouter 15 ml de tartrate double de sodium et de potassium **;**
- Faire la lecture au spectromètre UV-Visible à la longueur d'onde de 415 nm.



Figure II.8. Image prise lors de l'expérience de la détermination de nitrate.

II.5.14. Détermination de fer

II.5.14.1. Mode opératoire

- Prendre comme prise d'essai, 50 ml ;
- Transvaser la solution dans une fiole de 100 ml, ajouter 1 ml de la solution chlorhydrate hydroxylamine et mélanger soigneusement ;
- Ajouter 2.0 ml de tampon acétate pour obtenir un pH entre 3.5 et 5.5 de préférence 4.5 ;
- Ajouter 2.0 ml de la solution phénantroline et conserver pendant 15 min à l'obscurité ;
- Mesurer l'absorbance à l'aide d'un spectrophotomètre UV VIS à 510 nm en utilisant une cuve de 10 mm^2.

II.5.15. Détermination du manganèse

II.5.15.1. Mode opératoire

- Introduire 40 ml d'échantillon;
- ajouter 1 ml d'acide ascorbique + 2mL de réactif mélange dans une fiole jaugé de 50 ml;
- Compléter le volume;
- Une coloration rouge orangé apparaitre;
- Mesurer l'absorbance à l'aide d'un spectrophotomètre UV VIS à 450 nm.

II.5.16. Détermination de sodium et de potassium

II.5.16.1. Mode opératoire

Appareil Dr LANGE (JENWAY).

Il faut le suivre étape par étape :

- Allumer l'appareil à l'aide du bouton vert (power) ;

- Ouvrir le robinet de la bouteille du gaz ;
- Allumer la flamme à l'aide u bouton noir « IGNITION » sans lâcher le doigt jusqu'à l'affichage « FLM » en rouge sur l'écran ;
- Pipeter de l'eau distillée remplie dans une cuvette ;
- Optimiser la flamme si elle est jaune à l'aide du bouton « fuel » jusqu'à ce que la couleur devienne bleue violacée ;
- Optimiser à zéro à l'aide du bouton « Blank » ;
- Laisser se stabiliser 5 à 10 minutes ;
- Une fois qu'elle se stabiliser à zéro, activer la cuvette d'eau distillée et la remplacer par une autre cuvette remplie par une solution étalon de Na^+ ou du K^+à10 mg/ ;
- Optimiser à 10 mg/l à l'aide du bouton « FINE » ;
- Retirer la cuvette remplie par une solution étalon de Na^+ ou de K^+ à 10 mg/l et la remplacer par une cuvette remplie d'eau distillée et vérifier si l'écran affiche zéro (0.000) ;
- Retirer la cuvette remplier par l'eau distillée et la remplacer par une cuvette remplie par une solution étalon de Na^+ ou K^+ à 10mg/l et vérifier si l'écran affiche. ;
- Retirer la cuvette et la remplacer par une autre cuvette remplie d'eau distillée ;
- A la fin passer aux échantillons inconnus jusqu'à ce que la valeur affichée sur l'écran soit stable (3 essais pour chaque échantillon) ;
- A la fin du dosage et par mise de la sécurité, il faut toujours fermer la bouteille de gaz propane en premier lieu ensuite l'appareil el la pompe.

II.5.17. Détermination de l'oxydation au permanganate de potassium :

II.5.17.1. Mode opératoire

- Dans un erlenmeyer à col large on introduit 100 ml d'eau à analyser puis on ajoute 10ml d'acide sulfurique ½ N plus quelques grains de Potter et porter à ébullition (1er bulles) ;
- Ajouter 10 ml de permanganate de k^+ N/80 et porter à ébullition pendant 10 minutes (2éme ébullition) ;



Figure II.9. Image prise pendant la détermination de la matière organique.

- Refroidir puis ajouter 10ml de sulfate ferreux ammoniacal (N/80) ;

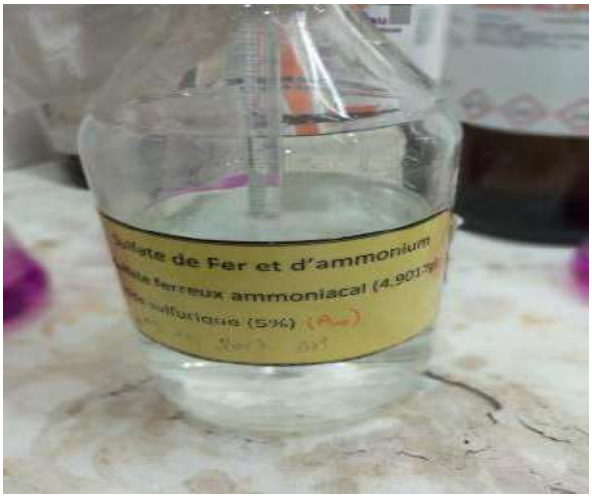


Figure II.10. Image prise de sulfate ferreux ammoniacal utilisé dans l'expérience.

- Titrer avec KM_nO_4 jusqu'au virage de la solution rose claire ;
- Faire un témoin avec l'eau distillé.

$O = V - V_0$ mg/l **(II.7)**

II.5.18. Détermination d'Aluminium Al^{3+}

II.5.18.1. Mode opératoire

- Dans une fiole jaugée de 50 ml, introduire à l'aide d'une pipette, un volume de10 ml de l'échantillon;
- ajouter 0.5 ml de thiosulfate de sodium 0.028 N avec agitation;
- 1 ml d'acide ascorbique à 1% (c)avec agitation;
- 1 ml d'acide sulfurique à 0.04 N(b);
- 10 ml de la solution tampon pH 5.9;

- 5 ml de la solution fille d'eriochrome cyanine, réajuster le volume à 50 ml avec de l'eau distillé et attendre 10 min. Une couleur rouge orange apparaitre en cas de présence d'aluminium;
- Mesurer enfin, l'absorbance à la longueur d'onde (environ 540 nm).

Partie III : Modélisation

L'eau traitée a été récupéré par différents sites de Médéa et analysée par différents paramètres physico-chimiques qui ont été écrits dans la 2éme partie afin de modéliser deux paramètres (sulfate avec réseau neurone et régression linaire multiple et bicarbonate avec réseau neurone).

III.1. Modélisation

III.1.1. Sulfate :

Plusieurs méthodes ont été appliquées pour résoudre les problèmes liés à la prédiction et à la modélisation de systèmes non linéaires complexes. Ces méthodes sont particulièrement utiles lorsque ces systèmes sont difficiles à modéliser à l'aide de méthodes classiques[97]. Dans cette étude, nous nous sommes intéressés à l'utilisation et à la comparaison de deux méthodes de prédiction des teneurs en sulfate soluble à partir des paramètres physico-chimiques de l'eau potable. Ces méthodes sont la régression linéaire multiple et les réseaux de neurones artificiels.

III.1.1.1. Base de données

Les données utilisées pour cette étude ont été obtenues à partir de l'analyse expérimentale d'échantillons d'eau prélevés à plusieurs reprises au cours de la période 2018 dans la région de Médéa, en Algérie (trois échantillons par semaine provenant de différentes régions).

La variable dépendante était la teneur en sulfate soluble dans l'eau potable. Les variables indépendantes étaient les paramètres physico-chimiques: conductivité, turbidité, potentiel hydrogène, dureté, calcium, magnésium, chlorures, titre métrique alcalin total (TAC), matière organique, dioxyde d'azote, nitrates, sodium, bicarbonate, potassium, métaux lourds (Mn^{2+}, Fe^{3+} et Al^{+}) et résidus secs. Les données sont divisées en trois phases (70% d'apprentissage, 15% de test et 15% de validation) selon le modèle RNA pour comparer les coefficients de corrélation et les erreurs entre les deux modèles[112].

III.1.1.2. Méthodes de prédiction

III.1.1.2.1. Réseau de neurone

Afin de relier la variable dépendante avec les variables indépendantes, on a construit un modèle prédictif à base de RN type feed-forward (RNFF). Le RNFF proposé renferme trois couches: une couche d'entrée avec 18 noeuds représentant les facteurs étudiés, une seule couche cachée (une seule couche cachée est jugée suffisante vue le nombre limité des points expérimentaux) et une couche de sortie avec un neurone générant la valeur estimée des doses de coagulant.

Afin de rendre l'apprentissage du RNFF plus efficace en empêchant la fonction de transfert de devenir saturé et rendre l'apprentissage très rapide, toutes les variables d'entrée et de sortie ont été normalisées dans l'intervalle (IN) -1 à 1 en utilisant l'équation (I.1) de normalisation.

La fonction de transfert linéaire "logsig et tansig" ont été attribuée à la couche de l'entrée et "purelin et tansig" ont été attribuée à la couche de sortie.

L'algorithme LM est utilisé dans l'apprentissage du réseau de neurones. Cette méthode d'apprentissage recherche le nombre de neurones dans la couche cachée qui produisent de bonnes qualités de prédictions par optimisation des poids et des biais L'algorithme de LM (Press et al. 1988) fait partie des méthodes d'optimisation. Le nombre d'itérations est nettement plus faible et les temps de calcul réduits d'autant. Vue le nombre limité des points expérimentaux, la couche cachée a été varié de 3 à 10.

Dans ce travail, l'optimisation de l'architecture RNA à l'aide du logiciel MATLAB R2013a est décrite par un organigramme représenté sur la **Figure I.2**[100]**.**

III.1.1.2.2. Régressions linéaires multiples

La régression linéaire multiple consiste à décrire les relations entre une variable dépendante et plusieurs variables appelées variables indépendantes, où n est le nombre de variables indépendantes. En effet, la régression linéaire multiple, qui est une méthode d'analyse de données, est couramment utilisée pour établir des modèles prédictifs des phénomènes observés en milieu aquatique [113]. Cette méthode permet de dessiner une fonction polynomiale décrivant la relation entre les variables dépendantes et indépendantes et permet de déterminer les variables d'entrée les plus significatives. Le modèle peut être écrit comme dans l'équation (III.1).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_i x_i + \cdots + \beta_n x_n + \varepsilon(x) \quad \textbf{(III.1)}$$

Avec :

- n : number of independent variables;
- $\varepsilon(x)$: random noise (error term or regression residual);
- y : dependent variable;
- $x_1, x_2, \cdots, x_i, \cdots, x_n$: independent variables;
- β_0 : estimated ordinate at the origin;
- $\beta_0, \beta_1, \cdots, \beta_i, \cdots, \beta_n$: model coefficients.

III.1.1.3. critères d'évaluation statistique

Le coefficient de corrélation (R), le coefficient ajusté, l'erreur quadratique moyenne (RMSE), l'erreur quadratique moyenne (MSE) et l'erreur absolue moyenne (MAE) ont été utilisés pour estimer les performances du modèle. Les valeurs correspondantes sont calculées à l'aide des équations (I.2 jusqu'à I.8) qui ont été décrites dans la partie (I) [101,102].

III.1.2. Bicarbonate

Dans cette partie, nous n'avons utilisé que réseau neurone pour prédiction des teneurs du bicarbonate

III.1.2.1. Base de données

Les données utilisées pour cette étude ont été obtenues à partir de l'analyse expérimentale d'échantillons d'eau prélevés à plusieurs reprises au cours de la période 2018 dans la région de Médéa, en Algérie (trois échantillons par semaine provenant de différentes régions).

Variables indépendantes sont les paramètres physico-chimiques qui sont à savoir: conductivité, turbidité, pH, dureté, calcium, magnésium, chlorures, titre d'alcalinité totale (TAC), matière organique, dioxyde d'azote, nitrates, sodium, sulfates, potassium, métaux lourds (Mn^{2+}, Fe^{3+} et Al^{+}) et les résidus secs.

Afin de relier la variable dépendante avec les variables indépendantes, on a construit un modèle prédictif à base de RN type feed-forward (RNFF). Le RNFF proposé renferme trois couches: une couche d'entrée avec 18 noeuds représentant les facteurs étudiés, une seule couche cachée (une seule couche cachée est jugée suffisante vue le nombre limité des points expérimentaux) et une couche de sortie avec un neurone générant la valeur estimée des doses de coagulant.

Afin de rendre l'apprentissage du RNFF plus efficace en empêchant la fonction de transfert de devenir saturé et rendre l'apprentissage très rapide, toutes les variables d'entrée et de sortie ont été normalisées dans l'intervalle (IN) -1 à 1 en utilisant l'équation (I.1) de normalisation[100].

La fonction de transfert linéaire "logsig et tansig" ont été attribuée à la couche de l'entrée et "purelin et tansig" ont été attribuée à la couche de sortie[100].

L'algorithme LM est utilisé dans l'apprentissage du réseau de neurones. Cette méthode d'apprentissage recherche le nombre de neurones dans la couche cachée qui produisent de bonnes qualités de prédictions par optimisation des poids et des biais L'algorithme de LM fait partie des méthodes d'optimisation. Le nombre d'itérations est nettement plus faible et les temps de calcul réduits d'autant. Vue le nombre limité des points expérimentaux, la couche cachée a été varié de 3 à 10.

Dans ce travail, l'optimisation de l'architecture RNA à l'aide du logiciel MATLAB R2013a est décrite par un organigramme représenté sur la **Figure I.2**[100]

III.1.2.2. critères d'évaluation statistique

Dans cette étude, nous présentons les différents paramètres statistiques utilisés pour évaluer l'efficacité de chaque réseau et sa capacité à faire des prédictions précises.

Le coefficient de corrélation (R), l'erreur quadratique moyenne (RMSE), l'erreur quadratique moyenne (MSE), l'erreur absolue moyenne (MAE), l'erreur de prédication standard (ESP) et

l'erreur de prédiction d'erreur de modèle (EPM) ont été utilisés pour estimer les performances du modèle. Ces paramètres statistiques sont calculés à l'aide des équations ((I.2 jusqu'à I.8)) qui ont été décrites dans la partie (II)[100]

Résultats et discussions

Partie I : Coagulation

I. Prédiction des doses de coagulation

RSM, RNA, SVM et ANFIS sont les quatre approches qui ont été utilisées pour la prédiction de la dose de coagulant à partir des paramètres physico-chimiques de l'eau. Les quatre méthodes ont été réalisées, évaluées et comparées.

I.1. RSM

Une analyse statistique par RSM a été réalisée à l'aide du logiciel "JMP 13" sur l'ensemble du logiciel de base de données. Cette méthode a permis de trouver la relation mathématique (Eq.I. 1) entre les doses de coagulant et les variables indépendantes qui correspondaient aux paramètres physico-chimiques. Premièrement, l'équation a pris en compte les 5 variables (tableau I.1), même celles qui ne semblaient pas avoir un impact significatif sur la variable dépendante. La relation obtenue a donc été évaluée afin de ne conserver que les variables indépendantes caractérisées par une valeur de probabilité élevée (tableau I.1).

Tableau I.1. Valeurs de β_i , xi i et caractérisation des variables indépendantes du RSM.

i	Terme	β_i	Erreur Std	t Rapport	Prob>|t|
0	Intercepter	29.44685	2.01046	14.64676	4.40E-08
1	X1	-3.88888	1.58898	-2.44740	0.03441
2	X2	-2.55555	1.58898	-1.60829	0.13884
3	X3	-11.94440	1.58898	-7.51703	2.02E-05
4	X4	16.38888	1.58898	10.31406	1.20E-06
5	X5	4.16666	1.58898	2.62222	0.02550
6	X1*X2	5.81250	1.68537	3.44879	0.00623
7	X1*X3	6.68750	1.68537	3.96796	0.00265
8	X2*X3	-7.56250	1.68537	-4.48713	0.00116
9	X1*X4	2.06250	1.68537	1.22376	0.24909
10	X2*X4	-14.18750	1.68537	-8.41802	7.52E-06
11	X3*X4	-1.31250	1.68537	-0.77875	0.45415
12	X1*X5	4.68750	1.68537	2.78128	0.01940
13	X2*X5	-2.56250	1.68537	-1.52043	0.15936
14	X3*X5	1.56250	1.68537	0.92709	0.37571
15	X4*X5	6.18750	1.68537	3.67129	0.00430
16	X1*X1	3.02386	4.29938	0.70332	0.49790
17	X2*X2	-5.97613	4.29938	-1.38999	0.19469
18	X3*X3	20.52386	4.29938	4.77367	0.00075
19	X4*X4	8.52386	4.29938	1.98257	0.07554
20	X5*X5	-13.47613	4.29938	-3.13443	0.01061

$$Y = \beta_0 + \beta_1.X_1 + \beta_2.X_2 + \beta_3.X_3 + \beta_4.X_4 + \beta_5.X_5 + \beta_6.X_1.X_2 + \beta_7.X_1.X_3 + \beta_8.X_2.X_3 + \beta_9.X_1.X_4 + \beta_{10}.X_2.X_4 + \beta_{11}.X_3.X_4 + \beta_{12}.X_1.X_5 + \beta_{13}.X_2.X5 + \beta_{14}.X_3.X_5 + \beta_{15}.X_4.X_5 + \beta_{16}.X_1^2 + \beta_{17}.X_2^2 + \beta_{18}.X_3^2 + \beta_{19}.X_4^2 + \beta_{20}.X_5^2 \quad \textbf{(I.1)}$$

Ensuite, seules 13 variables indépendantes qui ont un pouvoir explicatif élevé pour la variable dépendante ont été retenues (Pr <0,05); ainsi la relation peut être réduite à Eq (I.2):

$$Y = \beta_0 + \beta_1.X_1 + \beta_3.X_3 + \beta_4.X_4 + \beta_5.X_5 + \beta_6.X_1.X_2 + \beta_7.X_1.X_3 + \beta_8.X_2.X_3 + \beta_{10}.X_2.X_4 + \beta_{12}.X_1.X_5 + \beta_{15}.X_4.X_5 + \beta_{19}.X_4^2 + \beta_{20}.X_5^2 \quad \textbf{(I.2)}$$

La valeur du coefficient de détermination a légèrement diminué, mais l'équation est devenue plus simple après la suppression des variables avec peu de pouvoir explicatif de la variable dépendante. Ce coefficient R = 0,98691 signifie que la corrélation du modèle était modérément positive (figure I.2). La probabilité (Pr <.0001) était strictement inférieure à 0,5%, confirmant que le modèle était significatif.

La valeur du niveau de signification p et la valeur du rapport F, qui fournissent une mesure de la signification statistique du modèle de régression, ont également été déterminées. Une valeur élevée de F avec une valeur minimale de p signifie que l'équation est significative [114].

L'effet des facteurs (Xi), leurs interactions et leurs effets quadratiques peuvent être déduits simultanément par le modèle proposé. En effet, le tableau I.1 montre les effets de chaque facteur indépendant, son interaction avec les autres facteurs et son effet quadratique sur la dose de coagulant (Y).

En effet, les coefficients de chaque facteur dans le modèle permettent d'évaluer l'impact de chaque facteur sur la réponse [115]. Ainsi, il est évident que les changements de turbidité et de TAC dans l'effluent traité ont augmenté la dose de coagulant (figure I.1). En revanche, les changements de pH et de conductivité ont eu un effet négatif sur la dose de coagulant (figure I.1).

Les interactions et effets quadratiques des facteurs montrent que la relation entre la réponse et les facteurs n'est pas toujours linéaire [115]. Utilisé à différents niveaux dans le processus de coagulation et lorsque plusieurs facteurs sont modifiés simultanément, un facteur peut produire différents degrés de réponse. Ainsi, les interactions entre facteurs (pH - Température, pH - Conductivité, pH -TAC et Turbidité -TAC) ont montré des effets positifs sur les doses de coagulant (Figure I.1), alors que les interactions entre facteurs (Température - Conductivité et Température - Turbidité) ont eu des effets négatifs sur doses de coagulant (Figure I.1). De plus, les résultats ont montré deux facteurs avec des effets quadratiques intéressants, à savoir la conductivité et le TAC, indiquant respectivement des effets quadratiques positifs et négatifs (Figure I.1).

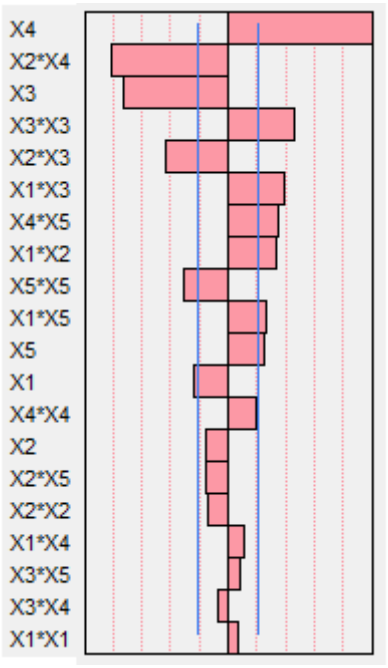


Figure I.1. Effets des facteurs de fonctionnement et de leurs interactions.

Au vu des résultats obtenus, il apparaît qu'à chaque augmentation du pH (7,25 à 8,78) et de la température (10,65 à 28,3 ° C), la dose de coagulant utilisé pour le traitement est passée de 109,51 mg / l à 83,15 mg / l pour pH et de 109,15 mg / l à 78,28 mg / l pour la température. La dose de coagulant a également diminué (de 109,51 mg / l à 75,24 mg / l) avec l'augmentation de la conductivité (de 106 µs / cm à 1243,54 µs / cm), puis a augmenté à 85,51 mg / l pour 1871 µs / cm. En revanche, les résultats obtenus montrent une augmentation de la dose de coagulant (de 47,16 mg / l à 109,51 mg / l) lorsque la turbidité augmente (de 0,86 NTU à 1039 NTU). Enfin, on peut noter que l'effet du TAC sur la dose de coagulant n'était pas linéaire. Ainsi, la dose de coagulant utilisé pour le traitement des effluents a augmenté (de 88,55 mg / l à 109,51 mg / l) pour des valeurs de TAC allant de 88,55 ° F à 19,71 ° F; puis diminué (101,88 g / l) pendant 24,4 ° F. Ces résultats confirment l'effet des facteurs de cette étude, à savoir le pH, la température, la conductivité, la turbidité et le TAC pendant le processus de coagulation [58,76,80,83].

Pour une dose optimale de coagulant choisie égale à 109,50 mg / l, une eau ayant les propriétés suivantes: pH 7,25, température 10,65 ° C, cond 106 µs / cm, turbidité 1039 NTU et TAC = 19,71 ° F a fourni un traitement efficace à 99,96%.

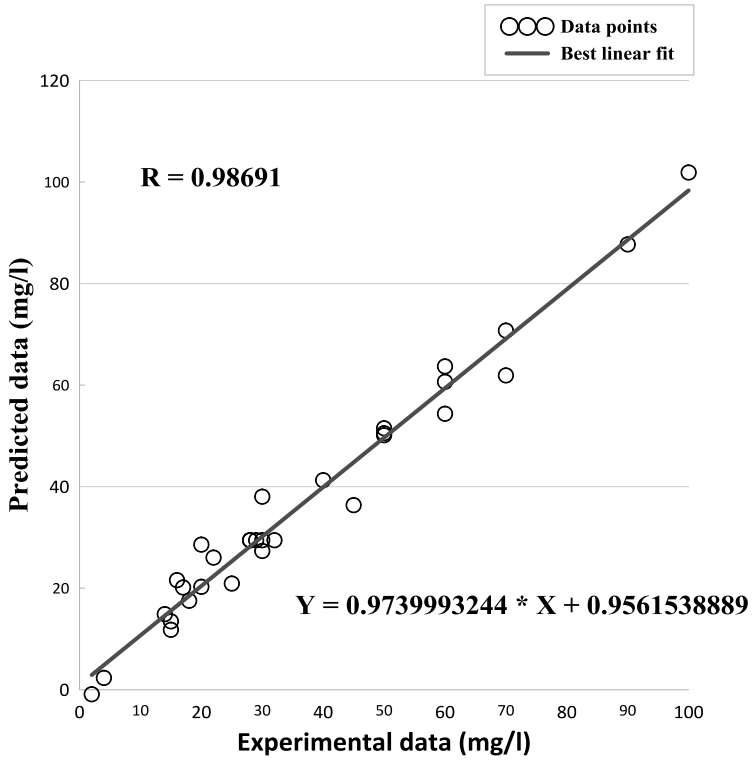


Figure I.2. Relation entre la dose observée de coagulant et celles estimées par le modèle RSM

Les résultats des performances du RSM en termes de toutes les erreurs et en termes de valeurs de vecteur d'accord (R, pente: α et intersection: β) sont donnés dans le tableau I.2.

Tableau I.2. RSM performances

	RMSE (mg/l)	*MSE (mg/l)*	*ESP* (%)	*EPM* (%)	MAE *(mg/)l*	*R*	A	β
Tous	*3.82890*	*14.66053*	*10.41194*	*1.47823*	*3.82890*	*0.98691*	*0.97399*	*0.95615*

I.2. RNA

Dans cette partie, nous avons optimisé 3 fonctions d'activation avec le nombre de neurones cachés (de 3 à 10). Les résultats de ces tests sont présentés dans le tableau I.3.

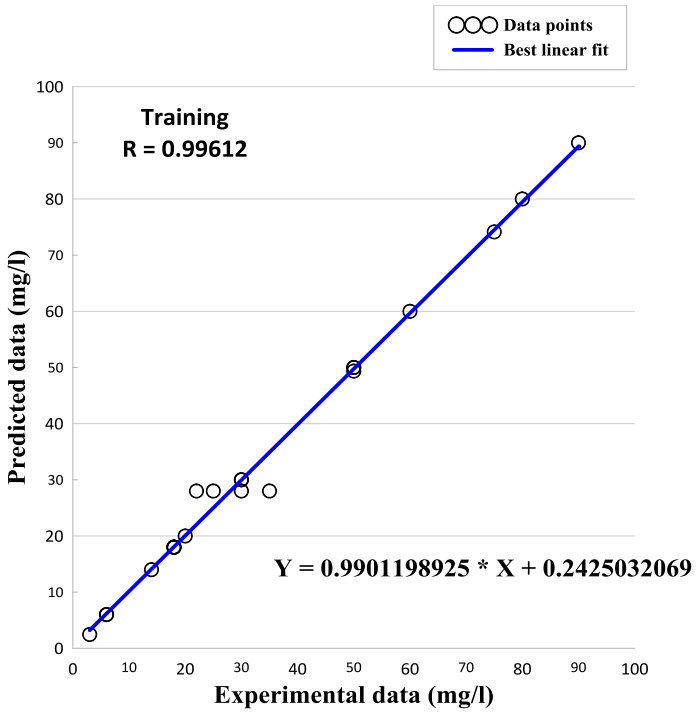
Le tableau I.3 présente les meilleures architectures trouvées. Il montre les coefficients de corrélation et l'erreur pour chaque apprentissage et validation en fonction du nombre de neurones dans la couche cachée et de la topologie du réseau. Il montre également les fonctions d'activation de la couche cachée et de la couche de sortie. L'architecture 1 (tableau I.3) semble être le modèle RNA le plus pertinent pour prédire les doses de coagulant. Le coefficient de corrélation et le RMSE étant quasiment égaux entre les architectures 1 et 3, le plus petit nombre

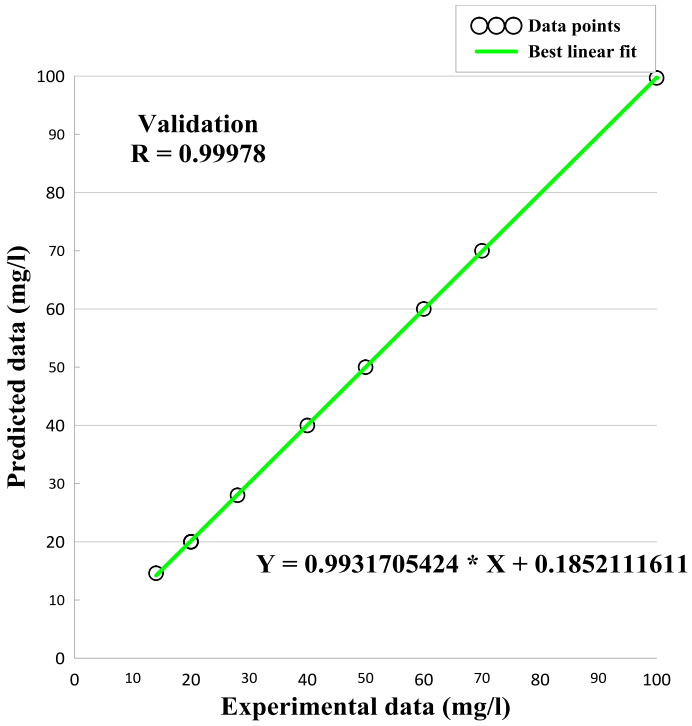
de neurones de la couche cachée a été choisi afin d'avoir le plus petit nombre de paramètres possible.

Tableau I.3. Performances des différentes architectures RNA testées

RNA		Fonction d'activation		Coefficients de correlation	RMSE
Nbr	**Neurones / couche**	**Couche cachée**	**Couche sortie**	**Totale**	**Totale**
1	**[5-6-1]**	**logsig**	**Purelin**	**0.9975**	**1.7954**
2	[5-9-1]	tansig	Purelin	0.99744	1.8132
3	[5-10-1]	tansig	Tansig	0.99751	1.7908

Une fois l'architecture de RNA synthétisée, elle a été validée; les valeurs prédites (valeurs estimées) ont été comparées aux valeurs expérimentales dans les deux étapes des phases d'apprentissage et de validation, et les résultats correspondants sont donnés dans la figure I.3. Les résultats ont été dénormalisés en valeurs réelles en vue de la comparaison avec les autres modèles.





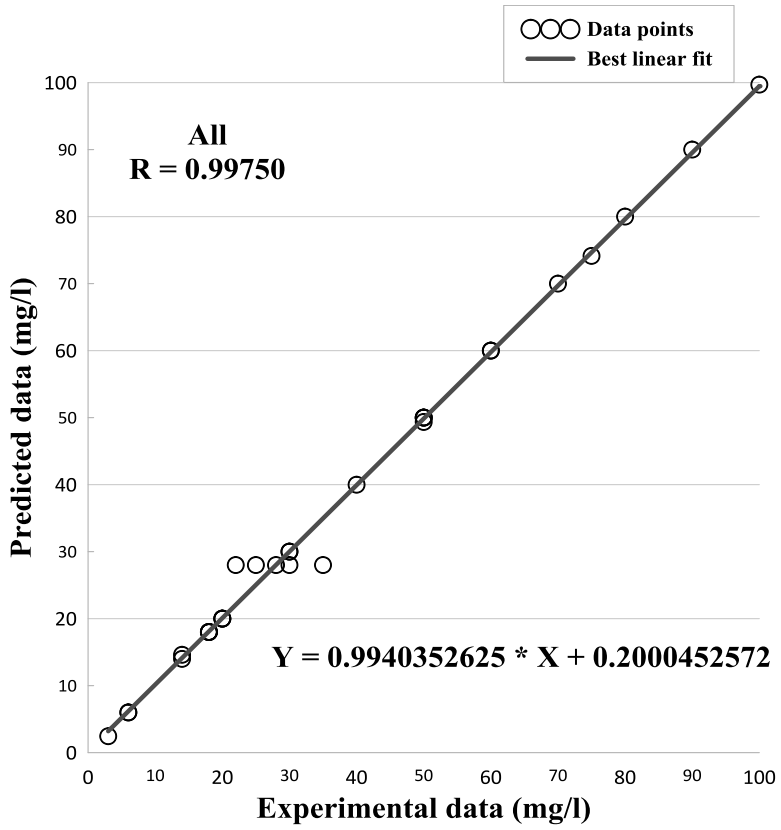


Figure I.3. Relation entre la dose observée de coagulant et celles estimées par le modèle RNA.

Les résultats de la performance RNA en termes de toutes les erreurs et en termes de valeurs de vecteur d'accord (R, pente: α et ordonnée à l'origine: β) sont donnés dans le tableau I.4.

Tableau I.4. Performances de la prédiction utilisant RNA

	RMSE (mg/l)	*MSE (mg/l)*	*ESP* (%)	*EPM* (%)	MAE *(mg/l)*	*R*	α	β
Tous	*1.79535*	*3.22331*	*4.83966*	*0.28386*	*0.67928*	*0.99750*	*0.99403*	*0.20004*
Formation	*2.12626*	*4.521011*	*6.25372*	*0.19098*	*0.91414*	*0.99612*	*0.99011*	*0.24250*
Validation	*0.22616*	*0.05115*	*0.50634*	*0.00428*	*0.10519*	*0.99978*	*0.99317*	*0.18521*

I.3. SVM

Dans cette partie, nous avons optimisé 3 fonctions du noyau, chaque noyau synthétisé par des paramètres qui doivent également être optimisés:

La fonction linéaire synthétisée par Box Constraint, epsilon et sigma, la fonction gaussienne synthétisée par Box Constraint, epsilon et sigma; la fonction polynomiale synthétisée par Box

Constraint, epsilon et Polynomial Order, sachant que le PolynomialOrder a été optimisé de 2 à 5.

Une fois le résultat de la phase d'apprentissage obtenu, il a été validé par la base de données de validation. Les résultats obtenus (valeurs prédites) ont été comparés aux valeurs expérimentales dans les deux étapes (la phase d'apprentissage et la phase de validation) afin d'avoir le coefficient de corrélation et RMSE. Les résultats de ces tests sont présentés dans le tableau I.5. Il convient de noter que les résultats ont été dénormalisés aux valeurs réelles pour les comparer aux autres modèles.

Tableau I.5. Performances des différentes fonctions du noyau SVM testées

Function de Noyau	C	ε	σ^2	d	Quantité de vecteurs de support	Coefficients de corrélation	RMSE
						Total	Total
Linear	846	0.0032	/	/	22	0.66564	19.04920
Gaussian	**479.0010**	**2.5551e-04**	**2.3000**	**/**	**22**	***0.99668***	***2.08927***
Polynomial	816	0.0052	/	2	22	0.92786	9.54540
Polynomial	23	0.0035	/	2.5	22	0.93081	9.3885
Polynomial	102	0.0012	/	3	22	0.98170	4.87540
Polynomial	614	0.0013	/	3.5	22	0.98311	4.67220
Polynomial	623	0.001	/	4	22	0.98672	4.17170
Polynomial	731	5.8782e-04	/	4.5	22	0.99272	3.16830
Polynomial	855	4.3554e-04	/	5	22	0.99474	0.9947

Il est évident que la fonction gaussienne donne un meilleur résultat en termes de coefficient de corrélation et de RMSE; ce résultat a été schématisé graphiquement par la figure I.4.

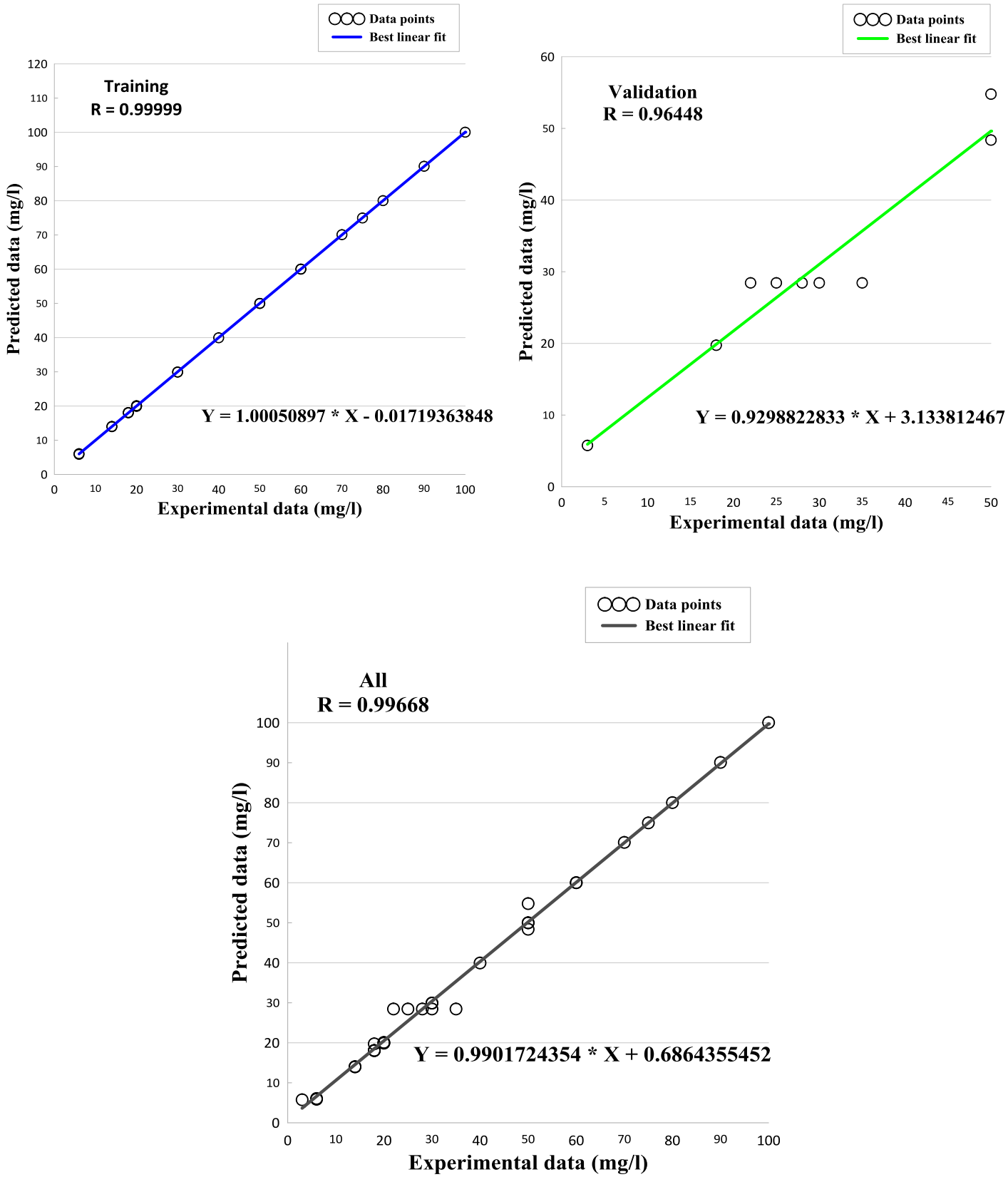


Figure I.4. Relation entre la dose observée de coagulant et celles estimées par le modèle SVM

Les résultats de la performance SVM en termes de toutes les erreurs et en termes de valeurs de vecteur d'accord (R, pente: α et intersection: β) sont donnés dans le tableau I.6.

Tableau I.6. Performances de la prédiction à l'aide de SVM

	RMSE (mg/l)	*MSE (mg/l)*	*ESP* (%)	*EPM* (%)	MAE *(mg/)l*	*R*	α	β
Tous	*2.08927*	*4.36505*	*5.63194*	*0.58983*	*0.98625*	*0.99668*	*0.99017*	*0.68643*
Formation	*0.06292*	*0.00395*	*0.15571*	*0.01155*	*0.05477*	*0.99999*	*1.00050*	*-0.01719*
Validation	*3.87627*	*15.02550*	*13.36646*	*0.16488*	*3.26318*	*0.96448*	*0.92988*	*3.13381*

I.4. ANFIS

Dans ce travail, nous avons optimisé les algorithmes de pondération, les fonctions (trimf, trapmf, gbellmf, gaussmf, gauss2mf, pimf, dsigmf et psigmf) pour l'entrée et (constante et linéaire) pour la sortie et aussi de nombreux nœuds pour chacun entrée pour obtenir le résultat le plus précis; il est à noter que ces algorithmes et fonctions se trouvent dans la boîte à outils de Matlab R 2013.

Comme décrit ci-dessus, deux étapes ont été suivies pour la modélisation ANFIS, où deux ensembles de données ont été utilisés, respectivement les données d'apprentissage et de vérification. Tout d'abord, la formation a été effectuée à l'aide de l'ensemble de données de formation. Deuxièmement, la vérification de l'ensemble de données a été envisagée pour vérifier l'exactitude et l'efficacité du modèle ANFIS.

Deux critères ont été retenus pour optimiser le modèle ANFIS: le nombre de fonctions d'appartenance affectées aux entrées et sorties et la valeur RMSE. En outre, les quatre aspects suivants ont été pris en compte pour la phase d'apprentissage de l'ANFIS, à savoir le nombre de fonctions d'appartenance utilisées, le type de fonctions d'appartenance, le surapprentissage et les options de formation.

Les résultats de ces tests sont présentés dans le tableau I.7.

Tableau I.7. Performances des différentes architectures ANFIS testées

ANFIS		*Type MF*		*Train FIS*	*Coefficients des corrélations*	*RMSE*
Nbr	*Nombre de MF*	*Entrée*	*Sortie*	*Méthode d'optimisation*	*Totale*	*Totale*
1	*[2 2 2 2 2]*	*trimf*	*linéaire*	*hybride*	*0.93691*	*9.24210*
2	*[2 2 2 2 2]*	*trapmf*	*linéaire*	*Hybride*	*0.95284*	*7.83380*
3	***[2 2 2 2 2]***	***gbellmf***	*linéaire*	***Hybride***	***0.97137***	***6.14477***

4	*[2 2 2 2 2]*	*gaussmf*	*linéaire*	*Hybride*	*0.96934*	*6.35520*
5	*[2 2 2 2 2]*	*gauss2mf*	*linéaire*	*Hybride*	*0.96842*	*6.45090*
6	*[2 2 2 2 2]*	*pimf*	*linéaire*	*Hybride*	*0.96839*	*6.45360*
7	*[2 2 2 2 2]*	*dsigmf*	*linéaire*	*Hybride*	*0.96875*	*6.42420*
8	*[2 2 2 2 2]*	*psigmf*	*linéaire*	*Hybride*	*0.96875*	*6.42420*

Les meilleurs résultats obtenus sont rassemblés dans le tableau I.7; ils sont illustrés sur la figure I.5. en dénormalisant les valeurs en valeurs réelles.

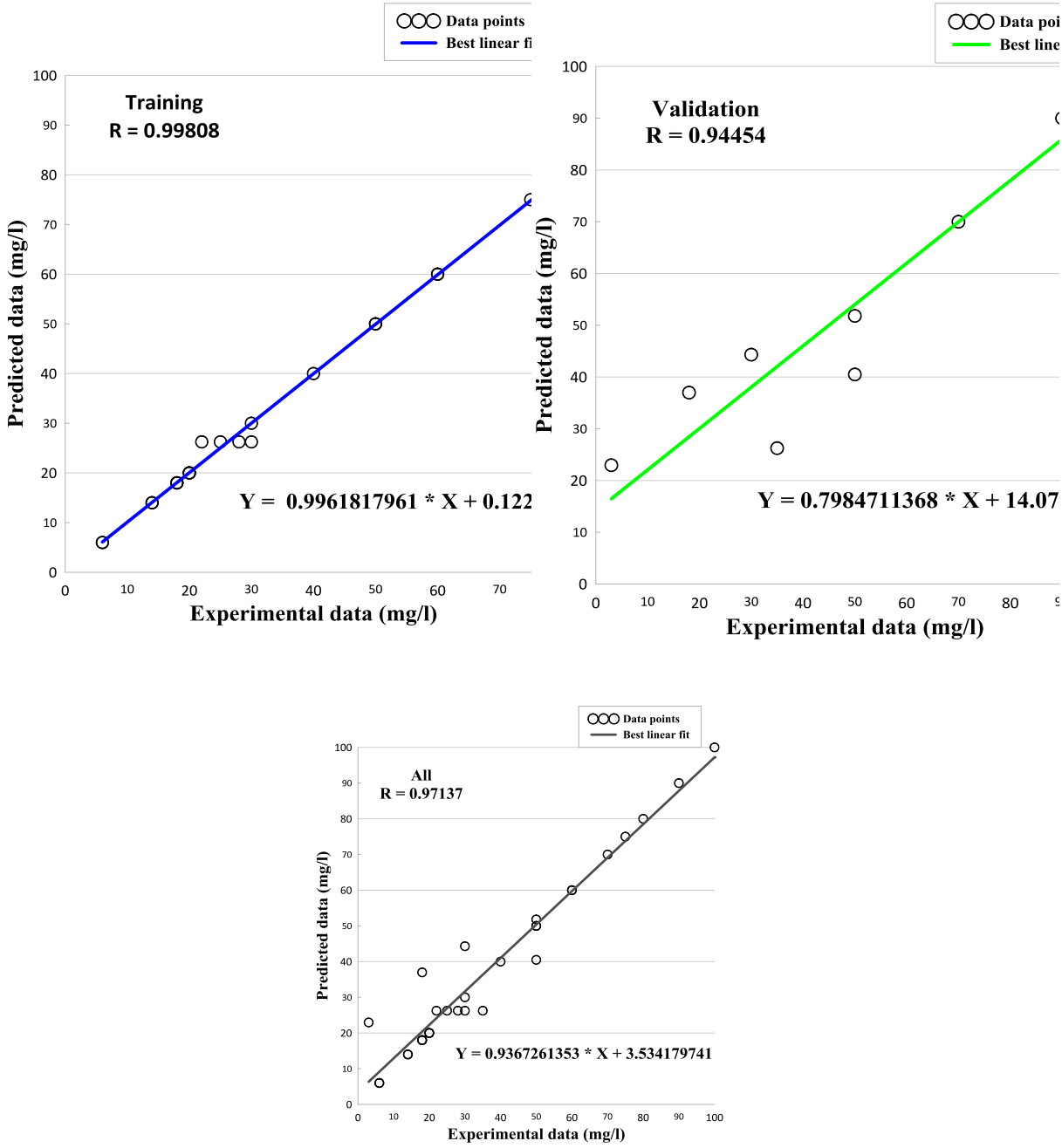


Figure I.5. Relation entre la dose de coagulant observée et celles estimées par le modèle ANFIS

Les résultats des performances de l'ANFIS en termes de toutes les erreurs et en termes de valeurs de vecteur d'accord (R, pente: α et intersection: β) sont donnés dans le tableau I.8.

Tableau I.8. Performances de la prédiction à l'aide de SVM

	RMSE (mg/l)	*MSE (mg/l)*	*ESP* (%)	*EPM* (%)	MAE *(mg/)l*	*R*	α	β
Tous	*6.14477*	*37.75828*	*16.56418*	*2.81801*	*2.71994*	*0.97137*	*0.93672*	*3.53417*
Apprentissage	*1.29246*	*1.67045*	*4.03893*	*0.09475*	*0.05000*	*0.99809*	*0.99618*	*0.12218*

I.5. Comparaison des modèles

La comparaison des coefficients de corrélation, de détermination, d'ajustement et d'indicateurs statistiques (RMSE, MSE, ESP, EPM et MAE) obtenus par les quatre modèles, RSM, RNA, SVM et ANFIS, pour l'ensemble des données est présentée dans le tableau I.9. Les coefficients du modèle RNA étaient légèrement supérieurs à ceux obtenus par le modèle SVM et significativement supérieurs à ceux obtenus par les modèles RSM et ANFIS. Ces résultats ont également été confirmés par les indicateurs statistiques, puisque ceux donnés par le modèle RNA étaient légèrement inférieurs à ceux obtenus par le modèle SVM, suivi du modèle RSM et enfin du modèle ANFIS.

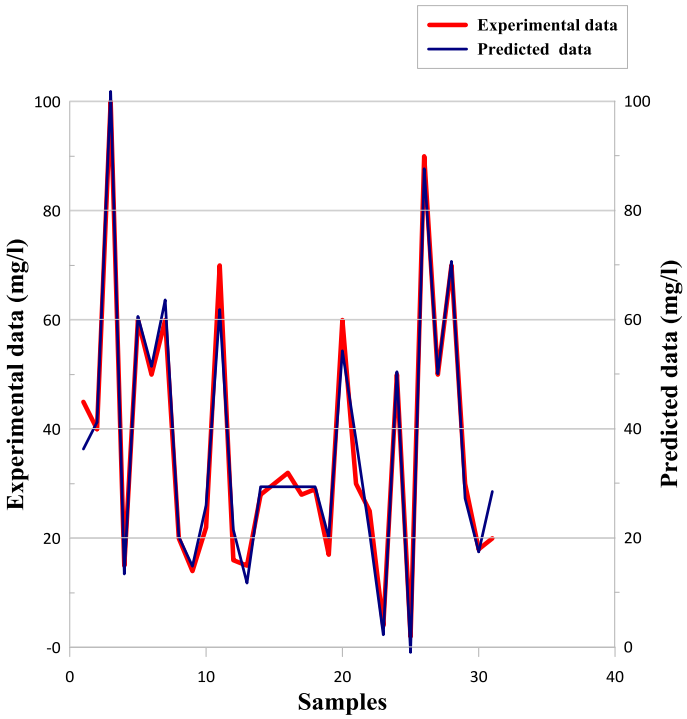
Tableau I.9. Comparaison des performances prédictives des modèles RSM, RNA, SVM et ANFIS pour toutes les données

Model	*RMSE (mg/l)*	*MSE (mg/l)*	*ESP* (%)	*EPM* (%)	MAE *(mg/l)*	*R*	A	B	R^2	R^2 adj	Nbre de parameters
RSM	*3.82890*	*14.66053*	*10.41194*	*1.47823*	*3.82890*	*0.98691*	*0.97399*	*0.95615*	*0.97399*	*0.96878*	*13*
RNA	*1.79535*	*3.22331*	*4.83966*	*0.28386*	*0.67928*	*0.99750*	*0.99403*	*0.20004*	*0.99500*	*0.99400*	*43*
SVM	*2.08927*	*4.36505*	*5.63194*	*0.58983*	*0.98625*	*0.99668*	*0.99017*	*0.68643*	*0.99337*	*0.99204*	*22*
ANFIS	*6.14477*	*37.75828*	*16.56418*	*2.81801*	*2.71994*	*0.97137*	*0.93672*	*3.53417*	*0.94355*	*0.93227*	*309*

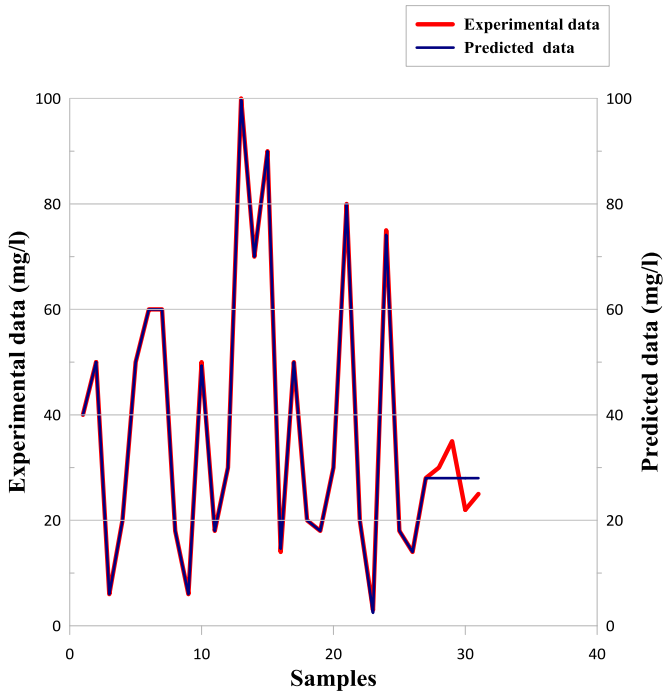
La comparaison des différents modèles est illustrée sur la figure I.6, confirmant la supériorité du modèle RNA et SVM sur les autres modèles (RSM, ANFIS) avec une très bonne superposition des valeurs expérimentales et celles données par les modèles. Bien que les modèles RNA et SVM soient presque égaux, le modèle RNA était légèrement supérieur. Le coefficient de corrélation et les indicateurs statistiques de la phase d'apprentissage du modèle RNA étaient presque égaux à ceux donnés par le modèle SVM avec une légère supériorité du modèle SVM (tableaux I.4 et I.6). Cependant, dans la phase de validation, le coefficient de corrélation était très élevé et les indicateurs statistiques étaient très petits pour le modèle RNA

(R = 0,99978, RMSE = 0,22616, MSE = 0,05115, ESP = 0,50634, EPM = 0,00428 et MAE = 0,10519) par rapport à le modèle SVM (R = 0,96448, RMSE = 3,87627, MSE = 15,02550, ESP = 13,36646, EPM = 0,16488 et MAE = 3,26318) (tableaux I.5 et I.7). Mais, en termes de réduction des coûts économiques, le modèle SVM était supérieur, puisque le nombre de paramètres de ce modèle était de 22, soit presque la moitié du nombre de paramètres du modèle RNA (43 paramètres). Pour réduire encore les coûts économiques, on peut également utiliser le modèle RSM qui reste très utile du fait de ses coefficients élevés (R = 0,98691, R^2 = 0,97399, R^2 adj = 0,96878) par rapport au nombre de paramètres, à savoir 13. De plus, le les indicateurs statistiques sont restés acceptables pour le modèle RSM (RMSE = 3,82890, MSE = 14,66053, ESP = 10,441194, EPM = 1,47823 et MAE = 3,82890) s'ils sont liés au nombre de paramètres du modèle.

Enfin, le modèle ANFIS qui était acceptable du point de vue des coefficients et des indicateurs statistiques (R = 0.97137, R^2 = 0.94355, R^2 adj = 0.93227, RMSE = 6.14477, MSE = 37.75828, ESP = 16.56418, EPM = 2.81801 et MAE = 2.71994), est apparu très coûteux car le nombre de paramètres était très élevé (309 paramètres) et les résultats moins précis par rapport aux autres modèles.



RSM Model



RNA Model

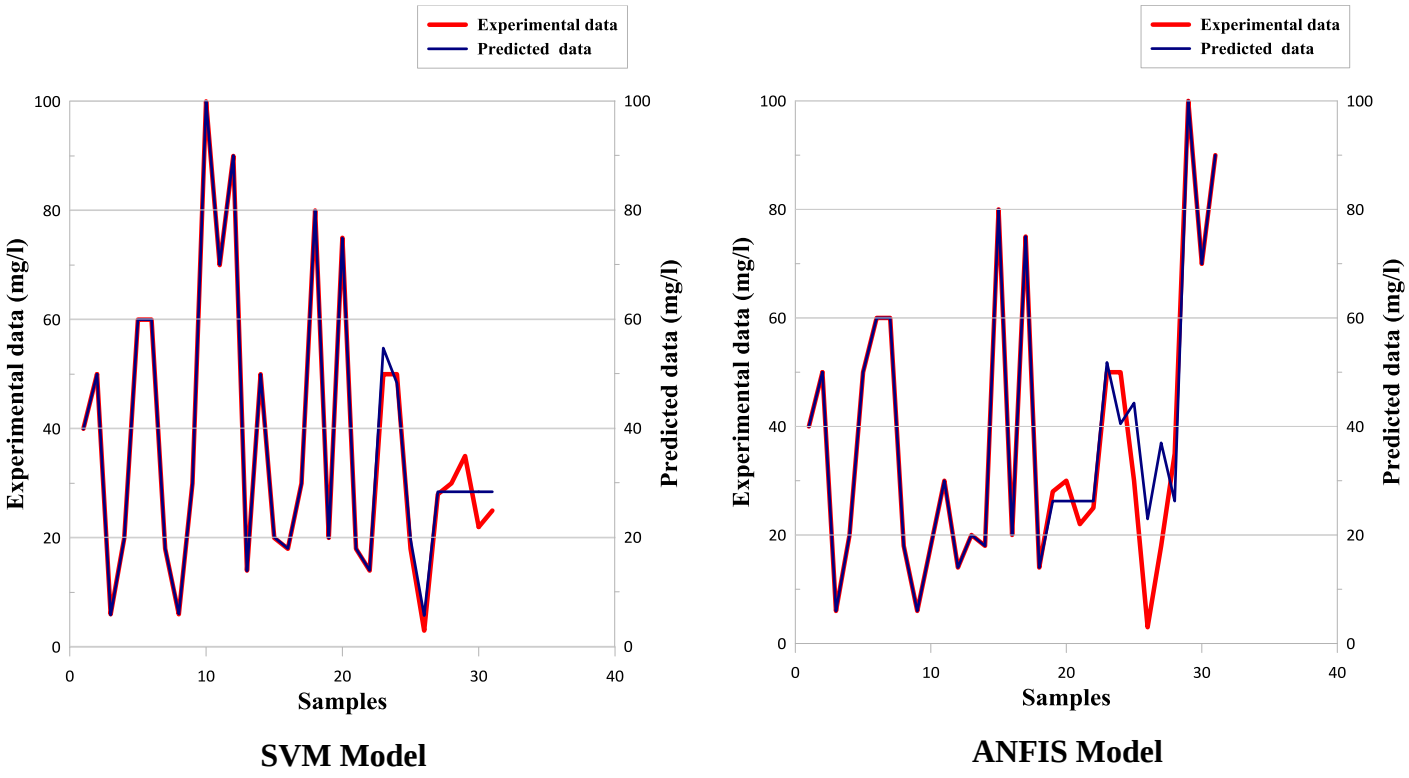


Figure I.6. Relation entre les données expérimentales et les données prédites à l'aide de la modélisation RSM, RNA, SVM et ANFIS

I.6. résidus

L'analyse des résidus est un moyen efficace de révéler les performances du modèle optimisé; il consiste en la mesure de l'erreur absolue ou relative entre les valeurs expérimentales et prédites [116]. Les méthodes d'analyse des résidus sont principalement des méthodes d'analyse graphique. La figure I.7 montre les résidus liés à chaque modèle (RSM, RNA, SVM et ANFIS). Cette figure montre que les résidus obtenus par la méthode des réseaux de neurones étaient moins dispersés (proches de zéro) par rapport à ceux obtenus par les autres modèles. Là encore, ce résultat montre la précision et la robustesse du modèle RNA par rapport aux autres modèles et justifie l'utilisation de l'approche RNA dans la prédiction des doses de coagulant.

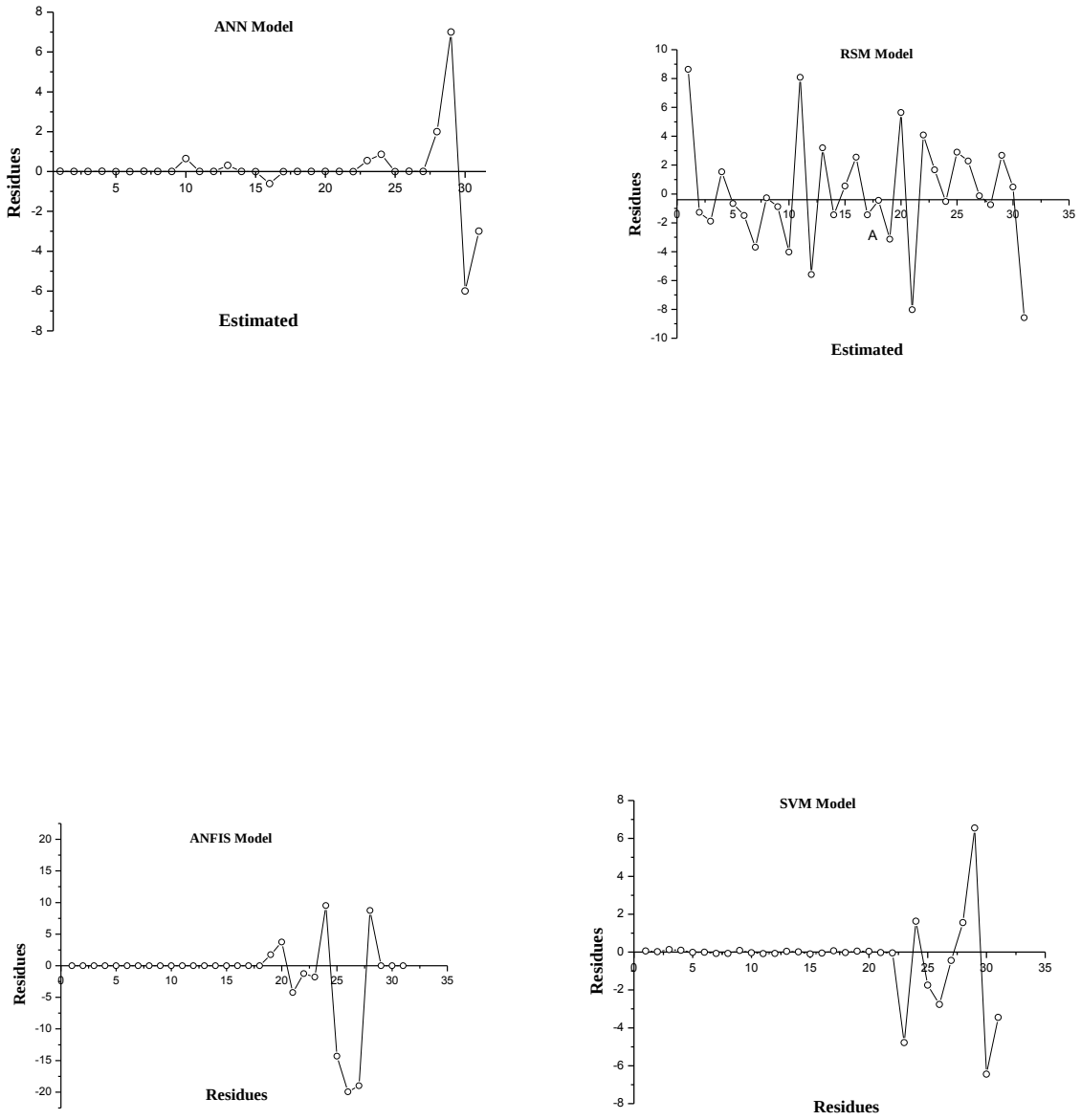


Figure I.7. Distribution des résidus pour les modèles RSM, RNA, SVM et ANFIS pour toutes les données.

Partie II : Caractérisation des eaux brutes et traitées

II.1. Résultats des analyses physico-chimiques de l'eau brute et traitées

Après une étude primaire, nous avons constaté que l'eau brute de chafa est la plus contaminée. Il faut mentionner que nous avons réalisé cette étude pendant une période de contamination de l'eau par choléra.

II.1.1. Paramètres physiques

II.1.1.1. Potentielle d'hydrogène (pH)

Les résultats de la mesure de pH sont représentés dans la figure II.1 de l'eau brute et traitée.

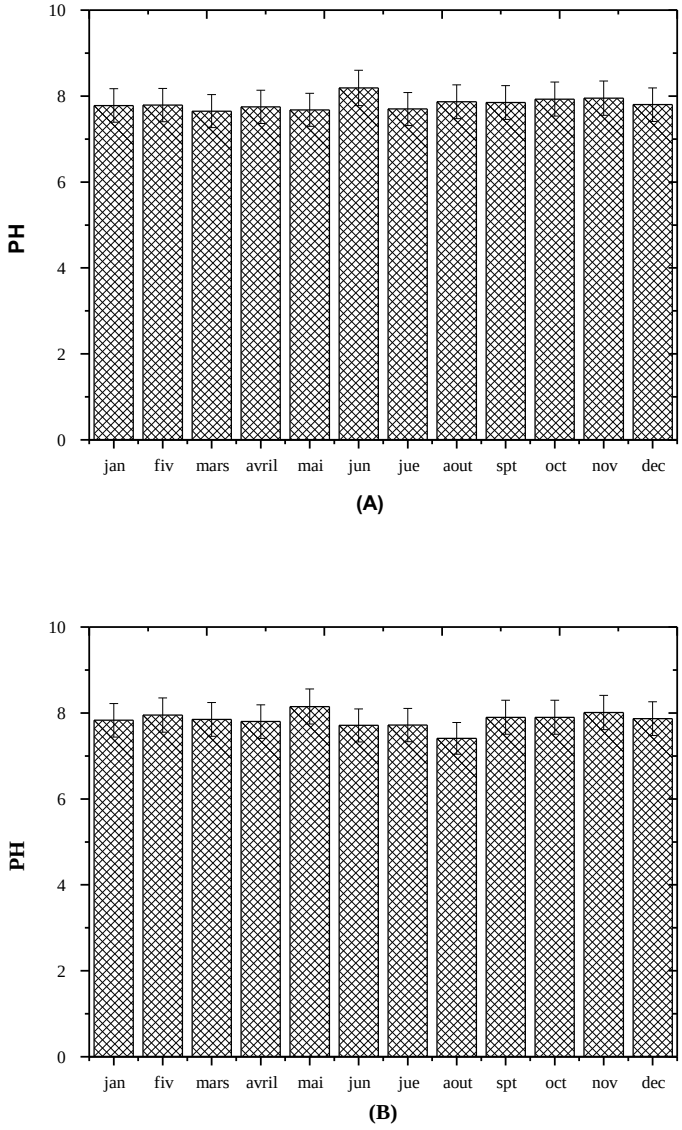


Figure II.1. Variation des moyennes mensuelles du pH en fonction du temps (mois) des eaux brutes (A) et traitées (B)

La figure (II.1) révèle que l'eau brute sont alcalines et dépassent les 7 unités de pH durant toute la période d'étude et une variabilité des valeurs des moyennes mensuelles de pH de 7.65 et 8.19, ces valeurs de pH montrent la conformité aux normes d'eau potable.

Nous avons remarqué que les valeurs obtenues dans la figure (B) pour l'eau traitée sont légèrement inférieures à celles de l'eau brute. Cette légèrement diminution est due au traitement appliqué pour l'élimination de certaines charges par l'injection de sulfate d'alumine $AL_2(SO_4)_3$. IL est à noter que le pH des eaux naturelles est lié à la nature des terrains traversés[81]**.**

Selon **JORA**, les valeurs de pH montent une conformité aux normes.

De plus, au-dessus de 8, il y a une diminution progressive de l'efficacité du chlore sur la contamination microbienne[81,117,118].

Des études ont montrés que le pH de l'eau est influencé par la température, l'oxygène dissous, la minéralisation totale et bien d'autre facteurs microbienne[81,117,118].

II.1.1.2. Conductivité électrique

Variation de la conductivité est illustrée dans la figure II.2.

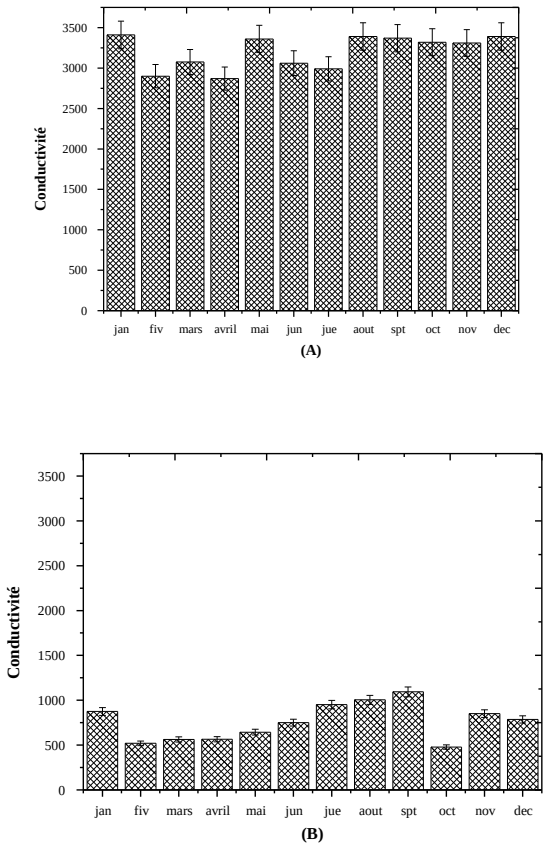


Figure II.2. Variation des moyennes mensuelles de la conductivité (µs/cm) en fonction du temps (mois) des eaux brutes (A) et traitées (B)

À partir de la Figure II.2, nous voyons que les valeurs de la conductivité électrique de les eaux brutes varient entre 2870 et 3410 µS / cm. Ces valeurs sont supérieures à la norme de l'OMS (valeur maximale de 2800 µS / cm). Nous avons remarqué aussi que les valeurs obtenues dans la figure II.2 (B) pour l'eau traitée sont inférieures à celles de l'eau brute.

D'après **Rodier et al (2009)**, une conductivité supérieure à 600 µS/cm implique une minéralisation importante des eaux. Ce qui est le cas de l'eau traité[81].

Lorsque la concentration des sels dissous (sels de sodium, de calcium et de magnésium, des bicarbonates, des chlorures) augmente dans l'eau, la conductivité électrique augmente également[119]**.** Cette augmentation du à l'injection de différents produits durant le traitement par l'hypochlorite de calcium.

D'après **USEPA (2003)**, la conductivité restera relativement constante tout au long d'un système de distribution aussi longtemps que l'eau est en équilibre avec le matériau de la conduite. Des changements de conductivité peuvent indiquer des problèmes de corrosion[120].

II.1.1.3. Turbidité

Variation de la turbidité est illustrée dans la figure II.3.

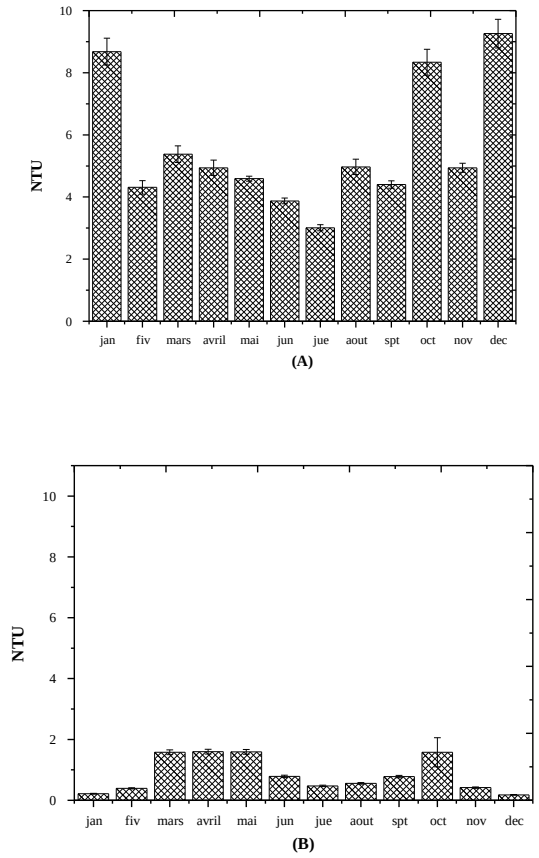


Figure II.3. Variation des moyennes de la turbidité (NTU) en fonctions du temps (mois) des eaux brutes (A) et traitées (B)

À partir de la Figure II.3, nous voyons que les valeurs de la turbidité de l'eau les brutes varient entre 1,59 et 9,26 NTU. Ces valeurs sont supérieures à la norme de l'OMS (valeur maximale de NTU).

Les valeurs retrouvées dans les eaux traitées sont inférieures à 2 NTU. Ces valeurs restent en deca du seuil toléré (<5NTU),[121]**.**

La turbidité est causé par les particules en suspension dans l'eau (débris organique : argile, organismes microscopiques….)[122]

En général, les valeurs augmentent fortement pendent et après une pluie. Une turbidité élevée peut être également à l'origine d'une augmentation de la température de l'eau et d'une diminution de l'oxygène dissous[123]

II.1.2. Paramètres de pollution

II.1.2.1. Matière organique

Variation de matière organique est illustrée dans la figure II.4.

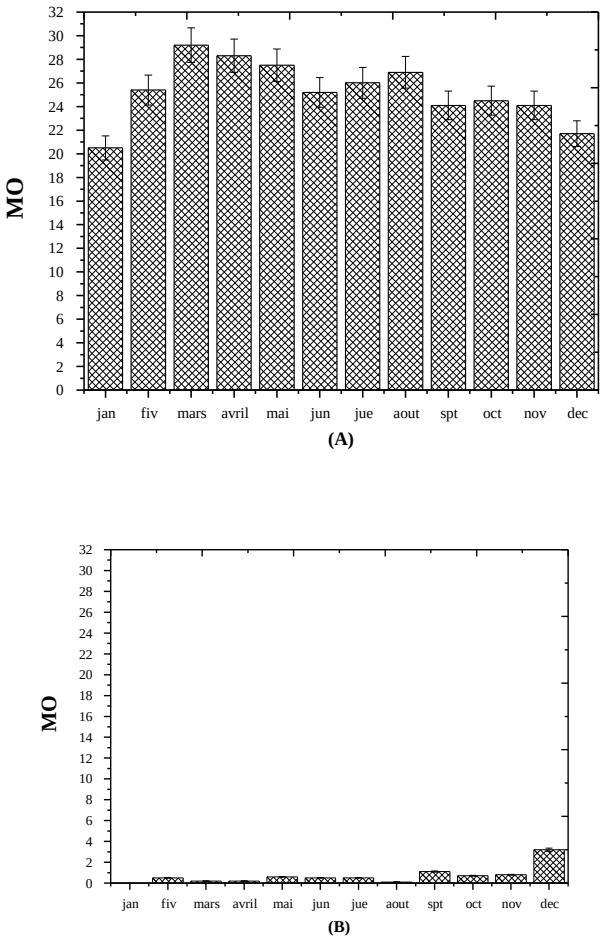


Figure II.4. Variation des moyennes mensuelles de matière organique (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B)

Les valeurs indiquées à la Figure. II.4 montrent des teneurs élevées en matière organique dans l'eau brute. Ces valeurs varient entre 20,5 et 29,2 mg / L, ce qui donne une incertitude aux normes algériennes.

Ces valeurs diminuent dans les eaux traitées, par le procédé de coagulation-floculation au sulfate d'alumine [124] **.**

D'après **Harrat et Achour (2010).** Le contact entre la matière organique des eaux de surface et de chlore induit la formation de sous-produits de chloration parmi lesquels le chloroforme dont les risques toxiques à long terme doivent être pris en considération[125].

II.1.2.2. Ammonium

À partir de la figure II.5, nous remarquons que l'eau brute ne répond pas aux normes d'eau potable. Par contre, les eaux traitées sont conformes aux normes de potabilité. Car, aucune concentration n'a été remarquée.

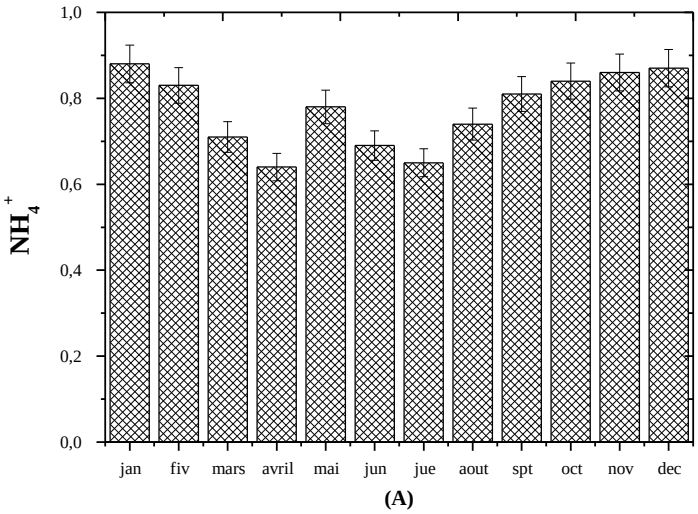


Figure II.5. Variation des moyennes mensuelles d'ammonium (mg/l) en fonction du temps (mois) des eaux brutes (A)

L'azote ammoniacal des eaux superficielles peut avoir pour origine la matière végétale des cours d'eau, la matière organique animale ou humaine (l'homme élimine 15 à 30 g d'urée par jour), les rejets industriels, les engrais, etc[81]

D'aprés **Rodier et al,(2009),** l'oxydation par le chlore permet d'éliminer l'azote ammoniacal dans l'eau. Selon **Rodier et al,(2009),** l'oxydation biologique et l'ammoniaque peut développer des zones anaérobies des certaines parties des réseaux de distribution et entériner ainsi des gouts désagréables en provoquant la corrosion des conduite[81]

II.1.2.3. Phosphate

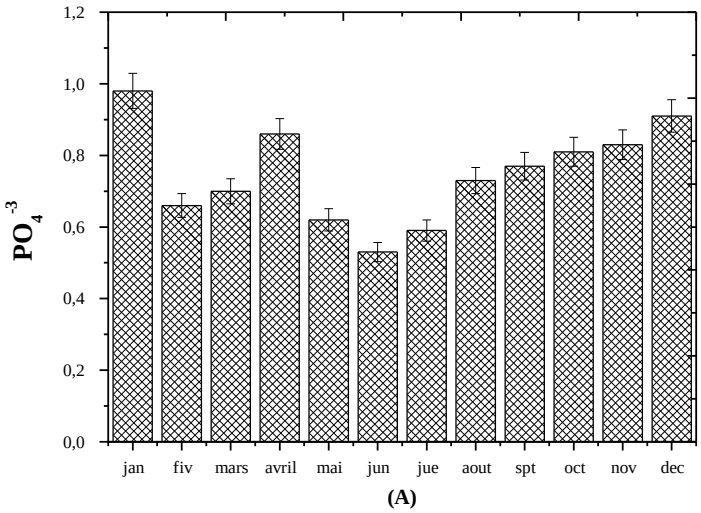


Figure II.6. Variation des moyennes mensuelles de phosphate (mg/l) en fonction du temps (mois) des eaux brutes (A)

La figure II.6, montre des teneurs enregistrées dépassent 0.5 mg/l dans les eaux brutes. En effet, la présence naturelle de phosphore dans les eaux est liée aux caractéristiques des terrains traversés et à la décomposition de la matière organique. Des teneurs supérieures à 0,5 mg/L doivent constituer un indice de pollution[81]. Par contre, dans l'eau traitée aucune concentration n'a été remarquée.

II.1.2.4. Nitrate

Les résultats de la mesure de nitrate sont représentés dans la figure II.7.

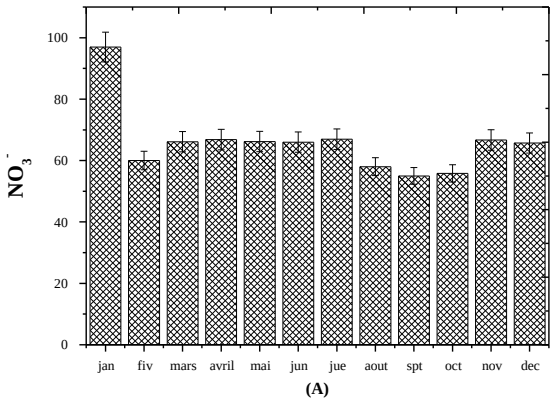


Figure II.7. Variation des moyennes mensuelles de nitrate (mg/l) en fonction de temps (mois) des eaux brutes (A).

La variation des concentrations de nitrate montre une grande différence dans les concentrations de nitrate d'eau brute variant entre 96,977 mg / l et 55 mg / l. Nous avons remarqué que les valeurs obtenues sont supérieures aux normes de l'eau algériennes. Par contre, dans l'eau traitée aucune concentration n'a été remarquée.

L'OMS recommande pour les eaux destinées à la consommation humaine une valeur limite de 50 mg/L (NO^-_3)[81]**.**

D'après **Rodier et al, (2005),** les nitrates proviennent de la décomposition naturelle, par des micro-organismes azotés tels que les protéines des végétaux, des animaux et des extra d'animaux. Ainsi, les eaux de pluies peuvent contenir des nitrates en provenance des oxydes d'azote et de l'ammoniac présents dans l'atmosphère. Elles peuvent également être liées à l'oxydation bactérienne de l'ammoniac[126]**.**

La concentration des eaux ayant des teneurs élevées en nitrate provoque l'irritation des muqueuses, des nausées, des vomissements, des selles sanguinolentes et des hypertensions[127]**.**

II.1.2.5. Nitrite

La figure II.8 montre la présence réduite de nitrite. Selon (Kurnaz, 2004), les nitrites se trouvent souvent à de faibles niveaux dans les eaux naturelles. Cependant, il faut noter que malgré ces valeurs faibles, elles restent néanmoins supérieures aux normes OMS et algériennes (valeur maximale 0,1 mg / L). Par contre, dans l'eau traitée aucune concentration n'a été remarquée.

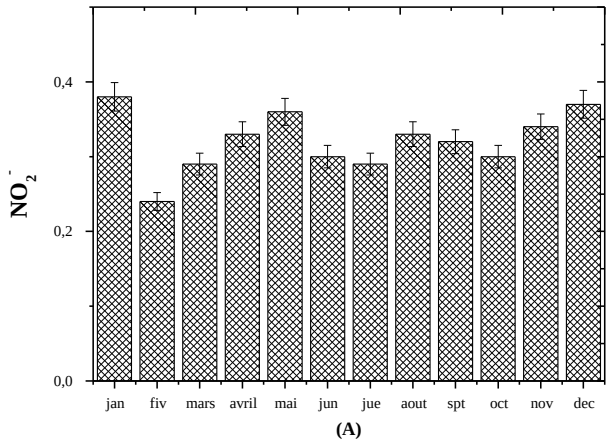


Figure II.8. Variation des moyennes mensuelles de nitrite (mg/l) en fonction de temps (mois) des eaux brutes (A).

L'origine des nitrites dans les eaux brutes peut être :

L'oxydation partielle de l'azote ammoniacale et sa transformation en nitrite sous l'action des bactéries nitrifiantes ;

La réduction des nitrates d'origine agricole en nitrites sous l'influence des bactéries dénitrifiantes[128] .

Selon **Rodier et al., (2009),** Les nitrites pouvant apparaître comme sous-produits lors de la chloration de l'eau, cette disposition permet de maintenir la chloration tout en garantissant un niveau adéquat de protection pour les consommateurs[81].

Pour l'eau destinée à la consommation humaine, l'OMS recommande une valeur guide provisoire de 3 mg/L (NO_2^-)[81]**.**

II.1.3. Minéralisation globale

II.1.3.1. Calcium, Magnésium et TH

Les résultats de la mesure de calcium, magnésium et la dureté sont représentés dans les figures II.9, II.10 et II .11.

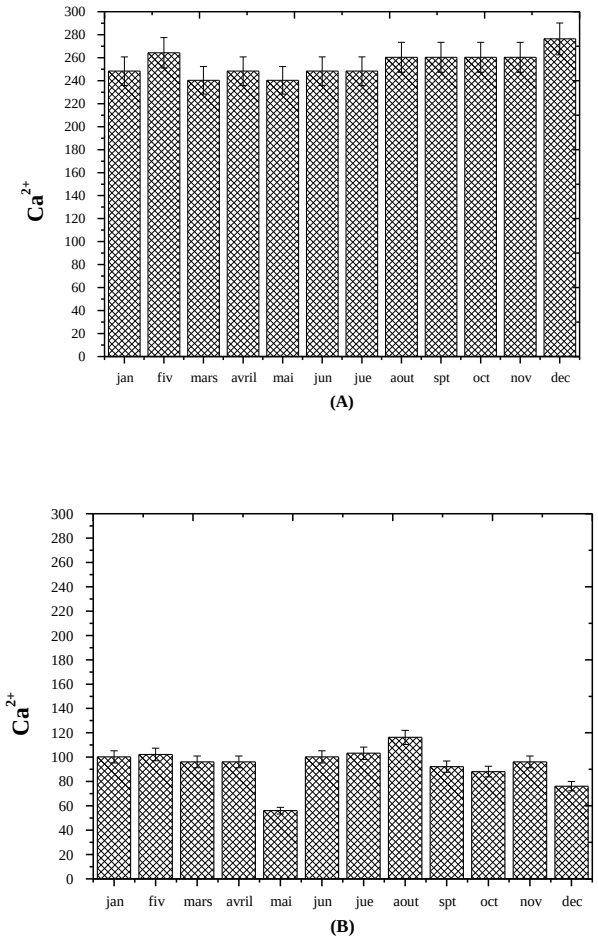


Figure II.9. Variation des moyennes mensuelles du calcium (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B).

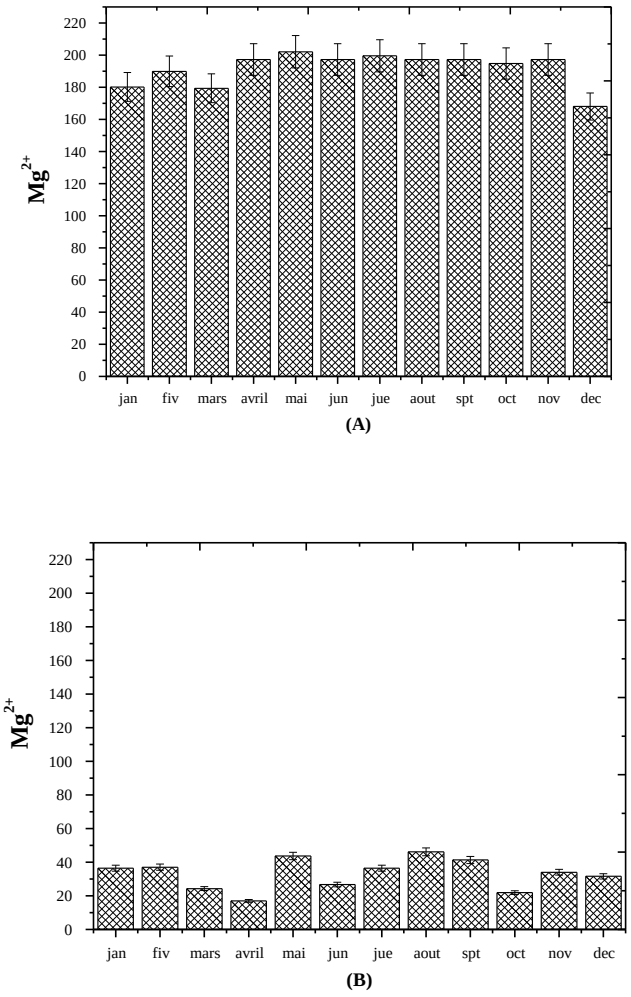
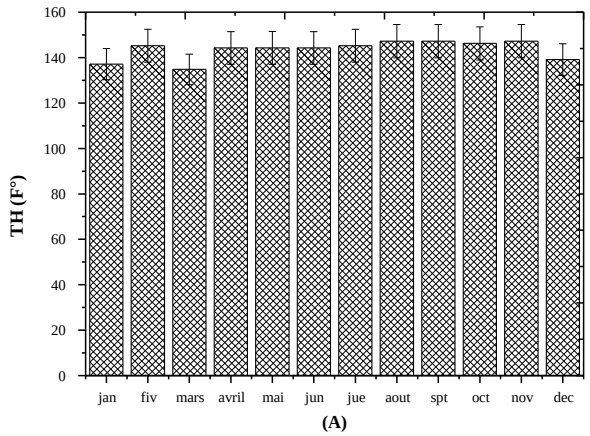


Figure II.10. Variation des moyennes mensuelles du Magnésium (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B).



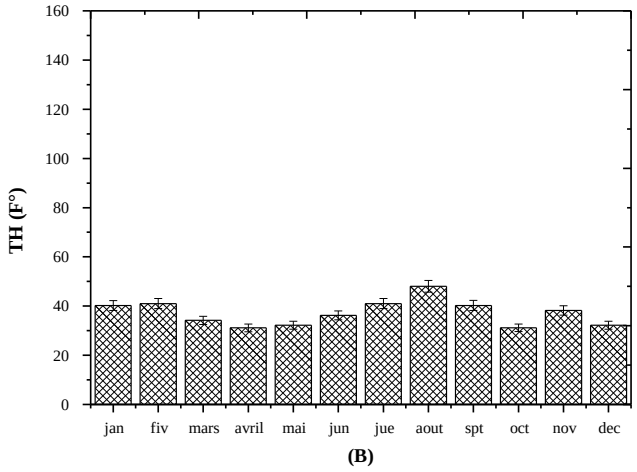


Figure II.11. Variation des moyennes mensuelles du la dureté (F°) en fonction du temps (Mois) des eaux brutes (A) et traitées (B).

On peut voir sur la figure II.9 que la teneur en calcium de l'eau brute varie entre 240,3 et 276,35 mg / l. En revanche, les concentrations mensuelles moyennes de magnésium de figure II.10 dans l'eau brute varient entre 168,04 et 202,1 mg / l. on remarque ces deux paramètres ne corresponde aux normes MSO et l'algériennes pour les eaux. Nous remarquons aussi que les concentrations en calcium et magnésium dans les eaux traitées ont des valeurs un peu élevées. Cela est dû à l'injection de l'hypochlorite de calcium ($Ca(ClO)_2$) durant l'étape de pré chloration. IL est impératif de constater que malgré ces valeurs soient un peu élevées, elles restent cependant Inférieures à la norme de l'OMS.

D'après **Rodier et al.,(2009),**la productivité des eaux douces est assez nettement liée à la teneur en sels de calcium et de magnésium ; son maximum correspondrait à une alcalinité totale en calcium de 25 mg/L[81].

Nous savons que la dureté est la somme du calcium et du magnésium. D'après les résultats très élevés qui ont été enregistrés avec le calcium et le magnésium, il est évident que l'eau est très dure, c'est ce qui a été confirmé par la figure II.11. Ces valeurs ont été démunies après traitement. Mais ont resté un peu élevées à cause de l'injection de l'hypochlorite de calcium. Malgré cette augmentation. Cependant, ces valeurs restent conforment aux normes recommandées par (**JORA, 2011)** (200°f)[121]

Selon **Rodier (2009),** il n'est pas facile de définir un niveau entre ce qui constitue une eau dure et une eau douce. Cependant, on peut considérer qu'une eau ayant une teneur inférieure à 75 mg/L de $CaCO_3$ (c'est-à-dire 7,5 ° de TH calcique) ou à 30 mg/L de Ca est une eau douce et qu'au-dessus, il s'agit d'une eau dure[81].

La dureté de l'eau est attribué à la présence d'ions calcium et magnésium. L'eau dure est généralement moins corrosive que l'eau douce en raison de l'augmentation des concentrations de calcium. Selon le pH de l'eau et la concentration d'alcalinité, le calcium va se combiner avec les carbonates pour former une couche protectrice sur la paroi du tube ce qui peut retard la corrosion [129,130]**.**

II.1.3.2. Chlorure

Variation de chlorure est représentée dans la figure II.12

La variation des concentrations des chlorures montre une différence dans les teneurs des chlorures dans les eaux brutes, varient entre 510 mg/l et 562,52 mg/l, ces valeurs ne corresponde pas aux normes MSO et l'algériennes pour les eaux. Ces valeurs ont été bien diminuées après le traitement à des valeurs inférieur à 90 mg/l sauf pour le mois novembre qui a été enregistré à une valeur de 124 mg/l. Cependant, cette valeur reste conforment aux normes recommandées par (**JORA, 2011)**[121].

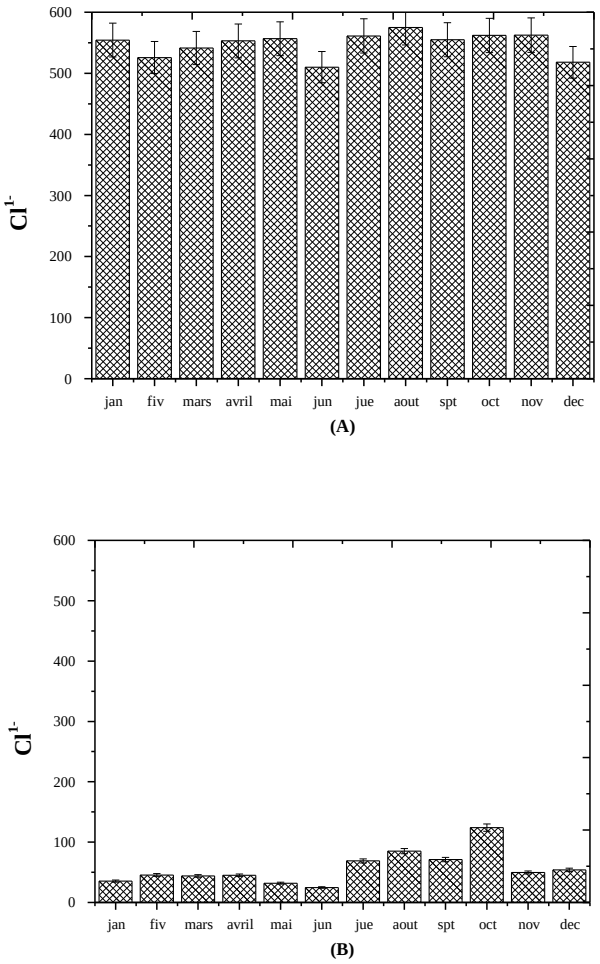


Figure II.12. Variation des moyennes mensuelles du chlorure (mg/l) en fonction de temps (mois) des eaux brutes (A) et traitées (B).

Le gros inconvénient des chlorures est la saveur désagréable qu'ils communiquent à l'eau à partir de 250 mg/L, surtout lorsqu'il s'agit de chlorure. Pour une quantité équivalente de chlorures, la saveur peut être moins marquée en présence de calcium et de magnésium. Ils sont aussi susceptibles d'amener une corrosion dans les canalisations et les réservoirs, en particulier pour les éléments en acier inoxydable, pour lesquels les risques s'accroissent à partir de 50 mg/L[81]**.**

II.1.3.3. Bicarbonate

Variation de bicarbonate est illustrée dans la figure II.13.

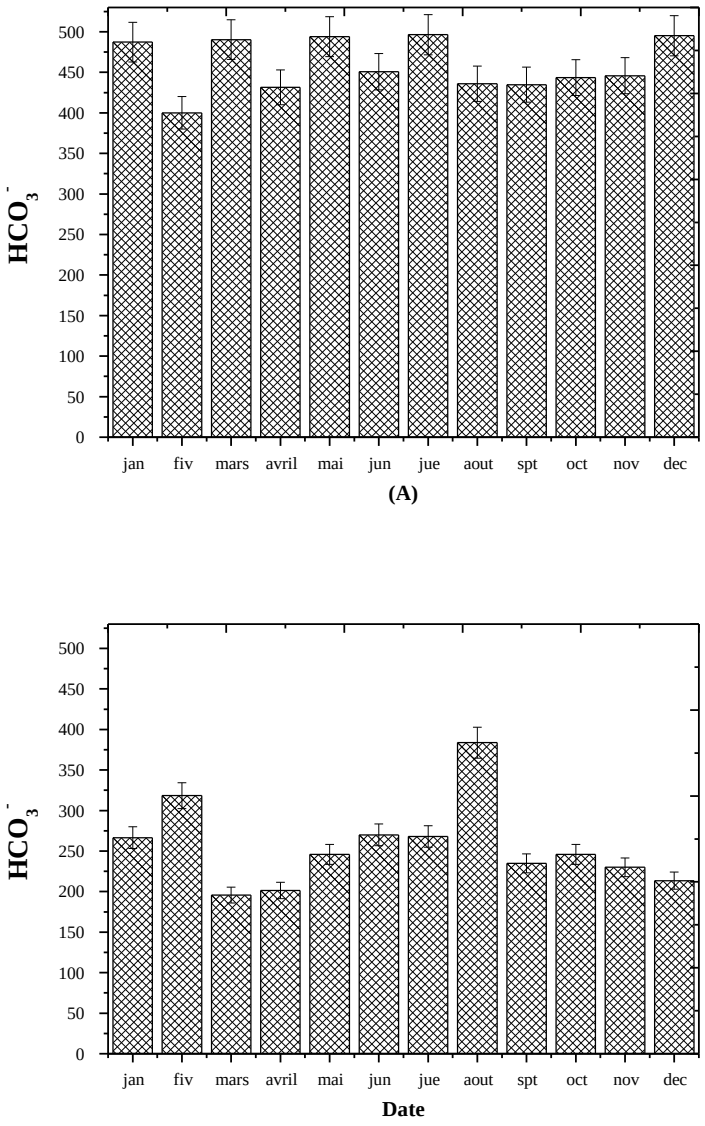


Figure II.13 variation des moyennes mensuelles du bicarbonate (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B).

À partir de la figure II.13, nous remarquons que l'eau brute ne répond pas aux normes d'eau potable. Par contre, les eaux traitées sont conformes aux normes de potabilité.

Selon **Rodier et al., (2005)**,la concentration des eaux en bicarbonates HCO^-_3est en fonction du pH de l'eau et de la température[81].

D'origine naturelle, les niveaux d'alcalinité totale allant jusqu'à 400 mg/L de $CaCO_3$ ne sont faibles peut entrainer des problèmes des corrosions du métal dans les systèmes de plomberie[131]**.**

II.1.3.4. Sulfate

Variation de sulfate est illustrée dans la figure II.14.

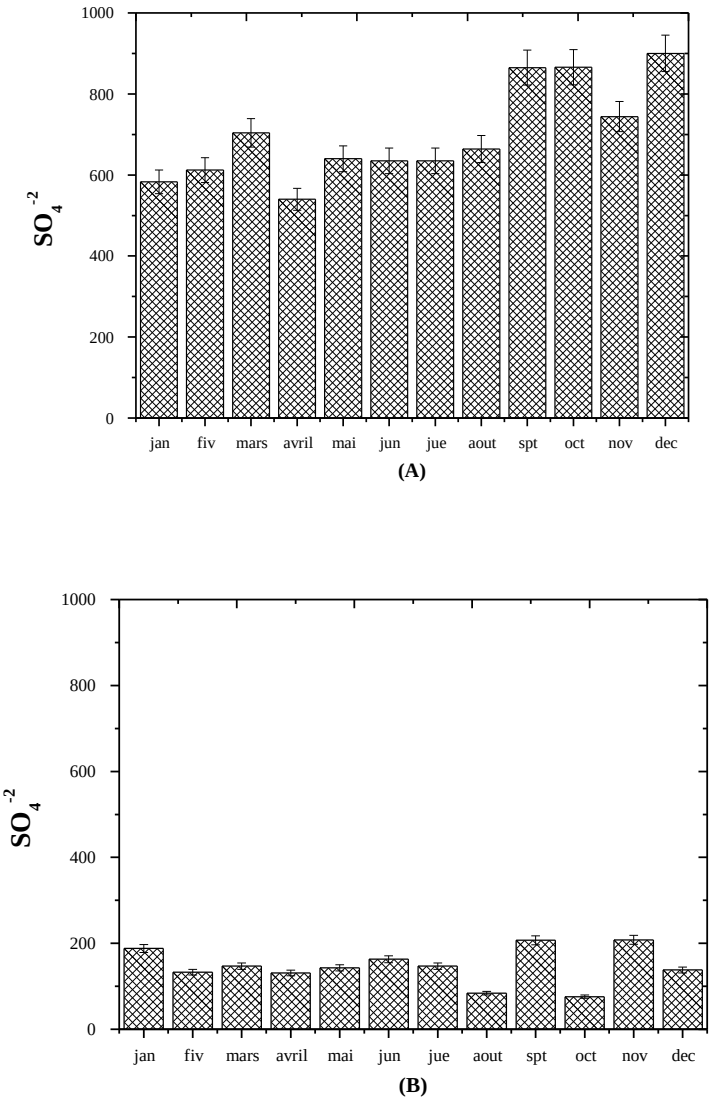


Figure II.14. Variation des moyennes mensuelles du sulfate (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B)

Les eaux brutes contiennent des concentrations très élevée en sulfates dépassent 540 mg/l, ne corresponde pas aux normes MSO et l'algériennes pour les eaux.

La concentration des sulfates dans l'eau traitée est un peu élevée. Mais, Elles sont conforment aux normes recommandées par **JORA, 2011** (<400mg/l).

D'prés **Rodier et al.,(2009)**,l'emploi de sulfate d'aluminium pour le traitement de coagulation des eaux peut contribuer à une élévation de la teneur en sulfates ce qui s'accorde aux résultats. Des teneurs en sulfates en excès peuvent donner à l'eau un gout amer et avoir un effet laxatif sur les personnes n'étant pas habituées à cette eau[81]**.**

II.1.4. Paramètres indésirables

II.1.4.1. Fer

Variation de sulfate est illustrée dans la figure II.15.

Les concentrations de fer dans les eaux brutes et traitées sont négligeables, elles sont à l'état de traces (<0.3mg/l).[121]

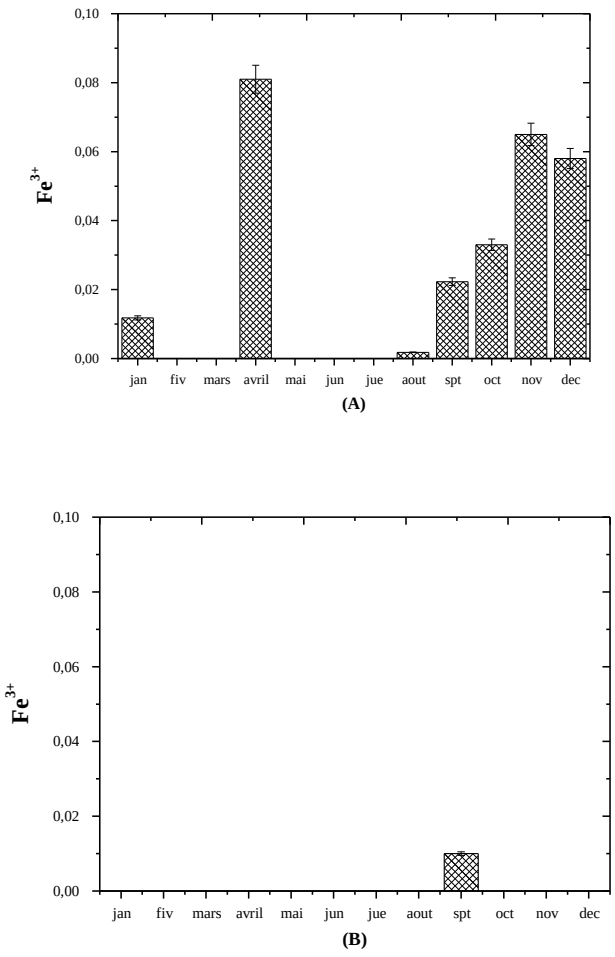


Figure II.15 variation des moyennes mensuelles du fer (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B).

D'après **Rodier et al., (2009)**,une saveur désagréable pouvant être perçue à partir de 0,05 mg/L, le fer développe dans l'eau une turbidité rougeâtre peu engageante pour le consommateur.

L'élimination du fer peut se faire par oxydation-aération suivie d'une filtration ou par oxydation biologique[81]

D'après **Friedman et al., (2004)**, la corrosion peut entrainer le rejet de contaminants, tels que le fer, dans le système de distribution. En effet il a été démontré que les conduites en fonte peuvent être une source de fer dans le système de distribution d'eau.

L'accumulation de fer dans le système de distribution peut être également le résultat de l'oxydation et de la décantation de fer[132]

II.1.4.2. Manganèse

D'après la figure II.16, il en ressort que les concentrations de manganèse dans les eaux brutes et traitées sont négligeables, elles sont à l'état de traces (<0.05mg/l).

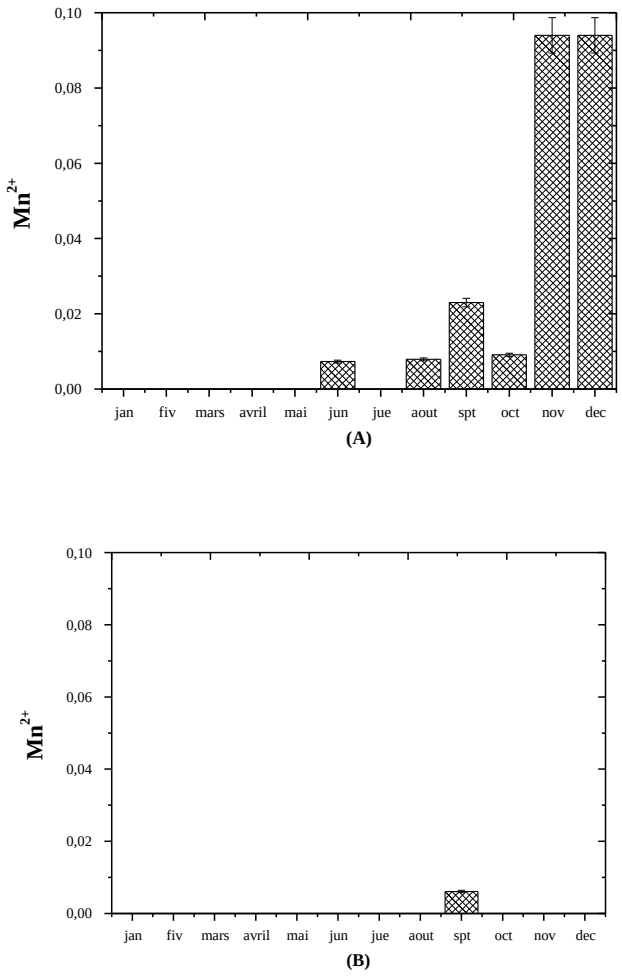


Figure II.16. Variation des moyennes mensuelles du manganèse (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B).

En effet la présence de manganèse dans les eaux superficielles provient essentiellement des industries et des usines d'engrais[81]

Dans les eaux de surface, le manganèse se trouve en général à l'état oxydé et précipité, il est donc éliminé par traitements classique de clarification[133]**.**

Au point de vue gustatif, le manganèse peut donner un goût désagréable à l'eau. Par ailleurs, même à des doses faibles (0,02 mg/L), il est susceptible de former une couche noire sur les canalisations qui lorsqu'elle se détache donne un aspect peu engageant à l'eau[81]**.**

II.1.4.3. Aluminium

Les résultats de l'analyse de l'eau brute montrent que l'aluminium est nul

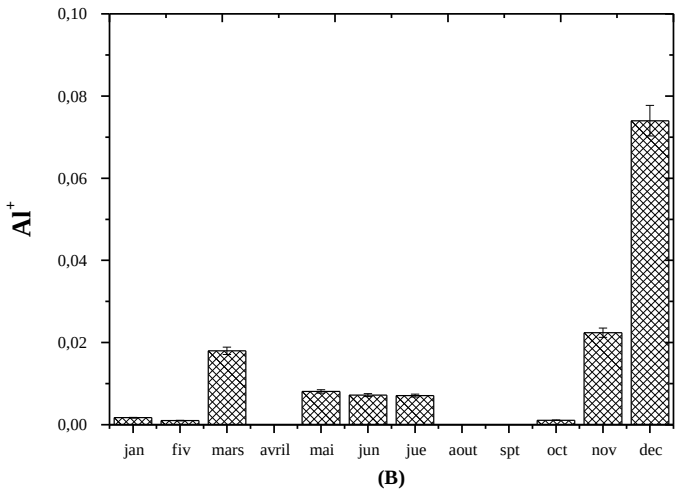


Figure II.17. Variation des moyennes mensuelles du l'aluminium en fonction du temps (mois) des eaux traitées (B)

Après traitement et du fait de l'utilisation de sulfates d'alumine dans les traitements de potabilisation il apparait à des teneurs négligeable qui ont acceptables par les normes dans les eaux traitées

Une station de traitement en bon état de fonctionnement devrait aboutir à une concentration en aluminium inférieure à 50 μg/L dans l'eau distribuée[81]. La présence d'aluminium dans les eaux de consommation est dû à l'utilisation de sulfate d'alumine[134].

Le sulfate d'alumine est utilisé comme coagulant pour le traitement de l'eau. Il y a une probabilité de 50 % que cela puisse conduire à des concentrations accrues d'aluminium dans l'eau traitée par rapport à l'eau brute[135]**.**

II.1.5. Paramètres des alcalins

II.1.5.1. Potassium

À partir de la figure II.18, nous remarquons que l'eau brute ne répond pas aux normes d'eau potable. Par contre, les eaux traitées sont conformes aux normes de potabilité.

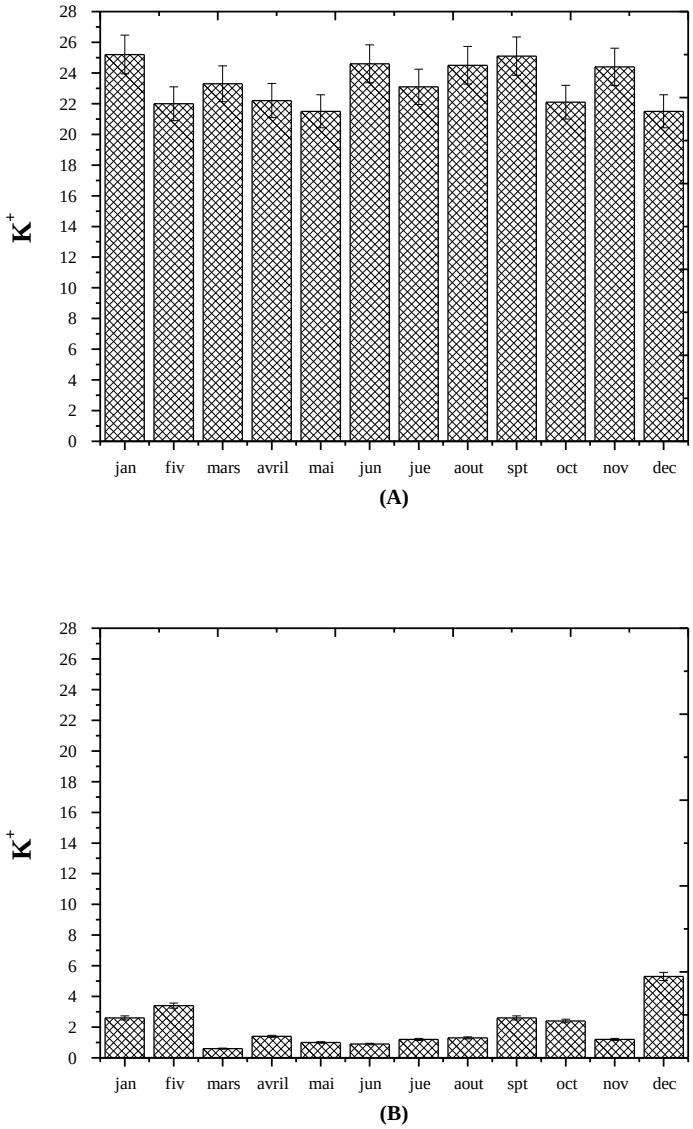


Figure II.18. Variation des moyennes mensuelles du potassium (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B).

II.1.5.2. Sodium

À partir de la figure II.19, nous remarquons que l'eau brute ne répond pas aux normes d'eau potable. Par contre, les eaux traitées sont conformes aux normes de potabilité.

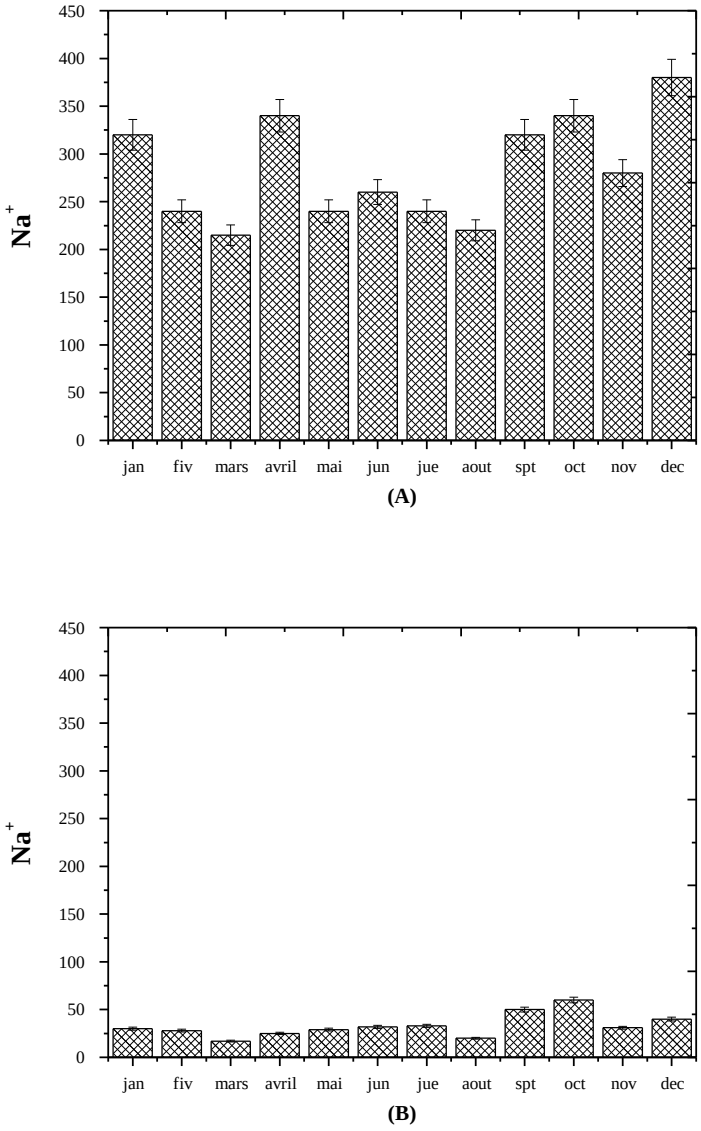


Figure II.19. Variation des moyennes mensuelles du sodium (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B).

II.1.5.2. Titre alcalimétrique complet (TAC)

À partir de la figure II.20, nous remarquons que les valeurs de TAC sont très élevées dans l'eau brute. Ces dernier, ne répond pas aux normes d'eau potable. Par contre, les eaux traitées nous sommes remarques qui les valeurs ont été diminuées et aussi sont conformes aux normes de potabilité.

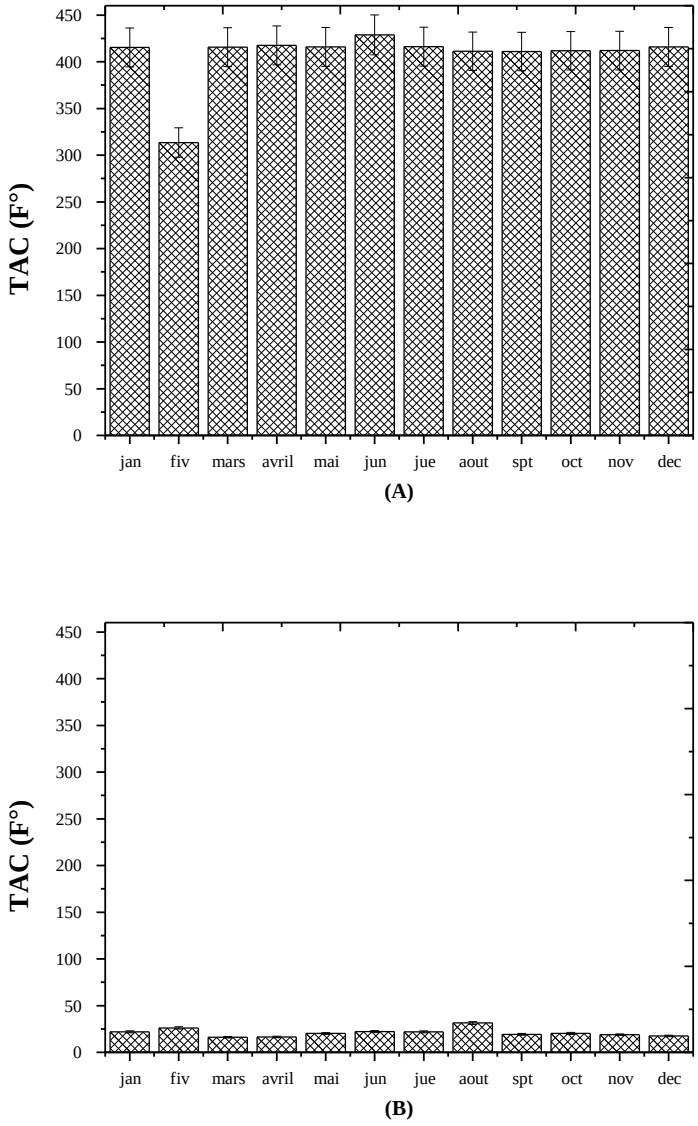


Figure II.20. Variation des moyennes mensuelles du TAC (mg/L de $CaCO_3$) en fonction du temps (mois) des eaux brutes (A) et traitées (B).

À l'inverse de l'acidité, l'alcalinité d'une eau correspond à la présence de bases et de sels d'acides faibles. Dans les eaux naturelles, l'alcalinité résulte le plus généralement à la présence d'hydrogénocarbonates, carbonates et hydroxydes. D'autres sels d'acides faibles peuvent aussi être dosés et interfèrent dans la mesure : acides humiques, phosphates, citrates, tartrates… La silice ionique peut aussi interférer notamment lorsque le pH est supérieur à 8,5.[81]

II.1.6. Résidu sec

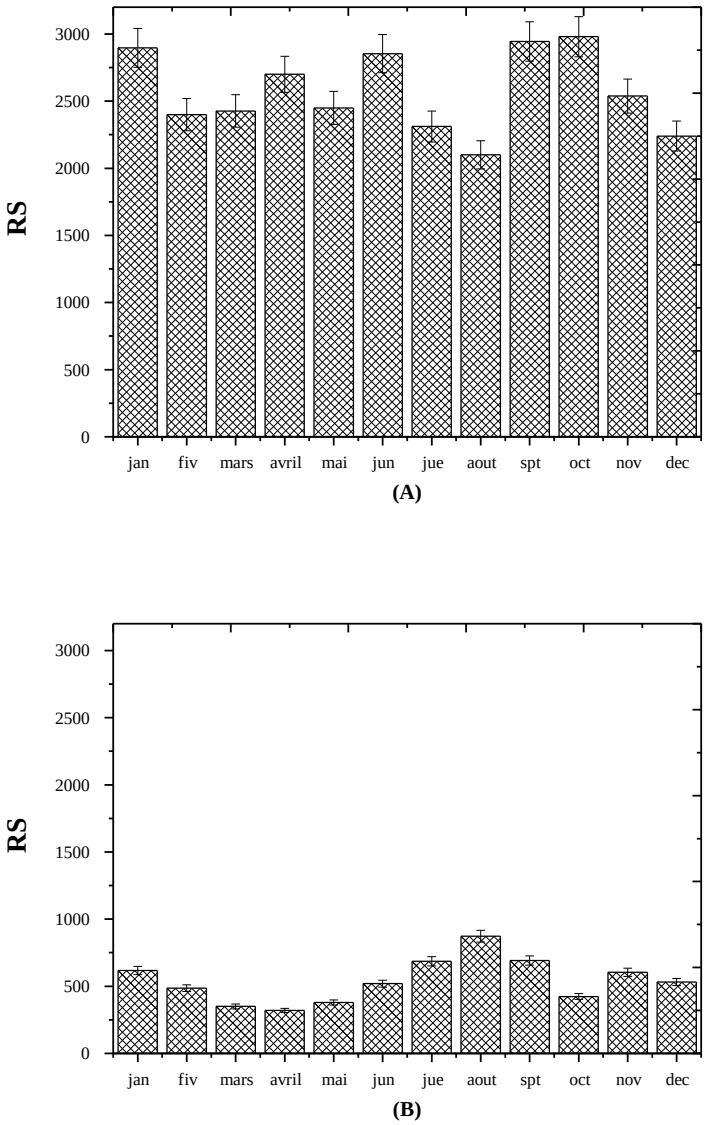


Figure II.21. Variation des moyennes mensuelles du résidu sec (mg/l) en fonction du temps (mois) des eaux brutes (A) et traitées (B).

La détermination du résidu sec dans l'eau non filtrée permet d'évaluer la teneur en matières dissoutes et en suspension, non volatiles, obtenues après une évaporation d'eau[81].

La figure suivante représente la variation spatiale de résidu sec dans les eaux brute et traitée pour site de chafa. Nous remarquons que l'eau brute ne répond pas aux normes d'eau potable qui ont été dépassé la valeur 2000 mg/L. Par contre, les eaux traitées sont conformes aux normes de potabilité. Le résultat obtenu est influencé par la température et la durée de la dessiccation.

En effet à 100-105 °C, une partie ou la totalité tant de l'eau interstitielle que de l'eau de cristallisation de certains sels peut s'évaporer. Vers 105-110 °C, les hydrogénocarbonates se

transforment en carbonates, ce qui entraîne une perte d'anhydride carbonique. Entre 105 et 180 °C, en plus de l'eau de cristallisation, il peut y avoir : perte d'ammoniaque à partir des combinaisons ammoniacales, d'acide sulfhydrique à partir de sulfures alcalins, de matières organiques à partir de composés organiques. En outre le chlorure de magnésium peut s'hydrolyser et précipiter sous forme d'hydroxyde. Le résidu sec déterminé à 180 °C est celui qui se rapprocherait le plus de la somme des différents résultats d'analyses, il est pratiquement exempt d'eau de cristallisation. Le résidu à 525 °C permet d'estimer (par différence avec le résidu à 105 °C) la teneur en matières organiques. Enfin le résidu sulfaté permet, en particulier, pour les eaux fortement minéralisées d'obtenir des résultats plus stables[81].

Partie III :

Modélisation

III.1. Modélisation de sulfate

Le réseau de neurones artificiels et la régression linéaire multiple sont les deux approches utilisées pour la prédiction de la concentration de sulfate soluble à partir des paramètres physico-chimiques de l'eau potable. Les deux méthodes ont été réalisées et évaluées puis comparées.

III.1.1 Multiple Linaire Régression

Une analyse statistique par la méthode de régression linéaire multiple a été réalisée à l'aide du logiciel «XLSTAT 2016» sur la base de données. Cette méthode permet de trouver la relation polynomiale mathématique (Eq.III.1) entre la teneur en sulfate soluble et les variables indépendantes, qui correspondaient aux paramètres physico-chimiques de 84 échantillons expérimentaux. Premièrement, l'équation prend en compte les 18 variables (tableau III.1), même celles qui ne semblent pas avoir d'impact significatif sur la variable dépendante. La relation obtenue a donc été évaluée afin de ne conserver que les variables indépendantes caractérisées par une valeur de probabilité élevée (tableau III.1).

$$y = \beta_0 + \sum_{i=1}^{18} \beta_i x_i \qquad \textbf{(III.1)}$$

Où β_i et x_i sont donnés dans le tableau III.1.

Tableau III.1. Valeurs de, xi et caractérisation des variables indépendantes du RLM.

i	x_i	β_i	*Erreur Standarde*	*T rapport*	*Prob.> |t|*
0	Constant	245.4060	435.0382	0.56	0.5746
1	Conductivity (Cond)	−0.0576	0.0676	–0.85	0.3976
2	Turbidity (NTU)	4.3994	4.8354	0.91	0.3663
3	Potential hydrogen (PH)	−28.6495	55.9304	–0.51	0.6102
4	Hardness (TH)	0.0700	2.4603	0.03	0.9774
5	Calcium (Ca)	1.7242	0.8677	1.99	0.0511
6	Magnesium (Mg)	4.1885	1.2180	3.44	0.0010*
7	Chlorides (Cl)	−1.1275	0.1141	–9.88	< 0.0001*
8	Total Alkalimetric Titer	−0.7729	0.2281	–3.39	0.0012*
9	Organic Materials (MO)	0.4933	0.4757	1.04	0.3035
10	Nitrogen dioxide (NO_2)	−3722.542	1532.262	–2.43	0.0179*
11	Nitrates (NO_3)	0.2332	2.4743	0.09	0.9252
12	Sodium (Na)	1.8031	0.2533	7.12	< 0.0001*
13	Potassium (K)	−3.3927	4.1309	–0.82	0.4145
14	Manganese (Mn)	–5154.41	19918.46	–0.26	0.7966
15	Iron (Fe)	135558.938	18425.29	0.74	0.4644
16	Aluminum (Al)	213.5702	447.0937	0.48	0.6345
17	Dry residues (Rs)	0.0828	0.0713	1.16	0.2500
18	Bicarbonate (HCO)	−0.6784	0.1533	–4.42	< 0.0001*

Ensuite, seules 6 variables indépendantes qui ont un pouvoir d'explication élevé pour la variable dépendante ont été retenues (Pr <0,05); ainsi la relation peut être réduite à l'équation (III.2)

$$y = \beta_0 + \sum_{i=6}^{18} \beta_i \times x_i \quad \text{for } i = 6,7,8,10,12 \text{ and } 18 \text{ only} \qquad \textbf{(III.2)}$$

En effet, les coefficients de chaque facteur dans le modèle permettent d'évaluer l'impact de chaque facteur sur la réponse [115]. D'après l'équation (III.2), il est évident que les changements de magnésium et de sodium dans l'eau potable augmentent la dose de sulfate. D'autre part, les changements dans les chlorures, le titre alcalimétrique total, le dioxyde d'azote et le bicarbonate dominent la dose de sulfate.

La valeur du coefficient de détermination a légèrement diminué, mais l'équation est devenue plus simple après l'élimination des variables peu explicatives de la variable dépendante. Le résultat de la phase d'apprentissage montre une efficacité acceptable, via leur coefficient de corrélation (figure III.1) (R = 0,93795), mais médiocre en termes d'erreur quadratique moyenne (RMSE = 84,0254 mg / l) et d'erreur absolue moyenne (MAE = 48,6575 mg / l) (tableau III.2). Le résultat de la phase d'apprentissage a été testé et validé par deux bases de données, l'une pour le test et l'autre pour la validation. Les résultats ont montré un coefficient de corrélation acceptable pour la phase de test (figure III.1) (R = 0,92066) et un coefficient de corrélation élevé pour la phase de validation (figure III.1) (R = 0,97100). Cependant, en termes d'erreur quadratique moyenne (RMSE = 117,933 et 70,5674 mg / l) et d'erreur absolue moyenne (MAE = 61,8972 et 46,6106 mg / l), ils sont toujours restés mauvais pour le test et la phase de validation respectivement (tableau III.2). Les trois étapes ont été combinées pour l'évaluation finale de ce modèle qui a donné un coefficient de corrélation acceptable (R = 0,941) (figure III.1), mais médiocre en termes d'erreur quadratique moyenne (RMSE = 88,3068 mg / l) et d'erreur absolue moyenne (MAE = 42,3274 mg / l) (tableau III.2). Au vu des résultats obtenus, on peut considérer que la corrélation du modèle était un peu positive via le coefficient de corrélation acceptable; tandis que les critères d'évaluation statistique sont restés élevés. La probabilité (Pr <0,0001) était strictement inférieure à 0,5% ce qui confirme que le modèle était significatif [114,136].

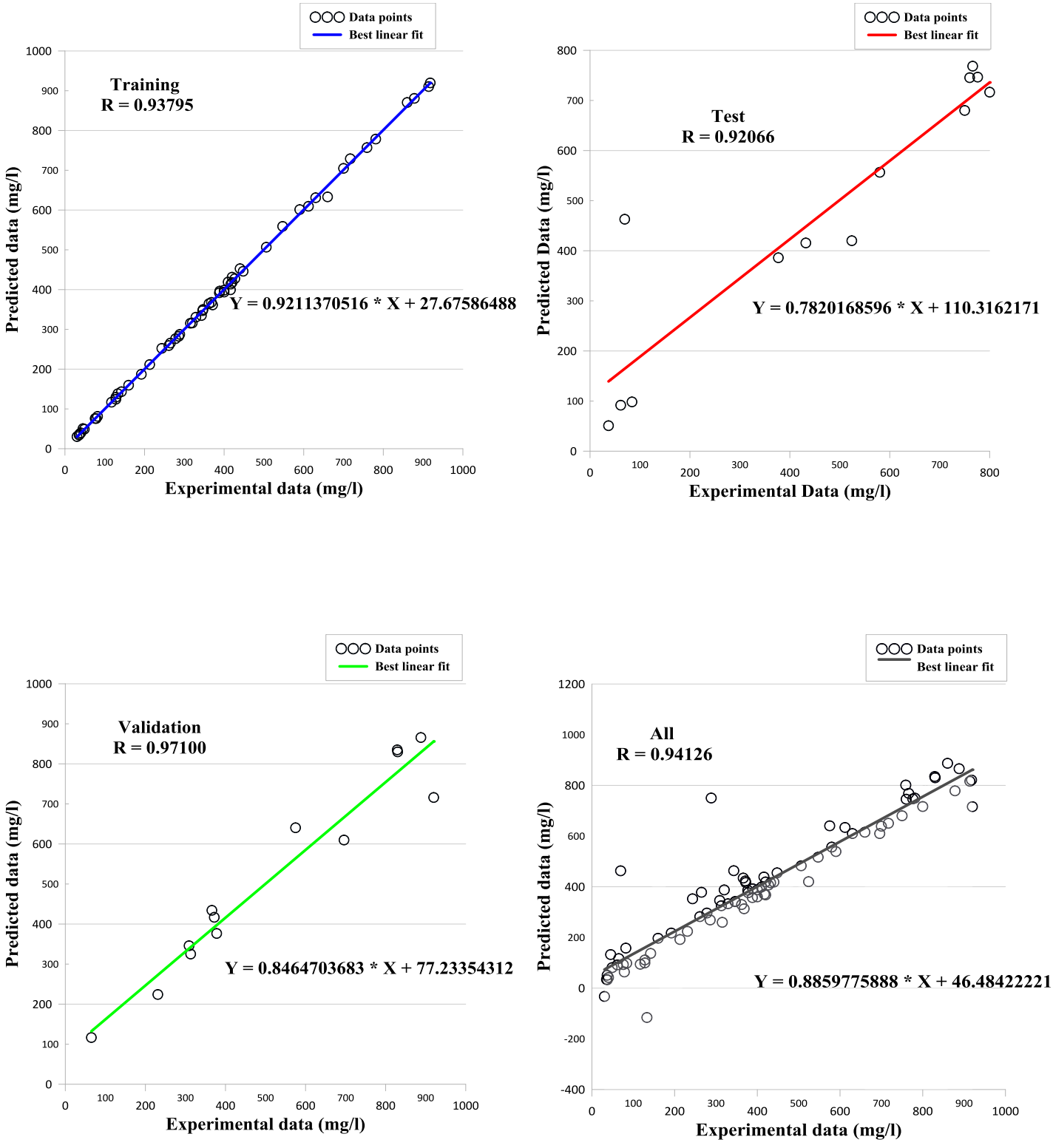


Figure III.1. Comparaison entre les valeurs expérimentales et calculées

Les résultats des performances MLR en termes de toutes les erreurs et en termes de valeurs du vecteur d'accord (R, pente: α et ordonnée à l'origine: β) sont donnés dans le tableau III.2.

Tableau III.2. Performances de la prédiction utilisant MLR

	RMSE (mg/l)	MAE *(mg/)l*	*R*	α	β
Tous	88.3068	42.3274	0.941	0.8859	46.480
Apprentissage	84.0254	48.6575	0.93795	0.9211	27.6758
Teste	117.9331	61.8972	0.92066	0.7820	110.3162
Validation	70.5674	46.6106	0.97100	0.8464	77.2335

III.1.2. Modélisation par RNA

Des tests préliminaires ont montré que pour améliorer les performances d'un modèle établi par RNA, il est nécessaire de modifier l'architecture du réseau, en changeant principalement le nombre de couches cachées, le nombre de neurones cachés et / ou le nombre de cycles d'apprentissage (nombre des itérations). Pour cela, nous avons successivement changé le nombre de neurones cachés (de 3 à 15). Les résultats de ces tests sont présentés dans le tableau III.3.

Tableau III.3. Performances des différentes architectures RNA testées

RNA		Function d'activation		Coefficients de corrélation				*MSE*		
Nb	Neurons/	**caché**	**sortie**	APP	Test	Val	Total	APP	Test	Val
1	[18-9-1]	tansig	purelin	0.99809	0.99964	0.99965	0.99863	2.34×10^{-4}	6.12×10^{-5}	7.15×10^{-5}
2	[18-13-1]	logsig	purelin	0.99961	0.99618	0.99723	0.99873	5.47×10^{-5}	5.27×10^{-4}	3.32×10^{-4}
3	**[18-8-1]**	**tansig**	**tansig**	**0.99953**	**0.99978**	**0.99992**	**0.9996**	$\mathbf{7.36\times10^{-5}}$	$\mathbf{2.55\times10^{-5}}$	$\mathbf{9.79\times10^{-6}}$

Le tableau III.3 présente les meilleures architectures trouvées. Il montre les coefficients de corrélation et l'erreur pour chaque apprentissage, test et validation en fonction du nombre de neurones dans la couche cachée et de la topologie du réseau. Il indique également les fonctions d'activation pour la couche cachée et la couche de sortie. L'architecture 3 (tableau III.3) semble être le modèle RNA le plus pertinent pour prédire la teneur en sulfate soluble.

Le réseau a été conduit jusqu'à atteindre le surapprentissage; ce phénomène a été rencontré après 1000 itérations. La phase de surapprentissage n'étant pas encore atteinte, il était donc intéressant de continuer à apprendre jusqu'à atteindre cette phase de test afin de diminuer encore le gradient et ainsi améliorer la précision de RNA. En utilisant la troisième architecture du tableau III.3, les trois courbes liées à l'évolution de l'erreur quadratique moyenne (MSE)

correspondant aux trois phases d'apprentissage ont été obtenues (apprentissage, validation et test) où elles convergent correctement vers un minimum satisfaisant (figure III.2)

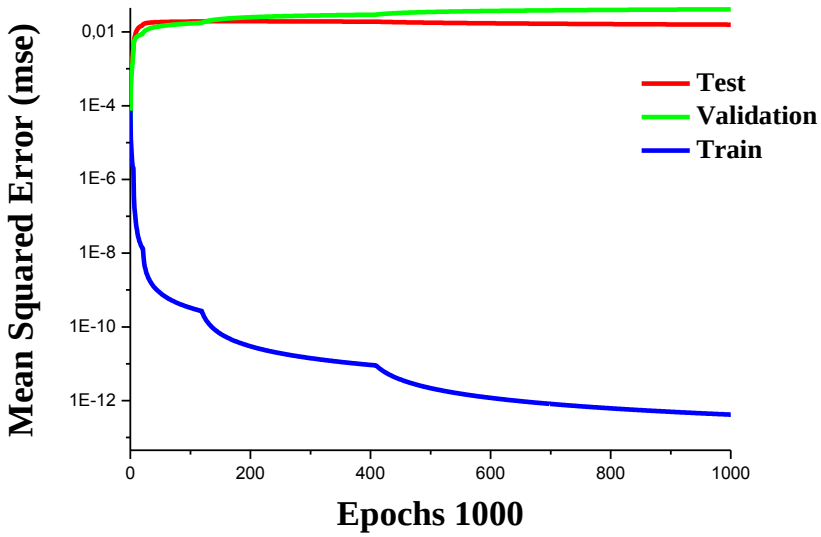
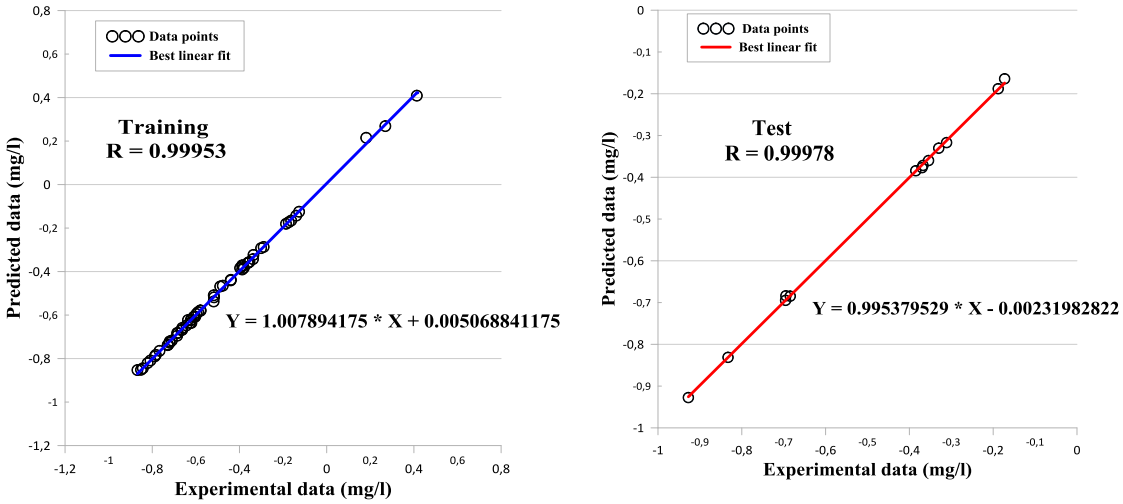


Figure III.2. Evolution du MSE correspondant aux trois phases d'apprentissage

Une fois l'architecture de RNA synthétisée, elle est testée et les valeurs prédites (valeurs estimées) sont comparées aux valeurs expérimentales dans les trois étapes des phases d'apprentissage; les résultats correspondants sont donnés à la figure III.1.



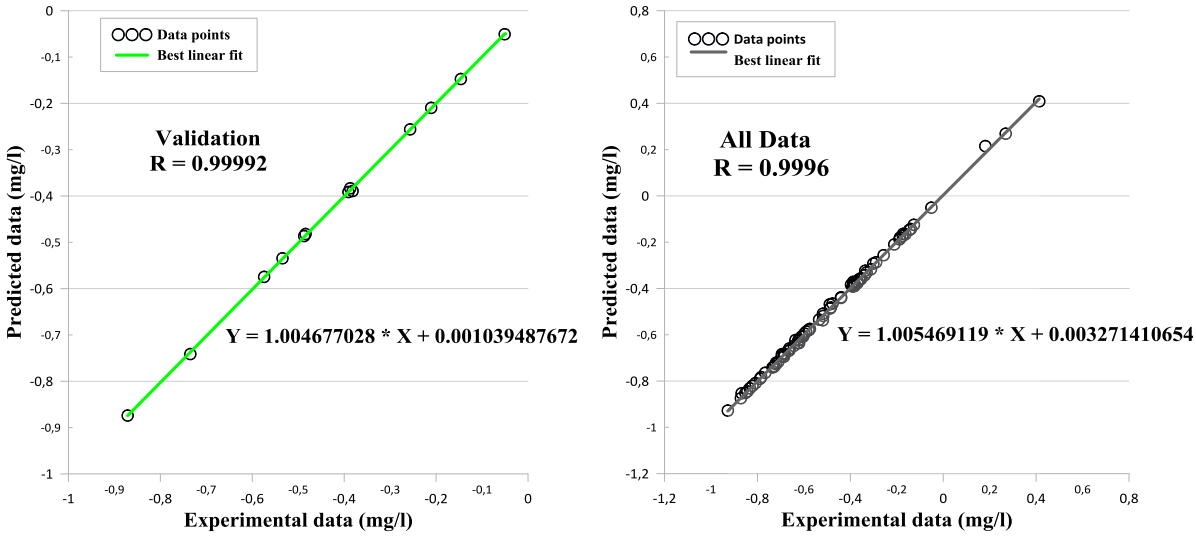


Figure III.3. Comparaison entre les valeurs expérimentales normalisées et calculées

La relation entre les valeurs observées et les valeurs estimées montre les performances et l'efficacité du modèle RNA développé. La pertinence du modèle a été confirmée par le coefficient de détermination obtenu R = 0,9996 pour l'ensemble des données.

III.1.3. Performances de prédiction

Le résultat de la phase d'apprentissage montre une excellente efficacité, via leur coefficient de corrélation très élevé (figure III.4) (R = 0,99963), également des valeurs très faibles en termes d'erreur quadratique moyenne (RMSE = 6,5382 mg / l et l'erreur absolue moyenne (MAE = 4,2437 mg / l) (tableau III.4). Le résultat de la phase d'apprentissage a été testé et validé par deux bases de données, l'une pour le test et l'autre pour la validation; le résultat a montré un coefficient de corrélation très élevé (R = 0,9998, 0.9999) pour les deux phases (test et validation) respectivement (figure III.4), également des valeurs très faibles en termes d'erreur quadratique moyenne (RMSE = 5,4650 et 3,1826 mg / l) et d'erreur absolue moyenne (MAE = 3,5362 et 1,9667 mg / l) pour les phases de test et de validation respectivement (tableau III. 4). Les trois étapes ont été réunies pour l'évaluation finale de ce modèle qui a donné un coefficient de corrélation très élevé (R = 0,99973) (figure III.4), et a également conduit à des valeurs en termes d'erreur quadratique moyenne (RMSE = 5,9755 mg / l) et d'erreur absolue moyenne (MA E = 3,1767 mg / l) (tableau III.4). Au vu des résultats obtenus, nous pouvons considérer que la corrélation du modèle était très positive, montrant l'efficacité du modèle.

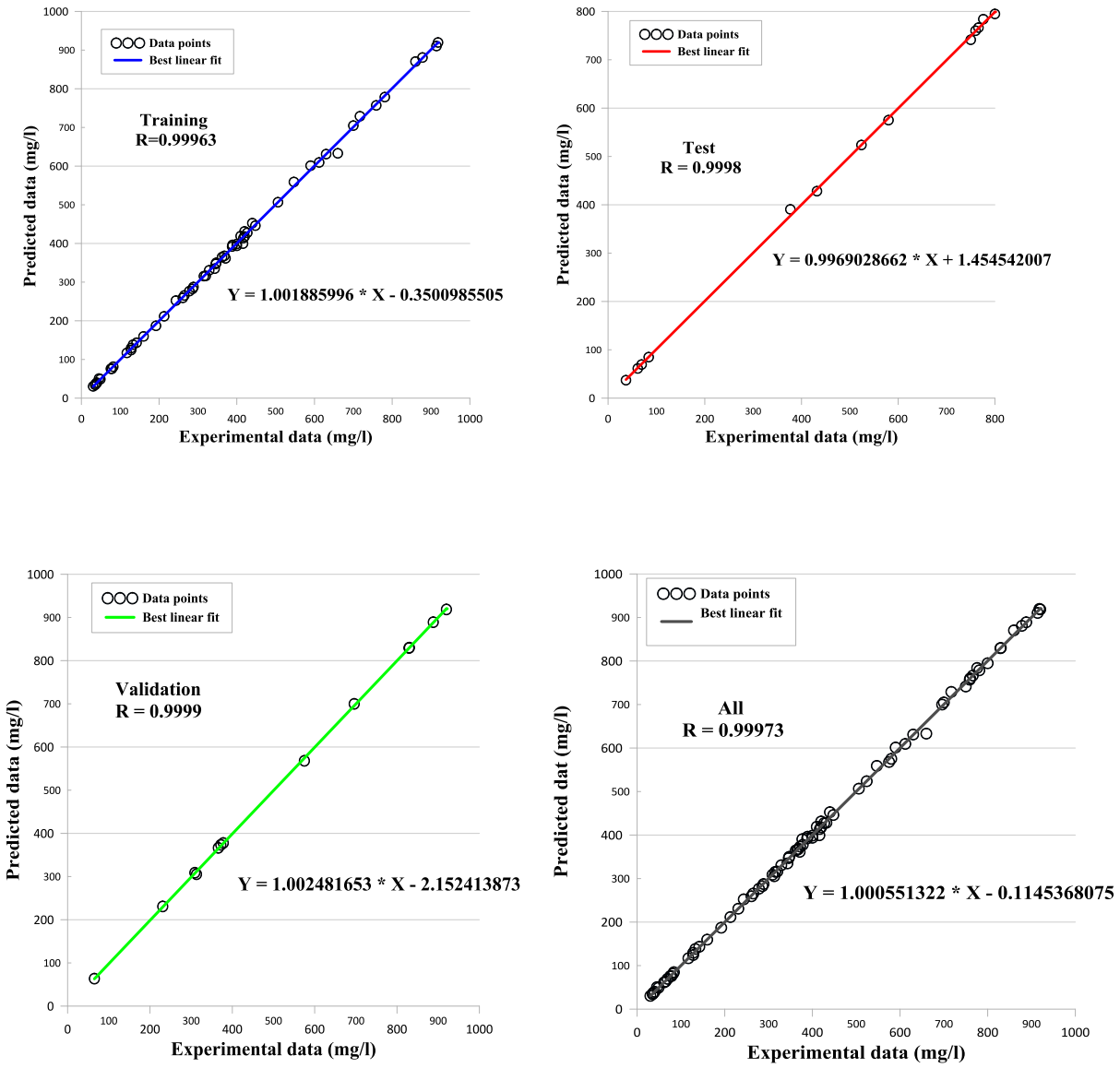


Figure III.4. Comparaison entre les valeurs expérimentales normalisées et calculées

Tableau III.4. Performances de la prédiction utilisant RNA

	RMSE *(mg/l)*	MAE *(mg/)l*	*R*	α	β
Tous	5.9755	3.1767	0.99973	1.0005	-0.1145
Apprentissage	6.5382	4.2437	0.99963	1.0018	-0.3500
Teste	5.4650	3.5362	0.9998	0.9969	1.4545
Validation	3.1826	1.9667	0.9999	1.0024	-2.1524

III.1.4. Performances d'interpolation

Afin de tester la précision du modèle RNA précédemment développé et optimisé, une interpolation a été réalisée. A cet effet, une base de données a été construite en 2019 contenant un ensemble de points de données localisés aux points expérimentaux. Les résultats montrant la courbe de régression entre les valeurs prédites et expérimentales et les performances de l'interpolation en termes d'erreur et de coefficient de corrélation sont présentés à la figure III.5.

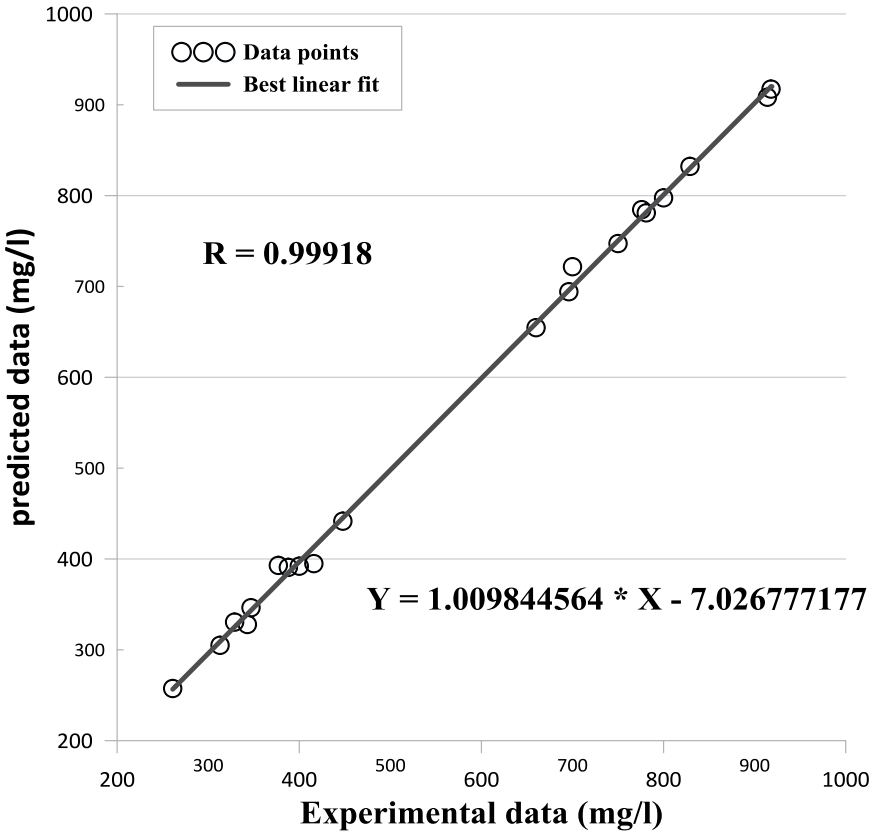


Figure III.5. Comparaison entre les valeurs expérimentales et calculées

Ces résultats montrent une bonne corrélation entre RNA prédit et les valeurs expérimentales avec un coefficient de corrélation très élevé (figure III.5) (R = 0,99918) et avec une erreur quadratique minimale (RMSE = 9,3595 mg / l) et une erreur absolue (MAE = 6,7598 mg / l) (tableau III.5). Ce résultat confirme à nouveau l'efficacité du modèle en utilisant de nouvelles données dans la plage de prédiction.

Les résultats des performances d'interpolation en termes de toutes les erreurs et en termes de valeurs du vecteur d'accord [R (coefficient de corrélation), α = (pente) et β (ordonnée à l'origine)] sont résumés dans le tableau III.5.

Tableau III.5. Performances de l'interpolation

RMSE	*MAE*	*R*	*α*	*β*
9.3595	6.7598	0.99918	1.001	-7.027

III.1.5. Performances d'extrapolation

Une prédiction a également été réalisée pour tester la précision du réseau de neurones artificiels développé. L'idée était de tester des données expérimentales sur la teneur en sulfate soluble, qui n'ont pas été utilisées lors de la formation de notre réseau; cette base de données a été construite en 2019. Les résultats de la figure III.6 montrent un coefficient très élevé (R = 0,9989), ainsi que des valeurs très faibles en termes d'erreur quadratique moyenne (RMSE = 12,2587 mg / l) et d'erreur absolue moyenne (MAE = 10,9054 mg / l) (tableau III.6). Une extrapolation d'une base de données a également été réalisée pour vérifier l'exactitude de notre modèle RNA optimisé. L'idée était de tester un ensemble de données expérimentales jamais exploité pendant les phases d'apprentissage et de test. L'ensemble de données expérimentales de ce système exploité dans notre base de données principale a été obtenu par analyse d'échantillons d'eau prélevés lors de plusieurs campagnes menées dans la région de Médéa. Les résultats des performances d'extrapolation en termes de toutes les erreurs et en termes de valeurs du vecteur d'accord [R (coefficient de corrélation), α = (pente) et β (ordonnée à l'origine)] sont résumés dans le tableau III.6. La qualité d'ajustement des données d'extrapolation l'ensemble est représenté sur la figure III.6. Les résultats ont montré un bon vecteur d'accord [R, α, β] entre les données expérimentales et les résultats prédits par RNA avec l'erreur quadratique moyenne acceptée. On peut observer que pour notre cas d'extrapolation, les résultats ont montré une bonne capacité prédictive de RNA à la fois dans le modèle développé et optimisé. Cela montre une bonne convergence entre la sortie expérimentale et la sortie prévue par RNA.

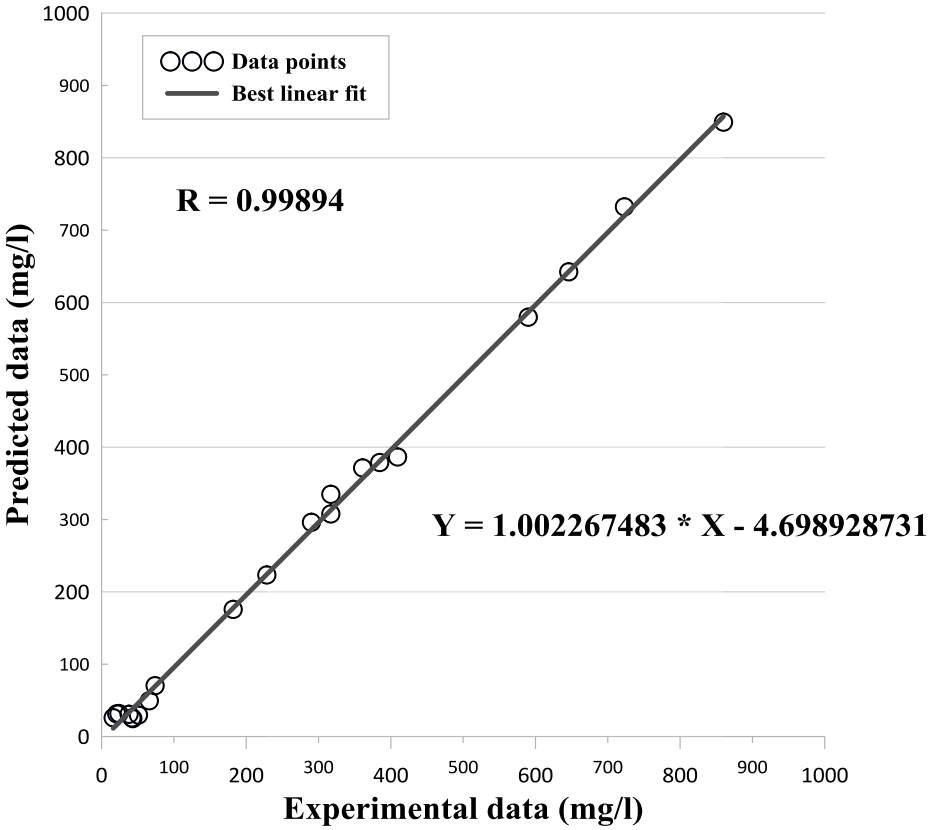


Figure III.6. Comparaison entre les valeurs expérimentales et calculées

Tableau III.6. Performances de l'extrapolation

RMSE (mg/l)	*MAE* (mg/l)	*R*	*A*	*β*
12.2587	10.9054	0.9989	1.0022	-4.6989

III.1.6. Comparaison de MLR et RNA pour prédire les concentrations de sulfate soluble

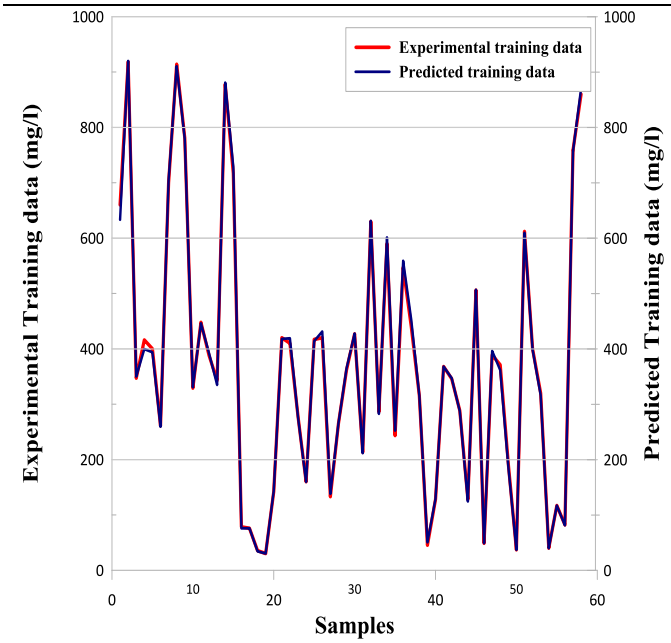
Afin d'évaluer les deux prédicteurs développés, toutes les données de MLR et RNA ont été comparées (tableau III.7). Les résultats présentés dans le tableau III.7 montrent la comparaison des coefficients de corrélation et des indicateurs statistiques obtenus par les deux modèles. Le coefficient de corrélation calculé par RNA était significativement plus élevé avec (R = 0,99973) tandis que celui donné par le MLR était moins faible (R = 0,941). De plus, les RMSE et MAE donnés par le modèle RNA (RMSE = 88,3068 mg / l et MAE = 42,3274 mg / l) étaient très faibles par rapport à celui du modèle MLR (RMSE = 5,9755 mg / l), (MAE = 3,1767 mg / l).

Tableau III.7. Comparaison des performances entre les modèles de prédiction MLR et RNA

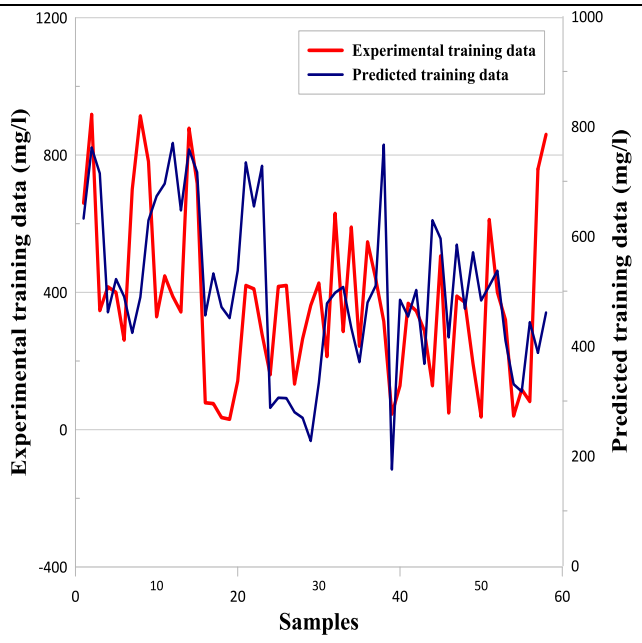
Modele	*RMSE (mg/l)*	MAE *(mg/l)*	*R*	A	B	R^2_{adj}
RNA	5.9755	3.1767	0.99973	1.0005	-0.1145	0.9996
MLR	88.3068	42.3274	0.941	0.8859	46.480	0.924

La figure III.7 confirme à nouveau la supériorité du modèle RNA par rapport au modèle MLR avec une très bonne superposition des courbes mesurées expérimentalement et celles estimées par le modèle RNA.

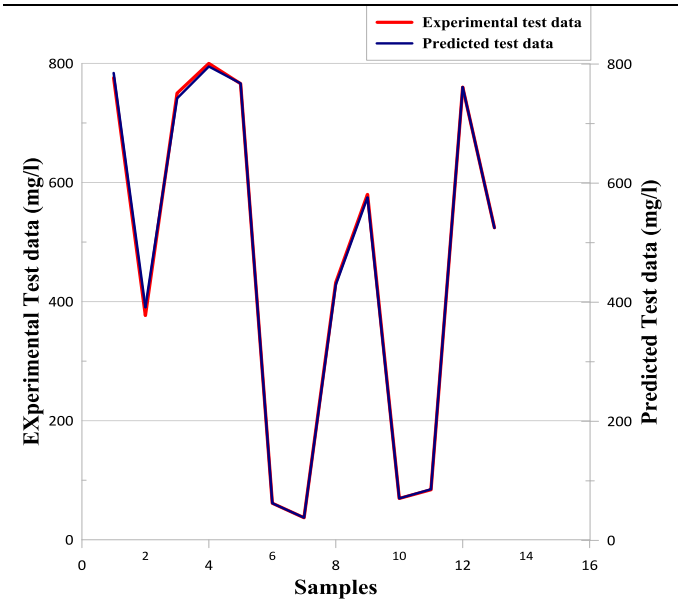
Les résultats obtenus montrent une très bonne concordance pour le modèle RNA, expliquée par un coefficient de corrélation élevé et de petites erreurs d'indicateurs statistiques pour la phase d'apprentissage, la phase de test et la phase de validation.



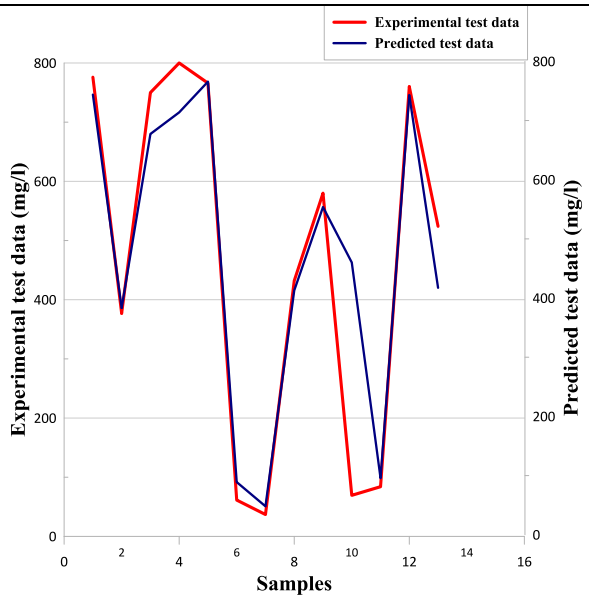
(1)



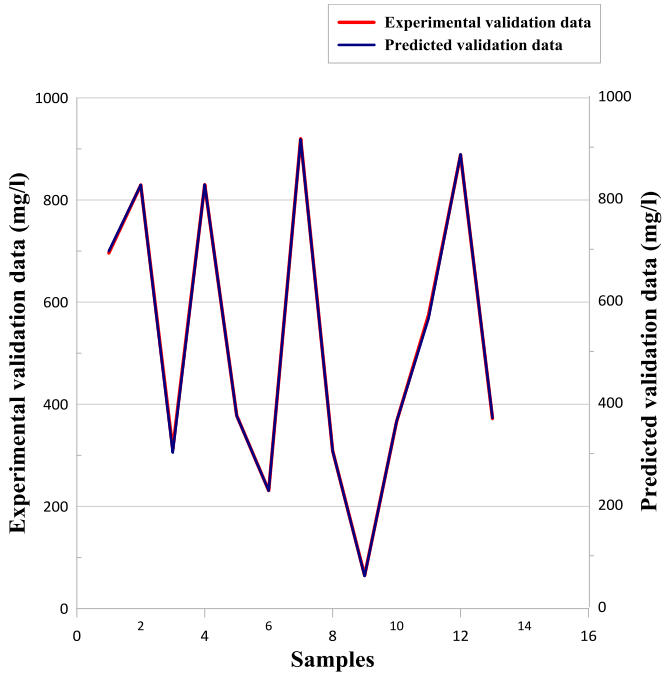
(1)


Experimental test data
Predicted test data
EXperimental Test data (mg/l)
Predicted Test data (mg/l)
Samples

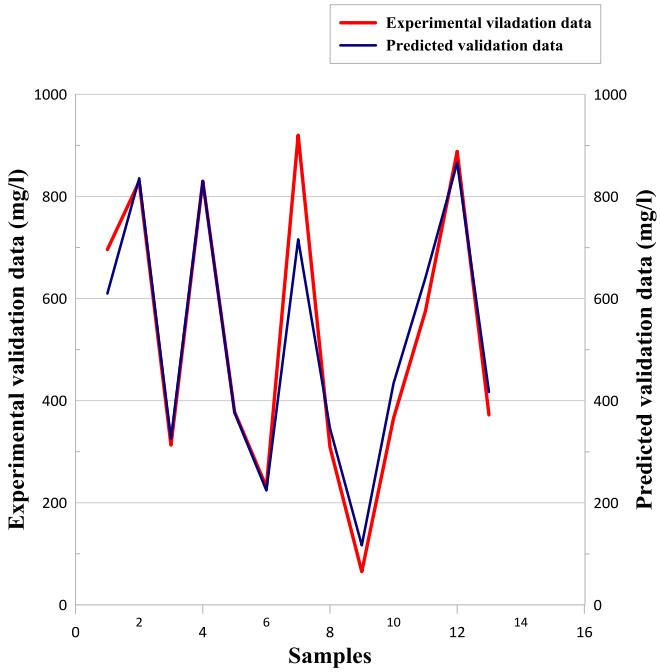
(2)


Experimental test data
Predicted test data
Experimental test data (mg/l)
Predicted test data (mg/l)
Samples

(2)


Experimental validation data
Predicted validation data
Experimental validation data (mg/l)
Predicted validation data (mg/l)
Samples

(3)


Experimental viladation data
Predicted validation data
Experimental validation data (mg/l)
Predicted validation data (mg/l)
Samples

(3)

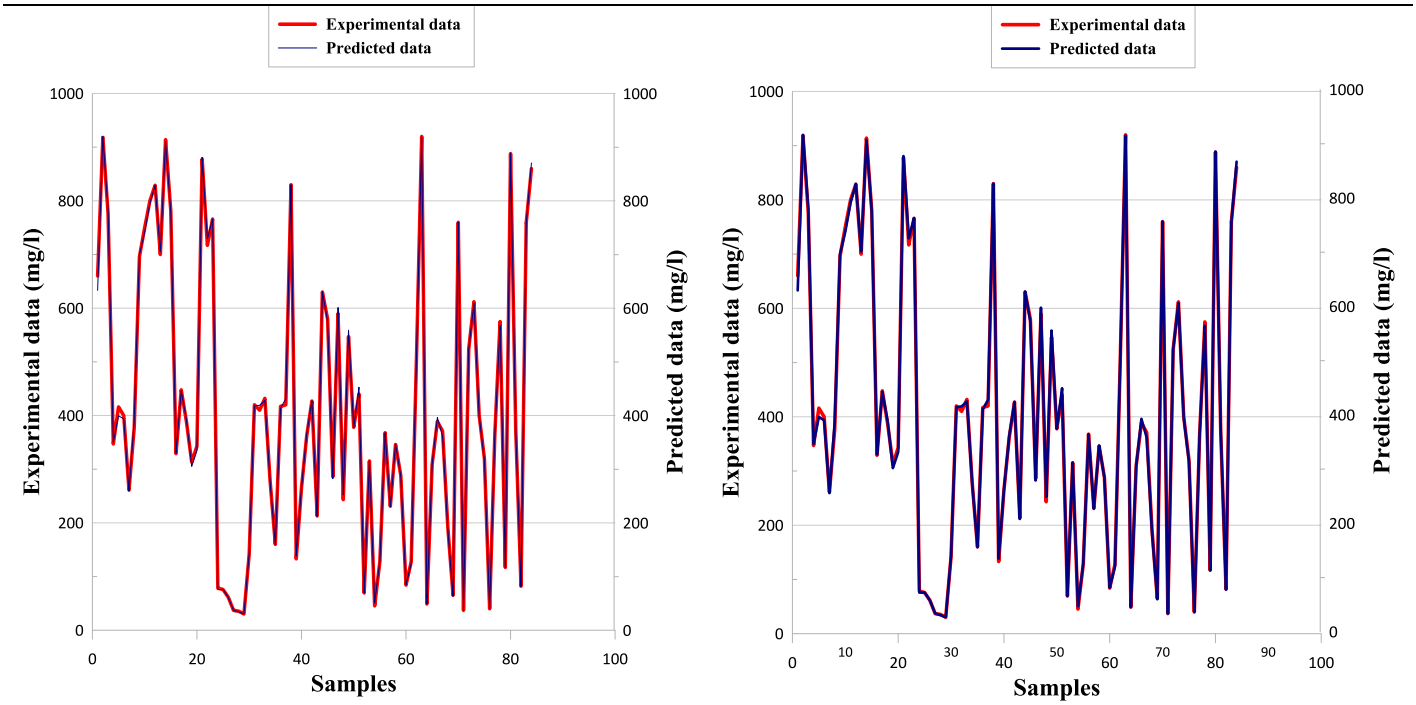


Figure III.7. Distribution des résidus pour les modèles RNA et MLR.

III.1.7 Etude des résidus

L'erreur commise par les modèles établis pour chaque méthode sur un individu de l'échantillon de construction du modèle est appelée résiduelle [137]. Ainsi, l'étude de la relation entre les niveaux estimés de la teneur en sulfate soluble par les modèles mathématiques et leurs résidus permet d'assurer la performance du modèle et de vérifier empiriquement entre eux, la validité des hypothèses des modèles.

Les méthodes d'analyse pour analyser les résidus sont principalement des méthodes d'analyse graphique. La figure III.8 montre les résidus liés au modèle établi par les réseaux de neurones artificiels et ceux liés au modèle établi par régression linéaire multiple basée sur les valeurs estimées.

Cette figure montre que les résidus obtenus par la méthode des réseaux de neurones étaient moins dispersés (proches de zéro) par rapport à ceux obtenus à partir de la régression linéaire multiple.

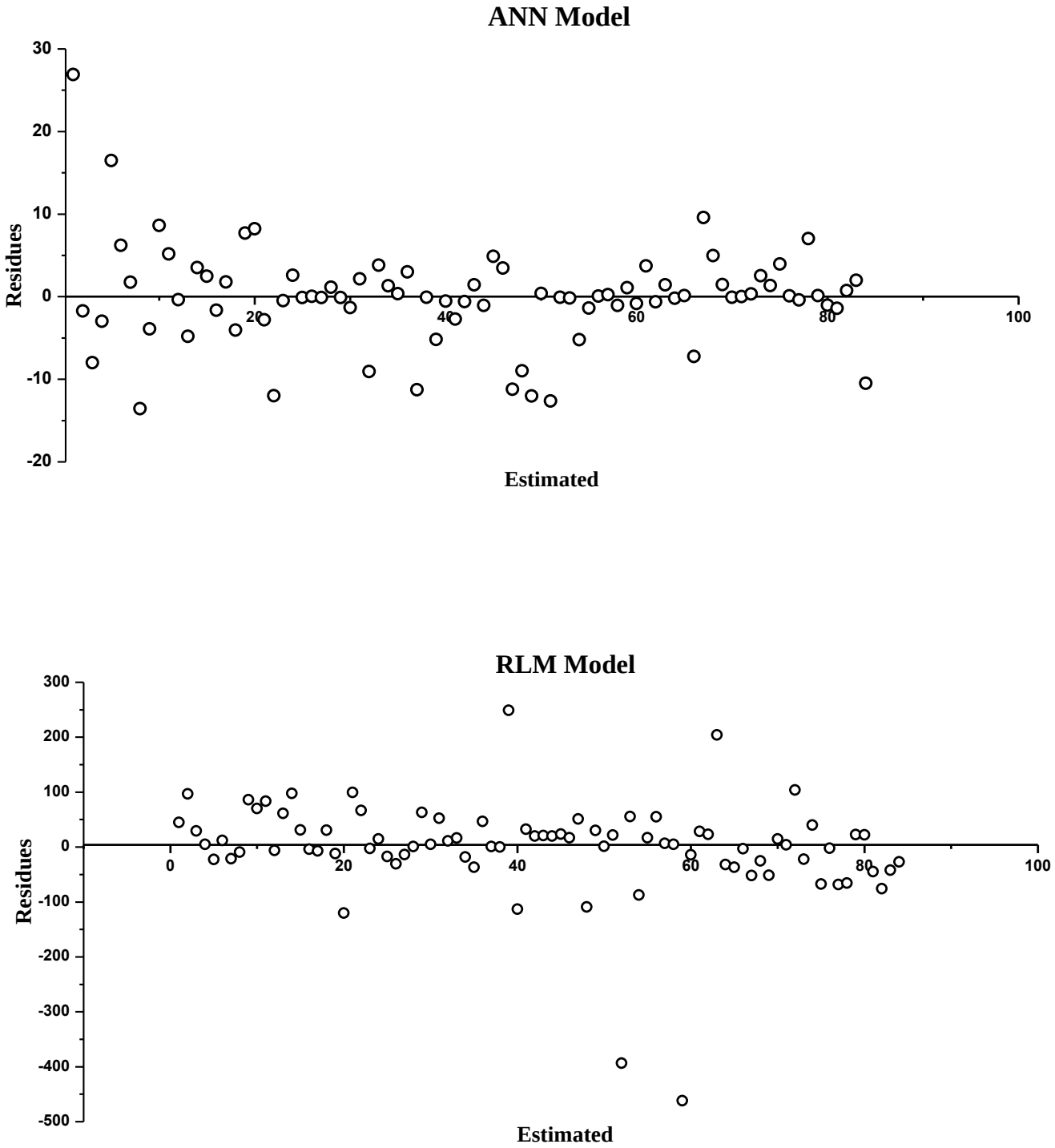


Figure III.8. Résidus relatifs aux modèles établis par régression linéaire multiple et réseau de neurones artificiels en fonction des valeurs estimées.

D'une manière générale, les résultats obtenus sont très satisfaisants et justifient l'utilisation de l'approche par réseau de neurones dans la prédiction des teneurs en sulfate soluble de la région de Médéa. Cela concorde avec les résultats de certaines études récentes qui ont montré que les

modèles de régression linéaire multiples fonctionnent mal par rapport à ceux établis par les modèles de réseaux de neurones artificiels [99,112,138].

III.2. Modélisation de bicarbonate

Des tests préliminaires ont montré que pour améliorer les performances d'un modèle établi par RNA, il est nécessaire de modifier l'architecture du réseau, en jouant principalement sur le nombre de couches cachées, ou sur le nombre de couches cachées, de neurones cachés et / ou le nombre de cycles d'apprentissage (nombre d'itérations). Pour cela, nous avons successivement modifié le nombre de neurones cachés (NH = 3–15). Les résultats de ces tests sont présentés dans le tableau III.8. On remarque que le minimum de la MSE, le nombre d'itérations et le coefficient de corrélation maximum sont atteints lorsque NH = 5. L'architecture optimale d'acquittements obtenue à partir d'RNA a une topologie de 18 –5–1 (tableau III.8) et où l'architecture est illustrée dans les Figure III.9 et III.10. On peut donc choisir cinq neurones de la couche cachée du réseau dans la présente étude, afin de prédire les concentrations de bicarbonate.

Tableau III.8. Structure des réseaux de neurones

Type du réseau	Algorithme d'apprentissage	Couche d'entrée	Couche cachée		Couche de sortie	
		Nombre de neurones	Nombre de neurones	Fonction d'activation	Nombre de neurones	Fonction d'activation
RNA	LM	18	5	Sigmoid	1	Linear

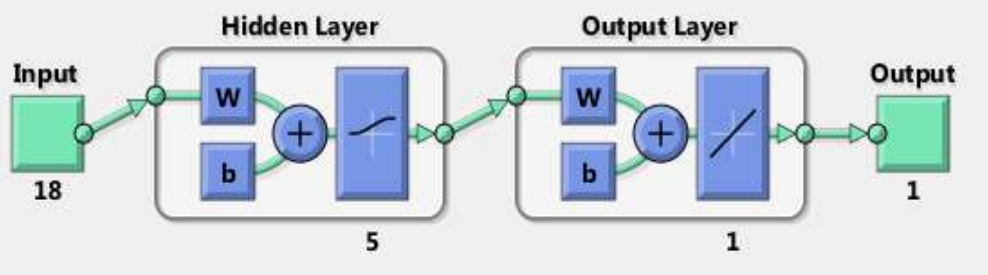


Figure III.9. Représentation schématique du modèle RNA

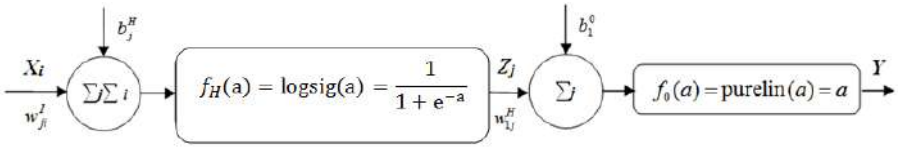


Figure III.10. Représentation schématique du modèle neuronal RNA

La Figure III.11 décrit les résultats de la phase de formation RNA. Il a montré qu'après 236 itérations, le résultat souhaité est atteint avec 5 neurones dans la couche cachée, les trois courbes liées à l'évolution de l'erreur des trois phases convergentes correctement vers une MSE acceptable.

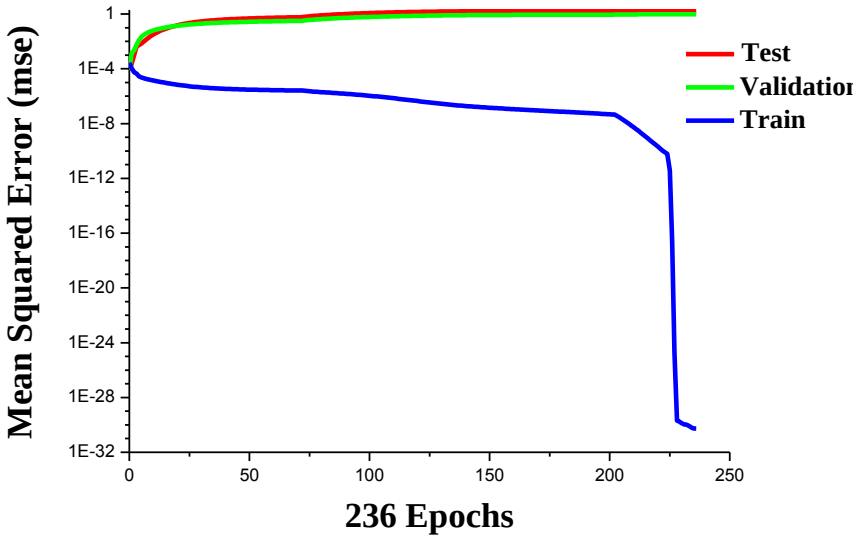


Figure III.11. Évolution du MSE

Les résultats obtenus (Figure III.12) permettent de tracer les différentes valeurs relatives aux paramètres d'apprentissage:

• Nombre maximum d'itérations (époques) = 236.

• Taux d'apprentissage (η) = 0,001.

• Pente minimale = 8,2405e-16.

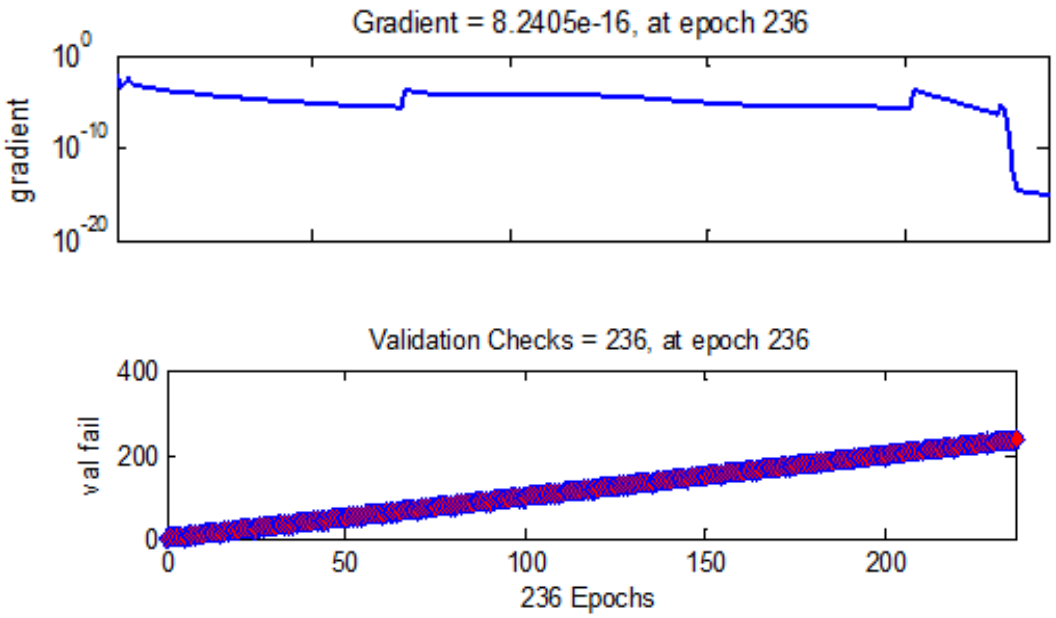
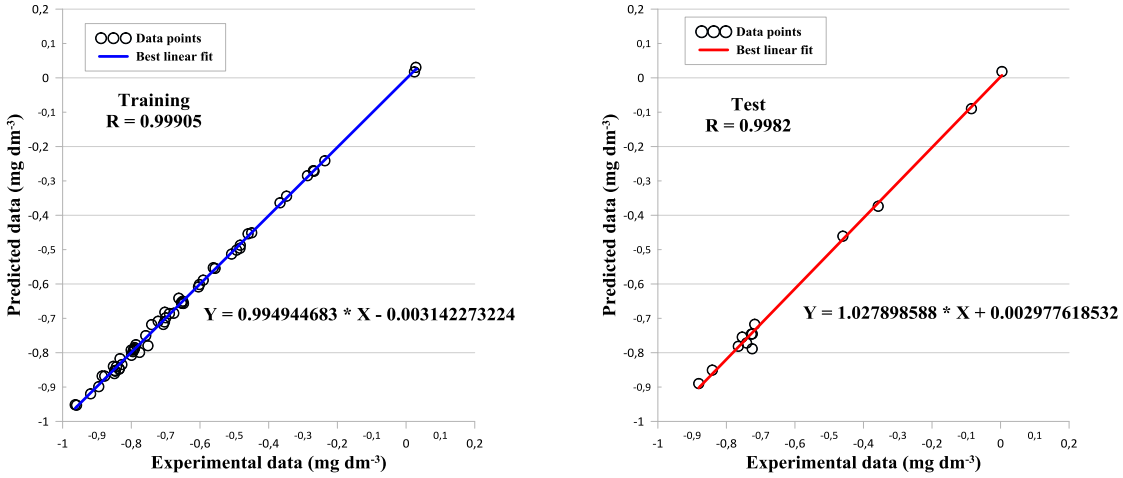


Figure III.12. Variations de gradient d'erreur, de taux d'apprentissage et d'erreur de validation en fonction du nombre d'itérations

Une fois l'architecture du modèle de réseau neuronal obtenue, le modèle doit être validé en comparant ses résultats avec les données expérimentales obtenues en laboratoire. Les résultats donnés par la Figure III.13 montrent les performances et l'efficacité du modèle RNA développé. La pertinence du modèle a été confirmée par le coefficient de corrélation obtenu R = 0,99839 avec l'erreur quadratique moyenne RMSE = 0,01402 pour l'ensemble de données.



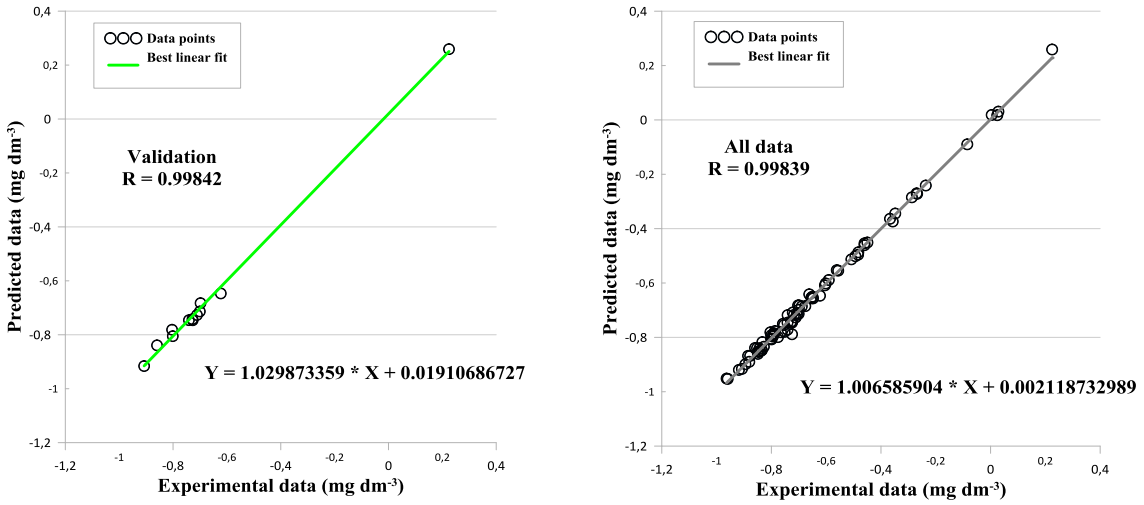
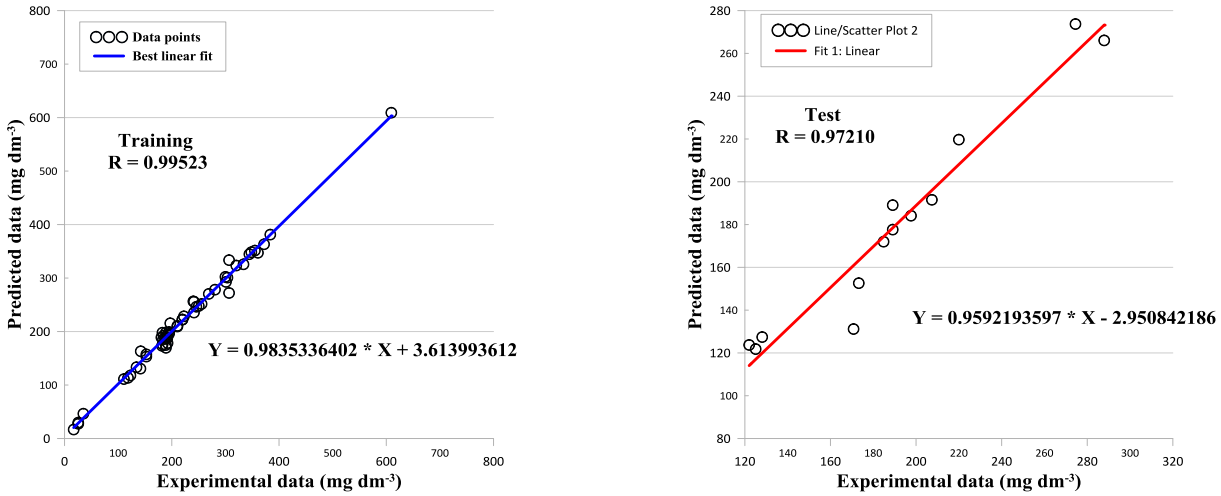


Figure III.13. Relation entre les niveaux de bicarbonate observés et ceux estimés par le modèle RNA

III.2.1. Performances de prédiction

Afin de tester la précision du modèle RNA précédemment développé et optimisé, une interpolation a été réalisée. Pour cela, une base de données a été construite contenant un ensemble de points de données situés au milieu entre les points expérimentaux de bicarbonate de l'eau potable dans la région de Médéa. Les résultats montrant la régression du bicarbonate entre les valeurs prédites et expérimentales et la performance de l'interpolation en termes d'erreur et de coefficient de corrélation sont présentés sur la Figure III.14.



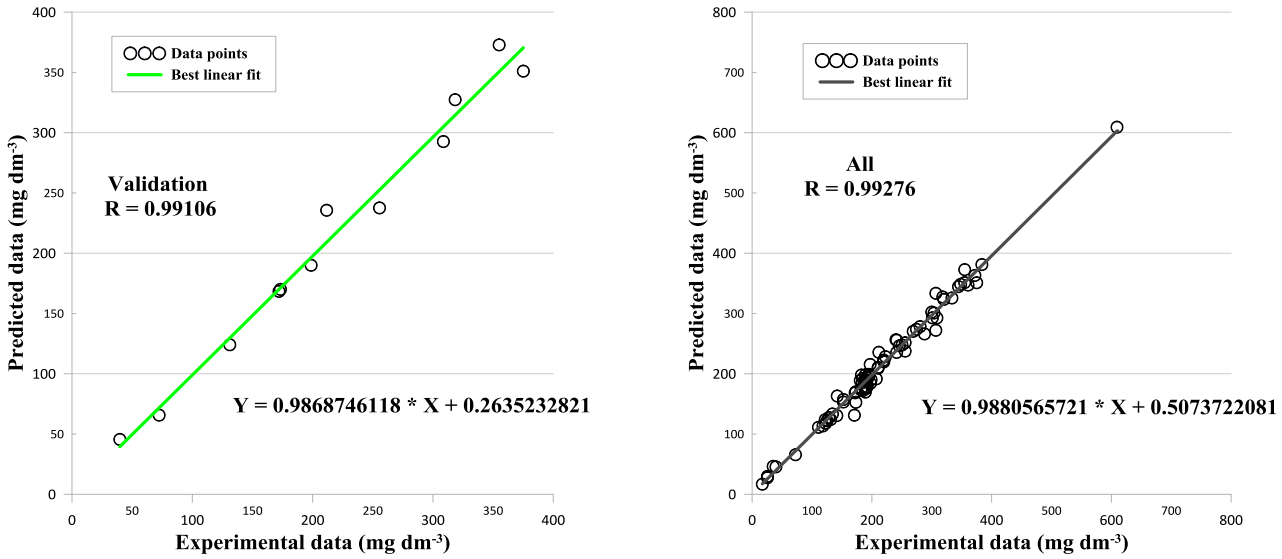


Figure III.14. Comparaison entre les données expérimentales et prédites estimées par le modèle RNA

Les résultats des performances de prédiction en termes de toutes les erreurs et en termes de valeurs du vecteur d'accord [α = (pente), β (ordonnée à l'origine), R (coefficient de corrélation)] sont résumés dans le tableau III.9.

Tableau III.9. Résultats des performances de prédiction en termes de toutes les erreurs et vecteur d'accord [α, β, R]

	RMSE / mg dm^{-3}	MSE / mg dm^{-3}	ESP / %	EPM / %	MAE / mg dm^{-3}	*R*	*α*	*β*
Apprentissage	9.68913055	93.8792508	4.3826748	1.31881052	6.54764823	0.99523	0.9835336	3.6139936
Teste	15.734532	247.575496	8.28240564	0.09771214	10.9639556	0.9721	0.9592193	-2.950842
Validation	14.0333691	196.935448	6.44782517	0.09058195	12.0114691	0.99106	0.9868746	0 .2635232
Tous	11.526138	132.85185	5.3433443	3.13764295	8.02930904	0.99276	0.9880565	0.5073722

Ces résultats montrent une bonne corrélation entre les valeurs prédit et les valeurs expérimentales avec un coefficient de corrélation élevé (R = 0,99276).

La Figure III.15 montre à nouveau les performances de la configuration RNA [18–5–1] avec une meilleure superposition des courbes tracées expérimentalement et celles prédites par le modèle déduit lors de notre étude. Les résultats obtenus par cette étude montrent une bonne concordance expliquée par un coefficient de corrélation élevé, qui a une valeur proche de l'unité. De plus, les valeurs des erreurs y compris (MSE, RMSE, ESP, EPM, MAE) sont très faibles.

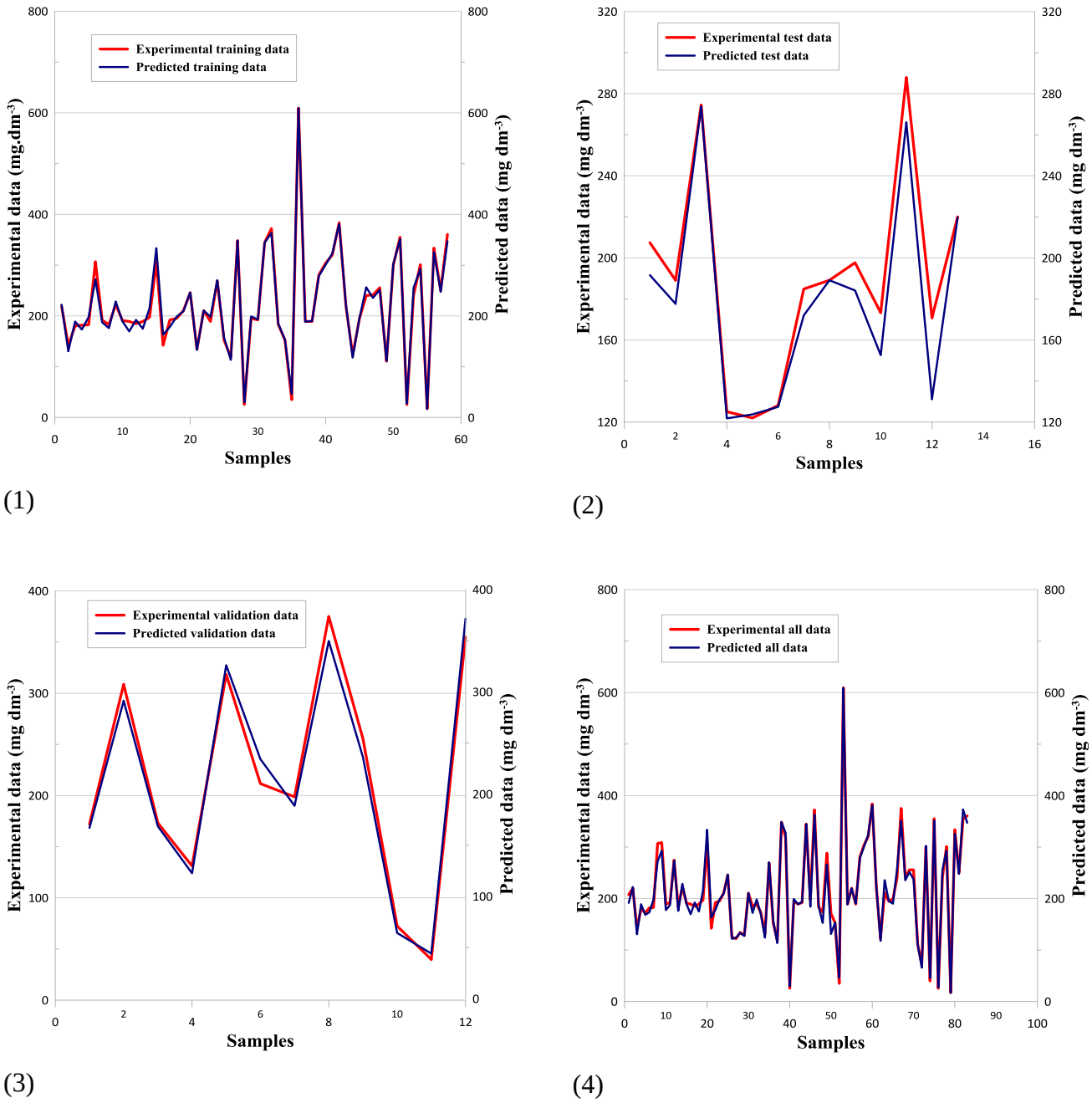


Figure III.15. Relation entre les données expérimentales et les données prédites d'échantillons à l'aide de la modélisation RNA: (1) formation, (2) test, (3) validation et (4) toutes les données

III.2.2 Etude des résidus

Les erreurs faites par le modèle des réseaux de neurones artificiels sur les échantillons qui ont été utilisés dans cette étude sont appelées résidus30. Ainsi, l'étude de la relation entre les teneurs estimées en bicarbonate par les modèles mathématiques et leurs résidus (yexp - ypred) nous a permis d'assurer les performances du modèle et de les vérifier expérimentalement.

La Figure III.16 montre les relations entre les niveaux estimés de bicarbonate dans l'eau de la région de Médéa et leurs résidus obtenus à l'aide de réseaux de neurones (RNA), respectivement. Il montre que les résidus obtenus par les réseaux de neurones sont beaucoup moins dispersés (plus proches de zéro) et une nette amélioration de la distribution des résidus en fonction de la nature des échantillons.

Cette distribution prouve le pouvoir prédictif de ce modèle établi par les réseaux de neurones dans la prédiction des teneurs en bicarbonate à partir des paramètres environnementaux. En général, le résultat obtenu est très satisfaisant et justifie l'utilisation de l'approche par réseau de neurones dans la prédiction des teneurs en bicarbonate dans la région de Médéa.

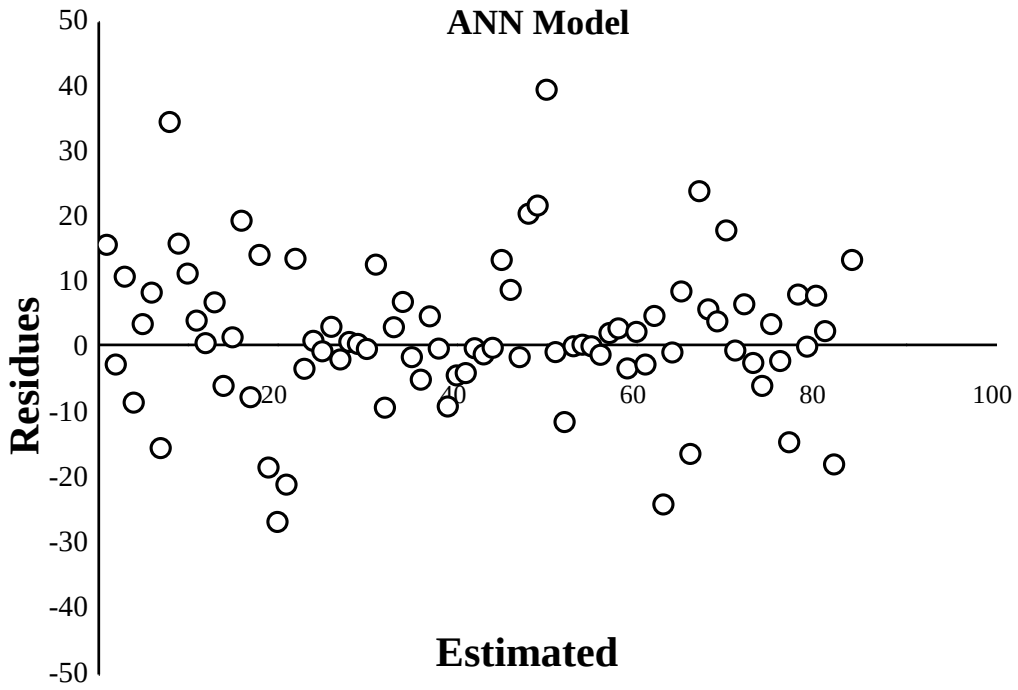


Figure III.16. Résidus relatifs aux modèles établis par le réseau de neurones artificiels en fonction de valeurs estimées

Conclusion générale

Conclusion générale

Conclusion générale :

- Dans cette étude, quatre modèles mathématiques (RSM, RNA, SVM et ANFIS) ont été utilisés pour construire des modèles pour prédire la dose de coagulant dans le but d'éliminer les composants organiques et la turbidité. Aussi, afin de réduire le gaspillage de produits chimiques coûteux et les coûts d'exploitation et d'atteindre les objectifs de qualité de l'eau potable.
- Les conclusions suivantes peuvent être tirées des résultats obtenus:
- Les modèles RSM, RNA, SVM et ANFIS se sont avérés être des modèles utiles et efficaces pour prédire les doses de coagulants du point de vue des coefficients et des indicateurs statistiques. Cependant, le modèle RNA semble être le modèle le plus efficace, car il conduit aux coefficients les plus élevés (R = 0,9975, R^2 = 0,99500, R^2 adj = 0,99400) et aux indicateurs d'erreur les plus faibles (RMSE = 1,79535, MSE = 3,22331, ESP = 4,83966) , EPM = 0,28386 et MAE = 0,67928) par rapport aux autres modèles. Suivi du modèle SVM (R = 0.99668, R^2 = 0.99337, R^2adj = 0.99204, RMSE = 2.08927, MSE = 4.36505, ESP = 5.63194, EPM = 0.58983 et MAE = 0.98625), le modèle RSM (R = 0.98691, R^2 = 0.97399 , R^2 adj = 0.96878, RMSE = 3.82890, MSE = 14.66053, ESP = 10.441194, EPM = 1.47823 et MAE = 3.82890) et enfin le modèle ANFIS (R = 0.97137, R^2 = 0.94355, R^2 adj = 0.93227, RMSE = 6.14477, MSE = 37.75828 , ESP = 16,56418, EPM = 2,81801 et MAE = 2,71994) .En outre, d'un point de vue économique, seuls trois modèles semblaient pertinents, le modèle RNA qui impliquait 43, le modèle SVM qui était presque similaire au modèle RNA en termes de coefficients et d'indicateurs statistiques mais n'impliquait que 22 paramètres, et enfin le modèle RSM, qui reste très utile en termes de réduction des coûts économiques du fait de ses coefficients élevés (R = 0,98691, R^2 = 0,97399, R^2 adj = 0,96878) s'il est lié au nombre de paramètres, seulement 13. En outre, cela a conduit à des indicateurs statistiques acceptables. Au contraire, le modèle ANFIS a été jugé inacceptable, en raison du nombre très élevé de paramètres, 309.laborieuse.
- Le modèle des doses de coagulation a été donné des résultats extraordinaires. Car, les doses qui ont injecté ont été éliminé les matières organiques parfaitement et aussi les autres paramètres physicochimiques de l'eau.
- Ce document a souligné l'importance de la modélisation de la qualité de l'eau potable basée sur la teneur en sulfate soluble dans le paramètre de l'eau potable. Dans cette étude, dans un premier temps, des échantillons ont été collectés en différents points de

la région de Médéa, en Algérie. Elle a été suivie d'une deuxième étape, qui a été consacrée à l'analyse de ces échantillons en laboratoire afin d'avoir les paramètres physico-chimiques (sulfate, conductivité, turbidité, potentiel hydrogène, dureté, calcium, magnésium, chlorures, alcali total métrique) titre (TAC), matière organique, dioxyde d'azote, nitrates, sodium, bicarbonate, potassium, métaux lourds (Mn^{2+}, Fe^{3+} et Al^{+}) et résidus secs). Ensuite, la base de données a été construite; une analyse statistique a été appliquée sur les paramètres physico-chimiques pour la prédiction de la teneur en sulfate soluble dans l'eau potable de la région de Médéa à l'aide de deux méthodes de modélisation: la régression linéaire multiple et les réseaux de neurones pour la prédiction de la teneur en sulfate soluble dans l'eau potable afin de les comparer. . De plus, l'erreur quadratique moyenne (RMSE) et l'erreur absolue moyenne (MAE) ont été utilisées pour évaluer l'efficacité de ces deux modèles.

- Les résultats ont montré que le modèle prédictif basé sur des réseaux de neurones artificiels de configuration [18-8-1] avec des fonctions de transfert de tangente hyperbolique dans la couche cachée et dans la couche de sortie et l'algorithme d'apprentissage de type Levenberg Marquardt étaient meilleurs que ceux établis par plusieurs régression linéaire. En effet, le coefficient de corrélation du modèle RNA était très élevé (R = 0. 99973) par rapport au modèle MLR. En outre, l'errLeur quadratique moyenne, l'erreur quadratique moyenne et l'erreur absolue moyenne du modèle RNA étaient inférieures à celles établies par régression linéaire multiple, 5,9755 et 88,3068 mg / l pour RMSE, 35,7067 et 7798,0935 mg / l pour MSE, et 3,1767 et 42,3274 mg / l pour le MAE, respectivement.
- le modèle RNA a été retesté en considérant deux bases de données pour l'interpolation et l'extrapolation pour évaluer son efficacité. Les résultats de l'interpolation et de l'extrapolation ont montré une efficacité élevée, en raison de leur coefficient de corrélation très élevé et des faibles valeurs de l'erreur quadratique moyenne (RMSE), de l'erreur quadratique moyenne (MSE) et de l'erreur moyenne absolue (MAE). Par conséquent, l'efficacité de notre modèle est confirmée. Les résultats de l'interpolation et de l'extrapolation étaient les suivants: coefficient de corrélation R = 0. 0,99918 et R = 0,9989, l'erreur quadratique moyenne (RMSE = 9,3595 mg / l et 12,2587 mg / l), l'erreur quadratique moyenne (MSE = 87,6019 mg / l et 150,2771 mg / l) et l'erreur absolue moyenne (MAE = 6,7598 mg / l et 10,9054 mg / l), respectivement.

- Cette performance semble être due au fait que la teneur en sulfate soluble était liée aux caractéristiques physico-chimiques de l'environnement avec des relations non linéaires. De plus, les graphiques de résidus ont montré la puissance des réseaux de neurones dans la modélisation des données.
- la prédiction des concentrations de bicarbonate dans l'eau potable de la région de Médéa a été obtenue à l'aide de réseaux de neurones artificiels (RNA) avec apprentissage supervisé, impliquant l'algorithme "Levenberg-Marquardt". Cet algorithme donne de meilleurs résultats en termes de vitesse, de convergence et de génération de performances. Les résultats ont montré une capacité d'apprentissage et de prédiction élevée pour les concentrations de bicarbonate avec un coefficient de corrélation très élevé de 0,99276 et une erreur quadratique moyenne très faible (11,52613 mg dm-3) pour l'ensemble de la base de données. De plus, ils ont montré un meilleur choix d'algorithme d'apprentissage, de fonctions d'activation et d'architecture de réseau [18-5-1] obtenu en appliquant des indicateurs statistiques de robustesse. De plus, la prédiction par les réseaux de neurones artificiels montre une bonne corrélation entre les valeurs expérimentales et prédites, ce qui signifie que le modèle RNA a un meilleur pouvoir prédictif. Cette performance semble être due au fait que les concentrations de bicarbonates dans l'eau potable de la région de Médéa, sont liées aux caractéristiques physico-chimiques de l'environnement par des relations non linéaires. Cela nous encourage à envisager de développer d'autres aspects de cette étude liés à l'avenir pour utiliser plus de paramètres et prédire d'autres propriétés.

Références bibliographiques

Références bibliographiques

[1] J.J. Milledge, S. Heaven, A review of the harvesting of micro-algae for biofuel production, Reviews in Environmental Science and Bio/Technology. 12 (2013) 165–178.

[2] E. Lefebvre, B. Legube, Coagulation par Fe (III) de substances humiques extraites d'eaux de surface: Effet du pH et de la concentration en substances humiques, Water Research. 24 (1990) 591–606.

[3] E.M. Thurman, Organic geochemistry of natural waters, Springer Science & Business Media, 2012.

[4] T. Hama, The seasonal variation of organic constituents in a eutrophic lake, Lake Japan Part II. Dissolved organic matter, Arch. Hydrobiol. 98 (1983) 443–462.

[5] A. Jadas-Hecart, A. El Morer, M. Stitou, P. Bouillot, B. Legube, Modélisation de la demande en chlore d'une eau traitée, Water Research. 26 (1992) 1073–1084.

[6] J.-S. Kim, E.S.K. Chian, F.M. Saunders, E.M. Perdue, M.F. Giabbai, Characteristics of humic substances and their removal behavior in water treatment, in: ACS Publications, 1987.

[7] B. Pan, B. Xing, Adsorption mechanisms of organic chemicals on carbon nanotubes, Environmental Science & Technology. 42 (2008) 9005–9013.

[8] E. Duursma, R. Dawson, Marine Organic Chernistry, Amsterdam. Elsevier Scientific. Publishing. Co, 1981.

[9] D. Harvey, Spaces of global capitalism, Verso, 2006.

[10] F. Birgand, E. Novince, Facteurs expliquant la présence de Matière Organique dans les Eaux Superficielles en Bretagne: Analyse des données existantes, irstea, 2004.

[11] P. Petitjean, O. Henin, G. Gruau, Dosage du carbone organique dissous dans les eaux douces naturelles. Intérêt, Principe, Mise en Oeuvre et Précautions Opératoires, 2004.

[12] J.-P. Croué, Contribution à l'étude de l'oxydation par le chlore et l'ozone d'acides fulviques naturels extraits d'eaux de surface, Poitiers, 1987.

[13] J. MALLEVIALE, E. SCHMITT, A. BRUCHET, Composés organiques azotés dans les eaux: inventaire et évolution dans différentes filières industrielles de production d'eau potable, Journées Informations Eaux. 1 (1982).

[14] C. Volk, C. Renner, P. Roche, H. Paillard, J.C. Joret, Effects of ozone on the production of biodegradable dissolved organic carbon (BDOC) during water treatment, Ozone: Science & Engineering. 15 (1993) 389–404.

[15] R.H.S. Jansen, Ozonation of humic substances in a membrane contactor: Mass transfer, product characterization and biodegradability., (2005).

[16] D.M. Owen, G.L. Amy, Z.K. Chowdhury, Characterization of natural organic matter and its relationship to treatability, Foundation and American Water Works Association, 1993.

Références bibliographiques

[17] M. CAMBREZY, V. EAU, A. HENRY, Les eaux de baignade: gestion active et nouveaux outils de détection des pollutions fécales, L'Eau, l'industrie, Les Nuisances. (2008) 81–91.

[18] A. Bergelin, P.A.W. Van Hees, O. Wahlberg, U.S. Lundström, The acid–base properties of high and low molecular weight organic acids in soil solutions of podzolic soils, Geoderma. 94 (2000) 223–235.

[19] F.J. Stevenson, Nature of divalent transition metal complexes of humic acids as revealed by a modified potentiometric titration method, Soil Science. 123 (1977) 10–17.

[20] F. Andreux, La matière organique des sols. Interactions avec les polluants, Comptes Rendus de l'Académie d'Agriculture de France. 83 (1997) 111–122.

[21] A. Hatira, T. Gallali, J. Rouiller, B. Guillet, Stabilité et solubilité des complexes formés entre le cuivre, le plomb, le zinc et les acides fulviques, Science Du Sol (1984). 28 (1990) 123–135.

[22] N. HARRAT, Elimination de la matière organique naturelle dans une filière conventionnelle de potabilisation d'eaux de surface, Université Mohamed Khider–Biskra, 2013.

[23] F.J. Stevenson, A. Fitch, M.S. Brar, Stability constants of Cu (II)-humate complexes: comparison of select models, Soil Science. 155 (1993) 77–91.

[24] E.M. Logan, I.D. Pulford, G.T. Cook, A.B. MACKENZIE, Complexation of Cu2+ and Pb2+ by peat and humic acid, European Journal of Soil Science. 48 (1997) 685–696.

[25] S.M. Ross, Toxic metals in soil-plant systems, Wiley Chichester, 1994.

[26] K.M. Elkins, D.J. Nelson, Spectroscopic approaches to the study of the interaction of aluminum with humic substances, Coordination Chemistry Reviews. 228 (2002) 205–225.

[27] A.A. Petrenko, B.H. Jones, T.D. Dickey, M. Lehaitre, C. Moore, Effects of a sewage plume on the biology, optical characteristics, and particle size distributions of coastal waters, Journal of Geophysical Research: Oceans. 102 (1997) 25061–25071.

[28] P. Rubini, A. Lakatos, D. Champmartin, T. Kiss, Speciation and structural aspects of interactions of Al (III) with small biomolecules, Coordination Chemistry Reviews. 228 (2002) 137–152.

[29] J. Buffle, ANALYSIS AND CHARACTERIZATION OF NATURAL ORGANIC MATTERS IN FRESHWATERS. II: COMPARISON OF THE PROPERTIES OF WATERS OF VARIOUS ORIGIN AND THEIR ANNUAL TREND, (1982).

[30] C. Rav-Acha, M. Rebhun, Binding of organic solutes to dissolved humic substances and its effects on adsorption and transport in the aquatic environment, Water Research. 26 (1992) 1645–1654.

[31] S.E. Cabaniss, Q. Zhou, P.A. Maurice, Y.-P. Chin, G.R. Aiken, A log-normal distribution model for the molecular weight of aquatic fulvic acids, Environmental Science & Technology. 34 (2000) 1103–1109.

Références bibliographiques

[32] J. Peuravuori, P. Ingman, K. Pihlaja, R. Koivikko, Comparisons of sorption of aquatic humic matter by DAX-8 and XAD-8 resins from solid-state 13C NMR spectroscopy's point of view, Talanta. 55 (2001) 733–742.

[33] C.H. Specht, M.U. Kumke, F.H. Frimmel, Characterization of NOM adsorption to clay minerals by size exclusion chromatography, Water Research. 34 (2000) 4063–4069.

[34] J.L. Zhou, S. Rowland, R. Fauzi, C. Mantoura, J. Braven, The formation of humic coatings on mineral particles under simulated estuarine conditions—a mechanistic study, Water Research. 28 (1994) 571–579.

[35] M. Schnitzer, S.U. Khan, Humic substances in the environment. Marcel Dekker, New York., Humic Substances in the Environment. Marcel Dekker, New York. (1972).

[36] G. Cathalifaud, J. Ayele, M. Mazet, Optimisation of micropollutant removal onto powdered activated carbon during the coagulation-flocculation step, Aqua- Journal of Water Supply: Research and Technology. 44 (1995) 55–59.

[37] L.J. Michot, I. Bihannic, M. Pelletier, E. Rinnert, J.-L. Robert, Hydration and swelling of synthetic Na-saponites: Influence of layer charge, American Mineralogist. 90 (2005) 166–172.

[38] A. Rezeg, S. Achour, Incidence des groupements fonctionnels acides dans les mécanismes d'élimination de la matière organique par coagulation-floculation, Séminaire International" L'Eau et Le Risque Dans Le Contexte Saharien. 19 (2004).

[39] M.J. Semmens, K. Ayers, Removal by coagulation of trace organics from Mississippi river water, Journal-American Water Works Association. 77 (1985) 79–84.

[40] E. Lefebvre, B. Legube, Coagulation-floculation par le chlorure ferrique de quelques acides organiques et phenols en solution aqueuse, Water Research. 27 (1993) 433–447.

[41] H. Linda, Effet de la dureté des eaux sur la coagulation-floculation de composés organiques phénoliques, (n.d.).

[42] M. Rahni, Coagulation-floculation de quelques composés organiques par le fer ferrique en milieu aqueux: étude du mécanisme et comparaison avec l'adsorption sur oxyde métallique, Poitiers, 1994.

[43] M.R. Jekel, Interactions of humic acids and aluminum salts in the flocculation process, Water Research. 20 (1986) 1535–1542.

[44] M. Doré, Chimie des oxydants et traitement des eaux, Technique et Documentation-Lavoisier, 1989.

[45] J.J. Rook, R. JJ, Formation of haloforms during chlorination of natural waters, (1974).

[46] H. Thibaud, Contribution a l'etude du mecanisme de formation de la chloropicrine au cours de traitements oxydants: application aux eaux a potabiliser, Poitiers, 1987.

Références bibliographiques

[47] S. Achour, S. Guergazi, Incidence de la minéralisation des eaux algériennes sur la réactivité de composés organiques vis-à-vis du chlore, Revue Des Sciences de l'eau/Journal of Water Science. 15 (2002) 641–660.

[48] K.L. Murphy, R. Zaloum, D. Fulford, Effect of chlorination practice on soluble organics, Water Research. 9 (1975) 389–396.

[49] J. De Laat, N. Merlet, M. Dore, Chloration de composés organiques: demande en chlore et réactivite vis-a-vis de la formation des trihalométhanes. Incidence de l'azote ammoniacal, Water Research. 16 (1982) 1437–1450.

[50] Z. Alouini, R. Seux, Cinétiques et mécanismes de l'action oxydative de l'hypochlorite sur les acides α-aminés lors de la désinfection des eaux, Water Research. 21 (1987) 335–343.

[51] G. Martin, Le problème de l'azote dans les eaux. Technique et documentation, Lavoisier. Paris. (1979).

[52] F. Afoufou, S. Achour, Incidence du procédé combiné préchloration–floculation sur l'élimination des substances humiques, Larhyss Journal. 58 (2004) 63–73.

[53] A.N. Van Breemen, T.J. Nieuwstad, G.C. Van Der Meent-Olieman, The fate of fulvic acids during water treatment, Water Research. 13 (1979) 771–779.

[54] O. Hoyer, H. Bernhardt, B. Lüsse, The effect of ozonation on the impairment of flocculation by algogenic organic matter, Zeitschrift Für Wasser-Und Abwasser-Forschung. 20 (1987) 123–131.

[55] A. Bazer-Bachi, E. Puech-Coste, R. Ben Aim, J. Probst, Modélisation mathématique du taux de coagulant dans une station de traitement d'eau, Revue Des Sciences de l'eau/Journal of Water Science. 3 (1990) 377–397.

[56] A. Amina, G. Khawla, Modélisation du système de coagulation-floculation: cas de la station de traitement de l'eau potable Ain Dalia-Souk Ahras, (2015).

[57] S. Heddam, A. Bermad, N. Dechemi, Modélisation de la dose de coagulant par les systèmes à base d'inférence floue (ANFIS) application à la station de traitement des eaux de Boudouaou (Algérie), Revue Des Sciences de l'eau/Journal of Water Science. 25 (2012) 1–17.

[58] R. Desjardins, Le traitement des eaux, Presses inter Polytechnique, 1997.

[59] R. Desjardins, Le traitement des eaux, Presses inter Polytechnique, 1997.

[60] M. Achak, N. Ouazzani, A. Yaacoubi, L. Mandi, Caractérisation des margines issues d'une huilerie moderne et essais de leur traitement par coagulation-floculation par la chaux et le sulfate d'aluminium., Revue Des Sciences de l'eau/Journal of Water Science. 21 (2008) 53–67.

[61] C. Cardot, Les traitements de l'eau: procédés physico-chimiques et biologiques: cours et problèmes résolus, Ellipses Edition Marketing, 1999.

[62] L. Sigg, P. Behra, W. Stumm, Chimie des milieux aquatiques, Dunod, 2001.

[63] L. Sigg, W. Stumm, P. Behra, Chimie des milieux aquatiques: chimie des eaux naturelles et des interfaces dans l'environnement, Masson, 1994.

[64] C. Lévêque, Ecosystèmes aquatiques, 1996.

[65] J. Rodier, C. Geoffray, L. Rodi, L'analyse de l'eau: eaux naturelles, eaux résiduaires, eau de mer: chimie, physico-chimie, bactériologie, biologie, Dunod, 1975.

[66] L. Chahid, A. Yaacoubi, A. Bacaoui, E. Lakhal, Valorization of drinking water treatment sludge (DWTS): Characterization and applications as coagulant and sorbent for Olive Mill Wastewater (OMW), J. Mater. Environ. Sci. 6 (2015) 2520–2533.

[67] C.D. Guedes, Coagulation-floculation des eaux issues des mines de fer: comparaison entre les coagulants minéraux et un coagulant organique issu de la moringa oléifera, Aix-Marseille 3, 2004.

[68] C. Sun, Q. Yue, B. Gao, B. Cao, R. Mu, Z. Zhang, Synthesis and floc properties of polymeric ferric aluminum chloride–polydimethyl diallylammonium chloride coagulant in coagulating humic acid–kaolin synthetic water, Chemical Engineering Journal. 185 (2012) 29–34.

[69] H.J. Rakotoniriana, N.R. Randriana, J. de Dieu Ramaroson, F. Randrianarivelo, A.M. Herihajaniavo, L. Andrianaivo, Etude comparative des coagulants dans le traitement des eaux., (n.d.).

[70] S. Achour, N. Guesbaya, Essais de floculation de substances humiques en milieux aqueux minéralisés, LARHYSS Journal P-ISSN 1112-3680/E-ISSN 2521-9782. (2006).

[71] A. Guyot, R. Audebert, R. Botet, B. Cabane, F. Lafuma, R. Jullien, E. Pefferkorn, C. Pichot, A. Revillon, R. Varoqui, Floculation de particules colloïdales par les polymères hydrosolubles, Journal de Chimie Physique. 87 (1990) 1859–1899.

[72] A. Rezeg, Elimination d'acides organiques hydroxylés et carboxylés par coagulation-floculation au sulfate d'aluminium, Université Mohamed Khider-Biskra, 2004.

[73] D. Bouyer, Analyse expérimentale de la floculation: influence de l'hydrodynamique sur les phénomènes d'agrégation et de rupture, Toulouse, INSA, 2002.

[74] J. Coma, A. Jabbouri, A. Grasmick, S. Elmaleh, Floculation en lit fluidisé, Water Research. 24 (1990) 361–366.

[75] M. Larakeb, Elimination du zinc par adsorption et par coagulation-floculation, Université Mohamed Khider-Biskra, 2015.

[76] B. LAKHDARI, Effet de la coagulation-floculation sur la qualité des eaux épurées de la STEP de Ain El Houtz, (2011).

[77] C. bel Houceine, M. Khalid, Contribution à l'étude de la floculation des levures: application à la champagnisation, Dijon, 1987.

[78] F. Zidane, K. Cheggari, J.-F. Blais, P. Drogui, J. Bensaid, S.I. Ahmed, Contribution à l'étude de l'effet de la coagulation avant chloration sur la formation des trihalométhanes (THM) et composés

organohalogénés (COX) dans les eaux alimentant la ville de Casablanca au Maroc, Canadian Journal of Civil Engineering. 37 (2010) 1149–1156.

[79] K. Djeffal, S. Bouranene, P. Fievet, S. Déon, A. Gheid, Treatment of controlled discharge leachate by coagulation-flocculation: influence of operational conditions, Separation Science and Technology. (2019) 1–16.

[80] S. Bouranene, N. Sedira, P. Fievet, N. Attia, Treatment of paint wastewater by coagulation process, Filtration+ Separation. 52 (2015) 42–45.

[81] J. Rodier, B. Legube, N. Merlet, R. Brunet, L'analyse de l'eau - 9e éd.: Eaux naturelles, eaux résiduaires, eau de mer, Dunod, 2009. https://books.google.dz/books?id=qUEGsUBZkL0C.

[82] D.M.A. DE LA VILLE, D.R.D.N. L'AIDE, PREDICTION DE LA TURBIDITE A L'EAU BRUTE ET A L'EAU FILTREE, (2007).

[83] F. Sekiou, Effets de la nature des particules solides dispersées et des conditions de formation du floc sur l'efficacité de la floculation, Mémoires de Magister. (2014).

[84] P.-A. Montenegro Rousseau, Identification de facteurs influençant les plaintes relatives à l'eau potable, (2006).

[85] J. Painchaud, La qualité de l'eau des rivières du Québec: état et tendances, Ministère de l'environnement et de la faune, 1997.

[86] L.A. HAMEL, Etude physico-chimique et bactériologique de l'eau de la source «Sidi Bouyashak» destinée à la consommation humaine de la population de Tlemcen., 2013.

[87] L. Aizel, Variabilité temporelle de la qualité organoleptique de l'eau potable issue du fleuve Saint-Laurent: contribution du manganèse et du COT, École de technologie supérieure, 2015.

[88] J.K. Morris, W.R. Knocke, Temperature effects on the use of metal-ion coagulants for water treatment, Journal-American Water Works Association. 76 (1984) 74–79.

[89] J. Rajonson, N. Rasolofonirina, J. Ratoaveloson, N. Ravaonindriana, Qualité des eaux, ELOUARD, 1992.

[90] M.F. Mohtadi, P.N. Rao, Effect of temperature on flocculation of aqueous dispersions, Water Research. 7 (1973) 747–767.

[91] K.N. Gan, A. Smolen, H.W. Eckerson, B.N. La Du, Purification of human serum paraoxonase/arylesterase. Evidence for one esterase catalyzing both activities., Drug Metabolism and Disposition. 19 (1991) 100–106.

[92] J.P. Julien, W.E. Mushynski, The distribution of phosphorylation sites among identified proteolytic fragments of mammalian neurofilaments., Journal of Biological Chemistry. 258 (1983) 4019–4025.

[93] J.-Y. Bottero, B. Lartiges, Séparation liquide/solide par coagulation-floculation: les coagulants-floculants, mécanismes d'agrégation, structure et densité des flocs./Liquid/solid separation by

coagulation-flocculation, coagulants-flocculants, flocs structure and density, Sciences Géologiques, Bulletins et Mémoires. 46 (1993) 163–174.

[94] R.D. Letterman, J.E. Quon, R.S. Gemmell, Influence of rapid-mix parameters on flocculation, Journal-American Water Works Association. 65 (1973) 716–722.

[95] A. Vilgé Ritter, ETUDE DES MECANISMES D'ELIMINATION DE LA MATIERE ORGANIQUE DES EAUX DE SURFACE PAR COAGULATION-FLOCULATION A L'AIDE DE SELS D'ALUMINIUM ET DE FER, 1997.

[96] H. Yamamura, E.U. Putri, T. Kawakami, A. Suzuki, H.D. Ariesyady, T. Ishii, Dosage optimization of polyaluminum chloride by the application of convolutional neural network to the floc images captured in jar tests, Separation and Purification Technology. 237 (2020) 116467. https://doi.org/10.1016/j.seppur.2019.116467.

[97] D. Jamin, Recherche du boson de Higgs du Modèle Standard dans le canal de désintégration ZH-> nu nu bb sur le collisionneur Tevatron dans l'expérience D0. Développement d'une méthode d'étiquetage des jets de quark b avec des muons de basses impulsions transverses, 2010.

[98] R. Yacef, A. Mellit, S. Belaid, Z. Şen, New combined models for estimating daily global solar radiation from measured air temperature in semi-arid climates: application in Ghardaïa, Algeria, Energy Conversion and Management. 79 (2014) 606–615.

[99] H. El Badaoui, A. Abdallaoui, I. Manssouri, L. Lancelot, Elaboration de modèles mathématiques stochastiques pour la prédiction des teneurs en métaux lourds des eaux superficielles en utilisant les réseaux de neurones artificiels et la régression linéaire multiple, Journal of Hydrocarbons Mines and Environmental Research. 3 (2012) 31–36.

[100] H. Tahraoui, A.E. Belhadj, A.E. Hamitouche, Prediction of the Bicarbonate Amount in Drinking Water in the Region of Médéa Using Artificial Neural Network Modelling, Kemija u Industriji. 69 (2020) 595–602. https://doi.org/10.15255/KUI.2020.002.

[101] D.A. Belsley, E. Kuh, R.E. Welsch, Regression diagnostics: identifying influential data and sources of collinearity, Wiley, 1980. https://books.google.dz/books?id=ALjuAAAAMAAJ.

[102] S.H. Hong, M.W. Lee, D.S. Lee, J.M. Park, Monitoring of sequencing batch reactor for nitrogen and phosphorus removal using neural networks, Biochemical Engineering Journal. 35 (2007) 365–370.

[103] O.R. Dolling, E.A. Varas, Artificial neural networks for streamflow prediction, Journal of Hydraulic Research. 40 (2002) 547–554.

[104] I. Manssouri, M. Manssouri, B. El Kihel, FAULT DETECTION BY K-NN ALGORITHM AND MLP NEURAL NETWORKS IN A DISTILLATION COLUMN: COMPARATIVE STUDY, Journal of Information, Intelligence and Knowledge. 3 (2011) 201.

[105] I. Manssouri, A. El Hmaidi, T.E. Manssouri, B. El Moumni, Prediction levels of heavy metals (Zn, Cu and Mn) in current Holocene deposits of the eastern part of the Mediterranean Moroccan margin (Alboran Sea), IOSR Journal of Computer Engineering. 16 (2014) 117–123.

Références bibliographiques

[106] D.C. Montgomery, Design and analysis of experiments, 8. ed, Wiley, Hoboken, NJ, 2013.

[107] A. Hammoudi, K. Moussaceb, C. Belebchouche, F. Dahmoune, Comparison of artificial neural network (ANN) and response surface methodology (RSM) prediction in compressive strength of recycled concrete aggregates, Construction and Building Materials. 209 (2019) 425–436. https://doi.org/10.1016/j.conbuildmat.2019.03.119.

[108] E. Ghafari, H. Costa, E. Júlio, RSM-based model to predict the performance of self-compacting UHPC reinforced with hybrid steel micro-fibers, Construction and Building Materials. 66 (2014) 375–383.

[109] D.K. Latifa, P.H. Salah, Etude de l'impacte de la modélisation et de l'optimisation de la fermentation des antibiotiques sur l'environnement, (n.d.) 94.

[110] H. Benimam, C. Si-Moussa, M. Laidi, S. Hanini, Modeling the activity coefficient at infinite dilution of water in ionic liquids using artificial neural networks and support vector machines, Neural Computing and Applications. (2019). https://doi.org/10.1007/s00521-019-04356-w.

[111] G.-D. Wu, S.-L. Lo, Predicting real-time coagulant dosage in water treatment by artificial neural networks and adaptive network-based fuzzy inference system, Engineering Applications of Artificial Intelligence. 21 (2008) 1189–1195. https://doi.org/10.1016/j.engappai.2008.03.015.

[112] H. El Badaoui, A. Abdallaoui, L. Lancelot, Application des réseaux de neurones artificiels et des régressions linéaires multiples pour la prédiction des concentrations des métaux lourds dans les sédiments fluviaux marocains, Université de Poitiers/France. (2012).

[113] A. Schmitt, B. Le Blanc, M.-M. Corsini, C. Lafond, J. Bruzek, Les réseaux de neurones artificiels. Un outil de traitement de données prometteur pour l'anthropologie, Bulletins et Mémoires de La Société d'Anthropologie de Paris. (2001).

[114] M. Naoual, A. Abdelaziz, E.H. Abdellah, Use of Artificial Neural Networks Type MLP for the Prediction of Phosphorus Level from the Physicochemical Parameters of Sediments, (n.d.) 10.

[115] S. Lefnaoui, N. Moulai-Mostefa, Investigation and optimization of formulation factors of a hydrogel network based on kappa carrageenan–pregelatinized starch blend using an experimental design, Colloids and Surfaces A: Physicochemical and Engineering Aspects. 458 (2014) 117–125.

[116] M. Bouhedda, S. Lefnaoui, S. Rebouh, M.M. Yahoum, Predictive model based on Adaptive Neuro-Fuzzy Inference System for estimation of Cephalexin adsorption on the Octenyl Succinic Anhydride starch, Chemometrics and Intelligent Laboratory Systems. 193 (2019) 103843. https://doi.org/10.1016/j.chemolab.2019.103843.

[117] Y. Laidani, G. Henini, B. Khatmi, A. Dellal, Evaluation de la pollution des eaux du sous bassin versant de l'oued mina Y. 2eme colloque international de chimie, CIC2 Algérie-Du Lau. 3 (2009).

[118] S. Ghizellaoui, Contribution à l'évaluation et l'inhibition du pouvoir entartrant des eaux dures par plusieurs procédés, (n.d.).

[119] K. Hu, Y. Huang, H. Li, B. Li, D. Chen, R.E. White, Spatial variability of shallow groundwater level, electrical conductivity and nitrate concentration, and risk assessment of nitrate contamination in North China Plain, Environment International. 31 (2005) 896–903.

[120] A. Rasekh, K. Brumbelow, Drinking water distribution systems contamination management to reduce public health impacts and system service interruptions, Environmental Modelling & Software. 51 (2014) 12–25.

[121] H. Debbih, B.E. Naili, Etude de qualité des eaux des barrages de l'Est Algérien, (2015).

[122] K. Bendada, M.W. Boulakradeche, Optimisation des conditions de dosage par spectroscopie d'absorption atomique (SAAF et SAAET) Application à la détermination de la pollution et de la bioaccumulation des métaux lourds [mémoire], Alger: Université Des Sciences et de La Technologie Houari Boumediene (USTHB). (2011).

[123] S. Behar, Definition of water quality parameters, Testing the Waters: Chemical and Physical Vital Signs of a River. River Watch Network, Montpelier. (1997).

[124] M. Franceschi, A. Girou, A.M. Carro-Diaz, M.T. Maurette, E. Puech-Costes, Optimisation of the coagulation–flocculation process of raw water by optimal design method, Water Research. 36 (2002) 3561–3572.

[125] N. Harrat, S. Achour, Pollution physico-chimique des eaux de Barrage de la région d'El Tarf. Impact sur la chloration, LARHYSS Journal P-ISSN 1112-3680/E-ISSN 2521-9782. (2010).

[126] M. Bengoumi, A. Traoure, N. Bouchriti, D. Bengoumi, A. Hraiki, Qualité de l'eau en aviculture, Revue Trimestrielle d'information Scientifique et Technique. 3 (2004) 5–25.

[127] M. CHABBI, N. NEZLI, Qualité des eaux de consommation de la vallée d'Oued Souf, (2018).

[128] S.K. Sarkar, M. Saha, H. Takada, A. Bhattacharya, P. Mishra, B. Bhattacharya, Water quality management in the lower stretch of the river Ganges, east coast of India: an approach through environmental education, Journal of Cleaner Production. 15 (2007) 1559–1567.

[129] N.K. Sharma, B. Goswami, B. Gajjar, C. Jain, D. Soni, K. Patel, Utilization of Amul dairy effluent for agriculture practices, International Journal of Environmental Sciences. 2 (2011) 22–32.

[130] C.R. Watson, Risk-Based Decision Making During Public Health Emergencies Involving Environmental Contamination, Johns Hopkins University, 2017.

[131] P.N. Soltanpour, J. Self, Soil, water, and plant testing, Crop Series. Soil; No. 0.507. (1996).

[132] G.J. Kirmeyer, Guidance manual for monitoring distribution system water quality, American Water Works Association, 2002.

[133] G. Degremont, Mémento technique de l'eau, Tome 1, 10éme édition, Edit, Tec et Doc. (2005).

[134] P.T. Srinivasan, T. Viraraghavan, K.S. Subramanian, Aluminium in drinking water: An overview, Water Sa. 25 (1999) 47–55.

[135] P.T. Srinivasan, T. Viraraghavan, Characterisation and concentration profile of aluminium during drinking-water treatment, Water SA. 28 (2002) 99–106.

[136] A. El Hmaidi, H. El Badaoui, A. Abdallaoui, B. El Moumni, Application des réseaux de neurones artificiels de type PMC pour la prédiction des teneurs en carbone organique dans les dépôts du quaternaire terminal de la mer d'Alboran, European Journal of Scientific Research. 107 (2013) 400–413.

[137] C. Voyant, M. Muselli, C. Paoli, M.-L. Nivet, Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation, Energy. 36 (2011) 348–359.

[138] M. Bélanger, N. El-Jabi, D. Caissie, F. Ashkar, J. Ribi, Estimation de la température de l'eau de rivière en utilisant les réseaux de neurones et la régression linéaire multiple, Revue Des Sciences de l'eau/Journal of Water Science. 18 (2005) 403–421.

Annexe

Predicting the Concentration of Sulfate (SO_4^{2-}) in Drinking Water using Artificial Neural Networks: a case study: Médéa-Algeria

Hichem TAHRAOUI[1], Abd-Elmouneïm BELHADJ[1], Adhya-eddine HAMITOUCHE[2], Mounir BOUHEDDA[3] and Abdeltif AMRANE[4]

[1]*Laboratory of Biomaterials and Transport Phenomena (LBMPT), University of MÉDÉA, 26000 Médéa, ALGERIA*
E-mail : tahraoui.hichem@univ-Médéa.dz, E-mail: belhadj_1@yahoo.fr, URL: www.univ-Médéa.dz

[2] *Centre de Recherche scientifique et technique en Analyses Physico-chimiques CRAPC, BP 384, Bou-Ismail, RP 42004, Tipaza, ALGERIA*
E-mail: ahamitouche2@yahoo.fr URL: www.crapc.dz

[3]*Laboratory of Advanced Electronic Systems (LSEA), University of MÉDÉA, 26000 Médéa, ALGERIA*
E-mail: bouhedda.m@gmail.com, URL: www.univ-Médéa.dz

[4]*Univ Rennes, Ecole Nationale Supérieure de Chimie de Rennes, CNRS, ISCR – UMR6226, F-35000 Rennes, France*
E-mail: abdeltif.amrane@univ-rennes1.fr URL: www.univ-Rennes1.dz

[*] *Corresponding Author:* Abdeltif AMRANE

E-mail address: abdeltif.amrane@univ-rennes1.fr

Abstract

The aim of this work was to use Artificial Neural Networks (ANN) and multiple linear regressions (MLR) models to predict the soluble sulfate content in drinking water. A set of 84 data points were used. For the ANN, 18 neurons were used in the input layer, 8 neurons at hidden layer and 1 was used in the output layer. Levenberg Marquardt learning (LM) algorithm with Hyperbolic tangent sigmoid transfer function logarithmic was used at the hidden and output layer. The comparison of the obtained results in term of root mean square error (RMSE) and correlation coefficient (R) using the ANN and MLR models revealed the superiority of the (ANN) model in predicting the soluble sulfate content in drinking water. Indeed, the statistical results showed a correlation coefficient R = 0.99973 with root mean square error RMSE = 5.9755 for the ANN model and R = 0.941 with RMSE = 88.3068 for the MLR model. A nonlinear relationship between the soluble sulfate content and the physico-chemical characteristics of drinking water (conductivity, turbidity, potential hydrogen, hardness, calcium, magnesium, chlorides, total alkali metric titre (TAC), material Organic, Nitrogen dioxide, nitrates, sodium, bicarbonate, potassium, heavy metals (Mn^{2+}, Fe^{3+} and Al^{+}) and dry residues) was demonstrated, showing that the soluble sulfate content concentration can be predicted.

Keywords: Drinking water, Physico-chemical parameters, Sulfate, Modelling, Artificial Neural Networks, multiple linear regressions.

1. Introduction

Water is a basic element for the sustainable development of the city. Drinking water is essential for human survival and regional stability[1]. Thus, water quality is an extremely important environmental factor as it affects human beings and their economic activities[2]. The main factor that affects the physical appearance such as the colour of the water is the concentration and distribution of suspended fine components and dissolved matter[2]. Industrial, agricultural and urban development is altering water quality and making it unsafe. This is the case of the Medea region, which is subject to various types of pollution and an increase in the quantity of pollutants released into the aquatic environment without treatment. Depending on the origin of the waste, the pollution can be of a chemical nature. Water quality tends to degenerate progressively with human interventions, such as hydrological alterations[3], land use changes[4], inputs of toxic chemicals and nutrients[5] and changes in other physico-chemical properties of water[6] this causes a range of environmental problems, such as the soluble sulfate content in drinking water levels. the soluble sulfate content in drinking water is ubiquitous and can be found not only in natural waters, but also in industrial wastewater[7,8]. Although the soluble sulfate content in drinking water is generally not considered a health problem, concentrations of soluble sulfate content in drinking water can cause a bitter taste and can cause diarrhoea when its concentration exceeds 600 mg/l[9]. The release of high levels of sulfate can significantly affect the water supply by causing corrosion and/or scaling of pipes and equipment. In addition, hydrogen sulphide (H_2S), which is toxic to the ecosystem, could be produced by sulfate reduction by sulfate-reducing bacteria under anaerobic conditions[10]. Because of these adverse effects on human health and the environment, many countries have set maximum soluble sulfate content in drinking water concentration values ranging from 250 mg/l to 500 mg/l, depending on the end use of the water source[8]. Traditional methods of measuring and studying water quality are both time-consuming and costly compared to numerical modelling techniques, especially if they deal with large areas. Recently, many authors have studied parameters affecting water quality using artificial intelligence. Several works have found great success in the simulation and prediction of environmental parameters [11] such as the prediction of several environmental parameters dealing with water quality in rivers of different countries [12–18], the prediction of the indicators of quality water for urban source management [19], analysis of surface water quality and identification of key water parameters [20], prediction of water quality via *Escherichia coli* levels [21], prediction of soil hard-setting and physical quality using water retention data [22] and the prediction of phosphorus and total nitrogen for lake in

Egypt [23].In particular, artificial neural networks (ANNs) are a method for approximating complex systems, especially useful when these systems are difficult to model using classical statistical methods[24]. ANNs provide interesting results due to their learning capability[25], their parallelism and their ability to solve many non-linear system problems[26]. Over the last decade, ANN research has been applied in the fields of hydrology, ecology and environment. ANN models have been shown to perform better than other models, the prediction of water quality [27]. ANN models have been requested for a variety of purposes; for example: for variations in water quality attributes [28], for prediction of water quality parameters [29–33], for prediction of water quality indices [34], for estimation of lake water quality using satellite images [35], to study water quality parameters of the Axios River in Northern Greece [36], to model nitrate concentrations in rivers [37], prediction of annual drinking water quality reduction based on Groundwater Resource Index [38], prediction of the groundwater remediation costs for drinking use based on the quality of water Resources [39], In addition, the evaluation of multivariate linear regression and artificial neural networks for predicting water quality parameters were examined [40].

The main objective of this research was therefore to conduct a comparative study between multiple linear regression (MLR) and artificial neural networks (ANN) for the prediction of soluble sulfate content in drinking water in the Médéa region in Algeria.

2. Materials and methods

2.1 Database

The data used for this study were obtained from the experimental analysis of water samples taken during several sampling times during the period 2018 in the region of Médéa, Algeria (three samples per week from different regions). Analyses were done according to the Jean Rodier's book of water analysis 9th edition[41].

The dependent variable was the soluble sulfate content in drinking water. The independent variables were the physicochemical parameters: conductivity, turbidity, potential hydrogen, hardness, calcium, magnesium, chlorides, total alkali metric titre (TAC), Organic material, Nitrogen dioxide, nitrates, sodium, bicarbonate, potassium, heavy metals (Mn^{2+}, Fe^{3+} and Al^{+}) and dry residues.

2.2. Prediction Methods

Several methods were applied to address problems related to prediction and modelling of complex nonlinear systems. These methods are particularly useful when these systems are difficult to model using classical methods[42]. In this study, we were interested in the use and the comparison of two methods for predicting the soluble sulfate content levels from the physicochemical parameters in drinking water. These methods are multiple linear regression and artificial neural networks.

The data are divided into three phases (70% learning, 15% testing and 15% validation) according to the ANN model to compare the correlation coefficients and the errors between the two models[43].

2.2.1 Multiple Linear Regressions

Multiple linear regression consists in describing the relationships between a dependent variable y and several variables called independent variables $x_1, x_2, \cdots, x_i, \cdots, x_n$, where *n* is the number of independent variables. Indeed, the multiple linear regression, which is a data analysis method, is commonly used for establishing predictive models to the phenomena observed in the aquatic environment [44]. This method allows to draw a polynomial function describing the relationship between the dependant and the

independent variables and allows to determine the most significant input variables. The model can be written as in Eq.(1).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_i x_i + \cdots + \beta_n x_n + \varepsilon(x) \quad (1)$$

With

- n : number of independent variables;
- $\varepsilon(x)$: random noise (error term or regression residual);
- y : dependent variable;
- $x_1, x_2, \cdots, x_i, \cdots, x_n$: independent variables;
- β_0 : estimated ordinate at the origin;
- $\beta_0, \beta_1, \cdots, \beta_i, \cdots, \beta_n$: model coefficients.

2.2.2 Artificial Neural Network (ANN) Model

Artificial neural networks provide an alternative to mathematical modelling and they can be classified as nonparametric nonlinear models [45,46].

The neuron is the fundamental cell of an artificial neural network, which can be considered as an elementary parallel operating processor. It is a computation unit which receives a number of inputs (x_i) directly from the environment or upstream neurons (Figure 1). When information comes from a neuron, a weight is given to this latter which represents the ability of the neuron upstream to excite or inhibit the neuron downstream through its unique output passing by the activation function [47,48].

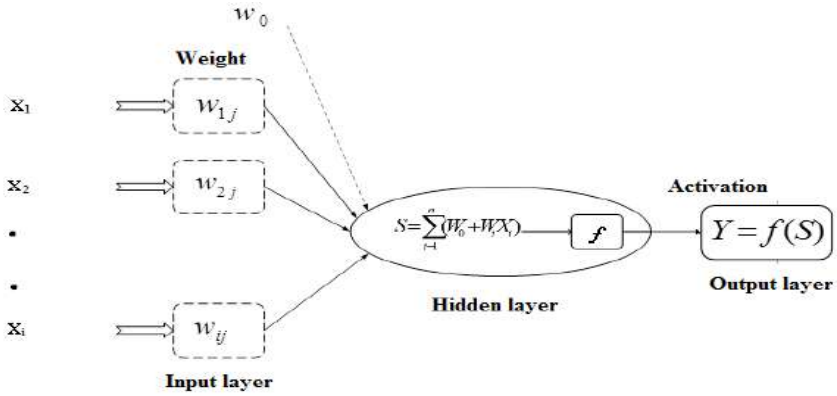


Figure 1: Artificial neuron model

The neuron's output is calculated using relation Eq. (2).

$$S_J = f\left(\sum_{i=1}^{N} w_{ij} X_i + b_j\right) \quad (2)$$

Where w_{ij} , is a synaptic weight, b_j is the bias input and X_i the i^{th} input. f is the activation function which can usually be sigmoid or hyperbolic tangent [24].

Neurons can be connected together in a way to form a multilayer artificial neural network. This latter is composed of an input layer, an output layer and one or more hidden layers. All the neurons of a layer are connected to the all neurons of the following layer through synaptic weights [49,50]. In figure 2, an

example of an ANN with one hidden layer is given where each neuron (See Figure 1) is represented by a circle.

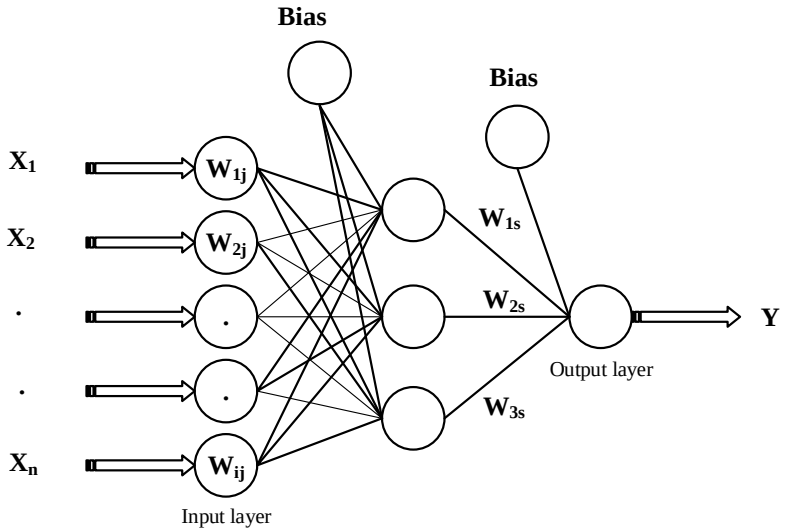


Figure 2: Multilayer neural network

The goal is introduce to the input and output data of the ANN and make it learns the relationship between them by a process called learning using specific algorithms like backpropagation or Levenberg-Marquardt [51]. The goal is to minimize the error between the model output and the desired output by adjustment of synaptic weights.

To obtain the optimal structure of neural network, we implement a strategy based on the design and optimization of the architecture of neural network.
The development of the neural network model entails the following stages[24,52]:
1: Collecting the experimental data.
2: Define the input variables and corresponding output variables.
3: Pre-treatment and analysis of the data.
4: Scaling and splitting of data for the phases of **learning** (with or without test) and validation.
5: Selection of a Neural Network model:
The selection of the ANN model is affected by four major factors:
5-1: Network type (recurrent networks, feed forward back propagation, wavelet neural network, radial basis functions, etc.). In our work we use the feed forward back propagation neural network.
5-2: Network structure (number of hidden layers, number of neurons per hidden layer).
5-3: Activation functions.
5- 4: **learning** algorithm. In our work we use **Levenberg Marquardt learning (LM) algorithm.**

First, transformation must be done in order to modify the distribution of input variables so that they can better match outputs. Before **learning** and validation, the inputs and targets are scaled using a normalized equation (Eq. 3) such that the data always fall within the interval [-1, 1][53].

$$x_N = (y_{\max} - y_{\min})\left(\frac{x - x_{\min}}{x_{\max} - x_{\min}}\right) + y_{\min} \quad (3)$$

Where x_N is the data value after normalization, $x_{\max}$ and $x_{\min}$ denote the maximum and the minimum of the data respectively; $y_{\max}$ and $y_{\min}$ are taken as 1 and -1; *x* denotes the data in question.

Data modelling was carried out by an ANN, 70 % of the dataset, chosen randomly among the totality of the samples, were used for the learning phase. The remaining 30%, which did not participate in model learning, were divided into two parts (15% for the test and 15% for the validation) to examine the validity and performance of the prediction of these models.[54].

2.3. statistical evaluation criteria

The correlation coefficient (*R*), the adjusted coefficient R^2_{adj}, The root mean square error (*RMSE*), the mean square error (*MSE)* and the Mean Absolute Error (*MAE*) were used out to estimate the performance of the model. The corresponding values are calculated using the equations [55,56]:

$$R = \frac{\sum_{i=1}^{N}\left(y_{\exp} - \overline{y}_{\exp}\right)\left(y_{pred} - \overline{y}_{pred}\right)}{\sqrt{\sum_{i=1}^{N}\left(y_{\exp} - \overline{y}_{\exp}\right)^2 \sum_{i=1}^{N}\left(y_{pred} - \overline{y}_{pred}\right)^2}} \quad (4)$$

$$R^2_{adj} = 1 - \frac{\left(1 - R^2\right)\left(N - 1\right)}{N - K - 1} \quad (5)$$

$$RMSE = \sqrt{\left(\frac{1}{N}\right)\left(\sum_{i=1}^{N}\left[\left(y_{\exp} - y_{pred}\right)\right]^2\right)} \quad (6)$$

$$MSE = \left(\frac{1}{N}\right)\left(\sum_{i=1}^{N}\left(y_{\exp} - y_{pred}\right)^2\right) \quad (7)$$

$$MAE = \left(\frac{1}{N}\right)\sum_{i=1}^{N}\left|y_{\exp} - y_{pred}\right| \quad (8)$$

Where *N* is the number of data; *K* is the number of variables (inputs); y_{ex} and y_{pr} are the experimental and the predicted values respectively; $\overline{y}_{ex}$and $\overline{y}_{pr}$ are respectively the average values of the experimental and the predicted values[57–59].

The significance level value p and the F ratio value which provide a measure of the statistical significance of the regression model were also determined. A high value of F with a minimum value of p means that the equation is significant[43].

3. Results and discussion

Artificial neural network and multiple linear regression are the two approaches, which were used for the prediction of **the soluble sulfate** concentration from the physicochemical parameters of drinking water. The two methods were performed and evaluated and then compared.

3.1 Multiple Linear Regression

A statistical analysis by the multiple linear regression method was performed using the “XLSTAT 2016” software on the database set. This method makes it possible to find the mathematical polynomial relation

(Eq. (9)) between **the soluble sulfate content** and the independent variables, which corresponded to the physico-chemical parameters from 84 experimental samples. First, the equation takes into account all 18 variables (Table 1), even those that do not seem to have a significant impact on the dependent variable. The relation obtained was therefore evaluated in order to keep only the independent variables which were characterized by a high probability value (Table 1). $y = \beta_0 + \sum_{i=1}^{18} \beta_i x_i$ (9)

Where β_i and x_i are given in Table 1.

Table 1: Values of β_i , x_i and characterization of independent variables of the RLM.

i	x_i	β_i	*Standard Error*	*T ratio*	*Prob.>* $\lvert t \rvert$
0	Constant	245.4060	435.0382	0.56	0.5746
1	Conductivity (Cond)	−0.0576	0.0676	−0.85	0.3976
2	Turbidity (NTU)	4.3994	4.8354	0.91	0.3663
3	Potential hydrogen (PH)	−28.6495	55.9304	−0.51	0.6102
4	Hardness (TH)	0.0700	2.4603	0.03	0.9774
5	Calcium (Ca)	1.7242	0.8677	1.99	0.0511
6	Magnesium (Mg)	4.1885	1.2180	3.44	0.0010*
7	Chlorides (Cl)	−1.1275	0.1141	−9.88	< 0.0001*
8	Total Alkalimetric Titer (TAC)	−0.7729	0.2281	−3.39	0.0012*
9	Organic Materials (MO)	0.4933	0.4757	1.04	0.3035
10	Nitrogen dioxide (NO_2)	−3722.542	1532.262	−2.43	0.0179*
11	Nitrates (NO_3)	0.2332	2.4743	0.09	0.9252
12	Sodium (Na)	1.8031	0.2533	7.12	< 0.0001*
13	Potassium (K)	−3.3927	4.1309	−0.82	0.4145
14	Manganese (Mn)	−5154.41	19918.46	−0.26	0.7966
15	Iron (Fe)	135558.938	18425.29	0.74	0.4644
16	Aluminum (Al)	213.5702	447.0937	0.48	0.6345
17	Dry residues (Rs)	0.0828	0.0713	1.16	0.2500
18	Bicarbonate (HCO)	−0.6784	0.1533	−4.42	< 0.0001*

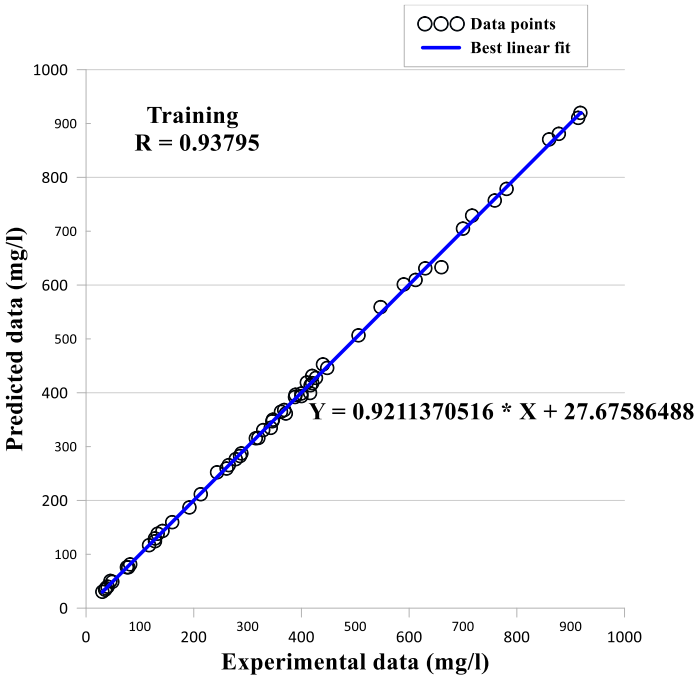
Then, only 6 independent variables which have a high power of explanation for the dependent variable were taken (Pr<0.05); thus the relation can be reduced to Eq.(10):

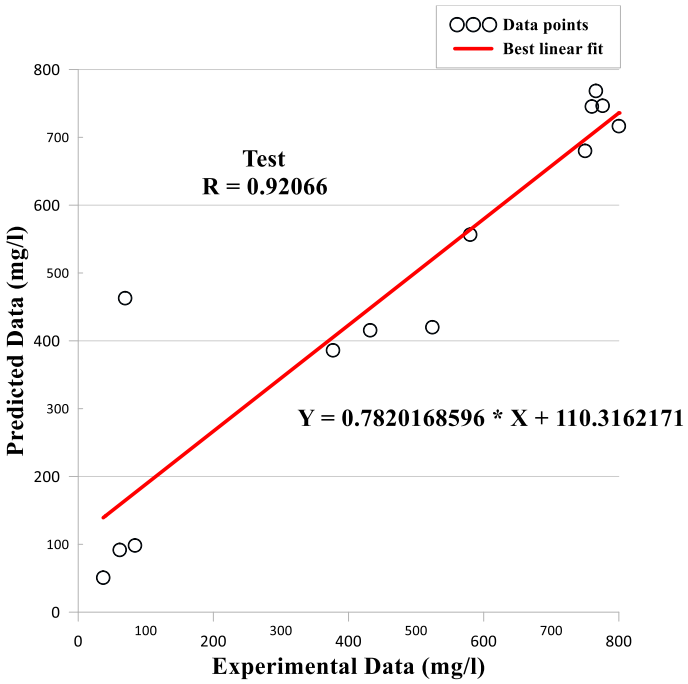
$$y = \beta_0 + \sum_{i=6}^{18} \beta_i \times x_i \quad \text{for } i = 6,7,8,10,12 \text{ and } 18 \text{ only} \tag{10}$$

Indeed, the coefficients of each factor in the model make it possible to assess the impact of each factor on the response[60].From equation (10), it is evident that changes in Magnesium and Sodium in drinking water increase the dose of sulfate. On the other hand, changes in Chlorides, Total Alkalimetric Titer, Nitrogen dioxide and Bicarbonate dominate the sulfate dose.

The value of the coefficient of determination decreased slightly, but the equation became more simple after the elimination of low-explanatory variables from the dependent variable. The result of the learning phase shows an acceptable efficiency, via their correlation coefficient (figure 3) (R = 0.93795), but poor in terms of mean square error (RMSE = 84.0254 mg/l) and mean absolute error (MAE = 48.6575 mg/ l) (Table

224 2). The result of the learning phase was tested and validated by two databases, one for the test and the
225 other for the validation. The results showed an acceptable correlation coefficient for the test phase (figure
226 3) (R = 0.92066) and a high correlation coefficient for the validation phase (figure 3) (R = 0.97100).
227 However, in terms of root mean square error (RMSE = 117.933 and 70.5674 mg/l) and absolute error
228 mean (MAE = 61.8972 and 46.6106 mg/l), they always remained bad for the test and the validation phase
229 respectively (Table 2). The three steps were combined for the final evaluation of this model which gave
230 an acceptable correlation coefficient (R = 0.941) (figure 3), but poor in terms of root mean square error
231 (RMSE = 88.3068 mg/l) and mean absolute error (MAE = 42.3274 mg/l) (Table 2). In view of the obtained
232 results, we can consider that the correlation of the model was a little positive via the acceptable
233 correlation coefficient; while the statistical evaluation criteria remained high. The probability (Pr<0.0001)
234 was strictly less than 0.5% which confirms that the model was significant [11,43].





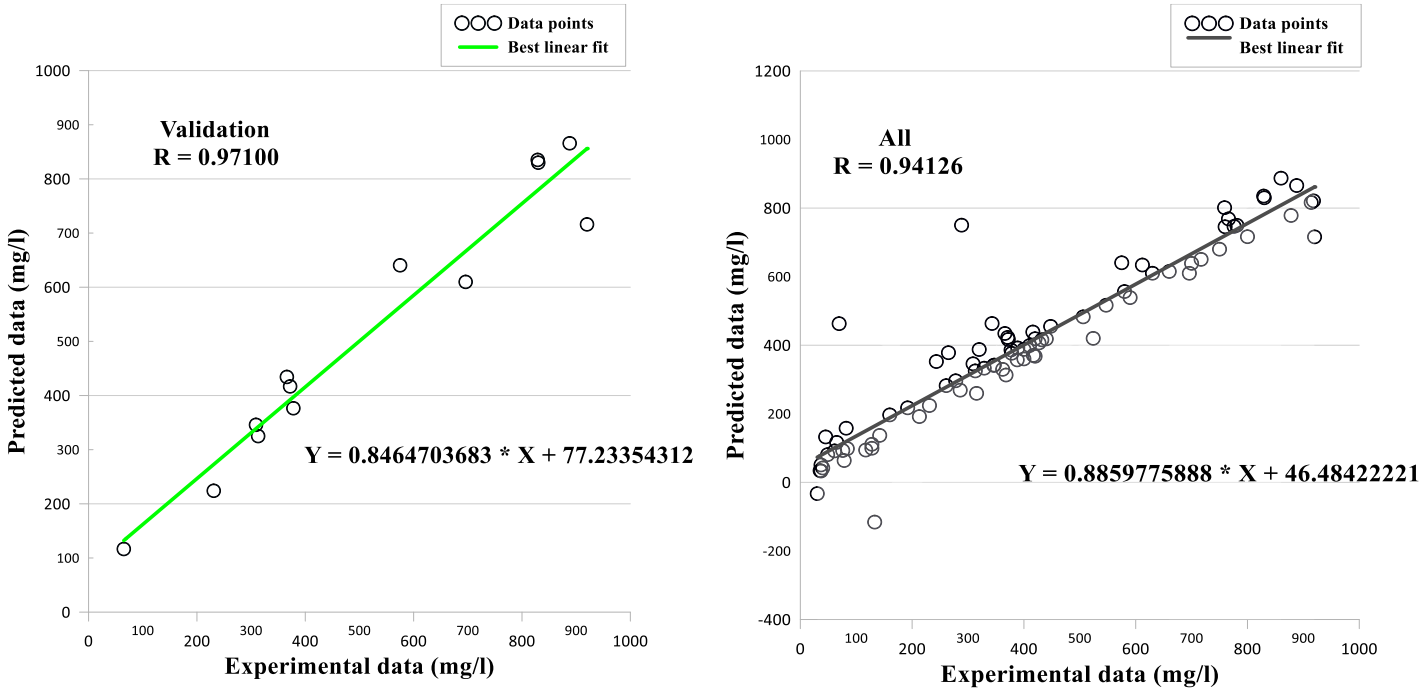


Figure 3: **Comparison between experimental and calculated values**

Results of MLR performances in terms of all errors and in terms of the agreement vector values (*R*, slope: α and *y* intercept: β) are given in Table 2.

Table 2: **Performances of the prediction using MLR**

	RMSE (mg/l)	MAE *(mg/)l*	*R*	α	β
All	88.3068	42.3274	0.941	0.8859	46.480
Training	84.0254	48.6575	0.93795	0.9211	27.6758
Test	117.9331	61.8972	0.92066	0.7820	110.3162
Validation	70.5674	46.6106	0.97100	0.8464	77.2335

3.2 ANN Modelling

Preliminary tests showed that to improve the performance of a model established by ANN, it is necessary to modify the architecture of the network, by changing mainly the number of hidden layers, the number of hidden neurons and/or the number of **learning** cycles (number of iterations). For this, we successively changed the number of hidden neurons (from 3 to 15). The results of these tests are shown in Table 3.

Table 3: Performances of the different tested ANN architectures

ANN		Activation function		Coefficients of correlation				*MSE*		
Nbr	Neurons /Layer	Hidden layer	Output layer	learning	Test	Validation	Total	Learning	Test	Validation
1	[18-9-1]	tansig	purelin	0.99809	0.99964	0.99965	0.99863	2.34×10^{-4}	6.12×10^{-5}	7.15×10^{-5}
2	[18-13-	logsig	purelin	0.99961	0.99618	0.99723	0.99873	5.47×10^{-5}	5.27×10^{-4}	3.32×10^{-4}
3	**[18-8-1]**	**tansig**	**tansig**	**0.99953**	**0.99978**	**0.99992**	**0.9996**	**7.36×10^{-5}**	**2.55×10^{-5}**	**9.79×10^{-6}**

Table 3 presents the best architectures found. It shows the correlation coefficients and the error for each learning, test and validation according to the number of neurons in the hidden layer and the network topology. It also indicates the activation functions for the hidden layer and the output layer. The architecture 3 (Table 3) appears to be the most relevant ANN model to predict the **soluble sulfate content**.

The network was driven until reaching over-learning; this phenomenon was met after 1000 iterations. The over-learning phase was not yet reached, so it was interesting to continue learning until reaching this phase for the test in order to decrease the gradient further and thus improve the precision of the ANN. Using the third architecture of Table 3, the three curves related to the evolution of the the mean square error (*MSE)* corresponding to the three learning phases were obtained (train, validation and test) where they converge correctly towards a satisfactory minimum (Figure 4).

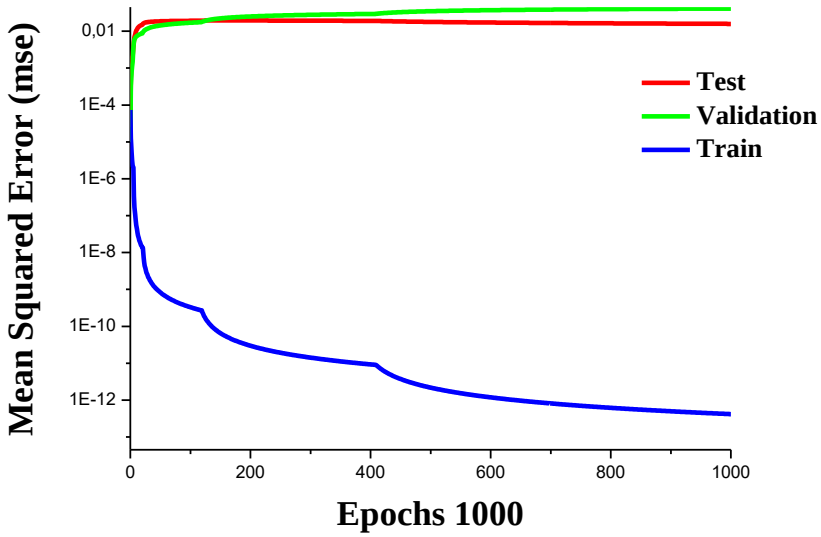


Figure 4: Evolution of the MSE corresponding to the three learning phases

Once the architecture of the ANN is synthesized, it is tested and the predicted values (estimated values) are compared to the experimental values in the three steps of learning phases; the corresponding results are given in Figure 5.

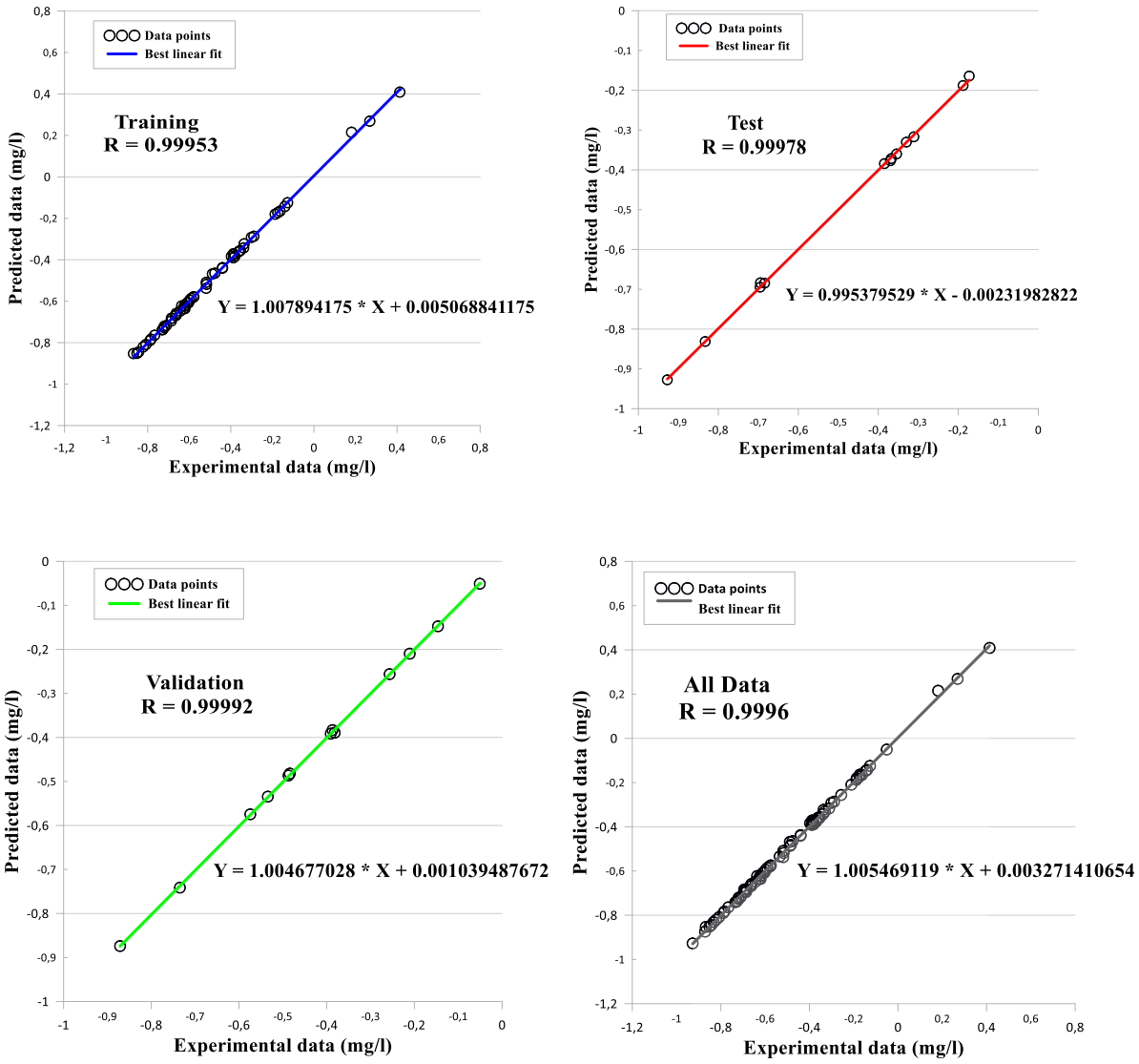


Figure 5: **Comparison between normalized experimental and calculated values**

The relation between the observed values and the estimated values shows the performance and the efficiency of the developed ANN model. The relevance of the model was confirmed by the obtained coefficient of determination *R*=0.9996 for the total data set.

3.3 Prediction performances

The result of the learning phase shows excellent efficiency, via their very high correlation coefficient (figure 6) (R = 0.99963), also very low values in terms of the mean square error (RMSE = 6.5382 mg/l and the mean absolute error (MAE = 4.2437 mg/l) (Table 4). The result of the learning phase was tested and validated by two databases, one for the test and the other for the validation; the result showed a very high correlation coefficient (R = 0.9998, 0.9999) for both phases (test and validation) respectively (figure 6), also very low values in terms of root mean square error (RMSE = 5.4650 and 3.1826 mg/l) and mean absolute error (MAE = 3.5362 and 1.9667 mg/l) for the test and the validation phases respectively (Table 4). The three steps were brought together for the final evaluation of this model which gave a very high correlation coefficient (R = 0.99973) (figure 6), and also led to very low values in terms of mean squared error (RMSE = 5.9755 mg / l) and mean absolute error (MAE = 3.1767mg / l) (Table 4). In view of the

obtained results, we can consider that the correlation of the model was very positive, showing the efficiency of the model.

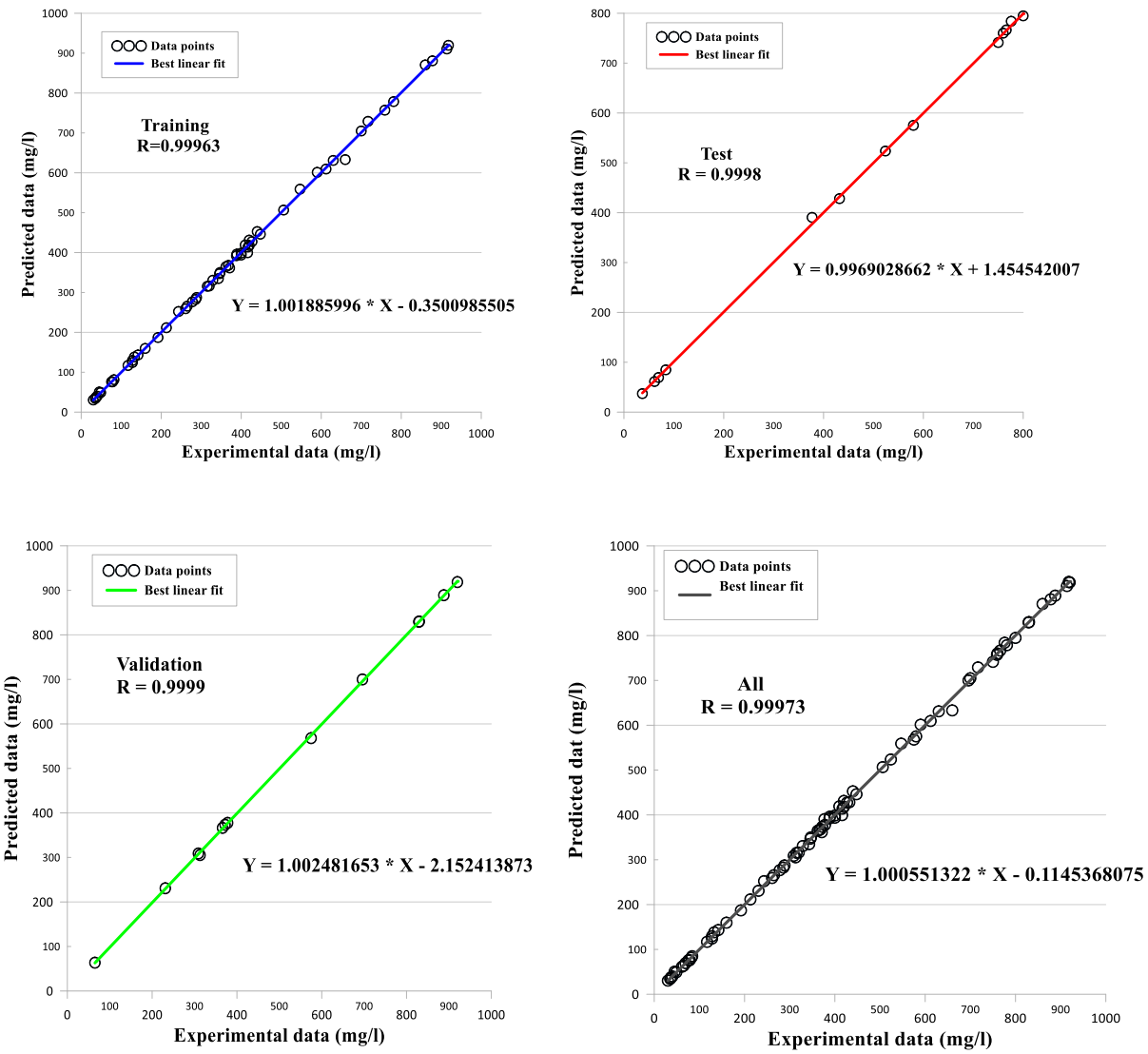


Figure 6: **Comparison between experimental and calculated values**

Table 4: Performances of the prediction using ANN

	RMSE (mg/l)	MAE *(mg/)l*	*R*	α	β
All	5.9755	3.1767	0.99973	1.0005	-0.1145
Training	6.5382	4.2437	0.99963	1.0018	-0.3500
Test	5.4650	3.5362	0.9998	0.9969	1.4545
Validation	3.1826	1.9667	0.9999	1.0024	-2.1524

3.4 Interpolation performance

In order to test the precision of the ANN model previously developed and optimized, an interpolation was performed. For this purpose, a data base was constructed **in 2019** containing a set of data points located at the experimental points. The results showing the regression curve between the predicted and experimental values and the performance of the interpolation in terms of error and correlation coefficient are shown in Figure 7.

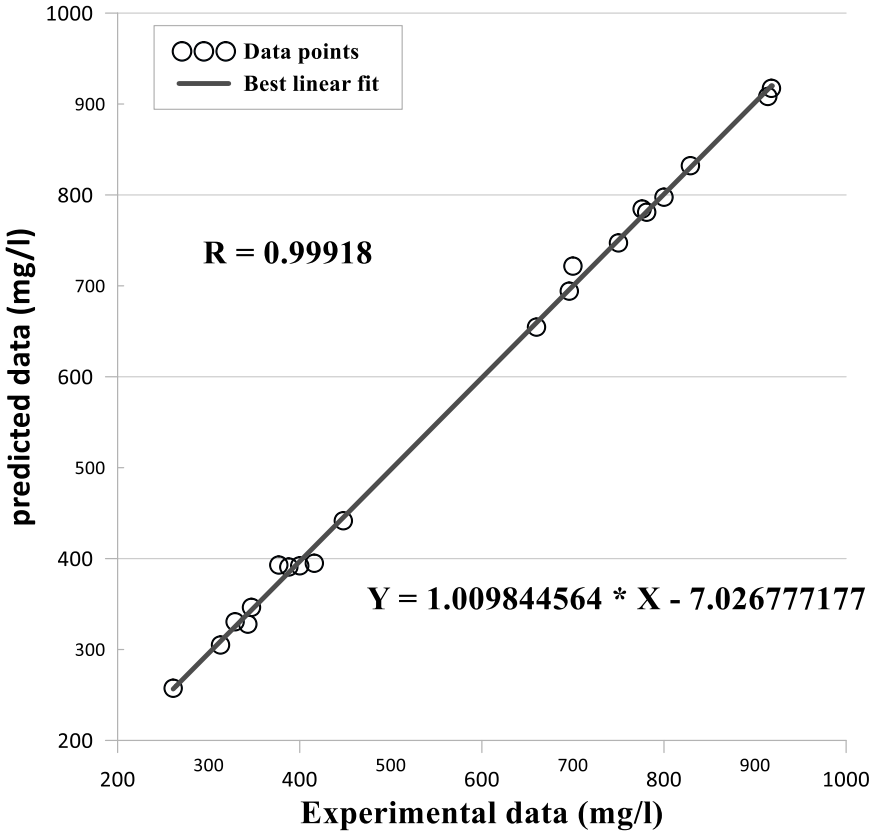


Figure 7: Comparison between experimental and calculated values

These results show a good correlation between the predicted ANN and the experimental values with a very high correlation coefficient (Figure 7) (R =0.99918) and with a minimum squared error (RMSE = 9.3595 mg/l) and absolute error (MAE = 6.7598 mg/l) (Table 5). This result confirms again model efficiency using new data in the prediction range.

Results of interpolation performances in terms of all errors and in terms of the agreement vector values [R (correlation coefficient), α= (slope) and β (y intercept)] are summarized in Table 5.

Table 5: **Performances of the Interpolation**

RMSE (mg/l)	*MAE (mg/l)*	*R*	*α*	*β*
9.3595	6.7598	0.99918	1.001	-7.027

3.5 Extrapolation performance

Prediction was also performed to test the accuracy of the artificial neural network developed. The idea was to test experimental data of the soluble sulfate content, which were not used during the training of our network; this database was built in 2019. The results in Figure8 show a very high coefficient (R= 0.9989), as well as very low values in terms of mean squared error (RMSE = 12.2587 mg/l) and mean absolute error (MAE = 10.9054 mg/l) (Table 6). Extrapolation of a database was also achieved to check

the accuracy of our optimized ANN model. The idea was testing an experimental data sets never exploited during the learning and the test phases. The experimental data set of this system exploited in our principal database was obtained by analysis of water samples collected during several campaigns carried out in the Médéa region. Results of extrapolation performances in terms of all error and in terms of the agreement vector values [R (correlation coefficient), α= (slope) and β (y intercept)] are summarized in Table 6. The quality of fit of the extrapolation data set is depicted in Figure8. The results showed a good agreement vector [R, α, β] between experimental data and the ANN predicted results with accepted root mean square error. It can be observed that for our extrapolation case, the results showed a good predictive ability of the ANN in both developed and optimized model. This shows a good convergence between the experimental output and the output predicted by the ANN.

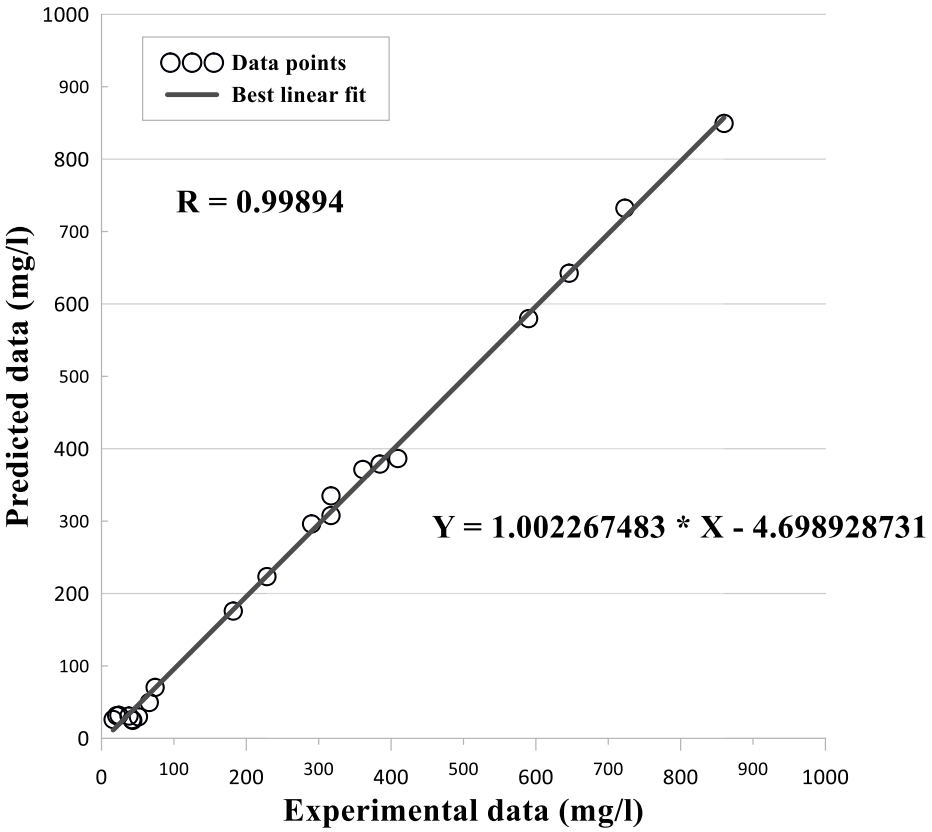


Figure 8: Comparison between experimental and calculated

Table 6: **Performances of the extrapolation**

RMSE (mg/l)	*MAE* (mg/l)	*R*	*A*	*B*
12.2587	10.9054	0.9989	1.0022	-4.6989

3.6 Comparison of MLR and ANN for predicting soluble sulfate concentrations

In order to evaluate the two developed predictors, all data for MLR and ANN were compared (Table 7). The results depicted in Table 7 show the comparison of correlation coefficients and statistical indicators obtained by the two models. The correlation coefficient calculated by the ANN was significantly higher with (R=0.99973) while that given by the MLR was less low (R=0.941). In addition, the RMSE and MAE given by the ANN model (RMSE= 88.3068 mg/l and MAE= 42.3274 mg/l) were very low compared to that of the MLR model (RMSE= 5.9755 mg/l), (MAE=3.1767 mg/l).

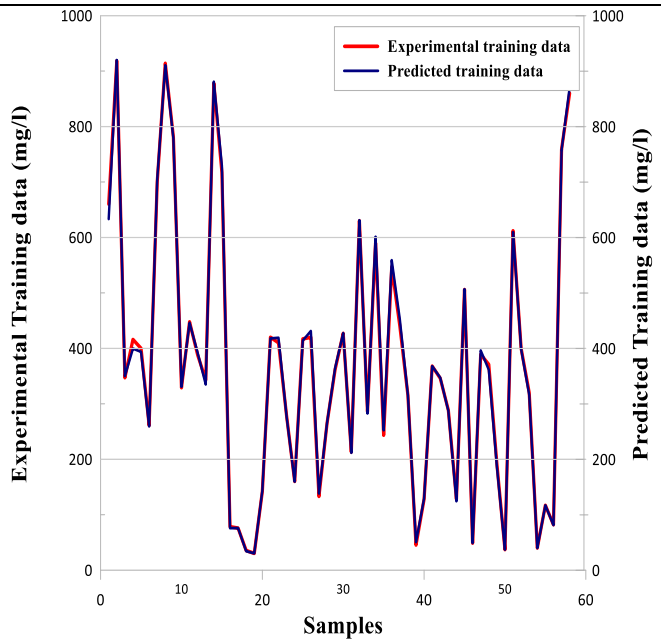
Table 7: Performance comparison between MLR and ANN prediction Models

Model	*RMSE (mg/l)*	MAE *(mg/l)*	*R*	A	B	R^2_{adj}
ANN	5.9755	3.1767	0.99973	1.0005	-0.1145	0.9996
MLR	88.3068	42.3274	0.941	0.8859	46.480	0.924

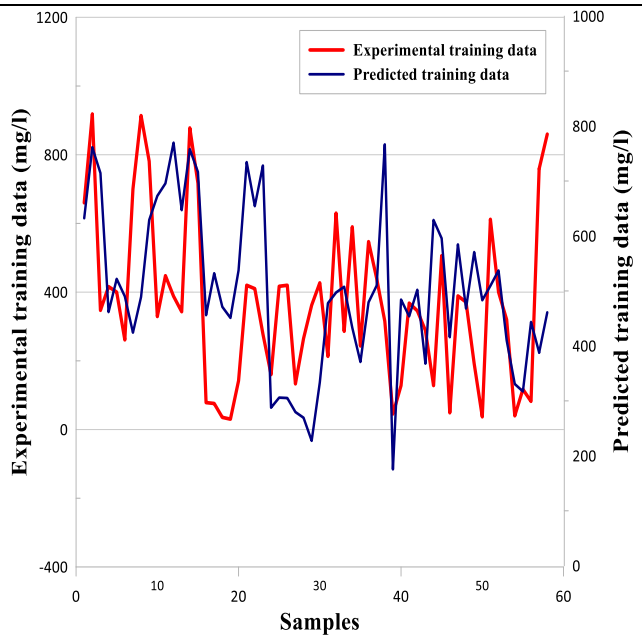
Figure 9 confirms again the superiority of the ANN model if compared to the MLR model with a very good superposition of the curves measured experimentally and those estimated by the ANN model.

The results obtained show a very good agreement for the ANN model, explained by a high correlation coefficient and small statistical indicators errors for the learning phase, the test phase and the validation phase.

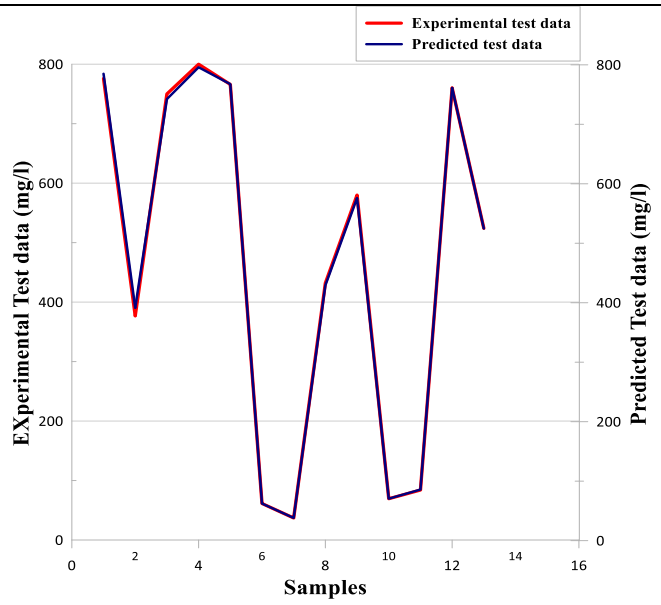
ANN Model **MLR Model**



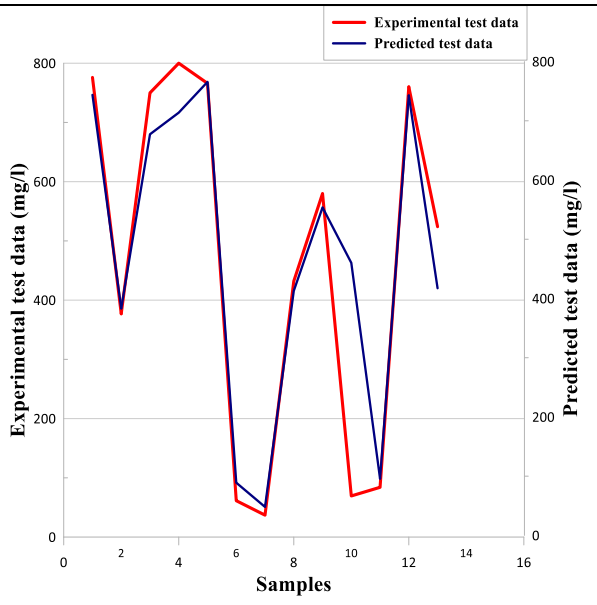
(1)



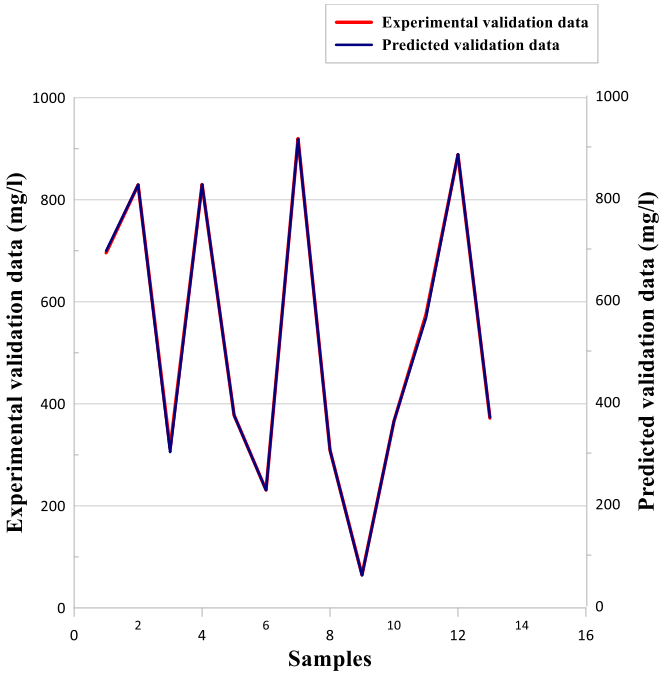
(1)


Experimental test data
Predicted test data
EXperimental Test data (mg/l)
Predicted Test data (mg/l)
Samples

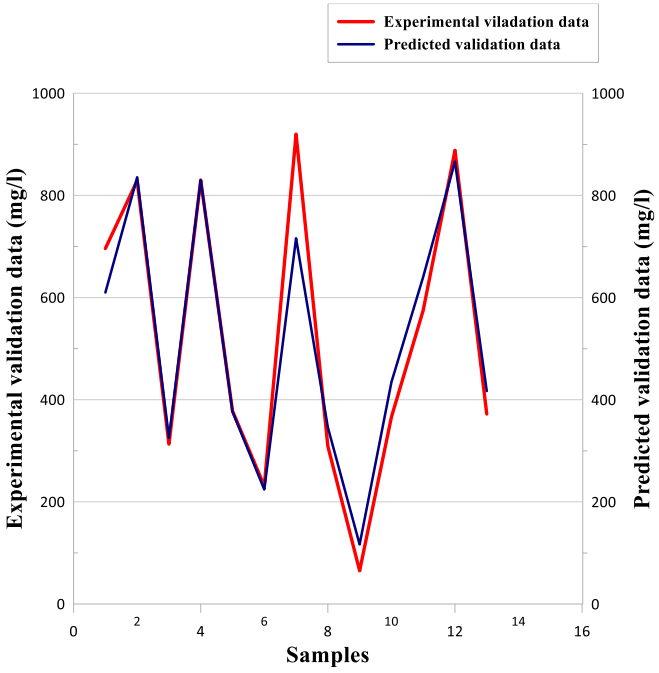
(2)


Experimental test data
Predicted test data
Experimental test data (mg/l)
Predicted test data (mg/l)
Samples

(2)


Experimental validation data
Predicted validation data
Experimental validation data (mg/l)
Predicted validation data (mg/l)
Samples

(3)


Experimental viladation data
Predicted validation data
Experimental validation data (mg/l)
Predicted validation data (mg/l)
Samples

(3)

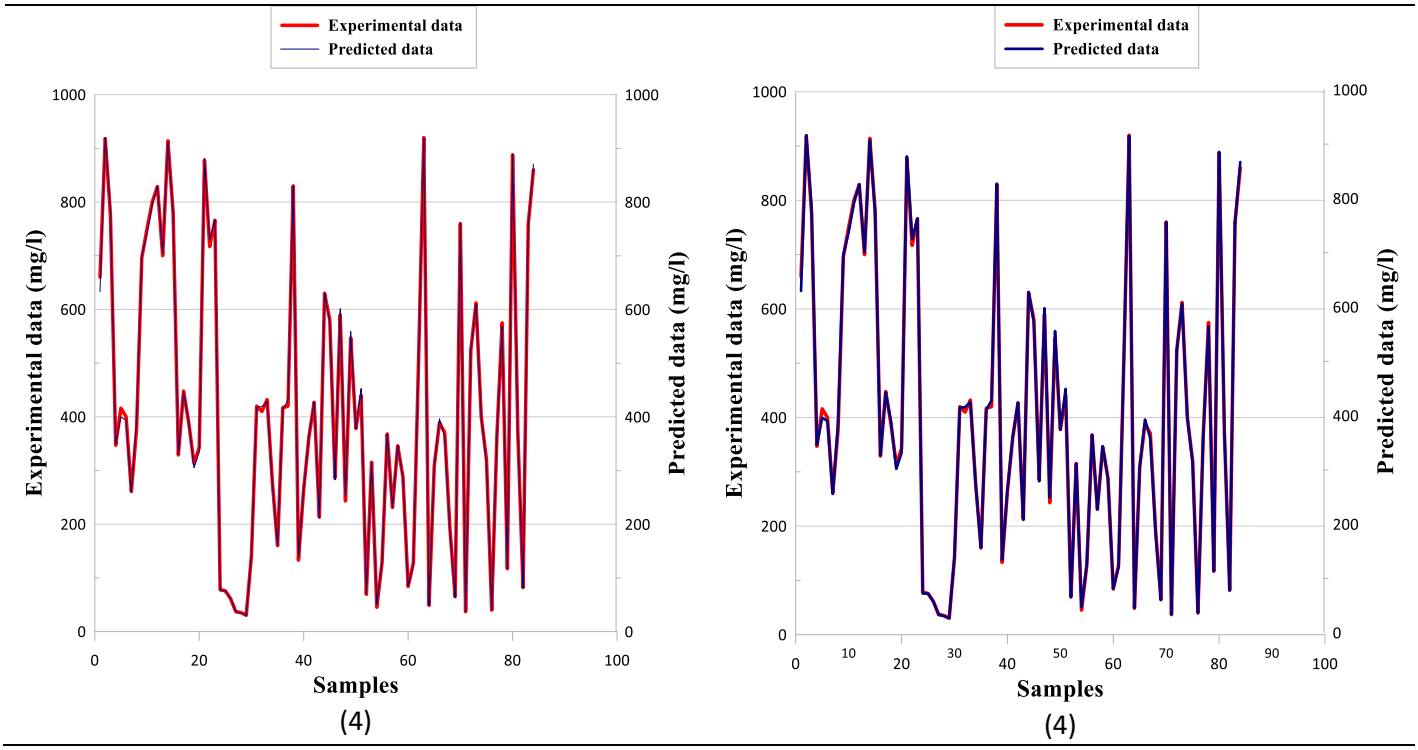


Figure 9: Relationship between experimental data and the predicted data of samples using multiple linear regression and artificial neural network modeling: (1) Training, (2) Test, (3) Validation and (4) All data.

3.7 Residues Study

The error made by the models established for each method on an individual of the model's construction sample is called residual [61]. Thus, the study of the relationship between the estimated levels of **the soluble sulfate content** by the mathematical models and their residuals $\left(y_{\exp} - y_{pred}\right)$ makes it possible to ensure the performance of the model and to verify empirically inter them, the validity of the hypotheses of the models.

The analytical methods to analyze the residues are mainly graphical analysis methods. Figure 10 shows the residuals related to the model established by artificial neural networks and those related to the model established by multiple linear regression based on the estimated values.

This figure shows that the residues obtained by the neural network method were less dispersed (close to zero) if compared to those obtained from the multiple linear regression.

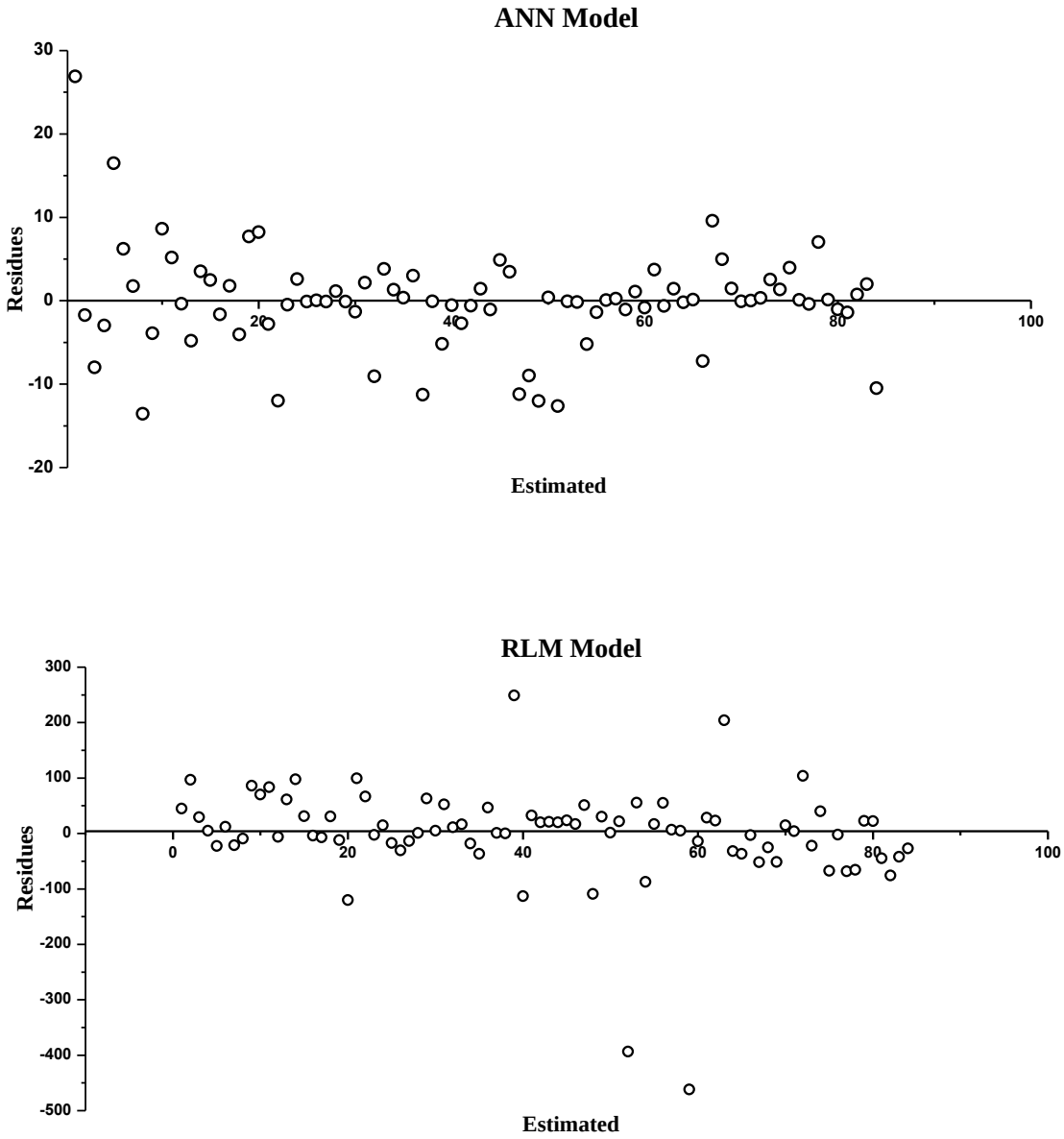


Figure 10: Residues relating to the models established by multiple linear regression and artificial neural network depending of estimated values.

In general, the results obtained were very satisfactory and justify the use of the neural network approach in the prediction of the soluble sulfate content levels of the Médéa region. It is consistent with the results of some recent studies that showed that multiple linear regression models perform poorly compared to those established by artificial neural network models [48,54,62].

4. Conclusion

This document highlighted the importance of drinking water quality modelling based on the soluble sulfate content in drinking water parameter. In this study, in a first stage, samples were collected at different points in the Medea region, Algeria. This was followed by a second stage, which was devoted to

the analysis of these samples in the laboratory in order to have the physico-chemical parameters (sulfate, conductivity, turbidity, potential hydrogen, hardness, calcium, magnesium, chlorides, total alkali metric titre (TAC), material Organic, Nitrogen dioxide, nitrates, sodium, bicarbonate, potassium, heavy metals (Mn2+, Fe3+ and Al+) and dry residues). Then the database was constructed; statistical analysis was applied on physico-chemical parameters for the prediction of soluble sulfate content in drinking water in the Medea region using two modelling methods: multiple linear regression and neural networks for prediction of the soluble sulfate content in drinking water in order to compare between them. In addition, the root mean square error (RMSE) and the Mean Absolute Error (MAE) were used to evaluate the effectiveness of these two models.

The results showed that the predictive model based on artificial neural networks of configuration [18-8-1] with hyperbolic tangent transfer functions in the hidden layer and in the output layer and the Levenberg-type learning algorithm Marquardt were better than those established by multiple linear regression. Indeed, the correlation coefficient of the ANN model was very high (R = 0. 99973) compared to the MLR model. Also, the root mean square error , the root mean square error and the mean absolute error of the ANN model were less than those established by multiple linear regression, 5.9755 and 88.3068 mg/l for RMSE, 35.7067 and 7798.0935 mg/l for MSE, and 3.1767 and 42.3274 mg/l for MAE, respectively..

the ANN model was re-tested considering two databases for interpolation and extrapolation to evaluate its efficiency. The results of the interpolation and extrapolation showed high efficiency, owing to their very high correlation coefficient and low values of the root mean square error (RMSE), the mean square error (MSE=) and the Mean Absolute Error (MAE). Consequently, the effectiveness of our model is confirmed. The results of for interpolation and extrapolation were as follows: correlation coefficient R=0. 0.99918 and R=0.9989, the root mean square error (RMSE=9.3595 mg/l and 12.2587 mg/l), the mean square error (MSE= 87.6019 mg/l and 150.2771 mg/l) and the Mean Absolute Error (MAE= 6.7598 mg/l and 10.9054 mg/l), respectively.

This performance seems to be due to the fact that the soluble sulfate content was linked to the physico-chemical characteristics of the environment with non-linear relationships. In addition, the residue graphs showed the power of neural networks in the modelling of the data.

References

[1] X. Ding, Q. Zhu, A. Zhai, L. Liu, Water quality safety prediction model for drinking water source areas in Three Gorges Reservoir and its application, Ecol. Indic. 101 (2019) 734–741.

[2] K.W. Abdelmalik, Role of statistical remote sensing for Inland water quality parameters prediction, Egypt. J. Remote Sens. Space Sci. 21 (2018) 193–200.

[3] D.J. Booker, R.A. Woods, Comparing and combining physically-based and empirically-based approaches for estimating the hydrology of ungauged catchments, J. Hydrol. 508 (2014) 227–239.

[4] U. Seeboonruang, A statistical assessment of the impact of land uses on surface water quality indexes, J. Environ. Manage. 101 (2012) 134–142.

[5] M. Gevrey, L. Comte, D. de Zwart, E. de Deckere, S. Lek, Modeling the chemical and toxic water status of the Scheldt basin (Belgium), using aquatic invertebrate assemblages and an advanced modeling method, Environ. Pollut. 158 (2010) 3209–3218.

[6] X. Xin, K. Li, B. Finlayson, W. Yin, Evaluation, prediction, and protection of water quality in Danjiangkou Reservoir, China, Water Sci. Eng. 8 (2015) 30–39.

[7] H. Runtti, S. Tuomikoski, T. Kangas, T. Kuokkanen, J. Rämö, U. Lassi, Sulphate Removal from Water by Carbon Residue from Biomass Gasification: Effect of Chemical Modification Methods on Sulphate Removal Efficiency, BioResources. 11 (2016) 3136–3152.

[8] C. Koschmann, A.-A. Calinescu, F.J. Nunez, A. Mackay, J. Fazal-Salom, D. Thomas, F. Mendez, N. Kamran, M. Dzaman, L. Mulpuri, ATRX loss promotes tumor growth and impairs nonhomologous end joining DNA repair in glioma, Sci. Transl. Med. 8 (2016) 328ra28-328ra28.

[9] D. Guimarães, V.A. Leão, Batch and fixed-bed assessment of sulphate removal by the weak base ion exchange resin Amberlyst A21, J. Hazard. Mater. 280 (2014) 209–215.

[10] W. Chen, R. Zheng, P.D. Baade, S. Zhang, H. Zeng, F. Bray, A. Jemal, X.Q. Yu, J. He, Cancer statistics in China, 2015, CA. Cancer J. Clin. 66 (2016) 115–132.

[11] A. El Hmaidi, H. El Badaoui, A. Abdallaoui, B. El Moumni, Application des réseaux de neurones artificiels de type PMC pour la prédiction des teneurs en carbone organique dans les dépôts du quaternaire terminal de la mer d'Alboran, Eur. J. Sci. Res. 107 (2013) 400–413.

[12] T. Rajaee, S. Khani, M. Ravansalar, Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review, Chemom. Intell. Lab. Syst. (2020) 103978.

[13] W. Deng, G. Wang, X. Zhang, A novel hybrid water quality time series prediction method based on cloud model and fuzzy forecasting, Chemom. Intell. Lab. Syst. 149 (2015) 39–49.

[14] A.H. Haghiabi, A.H. Nasrolahi, A. Parsaie, Water quality prediction using machine learning methods, Water Qual. Res. J. 53 (2018) 3–13.

[15] Z. Li, F. Peng, B. Niu, G. Li, J. Wu, Z. Miao, Water quality prediction model combining sparse auto-encoder and LSTM network, IFAC-Pap. 51 (2018) 831–836.

[16] H. Lu, X. Ma, Hybrid decision tree-based machine learning models for short-term water quality prediction, Chemosphere. 249 (2020) 126169.

[17] H.J. Mayfield, E. Bertone, C. Smith, O. Sahin, Use of a structure aware discretisation algorithm for Bayesian networks applied to water quality predictions, Math. Comput. Simul. 175 (2020) 192–201.

[18] A.N. Ahmed, F.B. Othman, H.A. Afan, R.K. Ibrahim, C.M. Fai, M.S. Hossain, M. Ehteram, A. Elshafie, Machine learning methods for better water quality prediction, J. Hydrol. 578 (2019) 124084.

[19] D. Wu, H. Wang, R. Seidu, Smart data driven quality prediction for urban water source management, Future Gener. Comput. Syst. 107 (2020) 418–432.

[20] K. Chen, H. Chen, C. Zhou, Y. Huang, X. Qi, R. Shen, F. Liu, M. Zuo, X. Zou, J. Wang, Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data, Water Res. 171 (2020) 115454.

[21] R. Avila, B. Horn, E. Moriarty, R. Hodson, E. Moltchanova, Evaluating statistical model performance in water quality prediction, J. Environ. Manage. 206 (2018) 910–919.

[22] E. Farahani, M.R. Mosaddeghi, A.A. Mahboubi, A.R. Dexter, Prediction of soil hard-setting and physical quality using water retention data, Geoderma. 338 (2019) 343–354.

[23] M. Khadr, M. Elshemy, Data-driven modeling for water quality prediction case study: The drains system associated with Manzala Lake, Egypt, Ain Shams Eng. J. 8 (2017) 549–557.

[24] H. Ousmana, A.E. Hmaidi, M. Berrada, B. Damnati, I. Etabaai, A. Essahlaoui, Development of a Neural Network approach for Predicting nitrate and sulfate concentration in three lakes: Ifrah, Iffer and Afourgagh, Middle Atlas Morocco, (2018) 10.

[25] L.-M.L. He, Z.-L. He, Water quality prediction of marine recreational beaches receiving watershed baseflow and stormwater runoff in southern California, USA, Water Res. 42 (2008) 2563–2573.

[26] W.-C. Liu, W.-B. Chen, Prediction of water temperature in a subtropical subalpine lake using an artificial neural network and three-dimensional circulation models, Comput. Geosci. 45 (2012) 13–25.

[27] A. Clementking, C.J. Venkateswaran, Prediction of Water Quality Attributes Variations Using Back Propagation Neural Network (BPNN) Model, in: Int. Conf. Technol. Bus. Manag., 2015: pp. 128–138.
[28] M. Heydari, E. Olyaie, H. Mohebzadeh, Ö. Kisi, Development of a neural network technique for prediction of water quality parameters in the Delaware River, Pennsylvania, Middle-East J. Sci. Res. 13 (2013) 1367–1376.
[29] H. Banejad, E. Olyaie, Application of an Artificial Neural Network Model to Rivers Water Quality Indexes Prediction—A Case Study, J. Am. Sci. 7 (2011) 60–65.
[30] Y.R. Ding, Y.J. Cai, P.D. Sun, B. Chen, The use of combined neural networks and genetic algorithms for prediction of river water quality, J. Appl. Res. Technol. 12 (2014) 493–499.
[31] N.S. Jaddi, S. Abdullah, A cooperative-competitive master-slave global-best harmony search for ANN optimization and water-quality prediction, Appl. Soft Comput. 51 (2017) 209–224.
[32] A. Beucher, R. Siemssen, S. Fröjdö, P. Österholm, A. Martinkauppi, P. Edén, Artificial neural network for mapping and characterization of acid sulfate soils: Application to Sirppujoki River catchment, southwestern Finland, Geoderma. 247 (2015) 38–50.
[33] N. Noori, L. Kalin, S. Isik, Water Quality Prediction Using SWAT-ANN Coupled Approach, J. Hydrol. (2020) 125220.
[34] S.S. Panda, V. Garg, I. Chaubey, Artificial neural networks application in lake water quality estimation using satellite imagery, J. Environ. Inform. 4 (2004) 65–74.
[35] H.Z. Abyaneh, Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters, J. Environ. Health Sci. Eng. 12 (2014) 40.
[36] J.-P. Suen, J.W. Eheart, Evaluation of neural networks for modeling nitrate concentrations in rivers, J. Water Resour. Plan. Manag. 129 (2003) 505–510.
[37] K. Ostad-Ali-Askari, M. Shayannejad, H. Ghorbanizadeh-Kharazi, Artificial neural network for modeling nitrate pollution of groundwater in marginal area of Zayandeh-rood River, Isfahan, Iran, KSCE J. Civ. Eng. 21 (2017) 134–140.
[38] S. Azimi, M.A. Moghaddam, S.H. Monfared, Prediction of annual drinking water quality reduction based on Groundwater Resource Index using the artificial neural network and fuzzy clustering, J. Contam. Hydrol. 220 (2019) 6–17.
[39] F. Qaderi, E. Babanezhad, Prediction of the groundwater remediation costs for drinking use based on quality of water resource, using artificial neural network, J. Clean. Prod. 161 (2017) 840–849.
[40] M.J. Diamantopoulou, D.M. Papamichail, V.Z. Antonopoulos, The use of a neural network technique for the prediction of water quality parameters, Oper. Res. 5 (2005) 115–125.
[41] J. Rodier, B. Legube, N. Merlet, R. Brunet, L'analyse de l'eau - 9e éd.: Eaux naturelles, eaux résiduaires, eau de mer, Dunod, 2009. https://books.google.dz/books?id=qUEGsUBZkL0C.
[42] D. Jamin, Recherche du boson de Higgs du Modèle Standard dans le canal de désintégration ZH-> nu nu bb sur le collisionneur Tevatron dans l'expérience D0. Développement d'une méthode d'étiquetage des jets de quark b avec des muons de basses impulsions transverses, 2010.
[43] M. Naoual, A. Abdelaziz, E.H. Abdellah, Use of Artificial Neural Networks Type MLP for the Prediction of Phosphorus Level from the Physicochemical Parameters of Sediments, (n.d.) 10.
[44] A. Schmitt, B. Le Blanc, M.-M. Corsini, C. Lafond, J. Bruzek, Les réseaux de neurones artificiels. Un outil de traitement de données prometteur pour l'anthropologie, Bull. Mém. Société D'Anthropologie Paris. (2001).
[45] S. Huo, Z. He, J. Su, B. Xi, C. Zhu, Using artificial neural network models for eutrophication prediction, Procedia Environ. Sci. 18 (2013) 310–316.

[46] N.D. Kaushika, R.K. Tomar, S.C. Kaushik, Artificial neural network model based on interrelationship of direct, diffuse and global solar radiations, Sol. Energy. 103 (2014) 327–342.
[47] K.D. Fausch, C.L. Hawkes, M.G. Parsons, Models that predict standing crop of stream fish from habitat variables: 1950-85., Gen Tech Rep PNW-GTR-213 Portland US Dep. Agric. For. Serv. Pac. Northwest Res. Stn. 52 P. 213 (1988).
[48] H. El Badaoui, A. Abdallaoui, I. Manssouri, L. Lancelot, Elaboration de modèles mathématiques stochastiques pour la prédiction des teneurs en métaux lourds des eaux superficielles en utilisant les réseaux de neurones artificiels et la régression linéaire multiple, J. Hydrocarb. Mines Environ. Res. 3 (2012) 31–36.
[49] E.M. Brakni, Réseaux de neurones artificiels appliqués à la méthode électromagnétique transitoire InfiniTEM, Université du Québec en Abitibi-Témiscamingue, 2011.
[50] R.P. Lippmann, Anintroduction to computing with neural nets, IEEE Assp Mag. 4 (1987) 4–22.
[51] N. Samani, M. Gohari-Moghadam, A.A. Safavi, A simple neural network model for the determination of aquifer parameters, J. Hydrol. 340 (2007) 1–11.
[52] M. Sediri, S. Hanini, H. Cherifi, M. Laidi, S.A. Turki, Dynamic Adsorption Modelling of P-nitrophenol in Aquous Solution Using Artificial Neural Network, (2017) 6.
[53] R. Yacef, A. Mellit, S. Belaid, Z. Şen, New combined models for estimating daily global solar radiation from measured air temperature in semi-arid climates: application in Ghardaïa, Algeria, Energy Convers. Manag. 79 (2014) 606–615.
[54] H. El Badaoui, A. Abdallaoui, L. Lancelot, Application des réseaux de neurones artificiels et des régressions linéaires multiples pour la prédiction des concentrations des métaux lourds dans les sédiments fluviaux marocains, Univ. PoitiersFrance. (2012).
[55] D.A. Belsley, E. Kuh, R.E. Welsch, Regression diagnostics: identifying influential data and sources of collinearity, Wiley, 1980. https://books.google.dz/books?id=ALjuAAAAMAAJ.
[56] S.H. Hong, M.W. Lee, D.S. Lee, J.M. Park, Monitoring of sequencing batch reactor for nitrogen and phosphorus removal using neural networks, Biochem. Eng. J. 35 (2007) 365–370.
[57] I. Manssouri, M. Manssouri, B. El Kihel, FAULT DETECTION BY K-NN ALGORITHM AND MLP NEURAL NETWORKS IN A DISTILLATION COLUMN: COMPARATIVE STUDY, J. Inf. Intell. Knowl. 3 (2011) 201.
[58] I. Manssouri, A. El Hmaidi, T.E. Manssouri, B. El Moumni, Prediction levels of heavy metals (Zn, Cu and Mn) in current Holocene deposits of the eastern part of the Mediterranean Moroccan margin (Alboran Sea), IOSR J. Comput. Eng. 16 (2014) 117–123.
[59] O.R. Dolling, E.A. Varas, Artificial neural networks for streamflow prediction, J. Hydraul. Res. 40 (2002) 547–554.
[60] S. Lefnaoui, N. Moulai-Mostefa, Investigation and optimization of formulation factors of a hydrogel network based on kappa carrageenan–pregelatinized starch blend using an experimental design, Colloids Surf. Physicochem. Eng. Asp. 458 (2014) 117–125.
[61] C. Voyant, M. Muselli, C. Paoli, M.-L. Nivet, Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation, Energy. 36 (2011) 348–359.
[62] M. Bélanger, N. El-Jabi, D. Caissie, F. Ashkar, J. Ribi, Estimation de la température de l'eau de rivière en utilisant les réseaux de neurones et la régression linéaire multiple, Rev. Sci. EauJournal Water Sci. 18 (2005) 403–421.

Optimisation and Prediction of the Coagulant Dose for the Elimination of Organic Micropollutants Based on Turbidity

Hichem Tahraoui,[a*] *Abd-Elmouneïm Belhadj*,[a] *Nassim Moula*,[b] *Saliha Bouranene*,[c] and *Abdeltif Amrane*[d]

[a] Laboratory of Biomaterials and Transport Phenomena (LBMPT), University of Médéa, Algeria, Nouveau Pôle Urbain, Médéa University, 26 000 Médéa, Algeria

[b] Fundamental and Applied Research in Animal and Health (FARAH) Department of Veterinary Management of Animal Resources, Faculty of Veterinary Medicine, University of Liege, Liege 4000, Belgium

[c] University of Souk Ahras, Dept. Process Engineering, STEE Lab., Rue d'annaba, BP 1553, 41000 Souk-Ahras, Algeria

[d] Univ Rennes, Ecole Nationale Supérieure de Chimie de Rennes, CNRS, ISCR – UMR6226, F-35 000 Rennes, France

[*] Corresponding Author: Hichem Tahraoui, tahraoui.hichem@univ-medea.dz

Abstract

In this study, four different mathematical models were considered to predict the coagulant dose in view of turbidity removal: response surface methodology (RSM), artificial neural networks (ANN), support vector machine (SVM), and adaptive neuro-fuzzy inference system (ANFIS). The results showed that all models accurately fitted the experimental data, even if the ANN model was slightly above the other models. The SVM model led to almost similar results as the ANN model; the only difference was in the validation phase, since the correlation coefficient was very high and the statistical indicators were very low for the ANN model compared to the SVM model. However, from an economic point of view, the SVM model was more appropriate than the ANN model, since its number of parameters was 22, *i.e.* almost half the number of parameters of the ANN model (43 parameters), while the results were almost similar in all the data phase. To further reduce the economic costs, the RSM model can also be used which remained very useful due to its high coefficients related to the number of parameters – only 13. In addition, the statistical indicators of the RSM model remained acceptable.

Keywords

Coagulation, physicochemical analysis, response surface methodology, artificial neural networks, support vector machine, adaptive neuro-fuzzy inference system

1. Introduction

Water treatment is very essential owing to the unavailability of drinking water and its lack in developing countries in Africa, which is mainly due to human and industrial pollution. Water treatment requires specific attention to be in agreement with the standards of consumption.[1] It includes several methods such as coagulation-flocculation, sedimentation or flotation, filtration and disinfection, *etc*.[2] Among these processes, coagulation-flocculation has been classified as the best available technology for large-scale drinking water production.[3] It has played and will continue to play an important role, directly or indirectly, in the control of particles, microorganisms, natural organic matter (NOM), synthetic organic carbon, disinfection by-product precursors (DBP), and certain inorganic ions and metals, and finally, in the control of drinking water quality.[4] Coagulation is a process of destabilisation of particles in water by neutralising the negative charge on the particle surface, followed by agglomeration of the neutralised micro-flakes to form settleable and filterable macro-flocculants.[5,6] The coagulation process is primarily used to remove turbidity in a water treatment plant.[7] It is used in more than half of the water purification plants in Algeria. Coagulation efficiency depends on the water characteristics, such as[8] colour, turbidity, temperature, pH, alkalinity, dissolved salts without forgetting the influence of mineralization. It also depends on the treatment itself, such as the nature and dosage of coagulant, the conditions of agitation, as well as on the treatment conditions.[9] Jar tests are generally conducted to determine optimal coagulant doses,[10,11] with the goal of reducing costly chemical wastage and operating costs and achieving drinking water quality objectives.[12] However, to date, this process poses a difficult control problem because it is non-linear and complex. Several studies have been conducted to apply different control strategies to the coagulation process.[13–15] In recent years, mathematical models have emerged as a viable method to model complex processes of water treatment.[16] These models have been very successful in modelling and predicting environmental parameters and in predicting coagulant quantity as a function of many physicochemical parameters.[17] For example, a study was focused on the use of *Moringa oleifera* seed as a coagulant for the treatment of surface water using response surface methodology (RSM).[18] Four parameters were varied *viz.* stabilization time, stirring time, stirring speed, and *Moringa oleifera* seed extract (MOSE) concentration. The model predicts the lowest turbidity of 5.49 NTU with optimal conditions of 120 min settling time, stirring speed of 100 rpm, 10 min stirring time, and 3 $g\,l^{-1}$ concentration by MOSE. The condition was verified in aftershocks and a turbidity of 5.51 NTU was obtained.[18]

Other work reported the effectiveness of three coagulants $CuSO_4$, $FeCl_3$, and $CuSO_4$ + $FeCl_3$ (the two salts mixed in 1 : 1 ratio (v/v)) for the treatment of wastewater from the effluent treatment plant of an oil refinery by coagulation and flocculation process.[19] Independent parameters such as pH and coagulant dosage were optimised using the response surface methodology with central composite design technique considering final pH, COD reduction, turbidity, TDS, and colour as dependent variables.[19] Results showed that the mixed coagulant $CuSO_4$ + $FeCl_3$ gave better results than $CuSO_4$ and $FeCl_3$ individually. $CuSO_4$ + $FeCl_3$ along with adsorption and sweep flocculation, forms tribasic copper chloride (TBCC) as an intermediate in the pH range 5 to 7. TBCC formation improved flocculation by destabilising colloidal and suspended particles due to its structural catalytic, bleaching, and octahedral properties. A maximum reduction in COD (76.77 %), turbidity (89.47 %), TDS (94.16 %) and colour (95.29 %) were observed at pH 7.12 and dosage 0.20 g L^{-1} of $CuSO_4$ + $FeCl_3$ coagulant. The final pH of the solution under optimal conditions for all three coagulants was below the removal limits.[19]

Another application by the response surface method (RSM) was to find the optimal combination of coagulant dose and pH relative to the highest removal efficiency of turbidity and dissolved organic carbon (COD).[20] The results obtained with polyaluminum chloride (PACl) were compared with those obtained using a conventional coagulant such as alum. Quadratic models developed for both responses (turbidity removal and DOC removal) indicated that the optimal conditions for a PAC1 concentration of 0.11 mM at pH 7.4 and alum concentration of 0.15 mM at pH 6.6. The compromise to optimise both responses simultaneously resulted in 91.4 % turbidity removal and 31.2 % DOC removal using PACl while 86.3 % turbidity and 34.3 % DOC. CODs were removed using alum.[20]

Another study was done to predict both turbidity and removal of dissolved organic matter (DOM) during the coagulation process at Akron Water Treatment Plant (Akron, Ohio, USA) by four different neural network models.[21] DOM was monitored and characterised using fluorescence spectroscopy and parallel factor analysis (PARAFAC), building on previous research that identified three fluorescence components (C1, C2, and C3). Neural network models were constructed using operational data to predict each of the components of fluorescence and turbidity after coagulation as a function of variable raw water quality and chemical doses. Correlation coefficients between measured and model predicted values for final turbidity models, C1, C2 and C3 on an invisible test data set were 0.91, 0.95, 0.97, and 0.51, respectively.[21]

Other researchers used extreme machine learning (ELM) coupled with radial base function (RBF) neural networks to predict coagulation doses.[22] The coagulation data was divided into two categories based on low and high turbidity. The optimal number of input parameters for low turbidity water coagulation modelling was found to be 3, while the optimal number of input parameters for turbidity water coagulation modelling were found to be 4. Re-selection of the number of input parameters were necessary, since the alkalinity of the raw water was an important factor in improving the performance of the high turbidity model. The low turbidity model was able to predict coagulant dosage with a correlation coefficient greater than 0.97.[22] The high turbidity model was able to predict coagulation dosage with a reasonably acceptable correlation coefficient of at least 0.80.[22]

Another study was done by a hybrid of *k*-signifies an adaptive neuro-fuzzy inference system (*k*means-ANFIS) for the turbidity of settled water prediction and optimal determination of the coagulant dose using historical data at large scale.[23] To construct a well adaptive model to different states of inflow water process, raw water quality data was classified into four groups according to its properties by a *k*-means clustering technique. The sub-models are developed individually based on each clustered data set. The results revealed that the sub-models constructed by a *k*-means-ANFIS hybrid perform better that not only a single ANFIS model, but also seasonal artificial neural network (ANN) models. The finally completed model composed of sub-models shows more precise and consistent prediction capacity than a single ANFIS model and a single ANN model based on all five evaluation indices.[23]

In another report, they used artificial neural network (ANN) and adaptive network-based fuzzy inference system (ANFIS) models were used to model the dosage of polyaluminum chloride (PAC), of surface water in northern Taiwan.[24] Each of them was built on the basis of 819 data sets controlled by the processes. Input parameters include yesterday's PAC dosage, the day before yesterday's PAC dosage, and temperature, turbidity, colour, pH in each of raw water, flocculation, sedimentation, and treated water.[24] The ANN model was better than the ANFIS model to be used to achieve the optimal prediction model for the optimal dosage of PAC in real time when stormwater brought high turbidity to the source water.[24]

In this work, response surface methodology (RSM), artificial neural networks (ANN), machine vector support (SVM), and adaptive neuro fuzzy inference system (ANFIS) were selected to predict coagulant doses as a function of physical and chemical parameters and to compare them to select the most efficient one. Indeed, to our knowledge, the ANN, SVM, and ANFIS models have never been used in this kind of data-based study built from RSM. In addition, such a comparison has not been made before.

2. Materials and methods

2.1. Data preparation

2.1.1. Jar test

Aluminum salt was used in this study as coagulant. The jar tests were carried out using beakers that were filled with 1 L of water, namely raw waters collected in different localities of the Médéa region, with various turbidity levels which were adjusted by the humic acid. Considering that acidic and oligotrophic waters often have a brown to yellow "tea" tint,[25] in particular due to humic acids (humic acids are one of the most important fractions of humus, thanks to their hydrophilic functions of the carboxylic acid type), they can retain about fifteen times their weight in water and thus play a fundamental role in water retention and the useful water reserve of a soil.[26]

Thus, after having created turbidity by humic acid, the water takes a "tea" hue, and without forgetting that the turbidity contains various mineral and organic matter in suspension or in solution. Dissolved and colloidal matter alone constitutes 60 to 80 % of the organic load (humic acids). If we manage to remove turbidity, we can say that we have removed organic and mineral matter (complexation),[27–29] because the removal of the colour created by humic acid means the removal of organic and dissolved colloidal organics and micropollutants.[30]

The samples were mixed at 180 rpm for 2 min after addition of coagulant to provide mixing; then the speed was reduced to 40 rpm for 30 min to ensure flocculation[31]. The time and speed of mixing were set with an automatic controller. After 45 min of settling, the supernatant from each pot was removed from the port sample and analysed for pH, temperature, turbidity, conductivity, and total alkali metric titre (TAC). pH, conductivity, temperature, and turbidity were measured by means of pH meter, conductivity meter, thermometer, and turbidimeter, respectively. TAC was analysed immediately following the assay method described by *J. Rodier*.[32]

2.2 Database

The data used in this study were those obtained from the analysis of samples purified with different doses of coagulant that were collected during several campaigns conducted in the Médéa region. The database that was built by RSM was used for the other models (ANN, SVM, and ANFIS), in order to be able to compare them.

The dependent variable was the coagulant dose. The independent variables were the physicochemical parameters: pH, temperature, conductivity, turbidity, and total alkalinity titre (TAC).

2.3 Prediction methods

Several methods have been applied to solve the problems related to the prediction and modelling of complex non-linear systems. These methods are particularly useful when such systems are difficult to model using the conventional methods.[33] In this study, four methods were tested for predicting coagulant doses from physicochemical water parameters, namely RSM, ANN, SVM, and ANFIS.

After constructing the database according to RSM, the database was normalised in the interval [−1,+1] according to the RSM method by using the *mapminmax* function (Eq. 1) of MATLAB software.[34] It was also divided in two parts for the three other models (ANN, SVM, ANFIS): 70 % of the dataset for training and the remaining 30 % of the samples, which did not participate in model learning, were used for validation and prediction performances of the models:[35,36]

$$x_N = (y_{\max} - y_{\min})\left(\frac{x - x_{\min}}{x_{\max} - x_{\min}}\right) + y_{\min} \quad (1)$$

where x_N is the data value after normalization, $x_{\max}$ and $x_{\min}$ denote the maximum and the minimum of the data, respectively; $y_{\max}$ and $y_{\min}$ are taken as −1 and 1; x represents the original values.

The correlation coefficient (R), determination coefficient (R^2), adjusted coefficient (R^2_{adj}), root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), error standard prediction (ESP), error prediction model (EPM) were used to estimate the performances of each model [36–41]

2.3.1 RSM

The principle of RSM has been described by *Khuri* and *Cornell* (1996) as a set of mathematical and statistical methods for evaluating the relationships between a group of independent variables and one or more responses.[42] To obtain adequate and reliable measures of the responses of interest, the design of the experiment is necessary. Normally, the relationship between the response and the independent variables cannot be well modelled by a linear function.[20]

A model that incorporates curvature is generally necessary to approximate the response in the region close to optimal, and in most cases a second-order model is adequate.[43] A central composite design (CCD) is considered, which is a very effective design tool for fitting second-order models.[43] This technique is generally used when there are several input parameters affecting the output (response). CCD was selected for use in this study.[43] Two levels, a value of a = 1 and 5 repetitions at the centre point were considered. Finally, the input variables were coded in intervals [−1, +1] (Table 1).

The number of experiments is determined according to the Eq. (2):[42]

$$N = 2^k + 2k + c \tag{2}$$

where k is the number of independent variables, 2^k is the points factor, 2k is the axial points, and c is the number of experiments in the centre points.[44] The model used to predict the responses and to describe the relationship between the independent variables was the second-order polynomial equation:[42]

$$Y = \mathrm{B}_0 + \sum_i^k \mathrm{B}_i X_i + \sum_i^k \mathrm{B}_{ii} X^2 + \sum_{ij}^k \mathrm{B}_{ij} X_i X_j + E \tag{3}$$

where, Y represents the response functions (in our case, it is force compression): B_0 is a constant coefficient; B_i, B_{ii}, and B_{ij} are the coefficients of the linear, quadratic, and interactive terms, respectively.[45]

Table 1 – Coding and real levels for CCD model

Variables	Symbol	Unit	Code factor Levels		
			−1	0	+1
pH	X_1	–	7.25	8.02	8.78
temperature	X^2	°C	10.65	19.48	28.30
conductivity	X_3	µS cm^{-1}	106	988	1871
turbidity	X_4	NTU	0.86	519.93	1039
TAC	X_5	°F	12.00	18.20	24.4

2.3.2. Artificial neural network

In this work, the optimisation of the ANN architecture using MATLAB R2013a software is described by a flowchart shown in Fig. 1.[36,46,47]

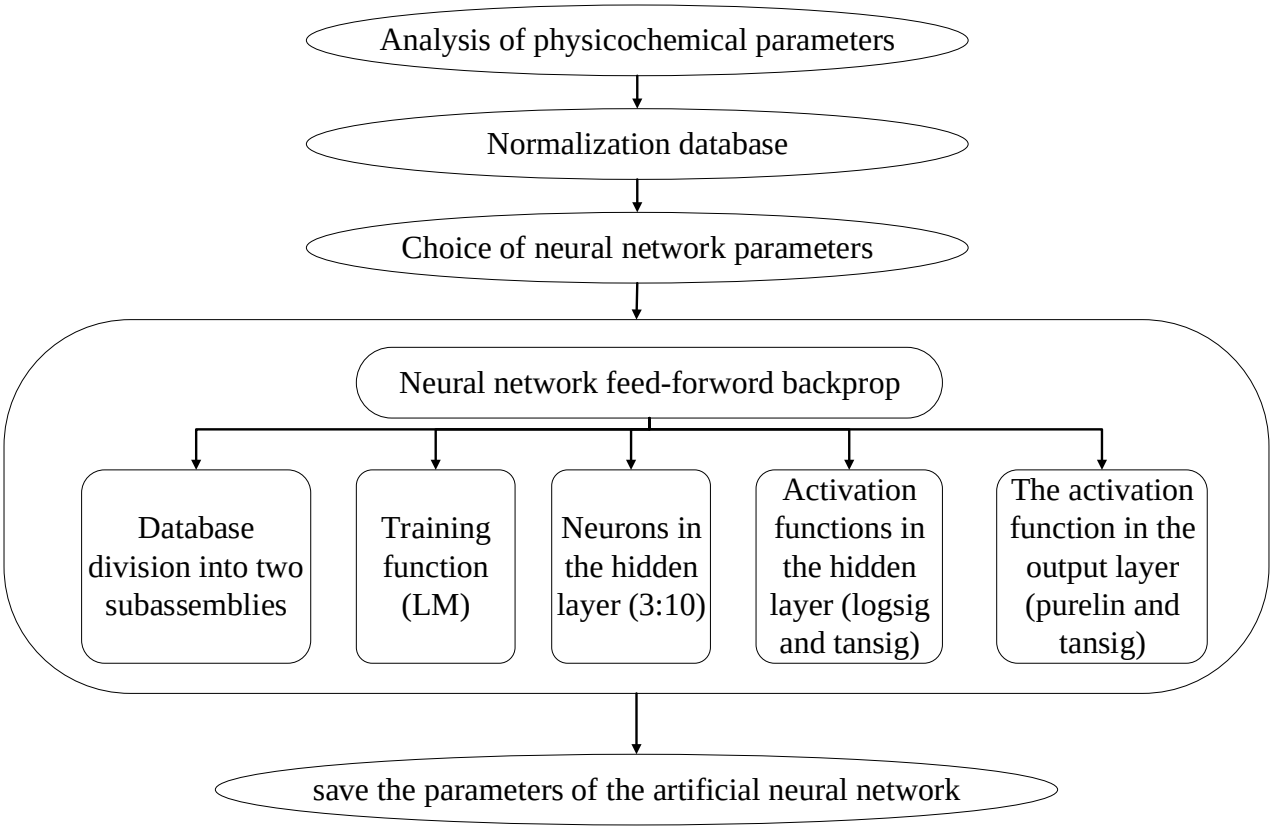


Fig. 1 – Organization chart for the development and optimisation of the ANN architecture

2.3.3 Support vector machine

Support vector machine was developed by *Vapnik* in the 1990s and is based on the statistical learning theory (SLT) tools and structural risk minimisation (SRM) theory.[48] It was mainly used for non-linear classification and regression analysis.[49] In this approach, a training data set of N points (X_i, Y_i) is considered, with $i = 1,..., N$. X represents the inputs of the model and Y its output. A SVM model takes the form:

$$y(x) = \omega^T \phi(x) + b \tag{4}$$

where $\phi(.)$: $R^n \rightarrow R^m$ is a non-linear function that maps the finite dimensional space entry into a higher dimensional space that is implicitly created, ω is a weight vector, and b is the bias.[50,51] SVM problems are solved using quadratic programming techniques, which however show some drawbacks:[52]

1. They are practically difficult to use.
2. They are time consuming.
3. They require large memory and CPU time.

In this article, SVM is used for non-linear modelling using the different functions of the kernel, since there is no guarantee that a given kernel appears more efficient than another one regarding

the specific data. From this, it is necessary to optimise different kernel functions and test their performances. These kernel functions can be found in the MATLAB Toolbox R 2018a.[53]
The kernel functions that have been selected in this article are as follows:[53]

- Poly (polynomial)

$$k\left(X_i, X_j\right) = (ax^T + c)^d \tag{5}$$

- Linear

$$k\left(X_i, X_j\right) = x^T y + c \tag{6}$$

- Gaussian

$$k\left(X_i, X_j\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-X_i - X_j^2}{2\sigma^2}\right\} \tag{7}$$

with C = BOxConstraint, ε = epsilon, d = Polynomial Order, and σ^2 = sigma
where d, c, ε, and σ^2 are user-defined kernel parameters.

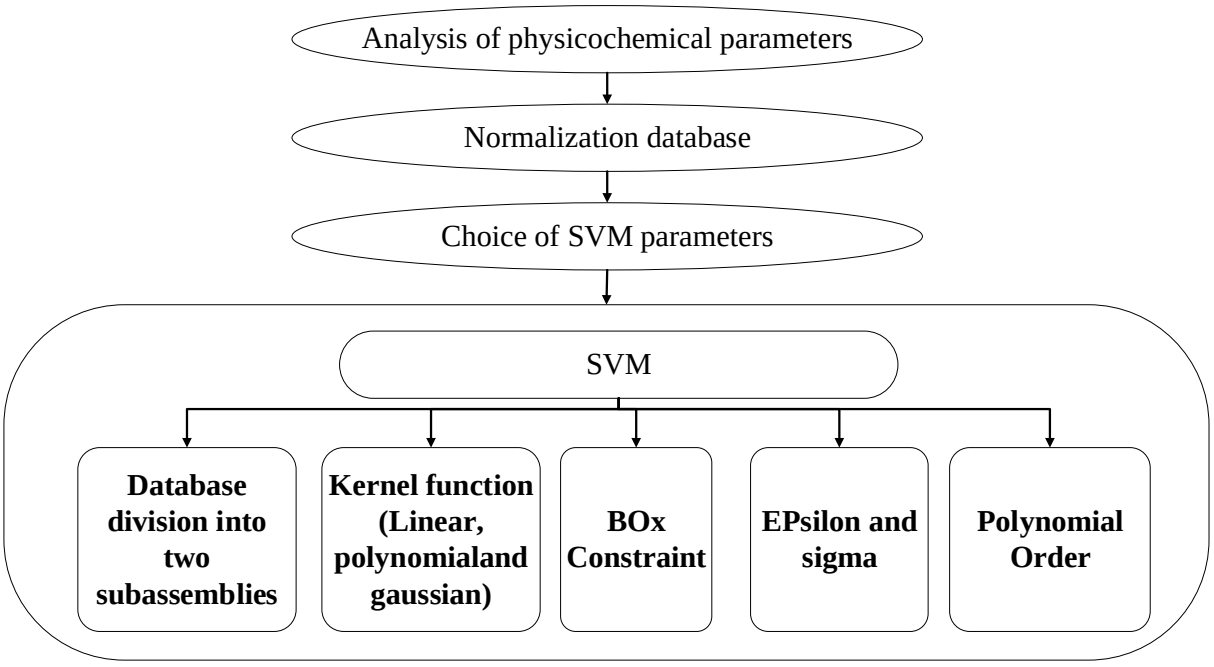


Fig. 2 – Organization chart for the development and optimisation of the SVM model

2.3.4 Adaptive Neuro Fuzzy Inference System

The adaptive ANFIS was developed by *Jang*.[54] The architecture of ANFIS is shown in Fig. 3. It consists of six key components: input and output data, data division unit, algorithm unit, fuzzy generating system, fuzzy inference system, and adaptive network neuron representing the fuzzy system. After dividing the processing data, each set of training and validation data is assigned

to the clusters formed by the algorithm.[23,24] Each algorithm is formed independently with training and validation together by ANFIS to obtain the optimal fuzzy inference system.[23] In addition, ANFIS is composed of 5 layers – each layer can include different nodes.[55]

Layer 1: Each neuron calculates the degree to which *X* and *Y* inputs belong to the different fuzzy sets (Eqs. 8 and 9) .[56]

The parameters inherent to these sets are called premise network parameters. For this system, we have:

$$X_{1,K} = \mu_{A(k)}(X) \qquad \text{for } k = 1 \text{ and } 2 \tag{8}$$

$$X_{1,K} = \mu_{B(k-2)}(Y) \qquad \text{for } k = 3 \text{ and } 4 \tag{9}$$

Layer 2: is used to calculate the degree of activation of the premises. Each neuron in this layer receives the outputs of the previous fuzzification neurons and calculates its activation. The conjunction of the antecedents is performed with the product operator, which uses the derivability constraint to deploy the learning algorithms. Each node performs a fuzzy T-norm.

$$X_{2,K} = \Pi_{(1,j)} \varepsilon Ant(2,k)(X) \tag{10}$$

or, *Ant* (2,*k*) indicating the antecedent nodes of node (2,*k*).

Layer 3: Each neuron calculates the normalized degree of truth of a given fuzzy rule. The value obtained represents the contribution of the fuzzy rule to the final result.

$$X_{3,k} = \frac{X_{2,k}}{\sum (2,j) \varepsilon Ant(2,k)^{X_{2,j}}} \tag{11}$$

Layer 4: Each i neuron of this layer is connected to a normalization neuron corresponding to the initial inputs of the network. The link in this layer fulfils the role of the consequence part of the rules; each node calculates the standard degree of truth of a given fuzzy rule. The value obtained represents the contribution of the fuzzy rule to the final result.

$$X_{4,k} = X_{3,k}\left[m_{k0} + m_{k1}.X + m_{k2}.Y\right] \tag{12}$$

where the parameters m_{k0}, m_{k1} and m_{k2}, are called neural fuzzy-system consequential parameters.

Layer 5: This is the output layer comprising a single neuron that provides the output of ANFIS by calculating the sum of the outputs of all output neurons.

$$X_{4,k} = X_{3,k}\left[m_{k0} + m_{k1}.X + m_{k2}.Y\right] \tag{13}$$

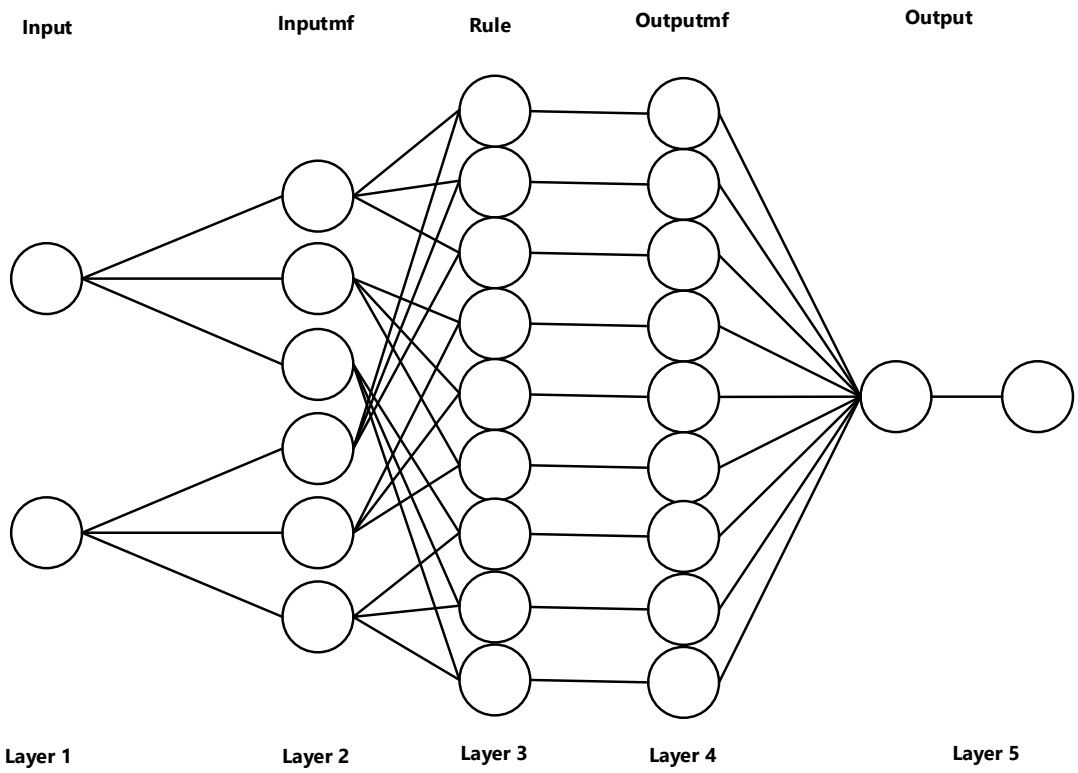


Fig. 3 – The structure of ANFIS

3 Results and discussion

RSM, ANN, SVM, and ANFIS are the four approaches that were used for the prediction of coagulant dose from physicochemical water parameters. All four methods were performed, evaluated, and compared.

3.1 RSM

A statistical analysis by RSM was performed using the "JMP 13" software on the entire database software. This method allowed finding the mathematical relationship (Eq. (14)) between the coagulant doses and the independent variables that corresponded to the physicochemical parameters. Firstly, the equation took into account all 5 variables (Table 2), even those that did not seem to have a significant impact on the dependent variable. The relationship obtained was, therefore, evaluated in order to keep only the independent variables characterised by a high probability value (Table 2).

Table 2: Values of βi, xi and characterization of independent variables of the RSM

i	Term	βi	Std Error	t Ratio	Prob > |t|
0	Intercept	29.44685	2.01046	14.64676	4.40E-08
1	X_1	−3.88888	1.58898	−2.44740	0.03441
2	X_2	−2.55555	1.58898	−1.60829	0.13884
3	X_3	−11.94440	1.58898	−7.51703	2.02E-05
4	X_4	16.38888	1.58898	10.31406	1.20E-06
5	X_5	4.16666	1.58898	2.62222	0.02550
6	X_1*X_2	5.81250	1.68537	3.44879	0.00623
7	X_1*X_3	6.68750	1.68537	3.96796	0.00265
8	X_2*X_3	−7.56250	1.68537	−4.48713	0.00116
9	X_1*X_4	2.06250	1.68537	1.22376	0.24909
10	X_2*X_4	−14.18750	1.68537	−8.41802	7.52E-06
11	X_3*X_4	−1.31250	1.68537	−0.77875	0.45415
12	X_1*X_5	4.68750	1.68537	2.78128	0.01940
13	X_2*X_5	−2.56250	1.68537	−1.52043	0.15936
14	X_3*X_5	1.56250	1.68537	0.92709	0.37571
15	X_4*X_5	6.18750	1.68537	3.67129	0.00430
16	X_1*X_1	3.02386	4.29938	0.70332	0.49790
17	X_2*X_2	−5.97613	4.29938	−1.38999	0.19469
18	X_3*X_3	20.52386	4.29938	4.77367	0.00075
19	X_4*X_4	8.52386	4.29938	1.98257	0.07554
20	X_5*X_5	−13.47613	4.29938	−3.13443	0.01061

$$Y = \beta_0 + \beta_1.X_1 + \beta_2.X_2 + \beta_3.X_3 + \beta_4.X_4 + \beta_5.X_5 + \beta_6.X_1.X_2 + \beta_7.X_1.X_3 + \beta_8.X_2.X_3 + \beta_9.X_1.X_4 + \beta_{10}.X_2.X_4 + \beta_{11}.X_3.X_4 + \beta_{12}.X_1.X_5 + \beta_{13}.X_2.X5 + \beta_{14}.X_3.X_5 + \beta_{15}.X_4.X_5 + \beta_{16}.X_1^2 + \beta_{17}.X_2^2 + \beta_{18}.X_3^2 + \beta_{19}.X_4^2 + \beta_{20}.X_5^2 \quad (14)$$

Next, only 13 independent variables that have a high explanatory power for the dependent variable were taken ($P_r < 0.05$); so the relationship can be reduced to Eq. (15):

$$Y = \beta_0 + \beta_1.X_1 + \beta_3.X_3 + \beta_4.X_4 + \beta_5.X_5 + \beta_6.X_1.X_2 + \beta_7.X_1.X_3 + \beta_8.X_2.X_3 + \beta_{10}.X_2.X_4 + \beta_{12}.X_1.X_5 + \beta_{15}.X_4.X_5 + \beta_{19}.X_4^2 + \beta_{20}.X_5^2 \quad (15)$$

The value of the coefficient of determination decreased slightly, but the equation became simpler after the removal of the variables with little explanatory power from the dependent variable. This coefficient $R = 0.98691$ means that the model's correlation was moderately positive (Fig. 5). The probability ($P_r < 0001$) was strictly less than 0.5 %, confirming that the model was significant.

The significance level value p and the F ratio value, which provide a measure of the statistical significance of the regression model, were also determined. A high value of F with a minimum value of p means that the equation is significant.[57]

The proposed model can infer the effect of the factors (X_i), their interactions and their quadratic effects simultaneously. Indeed, Table 2 shows the effects of each independent factor, its interaction with the other factors and its quadratic effect on the coagulant dose (Y).

Indeed, the coefficients of each factor in the model allow to assess the impact of each factor on the response.[58] Thus, it is evident that changes in turbidity and TAC in the treated effluent increased the dose of the coagulant (Fig. 4). On the other hand, changes in pH and conductivity had a negative effect on the dose of coagulant (Fig. 4).

The interactions and quadratic effects of the factors show that the relationship between the response and the factors is not always linear.[58] Used at different levels in the coagulation process and when several factors are modified simultaneously, one factor can produce different degrees of response. Thus, interactions between the factors (pH – temperature, pH – conductivity, pH – TAC and turbidity – TAC) showed positive effects on coagulant doses (Fig. 4), whereas the interactions between factors (temperature – conductivity and temperature – turbidity) had negative effects on coagulant doses (Fig. 4). In addition, the results showed two factors with interesting quadratic effects, namely conductivity and TAC, indicating positive and negative quadratic effects, respectively (Fig. 4).

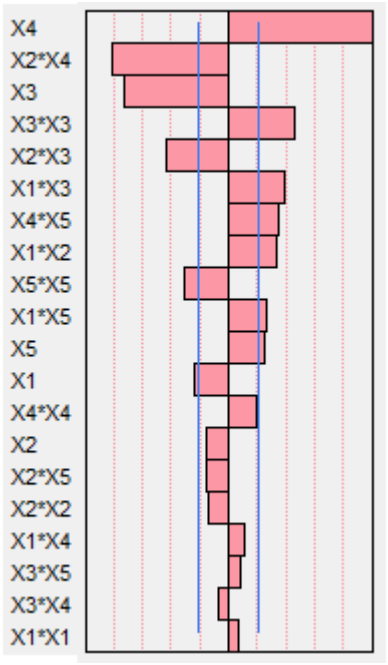


Fig. 4 – Effects of the operating factors and their interactions.

In the light of the obtained results, it appears that with each increase in pH (7.25 to 8.78) and temperature (10.65 to 28.3 °C), the dose of coagulant used for treatment decreased from

109.51 mg l^{-1} to 83.15 mg l^{-1} for pH and from 109.15 mg l^{-1} to 78.28 mg l^{-1} for temperature. The dose of coagulant also decreased (from 109.51 mg l^{-1} to 75.24 mg l^{-1}) with increasing conductivity (from 106 µS cm^{-1} to 1243.54 µS cm^{-1}) and then increased to 85.51 mg l^{-1} for 1871 µS cm^{-1}. On the other hand, the results obtained show an increase in the dose of coagulant (from 47.16 to 109.51 mg l^{-1}) when the turbidity increases (from 0.86 NTU to 1039 NTU). Finally, it can be noted that the effect of TAC on the coagulant dose was not linear. Thus, the dose of coagulant used for effluent treatment increased (from 88.55 to 109.51 mg l^{-1}) for TAC values ranging from 88.55 °F to 19.71 °F; then decreased (101.88 g l^{-1}) for 24.4 °F. These results confirm the effect of the factors in this study, namely pH, temperature, conductivity, turbidity, and TAC during the coagulation process.[8,59–61]

For an optimal chosen coagulant dose equal to 109.50 mg l^{-1}, water with the following properties: pH 7.25, temperature 10.65 °C, conductivity 106 µS cm^{-1}, turbidity 1039 NTU, and TAC = 19.71 °F provided a 99.96 % effective treatment.

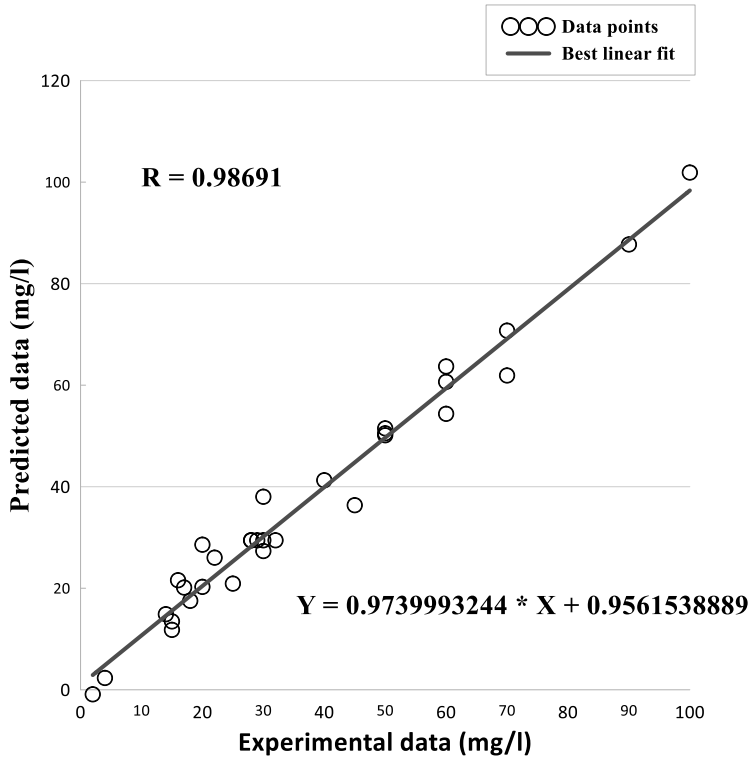


Fig. 5 – Relation between the observed dose of coagulant and those estimated by the RSM model

The results of the RSM performance in terms of all errors and in terms of agreement vector values (*R*, slope: α, and intercept: β) are given in Table 3.

Table 3 – RSM performances

	RMSE / mg l^{-1}	MSE / mg l^{-1}	ESP / %	EPM / %	MAE / mg l^{-1}	*R*	α	β
all	3.82890	14.66053	10.41194	1.47823	3.82890	0.98691	0.97399	0.95615

3.2. ANN

In this part, we have optimised 3 activation functions with the number of hidden neurons (from 3 to 10). The results of these tests are presented in Table 4.

Table 4 presents the best architectures found. It shows the correlation coefficients and the error for each learning and validation according to the number of neurons in the hidden layer and the network topology. It also shows the activation functions of the hidden layer and the output layer. Architecture 1 (Table 4) seems to be the most relevant ANN model for predicting coagulant doses. Since the correlation coefficient and RMSE were almost equal between architectures 1 and 3, the lowest number of neurons of the hidden layer was chosen in order to have the lowest possible number of parameters.

Table 4 – Performances of the different tested ANN architectures

ANN		Activation function		Coefficients of correlation	RMSE
Nbr	Neurons/Layer	Hidden	Output layer	Total	Total
1	**[5-6-1]**	**logsig**	**purelin**	**0.9975**	**1.7954**
2	[5-9-1]	tansig	purelin	0.99744	1.8132
3	[5-10-1]	tansig	tansig	0.99751	1.7908

Once the architecture of the ANN was synthesized, it was validated. The predicted values (estimated values) were compared to the experimental ones in the two stages of the learning and validation phases, and the corresponding results are given in Fig. 6. The results were denormalized to actual values in view of comparison with the other models.

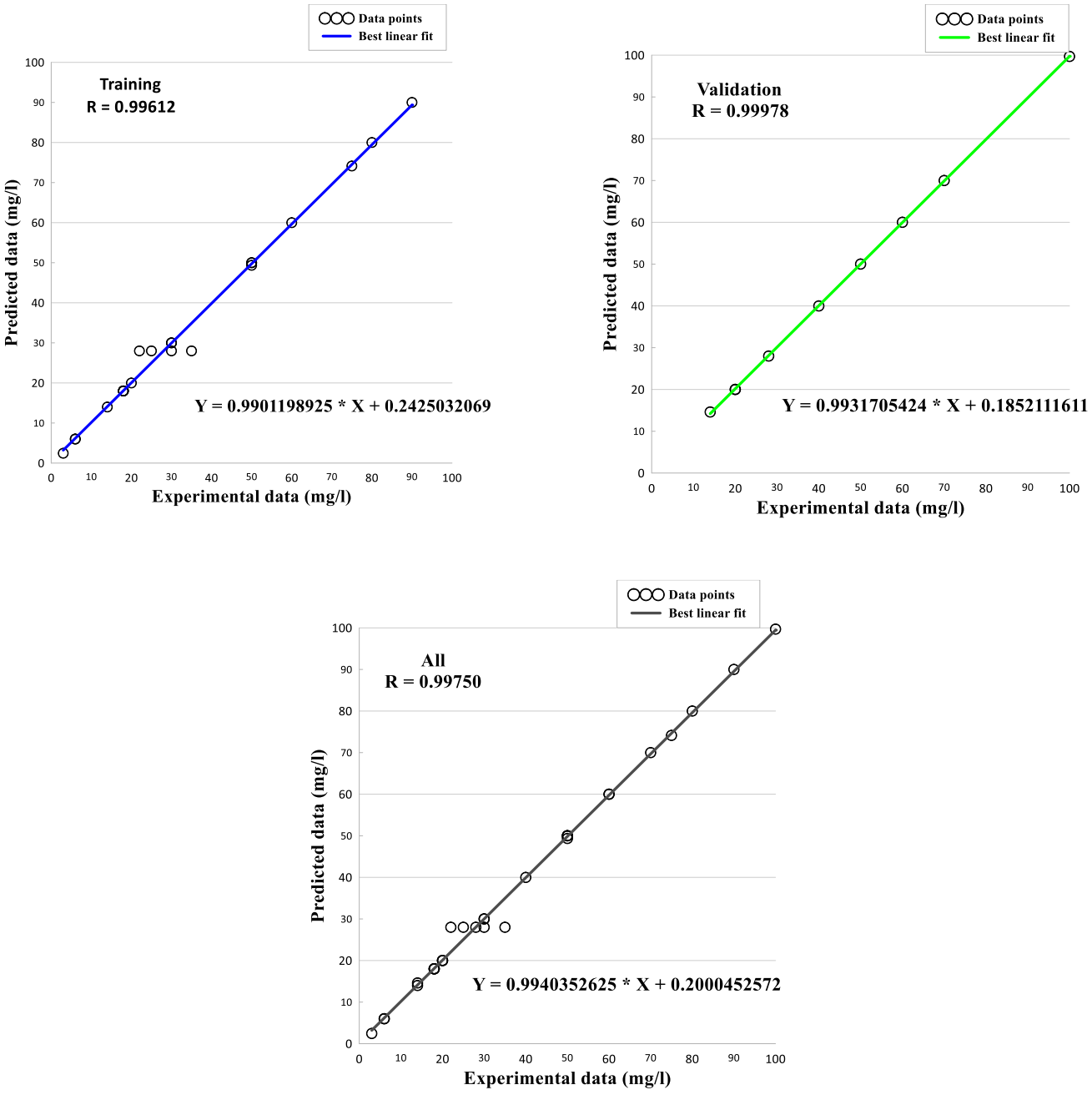


Fig. 6 – Relation between the observed dose of coagulant and those estimated by the ANN model

The results of the ANN performance in terms of all errors and in terms of agreement vector values (*R*, slope: α and y-intercept: β) are given in Table 5.

Table 5: Performances of the prediction using ANN

	RMSE／ mg l^{-1}	MSE／ mg l^{-1}	ESP／%	EPM／ %	MAE／ mg l^{-1}	*R*	α	β
all	1.79535	3.22331	4.83966	0.28386	0.67928	0.99750	0.99403	0.20004
training	2.12626	4.521011	6.25372	0.19098	0.91414	0.99612	0.99011	0.24250
validation	0.22616	0.05115	0.50634	0.00428	0.10519	0.99978	0.99317	0.18521

3.3. SVM

In this part, we have optimised 3 kernel functions, each kernel synthesised by parameters that must also be optimised:

The linear function synthesised by BOxConstraint, epsilon and sigma; the Gaussian function synthesized by BOxConstraint, epsilon and sigma; the polynomial function synthesized by BOxConstraint, epsilon and PolynomialOrder, knowing that the PolynomialOrder was optimised from 2 to 5 .

Once the result of the learning phase was obtained, it was validated by the validation database. The obtained results (predicted values) were compared with the experimental values in the two stages (the learning phase and the validation phase) in order to have the correlation coefficient and RMSE. The results of these tests are shown in Table 6. It should be noted that the results have been denormalized to the actual values to compare to the other models.

Table 6 – Performances of the different tested SVM Kernel function

Kernel function	*C*	*ε*	σ^2	*d*	Quantity of support vectors	Coefficients of correlation	RMSE
						Total	Total
Linear	846	0.0032	/	/	22	0.66564	19.04920
Gaussian	**479.0010**	**2.5551e-04**	**2.3000**	**/**	**22**	**0.99668**	**2.08927**
Polynomial	816	0.0052	/	2	22	0.92786	9.54540
Polynomial	23	0.0035	/	2.5	22	0.93081	9.3885
Polynomial	102	0.0012	/	3	22	0.98170	4.87540
Polynomial	614	0.0013	/	3.5	22	0.98311	4.67220
Polynomial	623	0.001	/	4	22	0.98672	4.17170
Polynomial	731	5.8782e-04	/	4.5	22	0.99272	3.16830
Polynomial	855	4.3554e-04	/	5	22	0.99474	0.9947

It is obvious that the Gaussian function gives better result in terms of correlation coefficient and RMSE; this result has been graphically schematized by Fig. 7.

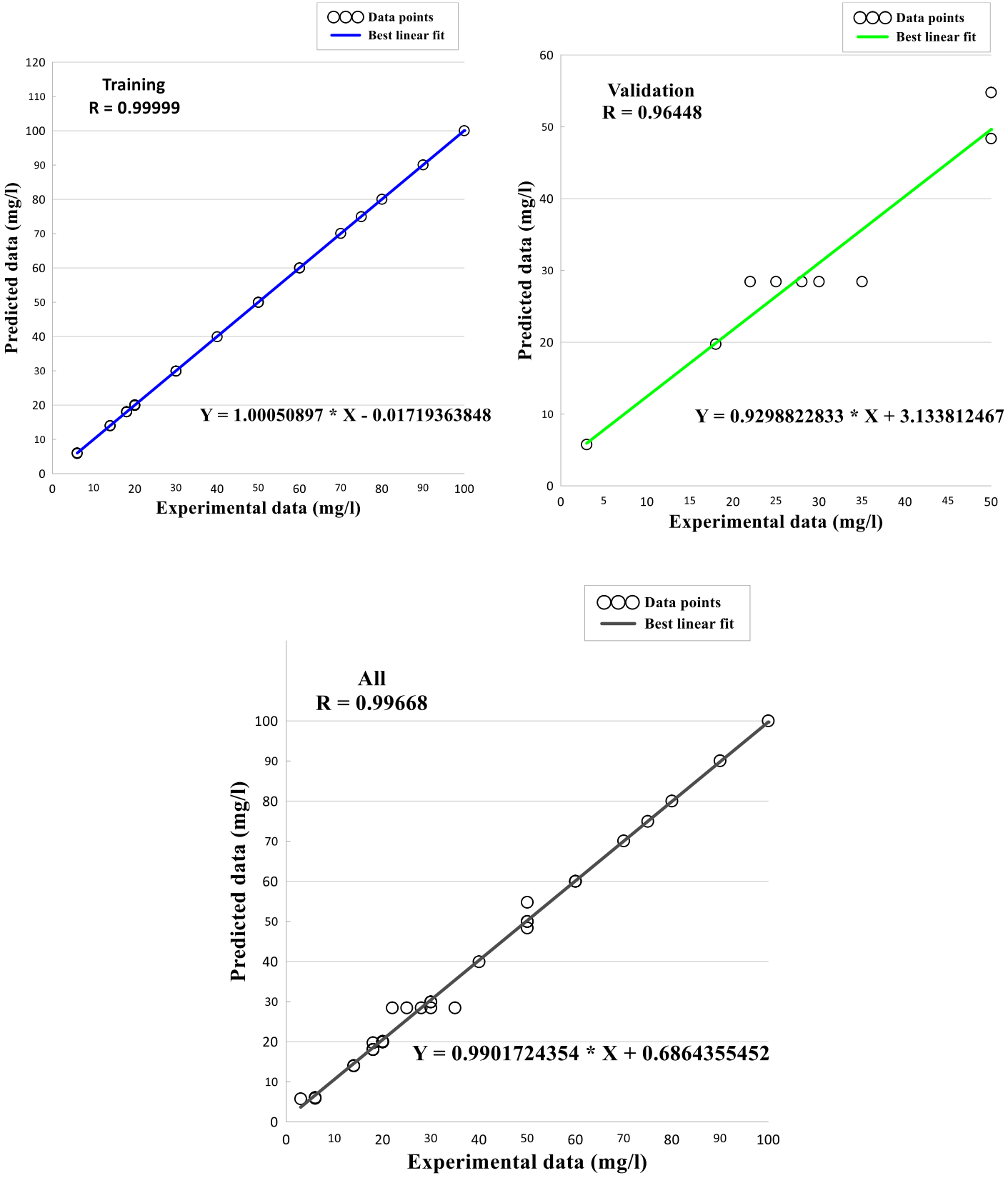


Fig. 7 – Relation between the observed dose of coagulant and those estimated by the SVM model

The results of the SVM performance in terms of all errors and in terms of agreement vector values (*R*, slope: α and intercept: β) are given in Table 7.

Table 7 – Performances of the prediction using SVM

	RMSE / mg l^{-1}	MSE / mg l^{-1}	ESP / %	EPM / %	MAE / mg l^{-1}	*R*	α	β
all	2.08927	4.36505	5.63194	0.58983	0.98625	0.99668	0.99017	0.68643
training	0.06292	0.00395	0.15571	0.01155	0.05477	0.99999	1.00050	−0.01719
validation	3.87627	15.02550	13.36646	0.16488	3.26318	0.96448	0.92988	3.13381

3.4. ANFIS

In this work, we have optimised the weighting algorithms, the functions (*trimf*, *trapmf*, *gbellmf*, *gaussmf*, *gauss2mf*, *pimf*, *dsigmf* and *psigmf*) for the input and (constant and linear) for the output and many of the nodes for each input to obtain the most accurate result. It should be noted that these algorithms and functions could be found in the toolbox of Matlab R 2013.

As described above, two steps were followed for the ANFIS modelling, where two sets of data were used, the training and verification data, respectively. First, the training was done using the training data set. Secondly, verification of the data set was considered to check the accuracy and effectiveness of the ANFIS model.

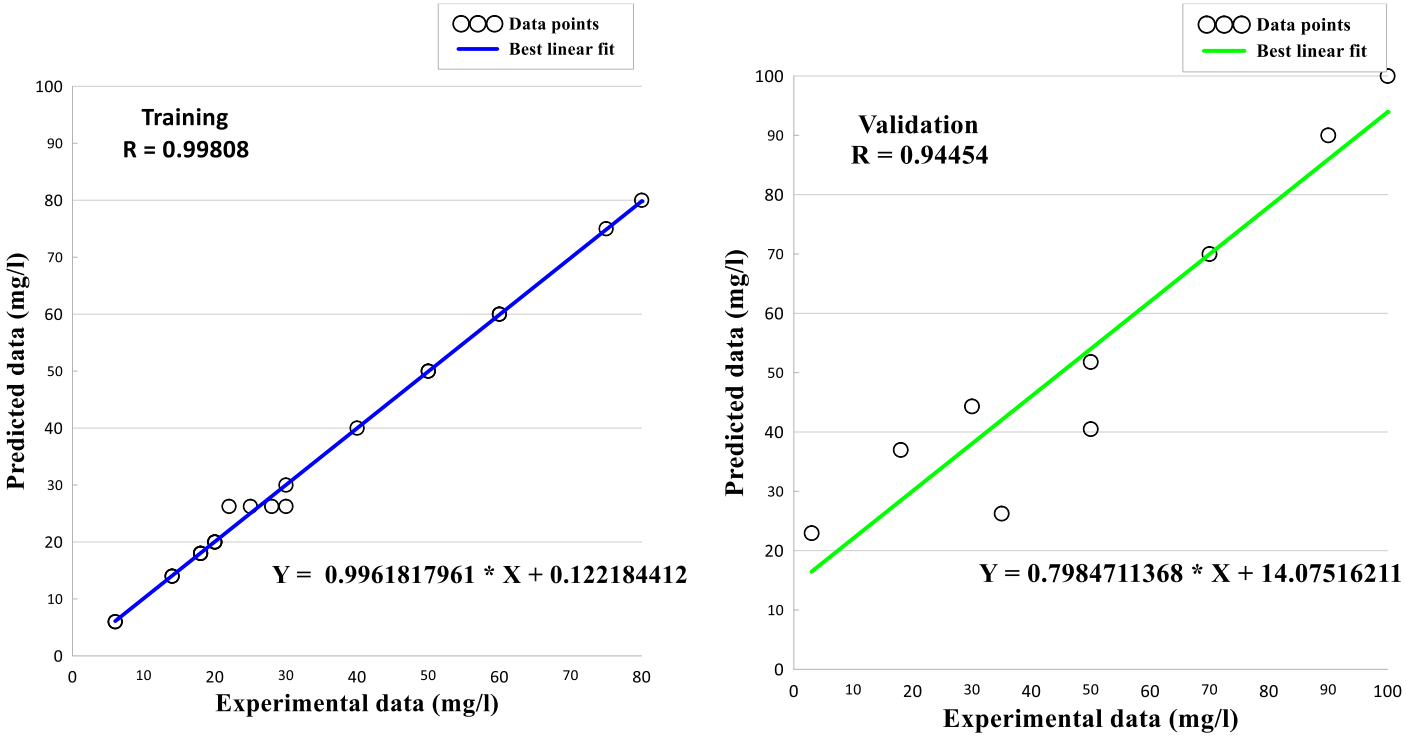
Two criteria were adopted for optimising the ANFIS model: the number of membership functions assigned to the inputs and output and the RMSE value. In addition, the following four aspects were considered for the ANFIS learning phase, which were the number of used membership functions, the membership functions type, the overfitting and the training options. The results of these tests are presented in Table 8.

Table 8 – Performances of the different tested ANFIS architectures

ANFIS		MF Type		Train FIS	Coefficients of correlation	RMSE
Nbr	Number of MFs	Input	Output	Optimisation method	Total	Total

1	[2 2 2 2 2]	*trimf*	linear	hybrid	0.93691	9.24210
2	[2 2 2 2 2]	*trapmf*	linear	hybrid	0.95284	7.83380
3	**[2 2 2 2 2]**	***gbellmf***	**linear**	**hybrid**	**0.97137**	**6.14477**
4	[2 2 2 2 2]	*gaussmf*	linear	hybrid	0.96934	6.35520
5	[2 2 2 2 2]	*gauss2mf*	linear	hybrid	0.96842	6.45090
6	[2 2 2 2 2]	*pimf*	linear	hybrid	0.96839	6.45360
7	[2 2 2 2 2]	*dsigmf*	linear	hybrid	0.96875	6.42420
8	[2 2 2 2 2]	*psigmf*	linear	hybrid	0.96875	6.42420

The best results obtained are gathered in Table 8 and they are illustrated in Fig. 8 by denormalizing the values to the actual values.



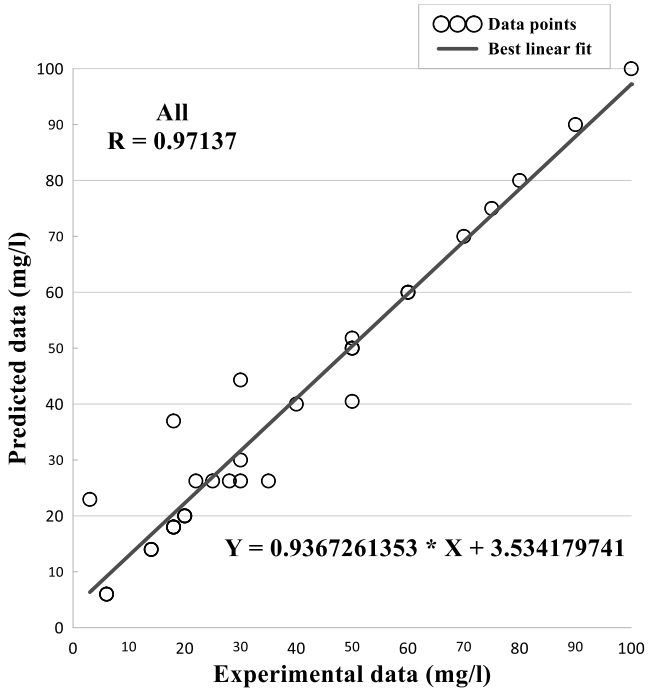


Fig. 8 – Relation between the observed dose of coagulant and those estimated by the ANFIS model

The results of the ANFIS performances in terms of all errors and in terms of agreement vector values (*R*, slope: α and intercept: β) are given in Table 9.

Table 9: Performances of the prediction using SVM

	RMSE／ mg l^{-1}	MSE／ mg l^{-1}	ESP／%	EPM／%	MAE／ mg l^{-1}	*R*	α	β
all	6.14477	37.75828	16.56418	2.81801	2.71994	0.97137	0.93672	3.53417
training	1.29246	1.67045	4.03893	0.09475	0.05000	0.99809	0.99618	0.12218
validation	11.22376	125.97298	22.64886	0.77937	8.14644	0.94454	0.79847	14.07516

3.5. Comparison of the models

The comparison of the coefficients of correlation, determination, adjustment and statistical indicators (RMSE, MSE, ESP, EPM and MAE) obtained by the four models, RSM, ANN, SVM and ANFIS, for all data is presented in table 10. The calculated coefficients by the model ANN was slightly higher than those obtained by the SVM model and significantly higher than those obtained by the RSM and the ANFIS models. These results were also confirmed by the

statistical indicators, since those given by the ANN model were slightly lower than those obtained by the SVM model, followed by the RSM model and finally the ANFIS model.

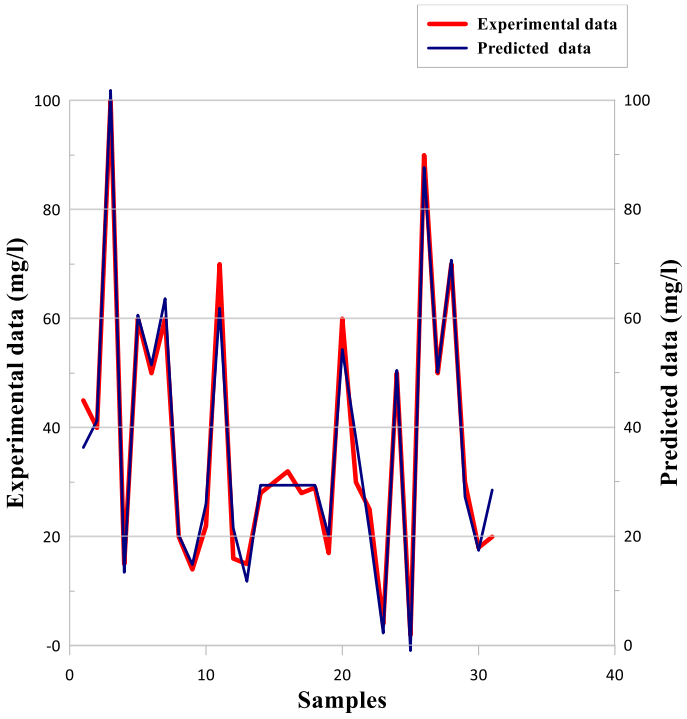
Table 10: Comparison of the predictive performance of the RSM, ANN, SVM and ANFIS Models for all data

Model	RMSE/ mg l^{-1}	MSE/ mg l^{-1}	ESP/%	EPM/ %	MAE/ mg l^{-1}	R	A	B	R^2	R^2_{adj}	No. of parameters
RSM	3.82890	14.66053	10.41194	1.47823	3.82890	0.98691	0.97399	0.95615	0.97399	0.96878	13
ANN	1.79535	3.22331	4.83966	0.28386	0.67928	0.99750	0.99403	0.20004	0.99500	0.99400	43
SVM	2.08927	4.36505	5.63194	0.58983	0.98625	0.99668	0.99017	0.68643	0.99337	0.99204	22
ANFIS	6.14477	37.75828	16.56418	2.81801	2.71994	0.97137	0.93672	3.53417	0.94355	0.93227	309

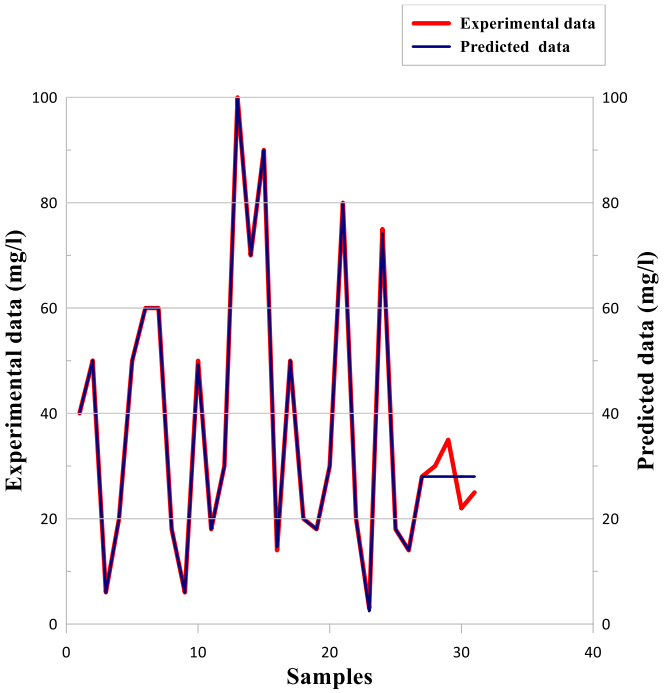
The comparison of the different models is illustrated in Fig. 9, confirming the superiority of the ANN and SVM model over the other models (RSM, ANFIS) with a very good superposition of the experimental values and those given by the models. Despite the ANN and SVM models were almost equal, the ANN model was slightly superior. The correlation coefficient and the statistical indicators of the apprenticeship phase of the ANN model were almost equal to those given by the SVM model with slight superiority of the SVM model (Tables 5 and 7). However, in the validation phase the correlation coefficient was very high and the statistical indicators were very small for the ANN model (R = 0.99978, RMSE = 0.22616, MSE = 0.05115, ESP = 0.50634, EPM = 0.00428 and MAE = 0.10519) compared to the SVM model (R = 0.96448, RMSE = 3.87627, MSE = 15.02550, ESP = 13.36646, EPM = 0.16488, and MAE = 3.26318) (Tables 5 and 7). However, in terms of reducing economic costs, the SVM model was superior, since the number of parameters of this model was 22, which is almost half of the number of parameters of the ANN model (43 parameters). To further reduce the economic costs, we can also use the RSM model which remains very useful because of its high coefficients (R = 0.98691, R^2 = 0.97399, R^2_{adj} = 0.96878) compared to the number of parameters, namely 13. In addition, the statistic indicators remained acceptable for the RSM model (RMSE = 3.82890, MSE = 14.66053, ESP = 10.441194, EPM = 1.47823 and MAE = 3.82890) if related to the number of model parameters.

Finally, the ANFIS model which was acceptable from the point of view of coefficients and statistical indicators (R = 0.97137, R^2 = 0.94355, R^2_{adj} = 0.93227, RMSE = 6.14477, MSE = 37.75828, ESP = 16.56418, EPM = 2.81801 and MAE = 2.71994), appeared very

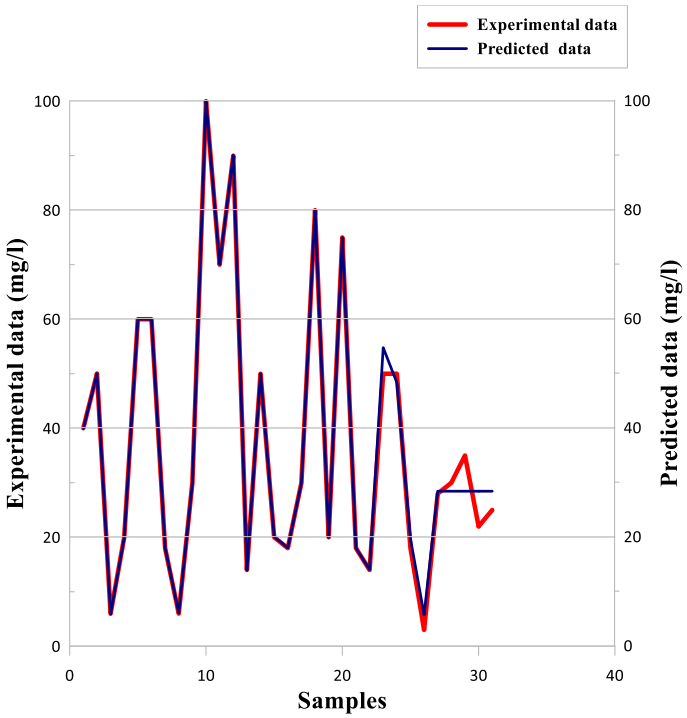
expensive because the number of parameters was very high (309 parameters) and less accurate results compared to other models.



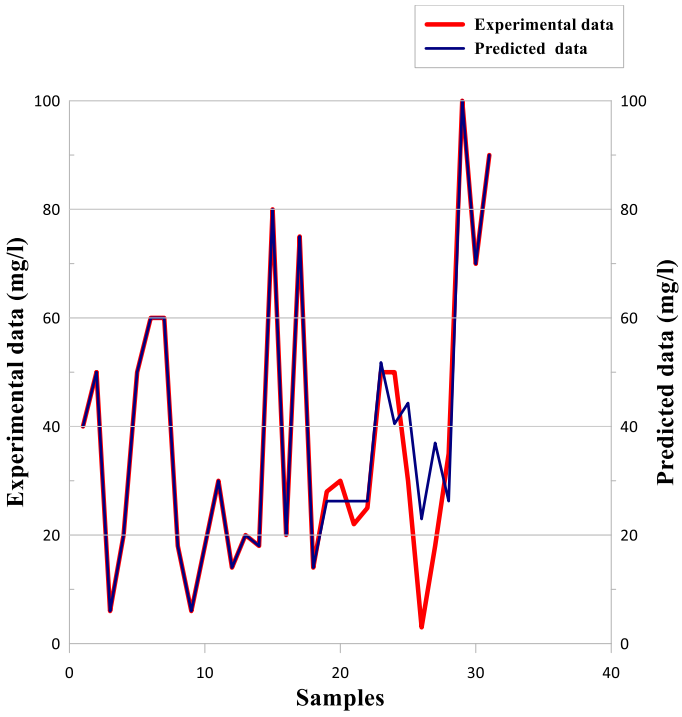
RSM Model



ANN Model



SVM Model

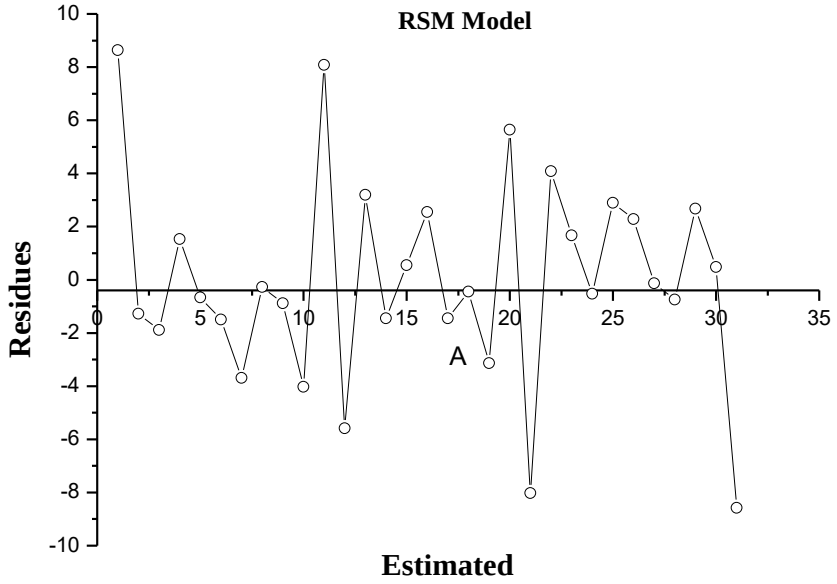


ANFIS Model

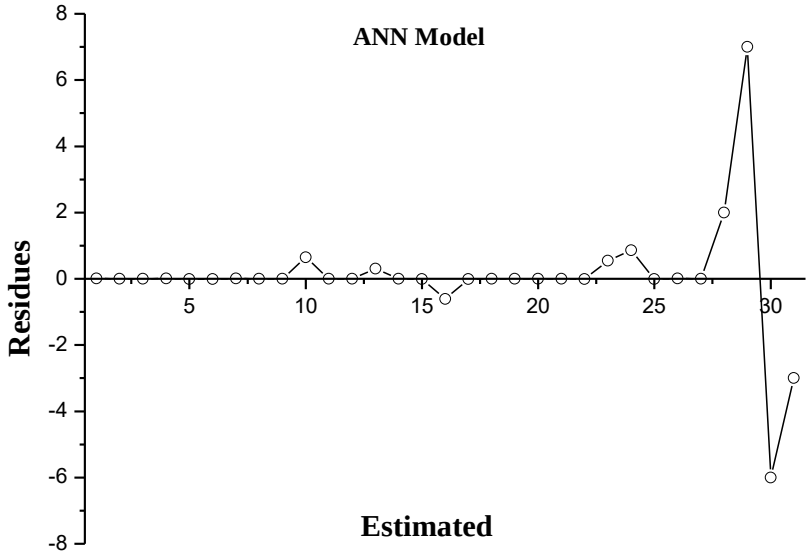
Fig. 9 – Relationship between experimental data and the predicted data of samples using RSM, ANN, SVM, and ANFIS modelling

3.6 Residues study

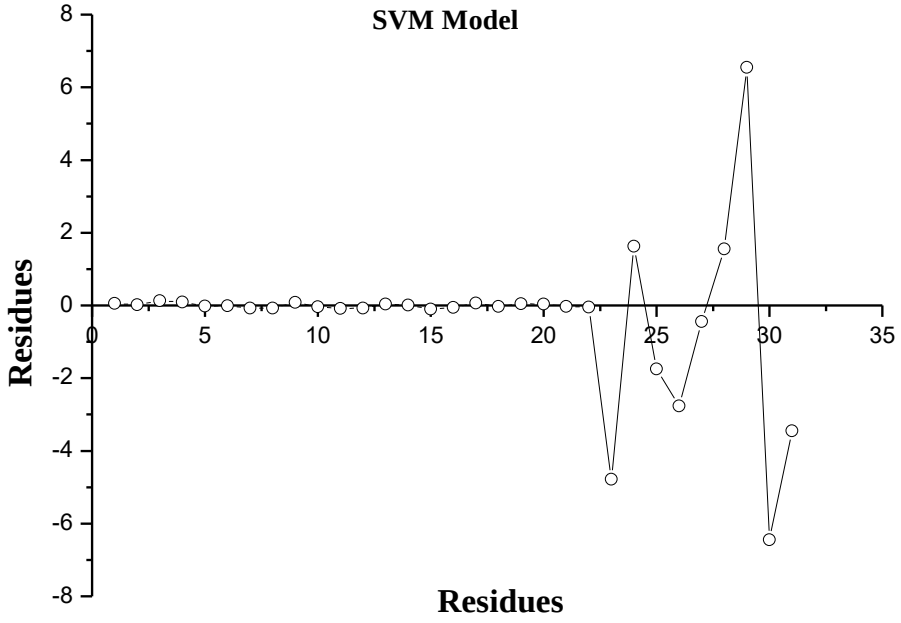
Residue analysis is an efficient way to reveal the performances of the optimised model. It consists in the measurement of the absolute or relative error between experimental and predicted values.[55] Residue analysis methods are mainly graphical analysis methods. Fig. 10 shows the residuals related to each model (RSM, ANN, SVM, and ANFIS). This figure shows that the residuals obtained by the neural network method were less scattered (close to zero) compared to those obtained by the other models. Again, this result shows the accuracy and robustness of the ANN model compared to the other models and justifies the use of the ANN approach in the prediction of coagulant doses.



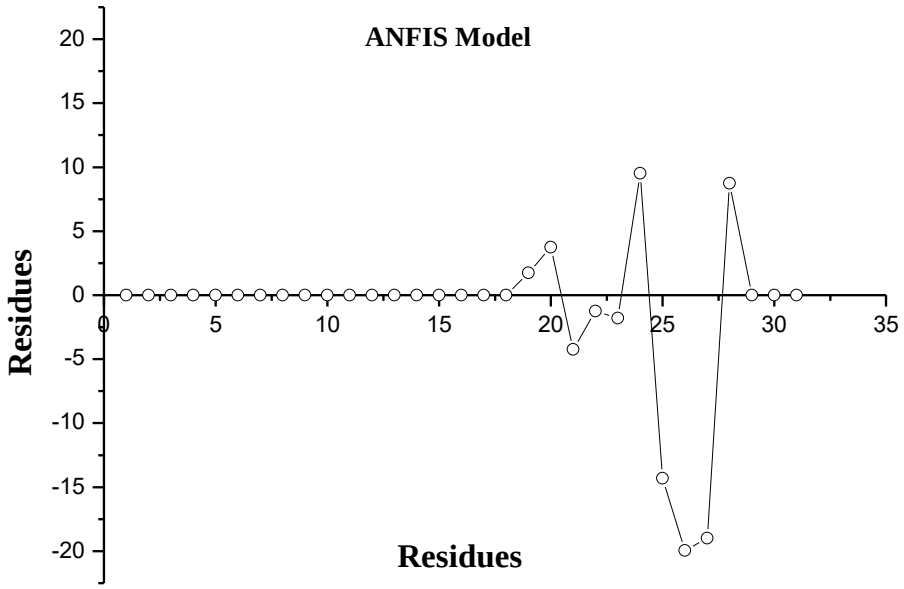
(a)


ANN Model
Residues
Estimated
8
6
4
2
0
-2
-4
-6
-8
5
10
15
20
25
30

(b)


SVM Model
Residues
Residues
8
6
4
2
0
-2
-4
-6
-8
0
5
10
15
20
25
30
35

(c)



(d)

Fig. 10 – Distribution of residuals for the RSM (a), ANN (b), SVM (c), and ANFIS (d) models for all the data.

4 Conclusion

In this study, four mathematical models (RSM, ANN, SVM, and ANFIS) were used to build models to predict the dose of coagulant for removing organic components and turbidity and also to reduce the waste of expensive chemicals and operating costs and achieve drinking water quality objectives. The results obtained from the RSM, ANN, SVM, and ANFIS models proved to be useful and effective models for predicting the coagulant doses from the point of view of coefficients and statistical indicators. However, the ANN model appeared to be the most effective model, as it leads to the highest coefficient (R^2_{adj} = 0.99400) and the lowest error indicator (RMSE = 1.79535) compared to the other models. Followed by the SVM model (R^2_{adj} = 0.99204, RMSE = 2.08927), the RSM model (R^2_{adj} = 0.96878, RMSE = 3.82890) and finally the ANFIS model (R^2_{adj} = 0.93227, RMSE = 6.14477). Furthermore, from an economic point of view, only three models appeared relevant: the ANN model which involved 43, the SVM model which was almost similar to the ANN model in terms of coefficients and statistical indicators but involved only 22 parameters, and finally the RSM model, which remains very

useful in terms of reducing economic costs due to its high coefficient (R^2_{adj} = 0.96878) if related to the number of parameters, only 13. In addition, it led to the acceptable statistical indicators. Contrarily, the ANFIS model was considered unacceptable, owing to the very high number of parameters, 309.

References

1. *A. T. A. Baptista, P. F. Coldebella, P. H. F. Cardines, R. G. Gomes, M. F. Vieira, R. Bergamasco, A. M. S. Vieira*, Coagulation–flocculation process with ultrafiltered saline extract of Moringa oleifera for the treatment of surface water, Chem. Eng. Technol. **276** (2015) 166–173, doi: https://doi.org/10.1016/j.cej.2015.04.045.
2. *J. Bratby*, Coagulation and flocculation in water and wastewater treatment, IWA publishing, 2016, doi: https://doi.org/10.2166/9781780407500.
3. *J. K. Edzwald, J. E. Tobiason*, Enhanced coagulation: US requirements and a broader view, Water Sci. Technol. **40** (1999) 63–70, doi: https://doi.org/10.2166/wst.1999.0444.
4. *J.-Q. Jiang*, Development of coagulation theory and pre-polymerized coagulants for water treatment, Sep. Purif. Technol. **30** (2001) 127–141, doi: https://doi.org/10.1081/SPM-100102986.
5. *B. Bolto, J. Gregory*, Organic polyelectrolytes in water treatment, Water Res. **41** (2007) 2301–2324, doi: https://doi.org/10.1016/j.watres.2007.03.012.
6. *R. Jiao, R. Fabris, C.W. Chow, M. Drikas, J. van Leeuwen, D. Wang*, Roles of coagulant species and mechanisms on floc characteristics and filterability, Chemosphere **150** (2016) 211–218, doi: https://doi.org/10.1016/j.chemosphere.2016.02.030.
7. *Z. Yang, B. Gao, Q. Yue*, Coagulation performance and residual aluminum speciation of $Al_2(SO_4)_3$ and polyaluminum chloride (PAC) in Yellow River water treatment, Chem. Eng. Technol. **165** (2010) 122–132, doi: https://doi.org/10.1016/j.cej.2010.08.076.
8. *B. Lakhdari*, Effet de la coagulation-floculation sur la qualité des eaux épurées de la STEP de Ain El Houtz, 2011, url: http://dspace.univ-tlemcen.dz/handle/112/132.
9. *M. Sillanpää, A. Matilainen*, NOM removal by coagulation, Natural Organic Matter in Water: Characterization and Treatment Methods. **55** (2015) 55–80, doi: https://doi.org/10.1016/B978-0-12-801503-2.00003-3.
10. *D.-S. Joo, D.-J. Choi, H. Park*, The effects of data preprocessing in the determination of coagulant dosing rate, Water Res. **34** (2000) 3295–3302, doi: https://doi.org/10.1016/S0043-1354(00)00067-1.
11. *A. Robenson, S. A. Shukor, N. Aziz*, Development of process inverse neural network model to determine the required alum dosage at Segama water treatment plant Sabah, Malaysia, Comput. Aided Chem. Eng. **27** (2009)525–530, doi: https://doi.org/10.1016/S1570-7946(09)70308-6.
12. *O. Bello, Y. Hamam, K. Djouani*, Control of a coagulation chemical dosing unit for water treatment plants using MMPC based on fuzzy weighting, J. Water Process. Eng. **4** (2014) 34–46, doi: https://doi.org/10.1016/j.jwpe.2014.08.007.
13. *A. Adgar, C. S. Cox, C. A. Jones*, Enhancement of coagulation control using the streaming current detector, Bioproc. Biosyst. Eng. **27** (2005) 349–357, doi: https://doi.org/10.1007/s00449-005-0413-5.
14. *B. Hua*, A fuzzy logic based coagulant real time control scheme for water purification system, in: 2009 International Conference on Mechatronics and Automation, IEEE, 2009: pp. 2123–2127, doi. https://doi.org/10.1109/ICMA.2009.5246439.
15. *P. S. Paz, J. P. V. Ocampo*, Automation and control of a water treatment plant process, in: Proceedings of 4th Mercosur Congress on Process System Engineering, Mecosur, 2009, p. 2009.

16. *J. S. Torrecilla, M. L. Mena, P. Yáñez-Sedeño, J. García*, Application of artificial neural network to the determination of phenolic compounds in olive oil mill wastewater, J. Food Eng. **81** (2007) 544–552, doi: https://doi.org/10.1016/j.jfoodeng.2006.12.003.
17. *P. Perez, A. Trier*, Prediction of NO and NO_2 concentrations near a street with heavy traffic in Santiago, Chile, Atmos. Environ. **35** (2001) 1783–1789, doi: https://doi.org/10.1016/S1352-2310(00)00288-0.
18. *O. A. Adesina, F. Abdulkareem, A. S. Yusuff, M. Lala, A. Okewale*, Response surface methodology approach to optimization of process parameter for coagulation process of surface water using Moringa oleifera seed, S. Afr. J. Chem. Eng. **28** (2019) 46–51, doi: https://doi.org/10.1016/j.sajce.2019.02.002.
19. *B. Singh, P. Kumar*, Pre-treatment of petroleum refinery wastewater by coagulation and flocculation using mixed coagulant: Optimization of process parameters using response surface methodology (RSM), J. Water Process. Eng. **36** (2020) 101317, doi: https://doi.org/10.1016/j.jwpe.2020.101317.
20. *T. K. Trinh, L. S. Kang*, Response surface methodological approach to optimize the coagulation–flocculation process in drinking water treatment, Chem. Eng. Res. Des. **89** (2011) 1126–1135, doi: https://doi.org/10.1016/j.cherd.2010.12.004.
21. *M. J. Kennedy, A. H. Gandomi, C. M. Miller*, Coagulation modeling using artificial neural networks to predict both turbidity and DOM-PARAFAC component removal, J. Environ. Chem. Eng. **3** (2015) 2829–2838, doi: https://doi.org/10.1016/j.jece.2015.10.010.
22. *C. Jayaweera, M. Othman, N. Aziz*, Improved predictive capability of coagulation process by extreme learning machine with radial basis function, J. Water Process Eng. **32** (2019) 100977, doi: https://doi.org/10.1016/j.jwpe.2019.100977.
23. *C. M. Kim, M. Parnichkun*, Prediction of settled water turbidity and optimal coagulant dosage in drinking water treatment plant using a hybrid model of k-means clustering and adaptive neuro-fuzzy inference system, Appl. Water Sci. **7** (2017) 3885–3902, doi: https://doi.org/10.1007/s13201-017-0541-5.
24. *G.-D. Wu, S.-L. Lo*, Predicting real-time coagulant dosage in water treatment by artificial neural networks and adaptive network-based fuzzy inference system, Eng. Appl. Artif. Intell. **21** (2008) 1189–1195, doi: https://doi.org/10.1016/j.engappai.2008.03.015.
25. *G. Rädlinger, K. G. Heumann*, Determination of halogen species of humic substances using HPLC/ICP-MS coupling, Fresenius' J. Anal. Chem. **359** (4) (1997) 430–433, doi: https://doi.org/10.1007/s002160050604.
26. *H. N. P. Dayarathne, M. J. Angove, R. Aryal, H. Abuel-Naga, B. Mainali*, Removal of natural organic matter from source water: Review on coagulants, dual coagulation, alternative coagulants, and mechanisms, J. Water Process. Eng. (2020) 101820 (in press), doi: https://doi.org/10.1016/j.jwpe.2020.101820.
27. *J. J. Milledge, S. Heaven*, A review of the harvesting of micro-algae for biofuel production, *Reviews in* Rev. Environ. Sci. Biotechnol. **12** (2) (2013) 165–178, doi: https://doi.org/10.1007/s11157-012-9301-z.
28. *E. Lefebvre, B. Legube*, Coagulation par Fe (III) de substances humiques extraites d'eaux de surface: Effet du pH et de la concentration en substances humiques, Water Res. **24** (5) (1990) 591–606, doi: https://doi.org/10.1016/0043-1354(90)90192-9.
29. *E. M. Thurman*, Organic geochemistry of natural waters. Springer Science & Business Media, Vol. 2, 2012.
30. *F. Zidane, K. Cheggari, J.-F. Blais, P. Drogui, J. Bensaid, S. I. Ahmed*, Contribution à l'étude de l'effet de la coagulation avant chloration sur la formation des trihalométhanes (THM) et

composés organohalogénés (COX) dans les eaux alimentant la ville de Casablanca au Maroc, Can. J. Civ. Eng. **37** (2010) 1149–1156, doi: https://doi.org/10.1139/L10-065.

31. *H. Yamamura, E.U. Putri, T. Kawakami, A. Suzuki, H.D. Ariesyady, T. Ishii*, Dosage optimization of polyaluminum chloride by the application of convolutional neural network to the floc images captured in jar tests, Sep. Purif. Technol. **237** (2020) 116467, doi: https://doi.org/10.1016/j.seppur.2019.116467.
32. *J. Rodier, B. Legube, N. Merlet, R. Brunet*, L'analyse de l'eau - 9e éd.: Eaux naturelles, eaux résiduaires, eau de mer, Dunod, 2009, url: https://books.google.dz/books?id=qUEGsUBZkL0C.
33. *D. Jamin*, Recherche du boson de Higgs du Modèle Standard dans le canal de désintégration ZH-> nu nu bb sur le collisionneur Tevatron dans l'expérience D0. Développement d'une méthode d'étiquetage des jets de quark b avec des muons de basses impulsions transverses, 2010.
34. *R. Yacef, A. Mellit, S. Belaid, Z. Şen*, New combined models for estimating daily global solar radiation from measured air temperature in semi-arid climates: application in Ghardaïa, Algeria, Energ. Conver. Manag. **79** (2014) 606–615, doi: https://doi.org/10.1016/j.enconman.2013.12.057.
35. *H. El Badaoui, A. Abdallaoui, I. Manssouri, L. Lancelot*, Elaboration de modèles mathématiques stochastiques pour la prédiction des teneurs en métaux lourds des eaux superficielles en utilisant les réseaux de neurones artificiels et la régression linéaire multiple, Int. J. Environ. Res. Public Health. **3** (2012) 31–36.
36. *H. Tahraoui, A. E. Belhadj, A. E. Hamitouche*, Prediction of the Bicarbonate Amount in Drinking Water in the Region of Médéa Using Artificial Neural Network Modelling, Kem. Ind. **69** (2020) 595–602, doi: https://doi.org/10.15255/KUI.2020.002.
37. *D. A. Belsley, E. Kuh, R. E. Welsch*, Regression diagnostics: identifying influential data and sources of collinearity, Wiley, 1980, doi: https://doi.org/10.1002/0471725153.
38. *S. H. Hong, M. W. Lee, D. S. Lee, J. M. Park*, Monitoring of sequencing batch reactor for nitrogen and phosphorus removal using neural networks, Biochem. Eng. J. **35** (2007) 365–370, doi: https://doi.org/10.1016/j.bej.2007.01.033.
39. *O. R. Dolling, E. A. Varas*, Artificial neural networks for streamflow prediction, J. Hydraul. Res. **40** (2002) 547–554, doi: https://doi.org/10.1080/00221680209499899.
40. *I. Manssouri, M. Manssouri, B. El Kihel*, Fault Detection by K-Nn Algorithm and Mlp Neural Networks in A Distillation Column: Comparative Study, J. Inform. Intell. Knowledge. **3** (2011) 201.
41. *I. Manssouri, A. El Hmaidi, T. E. Manssouri, B. El Moumni*, Prediction levels of heavy metals (Zn, Cu and Mn) in current Holocene deposits of the eastern part of the Mediterranean Moroccan margin (Alboran Sea), IOSR *J*. Comput. Eng. **16** (2014) 117–123, doi: https://doi.org/10.9790/0661-1618117123.
42. *A. Hammoudi, K. Moussaceb, C. Belebchouche, F. Dahmoune*, Comparison of artificial neural network (ANN) and response surface methodology (RSM) prediction in compressive strength of recycled concrete aggregates, Constr Build Maters. **209** (2019) 425–436, doi: https://doi.org/10.1016/j.conbuildmat.2019.03.119.
43. *D. C. Montgomery*, Design and analysis of experiments, 8th Ed, Wiley, Hoboken, NJ, 2013.
44. *E. Ghafari, H. Costa, E. Júlio*, RSM-based model to predict the performance of self-compacting UHPC reinforced with hybrid steel micro-fibers, Constr Build Maters. **66** (2014) 375–383, doi: https://doi.org/10.1016/j.conbuildmat.2014.05.064.
45. *F. Dahmoune, H. Remini, S. Dairi, O. Aoun, K. Moussi, N. Bouaoudia-Madi, N. Adjeroud, N. Kadri, K. Lefsih, L. Boughani, L. Mouni, B. Nayak, K. Madani*, Ultrasound assisted extraction of phenolic compounds from P. lentiscus L. leaves: Comparative study of artificial neural network (ANN) versus degree of experiment for prediction ability of phenolic compounds recovery, Ind. Crop. Prod. **77** (2015) 251–261, doi: https://doi.org/10.1016/j.indcrop.2015.08.062.

46. *H. Ousmana, A. E. Hmaidi, M. Berrada, B. Damnati, I. Etabaai, A. Essahlaoui*, Development of a Neural Network approach for Predicting nitrate and sulfate concentration in three lakes: Ifrah, Iffer and Afourgagh, Middle Atlas Morocco, Mor. J. Chem. 6 (2) (2018) 245–255, doi: https://doi.org/10.48317/IMIST.PRSM/morjchem-v6i2.5939.
47. *M. Sediri, S. Hanini, H. Cherifi, M. Laidi, S. A. Turki*, Dynamic Adsorption Modelling of P-nitrophenol in Aqueous Solution Using Artificial Neural Network, J. Mater. Environ. Sci. 8 (7) (2017) 2282–2287, url: https://www.jmaterenvironsci.com/Document/vol8/vol8_N7/245-JMES-Sediri.pdf.
48. *O. Bousquet, S. Boucheron, G. Lugosi*, Introduction to statistical learning theory, in: Summer School on Machine Learning, Springer, 2003: pp. 169–207, doi: https://doi.org/10.1007/978-3-540-28650-9_8.
49. *E. Pourbasheer, S. Riahi, M. R. Ganjali, P. Norouzi*, Application of genetic algorithm-support vector machine (GA-SVM) for prediction of BK-channels activity, Eur. J. Med. Chem. **44** (2009) 5023–5028, doi: https://doi.org/10.1016/j.ejmech.2009.09.006.
50. *S. J. AK, V. J. PL*, Least squares support vector machines, World scientific, 2002.
51. *V. Vapnik, S. E. Golowich, A. J. Smola*, Support vector method for function approximation, regression estimation and signal processing, in: Adv. Neural. Inform. Process Sys. 1997, pp. 281–287.
52. *X. Wang, J. Chen, C. Liu, F. Pan*, Hybrid modeling of penicillin fermentation process based on least square support vector machine, Chem. Eng. Res. Des. **88** (2010) 415–420, doi: https://doi.org/10.1016/j.cherd.2009.08.010.
53. *H. Benimam, C. Si-Moussa, M. Laidi, S. Hanini*, Modeling the activity coefficient at infinite dilution of water in ionic liquids using artificial neural networks and support vector machines, Neural. Comput. Appl. **32** (2019) 8635–8653, doi: https://doi.org/10.1007/s00521-019-04356-w.
54. *J.-S. Jang*, ANFIS: adaptive-network-based fuzzy inference system, IEEE Trans. Syst. Man Cyber. **23** (1993) 665–685, doi: https://doi.org/10.1109/21.256541.
55. *M. Bouhedda, S. Lefnaoui, S. Rebouh, M. M. Yahoum*, Predictive model based on Adaptive euro-Fuzzy Inference System for estimation of Cephalexin adsorption on the Octenyl Succinic Anhydride starch, Chemom. Intell. Lab. Syst. **193** (2019) 103843, doi: https://doi.org/10.1016/j.chemolab.2019.103843.
56. *R. Adeline, R. Gouriveau, N. Zerhouni*, Pronostic de défaillances: Maîtrise de l'erreur de prédiction., in: 2008.
57. *M. Naoual, A. Abdelaziz, E. H. Abdellah*, Use of Artificial Neural Networks Type MLP for the Prediction of Phosphorus Level from the Physicochemical Parameters of Sediments, (n.d.) 10.
58. *S. Lefnaoui, N. Moulai-Mostefa*, Investigation and optimization of formulation factors of a hydrogel network based on kappa carrageenan–pregelatinized starch blend using an experimental design, Colloids and Surfaces A: Colloids Surf. A Physicochem. Eng. Asp. **458** (2014) 117–125, doi: https://doi.org/10.1016/j.colsurfa.2014.01.007.
59. *R. Desjardins*, Le traitement des eaux, Presses inter Polytechnique, 1997.
60. *B. Lakhdari*, Effet de la coagulation-floculation sur la qualité des eaux épurées de la STEP de Ain El Houtz, 2011.
61. *F. Sekiou*, Effets de la nature des particules solides dispersées et des conditions de formation du floc sur l'efficacité de la floculation, Mémoires de Magister. (2014).
62. *S. Bouranene, N. Sedira, P. Fievet, N. Attia*, Treatment of paint wastewater by coagulation process, Filtr. Sep. **52** (2015) 42–45, doi: https://doi.org/10.1016/S0015-1882(15)30184-1.

https://doi.org/10.15255/KUI.2020.002
KUI-38/2020
Original scientific paper
Received January 1, 2020
Accepted February 19, 2020

Prediction of the Bicarbonate Amount in Drinking Water in the Region of Médéa Using Artificial Neural Network Modelling

H. Tahraoui,[a] *A.-E. Belhadj*,[a*] and *A.-E. Hamitouche* [b]

[a] Laboratory of Biomaterials and Transport Phenomenon (LBMPT), Faculty of Technology, University Yahia Fares of Médéa, 26 000, Algeria
[b] Center for Scientific and Technical Research in Physico-chemical Analysis CRAPC, BP 384, Bou-Ismail, RP 42 004, Tipaza, Algeria

This work is licensed under a Creative Commons Attribution 4.0 International License

Abstract

The region of Médéa (Algeria) located in an agricultural site requires a large amount of drinking water. For this purpose, the water analyses in question are imperative. To examine the evolution of the drinking water quality in this region, firstly, an experimental protocol was done in order to obtain a dataset by taking into account several physicochemical parameters. Secondly, the obtained data set was divided into two parts to form the artificial neural network, where 70 % of the data set was used for training, and the remaining 30 % was also divided into two equal parts: one for testing and the other for validation of the model. The intelligent model obtained was evaluated as a function of the correlation coefficient nearest to 1 and lowest mean square error (RMSE). A set of 84 data points were used in this study. Eighteen parameters in the input layer, five neurons in the hidden layer, and one parameter in the output layer were used for the ANN modelling. Levenberg Marquardt learning (LM) algorithm, logarithmic sigmoid, and linear transfer function were used, respectively, for the hidden and the output layers. The results obtained during the present study showed a correlation coefficient of $R = 0.99276$ with root mean square error RMSE $= 11.52613$ mg dm^{-3}. These results show that obtained ANN model gave far better and more significant results. It is obviously more accurate since its relative error is small with a correlation coefficient close to unity. Finally, it can be concluded that obtained model can effectively predict the rate of soluble bicarbonate in drinking water in the Médéa region.

Keywords

Drinking water, physicochemical parameters, bicarbonate, modelling, artificial neural networks

1 Introduction

Industrial, agricultural, and urban development alters the quality of water and makes it unsafe. This is the case in Médéa region, which is subjected to diversification and increase in the quantity of pollutants released in the aquatic environment without treatment. Depending on the origin of the waste, the pollution may be of a chemical nature, especially heavy metals. To predict the amount of a pollutant, depending on many physicochemical parameters, we refer to methods based on mathematical and artificial intelligence models, such as linear regression (RLM) and artificial neural networks (ANN).[1,2] Neural networks have been highly successful in modelling and predicting environmental parameters.[3] They have been applied for various purposes, e.g.: prediction of groundwater remediation costs for drinking use based on quality of water resource;[4] modelling for nitrate concentration in groundwater;[5] prediction of groundwater suitability for irrigation;[6] modelling of contaminated water treatment processes by homogeneous and heterogeneous nanocatalysis;[7] prediction of the quality of public water supply;[8] modelling of TDS concentrations in river water;[9] prediction of aluminium pitting in natural waters;[10] prediction of fluoride concentration;[11] prediction of bromate removal in drinking water;[12] estimating rainfall relationship with river pollution;[13] modelling of total dissolved solids;[14] forecasting nitrate concentration in groundwater;[15] effect of drought on pollution of river station by ANN;[16] and prediction of heavy metal concentrations.[17] All these works are linked to different physicochemical parameters of water.

In this work, we propose the use of ANN for the prediction of bicarbonate content of surface waters based on the physicochemical parameters.

2 Artificial neural network

Artificial neural network (ANN) represents a set of algorithms, the design of derives from and is schematically inspired by the functioning of biological neurons. ANNs are now used as a very powerful tool in modelling and analysing processes, as well as in predicting the behaviour of a given system.[18]

The ANN structure consists essentially of an input layer (independent variables), a number of hidden layers, and an output layer. Each of these layers consists of a number of interconnected processing units called neurons. These neurons interact by sending signals, and they are connected to all the neurons in the previous and next layers by links called weights and links.[19] The architecture of the ANN model is shown in Fig. 1.

* Corresponding author: Dr Abd-Elmouneïm Belhadj
Email: belhadj_1@yahoo.fr

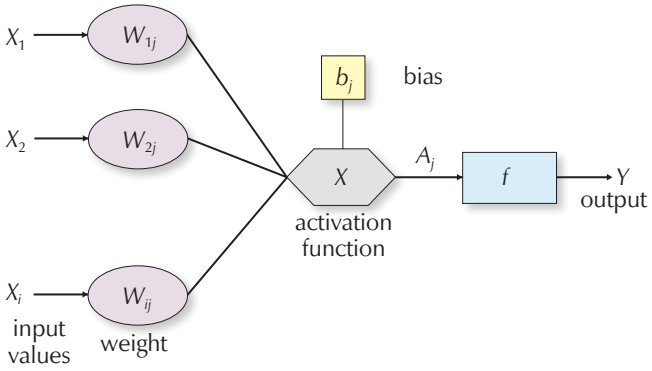


Fig. 1 – Presentation of a multilayer neural network

3 Materials and methods

In this work, a procedure based on the development and optimisation of the architecture of an ANN using software MATLAB R2013a, is described by a flowchart presented in Fig. 2.[20,21]

3.1 Database

The data used for this study were obtained from the experimental analysis of several water samples carried out for 84 different days in all seasons in the period from 2018 to 2019 at Médéa region, Algeria. Analyses were done according to the Jean Rodier's book of water analysis 9th Ed.[22]

The dependent variables are the contents of bicarbonate (HCO_3^-) soluble in water. The independent variables are the physicochemical parameters as follows: conductivity, turbidity, pH, hardness, calcium, magnesium, chlorides, total alkalinity titer (TAC), organic matter, nitrogen dioxide, nitrates, sodium, sulphates, potassium, heavy metals (Mn^{2+}, Fe^{3+}, and Al^{+}), and dry residues.

3.2 Normalisation and data pre-processing

The primary purpose of the data transformation is to modify the distribution of input variables so that they can better match outputs. Before training and validation, we scaled the inputs and targets using a normalised equation where the data always fall within a specified range:

$$x_N = (y_{max} - y_{min})\left(\frac{x - x_{min}}{x_{max} - x_{min}}\right) + y_{min} \quad (1)$$

The experimental data were normalised in the interval [−1,1] using the mapminmax function,[23] where x_N is the data value after normalisation, x_{max} and x_{min} denote the maximum and minimum of the data, respectively, y_{max} and y_{min} are taken as −1 and 1; and x represents the original values.

3.3 Data modelling techniques

Data modelling was carried out by ANNs that made it possible to justify the predictive quality of the models, by applying the same techniques on a set of data relating to 70 % of the samples, chosen at random, among the totality of the samples, which constituted the group for learning a predictive model of the dependent variable. The remaining 30 % of the samples, which did not participate in model learning, were divided in two (15 % for the test, and 15 % for the validation) to examine the validity and performance of the prediction of these models.[24]

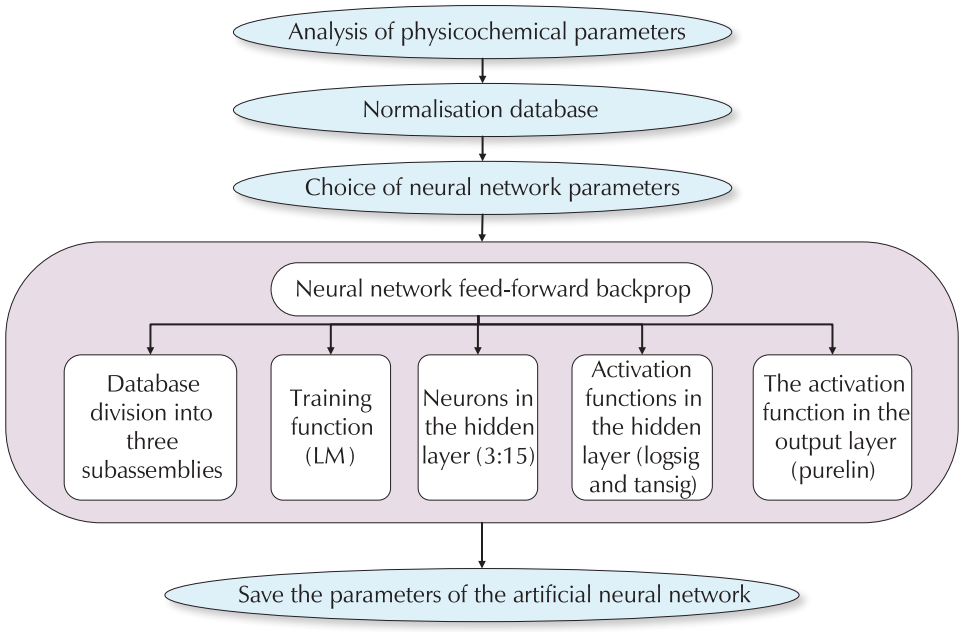


Fig. 2 – Organisation chart for the development and optimisation of the RNFF architecture

Table 1 – Structure of neural networks

Type of network	Learning algorithm	Input layer	Hidden layer		Output layer	
		Number of neurons	Number of neurons	Activation function	Number of neurons	Activation function
ANN	LM	18	5	Sigmoid	1	Linear

3.4 Modelling performance

In the current study, we present the different statistical parameters used in order to evaluate the effectiveness of each network and its ability to make precise predictions.[25–27]

The correlation coefficient (R), root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), standard preaching error (ESP), and error of prediction error model (EPM) were used to estimate the performance of the model. These statistical parameters were calculated using the following equations:[28,29]

$$R = \frac{\sum_{i=1}^{N}\left(y_{\text{exp}} - \overline{y}_{\text{exp}}\right)\left(y_{\text{pred}} - \overline{y}_{\text{pred}}\right)}{\sqrt{\sum_{i=1}^{N}\left(y_{\text{exp}} - \overline{y}_{\text{exp}}\right)^2 \sum_{i=1}^{N}\left(y_{\text{pred}} - \overline{y}_{\text{pred}}\right)^2}} \quad (2)$$

$$\text{RMSE} = \sqrt{\left(\frac{1}{N}\right)\left(\sum_{i=1}^{N}\left[\left(y_{\text{exp}} - y_{\text{pred}}\right)\right]^2\right)} \quad (3)$$

$$\text{MSE} = \left(\frac{1}{N}\right)\left(\sum_{i=1}^{N}\left(y_{\text{exp}} - y_{\text{pred}}\right)^2\right) \quad (4)$$

$$\text{MAE} = \left(\frac{1}{N}\right)\sum_{i=1}^{N}\left|y_{\text{exp}} - y_{\text{pred}}\right| \quad (5)$$

$$\text{ESP}(\%) = \frac{\text{RMSE}}{\overline{y}_{\text{exp}}} \cdot 100 \quad (6)$$

$$\text{EPM}(\%) = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{\left(y_{\text{exp}} - y_{\text{pred}}\right)}{y_{\text{exp}}}\right| \quad (7)$$

N is the number of data; y_{exp} and y_{pred} are the experimental and the predicted values, respectively, and $\overline{y}_{\text{exp}}$ and $\overline{y}_{\text{pred}}$ are the average values of the experimental and the predicted values, respectively.

4 Results and discussion

Preliminary tests have shown that, in order to improve the performance of a model established by ANN, it is necessary to modify the architecture of the network, by playing mainly on the number of hidden layers, or on the number of hidden layers, hidden neurons and/or the number of training cycles (number of iterations). For this, we successively changed the number of hidden neurons (NH = 3–15). The results of these tests are shown in Table 1. It was observed that the minimum of the MSE, the number of iterations, and the maximum correlation coefficient had been reached when NH = 5. The optimal acknowledgements architecture obtained from ANN has a topology of 18–5–1 (Table 1), and its architecture is illustrated in Figs. 3 and 4. We could therefore choose five neurons of the hidden layer of the network in the present study, in order to predict the bicarbonate concentrations.

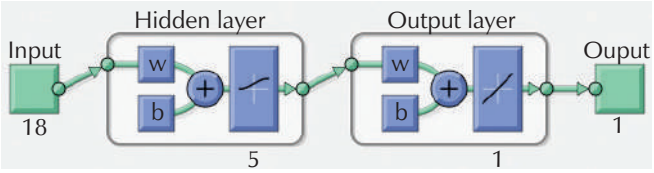


Fig. 3 – Schematic representation of the ANN model

Fig. 5 describes the results of the ANN training phase. It showed that, after 236 iterations, the desired result had been achieved with 5 neurons in the hidden layer, the three curves related to the evolution of the error of the three phases converged correctly towards an acceptable MSE.

The results obtained (Fig. 6) enabled extraction of the different values relating to the learning parameters:

- maximum number of iterations (Epochs) = 236,
- learning rate (η) = 0.001,
- minimum gradient = 8.2405e^{-16}.

Once the architecture of the neural network model is obtained, the model should be validated by comparing its results with the experimental data obtained in the laboratory. The results given in Fig. 7 show the performance and the efficiency of the developed ANN model. The relevance

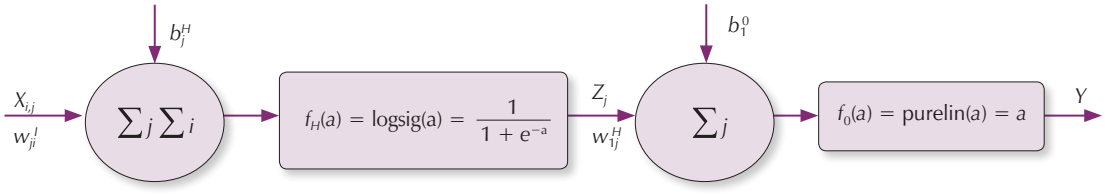


Fig. 4 – Schematic representation of the ANN neuronal model

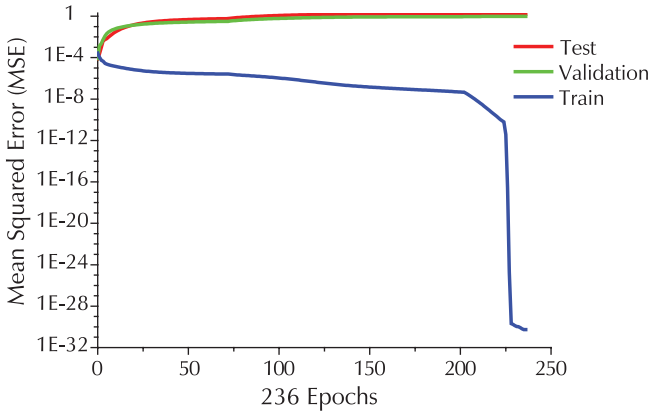


Fig. 5 – Evolution of the MSE

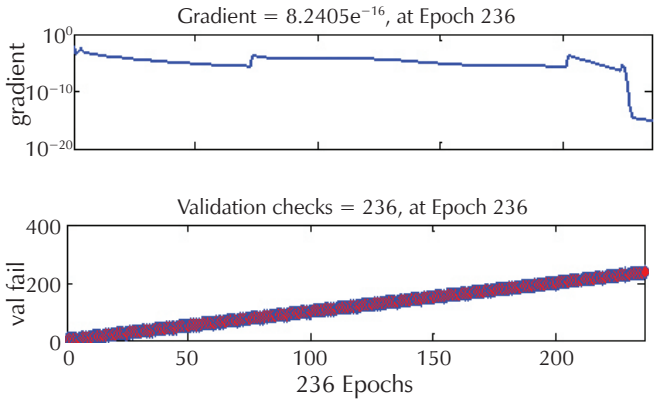


Fig. 6 – Gradient variations of error, learning rate, and validation error as a function of number of iterations

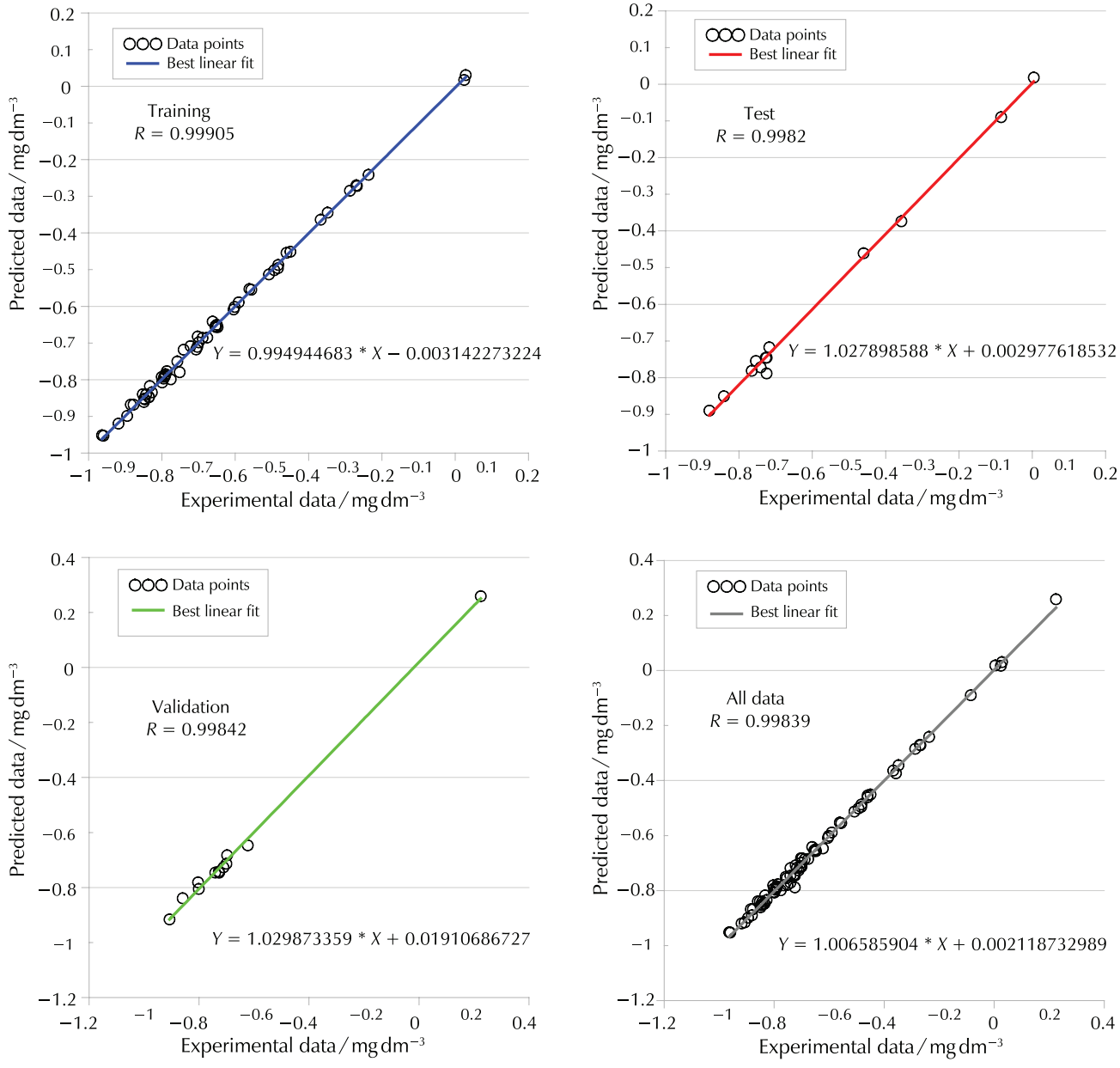


Fig. 7 – Relation between observed bicarbonate levels and those estimated by ANN model

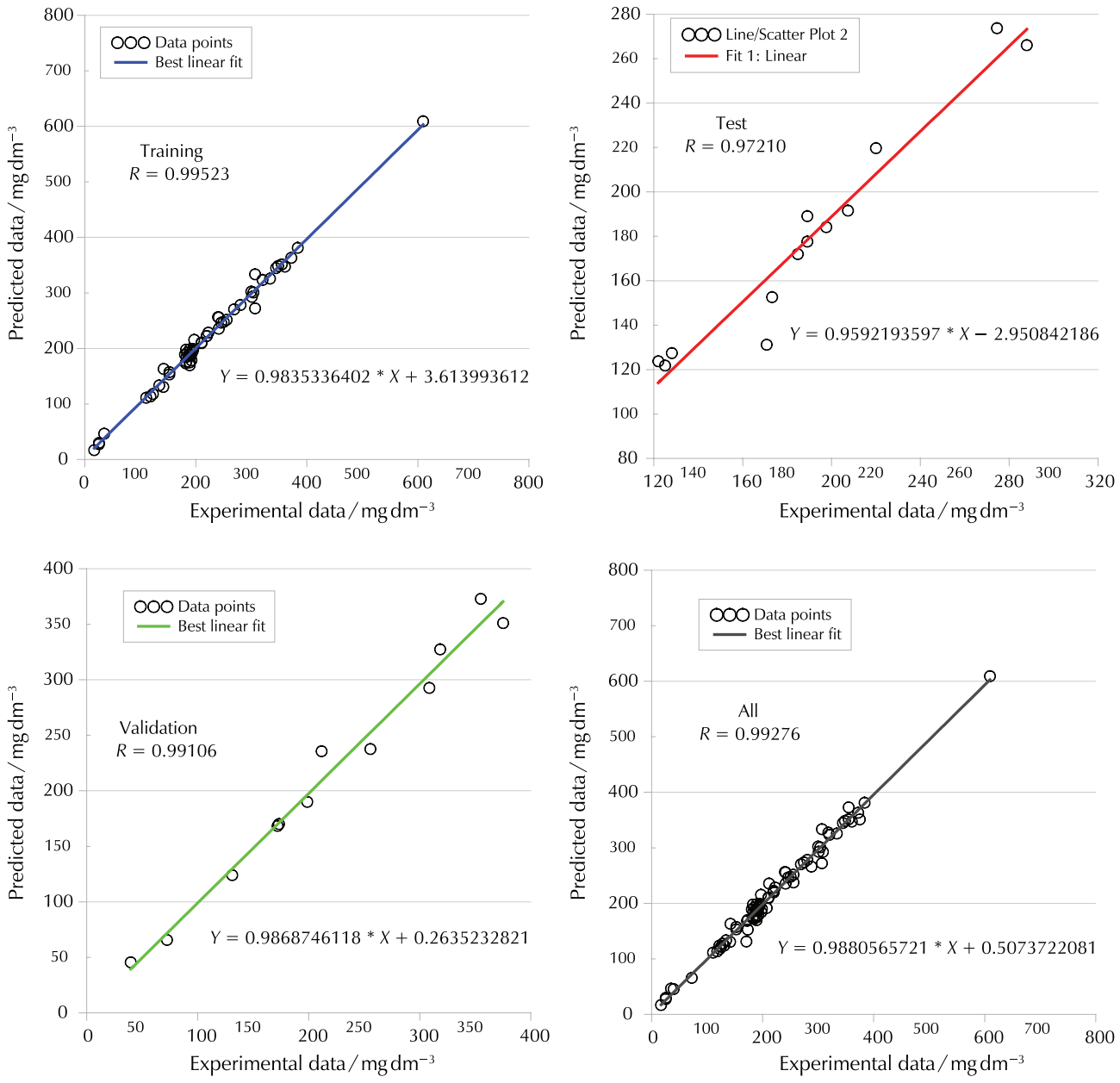


Fig. 8 – Comparison between experimental and predicted data estimated by ANN model

of the model was confirmed by the obtained coefficient of correlation $R = 0.99839$ with root mean square error RMSE = 0.01402 for the total data set.

4.1 Prediction performance

In order to test the precision of the previously developed and optimised ANN model, an interpolation was performed. For this purpose, a database was constructed containing a set of data points located in the middle between the experimental points of bicarbonate from drinking water in Médéa region. The results showing the regression bicarbonate between the predicted and experimental values, and the performance of the interpolation in terms of error and correlation coefficient are shown in Fig. 8.

Results of the prediction performances in terms of all errors and in terms of the agreement vector values [α = (slope), β (y intercept), R (correlation coefficient)] are summarised in Table 2.

These results show a good correlation between the predicted ANN and the experimental values with a high correlation coefficient ($R = 0.99276$).

Fig. 9 again shows the performance of the ANN configuration [18–5–1] with a better superposition of the curves plotted experimentally and those predicted by the model deduced during our study. The results obtained by this study show a good agreement explained by a high coefficient of correlation, which has a value close to unity. In addition, the values of errors including (MSE, RMSE, ESP, EPM, MAE) are very low.

Table 2 – Prediction performance results in terms of all errors and agreement vector [α, β, R]

	RMSE / mg dm^{-3}	MSE / mg dm^{-3}	ESP / %	EPM / %	MAE / mg dm^{-3}	R	α	β
Training	9.68913055	93.8792508	4.3826748	1.31881052	6.54764823	0.99523	0.9835336	3.6139936
Test	15.734532	247.575496	8.28240564	0.09771214	10.9639556	0.9721	0.9592193	−2.950842
Validation	14.0333691	196.935448	6.44782517	0.09058195	12.0114691	0.99106	0.9868746	0.2635232
All	11.526138	132.85185	5.3433443	3.13764295	8.02930904	0.99276	0.9880565	0.5073722

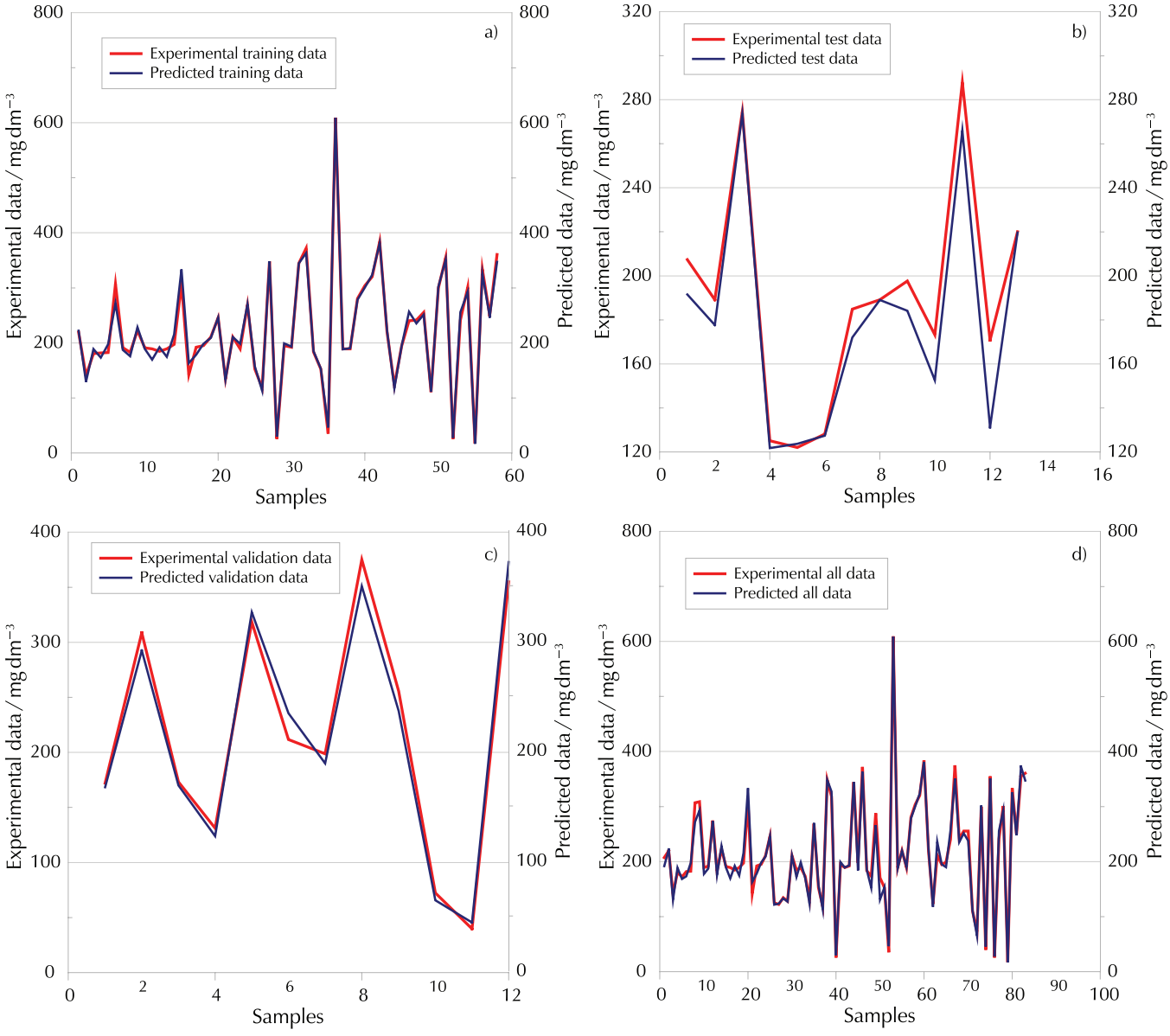


Fig. 9 – Relationship between experimental data and the predicted data of samples using ANN modelling: (a) training, (b) test, (c) validation, and (d) all data

4.2 Residues study

Errors made by the artificial neuron networks model on the samples that were used in this study are named residues.[30] Thus, the study of the relationship between the estimated contents of bicarbonate by the mathematical models and their residues ($y_{exp} - y_{pred}$) allowed us to ensure the performance of the model and verify them experimentally.

Fig. 10 shows the relationships between the estimated levels of bicarbonate in the water of Médéa region and their residues obtained using neural networks (ANN), respectively. It shows that the residues obtained by neural networks are much less dispersed (closer to zero), and a clear improvement in the distribution of residues depending on the nature of the samples.

This distribution proves the predictive power of this model established by the neural networks in the prediction of the contents of bicarbonate from the environmental parameters. In general, the result obtained is very satisfactory and justifies the use of the neural network approach in the prediction of bicarbonate contents in Médéa region.

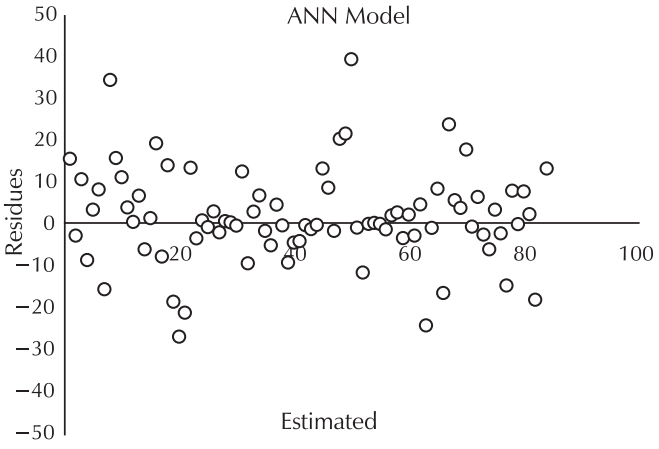


Fig. 10 – Residues relating to the models established by the artificial neural network depending on estimated values

5 Conclusion

In the present study, the prediction of bicarbonate concentrations in drinking water in the Médéa region was obtained using artificial neural networks (ANN) with supervised learning, involving the "Levenberg-Marquardt" algorithm. This algorithm gives better results in terms of speed, convergence, and generation of performance. The results showed a high learning and predictive capacity for bicarbonate concentrations with a very high correlation coefficient of 0.99276 and a very low mean square error (11.52613 mg dm^{-3}) for the whole database. In addition, they showed a better choice of learning algorithm, activation functions, and network architecture [18–5–1] obtained by applying statistical indicators of robustness. The prediction by artificial neural networks also shows a good correlation between the experimental and predicted values, meaning that the ANN model has better predictive power. This performance seems to be due to the fact that the concentrations of bicarbonates in drinking water in Médéa region are linked to the physicochemical characteristics of the environment by non-linear relationships.

This encourages us to consider in future the development of other aspects of this study related to the use of more parameters and prediction of other properties.

References

Literatura

1. *J. S. Torrecilla, M. L. Mena, P. Yáñez-Sedeño, J. García*, Application of artificial neural network to the determination of phenolic compounds in olive oil mill wastewater, J. Food Eng. **81** (3) (2007) 544–552, doi: https://doi.org/10.1016/j.jfoodeng.2006.12.003.
2. *J. Rude,* Développement d'un modèle statistique neuronal pour la description fine de la pollution atmosphérique par le dioxyde d'azote: application à la région Parisienne, Paris Est, 2008.
3. *P. Perez, A. Trier*, Prediction of NO and NO_2 concentrations near a street with heavy traffic in Santiago, Chile, Atmosph. Environ. **35** (10) (2001) 1783–1789, doi: https://doi.org/10.1016/S1352-2310(00)00288-0.
4. *F. Qaderi, E. Babanezhad,* Prediction of the groundwater remediation costs for drinking use based on quality of water resource, using artificial neural network, J. Clean. Prod. **161** (2017) 840–849, doi: https://doi.org/10.1016/j.jclepro.2017.05.187.
5. *V. Wagh, D. Panaskar, A. Muley, S. Mukate, S. Gaikwad,* Neural network modelling for nitrate concentration in groundwater of Kadava River basin, Nashik, Maharashtra, India, Groundwater Sustain. Develop. **7** (2018) 436–445, doi: https://doi.org/10.1016/j.gsd.2017.12.012.
6. *V. M. Wagh, D. B. Panaskar, A. A. Muley, S. V. Mukate, Y. P. Lolage, M. L. Aamalawar,* Prediction of groundwater suitability for irrigation using artificial neural network model: a case study of Nanded tehsil, Maharashtra, India, Model. Earth Syst. Environ. **2** (2016) 1–10, doi: https://doi.org/10.1007/s40808-016-0250-3.
7. *A. R. Khataee, M. B. Kasiri,* Artificial neural networks modeling of contaminated water treatment processes by homogeneous and heterogeneous nanocatalysis, J. Mol. Catal. A-Chem. **331** (2010) 86–100. https://doi.org/10.1016/j.molcata.2010.07.016.
8. *H. Vicente, S. Dias, A. Fernandes, A. Abelha, J. Machado, J. Neves* , Prediction of the quality of public water supply using artificial neural networks, J. Water Suppl. Res. Technol. – AQUA **61** (2012) 446–459, doi: https://doi.org/10.2166/aqua.2012.014.
9. *P. Mageshkumar, T. Pradeep, M. R. John, S. A. Raj, S. Anandakumar*, Neural network modelling of TDS concentrations in Cauvery River water, Tamilnadu, India, Int. J. Appl. Sci. Eng. Res. **1** (2012) 739–746, doi: https://doi.org/10.6088/ijaser.0020101074.
10. *J. Leifer, J. I. Mickalonis*, Prediction of aluminum pitting in natural waters *via* artificial neural network analysis, Corrosion **56** (2000) 563–571, doi: https://doi.org/10.5006/1.3280560.
11. *A. A. Nadiri, E. Fijani, F. T.-C. Tsai, A. Asghari Moghaddam*, Supervised committee machine with artificial intelligence for prediction of fluoride concentration, J. Hydroinform. **15** (2013) 1474–1490, doi: https://doi.org/10.2166/hydro.2013.008.
12. *E. Karadurmuş, N. Taşkın, E. Göz, M. Yüceer*, Prediction of bromate removal in drinking water using artificial neural networks, Ozone: Sci. Eng. **41**, (2019) 118–127, doi: https://doi.org/10.1080/01919512.2018.1510763.
13. *N. Mohsenifar, N. Mohsenifar, K. Mohsenifar*, Using Artificial Neural Network (ANN) for estimating rainfall relationship with river pollution, Adv. Environ. Biol. **5** (2011) 1202–1208.
14. *S. Nemati, L. Naghipour, M. H. F. Fard*, Artificial neural network modeling of total dissolved solid in the Simineh River, Iran, J. Civ. Eng. Urban. **4** (1) (2014) 8–13.
15. *A. H. Zare, V. M. Bayat, A. P. Daneshkare*, Forecasting nitrate concentration in groundwater using artificial neural network and linear regression models, Int. Agrophys. **25** (2011) 187–192.
16. *F. A. Pour, K. Mohsenifar, E. Pazira,* Affect of drought on pollution of Lenj station of Zayandehrood River by Artificial Neural Network (ANN), Adv. Environ. Biol. **5** (7) (2011)

1461–1464.

17. *A. Abdallaoui, H. El Badaoui*, Prédiction des teneurs en métaux lourds des sédiments à partir de leurs caractéristiques physico-chimiques, J. Phys. Chem. News **58** (2011) 90–97.
18. *L. Cavas, Z. Karabay, H. Alyuruk, H. Doğan, G. K. Demir*, Thomas and artificial neural network models for the fixed-bed adsorption of methylene blue by a beach waste *Posidonia oceanica* (L.) dead leaves, Chem. Eng. J. **171** (2) (2011) 557–562, doi: https://doi.org/10.1016/j.cej.2011.04.030.
19. *R. M. Aghav, S. Kumar, S. N. Mukherjee*, Artificial neural network modeling in competitive adsorption of phenol and resorcinol from water environment using some carbonaceous adsorbents, J. Hazard. Mater. **188** (1–3) (2011) 67–77, doi: https://doi.org/10.1016/j.jhazmat.2011.01.067.
20. *H. Ousmana, A. E. Hmaidi, M. Berrada, B. Damnati, I. Etabaai, A. Essahlaoui*, Development of a Neural Network approach for Predicting nitrate and sulfate concentration in three lakes: Ifrah, Iffer and Afourgagh, Middle Atlas Morocco, Mor. J. Chem. **6** (2) (2018) 245–255, url: https://revues.imist.ma/index.php?journal=morjchem&page=article&op=view&path%5B%5D=5939.
21. *M. Sediri, S. Hanini, H. Cherifi, M. Laidi, S. A. Turki,* Dynamic Adsorption Modelling of P-nitrophenol in Aqueous Solution Using Artificial Neural Network, J. Mater. Environ. Sci. **8** (7) (2017) 2282–2287.
22. *J. Rodier, B. Legube, N. Merlet, R. Brunet*, L'analyse de l'eau: Eaux naturelles, eaux résiduaires, eau de mer, Dunod, Paris, 2009, 1526 p.
23. *R. Yacef, A. Mellit, S. Belaid, Z. Şen*, New combined models for estimating daily global solar radiation from measured air temperature in semi-arid climates: application in Ghardaïa, Algeria, Energ. Convers. Manage. **79** (2014) 606–615, doi: https://doi.org/10.1016/j.enconman.2013.12.057.
24. *H. El Badaoui, A. Abdallaoui, I. Manssouri, L. Lancelot*, Elaboration de modèles mathématiques stochastiques pour la prédiction des teneurs en métaux lourds des eaux superficielles en utilisant les réseaux de neurones artificiels et la régression linéaire multiple, J. Hydrocarb. Mine. Environ. Res. **3** (2012) 31–36.
25. *I. Manssouri, M. Manssouri, B. El Kihel*, Fault detection by K-NN algorithm and MLP neural networks in a distillation column: Comparative study, J. Inf. Intellig. Knowl. **3** (3) (2011) 201.
26. *I. Manssouri, A. El Hmaidi, T. E. Manssouri, B. El Moumni*, Prediction levels of heavy metals (Zn, Cu and Mn) in current Holocene deposits of the eastern part of the Mediterranean Moroccan margin (Alboran Sea), IOSR J. Comp. Eng. **16** (2014) 117–123, doi: https://doi.org/10.9790/0661-1618117123.
27. *O. R. Dolling, E. A. Varas*, Artificial neural networks for streamflow prediction, J. Hydraul. Res. **40** (5) (2002) 547–554, doi: https://doi.org/10.1080/00221680209499899.
28. *D. A. Belsley, E. Kuh, R. E. Welsch*, Regression diagnostics: identifying influential data and sources of collinearity. Wiley, 1980, doi: https://doi.org/10.1002/0471725153.
29. *S. H. Hong, M. W. Lee, D. S. Lee, J. M. Park*, Monitoring of sequencing batch reactor for nitrogen and phosphorus removal using neural networks, Biochem. Eng. J. **35** (3) (2007) 365–370, doi: https://doi.org/10.1016/j.bej.2007.01.033.
30. *C. Voyant, M. Muselli, C. Paoli, M.-L. Nivet*, Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation, Energy **36** (1) (2011) 348–359, doi: https://doi.org/10.1016/j.energy.2010.10.032.

SAŽETAK

Predviđanje količine bikarbonata u pitkoj vodi regije Médéa modeliranjem umjetnom neuronskom mrežom

Hichem Tahraoui,[a] *Abd-Elmouneïm Belhadj*,[a*] and *Adhya-Eddine Hamitouche*[b]

Regija Médéa (Alžir) smještena na poljoprivrednom zemljištu zahtijeva veliku količinu pitke vode te je stoga analiza vode od iznimne važnosti. Da bi se ispitao razvoj kvalitete pitke vode u toj regiji, najprije je napravljen eksperimentalni protokol za dobivanje skupa podataka uzimajući u obzir nekoliko fizikalno-kemijskih parametara. Zatim je dobiveni skup podataka podijeljen na dva dijela za stvaranje umjetne neuronske mreže, gdje je 70 % skupova podataka upotrijebljeno za trening, a preostalih 30 % dodatno je podijeljeno na dva jednaka dijela: jedan za testiranje, a drugi za validaciju modela. Dobiveni inteligentni model procijenjen je kao funkcija koeficijenta korelacije najbližeg 1 i najnižeg korijena srednje kvadratne pogreške (RMSE). U ovom istraživanju upotrijebljen je skup od 84 podatkovnih točaka. Za modeliranje ANN-a upotrijebljeno je osamnaest parametara u ulaznom sloju, pet neurona u skrivenom sloju i jedan parametar u izlaznom sloju. Za skriveni i izlazni sloj upotrijebljeni su algoritam učenja Levenberg Marquardt (LM), logaritamski sigmoid i funkcija linearnog prijenosa. Rezultati dobiveni tijekom ovog istraživanja pokazali su koeficijent korelacije R = 0,99276 s korijenom srednje kvadratne pogreške RMSE = 11,52613 mg dm^{-3}. Ti rezultati pokazuju da je dobiveni model neuronske mreže dao daleko bolje rezultate, jer je točniji a njegova relativna pogreška je mala s koeficijentom korelacije blizu 1. Konačno, zaključeno je da taj model može učinkovito predvidjeti brzinu topljivosti bikarbonata u vodi za piće u regiji Médéa.

Ključne riječi
Pitka voda, fizikalno-kemijski parametri, bikarbonat, modeliranje, umjetne neuronske mreže

[a] *Laboratory of Biomaterials and Transport Phenomenon (LBMPT), Faculty of Technology, University Yahia Fares of Médéa, 26 000, Alžir*
[b] *Center for Scientific and Technical Research in Physico-chemical Analysis CRAPC, BP 384, Bou-Ismail, RP 42 004, Tipaza, Alžir*

Izvorni znanstveni rad
Prispjelo 1. siječnja 2020.
Prihvaćeno 19. veljače 2020.

جامعة زيان عاشور بالجلفة

الملتقى الدولي الأول للمواد، البيئـة و النظـم الميكانيكية والصناعية

30-29 جوان 2019

ICMEMIS'19

Djelfa on 29-30 June 2019,
Algeria

CERTIFICATE OF PARTICIPATION

This to certify that:

TAHRAOUI Hichem

Affiliation: Medea University, Medea, Algeria

has successfully presented a paper entitled: The application of the neural network method for the prediction of sulfate ion (SO4)2- from drinking water in the Médéa region

Co-Authored by: BELHADJ Abd-Elmouneïm, HAMITOUCHE Adh'ya-eddine

at The First International Conference on Materials, Environment, Mechanical and Industrial Systems.

General Chair

29-30 June 2019

ICMEMIS 2019

جامعة الجلفة - الجزائر

30-29 جوان 2019

Djelfa University
LAADT
Laboratoire d'Automatique Appliquée et Diagnostics Industriel

Joint Research Team
Djelfa University
Gas Turbine

Scanned by CamScanner

جامعة زيان عاشور بالجلفة

الملتقى الدولي الأول للمواد، البيئـة و النظـم الميكانيكية والصناعية

29-30 جوان 2019

ICMEMIS'19

Djelfa on 29-30 June 2019, Algeria

CERTIFICATE OF PARTICIPATION

This to certify that:

TAHRAOUI Hichem

Affiliation: Medea University, Medea, Algeria

has successfully presented a paper entitled: Artificial neural network model for the prediction of the bicarbonate amount in drinking water

Co-Authored by: BELHADJ Abd-Elmouneïm, BOUHEDDA Mounir, HAMITOUCHE Adhya-eddine

at The First International Conference on Materials, Environment, Mechanical and Industrial Systems.

General Chair

Scanned by CamScanner

Conférence Internationale sur les Sciences et Génie des Matériaux et leurs Impacts sur l'Environnement (ICMSE-19)
Sidi Bel-Abbès, 26-28 novembre 2019

جامعة الجيلالي اليابس بسيدي بلعباس
Université Djilali Liabès de Sidi Bel-Abbès
Faculté des Sciences Exactes
Département Physique

المؤتمر الدولي لعلوم المواد والهندسة وتأثيرهما على البيئة

تشرين الثاني 26-28/ نوفمبر 2019 ، سيدي بلعباس

Attestation de participation

Le comité scientifique de la conférence Internationale sur les Sciences et Génie des Matériaux et leurs impacts sur l'Environnement (ICMSE-19) , atteste que ***Malika AMARI*** *a participé(e) aux travaux de la conférence du 26 au 28 novembre 2019 et a présenté(e) une communication par affiche (poster) intitulée :*

Study of the modeling impact of raw water sulphate on the environment .

Co-auteurs : ***Hichem TAHRAOUI, Abdelmouneïm BELHADJ and Adhya-eddine HAMITOUCHE***

THE INTERNATIONAL CONFERENCE ON MATERIALS SCIENCE AND ENGINEERING AND THEIR IMPACT ON THE ENVIRONMENT
ICMSE-2019 - November 26-28,2019

P/Le Comité Scientifique
Le Président de la Conférence
Prof. M. Ameri

Président de Comité d'Organisation
Professeur AMERI Mohammed

Laboratoire Physico Chimie des Matériaux Avancés (LPCMA), Université Djilali Liabès de Sidi Bel-Abbès
Tél. : 048 79 90 02/048 74 56 19, Email: lltnsameri@yahoo.fr , https://www.univ-sba.dz/lpcma/

Scanned by TapScanner

الشعبية الديمقراطية الجزائرية لجمهورية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي و البحث العلمي
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
جامعة سعد دحلب البليدة1
UNIVERSITE SAAD DAHLAB BLIDA 1
FACULTE DES SCIENCES DE LA NATURE ET E LA VIE
Département de Biologie et physiologie Cellulaire (BPC) & Biologie et Physiologie des organismes (BPO)

Attestation De Communication

Journées Nationales virtuelles de Bioinformatique génomique et protéomique

Webinaires, JNBGP2020
9 et10 Octobre 2020

Le comité scientifique des Journées Nationales Virtuelles de Bioinformatique Génomique et protéomique (JNBGP2020) atteste que : ***Dr. Hichem TAHRAOUI***

A présenté un Poster intitulé « ***IDENTIFICATION DEUX MICROALGUES*** *» qui s'est tenu en ligne sous forme de Webinaires « visioconférences et débats », le 9 et 10 octobre 2020*

Co-auteurs : *Abd-Elmouneïm BELHADJ*

UNIVERSITE SAAD DAHLAB - BLIDA1 • JNBGP2020-FSNV-USDB • FACULTE DES SCIENCES DE LA NATURE ET DE LA VIE • FSNV •

Dr. EDDAIKRA Atika
Présidente
JNBGP-FFSNV-USDB

Algerian Symposium on Renewable Energy and Materials
ASREM2020, December 16-17, 2020 Médéa- Algeria
http://www.asrem2020.cf/

December 17th, 2020

CERTIFICATE

We hereby certify that

H. TAHRAOUI

University of Médéa, Médéa – Algeria

Has presented the paper referenced below in Algerian Symposium on Renewable Energy and Materials (ASREM2020), which took place in Médéa, Algeria from 16th to 17 th, December 2020

Title : **Predictive model based on a support vector machine for the estimation of the adsorption of clay diclofenac**

Co-author(s) : A. BELHADJ, K. SEGDI, B. IGHIRIZANE, M. HENTABLI, B. SOUMATI and R. FARGANI

Signature ASREM2020 Chair

ASREM 2020
General Chair
Dr. Younes Chiba

Algerian Symposium on Renewable Energy and Materials
ASREM2020, December 16-17, 2020 Médéa- Algeria
http://www.asrem2020.cf/

December 17th, 2020

CERTIFICATE

We hereby certify that

H. TAHRAOUI

University of Médéa, Médéa – Algeria

Has presented the paper referenced below in Algerian Symposium on Renewable Energy and Materials (ASREM2020), which took place in Médéa, Algeria from 16th to 17 th, December 2020

Title : **Predicting the diclofenac elimination rate by adsorption using machine vector support**

Co-author(s) : A. BELHADJ, Y. AYACHE, A. OULD LARBI, M. HENTABLI, B. SOUMATI and M. BENHAOUA

Signature ASREM2020 Chair

ASREM 2020
General Chair
Dr. Younes Chiba

PREMIER SEMINAIRE NATIONAL
DE L'ENVIRONNEMENT ET DES PHENOMENES DE TRANSPORT
SNEPT20, 03-04 Mars 2020

ATTESTATION DE PARTICIPATION

Cet attestation certifie que,

TAHRAOUI Hichem

A participé par une communication par affiche intitulée,

Optimization of the algorithm to soluble predict dry residue by artificial neural network

Au Premier Séminaire National de l'Environnement e énomènes de Transport (SNEPT20), organisé le 03-04 Mars 2020 à l'Université de Médéa.

Dr. FERHAT Samira
La présidente du SNEPT20

General Chair of SNEPT 20
Dr. FERHAT Samira

Université Yahia Fares de Médéa
SNEPT 20
03-04 Mars 2020
Faculté de Technologie