

الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE

المدرسة الوطنية العليا للفلاحة
الحراش - الجزائر
ECOLE NATIONALE SUPERIEURE AGRONOMIQUE (ENSA)
El Harrach – ALGER

THESE

En vue de l'obtention du diplôme de doctorat en sciences agronomiques

Thème

Modélisation de l'eutrophisation des eaux de surface

Réalisé par : BELOUZ Khaled

Soutenu le : 11 juin 2019

Devant le jury composé de :

Président du jury : Professeur HARTANI Tarik (Centre Universitaire de Tipaza)

Directeur de thèse : Professeur AIDAOUI Abdellah (ENSA d'Alger)

Examineurs : Professeur CHABACA Mohamed Nacer (ENSA d'Alger)

Professeur SOUAG-GAMANE Doudja (USTHB)

Année universitaire 2018/2019

Remerciement

Au terme de ce travail, je remercie Allah le tout Puissant qui m'a donné la force, la patience ainsi que le courage afin de parvenir à terminer ce dernier.

Je tiens à remercier mon directeur de thèse Monsieur **A. Aidaoui**, Professeur à l'Ecole Nationale Supérieure Agronomique (ENSA) d'avoir accepté de diriger cette thèse, de m'avoir encadré, il a été pour moi un enseignant privilégié de par sa gentillesse, son expérience, je le remercie pour sa rigueur, ses encouragements ainsi que pour ses critiques et conseils constructifs pour aboutir à ce travail.

Je remercie aussi Monsieur **T. Hartani**, Professeur au centre universitaire de Tipaza d'avoir accepté de présider mon jury de soutenance mais pour tout ce qu'il a fait pour la réussite de la promotion de post-graduation hydraulique agricole (2006-2007), il mérite tous les remerciements.

Mes remerciements vont aussi à monsieur **MN. Chabaca**, Professeur à l'Ecole Nationale Supérieure Agronomique (ENSA), à Madame **D. Souag-Gamane**, Professeur à l'USTHB, d'avoir donné de leurs temps, malgré leurs obligations, afin d'examiner et enrichir ce travail par leurs précieuses remarques et suggestions.

En fin je remercie toute personne ayant contribué de près ou de loin dans l'élaboration de ce travail.

ملخص

تعتبر جودة المياه السطحية أحد العوامل الرئيسية المستخدمة في إدارة الموارد المائية، إلا أن هذه الأخيرة لم تحظ بالاهتمام والعناية التي يستحقها.

نظرا إلى درجة تعقيد النظام الإيكولوجي المائي وعدم انتظامه، تنوع أنواع الطحالب ، سلوكها، دوراتها الموسمية التي تختلف من نوع إلى آخر، تفاعل أنواع الطحالب مع بيئتها و التفاعلات بين الأنواع، الفهم غير الكافي للعمليات ، تعدد الآليات المعنية ، نقص البيانات ذات الجودة ، والتغير الموسمي الكبير لأحمال الملوثات التي يتم تصريفها في خزانات السدود ، فإن كل هذه العوامل تجعل من تطوير نماذج تتنبأ بنوعية المياه وظاهرة التخثث تحديا كبيرا .

لقد أتاح ظهور تقنيات الذكاء الاصطناعي العديد من النتائج المشجعة في مجال إدارة ونمذجة نوعية المياه . يهدف هذا البحث إلى المساهمة في تطوير نماذج تستند على البيانات من أجل دراسة ظاهرة التخثث لخزانات السدود . تمت مناقشة بعض الاستنتاجات والتوصيات لهذه الدراسة في هذا الملخص . يمكن لنماذج الشبكات العصبونية الاصطناعية (ANN) وأنظمة الاستدلال الضبابي العصبي التكيفي (ANFIS) باستخدام بيانات نوعية المياه المتاحة التنبؤ بتركيز الكلوروفيل-الفا مع الدقة المقبولة باستخدام بيانات نوعية المياه المتاحة . ويمكن استخدام النماذج التي تم تطويرها كأدوات تنبؤية (محاكاة) لتركيز الكلوروفيل في دراسات سيناريوهات التخفيف من المغذيات في الأحواض الهيدروغرافية . كما أن نماذج الشبكات العصبونية الاصطناعية وأنظمة الاستدلال الضبابي العصبي التكيفي (ANFIS) قادرة على التنبؤ بدقة عالية ببعض مؤشرات ظاهرة التخثث مثل الفسفور الكلي والطلب الأحيائي الكيميائي للأكسجين . من أجل الحصول على نمذجة أكثر دقة لظاهرة التخثث في الخزانات ، يوصى بتخصيص فترات أقصر لأخذ العينات قصد تحسين نتائج النمذجة .

الكلمات الدالة:

ظاهرة تخثث المياه ; الشبكات العصبونية الاصطناعية ; أنظمة الاستدلال الضبابي العصبي التكيفي ; النمذجة ; جودة المياه ; الكلوروفيل ; الفوسفور الكلي ; الطلب الأحيائي الكيميائي للأكسجين ; النظام البيئي لخزانات السدود .

Résumé

La qualité des eaux de surface est l'un des facteurs clé utilisé dans la gestion et l'aménagement des ressources en eau, cependant, cet élément n'a pas reçu la considération et l'attention qu'il mérite.

Vu la complexité élevée et la non-linéarité de l'écosystème aquatique, la diversité des espèces d'algues, leurs comportement, leurs cycles saisonniers différents d'une espèce à l'autres, l'interaction des espèces d'algues avec leur environnement, les interactions entre les espèces, l'insuffisance de la compréhension des processus, la multiplicité des mécanismes impliqués, le manque de données de bonne qualité, et la haute variabilité saisonnière des charges polluantes déversées dans les retenues de barrages rendent le développement de modèles prédisant la qualité des eaux de barrages et l'eutrophisation un défi.

L'émergence des techniques de l'intelligence artificielle a donné de nombreux résultats encourageants dans le domaine de la gestion et la modélisation de la qualité de l'eau.

Cette recherche a pour objectif de contribuer au développement de modèles basés sur les données (data driven model) pour l'étude de l'eutrophisation de retenues de barrages. Certaines conclusions et recommandations de cette étude sont discutées dans ce résumé. Les réseaux de neurones artificiels (RNAs) et les systèmes d'inférence neuro-flou adaptatifs (ANFIS) permettent de modéliser avec une précision acceptable la concentration de chlorophylle-a en utilisant les données de la qualité de l'eau disponibles. Les modèles développés peuvent être utilisés comme des outils de simulation de la concentration de chlorophylle-a dans les études de scénarii de réduction des nutriments dans le bassin versant. Les modèles de RNAs et d'ANFIS sont également capables de simuler, avec une précision élevée, certains indicateurs de l'eutrophisation tels que le phosphore total et la demande biochimique en oxygène.

Pour une modélisation plus fine de l'eutrophisation des retenues de barrages, des intervalles d'échantillonnage plus courts (courts pas du temps) sont recommandés pour améliorer les résultats de la modélisation.

Mots clés:

Eutrophisation; Réseaux de neurones artificiels; ANFIS ; Modélisation; qualité de l'eau; chlorophylle-a; phosphore total; Demande biochimique en oxygène ; écosystème de retenue de barrage.

Abstract

The surface water quality is one of the key factors used in the management of water resources; however, this element has not received the consideration and attention it deserves.

Owing to the high complexity and non-linearity of the aquatic ecosystem, the diversity of algae species, their behavior, their seasonal cycles different from one species to another, the interaction of algal species with their environment, interactions between species, insufficient understanding of the processes, the multiplicity of mechanisms involved, the lack of good quality data, and the high seasonal variability of pollutant loads discharged into dam reservoirs make the development of models predicting reservoir water quality and eutrophication a challenge.

The emergence of artificial intelligence techniques has provided many encouraging results in the field of water quality management and modelling.

The objective of this research is to contribute to the development of data driven models for the study of eutrophication in reservoirs. Some conclusions and recommendations of this study are discussed in this abstract. Artificial neural networks (ANNs) and adaptive neuro-fuzzy inference systems (ANFIS) predict chlorophyll-*a* concentration with acceptable accuracy using available water quality data. The models developed can be used as tools for predicting (simulating) the concentration of chlorophyll-*a* in watershed nutrient reduction scenario studies. ANN and ANFIS models are also able to predict, with high accuracy, some indicators of eutrophication such as total phosphorus and biochemical oxygen demand.

For finer modelling of reservoir eutrophication, shorter sampling intervals at lower time steps are recommended to improve modelling results.

Keywords:

Eutrophication; Artificial neural networks; ANFIS; Modelling; water quality; chlorophyll-*a*; total phosphorus; biochemical oxygen demand; reservoir ecosystem.

Sommaire

Introduction générale et problématique	1
I. L'écosystème aquatique de retenue de barrage (réservoir).....	6
I.1 Introduction.....	6
I.2. Caractéristique physique de l'écosystème de retenue de barrage.....	7
I.2.1. Morphologie des barrages	7
I.2.1.1. La zone fluviale	7
I.2.1.2. Zone de transition	7
I.2.1.3. La zone lacustre	8
I.3. Caractéristiques abiotiques de l'écosystème de retenue de barrage.....	8
I.3.1. Substances organiques	8
I.3.2. Substances inorganiques.....	9
I.3.2.1. Les cations	10
I.3.2.2. Les anions	10
I.3.2.3. L'azote	10
I.3.2.4. Le phosphore.....	11
I.4. Caractéristiques biotiques de l'écosystème de retenue de barrage.....	12
I.4.1. Zones biologiques dans les lacs et les réservoirs	12
I.4.1.1. La zone littorale	13
I.4.1.2. La zone pélagique (limnétique)	13
I.4.1.3. La zone benthique	13
I.4.2. La chaîne alimentaire	14
I.4.3. Les producteurs primaires aquatiques.....	14
I.5. Eutrophisation	15
I.5.1. Introduction.....	15
I.5.2. Assimilation et nutriments limitants	17
I.5.3. Les facteurs de l'environnement.....	18
I.5.3.1. La Lumière	18
I.5.3.2. La température.....	18
I.5.3.3. Le temps de résidence de l'eau	19
I.5.4. Algues	19
I.5.4.1. Estimation de l'abondance de la biomasse algale.....	21
I.5.5. Effets de l'eutrophisation sur la qualité de l'eau.....	21

I.5.6. La méthode d'évaluation de l'eutrophisation	22
I.5.6.1. Introduction.....	22
I.5.6.2. Evaluation qualitative de degré d'eutrophisation.....	22
I.5.6.3. Evaluation quantitative du degré d'eutrophisation	23
I.5.6.3.1. L'état trophique.....	23
I.5.6.3.2. Indice de l'état trophique.....	24
I.5.6.3.3. Les indicateurs de l'eutrophisation	25
I.5.6.3.3.1. Le phosphore total :	25
I.5.6.3.3.2. La chlorophylle <i>alpha</i>	26
I.5.6.3.3.3. La transparence de l'eau	26
I.5.6.3.3.4. L'oxygène dissous.....	26
I.5.6.3.3.5. La demande biochimique en oxygène.....	27
I.5.7. Eutrophisation en Algérie : synthèse bibliographique	27
II. Modélisation de la qualité de l'eau et l'eutrophisation	29
II.1 Introduction.....	29
II.2 Les objectifs de modélisation de la qualité des eaux	29
II.3. Objectif de la prédiction de la qualité des eaux	30
II.4. Approche de modélisation de la qualité de l'eau.....	31
II.4.1. Approches conventionnelles de modélisation de la qualité de l'eau.....	31
II.4.1.1 Les modèles physiques	31
II.4.1.2 Les modèles empiriques (statistiques)	31
II.4.1.3 Les modèles mécanistes.....	32
II.4.1.3.1. Classification des modèles mécanistes de la qualité des eaux	33
II.4.1.4. Modélisation de l'eutrophisation en utilisant les approches conventionnelles: synthèse bibliographique	34
II.4.2. Approches basées sur l'intelligence artificielle	36
II.5 Justification de l'utilisation de l'intelligence artificielle pour modéliser la qualité des eaux et l'eutrophisation	38
II.6 Modélisation de la qualité de l'eau par l'intelligence artificielle : synthèse bibliographique.....	39
II.7 Modélisation de l'eutrophisation par l'intelligence artificielle : synthèse bibliographique	41
III. Matériels et méthodes	47
III. 1. La méthodologie de travail.....	47
III.2. Réseaux de Neurones Artificiels.....	48
III.2.1. Introduction.....	48
III.2.2. Historique des réseaux de neurones artificiels	49

III.2.3. Quelques domaines d'application des réseaux de neurones artificiels	50
III.2.4. Réseau de neurones: fondement biologique	50
III.2.4.1. Le neurone.....	50
III.2.4.2. Le corps cellulaire	51
III.2.4.3. Les dendrites	51
III.2.4.4. L'axone	51
III.2.5. Réseaux de neurones artificiels.....	52
III.2.5.1. Principe général.....	52
III.2.5.2. Structure d'un neurone artificiel (formel).....	53
III.2.5.3. Fonctionnement mathématique du neurone artificiel	54
III.2.5.3.1. Fonction de base (fonction discriminant)	54
III.2.5.3.2. Fonction dynamique	55
III.2.5.4. Classification des réseaux de neurones artificiels.....	55
III.2.5.4.1. Architecture générale des réseaux de neurones	55
III.2.5.4.1.1. Réseaux de neurones dynamiques.....	56
III.2.5.4.1.2. Réseaux de neurones non bouclés ou en couches (Perceptron)	57
III.2.5.4.1.2.1. Le Perceptron monocouche	57
III.2.5.4.1.2.2. Le Perceptron multicouches.....	57
III.2.5.4.1.2.3. Perceptron Multi-Couches (PMC) et approximation de fonctions.....	58
III.2.5.4.2. Apprentissage des réseaux de neurones.....	59
III.2.5.4.2.1. Apprentissage non supervisé	59
III.2.5.4.2.2. L'apprentissage semi-supervisé ou par renforcement.....	60
III.2.5.4.2.3. Apprentissage supervisé des réseaux de neurones artificiels.....	60
III.2.5.4.2.3.1. Algorithmes d'apprentissage supervisé	60
III.2.5.4.2.3.1.1. La Backpropagation	61
III.2.5.4.2.3.1.1.a. Algorithme de Backpropagation.....	61
III.2.5.4.2.3.1.1.b. Méthode d'optimisation avec la Backpropagation	63
III.2.5.5. Avantages et inconvénients des réseaux de neurones artificiels.....	64
III.2.5.5.1. Avantage des réseaux de neurones artificiels	64
III.2.5.5.2. Inconvénients des réseaux de neurones artificiels	65
III.2.5.6. Choix de paramètres du modèle neuronal.....	66
III.2.5.6.1. Choix d'un réseau de neurones.....	66
III.2.5.6.2. Choix de nombre de couches cachées	66
III.2.5.6.3. Choix de nombre de neurones par couche cachée	66

III.2.5.6.4. Paramètres d'apprentissage	67
III.2.5.7. Conclusion	67
III.3. Introduction a la logique floue	67
III.3.1. Introduction.....	67
III.3.2. Historique	69
III.3.3. Concepts de base de l'approche de la logique floue.....	69
III.3.3.1. Univers du discours	70
III.3.3.2. Les variables linguistiques	70
III.3.3.3. Les ensembles flous.....	71
III.3.3.3.1. Les opérations sur les ensembles flous	72
III.3.3.4. Fonctions d'appartenance.....	72
III.3.3.5. Les règles fous (Si.....Alors...)	73
III.3.4. Systèmes d'inférence floue (SIF)	74
III.3.4.1. Etapes de modélisation par les systèmes d'inférence flous	75
III.3.4.1.1. La fuzzification	75
III.3.4.1.1.1. Attribution d'étiquettes floues.....	75
III.3.4.1.1.2. Assignation de signification numérique	76
III.3.4.1.2. L'Inférence floue en utilisant les règles SI-ALORS	76
III.3.4.1.3. Implication des règles floues.....	77
III.3.4.1.4. Agrégation de différentes sorties de règles	77
III.3.4.1.5. Défuzzification	77
III.3.4.1.5.1. La technique du centre de gravité.....	78
III.3.4.2. Classification de systèmes d'inférence floue	78
III.3.4.2.1. Systèmes d'inférence floue créés par les experts	79
III.3.4.2.1.1. Modèles de Mamdani	79
III.3.4.2.1.2. Modèles de Takagi-Sugeno-Kang	81
III.3.4.2.2. Systèmes d'inférence floue créés par les données: ANFIS.....	82
III.3.4.2.2.1. Introduction.....	82
III.3.4.2.2.2. Structure.....	82
III.3.4.2.2.3. Apprentissage d'ANFIS	84
III.3.4.2.2.3.1. Subtractive clustering.....	85
III.3.4.2.2.3.2. Grid partition (Partitionnement en grille)	86
III.4. Phases d'élaboration d'un modèle.....	87
III.4.1. La collecte des données.....	87

III.4.2. Traitement de données	87
III.4.2.1. Analyse des données.....	87
III.4.2.2. Normalisation des données.....	87
III.4.2.3. La division de la base de données (apprentissage-validation)	88
III.4.3. La sélection des variables d'entrées.....	88
III.4.3.1. La corrélation statistique.....	88
III.4.3.2. L'analyse de sensibilité	89
III. 4.4. Elaboration de la structure du modèle.....	91
III. 4.5. Apprentissage.....	91
III.4.6. Validation	91
III. 4.7. Post validation ou phase de test	91
III.5. Implémentation des modèles	92
III.6. Critères d'évaluation de performance des modèles	92
III.6.1. Critère graphique.....	93
III.6.2. Critères statistiques	93
III.6.2.1. Coefficient de corrélation (CC).....	93
III.6.2.2. Erreur quadratique moyenne (RMSE).....	93
III.6.2.3. Erreur absolue moyenne (MAE)	94
III.7. Description des réservoirs étudiés et les données de qualité de l'eau.....	94
III.7.1. Etude de cas n° 01 : base de données mixte (barrages Lekhal, Tichy-Haf et Keddara)	94
IV.7.1.1. Justification de l'utilisation d'une base de données mixte	94
III.7.1. 2. Barrage Lekhal	95
III.7.1. 2.1. Situation géographique du barrage.....	95
III.7.1.2.2. Données de la qualité de l'eau utilisées	96
III.7.1.3. Barrage de Tichy Haf	98
III.7.1.3.1. Situation géographique du barrage	98
III.7.1.3.2. Données de la qualité de l'eau utilisées	98
III.7.1.4. Barrage de Keddara	100
III.7.1.4.1. Situation géographique du barrage	100
III.7.1.4.2. Données de la qualité de l'eau utilisées	101
III.7.2. Etude de cas n° 02 : barrage de Ouizert.....	102
III.7.2.1. Situation géographique du barrage de Ouizert.....	102
III.7.2.2. Données de la qualité de l'eau utilisées	103

IV. Résultats et discussions.....	106
IV.1. Introduction	106
IV.2. Modélisation de l'eutrophisation par les réseaux de neurones artificiels.....	107
IV.2.1. Première étude de cas: Simulation de la chlorophylle-a par les réseaux de neurones artificiels	107
IV.2.1.2. Développement du modèle	107
IV.2.1.2.1. Variables d'entrées (variables candidats)	107
IV.2.1.2.2. La sortie cible: la concentration en chlorophylle alpha (Chl-a).....	107
IV.2.1.2.3. Apprentissage.....	108
IV.2.1.2.4. La sélection des variables d'entrées	108
IV.2.1.2.4.1. La matrice de corrélations.....	109
IV.2.1.2.4.2. Analyse de sensibilité	109
IV.2.1.2.5. Le choix du modèle	111
IV.2.1.2.6. Amélioration du modèle développé	114
IV.2.2. Deuxième étude de cas: barrage de Ouizert.....	117
IV.2.2.1.2. Développement des modèles	117
IV.2.2.2. Simulation de phosphore total par les réseaux de neurones artificiels.....	118
IV.2.2.2.1. Introduction	118
IV.2.2.2.2. Elaboration des modèles.....	118
IV.2.2.2.3. Résultats de simulation	119
IV.2.2.3. Simulation de la demande biochimique en oxygène par les réseaux de neurones artificiels	122
IV.2.2.3.1. Introduction: Relation entre l'eutrophisation et la DBO ₅	122
IV.2.2.3.2. Elaboration des modèles.....	123
IV.2.2.3.3. Résultats de simulation	124
IV.3. Modélisation de l'eutrophisation par les systèmes d'inférence neuro-flous adaptatifs (ANFIS), comparaison avec les modèles de RNAs	128
IV.3.1. Introduction	128
IV.3.2. Première étude de cas: Simulation de la de chlorophylle-a par les modèles ANFIS	129
IV.3.2.1. Développement des modèles	129
IV.3.2.1. 1. Variables d'entrées/sortie.....	129
IV.3.2.1. 2. Apprentissage d'ANFIS	129
IV.3.2.1.3. Implémentation des modèle ANFIS	130
IV.3.2.2. Résultats de simulation	131
IV.3.2.3. Comparaison entre les résultats des modèles ANFIS et ceux de RNAs	135

IV.3.3. Deuxième étude de cas: barrage de Ouizert.....	137
IV.3.3.1. Simulation du phosphore total par les systèmes d'inférence neuro-flous adaptatifs..	137
IV.3.3.1.1. Introduction	137
IV.3.3.1.2. Développement des modèles	137
IV.3.3.1.3. Résultats de simulation.....	137
IV.3.3.1.4. Comparaison entre les résultats des modèles ANFIS et ceux de RNAs	140
IV.3.3.2. Modélisation de la demande biochimique en oxygène (DBO ₅) par les systèmes d'inférence neuro-flous adaptatifs.....	140
IV.3.3.2.1. Introduction	140
IV.3.3.2.2. Développement des modèles	141
IV.3.3.2.2.1. Variables d'entrées/sortie.....	141
IV.3.3.2.2.2. Implémentation des modèles ANFIS.....	141
IV.3.3.2.3. Résultats de simulation	142
IV.3.3.2.4. La comparaison entre les modèles ANFIS et les modèles de RNAs	145
IV.3.3.2.5. Conclusion	145
Conclusion générale	147
Références bibliographiques	152

Liste des figures

Figure 1 Zonage longitudinales dans les réservoirs (Daniel et al., 2005).....	8
Figure 2 Le cycle de l'azote dans l'écosystème aquatique (Ji, 2008)	11
Figure 3 Le cycle du phosphore dans l'écosystème aquatique (Ji, 2008)	12
Figure 4 Zones biologiques dans les lacs (Ji, 2008).....	13
Figure 5 Les causes et les effets de l'eutrophisation (Thomas et al., 2006).....	16
Figure 6 Cinétique algal (Ji, 2008).....	20
Figure 7 Neurone biologique.....	51
Figure 8 Modèle neuronal	52
Figure 9 Mise en correspondance neurone biologique / neurone artificiel	53
Figure 10 Neurone formel	54
Figure 11 Différents types de fonctions de transfert	55
Figure 12 Réseau de neurones artificiel	56
Figure 13 Réseau bouclé	56
Figure 14 Réseaux à une seule couche	57
Figure 15 Réseaux "FeedForward" multicouches	58
Figure 16 Rétro- propagation des gradients	61
Figure 17 La variable linguistique température.	71
Figure 18 Comparaison d'un ensemble classique (A) et d'un ensemble floue (B).....	71
Figure 19 Différents types de fonctions d'appartenance.....	73
Figure 20 Blocs fonctionnels d'un système d'inférence floue.....	75

Figure 21 Etiquettes pour la variable linguistique taille	76
Figure 22 Défuzzification par la méthode du centre de gravité.....	78
Figure 23 Schéma du système d'inférence flou de Mamdani.....	80
Figure 24 Schéma du système d'inférence flou de Sugeno.	81
Figure 25 Structure D'ANFIS.....	83
Figure 26 Situation géographique du barrage Lekhal	95
Figure 27 Situation géographique du barrage Tichy-Haf.....	98
Figure 28 Situation géographique du barrage Keddara	100
Figure 29 Situation géographique du barrage Ouizert.....	103
Figure 30 Schéma général du modèle neuronal développé pour estimer la chlorophylle-a.....	112
Figure 31 Diagramme de dispersion par groupe de valeurs de chlorophylle-a observées et celles simulées par le RNA pour l'ensemble de données (mélangées)	112
Figure 32 (a1) Valeurs mesurées et simulées de la chlorophylle-a ($\mu\text{g} / \text{L}$); (b1) Diagramme de dispersion des valeurs simulées et mesurées de la chlorophylle-a ($\mu\text{g}/\text{L}$) pour l'ensemble de données en utilisant le modèle neuronal M0 (16-23-1).....	113
Figure 33 (a2) Valeurs mesurées et simulées de la chlorophylle-a ($\mu\text{g} / \text{L}$); (b2) Diagramme de dispersion des valeurs simulées et mesurées de la chlorophylle-a ($\mu\text{g}/\text{L}$) lors de l'apprentissage pour modèle neuronal M0 (16-23-1).	114
Figure 34 (a3) Valeurs mesurées et simulées de chlorophylle-a ($\mu\text{g} / \text{L}$); (b3) Diagramme de dispersion des valeurs simulées et mesurées de la chlorophylle-a ($\mu\text{g}/\text{L}$) lors de la validation pour le modèle neuronal M0 (16-23-1).	114
Figure 35 (a1) Valeurs mesurées et simulées de phosphore total (mg / L); (b1) Diagrammes de dispersion des valeurs simulées et mesurées de phosphore total (mg/L) pour ensemble complet de données en utilisant le modèle M0 (12-8-1).....	121
Figure 36 (a2) Valeurs mesurées et simulées de phosphore total (mg/L); (b2) Diagrammes de dispersion des valeurs simulées et mesurées de phosphore total (mg/L) lors de l'apprentissage pour le modèle M0 (12-8-1)	121
Figure 37 (a3) Valeurs mesurées et simulées de phosphore total (mg / L); (b3) Diagrammes de dispersion des valeurs simulées et mesurées de phosphore total (mg/L) lors de la validation en utilisant le modèle M0 (12-8-1)	121
Figure 38 (a1) Valeurs mesurées et simulées de la DBO_5 ($\text{mg O}_2 / \text{L}$); (b1) Diagramme de dispersion des valeurs simulées et mesurées de la DBO_5 ($\text{mg O}_2 / \text{L}$) pour l'ensemble complet de données en utilisant le modèle M0 (12-5-1).....	127
Figure 39 (a2) Valeurs mesurées et simulées de DBO_5 ($\text{mg O}_2 / \text{L}$); (b2) Diagramme de dispersion des valeurs simulées et mesurées de la DBO_5 ($\text{mg O}_2 / \text{L}$) lors de l'apprentissage en utilisant le modèle M0 (12-5-1)	127
Figure 40 (a3) Valeurs mesurées et simulées de la DBO_5 ($\text{mg O}_2 / \text{L}$); (b3) Diagramme de dispersion des valeurs simulées et mesurées de la DBO_5 ($\text{mg O}_2 / \text{L}$) lors de la validation en utilisant le modèle M0 (12-5-1).	127
Figure 41 Schéma général du modèle ANFIS développé pour estimer la chlorophylle-a.....	130
Figure 42 Diagramme de dispersion par groupe de valeurs de chlorophylle-a observées et celles simulées par ANFIS-M5 pour l'ensemble de données.	133
Figure 43 (a1) Valeurs mesurées et simulées de la chlorophylle-a ($\mu\text{g} / \text{L}$); (b1) Diagramme de dispersion des valeurs simulées et mesurées de la chlorophylle-a ($\mu\text{g}/\text{L}$) pour l'ensemble de données en utilisant le modèle ANFIS-M5.....	134

Figure 44 (a2) Valeurs mesurées et simulées de Chl-a($\mu\text{g/L}$); (b2) Diagramme de dispersion des valeurs simulées et mesurées de la Chl-a ($\mu\text{g/L}$) lors de l'apprentissage en utilisant le modèle ANFIS-M5.	134
Figure 45 (a3) Valeurs mesurées et simulées de chlorophylle-a ($\mu\text{g / L}$); (b3) Diagramme de dispersion des valeurs simulées et mesurées de la chlorophylle-a ($\mu\text{g/L}$) lors de la validation pour le modèle ANFIS-M5	134
Figure 46 Résultats de modélisation des concentrations de chlorophylle-a (Chl-a) en utilisant deux techniques de machine learning.	135
Figure 47 (a1) Valeurs mesurées et simulées de phosphore total (mg / L); (b1) Diagrammes de dispersion des valeurs simulées et mesurées de phosphore total (mg/L) pour ensemble complet de données en utilisant le modèle ANFIS-M0	139
Figure 48 (a2) Valeurs mesurées et simulées de phosphore total (mg/L); (b2) Diagrammes de dispersion des valeurs simulées et mesurées de phosphore total (mg/L) lors de l'apprentissage pour le modèle ANFIS-M0.....	139
Figure 49 (a3) Valeurs mesurées et simulées de phosphore total (mg /L); (b3) Diagrammes de dispersion des valeurs simulées et mesurées de phosphore total (mg/L) lors de la validation en utilisant le modèle ANFIS-M0.....	139
Figure 50 Modèle ANFIS avec 5 entrées (Adaptée de Hosseini et al., 2012).....	142
Figure 51 (a1) Valeurs mesurées et simulées de DBO_5 (mg O_2/L); (b1) Diagramme de dispersion des valeurs simulées et mesurées de la DBO_5 (mg O_2/L) pour l'ensemble de données en utilisant le modèle ANFIS-M0.....	144
Figure 52 (a2) Valeurs mesurées et simulées de DBO_5 (mg O_2/L); (b2) Diagramme de dispersion des valeurs simulées et mesurées de la DBO_5 (mg O_2/L) lors de l'apprentissage en utilisant le modèle ANFIS-M0	144
Figure 53 (a3) Valeurs mesurées et simulées de la DBO_5 (mg O_2/L); (b3) Diagramme de dispersion des valeurs simulées et mesurées de la DBO_5 (mg O_2/L) durant la phase de validation en utilisant le modèle ANFIS-M0.....	145

Liste de tableaux

Tableau 1 Indice d'état trophique (TSI) et ses paramètres.....	25
Tableau 2 Analogie entre les neurones biologiques et artificiels	53
Tableau 3 Statistiques des paramètres de la qualité de l'eau pour le barrage Lekhal.....	96
Tableau 4 Statistiques des variables de la qualité de l'eau pour la retenue de barrage Tichy-Haf.....	99
Tableau 5 Statistiques des variables de la qualité de l'eau pour la retenue de barrage Keddara.....	101
Tableau 6 Statistiques des variables de la qualité de l'eau pour le barrage Ouizert.....	104
Tableau 7 Matrice de corrélations	109
Tableau 8 Ordre d'influence des variables de la qualité de l'eau sur la concentration en chlorophylle-a	110
Tableau 9 Performances des modèles élaborés pour modéliser la chlorophylle-a	116
Tableau 10 Performances des modèles élaborés pour modéliser le phosphore total.....	120
Tableau 11 statistiques de base des variables de la qualité de l'eau du barrage Ouizert.....	124
Tableau 12 les modèles de RNAS retenus pour modéliser la DBO_5	125
Tableau 13 Performances des modèles ANFIS élaborés pour modéliser la chlorophylle-a.....	132

Tableau 14 Résultats de simulation de la concentration en Chl-a par les meilleurs modèles de RNAs et d'ANFIS comparée avec les valeurs mesurées.....	136
Tableau 15 Performances des modèles ANFIS élaborés pour la simulation du phosphore total	138
Tableau 16 Critères de performance des modèles ANFIS élaborés pour la simulation du la DBO ₅ ...	143

Introduction générale et problématique

L'eau douce est une ressource limitée, indispensable à l'agriculture, l'industrie et même à l'existence humaine. Sans eau douce de qualité et de quantité suffisante, le développement durable est impossible. Cependant, ce produit dispensé par le jeu du climat est très fragile. La fragilité de cette ressource dépend essentiellement de sa vulnérabilité à la pollution d'origine urbaine, industrielle et agricole.

L'évaluation des ressources en eau nécessite la connaissance et la compréhension complète des processus affectant à la fois la quantité et la qualité de l'eau. La quantité de l'eau est relativement facile à évaluer, elle est déterminée par un seul paramètre, la masse de l'eau, le volume de l'eau ou le débit (m^3/s). Cependant, l'évaluation de la qualité de l'eau est difficile et représente une tâche fastidieuse à accomplir (**Abbasi et Abbasi, 2012**).

La qualité des eaux de surface est l'un des facteurs clés utilisés dans la gestion et l'aménagement des ressources en eau, elle est caractérisée par les diverses substances qu'elle contient, leur quantité et l'effet qu'elles ont sur l'écosystème et sur l'être humain. La surveillance de cette dernière est une tâche difficile, elle requiert l'observation et la mesure régulières et standardisées à long terme de l'environnement aquatique pour un objectif donné.

La qualité de l'eau peut être déterminée par la combinaison de plusieurs variables caractéristiques qui peuvent atteindre le nombre de 100 variables. Cependant, pour la plupart des objectifs, elle est suffisamment décrite par moins de 20 variables, elles sont classées en variables physiques, chimiques et biologiques.

En Algérie, actuellement les retenues de barrage fournissent la grande partie de l'eau potable et l'eau d'irrigation. Les charges polluantes déversées dans les barrages ne cessent d'augmenter en raison de développement excessif du bassin versant, ce qui accélère la détérioration de la qualité des eaux, et provoque l'apparition du phénomène de l'eutrophisation et l'efflorescence algale (algal bloom) associée. Ce phénomène impacte négativement la qualité de l'eau et l'utilisation des eaux de surface pour la production d'eau potable, ainsi que la qualité des écosystèmes dulcicoles et marines (**Diaz et Rosenberg 2008; Smith, 2003; Dodds et al., 2008**).

L'eutrophisation est un processus de surenrichissement d'un plan d'eau en éléments nutritifs, induisant une accélération de la production biologique primaire (croissance accélérée des algues et des plantes aquatiques). Ainsi, l'eutrophisation équivaut à la fois à des charges

nutritives élevées et à une production primaire aquatique élevée. Les charges élevées de nutriments sont la cause dominante, et la production élevée est la conséquence dominante, créant la plupart des problèmes de qualité de l'eau associés à l'eutrophisation (**Jorgensen et al., 2005**). Les symptômes de l'eutrophisation comprennent la prolifération d'algues, la réduction de la clarté de l'eau et l'appauvrissement en oxygène dans l'hypolimnion.

Depuis les années 1960, l'eutrophisation représente, à l'échelle planétaire, l'une des principales menaces pour la qualité de l'eau des lacs et des réservoirs, elle constitue une forme particulière de pollution.

La relation entre la charge en nutriments et l'eutrophisation est compliquée par la diversité de facteurs impliqués qui sont de nature physico-chimiques, biologiques et hydrodynamique.

Il est difficile de prédire la qualité des eaux et l'eutrophisation compte tenu du comportement des nutriments enrichissant les eaux de surface et la complexité des processus impliqués, le recours à la modélisation est une solution innovatrice dans la gestion et l'aménagement des écosystèmes aquatiques, la modélisation nous permet de simplifier la complexité du processus de l'eutrophisation par l'encapsulation des facteurs de forçage (variables d'entrée) dans le même cadre appelé "modèle" de sorte que le chercheur et le décideur puissent développer des outils d'aide à la décision ou modèles prévisionnels dans le but de préserver et de contrôler la qualité de l'eau. Plusieurs modèles de la qualité des eaux et de l'eutrophisation avec un large spectre de complexité ont été développés, la complexité de modèle dépend de la disponibilité des données et de l'écosystème aquatique.

Un modèle est un outil extrêmement utile dans le domaine de l'écohydraulique et la gestion des environnements aquatiques. Dans les études de modélisation, la qualité de l'eau et l'eutrophisation sont utilisées de façon interchangeable (l'une à la place de l'autre) pour représenter les processus d'enrichissement d'un plan d'eau en nutriments (**Ji., 2008**).

La nécessité de l'intégration de l'intelligence artificielle:

Vu la complexité élevée et la non-linéarité de l'écosystème aquatique, la diversité des espèces d'algues, leurs comportement, leurs cycles saisonniers différents d'une espèce à l'autre, l'interaction des espèces avec leur environnement, les interactions entre les espèces (**Recknagel, 1997; Lee et al., 2003**), l'insuffisance de la compréhension des processus, la multiplicité des mécanismes impliqués, le manque de données de bonne qualité (**Chen et Mynett, 2004**), et la haute variabilité saisonnière des charges polluantes déversées dans les retenues de barrages rendent le développement de modèles prédisant la qualité des eaux de barrage et l'eutrophisation un défi.

Les modèles classiques de la qualité des eaux que ce soient mécanistes ou empiriques montrent des points faibles de modélisation, particulièrement la disparité existante entre la rigidité des modèles (ces modèles sont basés sur des formules mathématiques) et la flexibilité de l'écosystème aquatiques (existence des mécanismes rétroactifs, régulation et interactions entre les composantes de l'écosystème), le nombre relativement élevé des paramètres à caler expérimentalement, ces modèles classiques de la qualité des eaux ne représentent que l'état de l'art de modélisation. (**Jørgensen, 2001**).

Les limites et la complexité des modèles mécanistes, ainsi que les approximations impliquées dans les méthodes empiriques utilisées afin de modéliser la qualité des eaux et l'eutrophisation ont motivé les scientifiques à développer de nouvelles méthodes de création de modèles comme reconnaissance de l'imperfection de nos données et la rigidité des modèles classiques, ces nouvelles méthodes de modélisation -qui appartiennent à l'intelligence artificielle tels que les réseaux de neurones, la logique floue et les systèmes d'inférence neuro-floues - constituent une alternative intéressante, un bon choix et une solution innovatrice à cause de leurs caractéristiques; flexibilité, généralisation, parcimonie et elles n'exigent pas la connaissance de la relation mathématiques liant entre les entrées (inputs) et les sorties (outputs) correspondantes.

La plupart des données environnementales ont certains genres de relations entre les variables (les unes aux autres), ces relations qui peuvent être découlées entre des variables environnementales pouvant avoir peu ou rien à voir avec les principes physiques et les mécanismes régissant les processus. Ces liaisons entre les variables n'impliquent pas forcément qu'une variable entraîne l'autre, mais qu'il existe des associations significatives entre elles, qui sont difficiles à exprimer par les équations mathématiques (**Jung, 2009**). Les modèles de type « data driven models » se sont avérés comme des outils puissants pour surmonter ce défi, en raison de leur aptitude de construire un modèle à partir des données, c'est-à-dire leur capacité de capturer les relations entre les variables explicatives et la variable à expliquer à partir de données qu'ils modélisent. C'est l'une des raisons pour lesquelles cette thèse fait appel aux modèles fondés sur l'intelligence artificielle pour la modélisation de la qualité de l'eau et l'eutrophisation.

Comme tout problème résolu par « data-driven models », le processus de l'eutrophisation est simplifié en un problème de type entrées/sortie et va être traité selon cette logique par l'utilisation de ces nouvelles techniques de modélisation.

L'avantage principal des réseaux de neurones artificiels est leurs capacités d'apprendre "par cœur" les données numériques de types entrées/sorties par le biais des poids synaptiques et de

cette façon ils sont capables d'emmagasiner l'information et le comportement du système aquatique.

L'utilisation de la logique floue permet de décrire le comportement du système en utilisant le raisonnement humain (expert) et en définissant les relations explicites à travers l'usage du langage naturel en forme de règles floues (Si.....Alors.....) ou règles linguistiques et la logique floue. Cette méthode est également capable d'intégrer et tient en compte les informations qualitatives, imprécises, vagues et floues qui entourent souvent la qualité de l'eau.

D'importants efforts ont été déployés, au cours des vingt dernières années, pour tenter de répondre aux défis scientifiques et technologiques que posent la compréhension et la prévision du comportement des systèmes aquatiques en réponse à des perturbations naturelles ou artificielles. Dans ce contexte, et pour répondre aux problèmes de la gestion environnementale et de la qualité des eaux de barrages, nous explorerons les capacités des outils de l'intelligence artificielle afin de créer des outils d'aide à la décision ou modèles prévisionnels permettant de comprendre, gérer et prédire le problème de l'eutrophisation de certains barrages en Algérie.

Les objectifs principaux de ce travail sont de (1) développer des modèles basés sur les réseaux de neurones artificiels (RNAs), la logique floue et système d'inférence adaptative neuro-floue (ANFIS) capables de simuler les indicateurs de l'eutrophisation dans certains barrages au Nord de l'Algérie et (2) démontrer la capacité et l'applicabilité de ces techniques dans les questions pratiques de la gestion de la qualité de l'eau et l'eutrophisation (3) choisir la combinaison optimale de variables d'entrée pour améliorer la performance des modèles développés, (4) Comparer les modèles développés et choisir le plus performant.

Nous avons structuré ce travail en quatre chapitres :

1^{ère} chapitre : L'écosystème aquatique de retenue de barrage (réservoir), ayant pour objectif de faire une description de cet écosystème et montrer les différents processus et mécanismes agissant sur la qualité de l'eau.

2^{ème} chapitre : Modélisation de la qualité de l'eau et l'eutrophisation, dans lequel on a présenté les différentes approches de modélisation de la qualité de l'eau et l'eutrophisation.

3^{ème} chapitre : Matériels et méthodes, permettant de présenter, la méthodologie de travail, la théorie des réseaux de neurones artificiels, la logique floue, les sites d'étude (barrages) ainsi que les outils utilisés.

4^{ème} chapitre : Résultats et discussions, comportant les discussions concernant les différents aspects analysés.

Nous terminons cette thèse par une conclusion générale dans laquelle nous présentons les résultats de notre travail, ainsi que les perspectives de recherche à développer.

I. L'écosystème aquatique de retenue de barrage (réservoir)

I.1 Introduction

Un écosystème est défini comme une unité structurelle et fonctionnelle de la biosphère ou une partie de la nature constituée d'une communauté d'êtres vivants et de l'environnement physique associé, qui interagissent et échangent des matériaux entre eux: les écosystèmes peuvent être considérés comme autorégulateurs et des unités autosuffisantes (**Odum, 1971; Zhuravlev et Avetisov, 2006**).

Les écosystèmes sont des systèmes vivants, ils ne réagissent pas de la même manière prévisible que les systèmes physiques rigides. En effet, les écosystèmes ont souvent montré des réactions surprenantes aux changements, ce qui complique évidemment les considérations de gestion environnementale (**Jørgenson et al., 2005**).

Un écosystème aquatique signifie une entité écologique constituée d'éléments vivants (la biocénose) et non vivants (le biotope) en interaction dans un milieu aqueux. Ces écosystèmes aquatiques sont caractérisés par trois interfaces: l'interface air-eau, l'interface sédiment-eau et l'interface organismes-eau où les échanges de la matière et de l'énergie se font de manière continue.

Les réservoirs sont des écosystèmes dynamiques présentant une diversité d'habitats aquatiques. La clé du développement de pratiques de gestion adéquates pour ces systèmes complexes est notre compréhension des interactions entre le bassin versant et le réservoir, et des processus qui se produisent à l'intérieur du réservoir.

Les retenues de barrage (réservoirs) sont des écosystèmes lacustres très importants, elles fournissent les ressources en eau pour plusieurs usages, y compris l'irrigation, l'alimentation en eau potable et l'industrie partout dans le monde (**Wen et al., 2016**). En Algérie, actuellement, les retenues de barrage offrent la grande partie de l'eau potable et l'eau d'irrigation particulièrement dans le Nord du pays.

Comme tout écosystème, les retenues de barrages recèlent un équilibre délicat entre le compartiment biotique et le compartiment abiotique. Les composantes du compartiment biotique sont essentiellement les plantes aquatiques, les phytoplanctons, les animaux, les micro-organismes. Les composantes abiotiques sont les substances organiques et inorganiques, en plus de ça s'ajoute le milieu aquatique lui-même (l'environnement). Elles constituent une entité organique et jouent différents rôles dans les processus de cycle de l'eau et d'autoépuration de l'eau.

I.2. Caractéristique physique de l'écosystème de retenue de barrage

I.2.1. Morphologie des barrages

Un barrage est un ouvrage hydraulique qui barre, sur toute la largeur, une section d'une vallée et qui crée ainsi une dépression topographique artificielle étanche à l'eau appelée retenue de barrage ou réservoir.

Les réservoirs possèdent trois zones distinctes le long du gradient longitudinal. Chaque zone possède des caractéristiques physiques, chimiques, biologiques et dynamiques uniques. Ces zones sont:

I.2.1.1. La zone fluviale

Elle est constituée d'un bassin étroit et ramifiés, le transport d'eau se fait par advection avec des débits relativement élevés. Elle est caractérisée par une turbidité et une concentration de matière en suspension élevées ce qui provoque une faible disponibilité de la lumière en profondeur, les apports en nutriments se font par advection, des niveaux d'éléments nutritifs relativement élevés, la productivité primaire est limitée par la lumière, les pertes cellulaires (phytoplancton) sont principalement causées par la sédimentation, les apports de la matière organique sont principalement allochtone, cette zone est plus eutrophique (**Elsheymy, 2010**).

I.2.1.2. Zone de transition

Elle est constituée par un bassin plus profond et plus large ce qui diminue la vitesse de l'eau par la dissipation d'énergie sur un vaste zone. La turbidité générale diminue et la sédimentation augmente. Elle est caractérisée par une faible concentration en matières solides en suspension, la disponibilité de la lumière en profondeur, une productivité primaire relativement élevée, les pertes cellulaires par broutage et sédimentation, les apports en matières organiques sont intermédiaires (autochtone et allochtone).

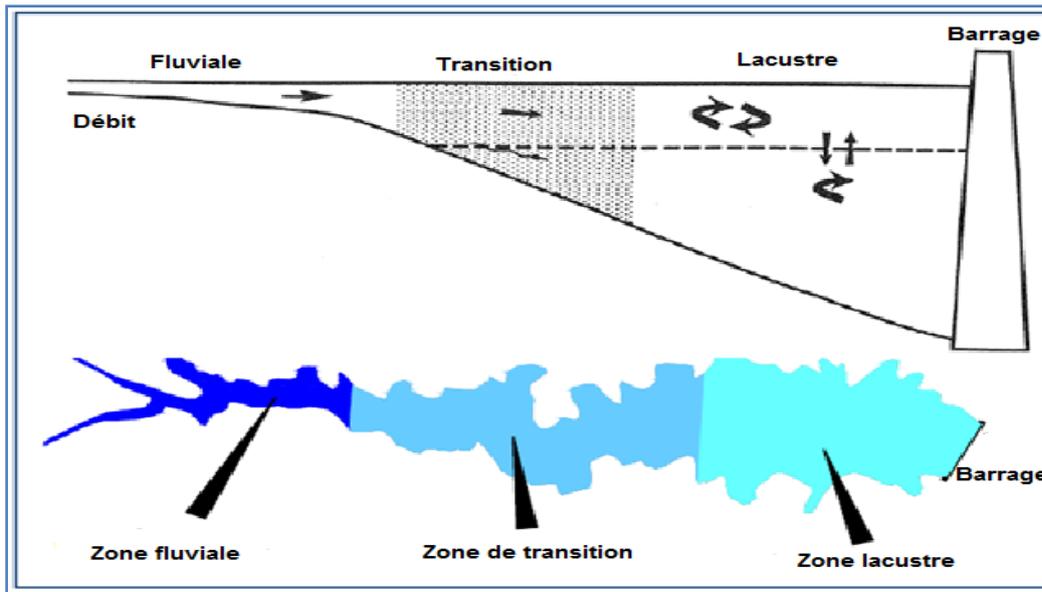


Figure 1 Zonage longitudinales dans les réservoirs (Daniel et al., 2005)

I.2.1.3. La zone lacustre

Les conditions caractéristiques de cette zone sont similaires d'un lac naturel, en particulier dans la partie du réservoir immédiatement derrière la digue. Dans cette zone, les forces de flottabilité dominant le régime d'écoulement. En raison de faible vitesse d'écoulement et la profondeur de l'eau, les concentrations de particules en suspension sont généralement faibles, et la pénétration de la lumière est suffisante ce qui stimule la croissance des algues (Ji, 2008). Les apports de nutriments se font par le recyclage interne, des niveaux d'éléments nutritifs relativement faibles, la productivité primaire est limitée par les nutriments, les pertes cellulaires sont principalement par le broutage, l'approvisionnement en matière organique est essentiellement autochtone.

I.3. Caractéristiques abiotiques de l'écosystème de retenue de barrage

Les composantes abiotiques sont constituées des substances inorganiques et organiques qui sont soumises au régime climatique dans l'écosystème d'eau douce. Ces composantes abiotiques comprennent les nutriments qui sont mis à disposition des producteurs primaires pour la production de la biomasse algale (composants biotiques).

I.3.1. Substances organiques

La matière organique est constituée de restes de matériel végétal et animal, et elle est composée d'un large éventail de composés. Elle peut rentrer dans l'eau à partir de la dissolution de la matière organique ou de la suspension de particules organiques dans les bassins hydrographiques, la chute des feuilles, etc. Les sources extérieures de matière

organique sont connues sous le nom de sources allochtones. La matière organique est également produite dans les écosystèmes aquatiques; ce matériau est connu sous le nom de matière organique autochtone. La matière organique se dégrade progressivement à la suite de processus physiques, chimiques et biologiques, et une fraction de cette matière devient suffisamment petite pour qu'elle soit soluble (**Boyd, 2015**).

En ce qui concerne ses effets sur la qualité de l'eau, la matière organique peut être essentiellement classée en deux catégories brutes; à savoir, matière organique résistante et matière organique facilement dégradable. La matière organique d'origine naturelle est de proportions variables entre ces deux catégories, mais la matière organique autochtone produite dans un plan d'eau par des organismes appartient largement à la deuxième catégorie (**Jørgenson et al., 2005**). La matière organique dans l'eau provoque la couleur et limite la pénétration de la lumière. Les composés organiques chélatent les métaux et augmentent la solubilité des éléments traces dans l'eau (**Boyd, 2015**).

La plupart des eaux douces contiennent de la matière organique (ces matériaux renferment du carbone, comme les phytoplanctons, les plantes et les animaux), qui peut être mesurée en carbone organique total (COT). La matière organique totale dans l'eau peut être une indication utile du degré de pollution. L'impact des polluants sur les ressources en oxygène d'un cours d'eau ou d'un lac est un facteur important dans l'élaboration de tout plan de gestion de la qualité de l'eau. L'impact est normalement mesuré en tant que demande d'oxygène, un paramètre qui peut être interprété comme une mesure brute de la concentration de matériaux oxydables présents dans un échantillon d'eau. Les deux tests les plus utilisés sont la mesure de la demande biochimique en oxygène (DBO) et la demande chimique en oxygène (DCO) (**Elshemy, 2010**).

Dans les eaux de surface non polluées, les concentrations de carbone organique total (COT) sont généralement inférieures à 10 mg / L, les valeurs de DBO₅ de 2 mg O₂/ L ou moins et les concentrations de DCO varient de 20 mgO₂/ L ou moins. Dans les eaux polluées, les concentrations de COT peuvent dépasser 100 mg / L, la DBO₅ peut dépasser 10 mg O₂/ L et la DCO peut dépasser 200 mg O₂/ L (**Chapman et Kimstach, 1996 in Elshemy, 2010**).

I.3.2. Substances inorganiques

Les substances inorganiques dans les eaux de surface sont principalement caractérisées par leur biodisponibilité. Elles sont composées des cations et des anions coexistant en équilibre dans l'écosystème aquatique, c'est ce qu'on appelle la balance ionique.

I.3.2.1. Les cations

Les cations prédominants sont essentiellement le calcium, le magnésium, le sodium et le potassium.

Cations: $\text{Ca}^{2+} > \text{Mg}^{2+} > \text{Na}^+ > \text{K}^+$

I.3.2.2. Les anions

Les anions prédominants sont essentiellement le bicarbonate, le carbonate, le sulfate, et le chlorure.

Anions: $\text{HCO}_3^- > \text{CO}_3^{2-} > \text{SO}_4^{2-} > \text{Cl}^-$.

En comparaison avec les substances inorganiques citées ci-dessus, les quantités d'azote et de phosphore disponibles comme nutriments sont très faibles. Cependant, ils sont très importants pour la production de composantes biotiques (phytoplancton..) dans les plans d'eau. Le faible apport naturel en phosphore est généralement le facteur limitant pour la production de composantes biotiques. Ces nutriments ainsi que de nombreux oligo-éléments sont essentiels pour le développement biologique, mais ils ne contribuent pas substantiellement à la modification de l'équilibre total entre les cations et les anions (**Overbeck, 1988**).

I.3.2.3. L'azote

L'azote est un élément nutritif très important car il est un élément nécessaire dans la structure des protéines. La chlorophylle, l'ARN, l'ADN, plusieurs coenzymes et certaines vitamines contiennent également de l'azote. Ainsi, l'azote est essentiel pour des fonctions aussi importantes que la photosynthèse, la respiration, la synthèse des protéines, la formation de gènes et la croissance. Les composés azotés dissous comprennent l'azote ammoniacal ($\text{NH}_3\text{-N}$ ou $\text{NH}_4^+\text{-N}$), le nitrite (NO_2^-), le nitrate (NO_3^-), l'azote moléculaire dissous et une grande diversité de composés organiques (acides aminés, protéines, nucléotides...etc) (**Jung, 2009**).

L'azote organique est une forme d'azote liée à un composé organique. L'azote inorganique peut exister à l'état libre sous forme de gaz (N_2) ou de nitrate (NO_3^-), de nitrite (NO_2^-) ou d'ammoniac (NH_3).

L'ammoniac est souvent la principale forme d'azote dissous dans un système aquatique et constitue la principale forme d'azote utilisée pour la croissance des algues.

Étant donné que les concentrations de NO_2^- sont relativement faibles, le nitrate est souvent considéré comme étant la somme de nitrite et de nitrate ($\text{NO}_2^- + \text{NO}_3^-$) dans les modèles de qualité de l'eau (**Ji, 2008**).

L'ammoniac est une forme inorganique dissoute d'azote. L'ammoniac total comprend l'ion ammonium (NH_4^+) et le NH_3 non ionisé. La concentration de NH_4 est normalement beaucoup plus élevée que la concentration de NH_3 (Ji, 2008).

Le NH_4 et $\text{NO}_2 + \text{NO}_3$ sont absorbés par les algues, mais le NH_4 est la forme préférée d'azote pour la croissance des algues, les algues utiliseront $\text{NO}_2 + \text{NO}_3$ pour la croissance lorsque la concentration de NH_4 devient très faible.

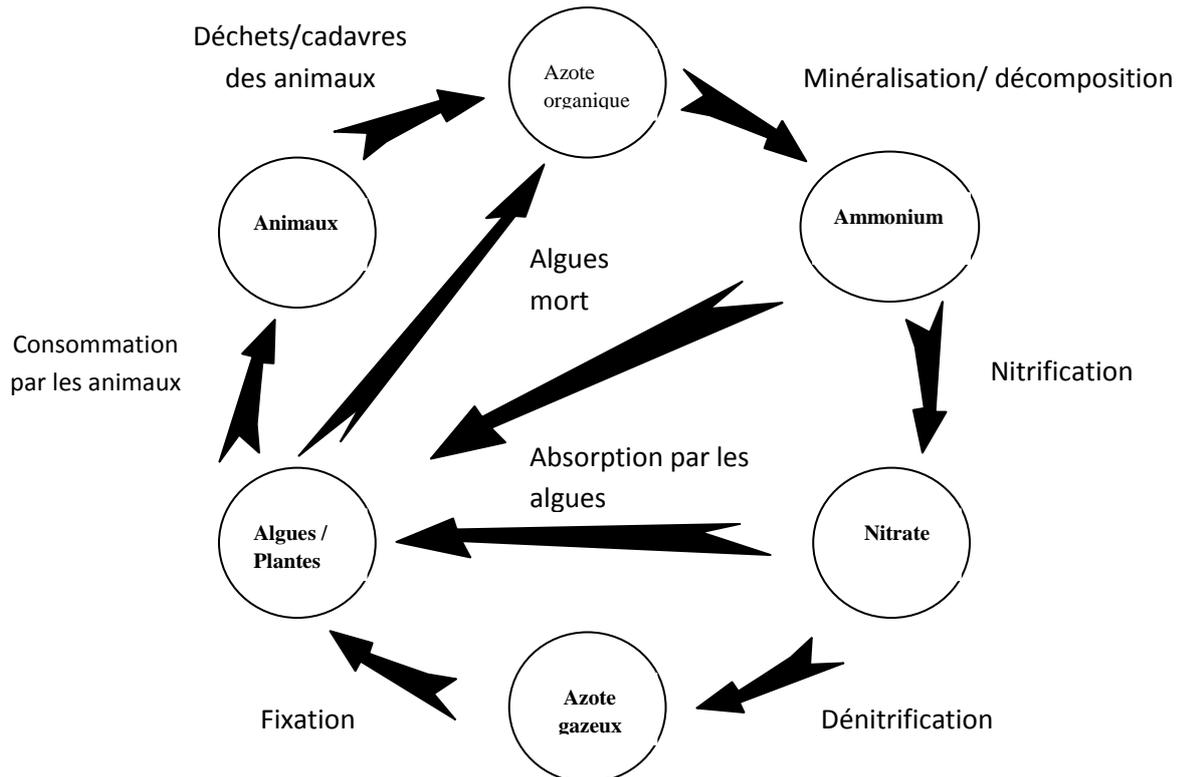


Figure 2 Le cycle de l'azote dans l'écosystème aquatique (Ji, 2008)

L'oxygène a une influence décisive sur l'ensemble du cycle, il détermine les taux de nitrification. L'ammonium est nitrifié en grande partie par *Nitrosomonas* au nitrite. Le nitrite est oxydé vers le nitrate par *Nitrobacter* (Witzel et Overbeck, 1979).



La dénitrification se produit uniquement dans des conditions anoxiques ou à faible teneur en oxygène comme suit : ($\text{NO}_3^- \rightarrow \text{NO}_2^- \rightarrow \text{N}_2$).

I.3.2.4. Le phosphore

Le phosphore est un constituant nutritif de la matière qui passe dans les compartiments abiotique et biotique d'une manière similaire à celle de l'azote. Le cycle du phosphore est

important sur le plan environnemental car la limitation de la croissance des algues dans les eaux douces est souvent due à un manque de cet élément sous une forme chimique pouvant être absorbée par les algues. Le phosphore total comprend le phosphore organique et celui inorganique, ces deux formes comprennent des phases particulières et dissoutes. Les formes organiques de phosphore sont généralement associées aux organismes vivants, et constituées des composés de phosphore facilement décomposables (Ji, 2008).

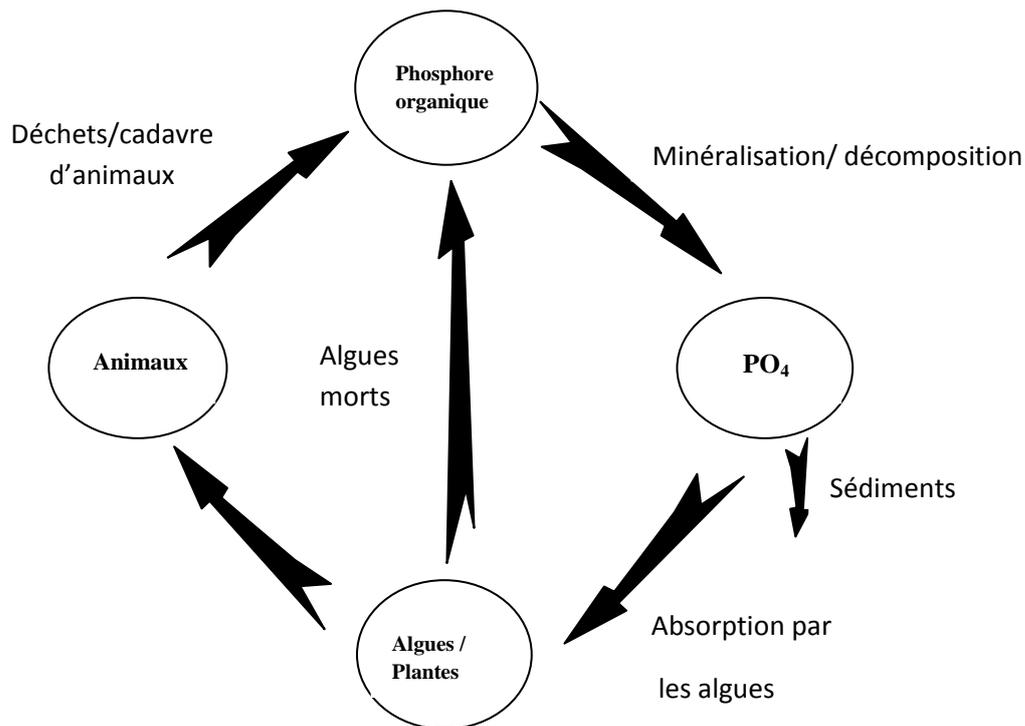


Figure 3 Le cycle du phosphore dans l'écosystème aquatique (Ji, 2008)

Le phosphore inorganique est résumé en phosphate total (PO_4), qui comprend à la fois les phases dissoutes et les phases particulières du phosphate. Le phosphate dissous (PO_4) représente la forme de phosphore qui est directement absorbée par les algues.

Le cycle du phosphore implique des interactions physiques, chimiques et biologiques dont le schéma général est présenté dans la figure 3.

I.4. Caractéristiques biotiques de l'écosystème de retenue de barrage

I.4.1. Zones biologiques dans les lacs et les réservoirs

En fonction de ses communautés biologiques et de sa structure physique, un lac peut être divisé en trois zones biologiques distinctes (Fig.4): zone littorale, zone pélagique (ou limnétique) et zone benthique (Ji, 2008).

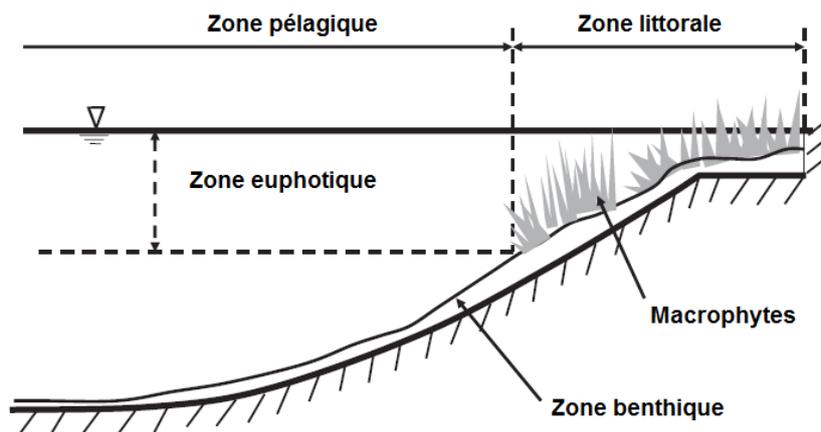


Figure 4 Zones biologiques dans les lacs (Ji, 2008)

I.4.1.1. La zone littorale

La zone littorale est un habitat unique au bord de la rive, dans laquelle au moins 1% de la lumière du soleil pénètre et atteint le fond du lac ce qui permet la croissance des plantes aquatiques enracinées et flottantes. Elle est très irrégulière dans les réservoirs en raison des fluctuations du niveau d'eau plus importantes que dans les lacs. Chaque affluent influence les zones littorales dans les réservoirs. Les plantes aquatiques émergentes, submergées et flottantes sont souvent abondantes dans la zone littorale. En plus d'être une source de nourriture, ces plantes aquatiques fournissent un habitat pour les poissons, les invertébrés et d'autres organismes. En été, les eaux de la zone littorale deviennent très chaudes (Ji, 2008).

I.4.1.2. La zone pélagique (limnétique)

La zone pélagique signifie tout le volume d'eau qui n'est ni à proximité du fond, ni à proximité de la rive. La zone pélagique est la zone d'eau libre où la lumière ne pénètre généralement pas jusqu'au fond. La zone *euphotique* est la couche dans laquelle il existe suffisamment de lumière ($\geq 1\%$) pour que la photosynthèse se produise.

I.4.1.3. La zone benthique

La zone benthique est une mince couche de sédiments au fond du lac. Elle a typiquement une épaisseur de quelques centimètres et contient une grande variété d'organismes benthiques, qui sont principalement des invertébrés. L'abondance et la composition spécifique des organismes benthiques sont fortement influencées par le niveau d'oxygène dans la zone benthique (Ji, 2008).

I.4.2. La chaîne alimentaire

La chaîne alimentaire dans un réservoir peut être divisée en producteurs, macroconsommateurs et microconsommateurs, les échanges de l'énergie et la matière se font sans cesse entre les maillons de la chaîne.

Les producteurs primaires sont à la base de la chaîne alimentaire, ce sont des organismes photosynthétiques formés essentiellement de phytoplanctons (micro-algues), des cyanobactéries et des plantes aquatiques qui, grâce à la photosynthèse, peuvent produire leur matière organique à partir des éléments nutritifs et du CO₂ en utilisant l'énergie lumineuse.

Les producteurs primaires forment la base de la chaîne alimentaire et servent de source de nourriture pour les organismes supérieurs dans les écosystèmes. Les algues sont l'un des deux principaux producteurs primaires dans les eaux de surface. Les autres producteurs primaires sont les plantes aquatiques enracinées ou flottantes (macrophytes), qui sont généralement limités aux eaux peu profondes. Dans la plupart des cas, les algues sont plus importantes dans la production alimentaire que les plantes aquatiques enracinées (**Ji, 2008**).

Les algues jouent un rôle clé dans le processus d'eutrophisation et sont essentielles pour la modélisation de la qualité de l'eau, ils affectent le cycle de l'azote, le cycle du phosphore, l'équilibre de l'oxygène dissous et la chaîne alimentaire, principalement par l'absorption des nutriments et la mort des algues. À mesure que les algues croissent et meurent, elles font partie des cycles des nutriments (**Ji, 2008**).

Les microconsommateurs, constitués de bactéries et de champignons qui sont responsables de la dégradation de la matière organique, ils sont également appelés les décomposeurs. **Les macroconsommateurs**, composés de zooplancton et de poissons, ils se nourrissent d'autres organismes ou de la matière organique (détritiques) (**Jung, 2009**).

I.4.3. Les producteurs primaires aquatiques

Les producteurs primaires aquatiques regroupent différents groupes d'organismes: les cyanobactéries, les algues, les végétaux supérieurs (phanérogames), les fougères (ptéridophytes), les mousses (bryophytes). Ce sont des organismes autotrophes ; ils sont capables de produire de la matière organique et de l'oxygène à partir du dioxyde de carbone et de l'eau grâce à l'énergie solaire. Du point de vue phylogénétique, ils présentent une très grande diversité, celle-ci se traduit par une très grande diversité des mécanismes physiologiques, des voies métaboliques et des structures morphologiques des organismes

photosynthétiques aquatiques (**Pinay et al., 2017**). Les producteurs primaires sont classés en microphytes et macrophytes. Les microphytes sont les producteurs primaires microscopiques regroupant les cyanobactéries et les microalgues. Les macrophytes sont les producteurs primaires visibles à l'œil nu, ce groupe inclut les macroalgues, les végétaux supérieurs, les fougères aquatiques et les mousses. Les producteurs primaires peuvent se développer librement dans la colonne d'eau (phytoplancton), ou bien attachés à un substrat (phytobenthos) ou à d'autres organismes (épiphytes ou épibiontes) (**Pinay et al., 2017**).

I.5. Eutrophisation

I.5.1. Introduction

La mise en eau des réservoirs transforme un écosystème lotique (la rivière) et des écosystèmes terrestres, en un écosystème lentique (un plan d'eau). Les fonctionnements physique et biologique ainsi modifiés conduisent à créer ou à augmenter la sédimentation, le temps de séjour de l'eau, et des conditions favorables à une production phyto-planctonique (**Félix-Faure, 2017**).

L'eutrophisation est un déséquilibre nutritionnel et écologique en milieux aquatiques induit par l'augmentation de la disponibilité des nutriments dans l'écosystème aquatique, provoquant une perturbation de l'équilibre entre le phytoplancton (algue) et les nutriments disponibles, en raison de l'accumulation continue des éléments nutritifs dans le plan d'eau. Celle-ci stimule une croissance accélérée des algues et des plantes aquatiques.

Dans la littérature scientifique, il existe plusieurs définitions de l'eutrophisation des eaux.

Ji (2008) a défini l'eutrophisation comme étant un processus de sur enrichissement d'un plan d'eau en éléments nutritifs, induisant une accélération de la production biologique primaire (croissance accélérée des algues et des plantes aquatiques). Les symptômes de l'eutrophisation comprennent la prolifération d'algues, la réduction de la clarté de l'eau et l'appauvrissement en oxygène.

Pinay et al., (2017) a définit l'eutrophisation des écosystèmes aquatiques comme une succession de processus biologiques enclenchés en réponse à un apport excessif de nutriments. Elle se traduit par des réponses complexes de l'ensemble des écosystèmes aquatiques d'eau douce, saumâtre ou salée. Elle est progressive ou brutale. Les effets les plus notables sont des proliférations de producteurs primaires (plantes aquatiques, algues, cyanobactéries), des phénomènes de toxicité et d'anoxie (absence d'oxygène), des pertes de biodiversité.

L'eutrophisation a été reconnue comme un problème de pollution dans les lacs et réservoirs partout dans le monde. Depuis lors, elle est devenue plus répandue. Des études ont montré que la plupart des lacs d'Asie, d'Europe, d'Amérique du Nord, d'Amérique du Sud et d'Afrique se trouvent dans un état eutrophique. L'eutrophisation entraîne des changements importants dans la qualité de l'eau (Ansari et al., 2010). Elle impacte négativement la qualité de l'eau et l'utilisation des eaux de surface pour la production d'eau potable, ainsi que la qualité des écosystèmes dulcicoles et marins (Diaz and Rosenberg, 2008; Smith, 2003; Dodds et al., 2008). Elle diminue la valeur des eaux de surface pour les usages industriels et récréatifs.

Les charges nutritives excessives produisent généralement une production primaire plus élevée par le phytoplancton et les macrophytes. Ainsi, l'eutrophisation équivaut à la fois à des charges nutritives élevées et à une production aquatique élevée. Les charges élevées de nutriments sont la cause dominante, et la production élevée est la conséquence dominante, créant la plupart des problèmes de qualité de l'eau associés à l'eutrophisation. Ce ne sont pas les concentrations de phosphore et de carbone dans un lac en soi qui causent des problèmes de qualité de l'eau; c'est plutôt la production de phytoplancton et de macrophytes résultant des nutriments qui crée les problèmes (Jorgensen et al., 2005). La figure ci-après représente les causes et les effets de l'eutrophisation (Thomas et al., 2006).

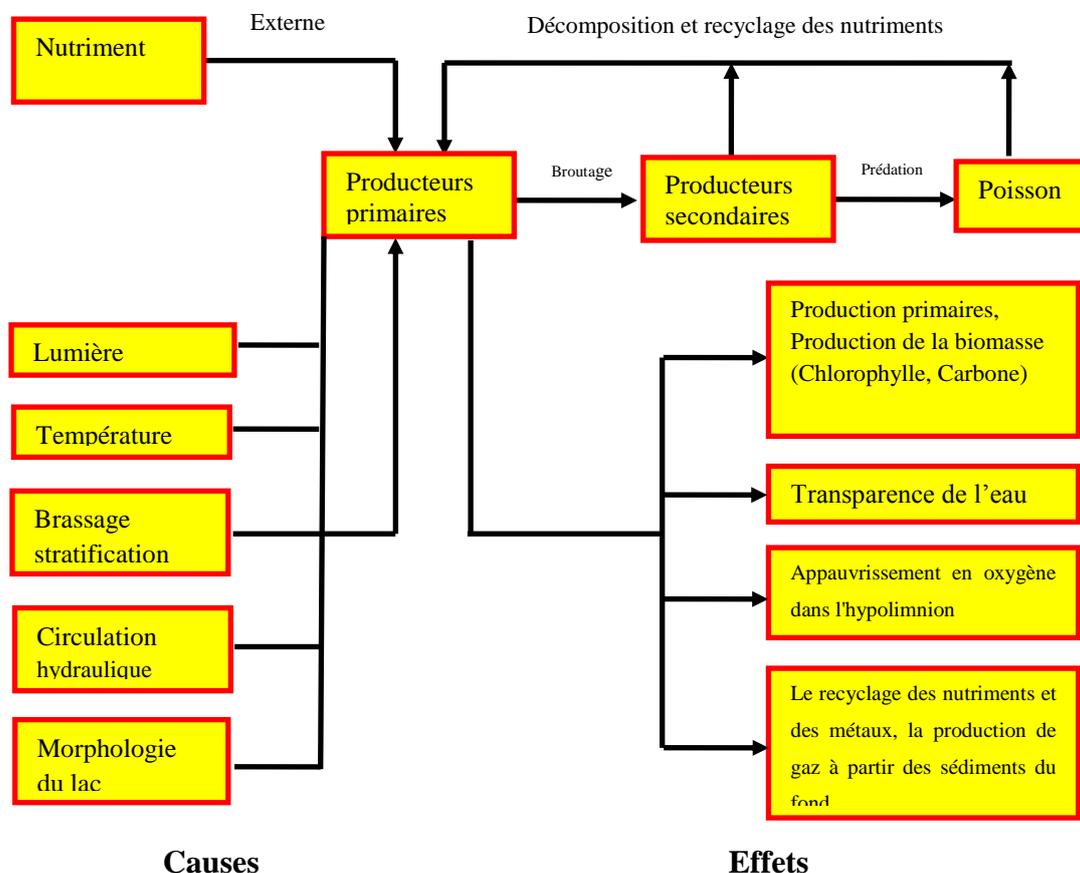
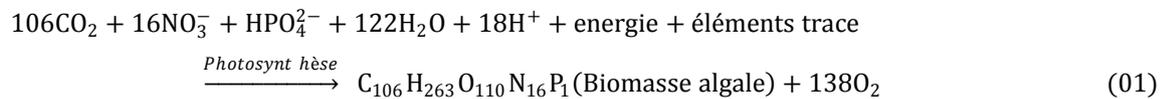


Figure 5 Les causes et les effets de l'eutrophisation (Thomas et al., 2006)

I.5.2. Assimilation et nutriments limitants

Les nutriments excessifs sont le facteur clé de l'eutrophisation, le phosphore étant le nutriment essentiel, suivi de l'azote, du carbone, des oligo-éléments et des vitamines. Le processus d'eutrophisation peut être exprimé selon l'équation suivante (**Wen et al., 2016**):



L'équation montre que le phosphore et l'azote dans les masses d'eau naturelles sont les facteurs critiques pour la production de biomasse de plancton. Les eaux usées industrielles, les eaux usées et les eaux de drainage agricoles contiennent de grandes quantités d'azote et de phosphore, ce qui pourrait entraîner une augmentation des nutriments dans l'eau et provoquer une croissance excessive (accélérée) des algues et des plantes vertes.

La biomasse maximale des algues est déterminée par l'apport en nutriments et la disponibilité de la lumière. Quand un nutriment essentiel à la croissance des plantes aquatiques devient indisponible (parce que sa réserve est épuisée ou qu'il est sous une forme indisponible), la croissance des algues devient limitée par ce nutriment (**Robertson, 1997**). Le concept de nutriment limitant a son fondement dans la réaction de photosynthèse.

L'équation précédente révèle que l'azote et le phosphore sont absorbés par les algues dans un rapport approximativement constant de 16 atomes d'azote par 1 atome de phosphore, ou de 7,2/1 en poids. Pour une première approximation, il faut environ sept fois plus d'azote que de phosphore pour produire une quantité donnée d'algues (**Jorgensen et al., 2005; Ji, 2008; Jung, 2009**). Si les quantités disponibles de N et P diffèrent largement de ce rapport, il en résulte une limitation de la production primaire. Le rapport critique N: P est souvent pris comme 10:1 (**Forsberg et al., 1978**). Cependant, des précautions doivent être prises dans l'utilisation pratique du concept de nutriments limitants (**Phoslock, 2008; Lee et al., 2000**). Premièrement, au sens propre du mot, le concept ne s'applique qu'en régime permanent. Une autre considération est qu'un plan d'eau contient normalement un certain nombre d'espèces d'algues différentes, chacune ayant des besoins nutritionnels relativement spécifiques (**Glibert, 2007**).

Différentes espèces d'algues peuvent assimiler des nutriments en quantités différentes et à des rythmes différents (**Tilman, 1982**). Dans les eaux naturelles, la prolifération de certaines espèces d'algues se fait pendant un certain temps et laissent ensuite la place à d'autres espèces

plus compatibles avec les nouvelles conditions (conditions modifiées), telles que la température, la lumière du soleil et / ou les concentrations en éléments nutritifs (**Ji, 2008**).

Les proportions relatives de nutriments des plantes aquatiques dans le réservoir peuvent également changer de façon saisonnière et d'année en année. Ainsi, un nutriment peut limiter les niveaux de biomasse algale en même temps, alors qu'un autre nutriment peut être limitant à un autre moment. Cependant, l'utilisation pratique du concept de nutriment limitant suppose qu'un seul élément nutritif requis par les algues sera le facteur limitant, du moins pendant la période considérée (**Jung, 2009**).

Compte tenu des considérations ci-dessus, il est généralement admis que l'azote est l'élément nutritif limitant la croissance des algues dans les écosystèmes marins côtiers, alors que dans les systèmes d'eau douce, il s'agit généralement de phosphore (**Howarth et Marino, 2006; Ji, 2008**).

I.5.3. Les facteurs de l'environnement

La biomasse du phytoplancton dans le plan d'eau semble être proportionnelle à la charge en nutriments, au moins jusqu'à un certain niveau. Au-delà de ce niveau, le phytoplancton présente une valeur relativement constante, ne présentant plus d'augmentation de la croissance des algues, indépendamment de l'augmentation de l'apport de nutriments (**Cowen et Lee, 1976 ; Sridharan et Lee, 1977 ; Lee et al., 1980**). Cela signifie que des facteurs autres que la disponibilité des nutriments peuvent également limiter la croissance des algues (**Jung, 2008**).

A part les nutriments, la croissance des algues est influencée par la lumière, la température de l'eau, le temps de résidence...etc.

I.5.3.1. La Lumière

Le phytoplancton a besoin de la lumière du soleil pour la croissance, mais une intense lumière solaire près de la surface de l'eau peut inhiber certaines espèces. Certaines eaux contiennent suffisamment de turbidité provenant de particules minérales ou organiques en suspension ce qui restreint considérablement la pénétration de la lumière et réduit la croissance du phytoplancton. La croissance du phytoplancton répond favorablement à la chaleur. Les taux de croissance les plus élevés se produisent habituellement au printemps et en été, mais la croissance se poursuit à un rythme plus lent durant les mois les plus froids (**Boyd, 2015**).

I.5.3.2. La température

La plupart des processus de qualité de l'eau dépendent de la température. La température influence significativement la cinétique des transformations nutritives; la vitesse de réactions

chimiques augmente avec la température. Le rôle de la température dans l'eutrophisation est à la fois physique et biologique. La température de l'eau influe fortement sur les taux de croissance des algues, la cinétique du recyclage des nutriments et la décomposition biologique. La température et, dans une moindre mesure, la salinité influencent la solubilité de l'oxygène dissous (**Jørgenson et al., 2013 ; Ji, 2008**).

I.5.3.3. Le temps de résidence de l'eau

Le temps de résidence du réservoir régule les cycles biogéochimiques, la production primaire et la biodiversité, il varie beaucoup de quelques jours à plusieurs mois ou années. Ce temps de résidence est également extrêmement variable entre les réservoirs. Il conditionne le développement des proliférations algales, des temps de résidence élevés (supérieurs à 1 an) retiennent les nutriments, dégradent la qualité de l'eau et influencent la biodiversité et augmentent potentiellement le risque d'eutrophisation en raison de l'augmentation des échanges entre les organismes photosynthétiques et les nutriments dissous (**Jørgensen et al., 2005 ; Pinay et al., 2017**). Un temps de résidence réduit est fondamental pour un réservoir qui fournit de l'eau potable, car il peut contenir une meilleure qualité de l'eau (**Jørgensen et al., 2013**).

I.5.4. Algues

Le phytoplancton (algues flottantes), les macroalgues et les plantes vasculaires sont généralement présents dans diverses proportions dans la plupart des écosystèmes aquatiques, mais le phytoplancton est plus biologiquement actif et a généralement une plus grande influence sur la qualité de l'eau que d'autres plantes (**Boyd, 2005**).

Les algues jouent un rôle clé dans le processus d'eutrophisation et sont essentielles pour la modélisation de la qualité de l'eau. Les algues affectent le cycle de l'azote, le cycle du phosphore, l'équilibre de l'oxygène dissous (DO) et la chaîne alimentaire, principalement par l'absorption des nutriments inorganiques dissous (PO_4 , NO_3 , NH_4 et silice disponible (SA)) par les algues pendant le processus de photosynthèse et recyclage des nutriments sous forme de phosphore organique (OP), azote organique (ON), et de silice non disponible (silice biogénique particulaire) (SU) (**Ji, 2008**).

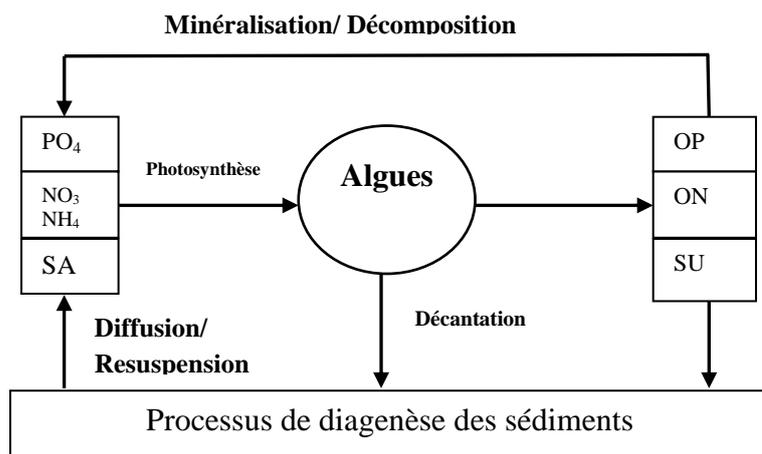


Figure 6 Cinétique algal (Ji, 2008)

Le processus de décantation élimine ces matières particulaires dans le fond. Grâce à la diagenèse des sédiments dans le fond et à la minéralisation et à la décomposition dans la colonne d'eau, les nutriments particulaires sont transformés en formes inorganiques disponibles pour le prochain cycle d'absorption d'algues, ceci complète le cycle des algues. À mesure que les algues croissent et meurent, elles font partie des cycles des nutriments (Ji, 2008).

La composition des phytoplanctons n'est pas constante, en général, elle reflète dans une certaine mesure la concentration de l'eau (Jørgensen et al., 2013).

Les algues sont un groupe d'organismes très divers, elles comprennent (1) les cyanobactéries (algues bleues), (2) les algues vertes et (3) les diatomées. Au cours de l'année, la dominance des espèces d'algues dans les lacs change au cours d'un cycle annuel, connu sous le nom de succession algale (Jung, 2009). Cette succession naturelle d'algues se produit en réponse aux changements de saison, de température, de vent, de précipitations et de cycles des éléments nutritifs (Moore et Thornton, 1988) et aux conditions d'absorption des nutriments pour les différentes espèces (Shapiro, 1970 ; Silvey et al., 1972).

Les principaux processus d'eutrophisation dans les lacs sont liés aux apports d'éléments nutritifs provenant du bassin versant et à l'activité des algues, des cyanobactéries, des macrophytes et des microbes et de la faune associés dans les lacs. L'influence des concentrations accrues d'éléments nutritifs sur la biomasse des algues dans les lacs est non linéaire. Avec des concentrations croissantes du nutriment limitant, la biomasse algale augmente d'abord lentement, puis plus rapidement, suivie d'une élévation presque linéaire jusqu'à ce qu'une certaine valeur asymptotique de la biomasse soit atteinte à des concentrations élevées du nutriment limitant (Jørgensen et al., 2005).

I.5.4.1. Estimation de l'abondance de la biomasse algale

La biomasse constitue la quantité d'un organisme vivant ou d'un ensemble d'organismes dans un volume ou une zone spécifique d'un écosystème, généralement mesurée en unités, telles que le poids humide, le poids sec, le biovolume et la teneur en azote (**Ji, 2009**).

Le nombre d'individus (cellules individuelles, filaments ou colonies) par millilitre ou litre d'eau peut être déterminé directement par examen microscopique. Différentes espèces ont des tailles caractéristiques, et le volume de cellules phytoplanctoniques par unité d'eau fournit une meilleure estimation de la biomasse que le dénombrement de l'abondance du phytoplancton. Cependant, il est extrêmement difficile de mesurer le volume de phytoplancton dans un échantillon (**Jørgensen et al., 2005; Boyd, 2005**). Le recours aux méthodes indirectes est indispensable. La concentration en chlorophylle-a est un indicateur fiable pour estimer la biomasse algale microscopique présente dans le plan d'eau, sa concentration dépend de l'état des nutriments, de la lumière et de la composition des espèces (**Dimberg et al., 2012**).

La quantité d'algues peut être mesurée beaucoup plus rapidement et commodément comme la concentration de chlorophylle-a dans les échantillons d'eau. Cela se fait soit par examen spectrophotométrique du matériau collecté sur des filtres à travers lesquels un volume connu d'eau de lac est passé, soit par mesure fluorimétrique d'échantillons d'eau contenant des algues vivantes. Les mesures peuvent même être faites dans un lac, en utilisant des fluorimètres submersibles modernes (**Jørgensen et al., 2005**).

Le rapport entre la biomasse algale et la Chlorophylle-a prend généralement la valeur de 145 pour la modélisation de la qualité de l'eau (**Cole et Wells, 2003**). Mais, par exemple, selon **Lee et al. (2000)**, le rapport moyen entre la biomasse et la Chlorophylle-a de *Microcystis aeruginosa* cultivé dans diverses conditions fixées par N et P, s'élève à 292.

I.5.5. Effets de l'eutrophisation sur la qualité de l'eau

Les effets de l'eutrophisation sur la qualité de l'eau comprennent généralement une détérioration de la qualité de l'eau, une hausse des coûts de traitement de l'eau potable et des problèmes de goût et d'odeur dans l'eau potable.

L'eutrophisation est une source de nombreux désagréments lors de la production d'eau potable dont la surconsommation des réactifs oxydants (ozone, chlore), le colmatage des filtres, l'apparition de mauvais goûts pour les consommateurs (présence de métabolites au goût de moisi), donc le coût de traitement s'élève et la qualité d'eau diminue. L'efflorescence algale peut faire tomber en panne une station de traitement de l'eau en pleine production.

Les filtres utilisés dans le traitement de l'eau peuvent s'encrasser et nécessiter des nettoyages fréquents (Jørgensen et al., 2005). Les efflorescences algales contribuent de manière significative à la teneur totale en matière organique (carbone) dans l'eau grâce aux processus métaboliques au sein des cellules algales et à leur autolyse. L'augmentation de la matière organique induite par la production de biomasse nécessite une plus grande dose de désinfectant dans le processus de traitement de l'eau potable (Jung, 2009). Les acides humiques et fulviques, qui sont des matières organiques pouvant être produites par les efflorescences algales, agissent comme précurseurs dans la formation de trihalométhanes tels que le chloroforme (CHCl₃), ainsi que d'autres sous-produits halogénés nocifs pour la santé (bromoforme CHBr₃, bromodichlorométhane CHBrCl₂) (Moore et Thornton, 1988).

Des difficultés mécaniques peuvent être créées par le colmatage des tubes à cause de la croissance des organismes et par la corrosion causée par les fortes concentrations de dioxyde de carbone et de méthane dans les masses d'eau extrêmement eutrophes. Les activités de loisirs peuvent être menacées par de fortes concentrations d'algues, y compris des changements dans la composition des espèces d'algues, et une présence accrue de cyanobactéries toxiques créant des problèmes de santé (y compris la mort des humains et des bétails). (Jørgensen et al., 2005).

I.5.6. La méthode d'évaluation de l'eutrophisation

I.5.6.1. Introduction

Tout au long de l'histoire des études sur l'eutrophisation, de la classification des lacs au développement de modèles prédictifs et de gestion, de nombreuses manifestations de l'eutrophisation ont été exprimées qualitativement. Des descriptions telles que la transparence de l'eau ou l'occurrence et la fréquence des «efflorescence algal nuisibles» sont toujours valables en tant qu'expression de l'impact de l'eutrophisation pour les non spécialistes, mais le développement de la compréhension scientifique de l'eutrophisation peut être considéré comme une progression des mesures qualitatives des causes et de leurs effets aux mesures quantitatives (Harper, 1992).

I.5.6.2. Evaluation qualitative de degré d'eutrophisation

Qualitativement, les limnologues classent les lacs selon l'équilibre des organismes et la concentration en nutriments contenus dans leurs eaux en différentes classes (oligotrophe, mésotrophique et eutrophique).

Oligotrophes: Eaux claires, pauvres en nutriments comportant très peu de plantes et d'animaux. Les lacs oligotrophes sont généralement profonds et contenant très peu de boue organique au fond.

Mésotrophique: eaux contenant des quantités modérées en nutriments et comportant des communautés saines et équilibrées de plantes, d'animaux et de bactéries. Les lacs mésotrophes reçoivent des quantités adéquates en eau douce et en nutriments, et les retournements saisonniers permettent aux couches pauvres en nutriments et celles riches en nutriments à se mélanger. Les lacs mésotrophes se situent à un point intermédiaire entre les lac oligotrophes et les lacs eutrophiques concernant la profondeur, la concentration en chlorophylle-a, la clarté de l'eau et les plantes aquatiques.

Eutrophique: Lacs riches en nutriments comportant une vie végétale abondante dans leurs surfaces, dont l'eau est souvent trouble par les plantes microscopiques, et leurs fonds sont couverts de couches épaisses de matériel végétal en décomposition. Les bactéries qui vivent sur la boue organique consomment de l'oxygène et provoquent un appauvrissement en oxygène des couches profondes ce qui cause la mort probable des poissons aquatiques.

I.5.6.3. Evaluation quantitative du degré d'eutrophisation

Quantitativement, le degré du phénomène de l'eutrophisation d'un lac et de l'ampleur de son enrichissement en nutriments ou de sa productivité primaire est quantifié par l'indice de l'état trophique.

I.5.6.3.1. L'état trophique

L'état trophique d'un lac est une mesure de la productivité biologique (niveau de qualité écologique de l'eau), c'est-à-dire la masse de plantes et d'animaux dans un lac. La diversité des plantes et des animaux dans l'eau est également déterminée par la qualité de l'eau des écosystèmes aquatiques et sert donc d'indicateur de qualité de l'eau. Les lacs clairs sont caractérisés par une grande diversité de plantes et d'animaux par rapport aux lacs troubles. Les variables utilisées pour déterminer l'état trophique comprennent: le phosphore total, la chlorophylle-a et la clarté de l'eau. Le niveau de qualité de l'eau écologique peut être exprimé par l'indice d'état trophique basée sur les variables d'état trophique. L'azote total et le phosphore total sont les nutriments qui limitent la croissance des algues dans les lacs. Plus les concentrations de nutriments sont élevées, plus l'indice d'état trophique du lac est élevé. L'état trophique est groupé en oligotrophe, mésotrophe, eutrophique et hypereutrophique selon la valeur de l'indice d'état trophique.

L'état trophique d'un lac donne le statut du lac en termes de productivité algale. Chaque lac se situe dans un état trophique particulier en fonction de son degré d'eutrophisation, et le statut trophique d'un lac peut changer avec le temps (**Moses, 2011**). L'augmentation des apports en phosphore stimule l'abondance du phytoplancton et des algues, entraînant une augmentation de la chlorophylle a, ce qui entraîne une diminution de la transparence de Secchi (**Carlson 1977; Haggard et al., 1999**).

I.5.6.3.2. Indice de l'état trophique

L'indice de l'état trophique (TSI) condense les données sur la qualité de l'eau en un indice numérique unique gérable pour déterminer l'état trophique d'un lac (**Moses, 2011**). La TSI exprime la productivité lacustre sur une échelle numérique continue de 0 à 100, il s'agit d'un nombre croissant indiquant davantage les conditions eutrophiques. Des mesures de n'importe quel critère trophique (phosphore, profondeur Secchi ou chlorophylle a) peuvent être utilisées pour déterminer l'indice de l'état trophique.

L'état trophique est indiqué au moyen d'une échelle de la productivité primaire de faible à élevée : l'état *oligotrophe* peu productif, l'état *mésotrophe* moyennement productif et l'état *eutrophique* hautement productif. Il ya également des classes intermédiaires.

La relation **PT-Chl-a** a été employée par **Carlson (1977)** pour développer un indice numérique de l'état trophique (TSI). C'est probablement l'indice le plus couramment utilisé, qui inclut trois variables : le phosphore total (TP), la chlorophylle-a (chl-a) et transparence de l'eau (mesurée à l'aide du disque de Secchi).

L'indice de l'état trophique de Carlson (TSI) (**Carlson, 1977**) résume l'état trophique d'un plan d'eau à un ou plusieurs nombres, dans une tentative d'éliminer la subjectivité inhérente aux termes oligotrophe, mésotrophique et eutrophique. Au lieu de cela, il indique le degré de l'eutrophisation dans chaque classe. La classe *eutrophique* par exemple, englobe un large éventail de conditions lacustres et ce terme lui-même ne précise pas le degré de l'eutrophisation, bien que l'utilisation de ces termes dans les communications sur la qualité du lac reste toujours nécessaire.

L'indice de l'état trophique de Carlson (TSI) représente les valeurs absolues de chlorophylle-a (chl-a), de phosphore total (TP) et de transparence (SD) applicables à n'importe quel lac (avec une turbidité minimale non algale), contrairement aux indices pour lesquels les valeurs sont relatives et limitées à un ensemble de données particulier. Les transformations des données par le \log_2 ont été utilisées pour relier ces trois indices sur une échelle de 0 à 100, de sorte qu'un doublement de TP est lié à une réduction de la moitié de la

transparence de l'eau (SD). Les valeurs représentatives pour TP, chl-a et SD, calculées à partir des équations suivantes pour le TSI, sont présentées dans le Tableau ci-dessous.

$$\text{TSI} = 10(6 - \log_2 \text{SD}) \quad (02)$$

$$= 10(6 - \log_2 7.7/\text{chl a}^{0.68}) \quad (03)$$

$$= 10(6 - \log_2 48/\text{TP}) \quad (04)$$

Tableau 1 Indice d'état trophique (TSI) et ses paramètres

Etat trophique	TSI	Disque de Secchi (m)	Phosphore de surface (mg /m ³)	Chlorophylle-a de surface (mg /m ³)
Oligotrophe	<40	64	0.75	0.04
		32	1.50	0.12
		16	3	0.12
		8	6	0.94
Mésotrophe	40-50	4	12	2.60
		2	24	7.30
Eutrophique	50-70	1	48	20
		0.5	96	56
Hyper-eutrophique	>70	0.25	192	154
		0.12	384	427
		0.062	768	1138

Source: **Carlson, R.E. 1977**. A trophic state index for lakes. *Limnol. Oceanogr.* 22: 361–368.

Il faut noter que la plus grande variation de SD se produit en dessous d'une concentration en chl-a d'environ 30 mg/L. Au-dessus de 30 mg/L, il y a relativement peu de changement de transparence avec l'augmentation de chl-a.

I.5.6.3.3. Les indicateurs de l'eutrophisation

Les variables généralement utilisées pour évaluer le degré de l'eutrophisation sont :

I.5.6.3.3.1. Le phosphore total :

La concentration en phosphore total dans les eaux de surface est généralement utilisée comme indicateur de l'état trophique de ces plans d'eau. Le phosphore est l'un des nutriments essentiels à la croissance des algues, il est considéré comme l'élément limitant principal déterminant la productivité biologique (productivité algale, et productivité ultérieure d'ordre supérieur) dans de nombreux systèmes aquatiques d'eau douce (**Scholten et al., 2005**).

L'indice de phosphore (voir formule précédente) est utile parce que les niveaux de phosphore demeurent relativement constants tout au long de l'année et peut donc fournir une mesure représentative de la quantité de nutriments dans le lac. Il est considéré comme un *predictor* du potentiel de la biomasse algale (Carlson, 1977).

I.5.6.3.3.2. La chlorophylle *alpha*

La variable biologique la plus communément mesurée pour définir l'état trophique et la qualité du lac est la chlorophylle-a (chl-a). La chl-a est un indicateur fiable de la biomasse algale microscopique présente dans le plan d'eau, sa concentration dépend de l'état des nutriments, de la lumière et de la composition des espèces. La teneur des cellules en chl-a peut varier avec les variables mentionnées ci-dessus. Elle est utilisée plus souvent que le biovolume pour indiquer l'état trophique (Dennis, 2005). La concentration en chlorophylle-a s'accroît avec l'augmentation de la concentration des nutriments. Il y a donc une relation entre la chlorophylle *a* et le niveau trophique du lac. Les lacs eutrophiques produisent une importante quantité d'algues (quantité importante de chl-a).

I.5.6.3.3.3. La transparence de l'eau

La transparence de l'eau, mesurée à l'aide de disque de Secchi (SD), qui est un moyen facile et peu coûteux, permet d'avoir une idée approximative de l'état trophique d'un plan d'eau. Lorsque la teneur de l'eau en particules en suspension augmente, la profondeur de disque de Secchi (transparence) diminue exponentiellement. Cependant, la transparence est habituellement liée à la concentration de particules, que celles-ci soient des algues ou d'autres matières en suspension. Il peut y avoir plus de variabilité horizontale de la transparence que de la température, surtout si les algues bleues flottantes sont abondantes dans le lac et sont distribuées de façon inégale par le vent. Des mesures sur plus d'un site, même dans les petits lacs, sont recommandées. (Dennis, 2005).

I.5.6.3.3.4. L'oxygène dissous

La matière organique sédimentée dans le fond, et provenant de la mort des algues et plantes aquatiques, lors de sa dégradation par l'activité microbienne induit la diminution de la concentration en oxygène dissous dans l'hypolimnion. Dans les lacs eutrophiques, la stratification mène souvent à l'épuisement de l'oxygène dans l'hypolimnion. En fait, les lacs sont classés en tant qu'eutrophique si l'épuisement de l'oxygène dissous se produit dans l'hypolimnion (Ansari et al., 2013). L'eutrophisation est l'un des sérieux problèmes de

pollution de l'eau affectant directement la faune aquatique (poisson...etc) en raison de la diminution de l'oxygène dissous au-dessous du niveau critique (anoxie).

I.5.6.3.3.5. La demande biochimique en oxygène

Plusieurs auteurs ont montré une corrélation élevée entre la biomasse algale exprimée en terme de chlorophylle-a et la demande biochimique en oxygène DBO₅ (**Xiao-long et al., 2007; Kaneko et al., 2014; Heiskary et Markus, 2001**). En effet, la charge totale de la DBO₅ d'un plan d'eau est constituée de la charge des apports extérieurs et de la charge résultant de la production primaire de matière organique par le phytoplancton dans le plan d'eau. La production primaire peut être convertie en DBO₅, avec un coefficient de 0,5 de la production primaire brute (exprimée en oxygène) (**Jørgensen, 2005**).

La modélisation du phénomène de l'eutrophisation se réfère donc à simuler et/ou à prédire l'un des indicateurs de l'état trophique d'un plan d'eau (lac du réservoir) que ce soient la concentration en chlorophylle-a (Chl-a), la concentration en phosphore total (TP), la transparence de l'eau exprimée en mètre et mesurée à l'aide de disque de Secchi (m) et parfois la concentration en oxygène dissous (**Kuo et al., 2007 ; Huo et al., 2013**), ou la modélisation de la DBO₅, et l'indice de l'état trophique (TSI) lui-même.

I.5.7. Eutrophisation en Algérie : synthèse bibliographique

En Algérie le problème de l'eutrophisation des retenues de barrages est répandu où les rejets en éléments nutritifs, tant ponctuels que diffus, ne cesse d'augmenter avec l'accroissement de la densité de population et l'intensification des activités agricoles et industrielles. Ceux-ci représentent une menace pour la qualité des eaux des barrages. En effet, plus de 20 lacs de barrages souffrent du problème de l'eutrophisation. **Remini (2005)** a rapporté dans son ouvrage « problématique de l'eau en Algérie » que les barrages de Beni Bahdel, Bakhada, Ouizert, Bouhanifia, Fergoug, Oued Iekhal, Hammam Grouz et Oued Harbil sont tous menacés par l'eutrophisation.

Bensafia et Remini (2014) ont mis l'accent sur l'influence de la vase décantée au fond des barrages sur le phénomène de l'eutrophisation. Ils ont révélé que la vase (la boue décantée) provoque un relargage intense du phosphore ce qui favorise le développement excessif des phytoplanctons et des plantes aquatiques.

Bouid-Lagha et Djelita (2012) ont étudié le phénomène d'eutrophisation des eaux du Barrage de Hammam Boughrara, situé au Nord-Ouest de l'Algérie, en utilisant une base de données s'étalant sur une période de neuf ans (1999-2008), relative à 19 variables physico-

chimiques, et biotiques. Cette étude a permis de déterminer l'évolution dans le temps de chaque variable, les corrélations entre variables, le degré d'eutrophisation des eaux, les espèces phytoplanctoniques dominantes (e.g. chlorophycées, diatomophycées) ainsi que leur nutriment principal, le phosphore. Les résultats ont montré que la qualité de l'eau de retenue de ce barrage, en ce qui concerne le processus de l'eutrophisation, appartient à la classe hyper-eutrophique, c'est-à-dire un degré très prononcé d'eutrophisation.

L'étude de **Hamaidi et al (2009)** a permis de suivre la variabilité temporelle du phytoplancton dans la retenue du barrage de Ghrib (Ain Defla-Algérie) afin de déterminer la diversité du plancton de cette retenue. Ces auteurs ont mis en évidence que l'eau du barrage Ghrib comprend 67 espèces appartenant à 39 genres qui se répartissent dans les embranchements suivants: Chlorophytes, Cyanobactéries, Chrysophytes, Euglénophytes et Pyrrophytes. La succession des populations phytoplanctoniques est largement conditionnée par les paramètres physico-chimiques.

Belabbes (2012) a souligné l'amplitude de la pollution et l'hyper-eutrophie (eutrophisation très prononcée) des eaux de la retenue du barrage de Cheurfa II, qui est un ouvrage situé en zone semi aride à hiver tempéré (Ouest Algérien). Il a également élaboré un modèle mathématique adapté au site permettant d'estimer les concentrations en phosphore à partir de sa charge et de gérer ainsi la prolifération d'algues par la limitation des teneurs en nutriments.

Lors de l'étude de la qualité de l'eau du barrage Keddara, **Regieg (2002)** a déterminé le degré de trophie, à partir des paramètres qui provoquent la pollution d'un plan d'eau. Les résultats ont montré une concentration en phosphore assez élevée et une production du phytoplancton assez prononcée, ce qui indique un degré d'eutrophisation très élevé durant la période étudiée.

II. Modélisation de la qualité de l'eau et l'eutrophisation

II.1 Introduction

La qualité de l'eau est largement définie comme la composition physique, chimique, et biologique de l'eau par rapport à son usage prévu pour différents objectifs tels que l'alimentation en eau potable, la récréation, la santé des écosystèmes, et la sécurité de la santé humaine (Veatch et Humphrys, 1966; Richard et McQuisten, 1968). Le terme peut être appliqué à une seule caractéristique de l'eau (une seule variable) ou à un groupe de caractéristiques combinées dans un indice de qualité de l'eau.

Un modèle est une représentation simplifiée, relativement abstraite, d'un processus, ou d'un système, permettant de décrire, expliquer, simuler et prévoir le comportement de ce dernier. Les modèles peuvent être employés comme un outil d'aide à la décision.

Dans le domaine des eaux de surface, un modèle de qualité de l'eau est une représentation mathématique d'une rivière, d'un cours d'eau, d'un lac ou d'un réservoir. Ces modèles comprennent des équations et des algorithmes qui décrivent les processus affectant la température, l'oxygène dissous, le pH, l'alcalinité, les nutriments, la matière organique, les substances toxiques, les plantes aquatiques, les algues et / ou la matière en suspension.

Les écosystèmes aquatiques sont très compliqués et hautement organisés. La gestion de qualité des eaux exige des actions et des mesures pour s'assurer que les charges polluantes déversées dans un plan d'eau n'excèdent pas sa capacité d'assimiler les charges (capacité auto-épuratrice) tout en maintenant les niveaux de la qualité indiqués par les normes de qualité. Dans ce contexte l'utilisation des outils d'aide à la décision tel que les modèles de la qualité des eaux devient une priorité absolue.

La modélisation de la qualité de l'eau dans les écosystèmes d'eau douce est considérée comme un problème complexe en raison du nombre élevé de facteurs qui interviennent, et les différentes formes de la pollution qui peuvent être observées dont l'eutrophisation.

II.2 Les objectifs de modélisation de la qualité des eaux

Les modèles de qualité de l'eau de surface sont des outils très importants pour contrôler les eaux de surface. En général, la modélisation de la qualité des eaux a pour objectifs de (Zheng et Bennett, 1995) :

- ✓ Aider les gestionnaires des ressources en eau à une meilleure compréhension comment les eaux de surface changent en réponse à la pollution et comment les protéger.

- ✓ Simuler les changements de la concentration de polluants dans un plan d'eau (retenue de barrage, lac...). La concentration d'un polluant dans l'environnement peut augmenter ou diminuer due à de nombreux mécanismes.
- ✓ Simuler les effets de différents scénarios (scénarios d'aménagement) sur la qualité de l'eau pour un bassin versant.
- ✓ Évaluer la ou les valeurs d'un variable de qualité, pour lequel il y a de l'information limitée (lacunes ou données manquantes).
- ✓ Dans un sens scientifique - développer un modèle conceptuel clair basé sur toutes les informations disponibles et mieux comprendre le régime de transport du polluant: tester les hypothèses, s'assurer qu'elles sont cohérentes avec les principes directeurs et les observations, et quantifier les processus de contrôle dominants. Sans cette compréhension, un code de simulation ne peut être utilisé que comme une boîte noire, ce qui limite clairement l'application intelligente du modèle.
- ✓ Evaluer la situation de la qualité de l'eau et les causes de leur détérioration.
- ✓ Aider les utilisateurs à interpréter et prévoir les réponses de qualité de l'eau aux phénomènes naturels et à la pollution d'origine anthropique pour différentes décisions de gestion.
- ✓ Evaluer (prévoir) les situations futures de qualité de l'eau résultant de différentes stratégies de gestion. Par exemple, évaluer l'amélioration de la qualité de l'eau après la construction d'une nouvelle station d'épuration (traitement des eaux usées), ou l'effet de la croissance industrielle accrue et les décharges en effluents (**Daniel et al., 2005**).

II.3. Objectif de la prédiction de la qualité des eaux

L'utilité pratique de la prédiction (au futur) de la qualité des eaux (variables ou indices de qualité) est de développer un *système d'alerte précoce* ou un *modèle prévisionnel* capable de prédire en avance les fluctuations de la qualité des eaux et prévoir leur dynamique et leur comportement avant que la détérioration aura lieu, ce qui permet aux opérateurs ou gestionnaires des écosystèmes aquatiques (retenue de barrage, lacs, oued ...) d'anticiper leurs actions et d'éviter des situations indésirable.

II.4. Approche de modélisation de la qualité de l'eau

II.4.1. Approches conventionnelles de modélisation de la qualité de l'eau

D'après **Cox (2003)**, un modèle de qualité de l'eau peut signifier n'importe quoi, une relation empirique, un ensemble d'équations de bilan de masses, ou un logiciel complexe. Les modèles de la qualité de l'eau peuvent être classés en trois classes fondamentales :

II.4.1.1 Les modèles physiques

Un modèle physique est une copie physique plus petite ou plus grande d'un système. La géométrie du modèle et l'objet qu'il représente sont souvent similaires dans le sens où l'un est un réajustement de l'autre (**Araghinejad, 2014**). Donc la modélisation physique ou la modélisation expérimentale consiste à représenter le système réel par un modèle à l'échelle géométriquement réduite et dynamiquement similaire et à effectuer des expériences sur celui-ci pour faire des observations et des mesures. Les résultats de ces expériences sont ensuite extrapolés aux systèmes réels.

II.4.1.2 Les modèles empiriques (statistiques)

Les modèles empiriques sont généralement exprimés en simples relations mathématiques provenant de l'ajustement statistique d'une équation avec les données observées indépendamment des lois régissant les mécanismes ou les processus impliqués. Par exemple la régression linéaire de la chlorophylle et les données de phosphore P est une simple relation empirique reliant la chlorophylle-a et le phosphore total (**Ji, 2008**). Parmi les inconvénients des modèles empiriques est qu'ils ont tendance à avoir une large erreur lors de la prédiction, notamment quand il n'y a aucune donnée spécifique au site (station d'échantillonnage) utilisées pour caler le modèle. (Plus le problème de l'extrapolation).

Dillon et Rigler (1974) ont développé une relation empirique pour l'estimation de la concentration moyenne en Chlorophylle-a (Chl-a) en été comme suit :

Si le ratio N/P de l'eau est > 12:

$$\log_{10}(\text{Chl. a}) = 1.45 * \log_{10}[(\text{P}). 1000] - 1.14 \quad (05)$$

Si le ratio N/P < 4, l'équation suivante a été établie, en se basant sur 8 cas étudiés:

$$\log_{10}(\text{Chl. a}) = 1.4 * \log_{10}[(\text{N}). 1000] - 1.9 \quad (06)$$

(N) et (P) sont exprimés en mg.l⁻¹ et (Chl-a) est exprimée en mg.l⁻¹. Si le ratio N/P est compris entre 4 et 12, la prise en compte de la plus petite valeur trouvée sur la base de deux équations est recommandée.

De nombreuses corrélations entre la concentration en phosphore et celle de chlorophylle-a ont été développées. **Dillon et Kirchner (1975)** a établi une relation entre transparence de l'eau exprimée par la profondeur de disque de Secchi (m) et la concentration de la chlorophylle-a (Chl.a). **Kristensen et al., (1990)** ont développé neuf équations, qui relient la concentration en phosphore avec la transparence moyenne de l'eau (la profondeur moyenne de disque de Secchi (Z_{eu})).

II.4.1.3 Les modèles mécanistes

Les modèles mécanistes de qualité de l'eau sont des modèles qui expriment les mécanismes et processus qui provoquent les changements sur la qualité de l'eau en utilisant des expressions mathématiques. Ils comprennent généralement un ou plusieurs groupes d'algues, des nutriments inorganiques et organiques (N, P, C) et l'oxygène dissous (DO). Ces modèles simulent les processus de qualité de l'eau, y compris les apports externes, le recyclage des nutriments et la croissance des algues dans les rivières, les lacs et les estuaires. Ils intègrent généralement des caractéristiques telles que la circulation et le mélange, les sources ponctuelles et non ponctuelles, la photosynthèse, la température de l'eau, la dynamique de l'oxygène dissous, les comportements des différentes formes de nutriments, les effets des charges atmosphériques et la demande en oxygène de sédiment (SOD) (**Ji, 2008**).

Ces modèles exigent la connaissance plus ou moins précise des processus biogéo-physico-chimiques et hydrodynamiques relatifs au fonctionnement du système à modéliser. Ils peuvent apporter de bonnes estimations de variable(s) de la qualité de l'eau, mais généralement ils demandent une longue procédure du calage. Le calage est une tâche délicate et fastidieuse, il s'agit d'une procédure itérative, la valeur ou les valeurs initiales de paramètre(s) sont choisies de la littérature, ensuite, un petit changement sur le paramètre du modèle se fait en vue d'obtenir une bonne concordance entre les données simulées et celles observées, ces changements se font par 'trial-and-error' jusqu'à l'obtention d'une réponse acceptable, puis le modèle est validé sur un jeu de données supplémentaires différent de celui du calage. Les prédictions peuvent alors être faites avec une confiance prédéfinie. Contrairement aux modèles empiriques, les modèles mécanistes reflètent comment les changements dans la performance du système sont liés aux changements des entrées. La modélisation mécaniste constitue le fondement de la modélisation informatique, tandis que les approches de modélisation physique et empirique fournissent des informations précieuses pour le processus de modélisation mécaniste (**Ji, 2008**).

Depuis de nombreuses années, la modélisation de l'eutrophisation basée sur le bilan de masses a été commencée (**Chen, 1970; Di Toro et Matysik, 1980; Thomann et Fitzpatrick, 1975**).

Avec l'amélioration de la compréhension des processus de l'eutrophisation et l'accroissement continu de puissance des PC et l'expansion et le développement de technologies informatiques, des modèles mécanistes de la qualité de l'eau des lacs et des retenues de barrage ont été développés et appliqués afin d'étudier les problèmes de la qualité de l'eau (**Chapra, 1997**).

Ces modèles utilisent des formulations mathématiques pour présenter les processus impliqués qui sont par la suite résolus par les méthodes numériques. Ils sont généralement basés sur l'hypothèse de l'homogénéité spatiale et le principe de la conservation des masses, des moments et de l'énergie. Dans ce cas là, ces modèles sont très compliqués en raison de la complexité de l'écosystème aquatique qui renferme en son sein des mécanismes de feedback (rétroactif, régulation et interaction). Ils ont besoin de nombreux paramètres expérimentaux pour qu'ils puissent fonctionner convenablement, le calibrage de ces modèles est généralement une tâche effrayante et fastidieuse et leur développement exige souvent la compréhension exacte des processus impliqués (**Jørgensen et Bendriccho, 2001**).

Le recours aux méthodes de modélisation telle que les techniques d'*intelligence artificielle* constitue une alternative intéressante.

II.4.1.3.1. Classification des modèles mécanistes de la qualité des eaux

D'après **Tsakiris et al., (2012)**, les modèles de la qualité des eaux peuvent être classés en plusieurs catégories selon:

La nature du modèle (déterministes ou stochastiques).

La dimension (Modèle 1D, 2DH, 2DV et 3D).

La nature de polluants (les nutriments, sédiments, sels etc..).

Le domaine d'application (bassin versant, eaux souterraines, système de fleuve, eaux côtières...etc).

L'état analysé (état d'équilibre (steady state) ou simulation dynamique).

Les besoins en matière de données (modèle exigeant une large base de données, modèles parcimonieux).

II.4.1.4. Modélisation de l'eutrophisation en utilisant les approches conventionnelles: synthèse bibliographique

Dans les écosystèmes aquatiques, l'eutrophisation et les produits chimiques toxiques sont des facteurs importants qui influent sur la qualité de l'eau ainsi que sur la santé des écosystèmes. Dans ce contexte, les modèles de simulation et de prévision de l'eutrophisation sont devenus de plus en plus utiles et populaires, ils permettent de comprendre et représenter les dynamiques écologiques et hydro-biogéochimiques de l'écosystème aquatique qui mènent à l'eutrophisation.

Les méthodes de modélisation de la qualité de l'eau et de l'eutrophisation existantes sont principalement divisées en deux catégories: Les méthodes de modélisation statistiques et les méthodes de modélisation dites mécanistes (**Xu et al., 2015; Ozgur and Singh, 2016**).

La première approche est basée sur des modèles dits statistiques. Ils cherchent à prévoir un ou des descripteurs de l'eutrophisation en fonction d'un certain nombre de variables causales mesurées sur le terrain (**Pinay et al., 2017**), ils permettent alors de construire des lois générales en se basant sur les observations indépendamment des mécanismes sous-jacents. Ces modèles bénéficient d'aides semi-automatiques (par tests de significativité) pour éliminer des variables d'entrée superflues. L'enrichissement des bases de données permettra de calibrer ces modèles peu génériques dans une grande diversité de sites (**Pinay et al., 2017**).

Plusieurs modèles statistiques ont été élaborés pour l'étude des relations entre la charge en phosphore provenant du bassin versant et la concentration de phosphore dans les lacs (**Vollenweider, 1968 in Harper, 1992**) et entre la concentration de phosphore dans les lacs et la chlorophylle phytoplanctonique (**Sakamoto, 1966 in Harper, 1992; Dillon et Rigler, 1974**).

Une deuxième approche est dite mécaniste, elle représente, au travers d'équations, les mécanismes hydrobiogéochimiques et écologiques et simule la dynamique de l'eutrophisation. (**Pinay et al., 2017**).

Les modèles mécanistes classiques de la qualité de l'eau comprennent: le modèle MIKE (**Xu et al., 2015; Charalampos et al., 2012**), le programme de simulation d'analyse de la qualité de l'eau (Water quality Analysis Simulation Program ou WASP) (**Wang et al., 2018; Quijano et al., 2017**), le modèle EFDC (Environmental Fluid Dynamics Code) (**Libin et al., 2016; Xi et Xuyong, 2018**), etc.

Au cours de ces dernières décennies, de nombreux modèles prédictifs ont été développés pour prévoir la prolifération des algues (par exemple, **Perez et al., 2000; Walsh et al., 2001; Baird et al., 2004; Ahn et al., 2007; Lui et al., 2007; Li et al., 2007; Jia et al., 2010;**

Salacinska et al., 2010). Ces modèles reposent principalement sur différents algorithmes et équations décrivant les processus dynamiques d'eutrophisation et la plupart d'entre eux utilisaient la chlorophylle-a comme étant la sortie des modèles.

Kuo et al. (2006) ont utilisé le modèle CE-QUAL-W2 pour simuler l'eutrophisation dans les réservoirs de Te-Chi et de Tseng-Wen à Taiwan. Après le calage du modèle, les effets des mesures de réduction de la charge en éléments nutritifs entrant dans le réservoir sur la qualité de l'eau de celui-ci ont été simulés dans différents scénarios. Les résultats de la simulation ont montré qu'une réduction de 30 à 55% de la charge en phosphore améliorerait les conditions de qualité du réservoir de Te-Chi de l'eutrophie-mésotrophie à l'oligotrophie.

Wu et Xu (2011) ont utilisé le modèle EFDC pour simuler le processus d'eutrophisation dans le lac Daoxiang, Beijing, la Chine. Pour faire fonctionner le modèle d'eutrophisation du lac, un échantillonnage sur le terrain a été effectué entre mars et octobre 2008, avec un intervalle de 10 à 20 jours. Les résultats montrent que les valeurs de chlorophylle-a simulées sont en concordance, pour la plupart des cas, avec celles observées, à l'exception de la dernière période de la station 2# et que la précision moyenne de la prévision de l'efflorescence algale est de 63,43%. Ces auteurs ont confirmé que le modèle EFDC pouvait être utilisé pour la simulation de la concentration de la chlorophylle-a et l'efflorescence algale dans le lac Daoxiang.

Afshar et Saadatpour (2009) ont appliqué le modèle CE-QUAL-W2 pour simuler les paramètres de température et de qualité de l'eau, notamment le phosphore total, les nitrates, la chlorophylle-a, l'oxygène dissous et l'ammonium dans le réservoir de Kharkheh en Iran. Après avoir calé le modèle, une analyse de sensibilité a été réalisée pour déterminer l'efficacité des paramètres significatifs dans la modélisation numérique. Les résultats obtenus étaient prometteurs.

Cependant, les modèles mécanistes de qualité de l'eau contiennent généralement un grand nombre de paramètres inconnus, qui possèdent une gamme de valeurs empiriques. Sur cette base, il est difficile de prévoir les changements de qualité de l'eau pour répondre aux exigences de précision. De nombreux modèles de l'eutrophisation combinent ces deux approches, selon la disponibilité des données sur une zone spécifique.

II.4.2. Approches basées sur l'intelligence artificielle

Les techniques de l'intelligence artificielle, y compris la logique floue et les réseaux de neurones artificiels (RNA), impliquent l'approximation des fonctions reliant les variables d'entrées aux variables de sortie (réponse) pour résoudre des problèmes complexes non linéaires (**Kisi et al., 2009**). Ces caractéristiques ont fait des techniques de l'intelligence artificielle une alternative viable aux méthodes conventionnelles.

L'objectif principal de la modélisation pilotée par les données (data-driven models) est de trouver des informations latentes sous la forme de régressions multiples en utilisant l'apprentissage automatique à partir des ensembles de données. La régression multivariée est classée sous deux rubriques: linéaire et non linéaire. La plupart des méthodes statistiques appartiennent à la catégorie linéaire. Dans le cas contraire, la régression fondée sur les connaissances, telle que la modélisation pilotée par les données, le data mining et l'intelligence artificielle, etc., sont des méthodes non linéaires.

Le champ de l'intelligence artificielle est très vaste, il inclut différentes catégories dont la Machine Learning, Data-Driven Models, le Soft Computing et Data Mining, ces groupes peuvent se chevaucher les uns des autres sans avoir des limites claires.

La modélisation d'un système aquatique est l'un des grands défis dans le domaine de la gestion de la qualité de l'eau et l'ingénierie environnementale, les modèles de type « data driven models » se sont avérés comme un puissant outil pour surmonter ce défi, en raison de leur aptitude de construire un modèle à partir des données, c'est-à-dire leur capacité de capturer les relations entre les variables explicatives et la variable à expliquer à partir des données elles-mêmes sans faire recours aux équations et lois physiques régissant le phénomène étudié.

Contrairement aux modèles traditionnels, les modèles de type « data driven models » sont fondés sur l'apprentissage numérique des données qu'ils modélisent dans le but de trouver une forme de relation (s) qui relie le mieux possible les ensembles de données (**Price, 2002**). Un modèle de type « data driven models » emprunte des techniques développées dans des domaines tels que les statistiques, l'informatique et l'intelligence artificielle (réseaux neuronaux artificiels et algorithmes génétiques), les Model Tree (MT) et le Soft Computing, qui comprend principalement la logique floue, les réseaux neurones artificiels et le raisonnement probabiliste. Parfois, des "modèles hybrides" sont construits combinant deux types de modèles.

Au cours des dernières années, on a eu de plus en plus tendance à utiliser les modèles de type « data driven models » pour compléter, voire remplacer, les modèles mécanistes, en particulier pour la prévision. Bien qu'un exemple simple de modèle « data driven models » soit un modèle de régression linéaire, les modèles plus complexes sont fortement non-linéaires et utilisent des techniques sophistiquées. Ces modèles basés sur les données se sont avérés attrayants dans la modélisation de la dynamique des espèces d'algues en raison des limites des connaissances écologiques (**Recknagel et al., 2002; Soyupak et Chen, 2004; Jung et Price, 2005; Karimipour et al., 2005**).

Les approches de modélisation pilotées par les données « data driven models », sont fondées sur l'exploration et l'analyse de grandes quantités de données par des moyens automatiques ou semi-automatiques afin de découvrir les structures des modèles (pattern), les relations et les règles significatifs impliquent l'opération d'apprentissage dans un sens pratique et non théorique (**Jung, 2009**).

Le problème de la gestion de la qualité de l'eau repose généralement sur des informations imprécises et insuffisantes. La plupart du temps, les objectifs ou les contraintes ne peuvent pas être définis précisément en raison du fait qu'ils sont basés sur des exigences subjectives et mal définies issues du jugement humain ou des préférences. Bien que les modèles numériques soient disponibles pour la simulation de la qualité de l'eau, les incertitudes et l'imprécision ne sont pas bien couvertes dans ces modèles. De plus, la nécessité de caler les modèles de qualité de l'eau rend les réseaux de neurones artificiels avantageux sur ces modèles conventionnels (**Lobrecht et al., 2002**).

La puissance et l'avantage pratique des réseaux neurones artificiels résident dans leur capacité de représenter des relations linéaires et non linéaires et dans leur aptitude d'apprendre ces relations directement des données qu'ils modélisent. Les modèles linéaires traditionnels sont inadéquats quand ils viennent à modéliser les données qui contiennent des caractéristiques non linéaires.

Les questions de la qualité de l'eau comportent de nombreuses incertitudes inhérentes et l'approche de la logique floue est particulièrement adaptée aux problèmes incertains avec des informations imprécises. Les normes de qualité de l'eau peuvent être définies par des transitions progressives entre un niveau de bonne qualité à un niveau de mauvaise qualité (inadaptés). Une étude comparative sur l'utilisation de la logique floue et la régression linéaire multiple pour l'analyse chimique de la qualité de l'eau et les tests de goût est réalisée par **Iwanaga et al (1997)**. L'étude a été effectuée sur un ensemble de données avec plusieurs

variétés d'eau de haute qualité au Japon. Le modèle de la logique floue était plus performant que le modèle de la régression linéaire multiple (**Lobbrecht et al., 2002**).

II.5 Justification de l'utilisation de l'intelligence artificielle pour modéliser la qualité des eaux et l'eutrophisation

La plupart des données environnementales ont certains genres de relations entre les variables (les unes aux autres), ces relations qui peuvent être dévoilées entre des variables environnementales pouvant avoir peu ou rien à voir avec les principes physiques et les mécanismes régissant les processus. Ces liaisons entre les variables n'impliquent pas forcément qu'une variable entraîne l'autre, mais qu'il existe des associations significatives entre elles, qui sont difficiles à exprimer par les équations mathématiques (**Jung, 2009**). C'est l'une des raisons pour lesquelles cette étude fait appel aux modèles fondés sur l'intelligence artificielle pour la modélisation de la qualité de l'eau et l'eutrophisation.

Comme mentionné ci-dessus, un grand nombre de facteurs affectant la qualité de l'eau ont des relations non linéaires compliquées avec les variables, de nouveaux types de modèles basés sur l'IA et les techniques de soft computing ont été récemment appliquées à la modélisation de la qualité de l'eau. Parmi ces outils on cite le soft computing, les RNAs, la LF et les AG qui sont plus populaires dans l'ingénierie des ressources en eau. La tendance récente est d'appliquer un système hybride tel que les RNA et le FL, les RNA et le Krigeage, RNA et AG etc. afin d'améliorer l'efficacité et performance de la modélisation. Cependant, leurs utilisation s'applique largement aux problèmes de qualité des eaux de surface et des eaux souterraines car elles offrent une approche attrayante et rentable pour la simulation et la prévision de la qualité de l'eau (**Sharma et al., 2003; Chaves et Koriji, 2007**). Contrairement aux approches conventionnelles, les RNA n'exigent pas la caractérisation et la quantification explicites des propriétés physiques et des conditions des systèmes étudiés. Les RNAs apprennent le comportement du système à partir de données représentatives. La possibilité d'apprendre et de généraliser à partir de paires de données adéquates permet aux RNAs de résoudre des problèmes complexes. La logique floue et particulièrement les systèmes d'inférence flous (FIS) constituent un moyen simple et commode pour construire des modèles en décrivant le comportement du système en utilisant le raisonnement humain (expert), et en définissant la relation explicite par l'utilisation du langage naturel sous forme de règles floues (règles linguistiques) sans utiliser des équations analytiques complexes.

La modélisation de la qualité de l'eau est la base du contrôle de la pollution de l'eau. Les modèles sont utilisés pour prédire les tendances de la qualité de l'eau en se basant sur les

conditions actuelles de l'eau, y compris les concentrations de polluants. Plusieurs modèles déterministes et stochastiques de qualité de l'eau ont été élaborés pour gérer les meilleures pratiques de conservation de la qualité de l'eau (**Hull et al., 2008; Einax et al., 1999**). La plupart de ces modèles sont très complexes et nécessitent une quantité importante de données de terrain pour soutenir l'analyse. De plus, de nombreux modèles de qualité de l'eau fondés sur des statistiques supposent que la relation entre les variables de sortie et celles d'entrées est linéaire et normalement distribuée. Comme la qualité de l'eau peut être affectée par de nombreux facteurs, les méthodes traditionnelles de traitement des données ne sont plus suffisantes pour l'analyse (**Xiang et al., 2006**) car de nombreux facteurs présentent des relations complexes non linéaires avec les variables de prévision de la qualité de l'eau. Par conséquent, l'utilisation d'une approche statistique n'offre généralement pas une grande précision. Récemment, des réseaux de neurones et les systèmes d'inférence flous ont été appliqués dans de nombreuses branches de la science, particulièrement, dans l'ingénierie hydraulique et environnementale.

II.6 Modélisation de la qualité de l'eau par l'intelligence artificielle :

synthèse bibliographique

L'évaluation des ressources en eau nécessite des connaissances et une compréhension complète des processus affectant à la fois la quantité et la qualité de l'eau. La quantité de l'eau est relativement facile à évaluer, elle est déterminée par un seul paramètre; la masse de l'eau, le volume de l'eau ou le débit (m^3/s). Cependant, l'évaluation de la qualité de l'eau est difficile et représente une tâche fastidieuse à accomplir (**Abbasi et Abbasi, 2012**). La qualité de l'eau est devenue un facteur de plus en plus important dans la gestion intégrée des ressources en eau. Elle est représentée par diverses variables de nature physico-chimiques et biologiques.

Des articles récents ont rapporté que les techniques de Machine Learning telles que les réseaux de neurones artificiels (**Csábrági et al., 2016; Mohammadpour et al., 2016; Keshtegar et al., 2017; Heddami, 2017**), les systèmes d'inférence flous (**Angulo et al., 2012; Emamgholizadeh et al., 2014**), la programmation d'expression de gène (GEP) (**Mohammadpour et al., 2016; Sattar et al., 2017**) et le support vector machine (**Noori et al., 2016; Ji et al., 2017**) peuvent offrir une alternative prometteuse pour modéliser la qualité de l'eau et l'eutrophisation.

Les réseaux de neurones artificiels (RNA) ont été utilisés avec succès pour modéliser la DBO₅, plusieurs auteurs ont testé efficacement les RNAs dans les rivières (**Dogan et al.,**

2009, Singh et al., 2009; Šiljić et al., 2015; Šiljić Tomić et al., 2016; Raheli et al., 2017) et dans les stations d'épuration (Ozkan et al., 2009; Dogan et al., 2008; Heddam et al., 2016). Les résultats obtenus étaient satisfaisants.

Chen & Lo (2010) ont utilisé les « gray models » et les réseaux de neurones avec l'algorithme de rétropropagation (BPNN) pour prédire la demande biochimique en oxygène (DBO₅), la demande chimique en oxygène (DCO) et la matière en suspension (MES) pour les stations d'épuration. Les résultats ont montré que la précision était meilleure lors de l'utilisation du RNAs pour la prédiction de la DBO₅.

Han et al. (2011) ont utilisé un réseau de neurones à fonction radiale et structure flexible (FS-RBFNN) pour la modélisation de la qualité de l'eau dans le processus de traitement des eaux usées, les résultats ont montré que les meilleures valeurs de MSE et RMSE étaient obtenues. En outre, le FS-RBFNN obtient la meilleure précision à la fois pour les valeurs minimales et maximales de la DBO₅.

La logique floue et, en particulier, les systèmes d'inférences floues (FIS) ont été utilisés pour créer différents types d'indices de qualité de l'eau pour les fleuves (**Lermontov et al., 2009; Ocampo-Duque et al., 2006**), les barrage-réservoir (**Gharibi et al., 2012**), l'eau souterraine (**Sahu et al., 2011**) et la ferme aquacole de crevettes (**Carbajal et al., 2012**), et système de distribution d'eau potable (**Aghaarabi et al., 2017**).

Kotti et al (2013) ont utilisé une méthodologie basée sur des systèmes d'inférence floue (FIS) en vue de développer des modèles de prédiction de la demande biochimique en oxygène (DBO₅) dans les zones humides construites en surface libre. Les prédictions du modèle flou ont montré une bonne concordance avec les données observées.

Les tendances actuelles de modélisation s'orientent vers l'utilisation de systèmes hybrides comme le système d'inférence neuro-flou adaptatif (ANFIS), ce système exploite la capacité d'apprentissage des réseaux de neurones artificiels (RNA) pour la génération automatique des règles et l'optimisation des paramètres directement à partir des données plutôt que de consulter un expert (**Yan et al., 2010**). Ces dernières années, plusieurs études (**Najah et al., 2014; Ay et al., 2016; Ahmed et al., 2017**) ont appliqué ANFIS, avec succès, pour modéliser la qualité de l'eau.

II.7 Modélisation de l'eutrophisation par l'intelligence artificielle : synthèse bibliographique

La qualité de l'eau est un élément crucial dans le domaine de la gestion de ressource de l'eau, pourtant, celle-ci n'a pas reçu l'attention et la préoccupation qu'elle mérite.

L'objectif principal de ce chapitre est de révéler l'état de l'art de la modélisation de l'eutrophication par les techniques de 'machine learning' tel que les réseaux de neurones artificiel, les systèmes d'inférence flous (FIS) et les systèmes d'inférence neuro-flous (ANFIS). Ces techniques peuvent être considérées comme alternative aux approches classiques. Elles contribuent à la gestion intégrée de la qualité des eaux.

L'eau est une ressource primordiale pour toutes formes de la vie et pour les activités humaines, elle devrait être disponible, en quantité et en qualité, en vue d'assurer le bien-être des citoyens.

Cependant, la croissance démographique, l'industrialisation, l'extension et l'intensification des activités d'agricoles ont contribué à l'augmentation continue des charges en nutriments déversés dans les oueds, les lacs et les réservoirs, ce qui peut entraîner des problèmes sérieux sur la qualité de l'eau telle que la pollution, l'eutrophication et l'efflorescence algale associée, provoquant la détérioration de qualité de l'eau.

L'eutrophisation est devenue un problème sérieux dans beaucoup de fleuves, lacs et réservoirs partout dans le monde. Ce problème suscite une grande préoccupation des chercheurs et des décideurs.

Les signes néfastes d'une efflorescence algale accélérée sont la formation de l'écume verte (mousse), menant à l'abaissement du taux d'oxygène au-dessous du seuil acceptable, la mort probable de poissons, la diminution de la clarté de l'eau, des odeurs nauséabondes et du mauvais goût de l'eau, ce qui causent un coût surélevé dans l'opération de traitement de l'eau potable, pour cette raison le développement des outils (modèles) permettant de comprendre, contrôler et prévoir le processus de l'eutrophication est d'un intérêt crucial.

Avec l'accroissement accru dans la capacité des ordinateurs et l'expansion et le développement des technologies informatiques, beaucoup d'outils ont été conçus pour gérer et contrôler la qualité de l'eau. Actuellement, la modélisation constitue un outil fondamental dans presque tous les domaines puisque les modèles nous permettent de mieux comprendre un système, de simuler et prévoir son comportement (**Heddam et al., 2011**).

Un modèle est une représentation simplifiée, relativement abstraite, d'un processus, ou d'un système, permettant de décrire, expliquer, simuler et prévoir ce dernier. Les modèles peuvent être employés comme étant un outil d'aide à la décision.

L'eutrophisation a été simplifiée en un problème de type entrées-sortie (inputs-output) (**Karul et al., 2000**), et elle a été traitée par les RNA selon cette philosophie, car la présentation des entrées-sortie au réseau pendant la phase d'apprentissage permet de stocker les informations sur le processus d'eutrophisation sous forme des connexions inter neuronales connues sous le nom de poids synaptiques, après l'apprentissage, le réseau sera validé avec un jeu de données différent du premier, si les résultats de validation sont satisfaisants, le modèle neuronal créé peut être utilisé pour simuler l'eutrophisation.

French et Recknagel (1994) étaient les pionniers qui ont examiné les RNA afin de prévoir la formation de l'efflorescence algale à partir des variables de la qualité de l'eau. Dans leur article, un RNA de type 'feed-forward' a été testé pour prévoir l'abondance des espèces phytoplanctoniques dans le réservoir de Saldenbach, Allemagne.

Les RNA ont été largement utilisés pour la modélisation des efflorescences algales. Plusieurs auteurs ont testé RNA dans des lacs et réservoirs (**Recknagel et al., 1997, Yabukana et al., 1997**), des rivières (**Whitehead et al., 1997, Maier et al., 1998**) et des systèmes côtiers (**Belcano et al., 2001, Lee et al., 2003**).

De nombreux articles ont traité la modélisation des changements temporels d'espèces d'algues dans les systèmes d'eau douce par le biais de réseaux de neurones artificiels (**Whitehead et al., 1997, French et al., 1998**). Ils ont montré que les RNAs sont capables de simuler les tendances de la dynamique de la croissance algale. Plus tard, **Karul et al. (2000)** ont appliqué une technique de réseau de neurones artificiels, un algorithme d'apprentissage par feed-back de type Levenberg-Marquardt à trois couches a été utilisé dans trois plans d'eau en Turquie (le barrage de Keban, les lacs de Mogan et Eymir), afin de modéliser le processus d'eutrophisation représenté par la chlorophylle-a.

Malgré la nature très complexe et extraordinaire du barrage de Keban, une corrélation relativement bonne a été obtenue entre les valeurs mesurées et celles calculées de la chlorophylle-a, avec un coefficient de corrélation oscillant entre 0,60 et 0,75. Pour les lacs Mogan et Eymir, une corrélation élevée a été atteinte ($R = 0,95$). De plus, les modèles ont pu estimer avec succès certaines valeurs extrêmes provenant des jeux de données de validation et de test qui n'ont pas été utilisés dans la phase d'apprentissage du RNA. À partir de ces exemples, on peut conclure que les RNA offrent de bons résultats, indiquant que ces modèles

peuvent être utilisées avec succès pour modéliser le processus de l'eutrophisation dans les écosystèmes d'eau douce.

Chen et al. (2006) dans leur étude, ont utilisé les RNA pour extraire les facteurs dominants qui déterminent la présence d'efflorescences algales nuisibles (EAN) au moyen de huit combinaisons de variables de qualité de l'eau (huit scénarios), en vue de prévoir les EAN une semaine à l'avance dans la baie de Xiamen. Les RNA se sont avérés être des outils simples et appropriés pour la sélection des facteurs de forçage en vue de prédire l'apparition de l'efflorescence algale nuisible.

Kuo et al. (2007) ont estimé la concentration en chlorophylle-a (Chl-a) dans le réservoir de Te-Chi (Taïwan) en utilisant un modèle de réseau de neurones à propagation arrière à trois couches (BP-NN). Les variables d'entrée étaient le mois, la température de l'eau, le pH, l'oxygène dissous, la profondeur de Secchi, la matière en suspension (SS), le NO₃-N et le PO₄-P. Des coefficients de corrélation de 0,92 pour l'apprentissage et de 0,86 pour l'ensemble de test ont été enregistrés entre les valeurs mesurées et simulées de la chlorophylle-a (Chl-a). Ces résultats ont démontré que les réseaux de neurones sont capables de modéliser la concentration en chlorophylle-a avec une précision raisonnable, suggérant que le réseau de neurones est un outil fiable pour la gestion des réservoirs.

Plus récemment, **Millie et al. (2012)** ont employé un réseau de neurones artificiels holistiques, un modèle de Gray-Box et la régression linéaire multiple (RLM) pour modéliser l'abondance en microalgues (en termes de Chl-a), dans la baie de Sarasota (sud-ouest de la Floride, USA) -qui est un estuaire lagunaire productif-, en utilisant des variables de qualité de l'eau comme inputs. Le modèle RNA holistique et le modèle Gray-Box ont clairement surmonté la RLM dans la modélisation de la concentration en chlorophylle-a (Chl-a), mais les RNA holistiques ont montré des résultats légèrement meilleurs que le modèle de Gray-Box.

Les réseaux de neurones artificiels appartiennent à la catégorie de modèle boîte noire (**Babovic, 2009**), par conséquent, il y a un manque d'interprétabilité de modèle de RNA, en d'autres termes, ils n'ont pas une structure lisible et compréhensible (les modèles ANN ne sont pas transparents), parce que les réseaux de neurones artificiels permettent de représenter des relations inconnues incorporées dans les données sous forme de poids synaptiques. Ils ne sont pas capables d'expliquer la relation causale entre les entrées et la réponse (sortie) (**Arthur, 2010**). En plus de ça, la majeure partie de nos informations sur l'écosystème aquatique est de nature qualitative plutôt que quantitative (**Chen et Mynett, 2004**). L'utilisation de la logique floue permet de pallier ces inconvénients en décrivant le comportement du système en utilisant le raisonnement humain (expert), en définissant une

relation explicite par l'utilisation du langage naturel sous la forme de règles floues «si.....alors.....» (ou règles linguistiques). Cette méthode est également capable d'intégrer et de prendre en compte les informations inexacts, vagues, qualitatives et floues qui entourent souvent la qualité de l'eau.

Un nombre important d'activités de recherche ont été menées sur l'utilisation de la logique floue dans la modélisation de l'eutrophisation et la gestion des lacs, en particulier au cours de ces dernières années. **Chen et Mynett (2003)**; **Marsili-Libelli (2004)**; **Laanemets et al. (2006)** ont appliqué la logique floue (système d'inférence floue) avec succès en vue de modéliser le processus d'eutrophisation.

Sylaios et al. (2008) ont développé un modèle flou de type Takagi-Sugeno nommé CHLfuzzy sous l'environnement MS-Excel (tableur), en utilisant la température de l'eau, la teneur en oxygène dissous, la concentration en azote inorganique dissous et le rayonnement solaire comme inputs pour modéliser la concentration en chlorophylle-a (output) dans la lagune de Vassova (Nord de la Grèce). Les données sont divisées en deux parties, la première est utilisée pour l'apprentissage des modèles CHLfuzzy et les modèles ANFIS, la seconde sert à la validation. Les valeurs de sortie du CHLfuzzy sont comparées avec celles obtenues par ANFIS. Les deux méthodes ont montré un niveau de performance satisfaisant au cours de la période d'apprentissage, mais ils ont montré une performance plus ou moins bonne durant la phase de validation avec un R^2 oscillant entre 0,67 et 0,77 pour ANFIS et CHLfuzzy, respectivement. Par conséquent, le modèle CHLfuzzy a démontré des résultats légèrement meilleurs plutôt que le modèle ANFIS pendant la validation.

Marsili-Libelli (2004) a décrit la conception d'un modèle flou servant à prévoir les efflorescences macro-algales dans un lagon côtier et à décider la stratégie de la récolte des algues en conséquence. Ce modèle flou est basé sur un ensemble de règles floues dérivées d'observations expérimentales et de connaissances de spécialistes, tout en utilisant les fluctuations quotidiennes de variables simples de qualité de l'eau tels que l'oxygène dissous, le potentiel d'oxydoréduction, le pH et la température. Les résultats révèlent que ce modèle a de bonnes capacités prédictives, qui pourraient être améliorées lorsque plus de données seront disponibles.

Shen et Chouchoulas (2001) ont développé un estimateur « fuzzy-Rough » de la population algale (FuREAP), qui est un système hybride intégrant la théorie des ensembles flous et celle d'ensembles rugueux (rough set). Ce modèle a permis d'estimer la concentration de sept espèces d'algues dans la rivière en utilisant la concentration chimique des variables de la qualité de l'eau et d'autres paramètres comme la saison, la taille de la rivière et le débit. Les

résultats prouvent que le modèle développé est très performant et l'erreur oscille entre 8 et 16%.

L'étude de **Soyupak et Chen (2004)** a examiné l'applicabilité de la logique floue pour estimer la production primaire pseudo-stationnaire dans les réservoirs complexes - le réservoir du barrage de Keban- qui est l'un des grands réservoirs au monde. Un ensemble de données de dix variables de qualité de l'eau a été utilisé. Le modèle flou surpasse les modèles de régression linéaire multiple, et les résultats obtenus prouvent que ce modèle est un outil utile dans la gestion des lacs et des réservoirs, puisque une fois le modèle flou est développé, les valeurs de chlorophylle-a peuvent être prédites en temps réel au moyen de simples variables de qualité de l'eau.

La logique floue est également utilisée pour créer différents types d'indices de qualité de l'eau pour les rivières (**Lermontov et al., 2009, Ocampo-Duque et al., 2006**), les réservoirs (**Gharibi et al., 2012**) les eaux souterraines (**Sahu et al., 2011**) et les fermes d'élevage de crevettes (**Carbajal et al., 2012**), car un expert humain ne peut à lui seul gérer une grande quantité de données sur la qualité de l'eau.

Cependant, la détermination de paramètres des fonctions d'appartenance et les règles floues demeure difficile (**Chen et Mynett, 2004 ; Lau et al., 2001; Lin et al 2001 ; Shen et Chouchoulas, 2001**), l'utilisation de systèmes hybrides comme le système d'inférence neuro-floue adaptative (ANFIS) résout ce problème en utilisant efficacement la capacité d'apprentissage des réseaux de neurones artificiels (RNA) pour la génération automatique des règles et l'optimisation des paramètres directement à partir des données plutôt que de consulter un expert (**Yan et al., 2010**).

Une attention insuffisante a été accordée à l'utilisation des systèmes d'inférence neuro-flous adaptatifs (ANFIS) en vue de modéliser le processus de l'eutrophisation, l'efflorescence algale, la concentration en chlorophylle-a, la production primaire de phytoplancton et l'abondance de la biomasse.

Au cours des dernières années, certains auteurs ont utilisé le système d'inférence neuro-floue adaptative (ANFIS) pour classer l'état de la qualité de l'eau. **Yan et al. (2010)** dans leur article ont développés un modèle ANFIS pour évaluer et classer l'état de qualité de l'eau, en se basant sur les variables de qualité de l'eau telle que l'oxygène dissous, la demande chimique en oxygène et l'ammoniac (NH₃-N) comme étant inputs, la sortie est l'état de la qualité de l'eau représenté par six classes (1,2 ... 6). Le modèle ANFIS a montré des résultats satisfaisants, il peut modéliser 89,59% des états de qualité de l'eau de la rivière.

Chen et Mynett (2003) ont utilisé des techniques de « data mining » et les connaissances heuristiques pour la réduction de la dimensionnalité, la définition de fonction d'appartenance et l'induction de règles floues (si.....alors....) pour modéliser l'eutrophisation en utilisant la logique floue dans le lac eutrophique de Taihu. La méthode a donné des résultats prometteurs dans la modélisation de la biomasse algale.

Les modèles de l'eutrophisation sont la catégorie des modèles qui ont suscité un grand intérêt et une grande attention à la communauté scientifique. Divers modèles présentant un large spectre de complexité ont été développés et appliqués. La complexité du modèle dépend de la disponibilité des données et de l'écosystème (**Jørgensen et al., 2001**).

Une révision sur la modélisation de l'eutrophisation par l'intelligence artificielle a été présentée dans ce chapitre. Un grand nombre d'articles traitant de la modélisation de l'eutrophisation par les RNA ont été publiés. Une attention insuffisante a été accordée à l'utilisation du système d'inférence floue (FIS), et en particulier, du système d'inférence neuro-flou adaptatif (ANFIS) pour la modélisation de l'eutrophisation, d'où la nécessité d'explorer les capacités de ces deux techniques (FIS et ANFIS).

III. Matériels et méthodes

III. 1. La méthodologie de travail

D'importants efforts ont été déployés, au cours des vingt dernières années, pour tenter de répondre aux défis scientifiques et technologiques que posent la compréhension et la prévision du comportement des systèmes aquatiques en réponse à des perturbations naturelles ou artificielles. Dans ce contexte, et pour répondre aux problèmes de la gestion environnementale et de la qualité des eaux de barrages, nous explorerons les capacités des outils de l'intelligence artificielle afin de créer des outils d'aide à la décision ou modèles prévisionnels permettant de comprendre, gérer et prédire le problème de l'eutrophisation de certains barrages en Algérie.

Dans ce travail on va démontrer l'applicabilité des réseaux de neurones, des systèmes flous et d'autres techniques de Machine Learning dans les questions pratiques de la modélisation et de la gestion de la qualité de l'eau et de l'eutrophisation. Deux études de cas ont été sélectionnées: l'une concerne la base de données de qualité de l'eau de barrage de Ouizert et l'autre examine une base de données mixte relative aux barrages de Keddara, Lekhal, Tichy Haff et cela dans le but de démontrer la capacité des RNAs et des modèles neuro-flous dans la généralisation et l'approximation des fonctions et de surmonter certains problèmes relatifs au nombre limité de données d'apprentissage.

Tout d'abord, on va présenter en détail la théorie des réseaux de neurones artificiels, la logique floue et les systèmes d'inférences neuro-flous, ensuite on va aborder les différentes étapes pour la création des modèles basés sur ces nouvelles techniques, puis on va montrer les critères statistiques de performances utilisés pour juger la qualité des modèles développés, et à la fin on va présenter succinctement les barrages étudiés et les données de la qualité de l'eau utilisées.

Les modèles de réseaux de neurones artificiels (RNA) et les modèles neuro-flous (ANFIS) vont être étudiés dans ce travail pour la modélisation des indicateurs de l'eutrophisation. Les réseaux de neurones artificiels (RNA) sont des modèles mathématiques dont l'architecture est inspirée de réseaux de neurones biologiques. Ils sont fortement adaptés dans la modélisation des processus non linéaires comme la qualité des eaux et l'eutrophisation, l'avantage principal de la méthode de réseaux de neurones artificiels par rapport aux méthodes conventionnelles est leur capacité à résoudre des problèmes non linéaire qui sont difficiles à formuler

mathématiquement. Les modèles neuro-flous ont l'avantage d'intégrer les options des RNAs et celles des systèmes d'inférence flous dans un cadre unique.

Plusieurs modèles de réseaux de neurones et d'ANFIS (avec différentes combinaisons de données d'entrée) vont être implémentés et comparés avec les données observées de la qualité de l'eau choisie (indicateurs de l'eutrophisation ou variables de sortie). Les comparaisons sont basées sur le traitement statistique des résultats par le biais des critères statistiques de performance tels que le coefficient de corrélation et le root mean square error (RMSE).

III.2. Réseaux de Neurones Artificiels

III.2.1. Introduction

L'homme n'a rien inventé mais il a essayé depuis l'Antiquité de mimer la nature pour améliorer ses conditions de vie. La fabrication de l'avion en imitant les oiseaux, ensuite la modification de la séquence d'ADN des plantes après avoir découvert les processus biotechnologique chez *Agrobacterium tumefaciens*, l'invention de système d'irrigation par aspersion en simulant les pluies, dernièrement, il a essayé de reproduire la réflexion humaine en créant par analogie du système nerveux biologique ce que l'on appelle les réseaux de neurones artificiels (RNA) qui servent à résoudre diverses fonctions ou problèmes non linéaires, récemment, il a inventé les algorithmes génétiques (AG) par analogie du processus génétiques qui se déroulent chez les êtres vivants.

Les réseaux de neurones artificiels sont des modèles inspirés de la neurobiologie qui imitent le fonctionnement du cerveau. Ils sont basés sur la fonction neuronale, parce que les neurones sont identifiés comme éléments cellulaires responsables du traitement de l'information dans le cerveau. Les réseaux de neurones artificiels se sont donc basés sur l'hypothèse disant que le raisonnement intelligent des êtres humains a pour origine la structure de système nerveux et donc on peut l'inculquer à un ordinateur en lui implémentant un réseau de neurones artificiels pour le rendre « intelligent », tout en sauvegardant sa puissance et sa rapidité d'exécution.

L'approche neuronale s'est développée avec l'expansion de l'utilisation des outils informatiques qui facilitent l'analyse numérique de l'information et la mise en œuvre de cette méthode.

III.2.2. Historique des réseaux de neurones artificiels

Un bref historique des travaux menés sur le réseau de neurones est énuméré ci-dessous dans l'ordre chronologique.

McCulloch et Pitts (1943) montrent que le réseau de neurones artificiels peut, en principe, calculer n'importe quelle fonction arithmétique et logique, ces travaux sont souvent considérés comme l'origine des réseaux de neurones artificiels (**Martin et Hargan , 1996**).

En 1949, dans "*The Organization of Behavior*", un livre influent, Hebb suit les suggestions de Lashley et de Cajal, et il présente sa règle célèbre d'apprentissage : l'activation répétée d'un neurone par les autres, à travers une synapse particulière, augmente sa conductibilité.

Von Neumann (1956) montre comment présenter la redondance et la tolérance de fautes dans les réseaux de neurones, et il montre comment l'activation synchrone de beaucoup de neurones peut être employée pour représenter chaque bit de données.

Rosenblatt (1958) invente la "perceptron" présentant une méthode d'étude pour le modèle neuronal de McCulloch et de Pitts.

Widrow et Hoff (1960) présentent "l'Adaline", un réseau simple qualifié par une règle de descente de gradient pour réduire au minimum l'erreur carrée moyenne.

Rosenblatt (1961) propose la Backpropagation pour l'apprentissage des réseaux multicouches; cette tentative était non réussie parce qu'il a employé des fonctions non différentiables. (**Kishan et al., 1996**).

Hubel et Wiesel (1962) ont entrepris des études biologiques importantes sur les propriétés des neurones dans le cortex visuel des chats, stimulant le développement des modèles neuronaux artificiels à organisation automatique qui ont simulé ces propriétés (**Kishan et al., 1996**).

Minsky et Papert (1968) publient le livre *Perceptrons* dans lequel ils utilisent une solide argumentation mathématique pour démontrer les limitations des réseaux de neurones à une seule couche. Ce livre a une influence telle que la plupart des chercheurs quittent le champ de recherche sur les réseaux de neurones, et c'était un grand tournant dans la modélisation par les réseaux de neurones, entre 1969 à 1985, la recherche s'est complètement arrêté.

Rumelhart (1986) publie l'algorithme de la rétropropagation de l'erreur qui permet d'optimiser les paramètres d'un réseau de neurones à plusieurs couches, c'est elle qui a permis de donner un nouvel élan aux réseaux de neurones artificiels. À partir de ce moment, la recherche sur les réseaux de neurones connaît un essor fulgurant et les applications commerciales de ce succès académique suivent au cours des années 90.

Aujourd'hui, les réseaux de neurones artificiels possèdent un fondement théorique solide, constituent un ensemble d'opérateurs linéaires et non linéaires permettant grâce à leurs caractéristiques fonctionnelles une large utilisation dans différentes disciplines scientifiques.

Dans la gestion des ressources en eau, et vu la non linéarité des processus de la qualité des eaux et de l'eutrophisation, plusieurs auteurs ont utilisé les réseaux de neurones artificiels pour modéliser ces phénomènes (**Karul et al., 2000; Kuo et al., 2007; Csábrági et al., 2016; Mohammadpour et al., 2016; Keshtegar et al., 2017; Heddam, 2017**).

III.2.3. Quelques domaines d'application des réseaux de neurones artificiels

On utilise les réseaux de neurones artificiels dans différents domaines (informatique, électronique, science cognitive, hydrologie, neurobiologie), l'étude des réseaux de neurones est une voie prometteuse de l'Intelligence Artificielle, qui a des applications dans de nombreux domaines:

Industrie : contrôle qualité, diagnostic de panne, corrélations entre les données fournies par différents capteurs, analyse de signature ou d'écriture manuscrite.

Finance : prévision et modélisation du marché (cours de monnaies...), sélection d'investissements, attribution de crédits.

Télécommunications et informatique : analyse du signal, reconnaissance de formes (bruits, images, paroles), compression de données.

Environnement : évaluation des risques, analyse chimique, prévisions et modélisation de la qualité de l'eau, gestion des ressources en eau.

III.2.4. Réseau de neurones: fondement biologique

Les modèles neuronaux ont été développés par analogie avec les neurones biologiques, il est donc important de rappeler au préalable son fonctionnement.

III.2.4.1. Le neurone

Les cellules nerveuses, appelées neurones, constituent les éléments structuraux et fonctionnels du système nerveux central, l'Organisme humain possède environ cent milliards de neurones. Celles-ci sont des cellules somatiques et ont plusieurs points en commun dans leurs organisations avec les autres cellules, cependant elles sont extrêmement spécialisées, puisqu'elles assurent plusieurs fonctions :

Recevoir des signaux en provenance de neurones voisins ;

Intégrer ces signaux ;

Engendrer un flux nerveux ;

Conduire le flux nerveux et le transmettre à un autre neurone capable de le recevoir.

Du point de vue anatomique, le neurone est une cellule composée d'un corps cellulaire et d'un noyau, d'un axone, le corps cellulaire se ramifie pour former les dendrites.

C'est par les dendrites que les informations sont acheminées de l'extérieur vers le soma, qui présente le corps du neurone (**Romain, 2003**).

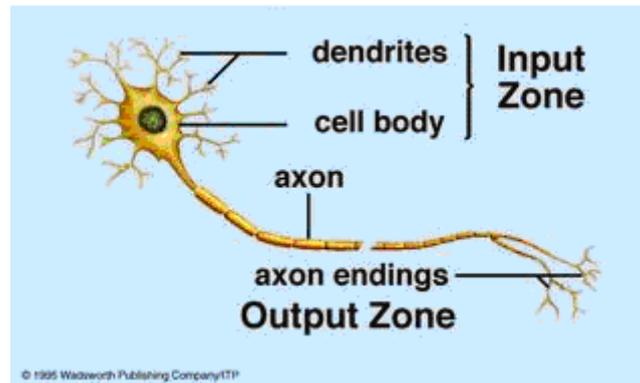


Figure 7 Neurone biologique

III.2.4.2. Le corps cellulaire

Il contient le noyau du neurone et effectue les transformations biochimiques nécessaires à la synthèse des enzymes et des autres molécules qui assurent la vie du neurone. Sa forme est pyramidale ou sphérique dans la plupart des cas, elle dépend souvent de sa position dans le cerveau.

III.2.4.3. Les dendrites

Chaque neurone possède une « chevelure » de dendrites, celles-ci sont de fines extensions tubulaires, elles se ramifient, ce qui les amènent à former une espèce d'arborescence autour du corps cellulaire, elles sont les récepteurs principaux du neurone pour capter les signaux qui leur parviennent.

III.2.4.4. L'axone

L'axone, qui est la fibre nerveuse, sert de moyen de transport pour les signaux émis par le neurone. Il se distingue des dendrites par sa forme et par les propriétés de sa membrane externe. En effet, il est généralement plus long que les dendrites, et se ramifie à l'extrémité, là où il communique avec d'autres neurones, alors que les ramifications des dendrites se produisent plutôt près du corps cellulaire.

Pour former le système nerveux, les neurones sont connectés les uns aux autres suivant des répartitions spatiales complexes, les connexions entre deux neurones se font en des endroits appelés synapses où ils sont séparés par des espaces synaptiques.

La communication entre les neurones à travers l'espace synaptique se fait par des substances chimiques appelées « neurotransmetteurs ».

III.2.5. Réseaux de neurones artificiels

III.2.5.1. Principe général

Les réseaux de neurones artificiels (RNA) sont des modèles mathématiques non linéaires, de type « boîte noire », capables de déterminer des relations entre données par la présentation (l'analyse) répétée d'exemples (à savoir des couples constitués par des informations d'entrée et une valeur de sortie que l'on voudrait approcher par le modèle).

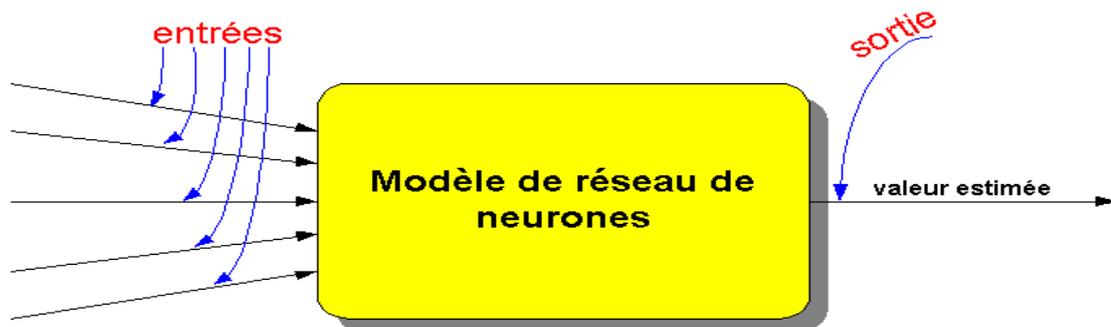


Figure 8 Modèle neuronal

La modélisation à l'aide de réseaux de neurones artificiels suppose l'adaptation des paramètres du réseau, afin de mettre en évidence les relations qui portent sur les exemples présentés. Les réseaux de neurones artificiels sont constitués d'un ensemble d'éléments de calcul (neurones artificiels), organisés dans une structure spécifique, les paramètres du réseau (les poids) étant représentés par les valeurs associées aux connexions de ces éléments de calcul, un élément de calcul du réseau de neurone artificiel (RNA) comporte plusieurs entrées et une sortie.

La valeur de sortie est obtenue par l'application d'une relation mathématique (fonction d'activation) sur la somme pondérée d'entrées.

Dans la modélisation à l'aide de réseaux de neurones artificiels, on peut choisir le type de fonctions d'activation, le nombre de neurones et l'arrangement de leurs connexions (à savoir la structure du réseau). Généralement, on utilise des fonctions d'activation de type sigmoïde

III.2.5.2. Structure d'un neurone artificiel (formel)

Un neurone formel, ou neurone, au sens informatique du terme, est une fonction algébrique non linéaire et bornée, dont la valeur dépend de paramètres appelés coefficients ou poids. Les variables de cette fonction sont habituellement appelées « entrées » du neurone, et la valeur de la fonction est appelée sortie. Un neurone est donc avant tout un opérateur mathématique.

La figure III.3 montre la structure d'un neurone artificiel, ce dernier est un processeur élémentaire, il reçoit un nombre variable d'entrées en provenance de neurones précédents. A chacune de ces entrées est associé un poids w représentant de la force de la connexion. Chaque processeur élémentaire est doté d'une sortie unique, qui se ramifie ensuite pour alimenter un nombre variable de neurones avants, à chaque connexion est associée un poids (Heddam, 2006).

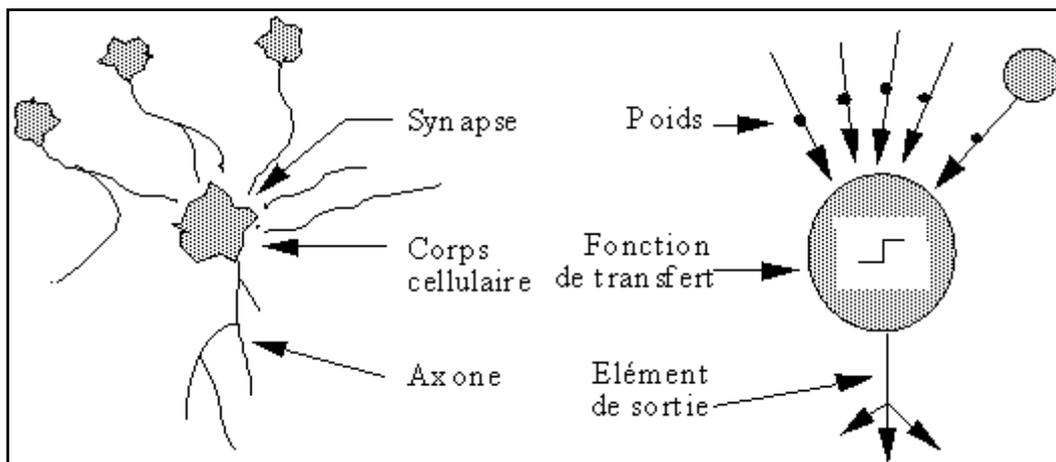


Figure 9 Mise en correspondance neurone biologique / neurone artificiel

Le tableau suivant représente une comparaison entre les deux réseaux de neurones.

Tableau 2 Analogie entre les neurones biologiques et artificiels

Neurone biologique	Neurone artificiel	Représentation symbolique
Soma	Neurone	
Dendrite	Intrant ou entrée (<i>Input</i>)	x_i, x (intrant net)
Axone	Extrant ou sortie (<i>Output</i>)	Y
Synapse	Poids (weight)	w_i

Chaque neurone artificiel reçoit un nombre variable d'entrées (x_i) en provenance de neurone amont. A chacune de ces entrées est associé un poids (w_i) représentatif de la force de la connexion.

Un poids d'un neurone artificiel représente donc l'efficacité d'une connexion synaptique, un poids négatif vient inhiber une entrée, alors qu'un poids positif vient l'accentuer (**Benkaci, 2006**).

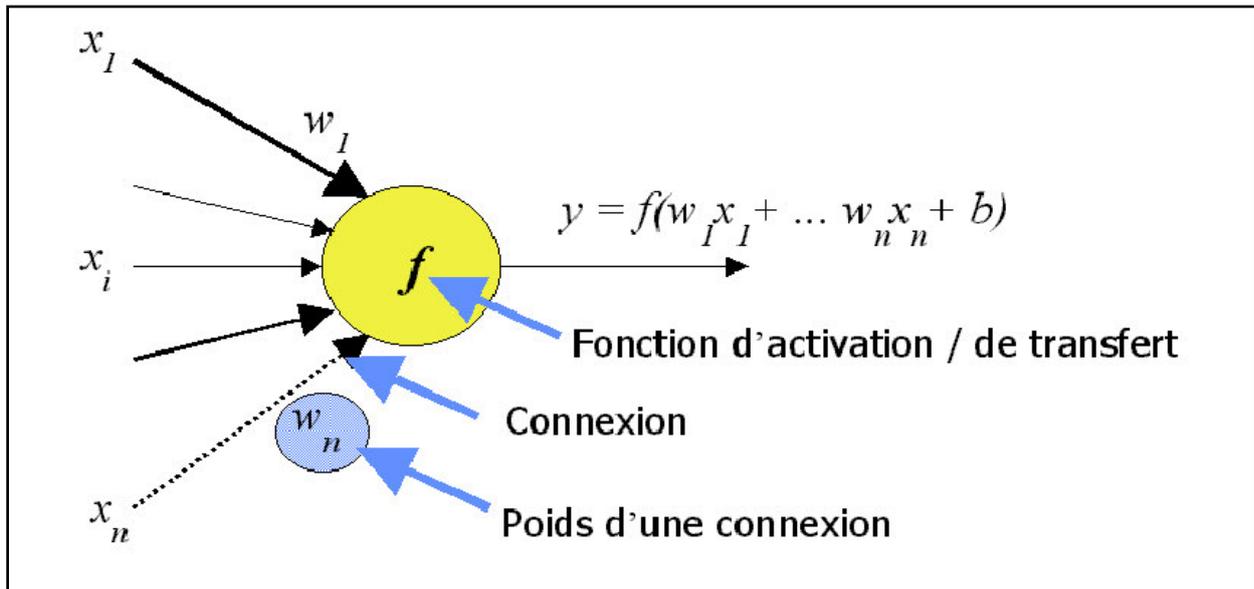


Figure 10 Neurone formel

III.2.5.3. Fonctionnement mathématique du neurone artificiel

Le fonctionnement mathématique du neurone artificiel est contrôlé par deux fonctions à savoir :

La fonction de base (fonction discriminant) ;

Et la fonction dynamique.

III.2.5.3.1. Fonction de base (fonction discriminant)

Elle définit l'activité du neurone, la plus utilisée est la fonction linéaire (la somme pondérée), elle est définie par :

$$\phi(w_i, x_i) = w_1 * x_1 + \dots + w_n * x_n - b = \sum_{i=1}^{i=n} (x_i * w_i) - b \quad (07)$$

Avec :

x_1, x_2, \dots, x_n : Entrées, ou encore les excitations ;

w_1, w_2, \dots, w_n : Coefficients synaptiques ou poids synaptiques ;

b : Représente le biais (constant).

V.2.5.3.2. Fonction dynamique

C'est la fonction dite d'activation (transfert) qui a pour objectif de rendre l'activité du neurone bornée. Cette fonction doit être nonlinéaire, continue dérivable et monotone. La fonction d'activation calcule la valeur de l'état du neurone. C'est cette valeur qui sera transmise aux neurones avals (**Benkaci, 2006**). Le choix d'une fonction de transfert se révèle être un élément constitutif important des réseaux de neurones.

Il existe de nombreuses formes possibles pour la fonction de transfert, les plus courantes sont les fonctions linéaires, binaires, sigmoïdes, et radiales (gaussiennes).

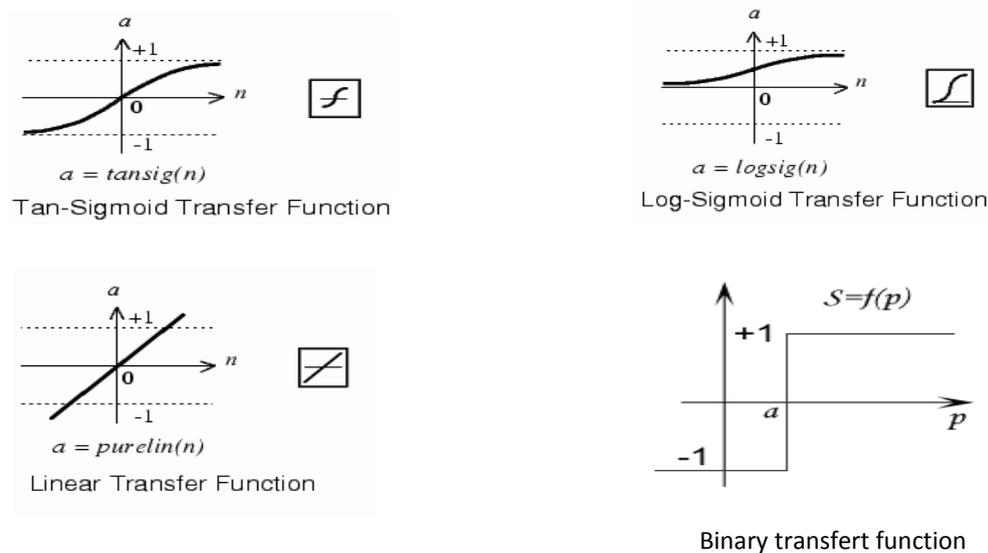


Figure 11 Différents types de fonctions de transfert

On remarquera qu'à la différence des neurones biologiques dont l'état est binaire, la plupart des fonctions de transfert sont continues, offrant une infinité de valeurs comprises dans l'intervalle $[-1, +1]$; $[0, +1]$ ou même $[-\infty, +\infty]$.

III.2.5.4. Classification des réseaux de neurones artificiels

Il existe de nombreux types de réseaux de neurones, on peut les classer selon leur architecture et leur mode d'apprentissage.

III.2.5.4.1. Architecture générale des réseaux de neurones

Un réseau de neurones est un maillage de plusieurs neurones, généralement organisé en couches, effectuant un calcul selon une fonction d'activation (neurones) et des connections reliant les neurones entre eux (poids).

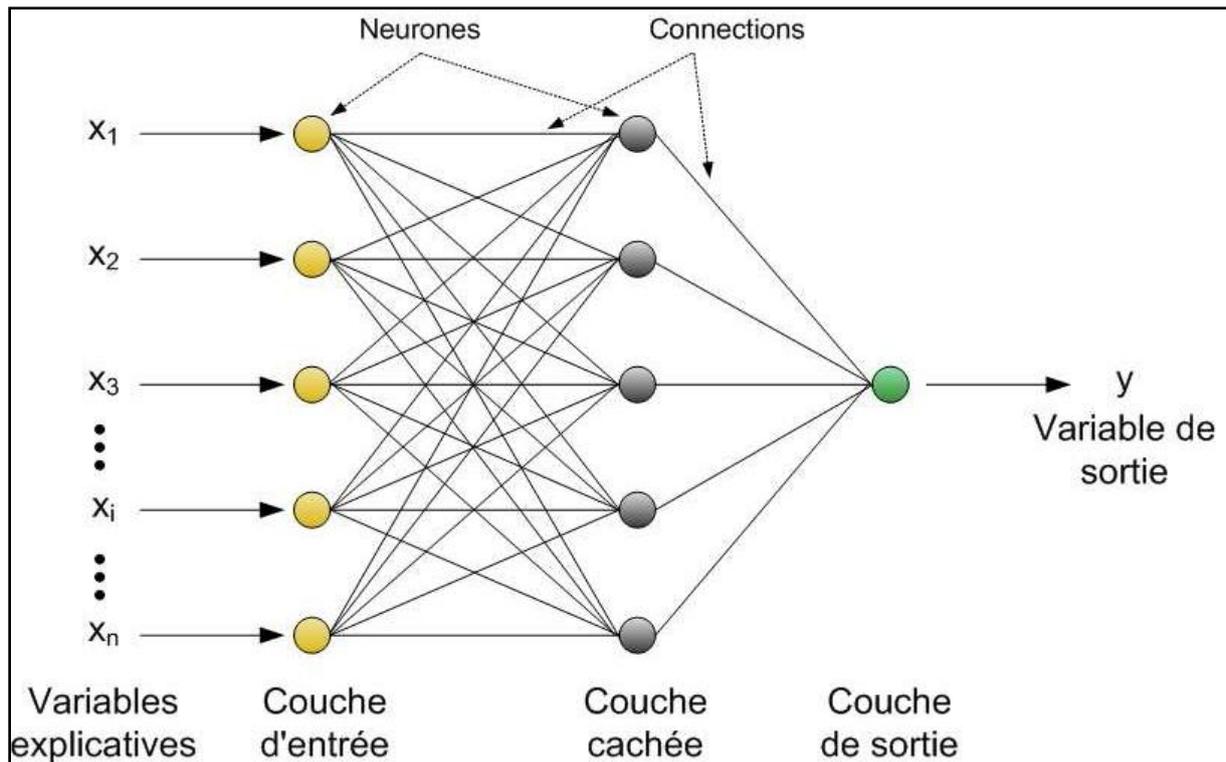


Figure 12 Réseau de neurones artificiel

Selon l'architecture, les réseaux de neurones peuvent être classés en deux types, Les réseaux de neurones bouclés et les réseaux de neurones non bouclés (en couches).

III.2.5.4.1.1. Réseaux de neurones dynamiques

Ils sont appelés aussi, réseaux de neurones bouclés ou récurrents, contrairement aux réseaux de neurones non bouclés, les réseaux de neurones bouclés peuvent avoir une topologie de connexions quelconque, comprenant notamment des boucles qui ramènent aux entrées la valeur d'une ou plusieurs sorties.

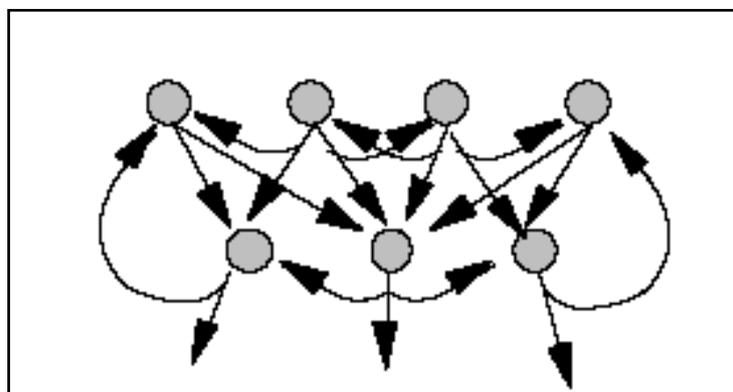


Figure 13 Réseau bouclé

III.2.5.4.1.2. Réseaux de neurones non bouclés ou en couches (Perceptron)

Appelés aussi "réseaux "Feed-Forward", ce sont des réseaux dans lesquels l'information se propage de couche en couche sans retour en arrière possible (feed back). *Un réseau de neurones non bouclé (en couches) réalise une (ou plusieurs) fonctions algébriques de ses entrées, par composition des fonctions réalisées par chacun de ses neurones.*

III.2.5.4.1.2.1. Le Perceptron monocouche

C'est historiquement le premier RNA, c'est le Perceptron de Rosenblatt. C'est un réseau simple, puisque il ne se compose que d'une couche d'entrée et d'une couche de sortie. Il est calqué, à la base, sur le système visuel et de ce fait a été conçu dans un but premier de reconnaissance des formes. Cependant, il peut aussi être utilisé pour faire de la classification et pour résoudre des opérations logiques simples (telle "ET" ou "OU"). Sa principale limite est qu'il ne peut résoudre que des problèmes linéairement séparables. Il suit généralement un apprentissage supervisé selon la règle de correction de l'erreur (ou selon la règle de Hebb).

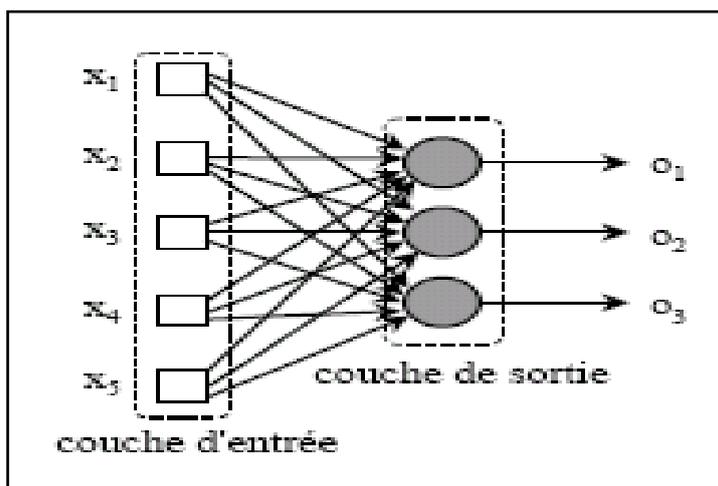


Figure 14 Réseaux à une seule couche

III.2.5.4.1.2.2. Le Perceptron multicouches

Le Perceptron multicouches est sans doute le plus simple et le plus connu des réseaux des neurones, la structure est relativement simple : une couche d'entrée, une couche de sortie et une ou plusieurs couches cachées. Chaque neurone n'est relié qu'aux neurones de couche précédente.

Réseau multicouche : les neurones sont arrangés par couche. Il n'y a pas de connexion entre neurones d'une même couche et les connexions ne se font qu'avec les neurones de couches avales. Habituellement, chaque neurone d'une couche est connecté à tous les neurones de la couche suivante et celle-ci seulement. Ceci nous permet d'introduire la notion de sens de

parcours de l'information (de l'activation) au sein d'un réseau et donc définir les concepts de neurone d'entrée, neurone de sortie. Par extension, on appelle couche d'entrée l'ensemble des neurones d'entrée, couche de sortie l'ensemble des neurones de sortie, les couches intermédiaires n'ayant aucun contact avec l'extérieur sont appelées couches cachées.

Chaque neurone dans une couche est connecté à tous les neurones de la couche précédente et de la couche suivante (excepté pour les couches d'entrée et de sortie) et il n'y a pas de connexions entre les cellules d'une même couche. Les fonctions d'activation utilisées dans ce type de réseaux sont principalement les fonctions à seuil ou sigmoïdes, il suit aussi un apprentissage supervisé selon la règle de correction de l'erreur.

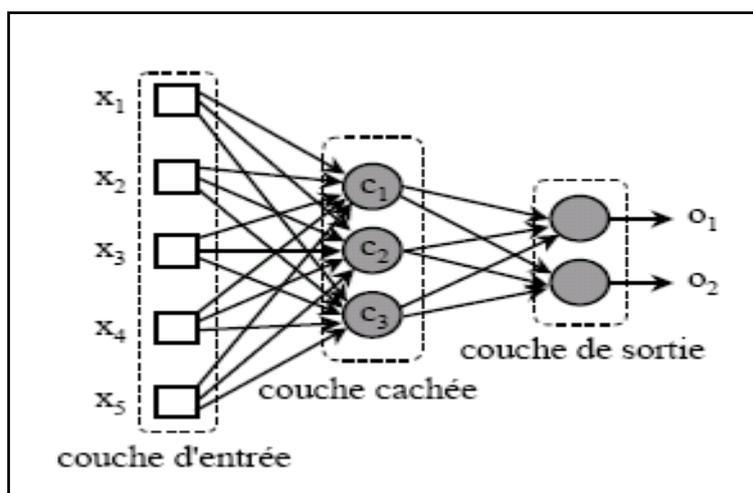


Figure 15 Réseaux "FeedForward" multicouches

III.2.5.4.1.2.3. Perceptron Multi-Couches (PMC) et approximation de fonctions

Les réseaux de neurones artificiels sont un outil extrêmement précieux pour résoudre les problèmes non linéaires, surtout si on dispose d'un large ensemble d'apprentissage.

Lorsque l'on veut modéliser un processus à partir des données, on cherche toujours à obtenir les résultats les plus satisfaisants possibles avec un nombre minimum de paramètres ajustables.

Dans cette optique, Hornik [Hornik, 1994, in Rajai., 2002] a montré que :

«Si le résultat de l'approximation (c'est à dire la sortie du réseau de neurones) est une fonction non linéaire des paramètres ajustables, elle est plus parcimonieuse que si elle était une fonction linéaire de ces paramètres. De plus, pour des réseaux de neurones à fonction d'activation sigmoïdale, l'erreur commise dans l'approximation varie comme l'inverse du nombre de neurones cachés, et elle est indépendante du nombre de variables de la fonction à

approcher. Par conséquent, pour une précision donnée, le nombre de paramètres du réseau est proportionnel au nombre de variables de la fonction à approcher ».

Ces résultats s'appliquent aux Perceptrons Multi-Couches à fonction d'activation sigmoïdale.

Ces propriétés (approximation et parcimonie) sont démontrées de manière générale à condition d'avoir une architecture adéquate. Dans le cas où l'architecture ne serait pas adaptée, celles-ci ne peuvent pas s'appliquer, elles constituent néanmoins une justification fondamentale de l'utilisation des Perceptrons Multi-Couches.

III.2.5.4.2. Apprentissage des réseaux de neurones

Le calage (apprentissage) : au sens strict du terme, c'est l'opération qui consiste à trouver les valeurs des paramètres du modèle qui minimisent l'erreur de modélisation.

L'apprentissage du modèle neuronal est une phase du développement d'un réseau de neurones durant laquelle le comportement du réseau est modifié en réaction avec les stimuli qu'il reçoit de son environnement jusqu'à l'obtention du comportement désiré (**Heddam, 2006**). Généralement, les paramètres modifiés sont les poids synaptiques qui relient les neurones entre eux.

L'objectif de l'apprentissage est de réduire l'erreur globale jusqu'à l'obtention d'une erreur acceptable.

Il existe plusieurs règles d'apprentissage qui peuvent être regroupées en trois grandes catégories, l'apprentissage non supervisé et l'apprentissage semi-supervisé ou par renforcement et l'apprentissage supervisé, (**Dreyfus et al., 2004**).

III.2.5.4.2.1. Apprentissage non supervisé

Il s'agit d'apporter au réseau de neurone les entrées sans faire appel à la réponse désirée.

Dans ce cas, les exemples présentés à l'entrée provoquent une autoadaptation du réseau afin de produire des valeurs de sortie. Ce type d'apprentissage est moins utilisé par rapport aux autres modes d'apprentissage.

«Un réseau de neurones non bouclé peut être utilisé dans un but de visualisation ou d'analyse de données : on dispose d'un ensemble de données, représentées par des vecteurs de grande dimension, et l'on cherche à les regrouper, selon des critères de ressemblance qui sont inconnus a priori. Ce type de tâches est connu en statistique sous le nom de méthodes d'agrégation » (en anglais clustering). (Dreyfus et al., 2004).

III.2.5.4.2.2. L'apprentissage semi-supervisé ou par renforcement

L'apprentissage semi-supervisé ou par renforcement (apprentissage par Récompense et Pénalité) : représente une technique nouvelle, dans ce cas on fournit des exemples au réseau et sur le résultat des indications par un superviseur qui est régi par un de degrés de performance (vrai ou faux). (Benkaci, 2005).

III.2.5.4.2.3. Apprentissage supervisé des réseaux de neurones artificiels

Dans ce cas on apporte au modèle neuronal les entrées mais aussi la ou les sorties désirées c'est-à-dire qu'on présente au réseau un ensemble de couples (entrées-sorties désirées), L'apprentissage neuronal fait appel donc à des exemples de comportement. La réponse de neurone de sortie est alors comparée aux valeurs attendues. Le but de L'apprentissage supervisé est d'inculquer un comportement de référence au réseau.

III.2.5.4.2.3.1. Algorithmes d'apprentissage supervisé

Les algorithmes d'apprentissage itératifs utilisés dans les réseaux de neurones explorent l'espace des paramètres (i.e. poids et biais du réseau) à partir d'un point de départ fixé par les valeurs initiales de ces paramètres et selon une trajectoire qui sera fonction de l'algorithme utilisé et des données.

L'apprentissage "supervisé", pour les réseaux de neurones formels, consiste à modifier les coefficients (poids synaptique W) de telle manière que les sorties du réseau de neurones soient, pour les exemples utilisés lors de l'apprentissage, aussi proches que possible des sorties "désirées", qui peuvent être la valeur de la fonction que l'on veut approcher ou de la sortie du processus que l'on veut modéliser, ou encore la sortie souhaitée du processus à commander. La plupart des algorithmes d'apprentissage des réseaux de neurones formels sont des algorithmes d'optimisation, s cherchent à minimiser, par des méthodes d'optimisation non linéaire, une fonction de coût qui constitue une mesure de l'écart entre les réponses réelles du réseau et ses réponses désirées.

Cette optimisation se fait de manière itérative, en modifiant les poids en fonction du gradient de la fonction de coût: le gradient est estimé par une méthode spécifique aux réseaux de neurones, dite dans notre cas, la méthode de rétropropagation, puis il est utilisé par l'algorithme d'optimisation proprement dit. Les poids sont initialisés aléatoirement avant l'apprentissage, puis modifiés itérativement jusqu'à obtention d'un compromis satisfaisant entre la précision de l'approximation sur l'ensemble d'apprentissage et la précision de l'approximation sur un ensemble de test disjoint du précédent. Contrairement à des affirmations maintes fois répétées, l'apprentissage des réseaux de neurones n'est pas

spécialement lent: il existe des algorithmes d'optimisation non linéaires extrêmement rapides qui permettent de faire des développements industriels sur de simples PC. L'apprentissage des réseaux de neurones bouclés (pour réaliser des modèles dynamiques) est très semblable à celui des réseaux non bouclés.

III. 2.5.4.2.3.1.1. La Backpropagation

La rétro propagation du gradient était à la base de la réussite réalisée par les réseaux de neurones artificiels, c'est la méthode la plus utilisée aujourd'hui dans les algorithmes d'apprentissage des réseaux multicouches. Il s'agit d'une méthode d'apprentissage supervisée fondée sur la modification des poids synaptiques dans le sens inverse de propagation de l'information dans le réseau et en fonction de l'erreur produite, d'où la nomenclature de rétro propagation ou Backpropagation.

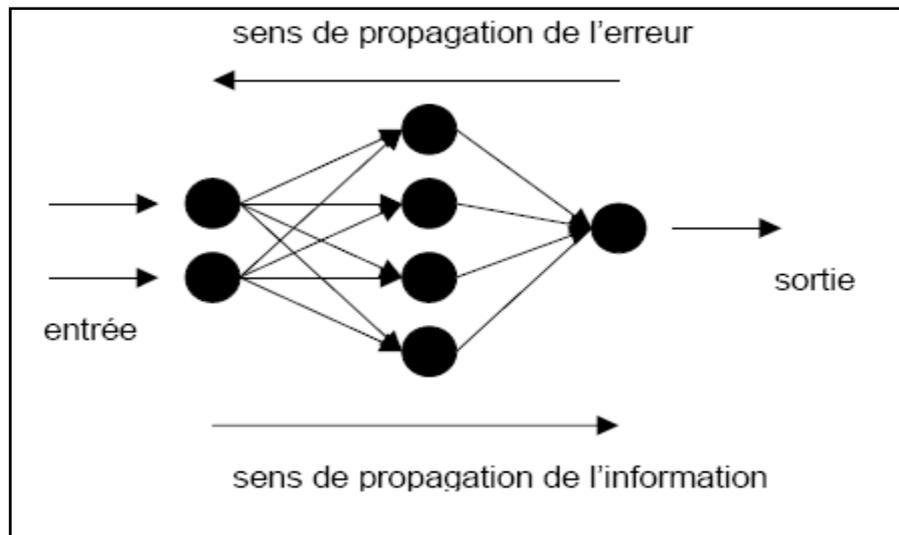


Figure 16 Rétro- propagation des gradients

La rétro- propagation des gradients se fait “en amont” dès la couche de sortie vers la couche d’entrée.

III. 2.5.4.2.3.1.1.a. Algorithme de Backpropagation

L’algorithme de rétropropagation du gradient est un algorithme itératif, les poids sont modifiés à chaque étape selon les instructions suivantes (**Benkaci, 2006**) :

- 1) Initialiser les poids synaptiques à de petites valeurs aléatoires comprises en général entre -1 et 1 .
- 2) Présenter un nouvel exemple X_p à l’entrée du réseau.

Calculer la sortie du réseau et l’erreur en sortie:

$$Ep = \sum_{i=1}^{nL} (d_i^p - y_i^L)^2 \quad (08)$$

Où i parcourt les indices des neurones de sortie, d_i et y_i représentent respectivement l'activation mesurée et l'activation désirée pour les neurones ;

3) les poids du réseau sont modifiés suivant la règle :

$$\left[\Delta w_{ij}^l \right] = \eta f'(u_i^l(t)) \delta_i^l y_j^{l-1} \quad (09)$$

Où η est une constante positive appelée pas du gradient (Taux d'apprentissage), et contrôle la vitesse de l'apprentissage.

$$\text{Avec : } \delta_i^l = \begin{cases} (d_p^i - y_i^L) & \text{Pour les neurones de la couche de sortie.} \\ \sum_{k=1}^{k=n^{l+1}} \delta_k^{l+1} w_{ki}^{l+1} & \text{Pour les neurones des autres couches.} \end{cases} \quad (10)$$

S'il s'agit d'un Data Learning (FFN pattern), réadapter les poids :

$$4) \text{ Poser : } \Delta w_{ij}^l(t) = \left[\Delta w_{ij}^l(t) \right]_p \quad (11)$$

S'il s'agit d'un Block Learning (FFN Batch):

Aller à 5.

Si $p \neq M$ (M étant le nombre d'exemple d'entraînement), retourner à 1.

$$\text{Sinon : } \Delta w_{ij}^l(t) = \sum_{p=1}^{p=M} \left[\Delta w_{ij}^l(t) \right]_p \quad (12)$$

5) Réadapter les poids.

$$\Delta w_{ij}^l(t+1) = w_{ij}^l(t) + \Delta w_{ij}^l(t) \quad (13)$$

6) Répéter de 1 à 5 autant de fois que nécessaire jusqu'à convergence vers le seuil d'erreur fixé.

L'apprentissage avec cet algorithme (réadaptation des poids) se fait par deux manières différentes, à savoir Block Adaptive Learning, et l'apprentissage par données Data Adaptive Learning.

Block Adaptive Learning (Batch learning):

En Batch Training les poids et les biais de réseau de neurones ne sont ajustés qu'après le passage total sur la base d'entraînement. Les gradients calculés à chaque passage sur la base d'entraînement sont ajoutés ensemble pour déterminer le changement des poids et des biais.

La fonction de coût qui calcule l'erreur est :

$$Ep = \sum_{i=1}^{nL} (d_i^p)^2 - \sum_{i=1}^{nL} (y_i^p)^2 \quad (14)$$

Cette méthode est jugée plus robuste, cependant les données simulées correspondent à une moyenne de données observées (**Benkaci, 2006**).

Data Adaptive Learning (Incremental Learning):

Les poids sont ajustés itérativement après chaque exemple. Lors du passage sur la base d'entraînement et le calcul des erreurs, ces dernières sont présentées immédiatement pour adapter les poids, sans accumuler tous les changements jusqu'à la fin de l'itération (**David et al., 2006**). Pour cette méthode d'apprentissage, les poids sont réadaptés au passage de chaque exemple :

$$Ep = \sum_{i=1}^{nL} (d_i^p - y_i^L)^2 \quad (15)$$

III.2.5.4.2.3.1.1.b. Méthode d'optimisation avec la Backpropagation

Il existe de nombreuses méthodes (variantes) d'optimisation avec la Backpropagation, on cite à titre indicatif celle de Levenberg-Marquardt (la commande *trainlm* in Matlab).

L'algorithme de Levenberg-Marquardt

C'est une variante de l'algorithme d'apprentissage de rétropropagation du gradient (Backpropagation) visant à améliorer les performances de cet algorithme. Il est utilisé par défaut sous MATLAB. C'est la méthode la plus rapide dans la « boîte à outils » de réseau de neurones (ANN toolbox), mais elle exige de large espace de mémoire.

Une étude menée par l'auteur de guide d'utilisateur de la boîte à outil de réseaux de neurones sous Matlab (*Neural Network Toolbox*), a montré que dans les problèmes d'approximation de fonction et pour les réseaux qui contiennent jusqu'à quelques centaines poids synaptiques, l'algorithme de Levenberg-Marquardt aura la convergence la plus rapide. L'avantage est particulièrement apparent si l'apprentissage très précis est exigé. Dans beaucoup de cas, l'algorithme de Levenberg-Marquardt permet d'obtenir des erreurs quadratiques moyennes inférieures à celles trouvées par d'autres algorithmes utilisés. Cependant, à mesure que le nombre de poids dans le réseau augmente, l'avantage de l'algorithme de Levenberg-Marquardt diminue.

Dans ce travail, vu ses propriétés, on va utiliser cette variante de Backpropagation.

III. 2.5.5. Avantages et inconvénients des réseaux de neurones artificiels

III. 2.5.5.1. Avantage des réseaux de neurones artificiels

Récemment, les réseaux de neurones artificiels sont appliqués en modélisation météorologique, agro écologique et hydrologique ; la plupart de publications éditées dans la littérature concernent l'estimation, la prédiction et la classification. Les applications de RNA sont rapidement diffusées vu leurs caractéristiques fonctionnelles, qui fournissent de nombreux avantages par rapport aux approches analytiques traditionnelles. (**Demetris, 2007**).

Les réseaux de neurones artificiels deviennent populaires ces dernières années, cette popularité peut être attribuée à leur apparente facilité d'utilisation. Comme pour d'autres modèles dits « boîtes noires » et contrairement aux modèles « conceptuels » la construction d'un modèle pour une étude de cas particulier ne demande pas d'analyse détaillée et de compréhension des relations entre variables de forçage et variables dépendantes. Ce sont les algorithmes d'apprentissage (de calage) des réseaux qui analyseront automatiquement ces relations. D'autre part, comme les fonctions polynomiales, et contrairement au modèle linéaire, les réseaux de neurones sont des interpolateurs universels (universal approximators) ce qui veut dire qu'ils sont capables d'approcher n'importe quelle fonction continue et ceci quel que soit le niveau d'approximation souhaité (**Hornick et al., 1989 in Eric, 2000**).

Une propriété fondamentale de l'apprentissage réalisé concerne les capacités de généralisation de ces réseaux. Dans le cas où l'architecture initiale est correctement choisie, on constate généralement que les exemples ne sont pas appris, mais que le réseau est capable d'étendre les connaissances acquises à des exemples proches ou intermédiaires (généralisation=interpolation + extrapolation) (**Benkaci, 2006**).

Les réseaux de neurones artificiels offrent des avantages significatifs par rapport à la régression linéaire multiple et d'autres méthodes conventionnelles dans la prévision de la qualité de l'eau et l'eutrophisation parce qu'il n'est pas nécessaire d'indiquer la forme du modèle mathématique avant d'adapter les données. C'est important puisque beaucoup de processus relatifs à la qualité de l'eau ont des facteurs multiples qui changent avec le temps. En conséquence il est souvent difficile de trouver un modèle mathématique approprié pour décrire ces processus. Cependant, ces réseaux neuronaux devront être examinés contre d'autres modèles et avec des données variées provenant de diverses régions géographiques, nous nous sentons que ce réseau neuronal est une technique prometteuse à aider dans la gestion de la qualité de l'eau.

L'avantage pratique des réseaux de neurones par rapport aux techniques classiques de modélisation non linéaire réside donc dans leur capacité à réaliser des modèles de précision équivalente avec moins de données expérimentales (ou dans leur capacité à constituer des modèles plus précis à partir du même nombre d'exemples). De manière générale, les réseaux de neurones permettent de tirer le meilleur parti des données numériques disponibles, pour construire des modèles à plusieurs entrées ; c'est la propriété parcimonieuse de RNA (Dreyfus, 2004).

III. 2.5.5.2. Inconvénients des réseaux de neurones artificiels

Malgré les avantages marqués, les RNAs comportent certains inconvénients.

La performance de modèle neuronal est souvent tributaire à son architecture. Or, le choix de l'architecture est un problème difficile à résoudre, l'inconvénient majeur de réseaux de neurones artificiels est l'absence d'une méthodologie permettant de déterminer le nombre de couches cachées, le nombre de neurones par couche cachée et le nombre d'itérations, seule les essais permettent de répondre à ces questions, et on ne sait pas si l'optimum sera atteint. Un autre inconvénient c'est le choix de la méthode d'apprentissage, celle-ci doit également être choisie de façon judicieuse.

Les réseaux de neurones artificiels comme les autres modèles « boîte noire » ont un inconvénient majeur par rapport aux modèles conceptuels simples : on ne maîtrise pas leur comportement en extrapolation pour des valeurs de variables de forçage sortant considérablement de la gamme utilisée pour le calage du modèle.

Si la fonction d'activation des neurones de la couche cachée est la fonction sigmoïde et le neurone de la couche de sortie est linéaire, alors la donnée de sortie du réseau est nécessairement bornée. Ceci peut poser un problème lorsque le réseau simule des données non bornées. Il peut ne pas être en mesure de simuler des valeurs nettement supérieures aux valeurs présentes dans l'échantillon de données utilisé pour le calage du réseau de neurones. (Eric, 2000).

On ne maîtrise pas l'évolution des variables de sortie du réseau de neurones lorsque les variables de forçage sortent sensiblement de la gamme des valeurs pour lesquelles le modèle a été calé. (Eric, 2000).

Les réseaux de neurones artificiels (RNA) présentent l'avantage de ne requérir qu'une structure de modélisation formelle pour être ajustés aux données expérimentales ; cependant, leur principal inconvénient est la perte de transparence du modèle (de type « boîte noire »), et

le manque d'indicateurs pour l'évaluation de la précision et de la fiabilité de la réponse des RNA quand on est en présence de comportements nouveaux (**Dreyfus, 2004**).

III.2.5.6. Choix de paramètres du modèle neuronal

III.2.5.6.1. Choix d'un réseau de neurones

Il existe un grand nombre de types de réseaux de neurones qui sont utilisés dans de nombreux domaines, mais chaque type a son domaine propice d'utilisation, cependant, l'utilisation de plusieurs types et la comparaison entre eux constituent le seul moyen pour déterminer le meilleur type en fonction de champs de recherche.

D'après des recherches récentes (**Karul et al., 1999 ; Kuo et al., 2007 ; Huo et al., 2013**), les PMC apportent de bons résultats pour la modélisation de l'eutrophisation à partir des variables de la qualité de l'eau.

III.2.5.6.2. Choix de nombre de couches cachées

Les résultats obtenus par plusieurs chercheurs (**Kuo et al., 2007 ; Huo et al., 2013**), en utilisant le PMC avec l'algorithme de Backpropagation, et les variables de la qualité de l'eau comme entrées de réseaux pour estimer les indicateurs de l'eutrophisation (Chl-a.), confirment que les modèles neuronaux à une seule couche cachée sont suffisants pour rendre compte à la relation non linéaire existante entre les données de la qualité de l'eau et l'eutrophisation.

III.2.5.6.3. Choix de nombre de neurones par couche cachée

Le nombre d'unités cachées joue un rôle crucial dans le contrôle de la capacité du réseau de neurones.

L'erreur commise par un réseau de neurones est inversement proportionnelle au nombre de neurones cachés (**Dreyfus, 2004**) ; malheureusement, ce résultat, comme d'autres considérations théoriques n'est pas constructif, en ce sens qu'il ne peut que donner des estimations, ou des bornes larges, du nombre de neurones cachés nécessaires. Il n'existe pas, à l'heure actuelle, de résultat théorique permettant de prévoir le nombre de neurones cachés nécessaires pour obtenir une performance spécifiée du modèle, compte tenu des données disponibles. Il faut donc nécessairement mettre en œuvre une procédure numérique de conception de modèle (**Dreyfus, 2004**).

III.2.5.6.4. Paramètres d'apprentissage

Les paramètres d'apprentissage sont :

- La fonction coût est l'erreur quadratique, inférieure ou égale à 0.01
- Taux d'apprentissage : L'exécution de l'algorithme de Backpropagation est très sensible au taux d'apprentissage. S'il est très grand, l'algorithme peut osciller et devient instable, s'il est trop petit, l'algorithme mettra longtemps avant de converger.
- Le nombre d'itération est un paramètre très important, il influence considérablement les résultats obtenus. Le nombre d'itération optimum est celui qui permettra d'atteindre un minimum sur l'erreur commise. Il n'y a pas de règle utilisée pour déterminer ce nombre, les essais et la comparaison (trial-and-error, step-by-step, one-by-one) constituent une méthode fiable afin de le déterminer.

III.2.5.7. Conclusion

Dans ce chapitre nous avons mis en évidence la démarche de recherche dans le domaine des réseaux des neurones artificiels, ainsi que leur mode de fonctionnement.

L'utilisation de réseaux de neurones artificiels est avantageuse chaque fois que l'on cherche à établir une relation non linéaire entre les valeurs numériques à condition que l'on dispose de données suffisamment nombreuses et représentatives.

Les réseaux de neurones artificiels ouvrent une nouvelle ère dans la modélisation des phénomènes non linéaires tel que l'eutrophisation, il suffit d'entraîner le RNA en se basant sur les entrées et les targets (sorties désirées) et par la suite on n'apporte au RNA que les entrées, il va prédire les sorties et tous ça se fait sans connaître le formalisme intérieur de RNA d'où le nom de modèle « boîte noire » ou encore « black box », ils sont également appelés « data-driven-models ».

III.3. Introduction a la logique floue

III.3.1. Introduction

La logique floue est une méthode mathématique pour répondre à des questions avec des informations imprécises. Elle traite le raisonnement qui est approximatif plutôt que fixe et précis. Contrairement à la théorie traditionnelle de logique où il y a seulement vrai ou faux, la logique floue traite l'intermédiaire entre le vrai et le faux. L'approche de base est d'attribuer une valeur entre zéro et un pour décrire le domaine (area) entre la limite supérieure et celle inférieure. Principalement, la logique floue (FL) est une logique multi-valeur qui définit des valeurs intermédiaires entre les évaluations traditionnelles comme vrai / faux, oui / non, haut /

bas, etc. Ces valeurs intermédiaires peuvent être formulées mathématiquement et traitées par les ordinateurs, afin d'appliquer une façon de penser plus humaine à la machine (**Sumathi & Surekha, 2010**).

La logique floue imite la pensée intuitive humaine en utilisant une série d'ensembles flous, et les règles floues de type (Si...Alors....) pour modéliser les processus. La théorie des ensembles flous de Zadeh (**Zadeh, 1965**) est à la base de la logique floue (**Kovacic et Bogdan, 2006**), et la première application de la théorie de Zadeh a été développée par Mamdani en 1974, quand il a conçu un système de contrôle flou expérimental pour une chaudière à vapeur en synthétisant un ensemble de règles de contrôle linguistiques (règles floues) obtenues à partir d'opérateurs humains expérimentés (**Nguyen et al., 2003**).

Les systèmes d'inférence floue sont un cadre populaire, largement utilisé dans l'ingénierie hydraulique et environnementale, basé sur des concepts tels que la théorie des ensembles flous, le raisonnement flou et les règles floues. En général, cela aide les ordinateurs à prendre des décisions plus proches ou plus similaires à des êtres humains. L'un des principaux avantages de la logique floue est qu'elle peut résoudre plusieurs ambiguïtés et incertitudes dans la prise de décision et qu'elle peut être appliquée dans tous les domaines où les relations empiriques ne sont pas bien définies ou impossibles à modéliser.

Contrairement aux modèles mathématiques qui nécessitent une connaissance précise de toutes les variables contributives, la logique floue offre une approche plus souple, moins dépendante des hypothèses et auto-adaptative pour modéliser des processus de qualité de l'eau qui est intrinsèquement complexe, non linéaire et dynamique. Les modèles basés sur la logique floue peuvent être utilisés pour modéliser le comportement du processus même avec des informations incomplètes.

La technique de logiques floues peut être utilisée lorsque les approches conventionnelles deviennent plutôt complexes et coûteuses, ou que des informations vagues et imprécises sont directement intégrées au processus de modélisation. Avec la logique floue, il est possible de décrire les connaissances disponibles directement en termes linguistiques et selon les règles, ce qui permet d'introduire une interprétabilité au système modélisé.

Les caractéristiques quantitatives et qualitatives peuvent être combinées directement dans un modèle flou. Cela conduit à un processus de modélisation qui est souvent plus simple, plus facilement gérable et plus proche de la pensée humaine par rapport aux approches conventionnelles (**Wheater et al., 2007**).

La logique floue est un paradigme efficace pour traiter l'imprécision, elle est utilisée pour prendre des observations floues ou imprécises pour les entrées et arriver à des valeurs précises et réelles pour les sorties.

III.3.2. Historique

Lotfi A. Zadeh, professeur en informatique à l'Université de Californie à Berkeley était le pionnier qui a inventé la logique floue. Ce concept est apparu pour la première fois sur la revue de « Information and Control journal » en 1965 avec le titre provocateur "Fuzzy sets" ou "ensembles flous" en français.

L'évolution de la logique floue peut être discernée en trois phases. La première phase, de 1965 à 1973, concernait principalement la fuzzification, c'est-à-dire la généralisation du concept d'ensemble flou, avec une fonction caractéristique à deux valeurs généralisée à une fonction d'appartenance prenant des valeurs dans l'intervalle unitaire. Les questions de base et les applications abordées étaient, pour la plupart, de nature théorique, et la logique et le raisonnement n'étaient pas au centre d'intérêt des scientifiques.

La deuxième phase, de 1973 à 1999, a commencé avec deux concepts clés: (a) le concept de variable linguistique; et (b) le concept de règle floue (Si...alors...). Aujourd'hui, presque toutes les applications de la théorie des ensembles flous et de la logique floue impliquent l'utilisation de ces concepts. De nos jours, la logique floue est utilisée dans plusieurs domaines. Le développement le plus important durant la deuxième phase avait probablement pour origine la croissance rapide du contrôle flou et l'essor des applications de la logique floue, particulièrement au Japon (**Sumathi & Surekha, 2010**).

Un développement important faisant le début de la troisième phase avec l'apparition de l'article intitulé "From Computing with Numbers to Computing with Words" (**Zadeh, 1999**) où l'auteur a principalement indiqué que le développement de l'informatique avec les mots et les perceptions rassemble les premiers éléments de la logique floue et suggère que les théories scientifiques devraient être basées sur la logique floue plutôt que sur la logique Booléenne bivalente.

III.3.3. Concepts de base de l'approche de la logique floue

Dans ce chapitre, nous commençons par la clarification de plusieurs termes clés, tels que l'univers du discours, la variable linguistique, les termes ou étiquettes linguistiques, les ensembles flous, les fonctions d'appartenance et les règles floues.

III.3.3.1. Univers du discours

L'univers du discours est l'ensemble de référence d'un mot du langage naturel. Il contient des termes qui évoquent à différents degrés, le même concept que la variable floue et couvre l'ensemble des valeurs prises par cette variable. L'univers de discours est appelé aussi référentiel (**Aouchar & Bekhaled, 2005**).

III.3.3.2. Les variables linguistiques

Une variable linguistique (dite aussi variable floue) est une variable dont les valeurs sont des mots plutôt que des nombres. En effet, une grande partie de la logique floue peut être considérée comme une méthodologie pour modéliser avec des mots plutôt qu'avec des nombres. Bien que les mots soient intrinsèquement moins précis que les nombres, leur utilisation est plus proche de l'intuition humaine. De plus, le calcul avec des mots exploite la tolérance à l'imprécision et diminue ainsi le coût de la solution.

Une variable linguistique est définie par le quadruplet suivant (**Driankov et al., 1993**):

$$\{X(\text{nom}), L(X), (U), \mu(X)\} \quad (16)$$

Où :

$X(\text{nom})$ est le nom de la variable linguistique, comme taille, température, pluie, âge, concentration, etc.

$L(x) = \{L_1, L_2, \dots, L_n\}$ est l'ensemble des valeurs linguistiques que peut prendre la variable $X(\text{nom})$. Par exemple $L(X) = \{\text{grand, petit, moyen}\}$. $L(x)$ est aussi appelé ensemble des étiquettes associé à la variable $X(\text{nom})$ ou ensemble des termes (**Benkaci, 2006**).

U est dénommé univers du discours, il représente le domaine associé à la variable $X(\text{nom})$. Pour la variable linguistique taille, U peut être défini dans l'intervalle [1,60 m – 1,99 m]. Et finalement $\mu(X)$ sont les fonctions d'appartenance associées à l'ensemble de termes linguistiques.

Une variable linguistique peut être plus informative et plus généralisable qu'une variable numérique (**Nakoula, 1997 in Benkaci, 2006**).

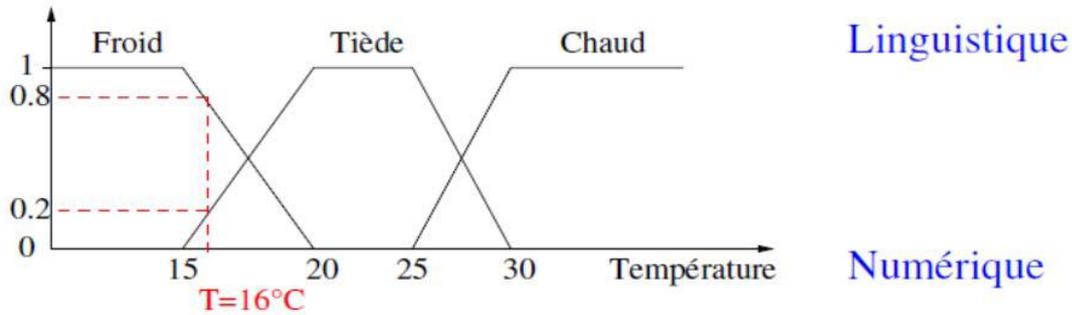


Figure 17 La variable linguistique température.

La Figure 17 montre un exemple de variable linguistique « température » avec trois termes (étiquettes) linguistiques, « froid », « tiède », « chaud ». Ainsi, la valeur numérique de température 16°C peut être considérée comme « froide » avec un niveau de 0,8 et comme « tiède » à un niveau de 0,2. Elle n’est pas considérée comme chaude (Fusco, 2016).

III.3.3.3. Les ensembles flous

À la différence des ensembles classiques, un ensemble flou, comme son nom l’indique, est défini avec des limites incertaines. En fait, dans ce type des ensembles, la transition de l’inclusion à la non-inclusion est effectuée progressivement.

Contrairement aux ensembles classiques, basés sur la logique binaire (0: $x \notin A$, 1: $x \in A$), La théorie des ensembles flous repose sur la notion d’appartenance partielle : chaque élément appartient partiellement ou graduellement aux ensembles flous qui ont été définis.

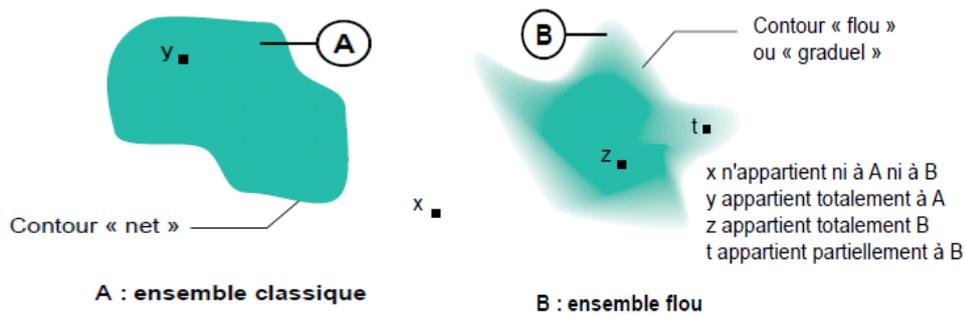


Figure 18 Comparaison d’un ensemble classique (A) et d’un ensemble flou (B).

Un ensemble flou est un ensemble d’objets sans limites claires; contrairement aux ensembles ordinaires où, pour chaque objet, on peut décider s’il appartient à l’ensemble ou non, une appartenance partielle à un ensemble flou est possible. Formellement, un ensemble flou est défini comme suit:

Définition1: Soit U l’univers du discours. A est appelé un sous-ensemble flou de U si A est un ensemble de paires ordonnées tel que:

$$A = \{(x, \mu_A(x)) \mid x \in U \text{ et } \mu_A(x) \in [0,1] \}. \tag{17}$$

Où $\mu_A(x)$ est le degré d'appartenance de x dans A . La fonction μ_A est appelée fonction d'appartenance de A . Plus $\mu_A(x)$ est proche de 1, plus x appartient à A - plus il est proche de 0, moins il appartient à A . Si $[0,1]$ est remplacé par l'ensemble à deux éléments $\{0,1\}$, alors A peut être considéré comme un sous-ensemble ordinaire de U . Dans ce texte, pour simplifier, nous utilisons la notion d'ensemble flou à la place du sous-ensemble flou.

L'un des exemples les plus couramment utilisés d'un ensemble flou est l'ensemble des personnes de grande taille. Les ensembles flous fournissent une transition en douceur entre les membres et les non-membres d'un ensemble.

Définition 2: Un ensemble flou A de U est défini par une fonction d'appartenance qui associe à chaque élément x de U , le degré $\mu_A(x)$, compris entre 0 et 1 avec lequel x appartient à A . Lorsque $\mu_A(x)$ s'approche de 1, le «degré d'appartenance» de x dans A augmente.

III.3.3.3.1. Les opérations sur les ensembles flous

Les opérations d'ensembles flous sont une généralisation des opérations d'ensemble classiques. Les opérations les plus utilisées sont appelées opérations d'ensemble floues standard. Soit U un univers de discours et A, B soit les ensembles flous sur U . Il y a quatre opérations de base sur les ensembles flous:

Egalité des ensembles flous: Deux sous-ensembles flous sont égaux si leurs fonctions d'appartenance prennent la même valeur en tout point de U .

A et B sont égaux si:

$$\mu_A(x) = \mu_B(x) \quad \forall x \in U \quad (18)$$

Complément flou standard: Il est défini comme un ensemble flou sur U pour lequel $A'(x) = 1 - A(x)$ pour chaque x dans U .

Intersection floue standard: Elle est définie comme un ensemble flou sur U pour lequel:

$$(A \cap B)(x) = \min [A(x), B(x)] \quad \text{pour chaque } x \text{ dans } U. \quad (19)$$

Union floue Standard: Elle est définie comme un ensemble flou sur U pour lequel

$$(A \cup B)(x) = \max [A(x), B(x)] \quad \text{pour chaque } x \text{ dans } U. \quad (20)$$

III.3.3.4. Fonctions d'appartenance

Définition : Fonction $\mu_A(x)$ qui à toute valeur d'entrée x fait correspondre son degré d'appartenance à l'ensemble A . Cette valeur graduelle est comprise entre 0 et 1.

Une fonction d'appartenance est une courbe qui définit comment chaque point dans l'espace d'entrée de la variable linguistique est associé à une valeur d'appartenance entre 0 et 1. Les valeurs de la fonction d'appartenance mesurent le degré auquel les objets satisfont des propriétés définies de manière imprécise. Les fonctions d'appartenance peuvent

être de différentes formes singleton, triangulaires, trapézoïdales, sigmoïdes, etc. Les ensembles flous sont clairement définis par la fonction d'appartenance et la définition elle-même n'a rien d'ambigu.

Les fonctions d'appartenance doivent être correctement construites, car elles sont à la base du système. Une courbe possède en abscisse les valeurs de la variable d'entrée (x) qui peut être une distance, une vitesse ou une mesure de débit, et en ordonnée le degré d'appartenance à l'état $\mu_A(x)$ compris entre 0 et 1. Dans ce cas, la transition de l'appartenance à la non appartenance est graduelle (**Kauffman, 1973 in Benkaci, 2006**).

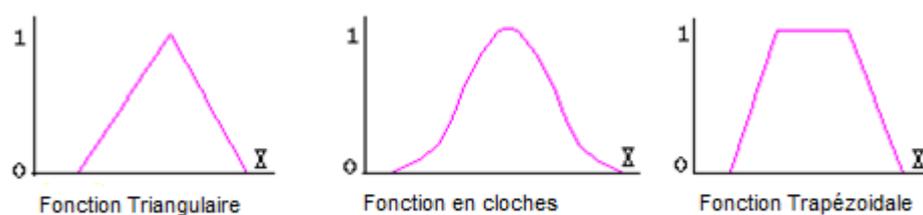


Figure 19 Différents types de fonctions d'appartenance.

Le plus souvent, des fonctions d'appartenance triangulaires, trapézoïdales et gaussiennes (bell-shaped) sont utilisées.

III.3.3.5. Les règles flous (Si.....Alors...)

Un système d'inférence flou a pour objectif d'inculquer le raisonnement intuitif humain à un ordinateur en vue d'exécuter des tâches telles que l'approximation de fonctions, la prévision et le control des processus, en utilisant plusieurs concepts à savoir : les variables linguistiques, les ensembles flous, les fonctions d'appartenances et les règles floues.

Les règles floues sont généralement du type "SI ... ALORS...." et permettent de représenter les relations entre les variables d'entrée et de sortie. Plus précisément une règle floue prend la forme suivante:

SI (x est A) **ET** (y est B) **ALORS** (z est C).

Où

x , y , z représentent les variables, et A, B, C sont des valeurs linguistiques ou des étiquettes (termes) linguistique dans l'univers de discours;

Ici, la partie « SI... » est appelée antécédent ou prémisse et la partie « ALORS... » est dite conséquent ou conclusion. L'opérateur booléen « ET » relie deux ou plusieurs antécédents.

La partie antécédente exprimée par "SI" (prédicat ou prémisse) décrit l'état du système. C'est une combinaison de propositions par des opérateurs ET, OU. La partie conséquente exprimée

par "ALORS" (conclusion) correspond à l'action que l'opérateur doit effectuer (**Aouchar & Bekhaled, 2005**).

Les règles floues constituent le noyau du système d'inférence flou, elles permettent de relier les variables floues d'entrée d'un système aux variables de sortie de celui-ci.

III.3.4. Systèmes d'inférence floue (SIF)

L'inférence floue est le processus de formulation de la relation reliant une entrée donnée à une sortie en utilisant la logique floue. Cette formulation fournit alors une base à partir de laquelle des décisions peuvent être prises. Il implique tous les sujets tels que la fuzzification, la défuzzification, l'implication et l'agrégation. Le contrôle expertes/ les connaissances en modélisation, l'expérience et l'association des variables d'entrée des modèles flous à la variable de sortie (ou aux variables) sont principalement basées sur des règles floues (**Sumathi & Surekha, 2010**).

Imitant le jugement humain dans le raisonnement logique (raisonnement humain de bon sens), le SIF utilise des valeurs linguistiques encadrées dans les règles floues de type (SI...ALORS...). Par exemple:

Si la température de l'eau est modérée et la concentration en ortho phosphate (PO_4) est élevée, Alors la concentration en chlorophylle-a est élevée.

Ici, les variables linguistiques la température de l'eau et la concentration en ortho phosphate sont des entrées; la variable linguistique « concentration en chlorophylle-a » est la sortie; Les valeurs *modérée*, *élevée* et *élevée* sont les valeurs (termes ou étiquettes) des variables linguistiques. La mise en œuvre du SIF nécessite le développement d'une base de connaissances qui permettrait de stipuler les règles floues (SI...ALORS...) en utilisant des ensembles flous. L'expérience et la connaissance des experts humains jouent un rôle important.

Les systèmes d'inférence flous sont uniques dans le sens qu'ils peuvent traiter simultanément des données numériques et des connaissances linguistiques. Du point de vue mathématique, un système d'inférence flou est une application non linéaire permet de relier les caractéristiques d'entrée à une sortie donnée. Le schéma général du modèle d'inférence floue est présenté dans la figure ci-après

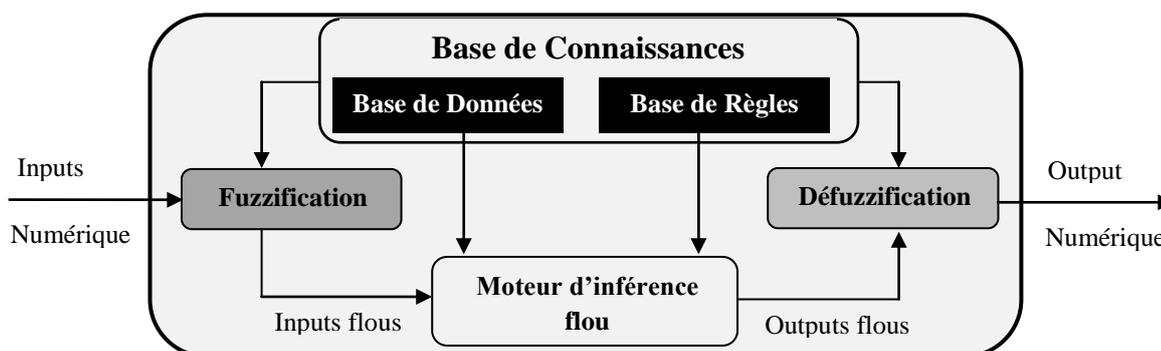


Figure 20 Blocs fonctionnels d'un système d'inférence floue

III.3.4.1. Etapes de modélisation par les systèmes d'inférence flous

III.3.4.1.1. La fuzzification

La fuzzification est la première étape dans la création d'un système d'inférence floue, il s'agit de passer du domaine numérique au domaine symbolique par la transformation des entrées numériques disponibles en parties floues définie sur un espace de représentation lié à l'entrée. Cet espace de représentation est normalement un sous-ensemble flou. Durant l'étape de la fuzzification, chaque variable d'entrée et de sortie est associée à des sous-ensembles flous. Cette étape est nécessaire dès lors que l'on veut manipuler à l'aide de la théorie des ensembles flous des grandeurs physiques mesurables (Lefkir, 2009). La fuzzification est un processus en deux étapes: attribuer des étiquettes floues et assigner une signification numérique à chaque étiquette.

III.3.4.1.1.1. Attribution d'étiquettes floues

Chaque entrée nette est affectée d'une étiquette floue dans l'univers du discours. Par exemple, pour le paramètre d'entrée taille, les étiquettes floues peuvent être «Petit», «moyen» et «Grand». Chaque entrée nette peut recevoir plusieurs étiquettes. Comme le nombre d'étiquettes augmente, la résolution du processus est meilleure. Dans certains cas, l'attribution d'un grand nombre d'étiquettes entraîne un temps de calcul important et rend le système flou instable. Donc en général le nombre d'étiquettes pour un système est limité à un nombre impair dans la gamme [3, 9], de sorte que la surface soit équilibrée et symétrique (Sumathi & Surekha, 2010).

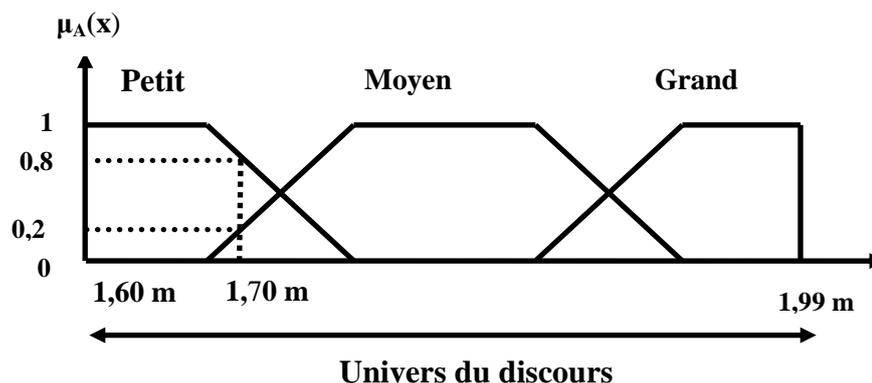


Figure 21 Etiquettes pour la variable linguistique taille

III.3.4.1.1.2. Assignation de signification numérique

Ici, les fonctions d'appartenance sont formées pour attribuer une signification numérique à chaque étiquette. La plage de la valeur d'entrée qui correspond à une étiquette spécifique peut être identifiée par la fonction d'appartenance. Bien qu'il existe différentes formes de fonctions d'appartenance, les fonctions d'appartenance triangulaires et trapézoïdales sont couramment utilisées pour éviter la complexité temporelle et spatiale.

Pour chaque ensemble flou et pour chaque variable linguistique, le degré d'appartenance d'une mesure nette dans chaque ensemble flou est déterminé. A titre d'exemple, la variable numérique taille qui a une valeur donnée de 1,7 m a été fuzzifiée en utilisant les fonctions d'appartenance trapézoïdales définies pour chaque ensemble flou pour la variable linguistique taille. À la suite du fuzzification, la variable linguistique taille a des valeurs linguistiques «Petit» et «Moyen», etc.

Dans une application de système expert flou, la valeur brute de chaque variable d'entrée est d'abord fuzzifiée en valeurs linguistiques avant que le moteur d'inférence ne procède au traitement avec la base de règles.

III.3.4.1.1.2. L'Inférence floue en utilisant les règles SI-ALORS

Les règles SI-ALORS peuvent être élaborées sur la base de connaissances des experts, de la littérature disponible ou de l'apprentissage automatique à partir de l'ensemble de données, la partie (SI) de chaque règle peut être calculée dans cette étape. L'étape d'inférence (produit de règles) affecte une ou plusieurs fonctions d'appartenance floue d'entrée (sous-ensemble) à un sous-ensemble de variables de sortie dans chaque règle.

La plupart du temps, chaque point de données de variable d'entrée fait partie de deux fonctions d'appartenance. Dans le cas de plusieurs variables d'entrée, chacune de ces valeurs d'appartenance est utilisée dans une règle différente avec la valeur d'appartenance d'une autre

variable. La partie "SI" des règles floues est liée aux fonctions d'appartenance aux variables d'entrée et est appelée prémisses ou antécédent; la partie "ALORS", qui est liée aux fonctions d'appartenance à la variable de sortie et s'appelle la conclusion ou la partie conséquente de la règle (Cobaner, 2011).

Le nombre de règles SI-ALORS dans un système dépend du nombre de fonctions d'appartenance et du nombre de variables d'entrée. En général, pour un système, le nombre de règles SI-ALORS floues est égal à la somme du produit de la multiplication de chaque variable d'entrée par le nombre de ses FAs.

III.3.4.1.3. Implication des règles floues

L'étape d'implication calcule la partie ALORS de chaque règle floue en utilisant le résultat de l'étape précédente. Chaque règle floue attribue une ou plusieurs fonctions d'appartenance à une seule fonction d'appartenance floue en sortie. Le résultat unique de la partie entrée de chaque règle est utilisé pour calculer la valeur d'appartenance ou pour activer la fonction d'appartenance en sortie.

Les opérateurs d'implication floue : Les définitions les plus courantes sont:

$$\text{Et (intersection): } \mu_{A \cap B}(x, y) = \mu_{AND}(x, y) = \min(\mu_A, \mu_B) \quad (21)$$

$$\text{Ou (union): } \mu_{A \cup B}(x, y) = \mu_{OR}(x, y) = \max(\mu_A, \mu_B) \quad (22)$$

III.3.4.1.4. Agrégation de différentes sorties de règles

L'agrégation est le processus par lequel les ensembles flous qui représentent les sorties de chaque règle sont combinés (agrégé) en un seul ensemble flou. A partir de l'étape précédente, différents résultats sont obtenus à partir de chaque règle, cette étape recueille toutes ces différentes valeurs d'appartenance de sortie (Toprak et al., 2009). L'agrégation ne se produit qu'une fois pour chaque variable de sortie, juste avant la dernière étape, la défuzzification. L'entrée du processus d'agrégation est la liste des fonctions de sortie tronquées renvoyées par le processus d'implication pour chaque règle. La sortie du processus d'agrégation est un ensemble flou pour chaque variable de sortie. Sen (2010) a considéré le processus d'implication et d'agrégation comme une étape unique de création d'un modèle flou.

III.3.4.1.5. Défuzzification

Le résultat agrégé, dans l'étape précédente, apparaît comme une valeur floue, donc il est nécessaire de transformer cette valeur floue à une valeur numérique normale, cette transformation d'un ensemble flou en un nombre net est appelée la défuzzification.

La défuzzification est le processus inverse de la fuzzification, elle convertit les parties floues relatives aux sorties du mécanisme d'inférence en sorties numériques. Il existe plusieurs techniques de défuzzification telles que le centroïde (centre de gravité), la bissectrice, le dernier des maxima, le milieu des maxima et le plus petit des maxima (Cobaner, 2011), mais la méthode la plus populaire et la plus utilisée est celle de centre de gravité.

III.3.4.1.5.1. La technique du centre de gravité

Cette méthode de défuzzification a été développée par Sugeno en 1985, c'est la méthode la plus utilisée. Elle est définie par l'expression algébrique suivante:

$$z^* = \int \frac{\mu_A(z) z dz}{\int \mu_A(z) dz} \quad (23)$$

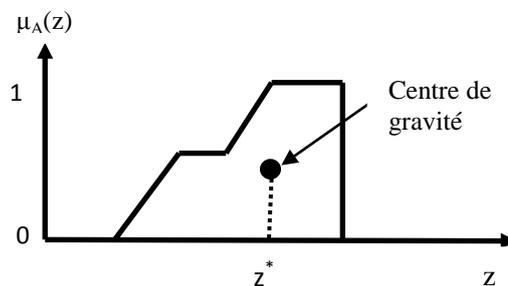


Figure 22 Défuzzification par la méthode du centre de gravité.

III.3.4.2. Classification de systèmes d'inférence floue

Il existe deux approches différentes pour créer un modèle flou (SIF). La première approche consiste à utiliser des informations provenant d'experts ou de connaissances de base sur le problème. La deuxième approche utilise l'apprentissage automatique tel que le réseau de neurones, ANFIS et les algorithmes génétiques pour en apprendre davantage sur les fonctions d'appartenance et les règles floues. La première méthode n'utilise pas l'historique des données disponibles pour faire l'apprentissage du modèle flou, elle utilise les informations et l'expérience acquises par des experts dans ce domaine. La seconde méthode s'appuie sur les données disponibles et utilise ces informations pour générer les règles floues. De plus, la première méthode de construction d'un modèle flou, appelée «raisonnement flou», fonctionne bien lorsque les données observées sont insuffisantes alors que le comportement du système est connu grâce aux connaissances de base sur le problème. La seconde méthode d'apprentissage automatique, telle que ANFIS, est utile lorsque les connaissances de base sur la structure et le comportement du système ne sont pas disponibles (Nayak et al., 2005; Chen et al., 2006 ; Sivanandam et al., 2007).

III.3.4.2.1. Systèmes d'inférence floue créés par les experts

Nous pouvons classer les modèles flous créés par les experts en deux catégories; les modèles linguistiques ou modèles de Mamdani, et les modèles à conséquences fonctionnelles ou modèles de Takagi-Sugeno-Kang.

III.3.4.2.1.1. Modèles de Mamdani

Parmi les méthodes d'inférence mentionnées ci-dessus, la technique de Mamdani est la plus utilisée en raison de sa structure simple (min-max). Le modèle a été proposé par Mamdani (1975) pour tenter de contrôler une chaudière à vapeur en synthétisant un ensemble de règles de contrôle linguistique obtenues d'opérateurs humains expérimentés. Un exemple d'un système d'inférence Mamdani est présenté dans la figure 23 expliquant les opérations impliquées dans le calcul à l'aide du modèle de Mamdani.

Soient X, Y deux variables linguistiques d'entrées, et Z la variable linguistique de sortie.

Le degré d'activation de x dans la fonction d'appartenance portant l'étiquette A_1 , dans la règle R_1 est défini comme $\mu_{A_1}(x)$, et le degré d'activation de y dans la fonction d'appartenance portant l'étiquette B_1 dans la règle R_1 est défini comme $\mu_{B_1}(y)$. Le degré d'activation de la règle R_1 est défini comme le minimum parmi les degrés d'activation de toutes les entrées dans la règle R_1 si l'opérateur "ET" est appliqué pour relier les valeurs indépendantes (entrées). La même chose se fait avec la règle R_2 . C'est aussi le maximum parmi les degrés d'activation de toutes les entrées dans la règle si l'opérateur "OU" est appliqué. La combinaison de règles dans la méthode de Mamdani est obtenue par la superposition des sorties de chaque règle. Un exemple graphique de FIS utilisant la méthode de Mamdani est montré dans la figure 23.

Le SIF montré à la figure 23 est constitué de deux règles (si-alors). Deux variables indépendantes (X, Y) sont prises en compte dans ce système pour estimer les valeurs de variable de sortie (Z). Pour la simplicité, toutes les fonctions d'appartenance sont de type trapézoïdal.

Comme le montre cette figure, les valeurs observées de x pour la première variable d'entrée et y pour la seconde variable d'entrée activent la règle R_1 et la règle R_2 . Chaque règle fournit un ensemble flou comme variable de sortie, lié au degré minimal d'activation des variables d'entrées.

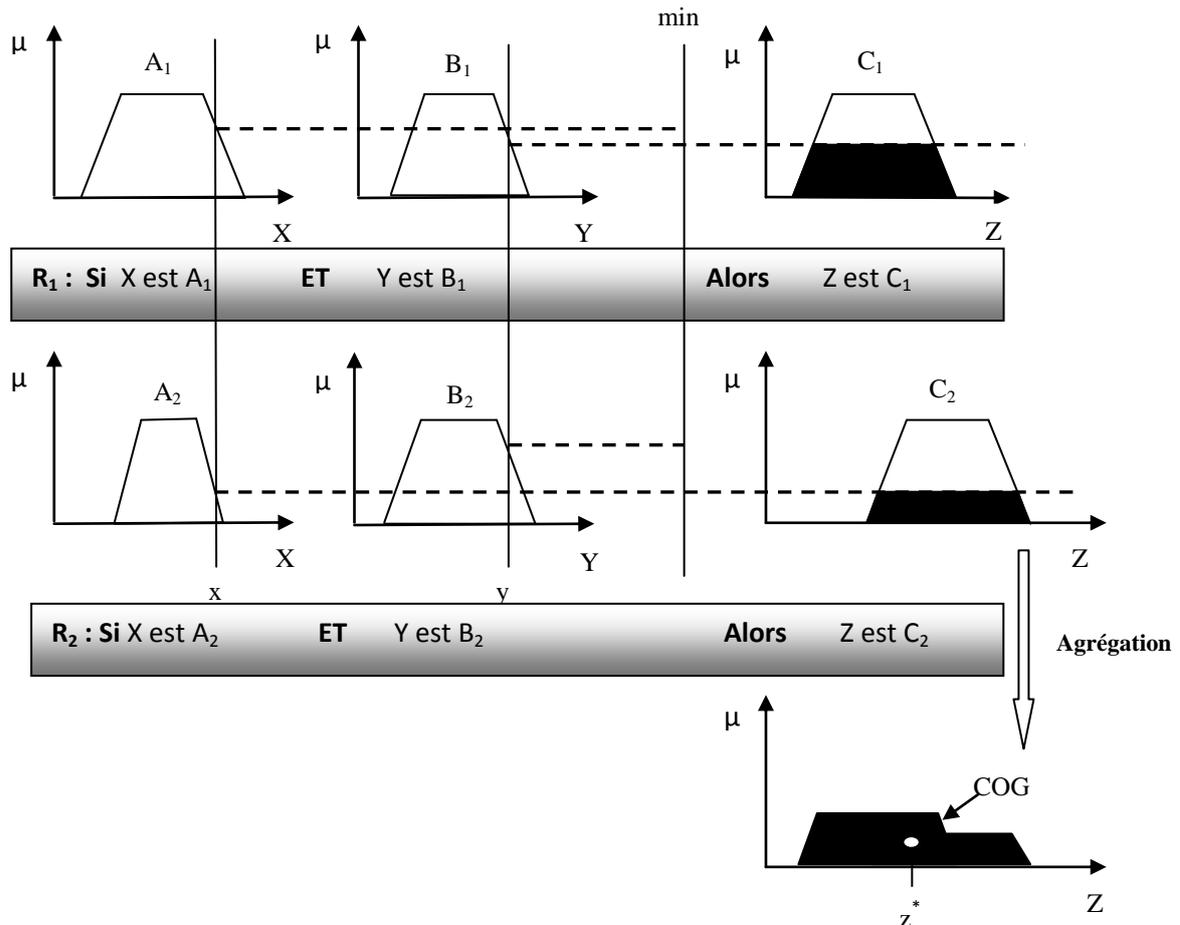


Figure 23 Schéma du système d'inférence flou de Mamdani

La contribution de la sortie de chaque règle est liée aux valeurs minimales du degré d'activation des variables d'entrées de celle-ci lorsque l'opérateur "ET" est utilisé. La réponse finale est obtenue par l'agrégation de ces sorties floues (superposition), qui représentent une variable floue et pourraient être défuzzifiées pour représenter la valeur représentative de la plage estimée de la variable dépendante.

La défuzzification est le processus consistant à produire un résultat quantifiable dans la logique floue, à partir d'ensembles flous et de degrés d'appartenance correspondants.

Une technique de défuzzification commune est celle du centre de gravité. Tout d'abord, les résultats des règles doivent être agrégés. Maintenant, les fonctions d'appartenance de sorties de chaque règles devaient être coupé en ligne droite horizontale quelque part entre le haut et le bas et que la partie supérieure devait être enlevée, la partie restante forme un trapèze (rempli et coloré en noir dans la figure).

Tous ces trapèzes se superposent alors, formant une seule forme géométrique. Ensuite, le centre de gravité de la forme obtenue est calculé. La coordonnée du centre de gravité est la valeur défuzzifiée (z^*) ou la valeur finale du SIF.

III.3.4.2.1.2. Modèles de Takagi-Sugeno-Kang

Le modèle Takagi-Sugeno, également connu sous le nom de méthode TS, a été proposé par Takagi et Michio Sugeno en 1985 afin de développer une approche systématique pour générer des règles floues. Il est similaire à celui de Mamdani dans plusieurs aspects. Les deux premières parties du processus d'inférence floue à savoir; la fuzzification des entrées et l'application d'opérateurs flous, sont exactement les mêmes. La principale différence entre Mamdani et Sugeno est que les fonctions d'appartenance à la sortie de modèle Sugeno sont linéaires ou constantes.

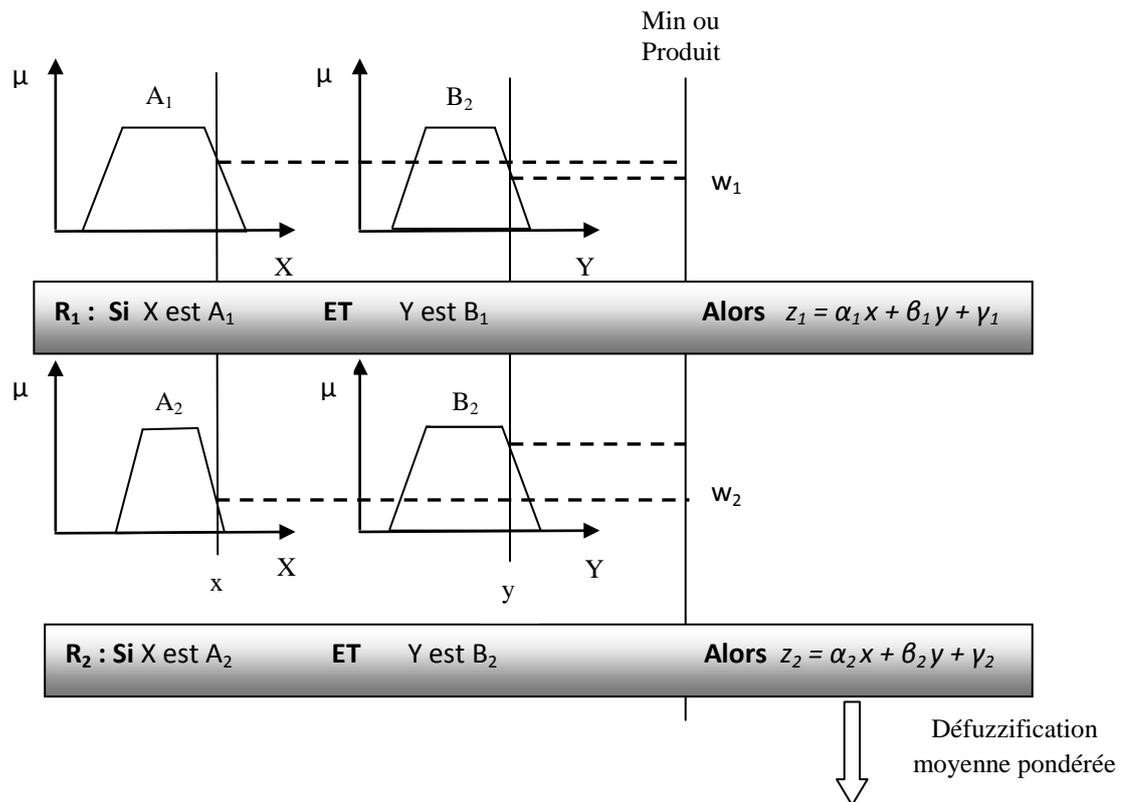


Figure 24 Schéma du système d'inférence flou de Sugeno.

$$z^* = \frac{w_1 z_1 + w_2 z_2}{w_1 + w_2}$$

Dans la méthode TS, également appelée méthode paramétrique, les conséquences sont des équations paramétriques linéaires représentées en fonction des entrées du système. La forme générale d'une règle floue de TS est:

SI antécédent₁ **ET** antécédent₂ **ALORS** sortie = f (x, y)

Ici, sortie = f (x, y) est une fonction. Cette fonction mathématique peut être linéaire ou non linéaire. Les fonctions linéaires sont le plus souvent utilisées et les techniques adaptatives sont utilisées pour les équations non linéaires.

La fonction d'appartenance de la règle qui en résulte est de type singleton dans la méthode TS.

Quelques exemples de la méthode d'inférence TS:

SI x est petit ALORS $y = 3x-2$

SI x est grand ALORS $y = x+ 5$

L'évaluation des règles varie par rapport à la méthode d'inférence de Mamdani. Comme chaque règle a une sortie nette, la sortie globale est obtenue par la méthode de défuzzification de moyenne pondérée. L'utilisation de la méthode de moyenne pondérée réduit le processus fastidieux de défuzzification effectué dans un modèle de Mamdani. De plus, la seule partie floue du modèle Sugeno étant son antécédent, une distinction claire peut être faite entre la partie floue et la partie non floue.

La méthode Sugeno est efficace sur le plan informatique et convient parfaitement aux problèmes d'adaptation et d'optimisation, ce qui la rend plus efficace dans le domaine des systèmes de contrôle. Les contrôleurs Sugeno ont généralement des paramètres beaucoup plus ajustables dans la partie conséquente de la règle et le nombre de paramètres augmente de manière exponentielle avec l'augmentation du nombre de variables d'entrée.

III.3.4.2.2. Systèmes d'inférence floue créés par les données: ANFIS

III.3.4.2.2.1. Introduction

Système d'inférence neuro-floue adaptatif (ANFIS), développé par **Jang (1993)**, est un système hybride intégrant la capacité d'apprentissage des réseaux de neurones et l'interprétabilité linguistique de la logique floue dans un cadre unique afin d'approximer les fonctions non linéaires et prendre en compte l'incertitude et la non linéarité, en d'autres termes, les systèmes neuro-flous ont été développés en combinant la transparence sémantique des règles basée sur les systèmes d'inférence floue avec la capacité d'apprentissage des réseaux de neurones (**Kurtulus et Razack, 2010**).

III.3.4.2.2.2. Structure

La structure d'ANFIS est semblable au réseau de neurones de type feed-forward à cinq couches qui combine des algorithmes d'apprentissage de RNAs et une logique floue afin de mapper l'espace d'entrée avec l'espace de sortie. ANFIS est un système d'inférence floue de type *Sugeno* qui a une entrée multiple et une sortie unique (MISO). Pour simplifier, nous supposons que le système d'inférence floue considéré a deux entrées x et y et une sortie f et contient deux règles de Takagi-Sugeno (Si...Alors...) comme suit:

Règle 1: Si (x est A_1) et (y est B_1) alors: $f_1 = p_1 x + q_1 y + r_1$ (24)

Règle 2: Si (x est A_2) et (y est B_2) alors: $f_2 = p_2 x + q_2 y + r_2$ (25)

Avec : A_1 , A_2 et B_1 , B_2 sont les fonctions d'appartenance (MF) pour les entrées x et y respectivement. p_1 , q_1 , r_1 et p_2 , q_2 , r_2 sont les paramètres des fonctions de sortie. Le système de raisonnement flou de Sugeno (ANFIS) qui en résulte est illustré dans la figure 25.

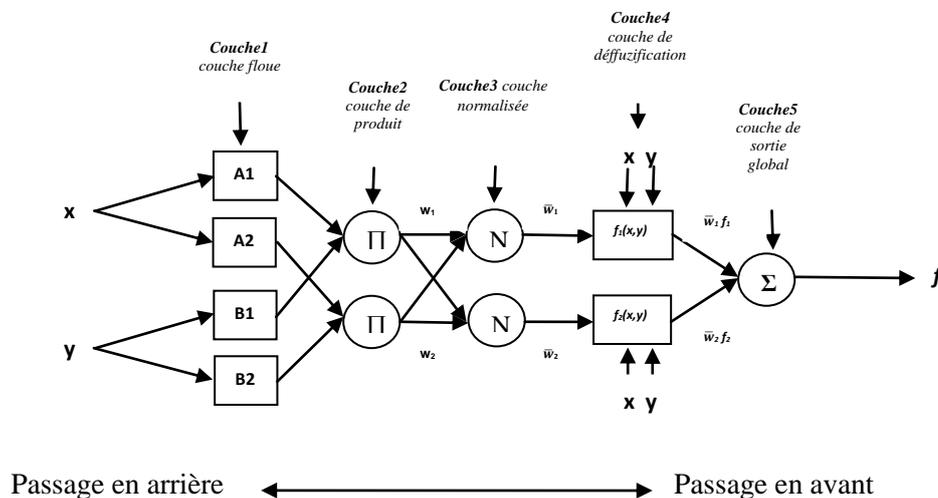


Figure 25 Structure D'ANFIS

Chaque couche est composée de nœuds, les nœuds carrés sont adaptatifs, les règles et la structure de chaque couche sont expliqués comme suit:

Couche 1: couche floue

Elle est appelée couche floue dans laquelle les variables d'entrée sont fuzzifiées, la fuzzification implique que chaque nœud de cette couche génère des degrés d'appartenance des entrées réelles et que la sortie de chaque nœud O_i^1 est calculée par:

$$O_i^1 = \mu_{A_i}(x) \quad i = 1, 2; \quad (26)$$

$$O_i^1 = \mu_{B_{i-2}}(y) \quad i = 3, 4; \quad (27)$$

Où x est l'entrée du nœud i , et A_i est l'étiquette linguistique (linguistic label) (bas, élevé, etc.)

associée à cette fonction de nœud. Cela signifie que O_i^1 est la fonction d'appartenance de A_i et spécifie le degré auquel le x donné satisfait le quantificateur A_i . Habituellement, nous choisissons $\mu_{A_i}(x)$ pour être une fonction d'appartenance gaussienne, telle que:

$$\mu_{A_i} = \exp[-0.5 * \{(x - c_i) / \sigma_i\}^2], \quad (28)$$

$$\mu_{B_i} = \exp[-0.5 * \{(y - c_i) / \sigma_i\}^2]. \quad (29)$$

Avec : c_i et σ_i sont les paramètres de la fonction d'appartenance d'entrée, ils sont également appelés paramètres de prémisse.

Couche 2: couche de produit

Le rôle de cette couche est de calculer des poids de règles nommés « firing strengths » (degrés d'activation) de règles floues particulières. Il se compose de nœuds circulaires marqués Π ,

indiquant qu'il fonctionne comme un multiplicateur simple, qui multiplie les signaux entrants et envoie le produit comme suit:

$$O_i^2 = w_i = \mu A_i(x) \times \mu B_i(y), \quad \text{for } i = 1, 2. \quad (30)$$

Chaque sortie de nœud (w_i) représente le degré d'activation d'une règle.

Couche 3: couche normalisé

Chaque nœud de cette couche normalise les poids de fonctions obtenues à partir de la couche de produit précédente,

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad \text{for } i=1, 2. \quad (31)$$

Les sorties de cette couche seront appelées degrés d'activation normalisés (normalized firing strengths).

Couche 4: couche de déffuzzification

La couche 4 est constituée de nœuds adaptatifs. Le nœud i de cette couche calcule la contribution de la $i^{\text{ème}}$ règle à la sortie du modèle comme suit:

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad (32)$$

Où \bar{w}_i est la sortie de la couche 3, et $\{p_i, q_i, r_i\}$ est l'ensemble des paramètres de sortie de la règle i . Les paramètres de cette couche sont appelés paramètres conséquents ou paramètres de la fonction de sortie.

Couche 5: couche de sortie totale

C'est une couche à nœud unique, étiquetée par Σ , et calcule la sortie globale comme la somme de toutes les entrées de la couche précédente. La sortie est calculée comme suit:

$$O_i^5 = \text{Overall output} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (33)$$

La figure 25 montre qu'ANFIS renferme deux couches adaptatives, la première couche et la quatrième couche qui sont constituées de nœuds carrés dans lesquels les paramètres de fonction d'appartenance sont ajustés.

III.3.4.2.2.3. Apprentissage d'ANFIS

ANFIS appartient à la catégorie des systèmes flous créés par les données, ils sont aussi appelés systèmes flous guidés par les données (data driven). ANFIS implémente un système d'inférence floue (TSK) de type *Sugeno* dont les règles et les paramètres de fonction d'appartenance sont extraits du jeu de données d'entrées/sortie en utilisant le processus d'apprentissage, autrement dit, ANFIS construit un système d'inférence flou de type Sugeno (FIS) dont les paramètres de fonction d'appartenances sont ajustés par l'algorithme de rétropropagation en combinaison avec la méthode des moindres carrés, cela signifie que ANFIS pratique un processus d'apprentissage hybride pour l'estimation des paramètres des parties prémisses et des parties conséquentes. Les paramètres de prémisse (paramètres des

fonctions d'appartenance d'entrée) sont ajustés à l'aide de l'algorithme de rétropropagation et les paramètres conséquents (paramètres des fonctions d'appartenance de sortie) sont ajustés par l'algorithme des moindres carrés afin d'identifier facilement les valeurs optimales. Ces ajustements permettent à nos systèmes flous d'apprendre à partir des données qu'ils modélisent.

Le développement du modèle ANFIS nécessite deux processus, l'identification de la structure et l'estimation des paramètres. L'identification de la structure implique le processus d'extraction des règles floues, tandis que l'identification des paramètres implique la recherche des paramètres de prémisses et des paramètres conséquents des systèmes ANFIS (**Jang et al., 1997**).

Dans cette partie, nous avons utilisé deux méthodes pour l'identification de la structure: la « grid partitioning » (GP) et la méthode de « subtractif clustering » (SC). ANFIS-GP indique que le modèle ANFIS utilise la grid partitioning et ANFIS-SC indique que le modèle ANFIS utilise le subtractif clustering pour l'identification de la structure.

III.3.4.2.2.3.1. Subtractive clustering

Le « Subtractive clustering » est proposé par **Chiu (1994)** en étendant la méthode de clustering en montagne (mountain clustering method). C'est une technique de génération automatique de systèmes d'inférence floue en détectant les clusters dans les données d'entrées-sorties de manière non supervisée, les points de données eux-mêmes sont considérés comme candidats pour les centres de cluster au lieu de points de grille. Il produit une partition dispersive (scattering partition) de l'espace d'entrée.

Le « Subtractive clustering » est basée sur une mesure de la densité des points de données dans l'espace des caractéristiques. L'idée derrière le Subtractive clustering est de déterminer les régions dans l'espace de caractéristiques avec des densités élevées de points de données. Le point avec le nombre maximum de voisins est sélectionné en tant que centre pour un cluster, les points de données dans un rayon flou préspecifié sont soustraits et l'algorithme recherche un nouveau point avec le plus grand nombre de voisins. Ce processus se poursuit jusqu'à ce que tous les points de données soient examinés (**Sumathi et Surekha, 2010**). La plage d'influence d'un cluster dans les différentes dimensions des données est connue sous le nom de rayon de cluster (radius).

Le rayon d'influence est un paramètre pour le subtractive clustering du système d'inférence floue. Le rayon est le seul paramètre qui influence le nombre de règles dans le modèle ANFIS-SC. Plus la valeur du rayon est basse, plus le nombre de règles est élevé, et vice versa.

Il est important de noter que le nombre de clusters est égal au nombre de règles. Les valeurs de rayon sont comprises entre 0 et 1, les meilleures valeurs de rayon sont généralement comprises entre 0,2 et 0,5. La valeur optimale du rayon est déterminée par la méthode de 'l'essai et l'erreur' (trial and error) en faisant varier le rayon (r) entre 0.1 et 1 avec un incrément de 0.1, les paramètres optimaux sont déterminés (voir tableau 3 colonne 2) en tenant compte des critères statistiques (RMSE) au cours des périodes d'apprentissage, de validation et de test en vue d'éviter le problème du surparamétrage (overfitting).

La combinaison d'ANFIS et le subtractive clustering a été appliquée pour l'évaluation de la dose de coagulant dans une station d'épuration (**Heddam et al., 2001**), la modélisation de l'évapotranspiration (**Cobaner, 2011**), la modélisation de la demande chimique en oxygène du système d'épuration (**Ay et Kişi, 2014**).

III.3.4.2.3.2. Grid partition (Partitionnement en grille)

Le 'grid partition' assure un partitionnement linéaire régulier de l'espace, il divise l'espace de données en sous-espaces rectangulaires en utilisant une partition à axe parallèle basée sur un nombre prédéfini de fonctions d'appartenance et leur type dans chaque dimension. Cette technique produit un partitionnement en grille de l'espace de données et génère des règles en énumérant toutes les combinaisons possibles de fonctions d'appartenance de toutes les entrées; ceci conduit à une explosion exponentielle de nombre de règles floues même lorsque le nombre d'entrées est modérément grand. Le grid partition n'est appropriée que pour les cas avec un petit nombre de variables d'entrée (**Wei et al., 2007**). Un inconvénient majeur de la technique de grid partition est appelé malédiction de dimensionnalité, cela signifie que le nombre de règles floues se multiplie rapidement et exponentiellement avec l'augmentation du nombre de variables d'entrée et du nombre de fonctions d'appartenance pour chaque variable, donc l'application de grid partition est menacé par le grand nombre de règles floues, ce qui fait que l'ordinateur s'arrête de fonctionnement.

Le nombre de règles (NR) est calculé comme suit:

$$\mathbf{NR} = (\mathbf{NMF})^{NI} \quad (34)$$

Avec:

NMF est le nombre de fonctions d'appartenance et NI est le nombre de variables d'entrée. Un autre problème concernant l'application de cette méthode est que la construction des fonctions d'appartenance de chaque variable n'est pas dépendante les unes des autres, donc la relation entre les variables est omise.

Plusieurs auteurs ont combiné ANFIS et la technique de grid partition pour prévoir les fluctuations de profondeur de la nappe phréatique (Shiri et Kişi, 2011), l'évapotranspiration (Ladlani et al., 2013, Cobaner, 2011) et l'analyse épidémiologique de la rouille du soja (Alves et al., 2011).

III.4. Phases d'élaboration d'un modèle

Elles se subdivisent en plusieurs étapes, chacune dépend de la performance de la phase précédente.

III.4.1. La collecte des données

L'acquisition des données est la phase essentielle de toute recherche appliquée. La qualité des données est une condition *sine qua non* de la qualité d'une étude. En modélisation, les données servent non seulement à élaborer un modèle mais aussi à le tester. Sans données en quantité et qualité suffisantes, le modèle aura donc peu de chances d'être performant et opérationnel (Oudin, 2004).

L'objectif est à la fois de caler et valider le modèle. Il consiste à rassembler un nombre de données suffisant pour constituer une base représentative nécessaire à la modélisation.

III.4.2. Traitement de données

Cette étape comprend trois phases, nettoyage, normalisation et division de la base de données;

III.4.2.1. Analyse des données

L'analyse des données doit inclure l'opération de la modification des valeurs aberrantes, l'identification et la correction des erreurs apparentes, le comblement de lacunes (valeurs manquantes). Le graphique de la série temporelle ne doit présenter aucunes anomalies telles que les valeurs aberrantes et les discontinuités. Une aberrantes additive est une valeur aberrante isolée, un seul point isolé du graphique de la série, elle ne semble pas s'ajuster avec la forme 'normale' du graphique. Ce type d'aberrantes est typiquement une erreur d'enregistrement ou erreur de mesure.

III.4.2.2. Normalisation des données

Avant l'utilisation de l'ensemble de données pour la construction du modèle, les entrées et la sortie cible ont été normalisées linéairement entre 0 et 1 afin de libérer les données de l'arbitraire des unités de mesure utilisées pour chaque variable (Brigitte et Jérôme, 2008). "La normalisation vise également à ramener la plage d'évolution des valeurs prises par les variables à l'intérieur d'un intervalle standardisé, fixé a priori. Elle est souhaitable car elle

évite au système de se paramétrer sur une plage de valeurs particulières, ignorant ainsi les valeurs extrêmes'' (Paquet, 1997).

La formule suivante a été utilisée pour normaliser l'ensemble de données:

$$X_{\text{norm}} = \frac{X_i - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad (35)$$

Où : X_{norm} est la valeur normalisée, X_i est la valeur observée, X_{max} et X_{min} sont respectivement la valeur maximale et la valeur minimale de chaque variable. Lorsque l'apprentissage, la validation sont terminés, les valeurs observées et simulées sont dé-normalisées par l'opération inverse.

III.4.2.3. La division de la base de données (apprentissage-validation)

L'un des avantages majeurs des réseaux de neurones artificiels est leur aptitude de généralisation ; extrapolation pour les valeurs proches et interpolation pour les valeurs intermédiaires. Pour avoir une bonne généralisation, la base de donnée doit être divisée en deux parties, la première sert pour l'apprentissage (calage), et la deuxième servant pour la validation de RNA (performances du modèle). Il n'y a pas de règle pour estimer la longueur de chacune des deux parties, il faut effectuer donc plusieurs essais afin de trouver la meilleure combinaison (calage-validation).

Dans cette étude, l'ensemble de données sur la qualité de l'eau a été divisé en deux sous-ensembles en utilisant 70% des données pour l'apprentissage des modèles, 30% pour la validation. Le sous-ensemble d'apprentissage sert à ajuster les paramètres du modèle, et les sous-ensembles de validation ont été utilisés pour vérifier la capacité des modèles.

III.4.3. La sélection des variables d'entrées

La sélection des variables d'entrées est une tâche délicate et fastidieuse, elle constitue une étape très importante dans la modélisation de n'importe quel phénomène ou processus.

Pour évaluer l'effet des variables d'entrée sur le modèle, deux processus d'évaluation ont été utilisés. Premièrement, une connaissance à priori soutenue par l'analyse de corrélation statistique. Le deuxième processus d'évaluation est fondé sur l'analyse de sensibilité des RNAs.

III.4.3.1. La corrélation statistique

Le choix de variables d'entrée basés sur l'analyse de corrélation linéaire est la technique analytique la plus populaire pour sélectionner les entrées. L'inconvénient de cette technique est qu'elle est capable seulement de capturer la dépendance linéaire entre deux variables. Par

conséquent, elle peut entraîner l'omission de variables d'entrées importantes liés à la sortie de façon non linéaire, pour cela on fait appel à l'analyse de sensibilité par le biais des RNAs, celle-ci permet de dévoiler les liaisons de type linéaire et non-linéaire entre les entrées et la sortie correspondante.

III.4.3.2. L'analyse de sensibilité

L'analyse de sensibilité est utilisée pour déterminer la ou les variables d'entrée les plus influençant sur la sortie du modèle. Le simple fait d'utiliser toutes les informations disponibles peut ne pas toujours améliorer les capacités de prévision, car les variables d'entrée peuvent être corrélées entre elles. S'il n'y a pas beaucoup de différence sur les performances avec ou sans variable, cette variable est dite moins importante pour la cible et peut donc être supprimée des entrées du réseau (**Kate et Jatinder, 2001**).

Une analyse de sensibilité a été effectuée pour évaluer l'importance relative de chacune des variables de la qualité de l'eau (17 WQVs) dans la prévision des indicateurs de l'eutrophisation (Chl-a, PT...). Lors de la création de modèles, l'analyse de sensibilité consiste à évaluer l'importance des prédicteurs (variable d'entrée) dans les modèles conçus (fitted models).

Cette analyse classe les variables d'entrée en se basant sur la détérioration de la performance du modèle qui se produit si une variable est éliminée du modèle. Par la suite, on identifie les variables qui peuvent être ignorées en toute sécurité dans les analyses ultérieures ainsi que les variables essentielles qui doivent être maintenues (**Olszewski et al., 2008**). Les résultats de l'analyse peuvent être utilisés à des fins purement informatives ou pour l'élagage des variables d'entrée (**Gazzaz et al., 2012**).

Dans cette étude, l'analyse de sensibilité a été basée sur l'examen des effets de variables d'entrée (17 VQEs) sur la variable dépendante (Chl-a par exemple) suivant la méthode dite 'leave-one-out' ou «éliminer une seule variables de l'ensemble de variables candidates». Elle correspond à évaluer le changement de l'erreur de réseau de neurones qui sera obtenue si une variable d'entrée est éliminée (élaguée) de l'ensemble des entrées (**Ha et Stenstrom, 2003; Pastor-Bárcenas et al., 2005**).

Cette méthode applique deux indicateurs: *le ratio* de l'erreur du réseau et *le rang*. Le ratio des erreurs de réseau indique les effets sur le fonctionnement du réseau lors de l'élimination (élagage) d'une variable d'entrée. C'est le ratio de l'erreur obtenue après avoir éliminée une seule variable à l'erreur ait obtenu en utilisant toutes les variables, c.-à-d le rapport entre l'erreur du modèle élagué et celle du modèle complet (modèle utilisant toutes les variable

d'entrées). Plus le ratio est élevé, plus la variable d'entrée du modèle est importante, et vice versa. Quand le ratio est ≤ 1.0 , la variable peut être rejetés et supprimés du modèle. Ensuite, sur la base des ratios, des rangs sont affectés aux variables indépendantes tels que la variable ayant le ratio le plus élevé va porter le rang numéro un (**Olszewski et al., 2008**). Le ratio est calculé par:

$$Ratio = \frac{RMSE_{\text{modèle élagué}}}{RMSE_{\text{modèle complet}}} \quad (36)$$

L'analyse de sensibilité nous permet donc de:

- ✓ Choisir le meilleur ensemble de variables d'entrée (les variables clés) pour améliorer la performance de modèle et assurer la parcimonie de ce dernier par l'élimination (élagage) des variables d'entrées auxquelles le modèle n'est pas sensible;
- ✓ Évaluer l'effet de chaque variable d'entrée sur le modèle par la détermination des variables d'entrées les plus influentes et celles les moins influentes (des entrées qui ne sont pas importants).

Au lieu de changer le nombre de variables d'entrée, une autre approche est de changer les valeurs d'une variable particulière. Plusieurs expériences sont entreprises en utilisant des variables perturbées. Un changement positif ou négatif est fait sur les valeurs originales d'une variable. De même, s'il n'y a pas beaucoup de différence sur la performance du modèle avec ou sans perturbation d'une variable, cette variable est moins important (**Kate et Jatinder, 2001**).

Évaluer l'effet de chaque variable d'entrée sur le modèle en déterminant un paramètre d'entrée sensible et des paramètres d'entrée insensibles. Un paramètre sensible est un paramètre qui modifie de façon significative les sorties d'intérêt du modèle par variation unitaire de sa valeur et un paramètre insensible est un paramètre qui a peu d'effet sur les sorties d'intérêt du modèle;

Le changement de sortie provoqué par le changement d'entrée est calculé.

La sensibilité de chaque entrée peut être donnée par:

$$Sensibilité = \frac{1}{N} \sum_{i=1}^N \left(\frac{\text{Changement en Output (\%)}}{\text{Changement en Input (\%)}} \right) * 100 \quad (37)$$

Où N est le nombre d'échantillon.

L'analyses de sensibilité est la méthode la plus couramment utilisée pour extraire l'information d'un modèle de RNAs (**Maier et al., 1998; Scardi et Harding, 1999; Lee et al., 2003**).

III. 4.4. Elaboration de la structure du modèle

La structure du réseau dépend étroitement du type des échantillons. Il faut d'abord choisir le type de réseau. Les réseaux de neurones utilisés dans ce mémoire sont de type « Perceptron multicouches », la fonction de transfert de la ou les couches cachées est sigmoïde par contre la fonction de la couche de sortie est de type linéaire.

Il est intéressant de signaler qu'on va jouer sur plusieurs paramètres: le nombre de couches cachées (1 ou 2), le nombre des inputs (combinaison entre les inputs), le nombre de neurones de la ou les couches cachées afin de mettre en évidence la meilleure structure neuronale.

III. 4.5. Apprentissage

L'apprentissage est l'avant dernière phase d'élaboration d'un réseau de neurones. Elle consiste tout d'abord à calculer les pondérations optimales des différentes liaisons, en utilisant un échantillon constitué de couples (entrées-sorties désirées).

L'apprentissage utilisé est de type supervisé, l'Algorithme de rétro propagation de l'erreur (backpropagation), variante de Livenberg-Marquardt en raison de sa rapidité de convergence et ses performances de généralisation. Il consiste en la modification des poids des connexions entre les différentes couches par rapport à une erreur calculée en sortie de réseau.

III.4.6. Validation

Une fois le réseau de neurone est entraîné (apprentissage), une tâche délicate c'est l'évaluation de ses capacités de généralisation, c'est-à-dire il faut voir s'il est prêt pour l'usage, pour cela on va le tester sur une base de données différente de celle utilisée pour l'apprentissage. Durant cette phase on n'apporte au réseau que les entrées (données climatiques), sans faire appel à la sortie désirée (target). Ce test permet d'apprécier les performances du modèle élaboré.

Si le modèle neuronal élaboré n'est pas performant, il faudra modifier la structure de réseau ou modifier la base d'apprentissage (la longueur de celle-ci).

III. 4.7. Post validation ou phase de test

Parfois on divise la base de données en 3 ensembles l'un pour l'apprentissage (calage du modèle), l'autre pour le valider, et le dernier pour le test, cette phase nous permet de bien confirmer la performance du modèle élaboré durant les deux phases citées auparavant, pendant celle-ci on n'apporte au réseau que les entrées (sans target).

III.5. Implémentation des modèles

Pour les applications pratiques des réseaux de neurones et des systèmes d'inférences flous, l'utilisateur doit comprendre les concepts théoriques de base, mais il doit également maîtriser des outils appropriés en vue de créer des modèles permettant la simulation et la prédiction de phénomènes ou de processus que l'on veut étudier.

Depuis le développement théorique des réseaux de neurones à la fin des années 1980-1990, plusieurs outils ont été mis à la disposition des utilisateurs. Les fonctionnalités et les statuts de ces outils sont très variables. (Dreyfus, 2004).

Aujourd'hui, les outils d'implémentation des modèles de RNAs disponibles se divisent en deux catégories:

- Des outils spécifiques développés incluant une méthodologie complète (simulateurs neuronaux), qui ne demandent aucun effort de programmation, on cite à titre d'exemple NeuroOne6 ; ce genre d'outils ne fournit pas une flexibilité mais il donne des résultats fiables dans un temps court.
- Les boîtes à outils de réseaux de neurones telle que « artificial neural network toolbox built-in Matlab » qui est un sous programme (built-in) intégré dans le logiciel MATLAB.

Pour l'implémentation des modèles neuro-flous (ANFIS), nous avons utilisé la boîte à outil de la logique floue existant dans Matlab « fuzzy logic toolbox ».

Le logiciel utilisé pour implémenter tous les modèles de RNAs et d'ANFIS est Matlab v7.4, précisément, la boîte à outils de réseaux de neurones artificiels et celle de la logique floue intégrées dans Matlab. Plusieurs codes de programmes ont été écrits en langage Matlab, et utilisé pour développer différents modèles de RNAs et d'ANFIS tout en faisant le lien *Excel-MATLAB* pour faciliter l'importation de données vers *MATLAB* et l'exportation des résultats vers Excel.

III.6. Critères d'évaluation de performance des modèles

Le seul critère de conformité et de performance d'un modèle est son adéquation avec les données observées.

Après avoir implémenté un modèle, une tâche délicate c'est l'estimation de sa qualité et sa performance, pour cela on utilise généralement plusieurs critères graphiques et statistiques.

III.6.1. Critère graphique

Ce critère donne une idée générale sur la performance de modèle du fait qu'il apporte un jugement visuel sur la qualité de modèle à reproduire fidèlement la réalité. Une superposition entre la courbe portant les valeurs simulées et celle portant les valeurs observées, signifie une parfaite performance de modèle et vice-versa. Mais ce critère n'est pas suffisant d'où la nécessité de faire appel aux critères statistiques.

III.6.2. Critères statistiques

Critères statistiques consistent à déterminer une fonction qui va minimiser l'erreur entre les données simulées par le modèle, et celles observées (**Benkaci, 2006**).

III.6.2.1. Coefficient de corrélation (CC)

Le coefficient de corrélation (CC) mesure la force de la relation entre deux variables. Les valeurs de CC varient de -1 à +1; la force de la relation est indiquée par la valeur absolue de coefficient, tandis que sa direction est indiquée par le signe. Le coefficient de corrélation (CC) peut être calculé à l'aide de la formule suivante:

$$CC = \left[\frac{\sum_{i=1}^n (Y_{obs} - \bar{Y}_{obs})(Y_{sim} - \bar{Y}_{sim})}{\sqrt{\sum_{i=1}^n (Y_{obs} - \bar{Y}_{obs})^2} \sqrt{\sum_{i=1}^n (Y_{sim} - \bar{Y}_{sim})^2}} \right] \quad (38)$$

Où CC est le coefficient de corrélation; Y_{obs} et Y_{sim} et sont les valeurs observées et simulées respectivement; et \bar{Y}_{obs} et \bar{Y}_{sim} sont la moyenne des valeurs observées et simulées, N est le nombre d'observations. Un coefficient de corrélation proche de l'unité signifie que le modèle conçu est très efficace et vice versa.

Selon **Legates et McCabe (1999)**, l'utilisation du coefficient de corrélation seul n'est pas suffisante pour évaluer la qualité d'un modèle. Pour cette raison, nous avons utilisé d'autres indicateurs comme Root Mean Square Error (RMSE).

III.6.2.2. Erreur quadratique moyenne (RMSE)

Le RMSE est la racine de la moyenne arithmétique des carrés des écarts entre les prévisions et les observations, il doit être aussi petit que possible. Une concordance parfaite entre les valeurs simulées et celles observées donnerait un RMSE = 0.

Le RMSE est un indicateur de la capacité du modèle, un faible RMSE indique que le modèle a une bonne performance, et vice-versa (Sungwon et Hungo, 2008). L'erreur quadratique moyenne (RMSE) a été utilisée comme une mesure de la qualité de l'ajustement du modèle, elle est calculée comme suit:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_{obs} - Y_{sim})^2}{N}} \quad (39)$$

Où : RMSE est l'erreur quadratique moyenne; et Y_{obs} et Y_{sim} sont les valeurs observées et simulées, respectivement; N est le nombre d'observations (échantillons).

III.6.2.3. Erreur absolue moyenne (MAE)

L'erreur absolue moyenne (MAE) est la moyenne des valeurs absolues des écarts entre les valeurs simulées et celles observées. Un MAE faible implique une bonne performance du modèle. MAE est calculé comme suit;

$$MAE = \frac{1}{N} \sum |X_{sim} - X_{obs}| \quad (40)$$

III.7. Description des réservoirs étudiés et les données de qualité de l'eau

III.7.1. Etude de cas n° 01 : base de données mixte (barrages Lekhal, Tichy-Haf et Keddara)

IV.7.1.1. Justification de l'utilisation d'une base de données mixte

La taille de la base de données d'apprentissage n'est pas très longue, notamment les données relatives aux concentrations de chlorophylle-a. Cela est dû aux lacunes de données, puisque les données de chlorophylles-a ne sont pas régulièrement mesurées au niveau des barrages, comme les autres paramètres de la qualité des eaux et même les données existantes ont un nombre de valeurs très faibles, pour remédier à ce problème, nous avons procédé au mélange de données provenant de 3 barrages, à savoir, le barrage de Lekhal, de Keddara et de Tichy-Haf. Par la suite, on va traiter ces données mélangées par les modèles de RNAs et d'ANFIS, comme étant constituées la même base de données, dans le but de créer un *modèle généralisé* permettant de simuler la concentration en chlorophylle-a pour les 3 barrages à la fois, et ceci grâce aux capacités et caractéristiques des techniques de machine learning, à savoir ; l'apprentissage numériques des données, la flexibilité et la généralisation (interpolation et extrapolation).

Plusieurs chercheurs ont utilisé la technique de combinaison de données pour modéliser l'évapotranspiration de référence par ANFIS en Iran et en Espagne (Shiri et al., 2013), et simuler la concentration en chlorophylle-a par les RNAs dans le détroit de Johor, Singapour, tout en utilisant des données provenant de plusieurs stations de mesure (Palani et al., 2008 ; Chen et Mynett, 2003). Les résultats obtenus étaient très satisfaisantes pour les deux cas.

Les modèles de Machine Learning ont l'avantage qu'à chaque fois on a de nouvelles données, on va alimenter la base de données d'apprentissage et celle de validation, ce qui permet l'élargissement de la base de données d'apprentissage et amélioration de performance du modèle développé.

Les variables de la qualité de l'eau utilisées dans cette partie ne forment pas des séries chronologiques au sens propre du mot, puisque ces dernières sont dispersées dans le temps (absence de pas de temps régulier entre les échantillons) et dans l'espace (données mixtes provenant de trois barrages).

III.7.1. 2. Barrage Lekhal

III.7.1. 2.1. Situation géographique du barrage

Le barrage Lekhal est situé sur le territoire de la Wilaya de Bouira, sur l'oued Lekhal à 30 km au sud de la ville de Bouira. C'est un ouvrage en terre construit au point de confluence de deux Oued: l'oued Lakhal et l'oued Fahem. Sa longueur est de 630 m et sa hauteur est de 45m, il comporte un déversoir à seuil libre pouvant débiter 550 m³/s et deux vidange de fond d'un débit 13,5 m³/s chacun. La retenue a une capacité utile de 30 hm³ et d'une surface de 2,5 Km², permettant la régularisation annuelle de 22 hm³, destiné à l'alimentation en eau potable et industrielle des régions d'Ain Bessem et Sour El-Ghozlane (Dotation annuelle en 2007 : 4,0 hm³), ainsi qu'à l'irrigation des plaines des Arribs (dotation annuelle pour la compagnie d'irrigation en 2007 est de 7,0 hm³).

Caractéristiques hydrologiques

Oued : Lekhal

Capacité initiale : 30 hm³

Apport moyen annuel : de 22 hm³

Envasement annuel : 0,17 hm³/an

Surface du BV : 189 Km²

Caractéristiques du barrage

Année de mise en eau : 1985

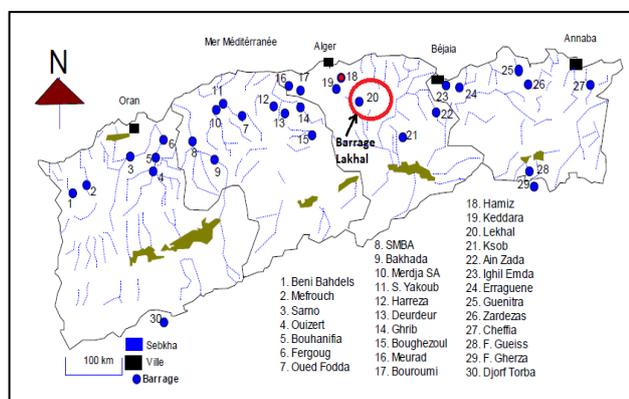


Figure 26 Situation géographique du barrage Lekhal

Type : Terre

Hauteur : 45m

Longueur : 630m

Côte de retenue normale (R.N) : 684,4m

Côte plus hautes eaux (P.H.E) : 687,00 m

Déversoir : seuil libre : 550 m³/s

Vidange de fond: 2X13.5 m³ /s

III.7.1.2.2. Données de la qualité de l'eau utilisées

Les données de la qualité de l'eau obtenues auprès de l'ANRH avec un pas de temps mensuel s'étalent sur période de 2 ans (2010 et 2011).

Tableau 3 Statistiques des paramètres de la qualité de l'eau pour le barrage Lekhal

	Moyenne	Ecart type	Min	Max	CV %
Conductivité (dS/m)	806	81.5977	628	1000	10.12
Résidu sec (mg/l)	524.691	59.2078	419	619	11.28
DCO (mgO₂/l)	16.909	7.9935	7	36	47.27
NH₄ (mg/l)	0.324	0.4182	0	1.42	129.07
NO₃ (mg/l)	2.9818	2.3035	0	8.9	77.25
NO₂ (mg/l)	0.2666	0.2591	0	0.89	97.18
TIN (mg/l)	3.5725	2.4914	0.13	9.914	69.73
Oxygène dissous (mg/l)	8.1818	1.9987	5.3	15.1	24.42
Saturation en oxygène%	94.037	37.2187	61.2	198.2	39.57
pH	8.0727	0.3757	7.2	8.8	4.65
P total (mg/l)	0.2678	0.2185	0.085	0.7822	81.59
PO₄⁻³ (mg/l)	0.1534	0.1839	0	0.673	119.88
TC	18.5773	7.0854	6.5	27	38.14
Turb	9.3156	7.894	2.4373	30.6	84.73
SD (m)	1.0636	0.3867	0.4	2	36.35
MO (mg/l)	6.5510	1.2999	4.287	8.6	19.84
Chl-a (µg/l)	15.1569	11.3411	4.6771	47.607	74.82

Les statistiques de base des paramètres de qualité de l'eau mesurés dans le réservoir de Lekhal sont illustrées dans le tableau 3. Dans ce tableau, max, min, et CV désignent respectivement le maximum, le minimum et le coefficient de variation.

Les variables de la qualité de l'eau utilisées sont : la conductivité électrique (CE), le résidu sec (RS), la demande chimique en oxygène (DCO), l'ammonium (NH_4^+), le nitrate (NO_3^-), le nitrite (NO_2^-), l'azote inorganique total (TIN), l'oxygène dissous (O_2), le taux de saturation ($\text{O}_2\%$), le pH, le phosphore total (Ptot), le phosphate (PO_4^{2-}), la température de l'eau, la turbidité, la visibilité de l'eau (disque de Secchi en m), la matière organique (MO) et la chlorophylle-a. L'azote inorganique total (TIN) a été calculé comme $\text{TIN} = \text{NH}_4^+ + \text{NO}_3^- + \text{NO}_2^-$.

Le coefficient de variation représente la dispersion des données autour de la moyenne. Il représente le ratio de l'écart-type rapporté à la moyenne.

Les variables de la qualité de l'eau de ce barrage affichaient des coefficients de variations oscillant entre 4.65 et 129.07 %.

Il ressort du tableau 3 que les différentes formes de l'azote inorganique (NH_4 , NO_2 , NO_3) enregistrent une large variation avec des coefficients de variation de 129.07%, 77.25% et 97.18% respectivement. Le phosphate et le phosphore total ont également enregistré une large variation autour de la moyenne (119.88 % et 81.59 % respectivement).

Le pH et la température de l'eau ont montré une faible variation avec des coefficients de variations de 4.65% et de 38.14% respectivement, cela est due au pouvoir tampon du réservoir (la retenue de barrage).

La concentration en chlorophylle-a varie entre 4.6771 $\mu\text{g/l}$ et 47.607 $\mu\text{g/l}$ avec une moyenne de 15.1569 $\mu\text{g/l}$. Cette variable de la qualité de l'eau montre une variation relativement élevée avec un coefficient de variation de 74.82%.

On peut conclure que les variables d'origine anthropique (azote et phosphore) ont montré des variations plus importantes que celles des variables d'origine naturelle (pH, température de l'eau, conductivité et RS). Cela peut être dû au fait que les processus géogéniques (d'origine naturels) sont presque à l'état d'équilibre, alors que les processus anthropogéniques dépendent du temps dans la nature.

III.7.1.3. Barrage de Tichy Haf

III.7.1.3.1. Situation géographique du barrage

Le barrage de Tichy-Haf (36° 23' 26" Nord, 4° 23' 25" Est) est situé sur l'oued de Boussellam à 10 Km au Sud Est de la ville d'Akbou, dans Wilaya de Bejaïa, sa construction a débuté en l'an 2000 et sa mise en eau était en 2007. C'est un ouvrage en béton de type voute, avec une longueur de 275 m et une hauteur de 83 ,5 m. Il comporte deux déversoirs seuil libres dont les débits sont 6400 m³/s et 2000 m³/s respectivement. La capacité de stockage initiale du réservoir de Tichy-Haf est de 80 hm³ avec un apport moyen annuels de l'ordre de 183 hm³, et un taux d'envasement annuel de 4,70 hm³/an. Le bassin versant du réservoir de Tichy-Haf couvre une superficie d'environ 3980 km². Le barrage de Tichy Haf a pour objectifs d'alimenter en eau potable et eau industrielle les agglomérations situées sur le couloir Akbou-Béjaïa, irriguer les périmètres du Sahel et la haute Soummam et écrêter les crues de l'oued Boussellam.

Caractéristiques hydrologiques

Oued : Boussellam

Capacité initiale : 80,00 hm³

Apport moyen annuel : 183,00 hm³

Envasement annuel : 4,70 hm³/an

Surface du bassin versant : 3980 km²

Caractéristiques du barrage

Barrage Voutes : Béton

Digue de Col : Béton compacté roulé

Hauteur : 83 ,5 m

Longueur : 275 m

Cote retenue normal (R.N) : 294.5 m

Cote plus hautes eaux (P.H.E) : 302,5 m

Déversoirs (02) : seuil libre

Débits : 6400 m³/s et 2000 m³/s

Vidange de fond : 160 m³/s



Figure 27 Situation géographique du barrage Tichy-Haf

III.7.1.3.2. Données de la qualité de l'eau utilisées

Le tableau 4 résume les statistiques de base des variables de qualité de l'eau mesurées pour le réservoir de Tichy Haf. Dans ce tableau, max, min, et CV désignent respectivement le maximum, le minimum et le coefficient de variation.

Tableau 4 Statistiques des variables de la qualité de l'eau pour la retenue de barrage Tichy-Haf

	Moyenne	Ecart type	Min	Max	CV %
Conductivité (dS/m)	1294.3	137.1668	1040	1510	10.59
Résidu sec (mg/l)	855.683	80.0454	750	990	9.35
DCO (mgO₂/l)	7.0154	3.27	1.8	13.2	46.61
NH₄ (mg/l)	0.1285	0.2116	0.01	0.82	164.66
NO₃ (mg/l)	2.2915	1.4171	0.42	4.99	61.84
NO₂ (mg/l)	0.0323	0.0273	0	0.07	84.52
TIN (mg/l)	2.4523	1.4535	0.52	5.085	59.27
Oxygène dissous (mg/l)	8.1877	1.4039	5.4	10.7	17.14
Saturation en oxygène%	89.4543	17.9613	60.25	114.19	20.07
pH	7.7385	0.3070	7.4	8.3	3.93
P total (mg/l)	0.2777	0.1513	0.03	0.56	54.48
PO₄⁻³ (mg/l)	0.1738	0.0883	0.07	0.32	50.80
TC	22.0615	6.8083	11.2	30.4	30.86
Turb	1.8846	1.1157	0.4	4.3	59.20
SD (m)	5.2277	3.4531	1.209	11.485	66.05
MO (mg/l)	3.4385	0.8461	1.8	4.3	24.60
Chl-a (µg/l)	1.265	1.7042	0.13	4.85	134.71

Les variables de la qualité de l'eau pour le barrage de Tichy Haf sont les mêmes que ceux du barrage Lekhal.

La concentration en chlorophylle-a a enregistré une large variation avec des valeurs oscillant entre 0.13 µg/l et 4.85 µg/l, et un coefficient de variation de 134.71%.

Le pH a montré la moindre variation avec un coefficient de variation de l'ordre de 3.93%, cela est dû au pouvoir tampon de l'eau. Le résidu sec, la conductivité électrique, la température de l'eau et la matière organique ont également enregistré une faible variation.

La profondeur de Secchi (SD) est influencée par le contenu des matières en suspension dans la colonne d'eau. Les algues et les fines particules générées par l'érosion des sols du bassin versant sont les principales causes de la variation de la profondeur de disque de Secchi. La

valeur moyenne du SD était de 5.22 m (max: 11.48 m, min: 1.2m). Le phosphore total et le phosphate ont montré une variation moyenne avec des coefficients de variation de 54.48 % et 50.80 % respectivement. Les différentes formes de l'azote inorganique (NH_4 , NO_2 , NO_3) enregistrent une large variation avec des coefficients de variation de 164.66 %, 84.52 % et 61.84% respectivement.

III.7.1.4. Barrage de Keddara

III.7.1.4.1. Situation géographique du barrage

Le barrage de Keddara ($36^\circ 42' 46''$ nord, $3^\circ 32' 50''$ Est) est construit durant la période de 1982-1986 à l'est de la ville de Keddara Bouzegza, et à 8 Km au Sud de la ville de Boudouaou, Wilaya de Boumerdès. Le barrage se trouve à 300 m en aval du point de confluence d'oued Keddara et oued d'El Haad. C'est un ouvrage en terre dont la hauteur de la digue est de 106 m sur une longueur de 468m, il comporte un déversoir labyrinthe pouvant débiter 750 m³/s et une vidange de fond. La superficie de la retenue à sa côte normale est de 5,2 x 106 m², sa profondeur moyenne est de 28 m, la profondeur maximale au niveau du barrage est de 150 m. Le remplissage de la retenue dépend du pompage des eaux de la prise de Beni-Amrane, des eaux de dérivation du barrage de Hamiz et des apports de son propre bassin versant. Les eaux de ce barrage sont destinées à l'alimentation en eau potable de la ville d'Alger. Le bassin versant de l'oued BOUDOUAOU couvre une superficie de 93 km² et dont la forme est presque arrondie, caractérisé par une topographie très raide. Le type de climat est méditerranéen, humide et doux en hiver et chaud et sec en été. Le bassin reçoit une moyenne interannuelle de 880 mm de pluie.

Caractéristique hydrologiques

Oued : Boudouaou

Capacité initiale : 145,6 hm³

Capacité de dernier levé (2004) : 142,32hm³

Apport moyen annuel : 32,3 hm³/an

Envasement : 0,05 hm³/an

Surface du bassin versant : 93 Km²

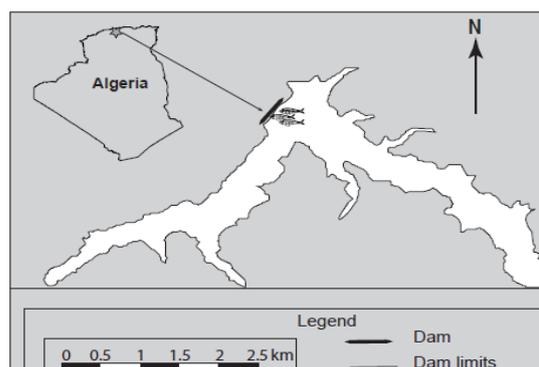


Figure 28 Situation géographique du barrage Keddara

Caractéristiques du barrage

Type : Terre

Hauteur : 106,00 m

Longueur : 468,00 m

Côte de retenue normale (R.N) : 145,00m

Côte plus hautes eaux (P.H.E) : 147,32 m

Déversoir : Labyrinthe : 750 m³/s. Vidange de fond: 55 m³/s

III.7.1.4.2. Données de la qualité de l'eau utilisées

Les données mensuelles sur la qualité de l'eau du barrage de Keddara ont été obtenues auprès de l'Agence Nationale des Ressources Hydriques (ANRH). Des échantillons d'eau ont été généralement prélevés une fois par mois à partir d'un point d'échantillonnage situé sur la retenue de barrage. Les statistiques de base des variables de qualité de l'eau mesurées pour le réservoir de Keddara sont illustrées dans le tableau 5.

Tableau 5 Statistiques des variables de la qualité de l'eau pour la retenue de barrage Keddara

	Moyenne	Ecart type	Min	Max	CV %
Conductivité (dS/m)	1046.6	164.1031	924	1457	15.67
Résidu sec (mg/l)	617.6667	102.0642	400	810	16.52
DCO (mgO₂/l)	8.78	5.7350	0.7	27	65.31
NH₄ (mg/l)	0.0449	0.0739	0	0.27	164.58
NO₃ (mg/l)	2.1640	1.9891	0	6.9	91.91
NO₂ (mg/l)	0.0487	0.0750	0	0.2710	154
TIN (mg/l)	2.2576	2.0384	0.0280	6.9920	90.29
Oxygène dissous (mg/l)	8.3133	1.3453	6.5	10.2	16.18
Saturation en oxygène%	99.5260	17.2806	77.2	137.7	17.36
pH	7.5667	1.9646	0.5	8.6	25.96
P total (mg/l)	0.3202	0.5185	0.0850	2.157	161.93
PO₄⁻³ (mg/l)	0.2259	0.5010	0	2	221.77
TC_{eau} (C°)	22.7333	4.3460	13.5	28	19.11
Turb (NTU)	3.0526	1.8527	1.5	8.3	60.69
SD (m)	2.4662	1.1959	0.8	5	48.49
MO (mg/l)	4.2821	1.0155	3.1	6	23.71
Chl-a (µg/l)	4.4457	2.2405	1.1609	7.78	50.39

La concentration en chlorophylle-a montre une variation moyenne avec des valeurs oscillant entre 1.1609 µg/l et 7.78 µg/l, et un coefficient de variation de 50.39 %.

Les concentrations de l'ammonium (NH₄) variaient sur trois ordres de grandeur, avec des valeurs minimales et maximales de 0 et de 0.27 mg /L et une moyenne de 0.0449 mg / L. le coefficient de variation atteint la valeur de 164.58%.

Les concentrations de nitrate varient entre 0 et 6.9 mg/L avec un coefficient de variation de 91.91 %.

Les valeurs de phosphate et de phosphore total ont enregistré une large variation avec des coefficients de variation de 119.88 % et 81.59 % respectivement.

La conductivité électrique, le résidu sec, la température de l'eau, la matière organique et le pH et ont montré une faible variation avec des coefficients de variations de 15.67 %, 16.52 %, 23.71% et 25.96% respectivement.

III.7.2. Etude de cas n° 02 : barrage de Ouizert

III.7.2.1. Situation géographique du barrage de Ouizert

Le barrage de Ouizert est situé en aval de l'Oued Sahouat dans le Nord-Ouest de l'Algérie, à environ 35 km au Sud-Ouest de la Wilaya de Mascara, il fait partie de triplex formé par les barrages (Ouizert-Bouhanifia-Fergoug). L'oued Sahouet est constitué de la jonction des oueds de Taira et de Saida et il est l'un des principaux affluents de l'oued El Hammam, sur lequel ont été édifiés les barrages de Bouhanifia et de Fergoug (**Benfeta et Remini, 2008**).

Les coordonnées de ce barrage sont X : 250.55 ; Y : 206.80 ; Z : 396.00 NGA. Initialement ce réservoir est construit pour stocker un volume d'environ $100 \times 10^6 \text{ m}^3$ destiné à l'alimentation en eau potable de la ville d'Oran et d'Arzew, l'industrie et à l'irrigation du périmètre d'El Habra situé au Nord de Mohammadia. L'apport annuel est d'environ $84 \times 10^6 \text{ m}^3$, le bassin versant est de 2100 km^2 . Le type de climat est semi-aride, humide et froid en hiver et chaud et sec en été. Le bassin reçoit entre 144 et 300 mm de pluie par an avec une moyenne interannuelle de 282 mm. Différentes sources de pollution, notamment les eaux usées domestiques, les engrais et les rejets industriels, ont été identifiées, **Hamlet (2014)** a reporté que la qualité de l'eau de barrage de Ouizert est menacée par ces rejets industriels.

Caractéristiques hydrologiques

Oued : Sahouat

Capacité initiale : 100 hm^3 Apport annuel : 84 hm^3 Envasement : $0,3 \text{ hm}^3/\text{an}$ Surface du bassin versant : 2100 Km^2 **Caractéristiques du barrage**

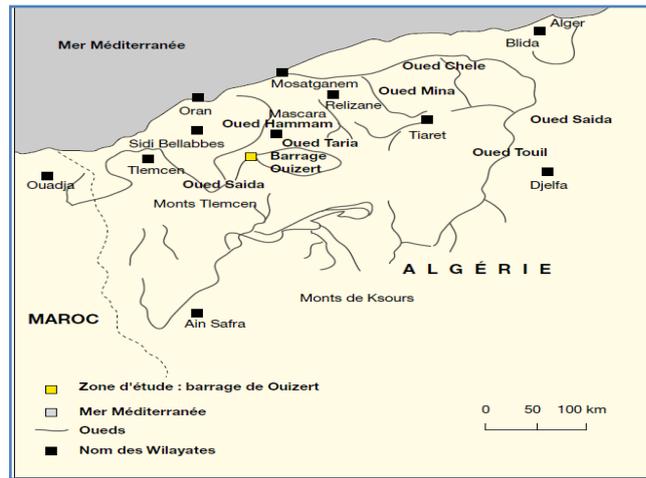
Type : Terre

Hauteur : 60m

Longueur : 950m

Côte retenue normale (R.N) : 448 m

Côte plus Hautes Eaux (P.H.E) : 449.8 m

Déversoir vanne : $4 \times 600 \text{ m}^3/\text{s}$ Vidange de fond : $100 \text{ m}^3/\text{s}$ **Figure 29** Situation géographique du barrage Ouizert**III.7.2.2. Données de la qualité de l'eau utilisées**

Les données sur la qualité de l'eau utilisées dans ce chapitre s'étalent de 2003 à 2013, obtenues auprès de l'Agence nationale des barrages et des transferts (ANBT).

Tableau 6 Statistiques des variables de la qualité de l'eau pour le barrage Ouizert

	Moyenne	Ecart type	Min	Max	CV %
Résidu sec (mg/l)	891.7460	208.9243	480	1360	23.43
DBO₅ (mgO₂/l)	9.6504	6.8715	1.3000	60.7	71.20
DCO (mgO₂/l)	42.1667	23.3343	10	130	55.33
NH₄ (mg/l)	0.3010	0.4226	0	3.15	140.39
NO₃ (mg/l)	8.9683	6.2494	1	31	69.68
NO₂ (mg/l)	0.3571	0.3500	0.0100	2.4	98.01
TIN (mg/l)	9.4970	6.4577	1.18	32.85	67.99
Oxygène dissous (mg/l)	8.4594	2.3720	4.5950	15.852	28.03
Saturation en oxygène%	94.4188	27.2649	50	179.4	28.87
pH	7.9756	0.3422	7.1000	9.05	4.29
P total (mg/l)	0.2302	0.2080	0.0400	1.813	90.35
PO₄⁻³ (mg/l)	0.1361	0.1521	0	1.3	111.75
MO (mg/l)	7.2544	2.8308	0	25	39.02

Les variables de la qualité de l'eau utilisés sont : l'azote inorganique total (TIN), L'ammonium (NH₄⁺) le nitrate (NO₃), le nitrite (NO₂) la demande chimique en oxygène (DCO), la DBO₅, le résidu sec (RS), l'oxygène dissous (O₂), le taux de saturation en oxygène (O₂%) et le phosphate (PO₄⁻²), le phosphore total (Ptot), le pH, la matière organique (MO). L'azote inorganique total (TIN) a été calculé comme $TIN = NH_4^+ + NO_3^- + NO_2^-$.

Les statistiques de base des variables de qualité de l'eau mesurées dans le réservoir de Ouizert sont illustrées dans le tableau 6. Dans ce tableau, max, min, et CV désignent respectivement le maximum, le minimum et le coefficient de variation.

Le tableau 6 montre que certaines variables de qualité de l'eau enregistrent une variations élevée entre les échantillons, avec un coefficient de variation de 140,39%, 111,75%, 98,01%, 90,35% pour NH_4 , PO_4^{-3} , NO_2^- et P_{total} , respectivement. Une raison derrière cette large variation de ces paramètres peut être due à la nature et le type des sources de pollution ponctuelles et non ponctuelles, la saisonnalité des événements hydrologiques, la longue période de mesure (dix ans), et le phénomène de l'évapoconcentration en été.

IV. Résultats et discussions

IV.1. Introduction

La relation entre la charge en nutriments et l'eutrophication est compliquée par la diversité de facteurs impliqués qui sont de nature physico-chimiques, biologiques et hydrodynamique.

La complexité de la relation nutriments-eutrophication (ou encore la relation éléments nutritifs-production primaire phytoplanctonique) peut être caractérisée par plusieurs éléments à savoir: la diversité des espèces d'algues, leurs comportements, leurs cycles saisonniers différents d'une espèce à l'autre, l'interaction des espèces avec leur environnement, les interactions entre les espèces, la non-linéarité de l'écosystème aquatique (**Recknagel, 1997; Lee et al., 2003**), l'insuffisance de compréhension des processus, la multiplicité des mécanismes impliqués, le manque de données de bonne qualité (**Chen et Mynett, 2004**), et la haute variabilité saisonnière des charges polluantes déversées dans les retenues de barrages. Tous ces éléments rendent le développement de modèles prédisant la qualité des eaux et l'eutrophication des retenues de barrages un défi.

Au cours des dernières années, l'émergence de techniques d'intelligence artificielle dans le domaine de modélisation des ressources en eau et la qualité de l'eau est devenue plus utile et plus populaire.

Les réseaux de neurones artificiels sont composés de simple éléments appelés neurones opérant en parallèle, peuvent résoudre de nombreux problèmes compliqués à peu près dans tous les domaines des sciences et technologies notamment dans la modélisation de qualité de l'eau.

Les systèmes d'inférence neuro-flous (ANFIS) est un système hybride intégrant, à la fois, la capacité d'apprentissage des réseaux de neurones et l'interprétabilité linguistique de la logique floue dans un cadre unique. Ils sont également utilisés dans la modélisation de qualité de l'eau.

Par le biais des modèles de réseaux de neurones artificiels et d'ANFIS, le phénomène de l'eutrophication est réduit en un problème de type "entrées-sortie" (input-output problem), par conséquent, ce phénomène va être traité selon cette logique par ces techniques de modélisation.

Nous rappelons que les objectifs principaux de ce chapitre sont de (1) développer des modèles basés sur les RNAs, la logique floue et système d'inférence adaptative neuro-floue (ANFIS) capables de simuler les indicateurs de l'eutrophication dans certains barrages au

Nord de l'Algérie et (2) démontrer la capacité et l'applicabilité de ces techniques dans les questions pratiques de la gestion de la qualité de l'eau et l'eutrophisation (3) choisir la combinaison optimale de variables d'entrée pour améliorer la performance des modèles développés, (4) Comparer les modèles développés et choisir le plus performant.

Nous avons divisé cette section en deux chapitres (étude de cas):

- ✓ Première étude de cas: Simulation de la de chlorophylle-a par les réseaux de neurones artificiels et les modèles neuro-flous (ANFIS), dans lequel nous avons utilisé une base de données mixte provenant de l'assemblage de données relatives aux barrages Lekhal, Tichy-Haf et Keddara dans le but de remédier le problème de l'insuffisance de données.
- ✓ Deuxième étude de cas est liée au barrage de Ouizert, dans lequel nous avons simulé l'indice de l'état trophique, le phosphore total et la DBO₅ par les par les réseaux de neurones artificiels et les modèles neuro-flous (ANFIS).

IV.2. Modélisation de l'eutrophisation par les réseaux de neurones artificiels

IV.2.1. Première étude de cas: Simulation de la chlorophylle-a par les réseaux de neurones artificiels

IV.2.1.2. Développement du modèle

IV.2.1.2.1. Variables d'entrées (variables candidats)

L'eutrophication et l'efflorescence algale associée sont dues à de nombreux facteurs d'ordre physiques, chimiques, biologiques et hydrodynamiques. Pour simuler l'eutrophication avec succès tous ces facteurs doivent être pris en compte directement ou indirectement.

Les variables de la qualité de l'eau qu'on a utilisé comme étant variables d'entrée sont : la température de l'eau, le pH, l'oxygène dissous (O₂) (mg/l), le taux de saturation en oxygène (%), la visibilité de l'eau représenté par la profondeur de disque de Secchi (m), la turbidité (NTU), la conductivité électrique (ds/m), le résidu sec (mg/l), l'ammonium (NH₄⁺), le nitrate (NO₃⁻²), le nitrite (NO₂⁻) l'azote inorganique total (TIN) le phosphore total (PT), le phosphate (PO₄⁻³) la demande chimique en oxygène (DCO), la matière organique (MO).

IV.2.1.2.2. La sortie cible: la concentration en chlorophylle alpha (Chl-a)

Il existe de fortes relations entre le niveau de qualité de l'eau, la concentration de pigment chlorophyllien alpha (Chlorophylle-a), l'eutrophisation et son efflorescence algale associée.

Dimberg et al. (2012) indique que la chlorophylle-a (Chl-a) est utilisée dans de nombreux programmes de surveillance environnementale comme indicateur de qualité de l'eau pour les lacs et les réservoirs, car elle est commune à presque tous les organismes photosynthétiques. Largement utilisé pour estimer la biomasse algale dans l'écosystème aquatique pour déterminer la qualité de l'eau et des études écologiques. C'est aussi l'un des indicateurs les plus importants de l'existence et du degré d'eutrophisation dans les plan d'eau, par conséquent, la modélisation de l'eutrophisation se réfère à la simulation des concentrations de chlorophylle-a, elle est utilisée comme la sortie des modèles.

L'application de techniques de modélisation (data driven) devrait permettre de capturer les relations significatives entre la chlorophylle-a et les variables de qualité de l'eau choisies.

IV.2.1.2.3. Apprentissage

L'apprentissage utilisé est de type supervisé, l'Algorithme de rétro propagation de l'erreur (backpropagation), variante de Livenberg-Marquardt en raison de sa rapidité de convergence et ses performances de généralisation. Il consiste en la modification des poids des connexions entre les différentes couches par rapport à une erreur calculée en sortie.

IV.2.1.2.4. La sélection des variables d'entrées

La sélection des variables d'entrées est une tâche délicate et fastidieuse, elle constitue une étape très importante dans la modélisation de n'importe quel phénomène ou processus.

Le choix de variables d'entrée basés sur l'analyse de corrélation statistique est la technique analytique la plus populaire pour sélectionner les entrées. L'inconvénient de cette technique est qu'elle est capable seulement de capturer la dépendance linéaire entre deux variables. Par conséquent, elle peut entraîner l'omission de variables d'entrées importantes liés à la sortie de façon non linéaire, pour cela on fait appel à l'analyse de sensibilité par le biais des RNAs, celle-ci permet de dévoiler les liaisons de type linéaire et non-linéaire entre les entrées et la sortie correspondante.

Pour évaluer l'effet des variables d'entrée sur le modèle, deux processus d'évaluation ont été utilisés. Premièrement, une connaissance à priori soutenue par l'analyse de corrélation statistique. Le deuxième processus d'évaluation est fondé sur l'analyse de sensibilité des RNAs.

IV.2.1.2.4.1. La matrice de corrélations

Tableau 7 Matrice de corrélations

	CE	RS	DCO	NH4	NO3	NO2	TIN	O2	O2p%	Ph	Ptot	PO4	TC	Turb	SD	MO	Chla
CE	1																
RS	0,79	1															
DCO	-0,46	-0,40	1														
NH4	-0,13	-0,11	0,15	1													
NO3	-0,22	-0,20	0,12	0,03	1												
NO2	-0,50	-0,40	0,49	0,13	0,58	1											
TIN	-0,27	-0,24	0,18	0,19	0,98	0,65	1										
O2	-0,16	-0,01	0,17	-0,01	0,32	0,10	0,30	1									
O2p %	-0,16	-0,08	0,17	-0,25	0,27	0,06	0,22	0,68	1								
Ph	-0,40	-0,31	0,22	-0,32	-0,13	0,07	-0,16	-0,05	0,15	1							
Ptot	-0,07	-0,07	-0,05	-0,12	-0,11	-0,02	-0,13	-0,21	-0,11	0,066	1						
PO4	-0,02	0,03	-0,11	-0,08	-0,19	-0,09	-0,20	-0,24	-0,13	0,024	0,95	1					
TC	0,12	0,06	-0,15	-0,40	-0,17	-0,21	-0,23	0,02	0,41	0,267	0,11	0,12	1				
Turb	-0,48	-0,35	0,33	0,18	0,17	0,52	0,24	0,05	-0,12	0,290	-0,04	-0,03	-0,44	1			
SD	0,57	0,57	-0,31	-0,03	-0,12	-0,33	-0,14	-0,14	-0,08	-0,201	0,01	0,09	0,34	-0,45	1		
MO	-0,71	-0,57	0,55	0,21	0,13	0,49	0,20	0,04	0,06	0,263	0,20	0,14	-0,17	0,48	-0,42	1	
Chla	-0,48	-0,41	0,26	0,53	0,12	0,31	0,22	0,22	0,28	0,015	-0,11	-0,13	-0,10	0,25	-0,40	0,50	1

On enregistre une forte corrélation entre la concentration en chlorophylle-a d'une part et l'ammonium et la matière organique d'autre part (puisque le NH₄ est la forme préférée de l'azote consommée par les algues (**Jung, 2009**)). Une corrélation moyenne négative existe entre la chlorophylle-a d'une part et le résidu sec et la conductivité électrique d'autre part, parce que la salinité affecte négativement le développement des algues d'eau douce. Une corrélation modérée enregistrée entre la DCO, Le NO₃, L'azote inorganique total, l'oxygène dissous et la turbidité d'un côté et la concentration en chlorophylle-a de l'autre. L'analyse de sensibilité va dévoiler les associations de type non linéaires.

IV.2.1.2.4.2. Analyse de sensibilité

La méthode de l'analyse de sensibilité choisie est dite 'leave-one-out' ou «éliminer une seule variables de l'ensemble de variables candidates», elle fait appel à deux indicateurs: *le ratio* de l'erreur du réseau et le *rang*. Le ratio des erreurs de réseau indique les effets sur le fonctionnement du réseau lors de l'élimination d'une variable d'entrée. C'est le ratio de l'erreur obtenue après avoir éliminé (élagué) une seule variable à l'erreur ait obtenu en

utilisant toutes les variables, c.-à-d le rapport entre l'erreur du modèle élagué et celle du modèle complet (modèle utilisant toutes les variables d'entrées). Le ratio est calculé par:

$$Ratio = \frac{RMSE_{\text{modèle élagué}}}{RMSE_{\text{modèle complet}}} \quad (41)$$

Ce ratio prend trois cas :

Si le **ratio = 1** : cela signifie que le modèle n'a pas enregistré ni détérioration, ni amélioration de la qualité de simulation après la suppression de la variable de la qualité de l'eau en question, donc son inclusion ou exclusion de la base de données est pareille.

Si le **ratio > 1** : cela traduit une détérioration de la qualité de simulation après avoir supprimé la variable de la qualité de l'eau en question, puisque l'élimination de variable la plus importante entraîne une valeur du RMSE plus élevée, donc un ratio plus élevé, ce qui indique que le réseau est davantage affecté lorsque ce variable n'est pas incluse (**Lee et al., 2003**). On peut conclure que plus le ratio est plus grand, plus la variable est plus importante, et que l'on doit prendre en compte lors de la création du modèle optimal.

Si le **ratio < 1** : cela indique que le modèle a connu une amélioration de la qualité de simulation après avoir supprimé la variable de la qualité de l'eau, donc plus le ratio est petit, plus la variable en question est moins important. Quand le ratio est inférieur à 1.0, la variable peut être rejetée et supprimée du modèle (**Olszewski et al., 2008**).

Tableau 8 Ordre d'influence des variables de la qualité de l'eau sur la concentration en chlorophylle-a

Variables	Ratio	Rang	Variables	Ratio	Rang
P _{total} (mg /L)	1,6685	1	NH ₄ ⁺ (mg /L)	1,3292	9
PO ₄ ⁻³ (mg /L)	1,6460	2	SD (m)	1,3207	10
RS (mg /L)	1,6408	3	MO (mg /L)	1,3188	11
pH (-)	1,6399	4	NO ₃ ⁻ (mg /L)	1,2709	12
O ₂ (mg /L)	1,5218	5	CE (ds/m)	1,2525	13
O ₂ % (%)	1,4357	6	TC _(eau) ^o (C°)	1,1428	14
NO ₂ ⁻ (mg /L)	1,4238	7	TIN (mg /L)	1,0995	15
DCO (mg O ₂ /L)	1,3371	8	Turbidité (NTU)	0,9566	16

L'analyse de sensibilité des variables de la qualité de l'eau dans des phases de validation (le tableau 8) en utilisant les RNAs indique que le phosphore total, est la variable la plus importante suivie par le phosphate PO₄⁻³ puisque le phosphore est premier facteur limitant de la production de la biomasse algale dans les écosystèmes aquatiques d'eau douce (**Pinay et**

al., 2017; Jung, 2009). Malgré que la matrice de corrélation (tableau 7) indique une faible corrélation linéaire entre la concentration en chlorophylle d'une part et le phosphore totale et le phosphate d'autre part, les RNAs ont permis de capturer une relation non linéaire entre les phosphore et la biomasse algale exprimée en terme de chlorophylle-a.

Les autres variables importantes selon leur ordre sont : le résidu sec (RS), le pH, la concentration en oxygène dissous (O_2), le pourcentage de saturation en oxygène ($O_2\%$), le nitrite (NO_2^-), la DCO, l'ammonium (NH_4^+), la visibilité de l'eau (SD), la matière organique (MO), le nitrate (NO_3^-), conductivité électrique (CE), la température de l'eau (TC°), l'azote inorganique total (TIN), respectivement.

Nous avons identifié le meilleur ensemble de variables d'entrée pour modéliser la chlorophylle-a par le biais des RNAs.

Ji (2008) a révélé que la croissance du phytoplancton est en réalité un processus en deux étapes. La première étape est l'absorption des nutriments et la deuxième étape est la croissance des phytoplanctons. **Jung (2009)** a rapporté que les algues emmagasinent le phosphore dans leurs cellules avant que l'absorption de celui-ci se traduit en termes de production primaire. C'est pour cela que la corrélation entre la chlorophylle-a (indicateur de la biomasse algale) et le phosphore total et l'orthophosphate est faible et négative. Les corrélations linéaires simples n'ont pas pu déterminer la liaison existant entre le phosphore et la biomasse, alors que les RNAs ont pu capturer cette liaison, et les résultats de l'analyse de sensibilité confirme cette conclusion.

IV.2.1.2.5. Le choix du modèle

Le meilleur résultat est obtenu par un réseau de neurones de type PMC à structure: **16-23-1**, c'est à dire un modèle contenant 16 variables d'entrées, 23 neurones dans la couche cachée, et une seule variable de sortie. Les critères de performance sont de 4,067 μ g/l pour le RMSE et de 0.7976 pour le coefficient de corrélation (CC) durant la phase de validation. Pour le modèle avec 23 neurones dans la couche cachée, lorsque 16 variables de qualité de l'eau ont été utilisés comme variables d'entrée du modèle MLP pour la simulation de la chlorophylle-a, l'erreur était moindre que les autres structures neuronales. De plus, les coefficients de corrélation dans l'ensemble des données, lors de l'apprentissage et la validation étaient respectivement de 0,9663, 0,9929 et 0,7976, ce qui montre une conformité raisonnable entre les données réelles de chlorophylle-a et celles obtenues par le modèle.

Les résultats obtenus par ce modèle neuronal étaient supérieurs à ceux enregistrés par **Karul et al. (2000)** où ces chercheurs ont testé la capacité des RNAs pour modéliser la concentration de la chlorophylle-a dans le barrage réservoir du Keban, et dans les lacs de Mogan et d'Eymir, en Turkey.

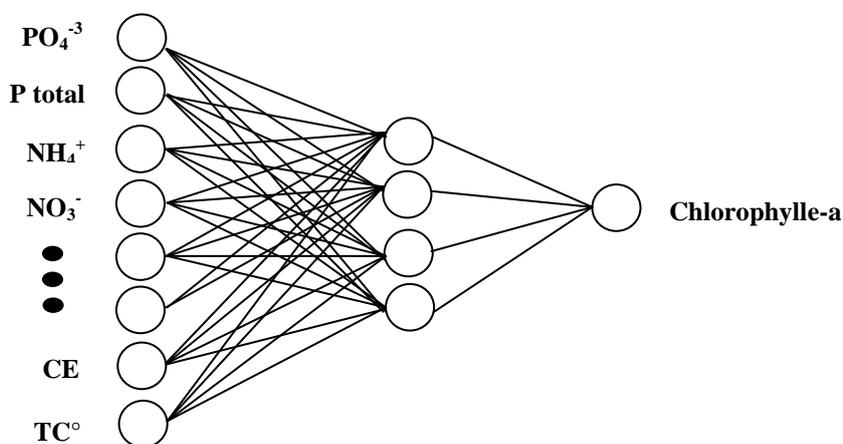


Figure 30 Schéma général du modèle neuronal développé pour estimer la chlorophylle-a

Afin de bien visualiser les résultats de simulation de ce modèle neuronal et de présenter sa performance pour chacun de trois barrages, nous avons tracé le diagramme de dispersion par groupe (*Scatter plot by group*) ci-après:

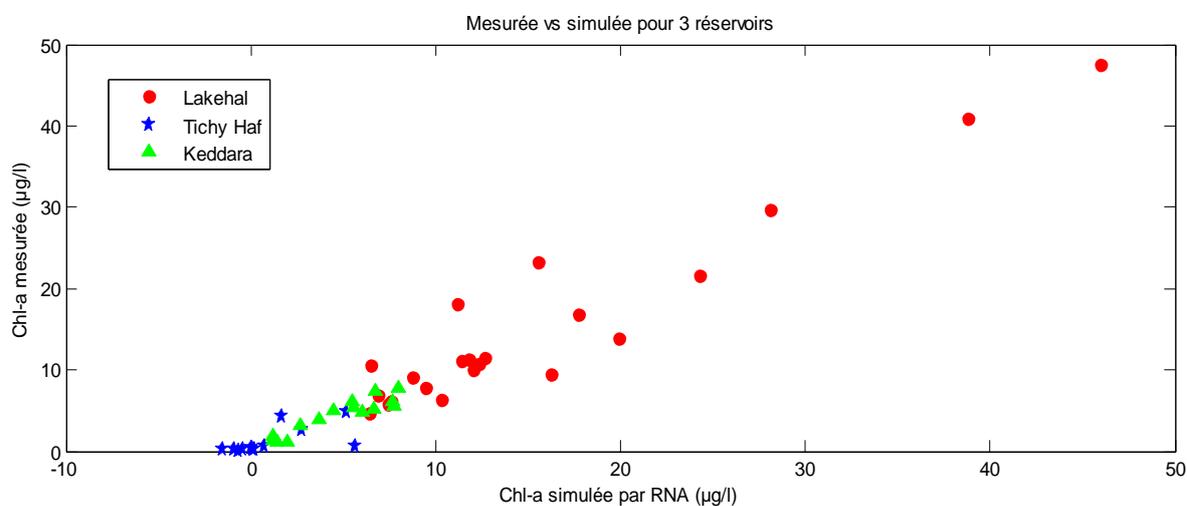


Figure 31 Diagramme de dispersion par groupe de valeurs de chlorophylle-a observées et celles simulées par le RNA pour l'ensemble de données (mélangées)

Il ressort de la figure 31 que le modèle M0 développées ne fait la distinction de l'origine de données, il permet simplement l'approximation d'une fonction existante entre les inputs et l'output sans tenir compte si les données proviennent d'un seul barrage ou de plusieurs barrages, puisque les RNAs sont des approximateurs universels (**Dreyfus, 2004**).

L'apparition de valeurs négatives pour les petites valeurs simulées de la chlorophylle est une indication de l'instabilité du modèle (Pereira et al., 2009), puisque la base de donnée d'apprentissage ne contient pas suffisamment de petites valeurs de la chlorophylle-a permettant au modèle de reproduire celles-ci convenablement, d'où la nécessité d'enrichir l'ensemble de données par de nouvelles valeurs mesurées.

Dans le but de bien illustrer les résultats de modélisation obtenus par le meilleur modèle neuronal (M_0 : 16-23-1), nous avons tracé les figures 32, 33 et 34 qui représentent les résultats de simulation de chlorophylle-a pour tout l'ensemble de données (a1 et b1), l'apprentissage (a2 et b2) et la validation (a3 et b3). L'observation visuelle de ces figures montrent que les valeurs de chlorophylle-a observées et celles simulées par le réseau de neurone (M_0 : 16-23-1) sont proches et les écarts oscillent autour de 0 à 5 $\mu\text{g/l}$.

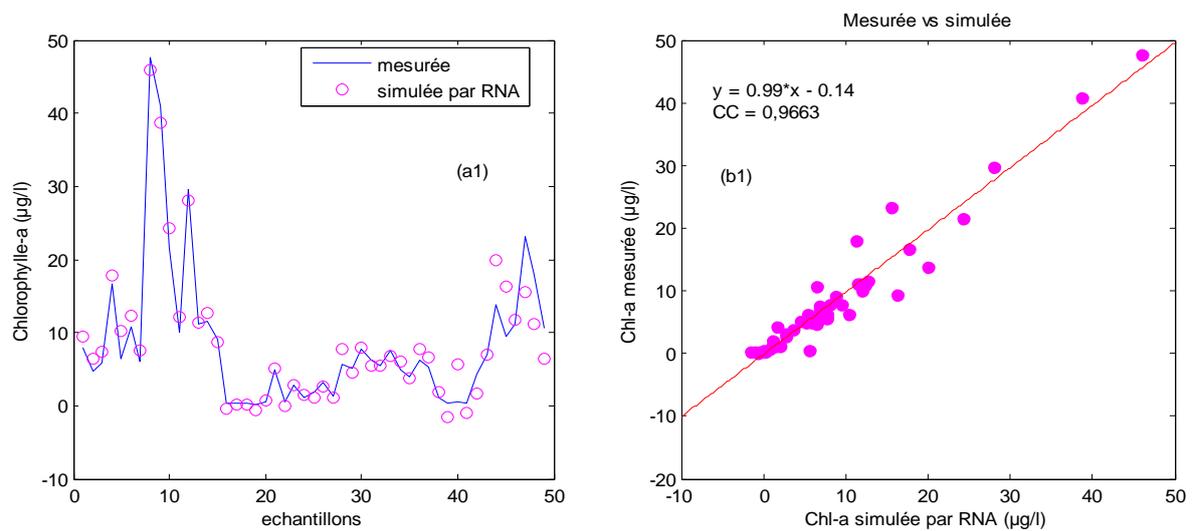


Figure 32 (a1) Valeurs mesurées et simulées de la chlorophylle-a ($\mu\text{g} / \text{L}$); (b1) Diagramme de dispersion des valeurs simulées et mesurées de la chlorophylle-a ($\mu\text{g/L}$) pour l'ensemble de données en utilisant le modèle neuronal M_0 (16-23-1)

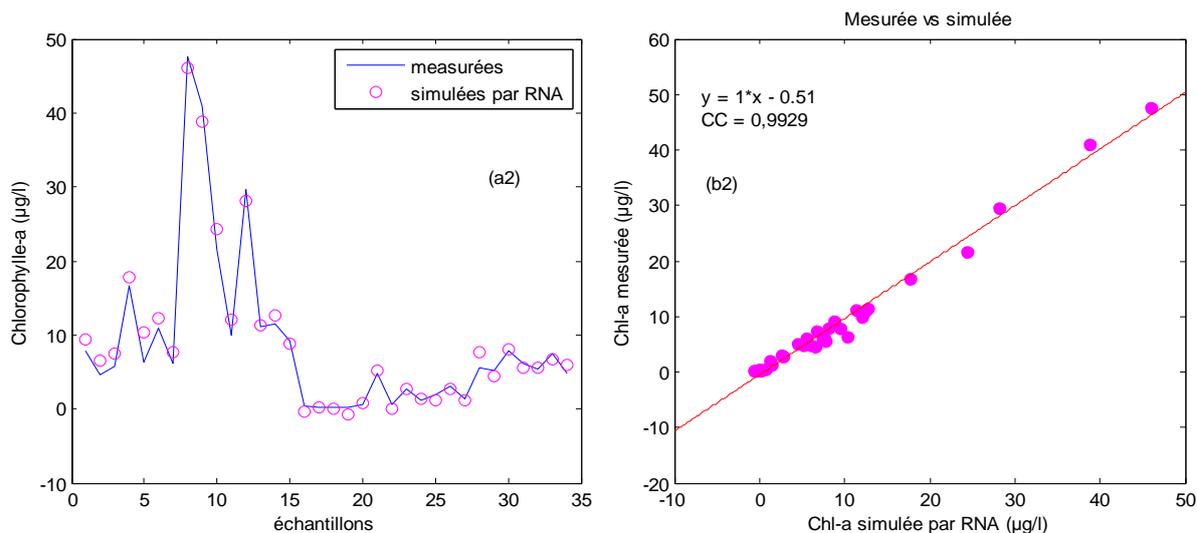


Figure 33 (a2) Valeurs mesurées et simulées de la chlorophylle-a ($\mu\text{g} / \text{L}$); (b2) Diagramme de dispersion des valeurs simulées et mesurées de la chlorophylle-a ($\mu\text{g}/\text{L}$) lors de l'apprentissage pour modèle neuronal **M0 (16-23-1)**.

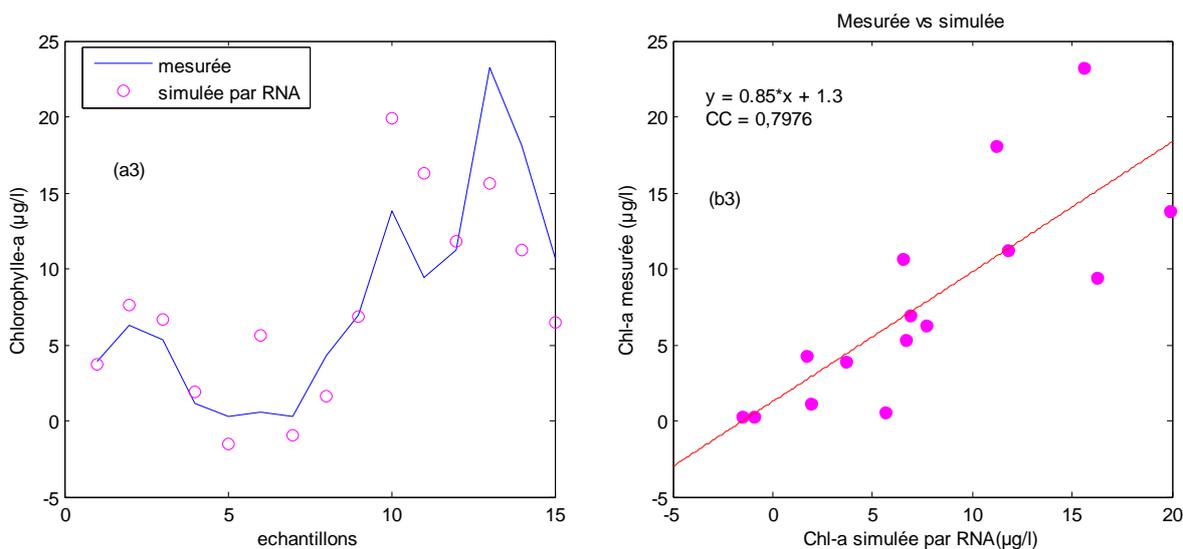


Figure 34 (a3) Valeurs mesurées et simulées de chlorophylle-a ($\mu\text{g} / \text{L}$); (b3) Diagramme de dispersion des valeurs simulées et mesurées de la chlorophylle-a ($\mu\text{g}/\text{L}$) lors de la validation pour le modèle neuronal **M0 (16-23-1)**.

IV.2.1.2.6. Amélioration du modèle développé

Les résultats obtenus par l'application des réseaux de neurones en utilisant toutes les variables d'entrées étaient satisfaisants, mais le fait d'inclure toutes les variables d'entrée ne permet pas toujours d'obtenir le meilleur modèle de RNA en raison de la multicollinéarité, c'est-à-dire que deux variables corrélées apportent à peu près la même information au modèle, dans ce cas on peut utiliser l'une ou l'autre.

On s'est posé alors la question de savoir si l'exclusion d'une ou plusieurs variables de la base de données, ne permettrait pas l'amélioration des performances du modèle.

Des tentatives d'améliorations ont été réalisées en faisant exclure à chaque fois à l'entrée du réseau, un ou plusieurs variable d'entrée, en se basant sur la matrice de corrélation et l'analyse de sensibilité, et la facilité de mesurer la variable d'entrée et on a suivi l'évolution des performances du modèle élaboré

Le tableau 9 résume les valeurs de critères de performance en calage et en validation (R^2 et RMSE respectivement) pour les modèles retenus pour différentes combinaisons de variables d'entrées.

Ce tableau montre clairement que les modèles apportent de bonne performance durant l'apprentissage (calage), alors qu'en validation on voit une certaine détérioration de performance des modèles, mais en générale ils apportent des résultats relativement acceptables.

En général, les réseaux neurones de type PMC est un outil approprié pour analyser la qualité de l'eau de surface et les commentaires liés à l'étude de cas sont comme suit:

Malgré l'insuffisance de données historiques pour l'apprentissage, et leurs origines diversifiées (données provenant de 3 différents barrages), la performance de modélisation à l'aide de 35 jeux de données d'apprentissage (35 échantillons pour l'apprentissage) et de 14 jeux de données de validation est très raisonnable.

A l'exception de modèle M_6 -qui montre une grande détérioration de ses performance lors de la validation ($R = 0.50$)-, les modèles de RNAs apportent des résultats satisfaisants.

Les résultats obtenus par les modèles RNAs sélectionnés en gras (tableau) donnent de résultats similaires à ceux fournis par **Karul et al. (1999)**; par **Kuo et al. (2007)**; par **Huo et al. (2013)** où ces chercheurs ont modélisé la concentration en chlorophylle-a dans différents barrages par le biais des RNAs. Ces résultats confirment, encore une fois, la robustesse, et l'efficacité des modèles de RNAs développés.

Les modèles RNAs développés sont capables de simuler assez bien la dynamique de concentration du phytoplancton (exprimé en termes de Chl-a) en utilisant des variables d'entrée minimales même si des facteurs inconnus affectaient la qualité de l'eau de barrages et des séries de données limitées. Les résultats montrent que le comportement complexe dans le processus d'eutrophisation pourrait être modélisé en utilisant la technique de RNA, et nous avons estimé avec succès quelques valeurs extrêmes à partir des ensembles de données de validation qui n'ont pas été utilisés dans l'apprentissage du réseau neuronal. Le modèle développé peut être utilisé pour (1) estimer la concentration de Chl-a lorsque la valeur réelle ne peut être obtenue (lacunes), (2) estimer les données interpolées entre deux échantillons consécutifs, et (3) simuler différents scénarios de qualité de l'eau pour les extrêmes variables

de sortie. Cette approche de modélisation de l'eutrophisation est utile pour simuler les niveaux de chlorophylle-a à n'importe quel endroit ou moment (t) dans le domaine d'intérêt.

Tableau 9 Performances des modèles élaborés pour modéliser la chlorophylle-a

Modèles	Variables d'entrée	Choix des entrées	Structure du modèle	R		RMSE ($\mu\text{g/l}$)	
				apprentissage	validation	apprentissage	validation
M₀** (modèle de base)	CE RS DCO NH ₄ NO ₃ NO ₂ TIN O ₂ O _{2p} Ph P _{tot} PO ₄ TC Turb SD MO.	Tous les variables d'entrées disponibles	16-23-1	0,9929	0,7976	1,3493	4,067
M₁**	CE NH ₄ NO ₂ O _{2p} SD MO P _{tot} PO ₄	Matrice de corrélation plus composées phosphatée	8-5-1	0,9967	0,8358	0,8772	4,4260
M₂**	CE NH ₄ NO ₂ NO ₃ TIN O _{2p} SD MO	Matrice de corrélation plus composées azotées	8-3-1	0,9714	0,7221	2,5711	5,6294
M₃**	CE NH ₄ NO ₂ O _{2p} SD MO	Matrice de corrélation	6-5-1	0,9949	0,7074	1,0934	4,7272
M₄**	CE RS Ph TC Turb SD	Paramètres physiques seules	6-3-1	0,8063	0,8136	6,4428	4,3463
M₅**	NH ₄ NO ₃ NO ₂ TIN P _{tot} PO ₄	Les nutriments seulement	6-22-1	0,9289	0,7570	4,4402	4,9333
M₆*	CE Ph TC Turb O _{2p}	Paramètres Facilement mesurables	5-8-1	0,9222	0,5026	4,5952	7,6705
M₇*	CE RS Ph TC	Paramètres physiques facilement mesurables	4-10-1	0,7353	0,6817	7,8249	6,2341

CE: conductivité électrique; RS: résidu sec; DCO ; NH₄ : ammoniaque, NO₃: nitrate, NO₂ : nitrite, TIN : azote inorganique totale ; O₂ : oxygène dissous ; O_{2p}: taux de saturation en oxygène ; pH ; P_{tot}: phosphore total ; PO₄: phosphate ; TC : température de l'eau ; Turb : turbidité ; SD : visibilité de l'eau (m) ; MO : matière organique.

Ces modèles neuronaux fournissent des informations précieuses pour les décideurs afin de prendre les mesures et les actions appropriées dans le but de préserver la qualité de l'eau des barrages. L'étude de cas pour la base de données mélangée démontre une large gamme d'applicabilité des modèles de réseaux de neurones artificiels.

IV.2.2. Deuxième étude de cas: barrage de Ouizert

IV.2.2.1.2. Développement des modèles

Dans ce chapitre, nous avons utilisé l'une des architectures de RNAs les plus courantes, le réseau neuronal de type perceptron multicouches. Cette architecture de RNAs est très populaire, car elle peut être utilisée dans différents domaines, y compris la modélisation de la qualité de l'eau. Nous avons utilisé des RNAs ordinaires, des modèles de RNA à trois couches constitués d'une couche d'entrée, une couche cachée, et une couche de sortie, ont été développés. L'algorithme de Levenberg-Marquardt a été utilisé pour entraîner les réseaux et les fonctions d'activation de 'Logsig' ont été appliquées dans les nœuds de couche cachés, le nombre optimal de neurones dans la couche cachées est obtenue par la méthode de 'l'essai et l'erreur' (trial and error), plusieurs modèles de RNA ont été développés et comparé en utilisant deux critères statistiques de la qualité et de performance de modèle (CC et RMSE) et l'architecture optimale des RNAs ont été retenues.

Dans ce chapitre, l'azote inorganique total (TIN), L'ammonium (NH_4^+) le nitrate (NO_3^-), le nitrite (NO_2^-) la demande chimique en oxygène (DCO), la DBO_5 , le résidu sec (RS), l'oxygène dissous (O_2), le taux de saturation en oxygène ($\text{O}_2\%$) et le phosphate (PO_4^{2-}), le pH, la matière organique (MO) du réservoir Ouizert sont utilisés. L'azote inorganique total (TIN) a été calculé comme $\text{TIN} = \text{NH}_4^+ + \text{NO}_3^- + \text{NO}_2^-$. La qualité de l'ajustement des modèles est évaluée à l'aide des valeurs du coefficient de corrélation (CC), des erreurs quadratiques moyennes (RMSE) obtenues pendant les phases d'apprentissage et de validation. Le logiciel utilisé pour implémenter tous les modèles de RNAs est Matlab v7.4, précisément, la boîte à outils de réseaux de neurones intégrée à Matlab. Un code de programme a été écrit en langage Matlab et utilisé pour développer différents modèles basés sur les RNA.

Tout d'abord, plusieurs architectures de RNAs ont été testées en utilisant ce code tout en faisant varier le nombre de variables d'entrées et le nombre de neurones dans la couche cachée, et les structures de modèles appropriées ont été déterminées pour chaque combinaison de variables d'entrées.

IV.2.2.2. Simulation de phosphore total par les réseaux de neurones artificiels

IV.2.2.2.1. Introduction

Le phosphore et l'azote sont des nutriments essentiels à la croissance des algues, cependant, le phosphore est souvent considéré comme le nutriment qui contrôle la production d'algues dans les lacs, il est couramment utilisé pour estimer l'état trophique des lacs. Pour cette raison, l'état trophique de la qualité de l'eau est classé selon la concentration de phosphore total (PT) dans l'eau comme suit (**Best, 1999**):

Ultra-Oligotrophique, si : $PT < 4 \mu\text{g} / \text{L}$.

Oligotrophique, si : $4 \mu\text{g/L} < PT < 10 \mu\text{g/L}$.

Mésotrophique, si : $10 \mu\text{g/L} \leq PT < 35 \mu\text{g/L}$.

Eutrophique, si : $35 \mu\text{g/L} \leq PT \leq 100 \mu\text{g/L}$.

Hypertrophique, si : $P > 100 \mu\text{g/L}$.

Le phosphore total (TP), la chlorophylle-a (Chl-a) et la profondeur du disque Secchi (SD) sont couramment utilisés comme indicateurs de l'eutrophisation des réservoirs (**Kuo et al., 2007**).

Afin de déterminer les relations non linéaires entre les facteurs de qualité de l'eau et les indicateurs d'eutrophisation (phosphore total dans ce cas), la technique de RNA basée sur un algorithme de rétro propagation de l'erreur a été choisie.

IV.2.2.2.2. Elaboration des modèles

Les modèles de RNAs ont été développés pour simuler les concentrations de phosphore total (PT) mensuelle pour le réservoir Ouizert. L'architecture de perceptron multicouche a été utilisée. Plusieurs scénarios de modèle avec différentes combinaisons de variables d'entrée ont été testés pour simuler les concentrations de phosphore total. Les variables d'entrée ont été optimisées en supprimant ou en ajoutant des variables une par une.

Le nombre de neurone optimal est déterminé par la méthode de 'l'essai et l'erreur' en faisant varier ce nombre de 1 à 30, ensuite on va choisir le meilleur modèle en se basant sur les critères statistiques de performance (R et RMSE).

Pendant la phase d'apprentissage, les modèles de RNAs ont été entraînés avec un ensemble de données d'apprentissage, contenant toutes les valeurs d'entrée ainsi que les valeurs de sortie désirées (output). Le meilleur modèle ayant la valeur minimale du RMSE et la valeur maximale du coefficient de corrélation a ensuite été utilisé pour modéliser les concentrations de phosphore total (PT) dans le réservoir de Ouizert. Plusieurs combinaisons de variables d'entrées, comme les montre le tableau suivant, sont testés.

IV.2.2.2.3. Résultats de simulation

Le tableau 10 présente différents ensembles de variables d'entrée utilisées et les critères de performances pour les modèles de RNAs développés.

Ce tableau révèle que les résultats durant la phase d'apprentissage et de validation pour le modèle M_0 -qui renferme toutes les variables d'entrées disponibles- convergent vers un RMSE égal à 0,0089 et 0,0177, respectivement.

Le tableau 10 montre clairement que les modèles de RNAs dont le phosphate PO_4 fait partie des variables d'entrée (M_0 , M_2 et M_5) apportent de très bons résultats avec des coefficients de corrélation supérieur à 0,97 pendant la phase de validation. Nous pouvons conclure que le phosphate (PO_4) est la principale variable pour la modélisation de phosphore total via les RNAs. Ces trois modèles doivent être retenus afin de simuler la concentration de phosphore total dans le barrage de Ouizert.

Une simple comparaison entre le modèle M_0 et le modèle M_1 montre que l'élimination du phosphate (PO_4) provoque une grande détérioration de performance du modèle M_1 , ce qui confirme, encore une fois, que le phosphate (PO_4) est la variable la plus efficace dans l'expression de la variabilité de phosphore total via les RNAs.

Les modèles M_1 , M_3 et M_4 apportent de mauvais résultats avec des coefficients de corrélations très faibles oscillant entre 0,13 à 0,4 pour la phase de validation, l'une des raisons derrière cette faible performance est que les variables d'entrées de ces modèles sont faiblement corrélés avec le phosphore total, ces modèles ne sont pas pris en compte lors de la phase d'exploitation de modèles.

Tableau 10 Performances des modèles élaborés pour modéliser le phosphore total

Modèles	Variables d'entrée	Choix des entrées	Structure du modèle	R		RMSE (mg/l)	
				Apprentissage	Validation	Apprentissage	Validation
M ₀ (modèle de base)	RS DBO DCO NH ₄ NO ₃ NO ₂ TIN O ₂ O _{2p} Ph PO ₄ MO.	Toutes les variables d'entrées disponibles	<u>12-8-1</u>	0,9993	0,9739	0,0089	0,0177
M ₁	RS DBO DCO NH ₄ NO ₃ NO ₂ TIN O ₂ O _{2p} Ph MO.	Toutes les variables à l'exception de phosphate	11-9-1	0,7249	0,4075	0,1909	0,1238
M ₂	NH ₄ NO ₃ TIN PO ₄	Matrice de corrélation	<u>4-6-1</u>	0,9961	0,9823	0,0209	0,0148
M ₃	RS NH ₄ NO ₃ NO ₂ TIN MO.	Composées azotées Plus MO et RS	6-14-1	0,9065	0,3055	0,1080	0,1355
M ₄	NH ₄ NO ₂ NO ₃ TIN	Composées azotées	4-11-1	0,7534	0,1303	0,1565	0,1151
M ₅	PO ₄	Phosphate seule	<u>1-5-1</u>	0,9958	0,9806	0,0216	0,0152

RS: résidu sec; DCO ; NH₄ : ammoniacque, NO₃: nitrate ; NO₂ : nitrite, TIN : azote inorganique totale ; O₂ : oxygène dissous ; O_{2p} : taux de saturation en oxygène ; pH ; Ptot : phosphore total ; PO₄ : phosphate; MO : matière organique.

En comparant la concentration de phosphore total estimées par le modèle M₀ avec celles observées, les figures 35, 36 et 37 montrent une quasi superposition entre les données observées et celles simulées par les modèle **M0 (12-8-1)**.

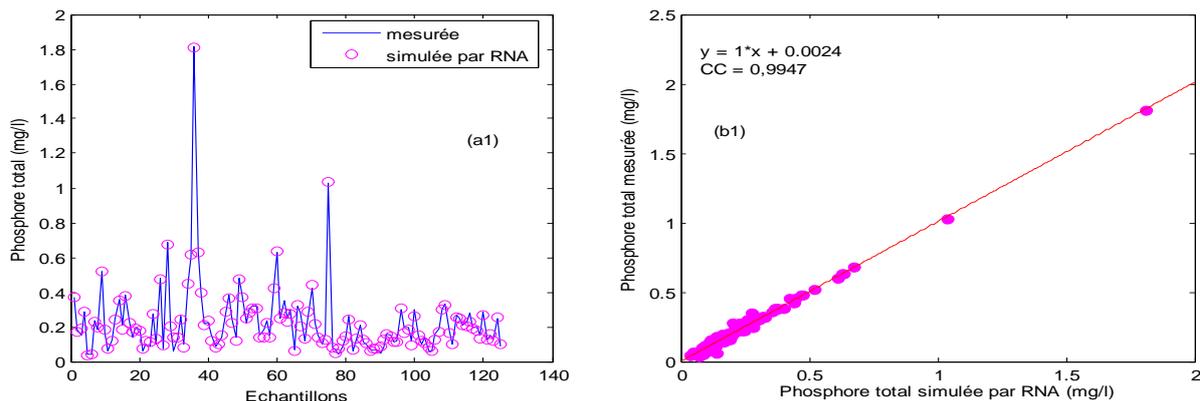


Figure 35 (a1) Valeurs mesurées et simulées de phosphore total (mg / L); (b1) Diagrammes de dispersion des valeurs simulées et mesurées de phosphore total (mg/L) pour ensemble complet de données en utilisant le modèle M0 (12-8-1)

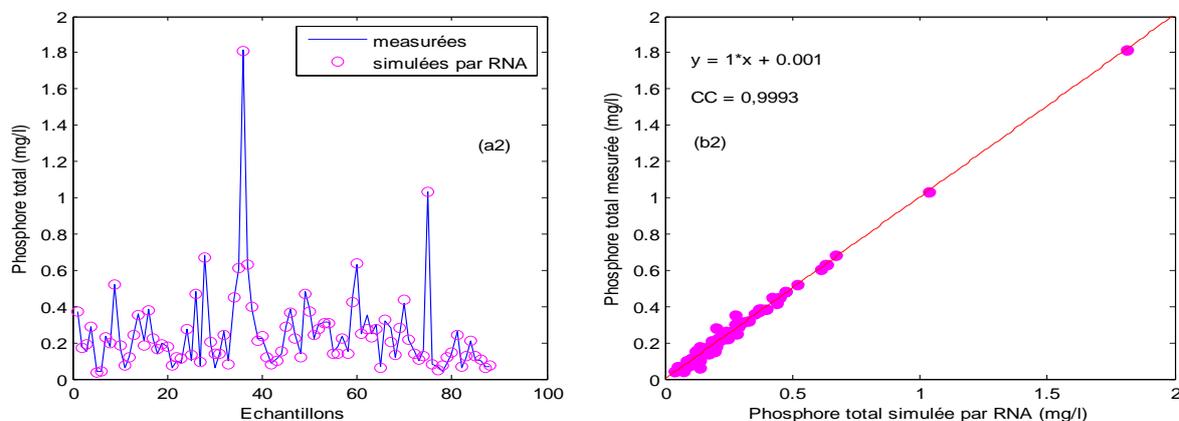


Figure 36 (a2) Valeurs mesurées et simulées de phosphore total (mg/L); (b2) Diagrammes de dispersion des valeurs simulées et mesurées de phosphore total (mg/L) lors de l'apprentissage pour le modèle M0 (12-8-1)

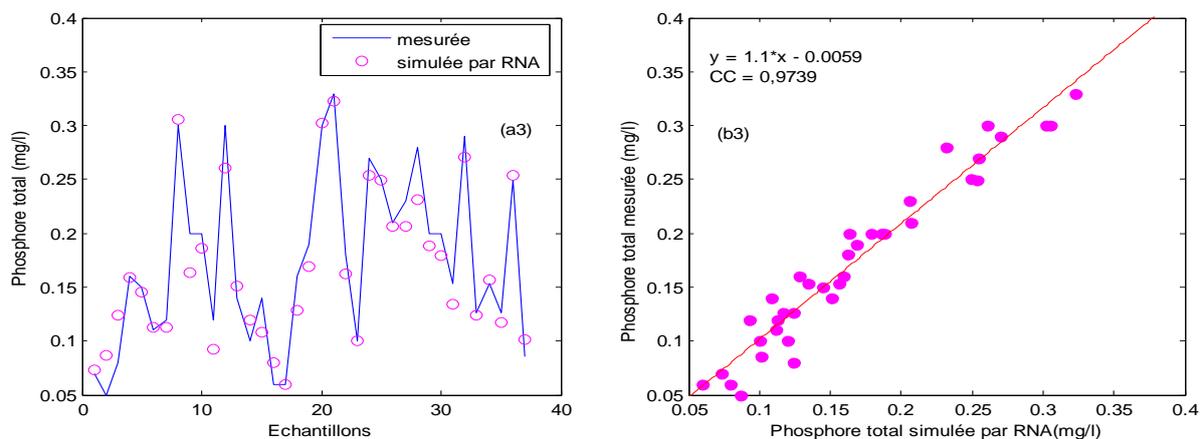


Figure 37 (a3) Valeurs mesurées et simulées de phosphore total (mg / L); (b3) Diagrammes de dispersion des valeurs simulées et mesurées de phosphore total (mg/L) lors de la validation en utilisant le modèle M0 (12-8-1)

Une approche utilisant des modèles de réseau neuronal artificiel (RNA) a été présentée pour analyser le comportement des variables de la qualité de l'eau dans le barrage de Ouizert. Des modèles simples basés sur les RNAs sont développés et appliqués avec succès pour quantifier les concentrations de phosphore total (TP) dans le barrage de Ouizert. Les RNAs sont capables d'approximer la réponse de type d'entrées-sortie d'un modèle de qualité de l'eau avec une précision très élevée en apprentissage et en validation. Les résultats montrent que les modèles de RNA peuvent simuler efficacement les variations saisonnières de la qualité de l'eau du réservoir. Ces résultats suggèrent également que les modèles de RNA sont des bons approximateurs du phosphore total (PT).

Des résultats encourageants obtenus dans ce chapitre révèlent que les outils basés sur l'intelligence artificielle peuvent être très utiles pour modéliser la qualité de l'eau et peuvent contribuer de manière significative pour une amélioration constante à la gestion de la qualité des ressources en eau.

IV.2.2.3. Simulation de la demande biochimique en oxygène par les réseaux de neurones artificiels

IV.2.2.3.1. Introduction: Relation entre l'eutrophisation et la DBO₅

L'eutrophisation est un concept qui indique une augmentation de la productivité primaire dans les écosystèmes aquatiques, principalement en termes de biomasse algale. Généralement, la croissance excessive des algues est liée aux apports des éléments nutritifs limitant (phosphore ou azote, selon le type d'écosystème et l'utilisation des terres dans le bassin hydrographique). L'excès des nutriments dans les lacs et réservoirs stimule la prolifération d'algues (algal bloom) qui meurent plus tard. Les microorganismes consomment de l'oxygène lorsqu'ils décomposent la matière morte (DBO élevée) ce qui provoque la diminution du taux d'oxygène dissous, les autres organismes ne peuvent pas respirer (faible DO) et meurent en contribuant encore plus au processus de décomposition. Par conséquent, L'eutrophication cause l'augmentation de la DBO₅ et la diminution du taux d'oxygène dissous (DO), donc l'eutrophisation est associée à une augmentation de la matière organique autochtone (produite à l'intérieur du réservoir) et de la DBO₅ et de la DCO.

La partie autochtone de l'apport total de composés organiques dans un réservoir augmente avec l'augmentation de degré d'eutrophisation et avec un temps de séjour d'eau prolongé (c'est-à-dire un temps de résidence de l'eau plus long). La charge totale de la DBO₅ d'un plan d'eau est constituée de la charge des apports extérieurs et de la charge résultant de la production primaire de matière organique par le phytoplancton dans le plan d'eau. La

production primaire peut être convertie en DBO, avec un coefficient de 0,5 de la production primaire brute (exprimée en oxygène). Le pourcentage de réduction de la DBO₅ de la charge totale dépend du temps de rétention d'eau et de la profondeur du plan d'eau (**Jorgensen et al., 2005**).

Plusieurs auteurs ont révélé une corrélation élevée entre le phénomène de l'eutrophisation exprimée en terme de chlorophylle-a et la DBO₅ (**Xiao-long et al., 2007; Kaneko et al., 2014; Heiskary et Markus, 2001**).

L'eutrophisation est un problème majeur pour les gestionnaires de l'eau en raison de la toxicité de nombreuses algues bleu-vert et les niveaux élevés de la demande biologique en oxygène (DBO₅) (**Eamus et al., 2006**).

IV.2.2.3.2. Elaboration des modèles

Le but de cette partie est de créer des modèles pour la simulation du niveau de la DBO₅ dans le réservoir de Ouizert, en utilisant un ensemble de données historiques comme inputs. Pour ce faire, un certain nombre de réseaux neuronaux artificiels basés sur l'algorithme d'apprentissage de rétropropagation ont été développés. Ces modèles ont été calés (apprentissage) et validés, en utilisant les données de 12 variables de qualité de l'eau contrôlées dans une station de surveillance dans le barrage de Ouizert et pour une période de 10 ans avec un pas de temps mensuel.

Les réseaux de neurones artificiels (RNA) ont été utilisés avec succès pour modéliser la DBO₅, plusieurs auteurs ont testé efficacement les RNA dans les rivières (**Dogan et al 2009, Singh et al., 2009; Šiljić et al., 2015; Šiljić Tomić et al., 2016; Raheli et al., 2017**) et dans les stations d'épuration (**Ozkan et al., 2009; Dogan et al., 2008; Heddham et al., 2016**). Les résultats obtenus étaient satisfaisants.

Les statistiques de base des variables de qualité de l'eau mesurées dans le réservoir de Ouizert sont illustrées dans le tableau 11.

Pendant son cours, l'oued de Sahouat traverse plusieurs municipalités et un certain nombre d'affluents déversant d'énormes quantités d'eaux usées non traitées dans le barrage de Ouizert. Les variables indépendantes ont également montré un coefficient de variation entre 4.29 % et 140.39 %. Cette variabilité entre les échantillons peut être attribuée aux grandes variations du climat (variation saisonnières et interannuelles des précipitations) et aux influences saisonnières dans la région étudiée. Le pH a montré la plus faible variation et cela peut être dû à la capacité tampon du barrage. Les variables d'origine anthropique ont montré des variations plus importantes que celles des variables d'origine naturelle (tableau). Cela peut être attribué

au fait que les processus géogéniques (d'origine naturels) sont presque à l'état d'équilibre, alors que les processus anthropogéniques dépendent du temps dans la nature.

Tableau 11 statistiques de base des variables de la qualité de l'eau du barrage Ouizert

	Moyenne	Ecart type	Min	Max	CV %	Corrélation avec DBO₅
DBO ₅ (mgO ₂ /l)	9.6504	6.8715	1.3000	60.7	71.20	1
RS (mg/l)	891.7460	208.9243	480	1360	23.43	0,033
DCO (mgO ₂ /l)	42.1667	23.3343	10	130	55.33	0,625
NH ₄ (mg/l)	0.3010	0.4226	0	3.15	140.39	0,037
NO ₃ (mg/l)	8.9683	6.2494	1	31	69.68	-0,099
NO ₂ (mg/l)	0.3571	0.3500	0.0100	2.4	98.01	0,152
TIN (mg/l)	9.4970	6.4577	1.18	32.85	67.99	-0,085
O ₂ (mg/l)	8.4594	2.3720	4.5950	15.852	28.03	0,155
O ₂ (%)	94.4188	27.2649	50	179.4	28.87	0,155
pH	7.9756	0.3422	7.1000	9.05	4.29	0,051
P total (mg/l)	0.2302	0.2080	0.0400	1.813	90.35	0,099
PO ₄ ⁻³ (mg/l)	0.1361	0.1521	0	1.3	111.75	0,094
MO (mg/l)	7.2544	2.8308	0	25	39.02	0,187

Le tableau 11 montre que la DBO₅ est fortement corrélée avec la DCO avec un coefficient de corrélation de l'ordre de 0,625, pour cette raison la DCO est la variable la plus explicative de la DBO₅, c'est-à-dire une grande part de la variabilité de la DBO₅ peut être expliquée par la DCO via les RNAs. Par conséquent, le DCO doit être utilisé comme variable commune pour chaque combinaison d'entrées (ensemble de variables d'entrée) lors de la création des modèles de RNAs. Les autres variables de la qualité de l'eau, comme le montre le tableau, sont faiblement corrélées avec la DBO₅.

IV.2.2.3.3. Résultats de simulation

Le tableau 12 montre les résultats de simulation de la DBO₅ lors de l'apprentissage et de la validation pour les modèles de RNAs développés. L'architecture finale des modèles de RNAs est également présentée dans ce tableau (colonne 4), le nombre optimal de neurones pour chaque modèle et le nombre correspondant de variables d'entrées sont également enregistrés.

Dans ce tableau, les valeurs en gras indiquent les résultats acceptables. Dans le tableau (M ***), (M **) et (M *) indiquent de très bons résultats, bons résultats et assez bons résultats, respectivement. Le nombre d'itérations a été fixé à 1000.

Le tableau 12 montre que les modèles donnent des résultats sensiblement différents, en fonction du nombre de variables d'entrées et du nombre de neurones dans la couche cachée. En effet, pour le modèle de base (M₀), celui qui contient toutes les variables d'entrées, les critères de performance ont respectivement des valeurs de 0,9039 et 2,1203 mg/L pour le coefficient de corrélation et le RMSE durant la phase de validation.

Tableau 12 les modèles de RNAS retenus pour modéliser la DBO₅

Modèles	Variables d'entrée	Choix des entrées	Structure du modèle	R		RMSE (mg/L)	
				Apprentissage	Validation	Apprentissage	Validation
M₀*** (modèle de base)	RS DCO NH4 NO3 NO2 TIN O2 O2p Ph Ptot PO4 MO.	Toutes les variables d'entrées disponibles	12-5-1	0,9750	0,9039	1,6399	2,1203
M ₁ *	RS NH4 NO3 NO2 TIN O2 O2p Ph Ptot PO4 MO	Toutes les variables d'entrées à l'exception de DCO	11-28-1	0,8910	0,4389	3,7858	4,8286
M₂***	TIN, DCO, O2%, RS, PO4.	Toutes les variables As my paper	5-8-1	0,9665	0,9359	1,9086	1,6831
M ₃ *	DCO NH4 NO3 NO2 TIN	DCO et composés azotés	5-6-1	0,5471	0,9534	6,1830	1,5753
M₄***	DCO PO4 Ptot	DCO et composés phosphatée	3-13-1	0,9573	0,9238	2,1538	2,1883
M₅***	DCO TIN Ptot	DCO et azote inorganique et Phosphore total	3-7-1	0,9282	0,9365	2,8225	2,0054
M ₆ *	DCO MO	Deux variables	2-7-1	0,5574	0,9322	6,1298	1,8725
M ₇ *	DCO TIN	Deux variables	2-5-1	0,5507	0,9532	6,1617	1,7360
M₈**	DCO PO4	Deux variables	2-12-1	0,7672	0,9127	4,8047	1,9563
M₉**	DCO	Une seule variable	1-9-1	0,6042	0,9005	5,8828	2,2715

Le modèle M_1 qui renferme toutes les variables d'entrées à l'exception de la DCO, donne de bons résultats durant l'apprentissage, mais durant de la validation on enregistre une large détérioration de ses performances avec un coefficient de corrélation de 0,4389 et un RMSE de 4,8286 mg/L. Cela signifie que l'élimination de la DCO de la base de données de variables d'entrées cause une diminution significative de la qualité de simulation du modèle. On peut conclure que la DCO est la variable la plus effective pour la modélisation de la DBO_5 .

Les modèles (M_0 , M_2 , M_4 , M_5) révèlent de très bons résultats en apprentissage et en validations avec des coefficients de corrélations oscillant entre 0,90 et 0,97. Ces modèles peuvent être utilisées, avec une fiabilité élevées, pour simuler la DBO_5 dans le barrage de Ouizert. Les résultats enregistrés avec ces modèles sont comparables avec ceux obtenus par **Dogan et al., 2009** ; par **Singh et al., 2009**; par **Šiljić et al., 2015**; par **Šiljić Tomić et al., 2016** et par **Raheli et al., 2017** où ces chercheurs ont testé la capacité des RNAs pour estimer la DBO_5 .

En phase d'apprentissage, les modèles (M_3 , M_6 , M_7 , M_8 et M_9), comme les montre le tableau précédent, enregistrent une qualité moyenne de simulation en comparant avec celles obtenues en période de validation.

En général, les meilleurs résultats sont enregistrés avec le modèles M_0 (modèle contenant toutes les variables d'entrées disponible) et le modèles M_2 ; modèle neuronal renfermant l'azote inorganique total, la DCO, le phosphate, le résidu sec et le taux de saturation en oxygène.

Il est à signaler que malgré des faibles corrélations entre la DBO_5 d'une part et les variables de la qualité de l'eau (à l'exception de la DCO) d'autre part, les RNAs ont pu « capturer » les associations (relations) de type *non linéaires* entre la DBO_5 et les autres variables de la qualité de l'eau de barrage.

Sur la base des résultats de cette partie, les RNAs peuvent être utilisé comme un outil efficace pour modéliser les variables de qualité de l'eau des barrages (la demande biochimique en oxygène dans ce cas) avec une performance très raisonnable.

Afin de bien illustrer les résultats de modélisation obtenus par le meilleur modèle neuronal (**M_0 : 12-5-1**), nous avons tracé les figures 38, 39 et 40 qui représentent les résultats de simulation de la DBO_5 pour toute l'ensemble de données (a1 et b1), l'apprentissage (a2 et b2) et l'ensemble de validation (a3 et b3). L'observation visuelle des graphiques de ces figures montre que les valeurs de la DBO_5 observées et celles simulées par le réseau de neurone (**M_0 : 12-5-1**) sont proches et les écarts oscillent autour de 0 à 2 mg O_2/l .

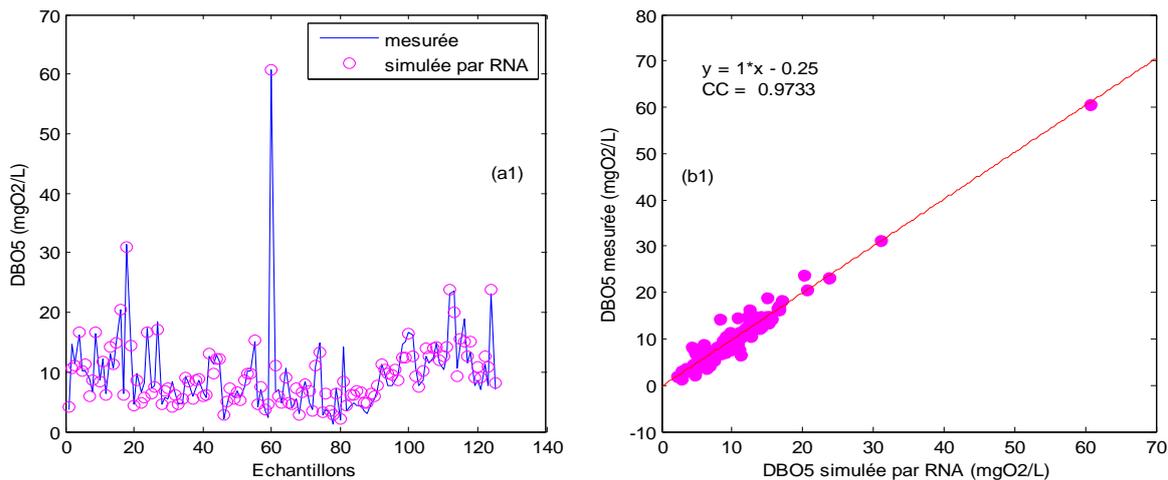


Figure 38 (a1) Valeurs mesurées et simulées de la DBO₅ (mg O₂ / L); (b1) Diagramme de dispersion des valeurs simulées et mesurées de la DBO₅ (mg O₂ / L) pour l'ensemble complet de données en utilisant le modèle M0 (12-5-1)

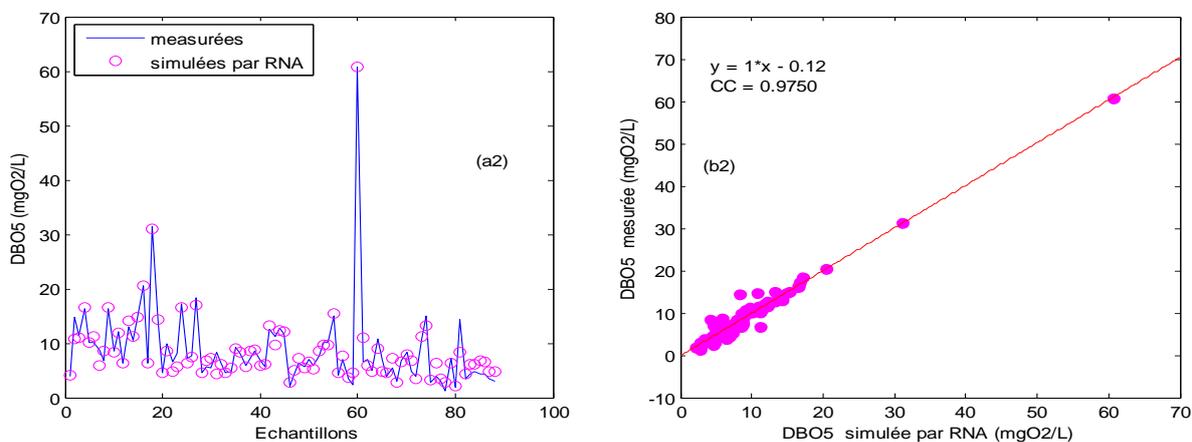


Figure 39 (a2) Valeurs mesurées et simulées de DBO₅ (mg O₂ / L); (b2) Diagramme de dispersion des valeurs simulées et mesurées de la DBO₅ (mg O₂ / L) lors de l'apprentissage en utilisant le modèle M0 (12-5-1)

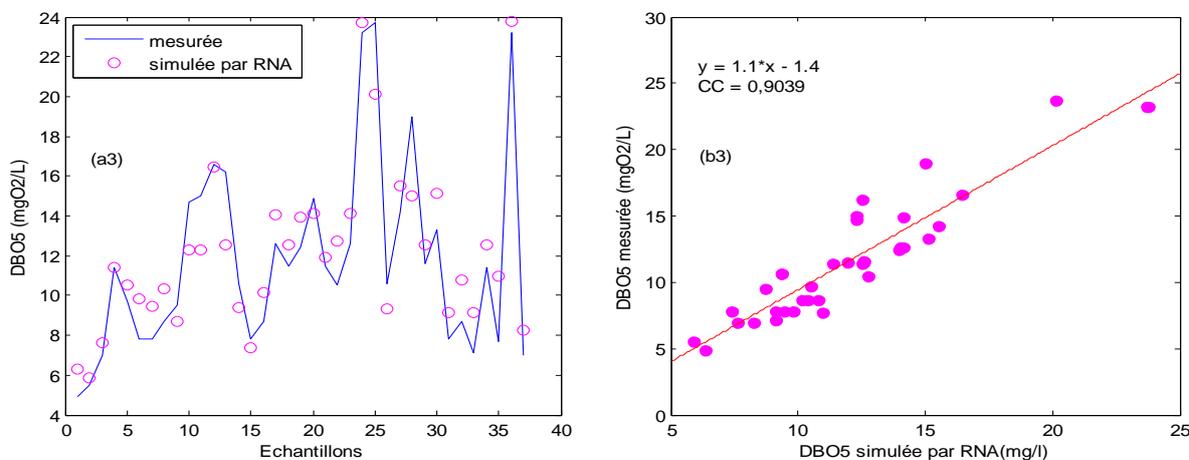


Figure 40 (a3) Valeurs mesurées et simulées de la DBO₅ (mg O₂ / L); (b3) Diagramme de dispersion des valeurs simulées et mesurées de la DBO₅ (mg O₂ / L) lors de la validation en utilisant le modèle M0 (12-5-1).

IV.3. Modélisation de l'eutrophisation par les systèmes d'inférence neuro-flous adaptatifs (ANFIS), comparaison avec les modèles de RNAs

IV.3.1. Introduction

Les réseaux de neurones artificiels appartiennent à la catégorie des modèles boîte noire (**Babovic, 2009**), par conséquent, il y a un manque d'interprétabilité de modèle de RNA, en d'autres termes, ils n'ont pas une structure lisible et compréhensible (les modèles ANN ne sont pas transparents), parce que les réseaux de neurones artificiels permettent de représenter des relations inconnue incorporées dans les données sous forme de poids synaptiques. Ils ne sont pas capables d'expliquer la relation causale entre les entrées et la réponse (sortie) (**Arthur, 2010**). En plus de ça, la majeure partie de nos informations sur l'écosystème aquatique est de nature qualitative plutôt que quantitative (**Chen et Mynett, 2004**). L'utilisation de la logique floue permet de pallier ces inconvénients en décrivant le comportement du système en utilisant le raisonnement humain (expert), en définissant une relation explicite par l'utilisation du langage naturel sous la forme de règles floues «si.....alors.....» (Ou règles linguistiques) et la floues logique. Cette méthode est également capable d'intégrer et de prendre en compte les informations inexacts, vagues, qualitatives et floues qui entourent souvent la qualité de l'eau.

Mais l'expert humain seul ne peut pas gérer une grande quantité de données traitant de la qualité de l'eau en utilisant des systèmes d'inférence floue. En outre, la définition des paramètres des fonctions d'appartenances et des règles floues demeure difficile dans les systèmes d'inférence floue (**Chen et Mynett, 2001; Lin et al., 2001 ; Shen et Chouchoulas, 2001**), l'utilisation de systèmes hybrides comme le système d'inférence neuro-flou adaptatif (ANFIS) résout ce problème en utilisant efficacement la capacité d'apprentissage des réseaux de neurones artificiels (RNA) pour la génération automatique des règles et l'optimisation des paramètres directement à partir des données plutôt que de consulter un expert humain (**Yan et al., 2010**). Ces dernières années, plusieurs études (**Najah et al., 2014; Ay et al., 2016; Ahmed et al., 2017**) ont appliqué ANFIS, avec succès, pour modéliser la qualité de l'eau.

L'objectif de cette partie est d'explorer la capacité des systèmes d'inférence neuro-flous adaptatifs (ANFIS) à modéliser certains indicateurs de l'eutrophisation, tout en utilisant les variables de la qualité de l'eau disponibles.

Comme dans le chapitre précédent, nous avons abordé deux études de cas, la première concerne la concentration en chlorophylle-a en utilisant une base de données mixte, et la deuxième touche la modélisation du phosphore total et la DBO₅ dans le barrage de Ouizert.

Comme il a été mentionné précédemment dans l'introduction générale, l'un des objectifs de cette étude est de comparer les performances de RNAs et d'ANFIS pour la modélisation des indicateurs de l'eutrophisation, pour cette raison, nous avons utilisé les mêmes combinaisons des variables d'entrées que l'on a utilisées lors de l'application des RNAs, ce qui a facilité la comparaison entre des deux modèles (RNA et ANFIS) contenant les mêmes variables d'entrées.

IV.3.2. Première étude de cas: Simulation de la de chlorophylle-a par les modèles ANFIS

IV.3.2.1. Développement des modèles

IV.3.2.1. 1. Variables d'entrées/sortie

Dans ce chapitre, la température de l'eau, le pH, l'oxygène dissous (O₂) (mg/l), le taux de saturation en oxygène (%), la visibilité de l'eau représenté par la profondeur de disque de Secchi (SD), la turbidité (NTU), la conductivité électrique (ds/m), le résidu sec (mg/l), l'ammonium (NH₄⁺), le nitrate (NO₃⁻²), le nitrite (NO₂⁻) l'azote inorganique total (TIN) le phosphore total (PT), le phosphate (PO₄⁻³) la demande chimique en oxygène (DCO), la matière organique (MO) sont utilisés comme entrées des modèles ANFIS en vue d'estimer la concentration en chlorophylle alpha (Chl-a).

IV.3.2.1. 2. Apprentissage d'ANFIS

Le développement du modèle ANFIS nécessite deux processus, l'identification de la structure et l'estimation des paramètres. L'identification de la structure implique le processus d'extraction des règles floues, tandis que l'identification des paramètres implique la recherche des paramètres de prémisses et des paramètres conséquents des systèmes ANFIS (**Jang et al., 1997**).

Dans cette partie, nous avons utilisé la méthode de «substractif clustering» pour l'identification de la structure, puisque cette méthode est plus performante que celle de 'grid partitioning' (**Belouz et al., 2018; Ay et Kişi, 2014; Heddam et al., 2011**).

L'estimation des paramètres est effectuée via un apprentissage hybride, les paramètres de conséquence étant ajustés en utilisant l'optimisation des moindres carrés tandis que les paramètres de prémisses sont ajustés à l'aide de l'algorithme de retro-propagation.

IV.3.2.1.3. Implémentation des modèle ANFIS

Le logiciel utilisé pour implémenter tous les modèles ANFIS est Matlab v7.4, précisément, la boîte à outils de logique floue intégrée à Matlab. Plusieurs codes de programmes ont été écrits en langage Matlab (m.file), pour développer différents modèles ANFIS.

Tout d'abord, plusieurs architectures ANFIS ont été essayées en utilisant ces codes, et les structures de modèles appropriées ont été déterminées pour chaque combinaison de variables d'entrées. La valeur optimale du rayon d'influence est déterminée, pour chaque modèle par la méthode de 'l'essai et l'erreur' (trial and error) en faisant varier le rayon (r) entre 0.1 et 1 avec un incrément de 0.01.

Un modèle ANFIS avec cinq variables d'entrée est illustré schématiquement sur la **figure 41**, cette figure représente toutes les composantes du modèle ANFIS.

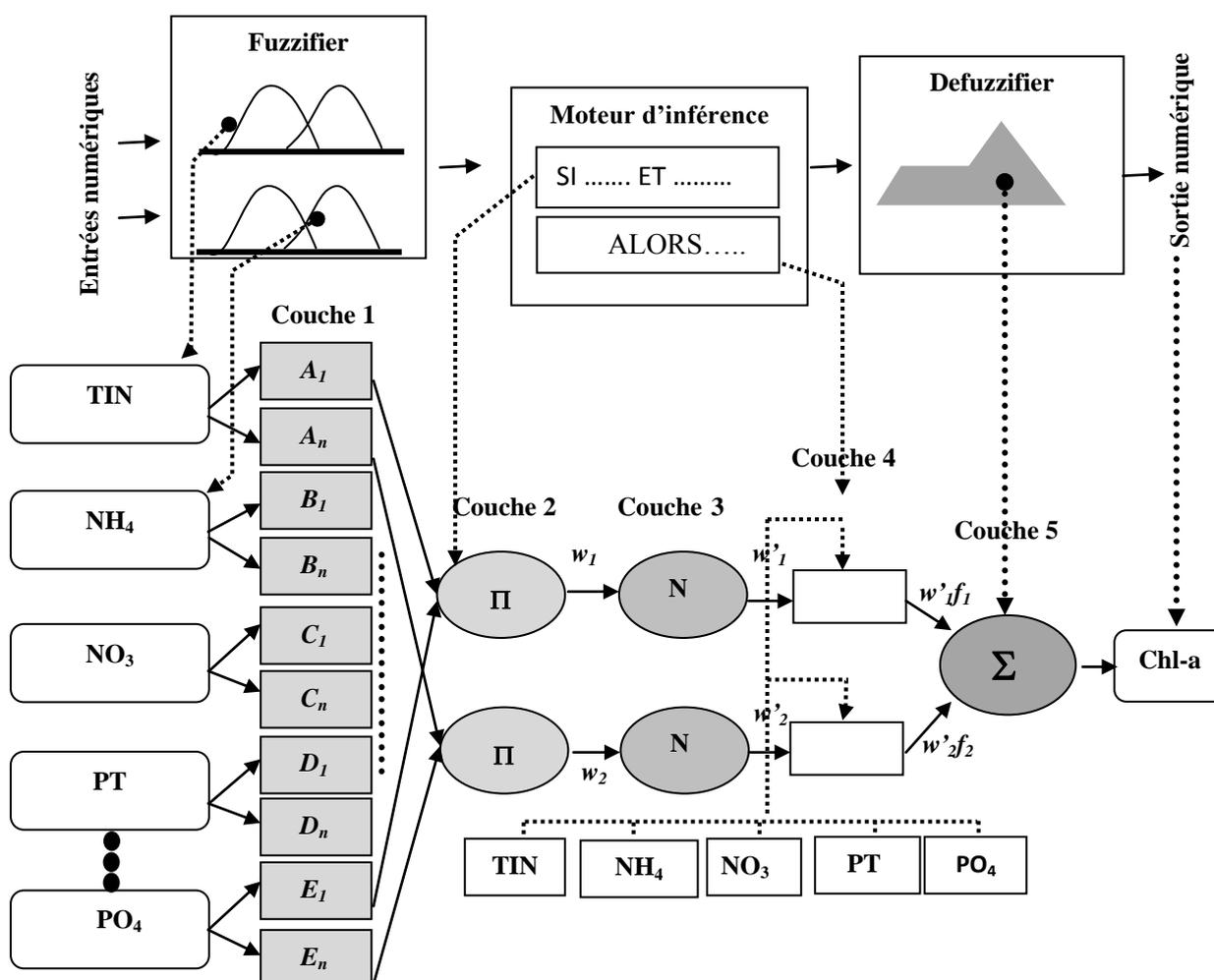


Figure 41 Schéma général du modèle ANFIS développé pour estimer la chlorophylle-a

IV.3.2.2. Résultats de simulation

Le tableau 13 résume les valeurs de critères de performance en calage et en validation (R^2 et RMSE respectivement) pour les modèles ANFIS retenus pour différentes combinaisons de variables d'entrées. La meilleure valeur du rayon d'influence pour chaque modèle est également enregistrée. Dans ce tableau (M ***), (M **) et (M *) indiquent de très bons résultats, de bons résultats et des résultats assez bons, respectivement. Le nombre d'itérations a été fixé à 100 itérations.

En général, le modèle ANFIS est un outil approprié pour analyser la qualité de l'eau de surface et les commentaires liés cette étude de cas sont les suivants:

Le tableau 13 montre clairement que les modèles apportent de bonne performance durant l'apprentissage (calage), alors qu'en validation on constate une certaine détérioration de performance des modèles, mais à l'exception du modèle M2, ils apportent des résultats relativement acceptables.

A l'inverse des RNAs, où nous avons montré que le meilleur résultat est enregistré avec un modèle utilisant toutes les variables d'entrée disponible (M0), les résultats obtenus par les modèles ANFIS révèlent que le meilleur résultat d'ANFIS est celui d'ANFIS-M₅ dont les variables d'entrées sont : l'azote inorganique total (TIN), l'ammonium (NH₄⁺), le nitrate (NO₃), le nitrite (NO₂), le phosphate (PO₄⁻²) et le phosphore total (PT) et les critères de performance sont de 1; 0,9073 pour le CC et de $7,7035 \cdot 10^{-6}$ mg/L; 3,1379 mg/L pour le RMSE pendant les phases d'apprentissage et de validation, respectivement, ce qui indique une compatibilité raisonnable entre les valeurs de chlorophylle-a mesurées et celles simulées par le Modèle ANFIS-M5. Le modèle ANFIS-M4 dont les variables d'entrées sont la conductivité électrique (CE), le résidu sec (RS), le pH, la température de l'eau, la turbidité et la profondeur de disque de Secchi est classé en deuxième place, les critères de performance sont de 1; 0.8967 pour le R et de $2,2836 \cdot 10^{-5}$ mg/L; 2.9643mg/L pour RMSE pendant les phases d'apprentissage et de validation, respectivement. Les résultats obtenus par les modèles ANFIS (M4 et M5) donnent de résultats supérieurs à ceux fournis par **Soyupak et Chen (2004)** où les auteurs ont testé le potentiel de la logique floue pour estimer la concentration de la chlorophylle-a dans le réservoir de Keban en Turquie, en utilisant les variables de la qualité de l'eau. Cela confirme la robustesse, la stabilité et l'efficacité des modèles ANFIS développés (modèle en gras).

Ces modèles présentent un grand avantage de simulation de la concentration de la chlorophylle-a, pratiquement, pour ces deux modèles testés, les valeurs du coefficient de corrélation en validation sont supérieures à 0.89, ce qui suppose une bonne aptitude des

modèles neuro-flous à reconstituer les valeurs manquantes de chlorophylle-*a* existantes dans les séries de la qualité de l'eau.

Tableau 13 Performances des modèles ANFIS élaborés pour modéliser la chlorophylle-*a*

Modèles	Variables d'entrée	Choix des entrées	Radius	R		RMSE ($\mu\text{g/l}$)	
				apprentissage	validation	apprentissage	validation
M₀* (modèle de base)	CE RS DCO NH ₄ NO ₃ NO ₂ TIN O ₂ O ₂ p pH Ptot PO ₄ TC° Turb SD MO.	Tous les variables d'entrées disponibles	1	1	0.7489	$4.386 \cdot 10^{-6}$	5,0511
M₁*	CE NH ₄ NO ₂ O ₂ p SD MO P _{tot} PO ₄	Matrice de corrélation plus composées phosphatée	0.49	0.9999	0.7863	0.00068	4.3922
M ₂	CE NH ₄ NO ₂ NO ₃ TIN O ₂ p SD MO	Matrice de corrélation plus composées azotées	0.48	1	0.5807	$7.437 \cdot 10^{-5}$	5.5208
M₃*	CE NH ₄ NO ₂ O ₂ p SD MO	Matrice de corrélation	0,2	1	0,6382	$5,6463 \cdot 10^{-5}$	5,1529
M₄***	CE RS Ph TC Turb SD	Paramètres physiques seules	0.41	1	0.8967	$2.2836 \cdot 10^{-5}$	2.9643
M₅***	NH ₄ NO ₃ NO ₂ TIN Ptot PO ₄	Les nutriments seulement	0,19	1	0,9073	$7,7035 \cdot 10^{-6}$	3,1379
M₆**	CE Ph TC Turb O ₂ p	Paramètres Facilement mesurables	0,19	1	0,8030	$8,3749 \cdot 10^{-6}$	4,1477
M₇*	CE RS Ph TC	Paramètres physiques facilement mesurables	0,07	1	0,7300	$8,909 \cdot 10^{-6}$	5,6869

CE: conductivité électrique; RS: résidu sec; DCO ; NH₄ : ammoniacque, NO₃: nitrate, NO₂ : nitrite, TIN : azote inorganique totale ; O₂ : oxygène dissous ; O₂p : taux de saturation en oxygène ; pH ; Ptot : phosphore total ; PO₄ : phosphate ; TC : température de l'eau ; Turb : turbidité ; SD : visibilité de l'eau (m) ; MO : matière organique.

Dans l'ensemble, les résultats de cette étude pourraient aider les décideurs dans la gestion de la qualité de l'eau et l'application de cette technique pourrait s'appliquer à une autre partie du monde.

Afin de bien visualiser les résultats de simulation de ce modèle d'ANFIS-M5 et de présenter sa performance pour chacun de trois barrage, nous avons tracé le diagramme de dispersion par groupe (*Scatter plot by group*) ci-après:

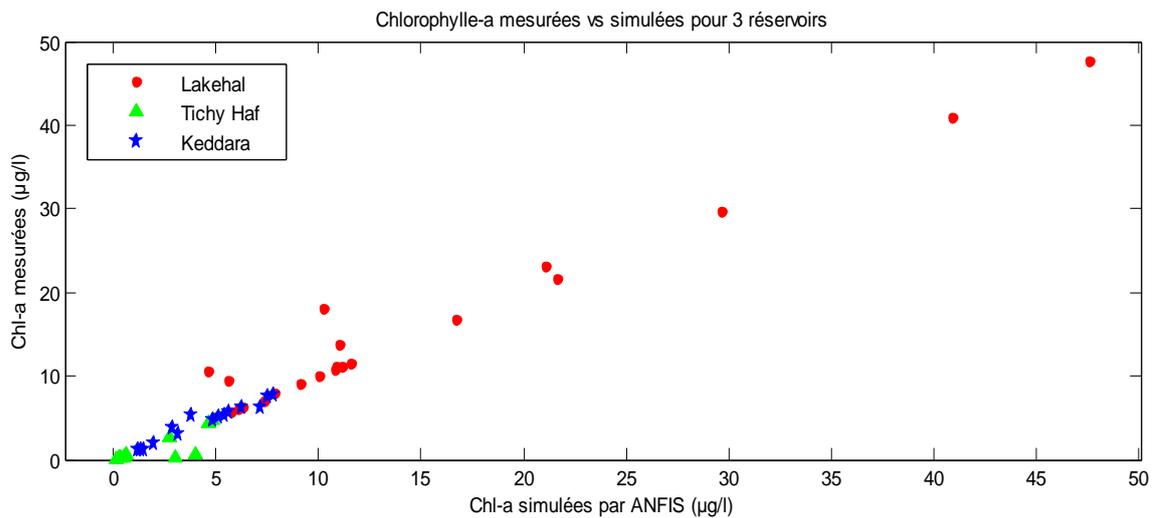


Figure 42 Diagramme de dispersion par groupe de valeurs de chlorophylle-a observées et celles simulées par ANFIS-M5 pour l'ensemble de données.

Malgré l'insuffisance de données historique pour l'apprentissage, et leurs origines diversifiées (données provenant de 3 barrages différents), les modèles proposés montrent une bonne efficacité dans la simulation des valeurs de chlorophylle-a dans les masses d'eau. En outre, ces modèles constituent des outils précieux pour étudier la dynamique de la qualité de l'eau des écosystèmes aquatique de retenue de barrages.

Dans le but de bien illustrer les résultats de modélisation obtenus par le meilleur modèle **M5** d'ANFIS, nous avons tracé les figures 43, 44 et 45 qui représentent les résultats de simulation de chlorophylle-a pour toute l'ensemble de données (a1 et b1), l'apprentissage (a2 et b2) et la validation (a3 et b3). L'observation visuelle de ces figures montrent que les valeurs de chlorophylle-a observées et celles simulées par le réseau de neurone (**M5**) sont proches.

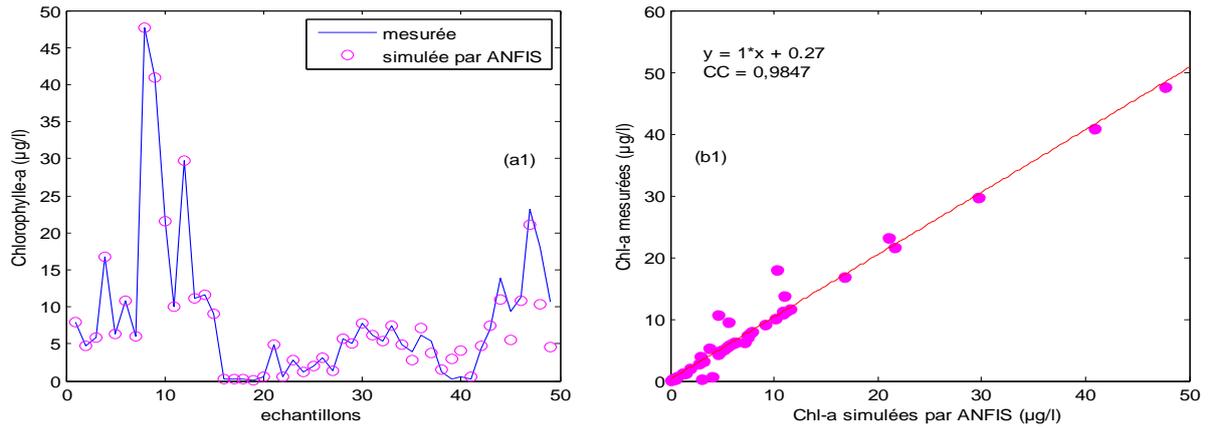


Figure 43 (a1) Valeurs mesurées et simulées de la chlorophylle-a (µg / L); (b1) Diagramme de dispersion des valeurs simulées et mesurées de la chlorophylle-a (µg/L) pour l'ensemble de données en utilisant le modèle ANFIS-M5

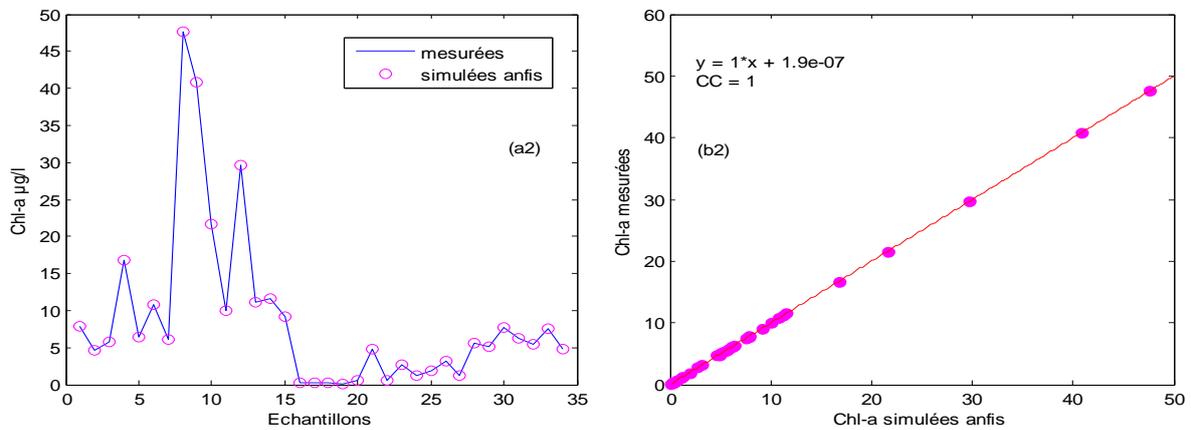


Figure 44 (a2) Valeurs mesurées et simulées de Chl-a(µg/L); (b2) Diagramme de dispersion des valeurs simulées et mesurées de la Chl-a (µg/L) lors de l'apprentissage en utilisant le modèle ANFIS-M5.

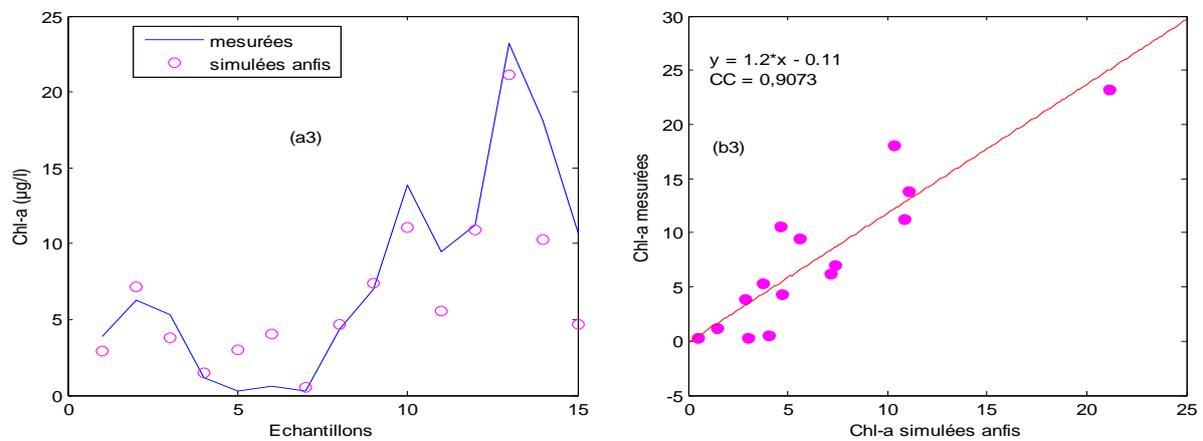


Figure 45 (a3) Valeurs mesurées et simulées de chlorophylle-a (µg / L); (b3) Diagramme de dispersion des valeurs simulées et mesurées de la chlorophylle-a (µg/L) lors de la validation pour le modèle ANFIS-M5

Le modèle de logique floue, une fois mis au point pour un plan d'eau, peut être utilisé de manière favorable lors de la surveillance ultérieure, en tant qu'outil de gestion prédictive. On peut en conclure que la modélisation neuro-floue peut être appliquée avec succès pour la modélisation de qualité de l'eau et l'eutrophisation dans les réservoirs, et peut remplacer les approches classiques en raison de sa simplicité. L'application d'ANFIS pourrait être utilisée à l'avenir pour étudier l'applicabilité de cette approche à d'autres réservoirs. En bref, ANFIS est un outil puissant pour la modélisation et l'évaluation de la qualité des eaux et de l'eutrophisation.

IV.3.2.3. Comparaison entre les résultats des modèles ANFIS et ceux de RNAs

Les Tableaux 09 et 13 montrent que les deux techniques (RNAs et ANFIS) donnent des résultats acceptables similaires et sont des algorithmes potentiels à utiliser comme outils de gestion de la qualité de l'eau. Les RNAs sont une approche plus simple dans le sens de la complexité des calculs, puisqu'ils renferment un nombre moindre de paramètres ajustables et n'exigent pas un temps de calcul considérable que les modèles ANFIS, mais d'un autre côté, les modèles ANFIS, avec le plus petit nombre d'entrées (modèles M5 et M4), offrent une précision plus élevée que les RNAs. De point de vue parcimonie, les modèles ANFIS semblent être plus parcimonieux que les modèles de RNAs, parce que les meilleurs résultats sont obtenus avec les modèles ANFIS qui utilisent un nombre de variable d'entrées moindre, six variables d'entrée pour les modèles M5 et M4.

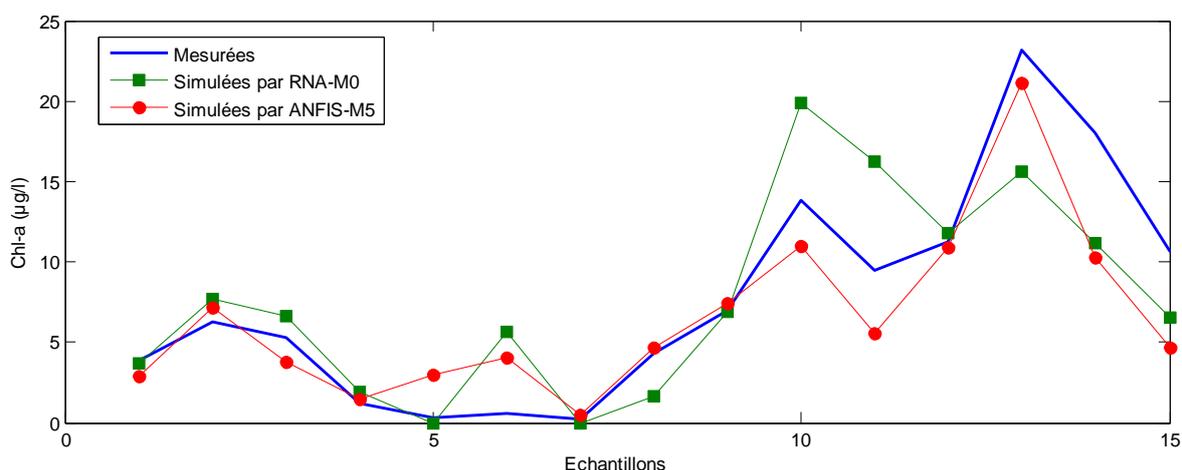


Figure 46 Résultats de modélisation des concentrations de chlorophylle-a (Chl-a) en utilisant deux techniques de machine learning.

Les résultats de simulation de la concentration en chlorophylle-a durant la phase de validation en utilisant le meilleur modèle de RNAs et celui d'ANFIS sont indiqués dans le tableau 14 et la figure 46. Il ressort de ce tableau que le modèle ANFIS-M5 simule les valeurs extrêmes de

chlorophylle-a mieux que le modèle de RNA-M0. On peut conclure que le modèle ANFIS est recommandé pour la modélisation éco-environnementale dont l'eutrophisation.

Tableau 14 Résultats de simulation de la concentration en Chl-a par les meilleurs modèles de RNAs et d'ANFIS comparée avec les valeurs mesurées.

N° d'échantillon	[Chl-a] Mesurées (µg/L)	[Chl-a] simulées par le RNA-M0	[Chl-a] simulées par le ANFIS-M5
01	3,87	3,69	2,88
02	6,25	7,65	7,15
03	5,33	6,67	3,76
04	1,16	1,93	1,45
05	0,3	0	3,02
06	0,58	5,64	4,008
07	0,26	0	0,492
08	4,33	1,65	4,659
09	6,98	6,90	7,39
10	13,82	19,91	11,03
11	9,45	16,26	5,58
12	11,22	11,77	10,86
13	23,21	15,58	21,08
14	18,05	11,20	10,28
15	10,64	6,51	4,63

IV.3.3. Deuxième étude de cas: barrage de Ouizert

IV.3.3.1. Simulation du phosphore total par les systèmes d'inférence neuro-flous adaptatifs

IV.3.3.1.1. Introduction

Le phosphore est le nutriment le plus influent sur la croissance et la productivité biologique des lacs et des réservoirs, il est le facteur limitant dans la plupart des lacs d'eau douce; ainsi, l'eutrophisation des lacs est généralement liée à l'abondance du phosphore (Haggard et al., 1999). L'augmentation des apports de phosphore stimule l'abondance du phytoplancton, entraînant une augmentation de la concentration de chlorophylle-a (Carlson, 1977; Haggard et al., 1999).

IV.3.3.1.2. Développement des modèles

Les variables d'entrées/sortie des modèles ANFIS sont les mêmes utilisées dans la partie simulation du phosphore total par les RNAs.

Pour déterminer les relations non linéaires entre les variables de qualité de l'eau et l'indicateur de l'eutrophisation (le phosphore total), le potentiel des modèles ANFIS basés sur la 'subtractive clustering' a été exploré en utilisant plusieurs combinaison de données d'entrées comme les montrent le tableau 15.

IV.3.3.1.3. Résultats de simulation

Le tableau 15 montre les résultats lors de l'apprentissage et de la validation des modèles ANFIS retenus pour simuler le phosphore total. Les meilleures valeurs du rayon d'influence (radius) pour chaque modèle sont également enregistrées. Dans ce tableau, les valeurs en gras indiquent des résultats acceptables.

Le modèle M2 est classé en première place, il dépasse tous les autres modèles en termes de divers critères de performance, et cela était attendu, car il prend en compte toutes les variables qui sont corrélées avec le phosphore total (PT) et les critères de performance sont de 0,9969; 0,9821 pour CC, de 0,0186 (mg/L); 0,0147 (mg/L) pour RMSE pendant les phases d'apprentissage et de validation respectivement.

Un regard attentif sur le tableau 15 montre que les modèles ANFIS dont le phosphate (PO_4) constitue une composante des données d'entrée (M0, M2 et M5), apportent de très bons résultats (valeurs en gras dans ce tableau). Nous pouvons conclure que le phosphate est la principale variable pour la modélisation du phosphore total via ANFIS.

Les résultats de l'étude indiquent que les modèles ANFIS peuvent simuler efficacement le comportement et la dynamique de la qualité de l'eau du réservoir.

Les modèles M_1 , M_3 et M_4 apportent de mauvais résultats avec des coefficients de corrélations très faibles oscillant entre 0,11 à 0,27 pour la phase de validation, l'une des raisons derrière cette faible performance est que les variables d'entrées de ces modèles sont faiblement corrélés avec le phosphore total.

Tableau 15 Performances des modèles ANFIS élaborés pour la simulation du phosphore total

Modèles	Variables d'entrée	Choix des entrées	Radius	R		RMSE (mg/l)	
				Apprentissage	Validation	Apprentissage	Validation
M_0^{***} (modèle de base)	RS DBO DCO NH ₄ NO ₃ NO ₂ TIN O ₂ O ₂ p pH PO ₄ MO.	Toutes les variables d'entrées disponibles	<u>0,85</u>	0,9974	0,9657	0,0168	0,0213
M_1	RS DBO DCO NH ₄ NO ₃ NO ₂ TIN O ₂ O ₂ p Ph MO.	Toutes les variables à l'exception de phosphate	0,84	0,8727	0,2712	0,1160	0,2294
M_2^{***}	NH ₄ NO ₃ TIN PO ₄	Matrice de corrélation	<u>0,40</u>	0,9969	0,9821	0,0186	0,0147
M_3	RS NH ₄ NO ₃ NO ₂ TIN MO.	Composées azotées Plus MO et RS	0,5	0,9048	0,1158	0,1012	0,4122
M_4	NH ₄ NO ₂ NO ₃ TIN	Composées azotées	0,4	0,8001	0,2406	0,1426	0,1599
M_5^{***}	PO ₄	Phosphate seule	<u>0,2</u>	0,9953	0,9834	0,0229	0,0148

Dans le but de bien illustrer les résultats de modélisation obtenus par le modèle M_0 d'ANFIS, nous avons tracé les figures 47, 48 et 49 qui représentent les résultats de simulation du phosphore total (PT) pour l'ensemble de données (a1 et b1), l'apprentissage (a2 et b2) et la validation (a3 et b3). L'observation visuelle de ces figures montrent que les valeurs du phosphore total (PT) observées et celles simulées par le modèle ANFIS- M_0 sont presque superposées.

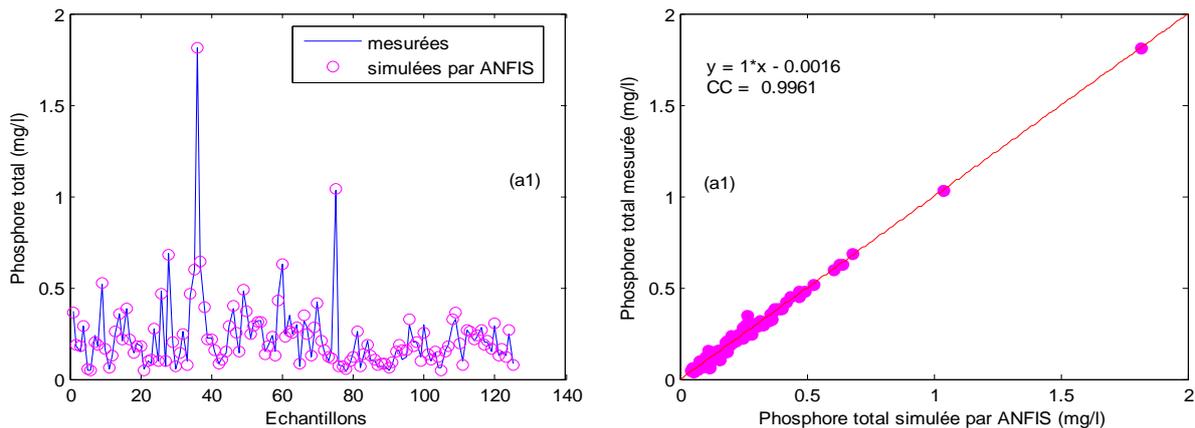


Figure 47 (a1) Valeurs mesurées et simulées de phosphore total (mg / L); (b1) Diagrammes de dispersion des valeurs simulées et mesurées de phosphore total (mg/L) pour ensemble complet de données en utilisant le modèle **ANFIS-M0**

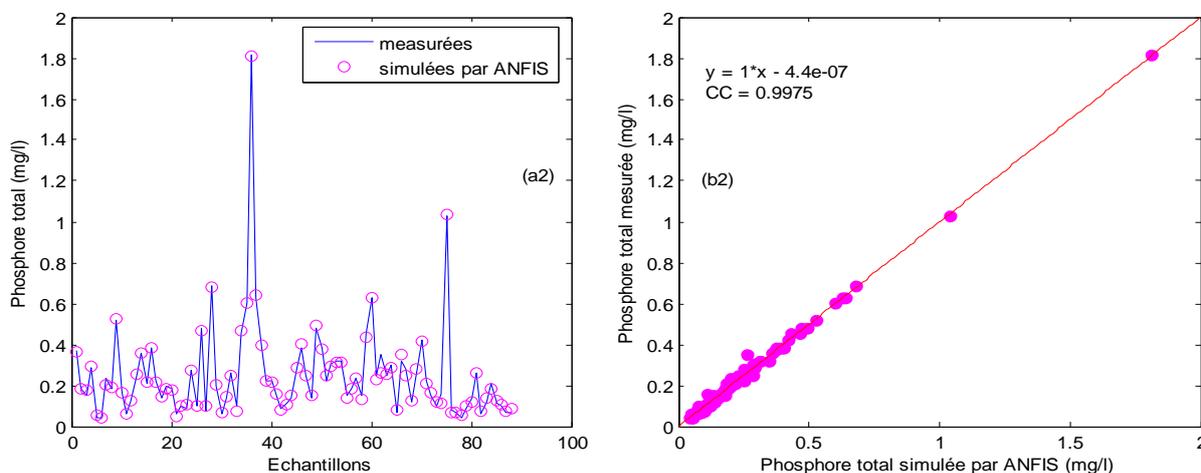


Figure 48 (a2) Valeurs mesurées et simulées de phosphore total (mg/L); (b2) Diagrammes de dispersion des valeurs simulées et mesurées de phosphore total (mg/L) lors de l'apprentissage pour le modèle **ANFIS-M0**

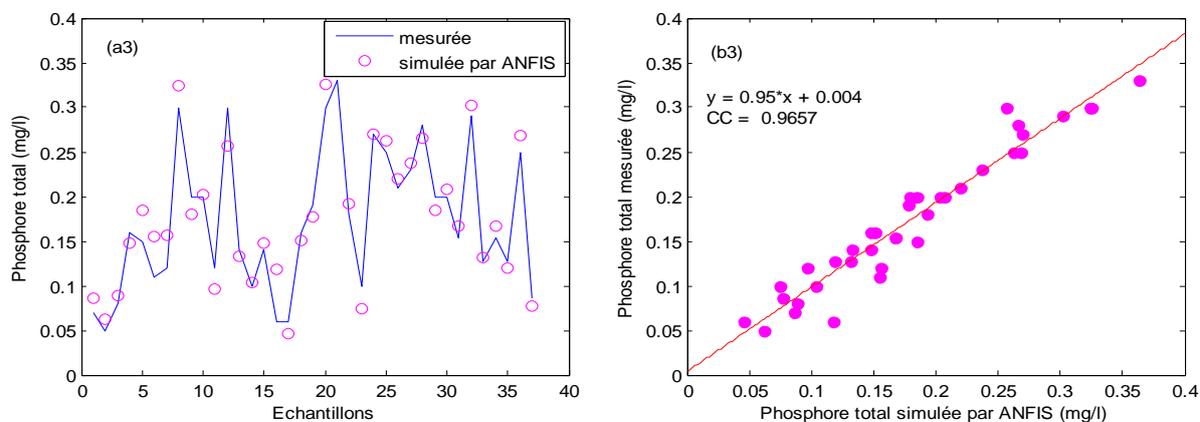


Figure 49 (a3) Valeurs mesurées et simulées de phosphore total (mg / L); (b3) Diagrammes de dispersion des valeurs simulées et mesurées de phosphore total (mg/L) lors de la validation en utilisant le modèle **ANFIS-M0**.

IV.3.3.1.4. Comparaison entre les résultats des modèles ANFIS et ceux de RNAs

Une simple comparaison entre les tableaux 10 et 15, en se basant sur les critères statistiques de performance (CC et RMSE) dans la phase d'apprentissage et celle de validation, montre que les deux techniques (RNAs et ANFIS) apportent des performances similaires concernant la simulation de la concentration du phosphore total. Dans l'ensemble, les modèles ANFIS et ceux de RNAs (sélectionnés en gras) constituent des outils appropriés pour simuler la qualité de l'eau de surface.

Du point de vue stabilité et rapidité de calculs, les modèles ANFIS sont plus stables que les modèles de RNAs puisque les valeurs finales des poids synaptiques des RNA dépendent des valeurs initiales d'une part, et les modèles de RNAs sont plus rapides que les modèles ANFIS parce que ces derniers contiennent un nombre relativement élevé de paramètres à ajuster par le biais du processus d'apprentissage d'autre part.

IV.3.3.2. Modélisation de la demande biochimique en oxygène (DBO₅) par les systèmes d'inférence neuro-flous adaptatifs

IV.3.3.2.1. Introduction

L'une des variables les plus importantes de la qualité de l'eau est la demande biochimique en oxygène (DBO); il s'agit d'une évaluation approximative de la teneur en matière organique biodégradable existante dans un écosystème aquatique. La DBO a été définie par « Environmental Protection Agency » des États-Unis comme étant «*la quantité d'oxygène (mg/L) nécessaire à l'oxydation de la matière organique par action biologique dans des conditions d'essai standards et spécifiques* » (<http://www.epa.gov/owow/monitoring/>). La DBO₅ est le paramètre le plus largement utilisé, il est défini comme la DBO obtenue après une période d'incubation de 5 jours.

La mesure de la DBO au laboratoire est coûteuse et demande beaucoup de temps, il faut au moins cinq jours comme période d'incubation. En outre, les méthodes de laboratoire couramment utilisées pour mesurer la demande biochimique en oxygène (DBO₅) sont fastidieuses et inadaptées en raison de la présence de substances toxiques dans l'échantillon pouvant inhiber l'activité biologique de certains germes, l'interférence de l'oxygène dissous avec l'oxygène produit par les algues présentes dans cet échantillon en raison de la photosynthèse, le processus de nitrification absorbe l'oxygène pour oxyder le NH₄⁺ en NO₂⁻ puis en NO₃⁻. Les conditions de laboratoire pour la détermination de la DBO diffèrent généralement de celles des systèmes aquatiques (Singh et al., 2009).

En raison des multiples inconvénients qui entourent les méthodes classiques de détermination de la DBO₅, le recours à des méthodes indirectes comme la modélisation constitue une alternative prometteuse.

Cette partie examine le potentiel des systèmes d'inférence neuro-flous adaptatifs (ANFIS) dans l'estimation de la demande biochimique en oxygène (DBO₅) en utilisant un nombre de variables de qualité de l'eau limitées.

L'objectif principal de ce chapitre est de (1) développer des modèles basés sur le système d'inférence adaptative neuro-floue (ANFIS) capable de simuler la demande biochimique en oxygène (DBO₅) dans le réservoir d'Ouizert au Nord de l'Algérie et (2) démontrer la capacité et l'applicabilité d'ANFIS dans les questions pratiques de la gestion de la qualité de l'eau (3) choisir la combinaison optimale de variables d'entrée pour améliorer la performance du modèle.

IV.3.3.2. Développement des modèles

IV.3.3.2.1. Variables d'entrées/sortie

Les modèles développés ont été créés en utilisant les données observées de la qualité de l'eau qui ont été recueillies pendant une période de dix ans pour le barrage de Ouizert, à un pas de temps mensuel. Les données d'entrée utilisées sont l'azote inorganique total (TIN), l'ammonium (NH₄), le nitrate (NO₃), le nitrite (NO₂), la demande chimique en oxygène (DCO), le résidu sec (RS), l'oxygène dissous (O₂), le taux de saturation en oxygène (O₂%) et le phosphore total (PT) et phosphate (PO₄⁻²). La sortie de modèles est la demande biochimique en oxygène (DBO₅).

IV.3.3.2.2. Implémentation des modèles ANFIS

L'implémentation des modèles ANFIS et les paramètres d'apprentissage sont déterminés de la même façon que la partie simulation de la de chlorophylle-a par les modèles ANFIS. Le nombre d'itérations a été fixé à 100. La performance des modèles (Goodness of the models fit) a été évaluée à l'aide du coefficient de corrélation (CC) et de l'erreur quadratique moyenne (RMSE).

Un modèle ANFIS avec cinq variables d'entrée est illustré schématiquement sur la figure 50, cette figure représente toutes les composantes du modèle ANFIS.

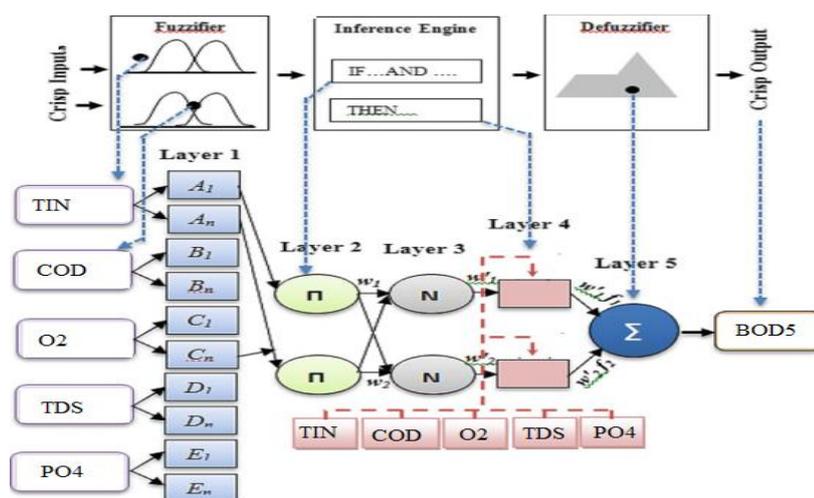


Figure 50 Modèle ANFIS avec 5 entrées (Adaptée de Hosseini et al., 2012)

IV.3.3.2.3. Résultats de simulation

Le tableau 16 montre les résultats lors de l'apprentissage et de la validation des modèles ANFIS retenus. Les meilleures valeurs du rayon d'influence (radius) pour chaque modèle sont également enregistrées. Dans ce tableau, les valeurs en gras indiquent des résultats acceptables.

A l'exception des modèles M_1 et M_9 , qui apportent des résultats moyens, les autres modèles d'ANFIS ont une performance élevée en apprentissage et en validation, alors les critères de performance varient de 0,91 à 0,97 pour le CC et de 0,04 mgO_2/L à 0,06 mgO_2/L pour le RMSE. Les résultats enregistrés avec ces modèles sont comparables avec ceux obtenus par **Ahmed et al. (2017)** où ces chercheurs ont testé la capacité des modèles ANFIS pour estimer la DBO_5 .

Les résultats obtenus par les modèles ANFIS sélectionnés en gras (tableau) donnent de résultats meilleurs que ceux fournis par **Noori et al. (2013)** où les critères de performance donnés par ces auteurs sont respectivement de 0,86 et 11,03 mg / L pour CC et RMSE, en phase de validation. Cela confirme la robustesse, la stabilité et l'efficacité des modèles ANFIS développés (modèle en gras).

Il ressort du tableau 16 que le modèle M_9 d'ANFIS qui utilise uniquement la demande chimique en oxygène (DCO) comme entrée enregistre des résultats moyens en apprentissage et de bons résultats en validation, alors que le CC varie de 0,7807 à 0,9238 et RMSE oscille de 0,1376 mg/L à 0,0674 mg/L pour les deux phases, respectivement, ce qui montre que la DCO est la variable la plus influente sur la DBO_5 par rapport aux autres variables, de sorte que la DCO est le principal facteur de forçage pour la modélisation de la DBO_5 par le biais d'ANFIS.

De point de vue parcimonie, les modèles M_6 , M_7 et M_8 sont plus parcimonieux par rapport aux autres modèles puisqu'ils apportent de très bons résultats en utilisant deux variables d'entrées seulement. Ces modèles d'ANFIS sont plus performants que ceux de réseaux de neurones artificiels correspondants.

Le modèle M_1 qui possède toutes les variables d'entrées disponibles, à l'exception des la DCO, montre qu'on peut modéliser la DBO_5 avec une précision moyen via la technique d'ANFIS sans faire appel à la demande chimique en oxygène (DCO).

Tableau 16 Critères de performance des modèles ANFIS élaborés pour la simulation du la DBO_5

Modèles	Variables d'entrée	Choix des entrées	Radius	R		RMSE (mgO ₂ /L)	
				Apprentissage	Validation	Apprentissage	Validation
M_0 *** (modèle de base)	RS DCO NH4 NO3 NO2 TIN O2 O2p Ph Ptot PO ₄ MO.	Toutes les variables d'entrées disponibles	<u>0,84</u>	0,9772	0,9349	0,0466	0,0536
M_1 *	RS NH4 NO3 NO2 TIN O2 O2p Ph Ptot PO ₄ MO	Toutes les variables d'entrées à l'exception de DCO	0,19	1	0,7237	1,6210*10 ⁻⁷	0,1351
M_2 ***	TIN, DCO, O2%, RS, PO4.	Toutes les variables As my paper	<u>0,47</u>	0,9785	0,9447	0,0453	0,0478
M_3 ***	DCO NH4 NO3 NO2 TIN	DCO et composés azotés	<u>0,33</u>	0,9682	0,9522	0,0551	0,0498
M_4 ***	DCO PO4 Ptot	DCO et Composées phosphatée	<u>0,25</u>	0,9705	0,9350	0,0530	0,0587
M_5 ***	DCO TIN Ptot	DCO et azote inorganique et Phosphore total	<u>0,44</u>	0,9652	0,9475	0,0575	0,0602
M_6 ***	DCO MO	Deux variables	<u>0,24</u>	0,9660	0,9453	0,0570	0,0614
M_7 ***	DCO TIN	Deux variables	<u>0,31</u>	0,9652	0,9459	0,0576	0,0570
M_8 ***	DCO PO4	Deux variables	<u>0,18</u>	0,9715	0,9198	0,0522	0,0664
M_9 **	DCO	Une seule variable	<u>0,17</u>	0,7807	0,9238	0,1376	0,0674

Les figures 51, 52 et 53 visualisent les résultats obtenus avec le modèle ANFIS-M0. Sur ces figures, les estimations de la DBO₅ du modèle sont représentées sous forme de séries temporelles (à gauche) et de nuages de points (à droite) pour ; l'ensemble complet des données (a1 et b1), durant l'apprentissage (a2 et b2) et pendant la validation (a3 et b3), respectivement. Les séries temporelles et les diagrammes de dispersion montrent clairement que les estimations du modèle M0 d'ANFIS sont plus proches des valeurs observées de la DBO₅ correspondante.

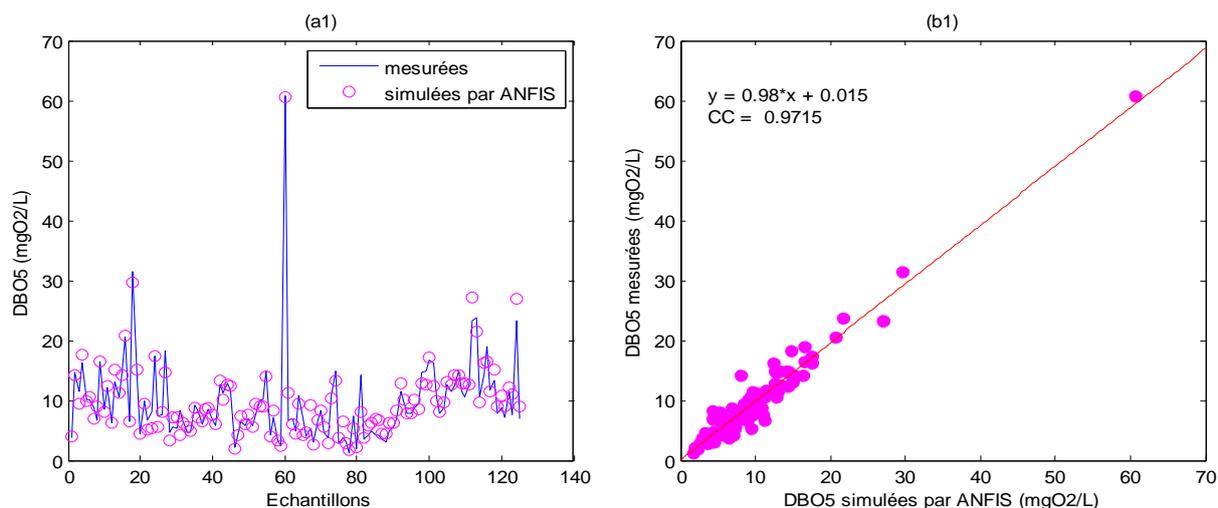


Figure 51 (a1) Valeurs mesurées et simulées de DBO₅ (mg O₂/L); (b1) Diagramme de dispersion des valeurs simulées et mesurées de la DBO₅ (mg O₂/L) pour l'ensemble de données en utilisant le modèle ANFIS-M0.

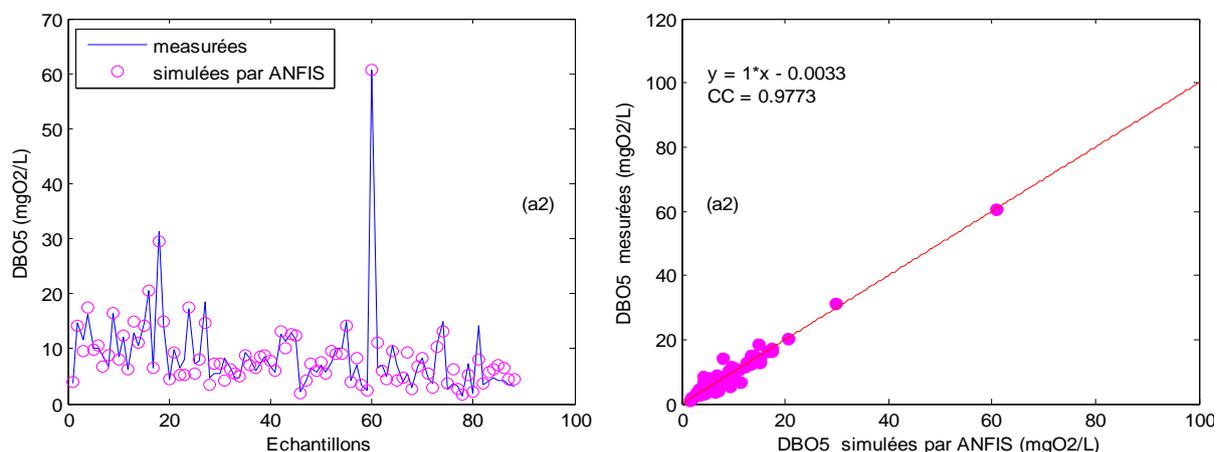


Figure 52 (a2) Valeurs mesurées et simulées de DBO₅ (mg O₂/L); (b2) Diagramme de dispersion des valeurs simulées et mesurées de la DBO₅ (mg O₂/L) lors de l'apprentissage en utilisant le modèle ANFIS-M0

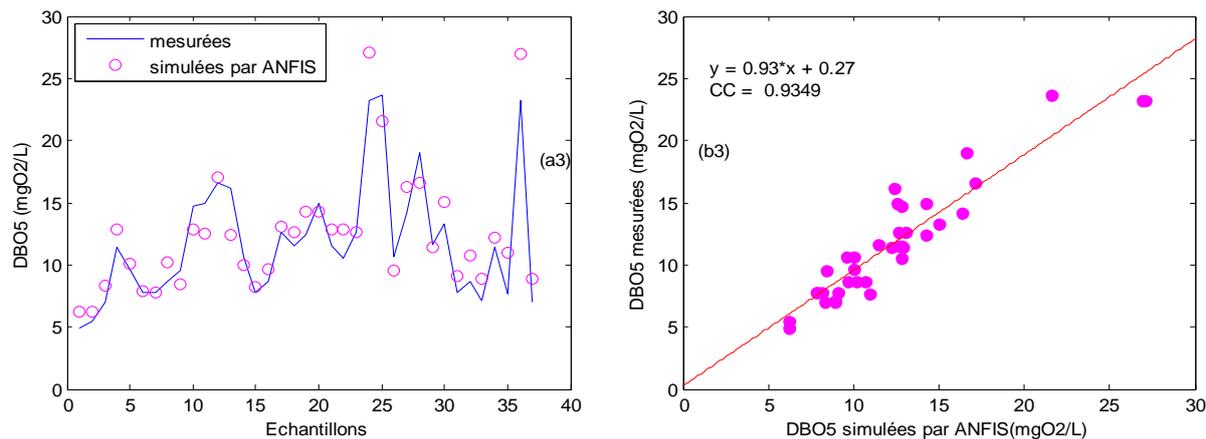


Figure 53 (a3) Valeurs mesurées et simulées de la DBO5 (mg O2/L); (b3) Diagramme de dispersion des valeurs simulées et mesurées de la DBO5 (mg O2/L) durant la phase de validation en utilisant le modèle ANFIS-M0

IV.3.3.2.4. La comparaison entre les modèles ANFIS et les modèles de RNAs

Un regard attentif dans les tableaux 12 et 16 révèle que tous les modèles ANFIS surpassent les modèles de RNAs correspondants, parce que les modèles ANFIS ont donné un coefficient de corrélation (CC) plus élevé, des erreurs quadratiques moyennes (RMSE) plus faibles que les modèles de RNAs correspondants. Nous pouvons conclure que les modèles ANFIS ont la supériorité sur les modèles de RNA en termes de critères de performance et de précision de modélisation de la DBO₅. Dans l'ensemble, les modèles ANFIS ont donné des performances significativement meilleures que les modèles de RNA correspondants en se basant sur les CC, RMSE. Ces résultats étaient attendus parce que les modèles ANFIS recèlent l'interprétabilité linguistique des systèmes d'inférence flous et la capacité d'apprentissage des RNAs.

IV.3.3.2.5. Conclusion

La détermination précise de la demande biochimique en oxygène (DBO₅) devient aujourd'hui très importante et cruciale dans la gestion et la planification de la qualité de l'eau (Noori et al., 2013). Cependant, la mesure au laboratoire de la demande biochimique en oxygène (DBO₅) est complexe, fastidieuse et chronophage. Le recours à des méthodes indirectes comme la modélisation constitue une alternative prometteuse.

Cette partie présente une technique d'intelligence artificielle pour évaluer la DBO₅ comme alternative des mesures au laboratoire (en cas d'absence de valeurs mesurées de DBO₅). Pour ce faire, le potentiel des modèles ANFIS pour l'estimation de la DBO₅ en utilisant les variables de qualité de l'eau a été examiné. L'avantage principal des modèles ANFIS est de combiner la capacité d'apprentissage des réseaux de neurones et l'interprétation linguistique de la logique floue dans un cadre unique.

Sur la base des résultats mentionnés ci-dessus, les conclusions suivantes peuvent être tirées: Les systèmes d'inférence neuro-floue adaptative peuvent être utilisés comme outil pour modéliser la demande biochimique en oxygène avec une grande fiabilité.

Les modèles ANFIS développés ont fournis des valeurs plus proches aux valeurs mesurées de la DBO₅ (modèles sélectionnés en gras dans le tableau 16). Les modèles ANFIS se sont avérés simples, efficaces et robustes pour modéliser la DBO₅ dans les réservoirs.

Les résultats de la simulation démontrent la supériorité des modèles ANFIS sur les modèles de RNAs en termes de critères de performance, ce qui signifie que le modèle ANFIS est plus fiable que le modèle de RNAs pour la modélisation de la demande biochimique en oxygène (DBO₅).

La DCO est la variable la plus effective sur la modélisation de la DBO₅. Les modèles ANFIS dont l'entrée est la DCO ont une bonne précision pour l'estimation de la DBO₅ en utilisant un nombre limitée de variable de qualité de l'eau (deux variables suffisent). L'étude suggère que la DBO₅ pourrait être calculée à partir de la DCO en utilisant la technique d'ANFIS. Par conséquent, le DCO doit être utilisé comme variable commune pour chaque combinaison d'entrées (ensemble de variables d'entrée) dans d'autres études que nous voudrions réaliser.

Tandis que ces modèles ANFIS devront être testés contre d'autres modèles et avec des ensembles de données contenant en plus de DCO d'autres variables telles que la matière en suspension, la température de l'eau, le pH, la turbidité et la chlorophylle-a, et provenant de différentes régions géographiques. Nous pensons que le modèle ANFIS est un outil prometteur pour aider à comprendre les processus et la dynamique des écosystèmes aquatiques complexes (rivières, lacs et réservoirs), mais il semble que le principal problème est d'obtenir des données expérimentales (en quantité et en qualité) permettant l'ajustement du modèle ANFIS et le problème de la méthode de l'essai et l'erreur (trials-and-errors) utilisée pour déterminer certains paramètres d'ANFIS.

Conclusion générale

Au cours de ces dernières années, les chercheurs ont déployé des efforts colossales pour tenter de répondre aux défis scientifiques et technologiques que posent la compréhension et la prévision du comportement des écosystèmes aquatiques en réponse à des perturbations naturelles (changement climatique) ou artificielles (pollution).

La réponse biologique de l'écosystème de retenue de barrage à des charges polluantes est compliquée, ce qui rend la compréhension de la relation nutriments-eutrophisation difficile.

Dans ce contexte, la modélisation fournit des outils permettant de comprendre, gérer et simuler le phénomène de l'eutrophisation, elle permet également d'interpréter comment un réservoir réagit aux scénarios de protection contre l'eutrophisation.

Les modèles de l'eutrophisation sont probablement la catégorie des modèles écologiques qui ont suscité beaucoup d'effort et de considération durant ces 25 dernières années, plusieurs modèles ont été développés avec un large spectre de complexité, celle-ci dépend de la disponibilité des données et de l'écosystème.

L'émergence des techniques d'intelligence artificielle a donné de nombreux résultats encourageants dans le domaine de ressources en eau et en particulier dans la gestion et la modélisation de la qualité de l'eau.

Les objectifs principaux de cette thèse sont de (1) développer des modèles basés sur les RNAs, et les systèmes d'inférence adaptative neuro-flous adaptatifs (ANFIS) capables de modéliser les indicateurs de l'eutrophisation dans certains barrages au Nord de l'Algérie et (2) démontrer la capacité et l'applicabilité de ces techniques dans les questions pratiques de la gestion de la qualité de l'eau et l'eutrophisation (3) Comparer les modèles développés et choisir le plus performant.

Bien que nous soyons pleinement conscients de la complexité de l'écosystème aquatique de retenue de barrage et en particulier du comportement et de la dynamique du phytoplancton, nous avons tenté de modéliser le mécanisme de l'eutrophisation, en utilisant les données de qualité de l'eau disponibles et certaines techniques de l'intelligence artificielle.

Les modèles de RNAs et d'ANFIS ont été appliqués à la modélisation de certains indicateurs de l'eutrophisation (Chl-a, PT et DBO₅) en utilisant des mesures mensuelles. Le travail de recherche présenté dans ce manuscrit de thèse nous a permis de tirer les conclusions suivantes :

L'analyse de sensibilité des variables de la qualité de l'eau par l'utilisation des réseaux de neurones artificiels a révélé que le facteur limitant de la production de la biomasse algale dans les écosystèmes aquatiques de retenues de barrages étudié est le phosphore (phosphore total et orthophosphate). Donc chaque travail qui porte sur la lutte contre l'eutrophisation doit obligatoirement prendre en compte la réduction de la charge en phosphore déversé dans les barrages.

Malgré l'hétérogénéité de données d'apprentissage (données mixte provenant de trois barrages), les modèles de RNAs développés pour simuler la chlorophylle-a peuvent exprimer entre 68 % à 83 % de la variabilité de la concentration de la chlorophylle-a, cette dernière est un proxy pour déterminer la biomasse algale.

Ces résultats révèlent que les meilleures performances sont observées quand on inclut toutes les variables de la qualité de l'eau disponible en entrée de modèles neuronaux. Du point de vue pratique on peut utiliser les modèles neuronaux alternatifs cités auparavant pour évaluer la chlorophylle-a en utilisant un nombre limité de variables de qualité de l'eau.

Les modèles RNAs développés sont capables de simuler assez bien la dynamique de concentration du phytoplancton (exprimé en terme de chlorophylle-a) en utilisant des variables d'entrée minimales même si des facteurs inconnus affectaient la qualité de l'eau de barrages et des séries de données limitées. Les résultats montrent que le comportement complexe du processus d'eutrophisation pourrait être modélisé en utilisant la technique de RNAs, et nous avons estimé avec succès quelques valeurs extrêmes à partir des ensembles de données de validation qui n'ont pas été utilisés dans l'apprentissage du réseau de neurones. Le modèle développé peut être utilisé pour (1) estimer la concentration de Chl-a lorsque la valeur réelle est manquante (lacunes), (2) estimer les données interpolées entre deux échantillons consécutifs. Cette approche de modélisation de l'eutrophisation est utile pour simuler les niveaux de chlorophylle-a à n'importe quel endroit ou moment dans le domaine d'intérêt.

Ces modèles neuronaux fournissent des informations précieuses pour les décideurs afin de prendre les mesures et les actions appropriées dans le but de préserver la qualité de l'eau des barrages. L'étude de cas pour la base de données mélangée démontre une large gamme d'applicabilité des modèles de réseaux de neurones artificiels. L'enrichissement de bases de données permettra de recalibrer ces modèles pour améliorer, encore une fois, la qualité de modélisation.

Les modèles de RNAs sont également capables de simuler, avec une précision élevée, certains indicateurs de l'eutrophisation tels que le phosphore total et la demande biochimique en

oxygène. Des modèles simples basés sur les RNAs sont développés et appliqués avec succès pour quantifier les concentrations de phosphore total (TP) et la DBO₅.

Nous avons également examiné le potentiel des systèmes d'inférence neuro-flous adaptatifs (ANFIS) pour la modélisation de l'eutrophisation en utilisant les variables de qualité de l'eau disponibles.

Les modèles ANFIS présentent un grand avantage pour la simulation de la concentration en chlorophylle-a, pratiquement, pour les modèles sélectionnés en gras avec ***, les valeurs du coefficient de corrélation en validation sont supérieures à 0.89, ce qui suppose une bonne aptitude des modèles neuro-flous à simuler la concentration en chlorophylle-a et reconstituer ses valeurs manquantes existantes dans les séries de la qualité de l'eau. Cela confirme la robustesse, la stabilité et l'efficacité des modèles ANFIS développés (modèle en gras).

Les modèles ANFIS, avec le plus petit nombre d'entrées (modèles M5 et M4), offrent une précision plus élevée que les RNAs. Ils sont plus parcimonieux, parce que les meilleurs résultats sont obtenus avec les modèles ANFIS qui utilisent un nombre de variable d'entrées moindre, six variables d'entrée pour les deux modèles.

Les résultats de l'étude indiquent également que les modèles ANFIS permettent de simuler la concentration en phosphore total avec une précision élevée.

Les modèles RNAs et ceux d'ANFIS apportent des performances similaires concernant la simulation de la concentration du phosphore total. Dans l'ensemble, les modèles ANFIS et ceux de RNAs (sélectionnés en gras) constituent des outils appropriés pour simuler le comportement et la dynamique de la qualité de l'eau des réservoirs.

Le potentiel des modèles ANFIS pour l'estimation de la DBO₅ en utilisant les variables de qualité de l'eau a été examiné. Les modèles ANFIS ont fournis des résultats très satisfaisants, et se sont avérés simples, efficaces et robustes pour modéliser la DBO₅ dans les réservoirs.

Nous pensons que le modèle ANFIS est un outil prometteur pour aider à comprendre les processus et la dynamique des écosystèmes aquatiques complexes (réservoirs).

Les résultats de la simulation révèlent la supériorité des modèles ANFIS sur les modèles de RNAs en termes de critères de performance, ce qui signifie que le modèle ANFIS est plus fiable que le modèle de RNAs pour la modélisation de la demande biochimique en oxygène (DBO₅).

Les travaux de recherche présentés dans ce rapport de thèse ont permis de fournir des résultats encourageants et révèlent que les outils basés sur l'intelligence artificielle sont très utiles pour modéliser la qualité de l'eau et peuvent contribuer de manière significative pour une amélioration constante de la gestion de la qualité des eaux de barrages.

Malgré que plusieurs facteurs contrôlant la variation de la qualité soient inconnus et que les processus complexes d'eutrophisation ne sont pas encore entièrement compris, les résultats obtenus sont très satisfaisants, en se référant à la bonne concordance entre les valeurs observées et celles simulées par les modèles développés.

Les modèles de réseaux de neurones artificiels et d'ANFIS contribuent au développement des outils d'aide à la décision pour la gestion de la qualité de l'eau des réservoirs, ils appuieront la prise de décision par la simulation, la prévision de processus d'eutrophisation des eaux de retenues de barrages.

Les modèles de l'eutrophisation développés peuvent être divisés en 2 catégories :

1. Modèles de dépistage:

Les modèles de dépistage sont fondés sur des données de terrain pour l'évaluation de l'état d'eutrophisation, ils permettent de :

- ✓ Encapsuler les processus complexes d'eutrophisation dans un cadre simplifié sous forme de paramètres ajustables incorporés dans ces modèles.
- ✓ Evaluer l'état de l'écosystème en se basant sur quelques paramètres mesurés (évaluer l'état de l'eutrophisation).
- ✓ Comblent les lacunes dans l'historique des données.
- ✓ Ils constituent le lien entre la collecte de données, l'interprétation et la gestion de l'écosystème aquatique.
- ✓ Ils sont utilisés par les gestionnaires pour comprendre le fonctionnement de l'écosystème et faire des comparaisons.

2. Modèles de recherche:

- ✓ Ces modèles aident à explorer des scénarii spécifiques et constituent un outil utile pour étudier les réponses écologiques (en termes de production primaire phytoplanctonique ou biomasse algale) de l'écosystème aquatique aux changements de pressions (pollution).

Pour continuer d'avancer sur cette voie, les points suivants seront recommandés pour de futures recherches:

- ✓ Pour une modélisation plus fine de l'eutrophisation des retenues de barrages, des intervalles d'échantillonnage plus courts (plus faibles pas de temps) sont recommandés pour améliorer les résultats de la modélisation, puisque les intervalles plus longs entre les données pour les variables de qualité de l'eau réduisent la capacité d'enregistrer la variation de la qualité de l'eau entre les intervalles de surveillance.

- ✓ Pour mieux comprendre la dynamique de la biomasse algale, il est recommandé d'observer quotidiennement la concentration chlorophylle-a ainsi que les autres variables de la qualité de l'eau afin de ne pas perdre de l'information précieuse pour la modélisation.
- ✓ Le champ de l'intelligence artificielle est très vaste, il englobe plusieurs techniques dont on peut citer les réseaux de neurones artificiels, la logique floue, et les algorithmes génétiques, le Support Vector machine, il nous semble très intéressant d'exploiter ces deux dernières techniques dans le domaine de la gestion et la modélisation de la qualité de l'eau.

Références bibliographiques

- Abbasi T., Abbasi SA., 2012.** Water Quality Indices. Oxford, UK: Elsevier, DOI: 10.1016/B978-0-444-54304-2.00001-4. ISBN: 978-0-444-54304-2.
- Afshar A and Saadatpour M., 2009.** "Reservoir eutrophication modeling, sensitivity analysis, and assessment: Application to Karkheh Reservoir, Iran", *Journal of Environmental Engineering Science*, 26, 1227-1238.
- Aghaarabi E., Aminravan F ., Sadiq R., Hoorfar M., Rodriguez M. J & Najjaran H., 2017.** Application of neuro-fuzzy based expert system in water quality Assessment. *Int J Syst Assur Eng Manag*, 8, 2137–2145.
- Ahmed A.A.M & Shah S.M.A., 2017.** Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River. *Journal of King Saud University – Engineering Sciences*, 29, 237–243.
- Ahn C.Y., Joung S.H., Yoon S.K., Oh H.M., 2007.** Alternative alert system for cyanobacterial bloom, using phycocyanin as a level determinant. *The Journal of Microbiology*, 45, 98–104.
- Alves M.C., Pozza A.E., João C.B.C., Luiz G.C., Alves LS., 2011.** Adaptive neuro-fuzzy inference systems for epidemiological analysis of soybean rust. *Environmental Modelling & Software* 26:1089-1096.
- Angulo C., Cabestany J., Rodríguez P., Batlle M., González A & de Campos S., 2012.** Fuzzy expert system for the detection of episodes of poor water quality through continuous measurement. *Expert Systems with Applications*, 39, 1011–1020.
- Ansari A.A., Singh S., Gill G.R., Rast L.W., 2010.** Eutrophication: causes, consequences and control. volume1. 10.1007/978-90-481-9625-8_7.
- Ansari. A. A., Gill. S. S., 2013.** Eutrophication: Causes, Consequences and Control. Volume2. ISBN 978-94-007-7813-9.
- Aouchar N & Bekhaled C., 2005.** Application des systèmes hybrides neuro-flous au diagnostic des huiles de transformateurs. Mémoire d'ingénieur, ENP, El Harrach, Alger.
- Araghinejad S., 2014.** Data-driven modeling: using MATLAB® in water resources and environmental engineering. Springer Science & Business Media.
- Ay M & Kisi O., 2016.** Estimation of dissolved oxygen by using neural networks and neuro-fuzzy computing techniques. *KSCE J Civ Eng*, 21(5), 1631–1639. doi:10.1007/s12205-016-0728-6.

- Ay M., Kisi O., 2014.** Modelling of chemical oxygen demand by using ANNs, ANFIS and k-means clustering techniques. *Journal of Hydrology* 511: 279–289.
- Babovic V., 2009.** Introducing knowledge into learning based on genetic Programming, *J. Hydroinform.* 2 (1), 35–60.
- Baird M.E., Oke P.R., Suthers I.M., Middleton J.H., 2004.** A plankton population model with biomechanical descriptions of biological processes in an idealized 2D ocean basin. *Journal of Marine Systems*, 50, 199–222.
- Belabbes H., 2012.** Étude du phénomène d'eutrophisation dans le barrage de Cheurfa II (wilaya de Mascara). Thèse magistère (USTHB), 109p.
- Belouz K., Aidaoui A., Dechemi N., Heddami S and Aguenini S., 2018.** Modelling of biochemical oxygen demand from limited water quality variable by ANFIS using two partition methods. *Water Quality Research Journal*. doi: 10.2166/wqrj.2017.015.
- Benfeta H et Remini B., 2008.** Les fuites d'eau à travers le barrage algérien de Ouizert. *Sécheresse*.19, 185-92
- Benkaci A.T., 2006.** Modélisation pluie-debit mensuelle et journalière par les modèles conceptuels et les systèmes neuro-flous (application aux bassins algériens). Thèse doctorat, INA. EL-Harrach, 242 p.
- Bensafia D et Remini B., 2014.** Le rôle de la vase dans l'accélération de l'eutrophisation des eaux de barrages. Etude expérimentale. *Larhyss Journal*, ISSN 1112-3680, n°19, pp. 161-181.
- Best G., 1999.** *Environmental Pollution Studies*, 167 pp. Liverpool University Press.
- Bouzaïd-Lagha S et Djelita B., 2012.** Etude du phénomène d'eutrophisation dans le Barrage de Hammam Boughrara (Wilaya de Tlemcen, Algérie). *Hydrological Sciences Journal*, 57, 186-201.
- Boyd C.E., 2015.** *Water Quality. An Introduction*. Springer Science & Business Media.
- Brigitte E et Jérôme P., 2008.** *Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation*. Dunod. 4th ed, Dunod, Paris, ISBN: 978-2-10-053809-6.
- Carbajal J., Sanchez-Fernandez P.L., Carrasco-Ochoa J.A., Martinez-Trinidad F.J., 2012.** Immediate water quality assessment in shrimp culture using fuzzy inference systems. *Expert Systems with Applications* 39, 10571–10582.
- Carlson R.E., 1977.** A trophic state index for lakes. *Limnology and Oceanography* 22:361-369. Cole, Gerald A. 1983. *Textbook of Limnology*. The C.V. Mosby Company. St. Louis, Missouri.
- Chapra S.C., 1997.** *Surface Water-Quality Modeling*. McGraw-Hill, New York, NY, 845 pp.

- Charalampos D., Pantazis G., Dimitris P., et al 2012.** Ecosystem approach to water resources management using the MIKE11 modeling system in the Strymonas River and Lake Kerkini. *J. Environ. Manag.* **94**, 132–143.
- Chavres P et Koriji T., 2007.** Conceptual fuzzy neural network model for water quality simulation. *Hydrological processes*, 21, 634-646.
- Chen C.W., 1970.** Concepts and utilities of ecological models. *Proc. Am. Soc. Civ. Eng., J. San. Eng. Div.* 96 (SA5): 1085-1097.
- Chen H.M & Lo S.L., 2010.** Prediction of the effluent from a domestic wastewater treatment plant of CASP using gray model and neural network. *Environmental Monitoring and Assessment*, 162(1-4), 265-75.
- Chen Q., Ouyang Z., 2005.** Integrated ecohydraulics model and the application. *Journal of Hydraulic Engineering*, 36(11): 1273–1279.
- Chen Q.W., Morales-Chaves Y., Li H & Mynett A.E., 2006.** Hydroinformatics techniques in eco-environmental modelling and management. *Journal of Hydroinformatics*, 297-316.
- Chen Q.W., Morales-Chaves, Y., Li, Hong., 2006.** Hydroinformatics techniques in eco-environmental modelling and management. *Journal of Hydroinformatics*, 297-316.
- Chen, Q., Mynett, A.E., 2004.** A robust fuzzy logic approach to modelling algal biomass. *Journal of Hydraulic Research*, 42: 303–309.
- Chen, Q.W., Mynett, A.E., 2003.** Integration of data mining techniques and heuristic knowledge in fuzzy logic modelling of eutrophication in Taihu Lake. *Ecological Modelling* 162 (1–2), 55–67.
- Chiu S., 1994.** Fuzzy model identification based on cluster estimation. *J Intell Fuzzy Syst* **2**:267–278.
- Cobaner M., 2011.** Evapotranspiration estimation by two different neuro-fuzzy inference systems. *Journal of Hydrology* 398:292–302.
- Cole T.M and Wells S.A., 2003.** Instruction Report EL-03-1. CE-QUAL-W2: A Two-Dimensional, Laterally Averaged, Hydrodynamic and Water Quality Model. Version 3.2. User Manual. U.S. Army Corps of Engineers, Washington, DC 20314-1000.
- Cowen W.F. and Lee G.F., 1976.** *Algal nutrient availability and limitation in Lake Ontario during IFGYL.* Part 1. Office of Research and Development, U.S. Environmental Protection Agency, Duluth, MN, EPA-600/3-76-094a:218.
- Cox B.A., 2003.** A review of currently available in-stream water quality models and their applicability for simulating dissolved oxygen in lowland rivers. *Sci Total Environ* 314–316:335–377.

- Csábrági A., Molnár S., Tanos P & Kovács J., 2016.** Application of artificial neural networks to the forecasting of dissolved oxygen content in the Hungarian section of the river Danube. *Ecological Engineering*, 100, 63–72.
- Daniel P. Loucks et Eelco van Beek., 2005.** Water resources systems planning and management. An Introduction to Methods, Models and Applications, P380.
- David E., Goldberg J., Koza R., 2006.** Adaptive Learning of Polynomial Networks. *Genetic Programming, Backpropagation and Bayesian Methods*. 328p.
- Demetris K., 2007.** Discussion of “Generalized regression neural networks for evapotranspiration modelling”. *Hydrological Sciences–Journal*, Vol 51, Issue 6.
- Dennis Cooke G., Welch Eugene B., Peterson Spencer A., Nichols Stanley A. 2005.** Restoration and management of lakes and reservoirs. ISBN 0-7484-0772-3 (alk. paper).
- Di Toro, D.M. and Matysik W.F., Jr., 1980.** Mathematical models of water quality in large lakes, Part I: Lake Huron and Saginaw Bay. U.S. Environmental Protection Agency, Ecological Research Series, EPA-600.
- Diaz, R. J., and R. Rosenberg. 2008.** Spreading dead zones and consequences for marine ecosystems. *Science*, 321, 926-929
- Dillon P.J. and Rigler F.H., 1974.** A test of a simple nutrient budget model predicting the phosphorus concentration in lake water. *J. Fish. Res. Board Can.* 31, 1771-1778.
- Dillon P.J. and Kirchner W.B., 1975.** The effects of geology and land use on the export of phosphorus from watersheds. *Water Res.* 9, 135-148.
- Dimberg P.H., Hytteborn J.K., Bryhn A.C., 2012.** Predicting median monthly chlorophyll-a concentrations, *Limnologica*. <http://dx.doi.org/10.1016/j.limno.2012.08.011>.
- Dodds W. K., Bouska W.W., Eitzmann J. L., Pilger T. J., Pitts K.L., Riley A. J., Schloesser J. T., and Thornbrugh D.J., 2008.** Eutrophication of US freshwaters: analysis of potential economic damages. *Environmental Science & Technology*, 43, 12–19.
- Dogan E., Ates A., Yilmaz E.C., 2008.** Application of Artificial Neural Networks to Estimate Wastewater Treatment Plant Inlet Biochemical Oxygen Demand. *Environmental Progress* 27 (4): 439-446.
- Dogan E., Sengorur B., Koklu R., 2009.** Modeling biochemical oxygen demand of the Melen River in Turkey using an artificial neural network technique. *J Environ Manag*, 90, 1229–1235.
- Dreyfus G., Samuelides M., Martinez J., Gordon M., Badran F., Thiria S., Herault L., 2004.** Réseaux de neurones - Méthodologies et applications. Editions Eyrolles, 200 p.

- Driankov D., Hellendoorn H and Reinfrank M., 1993.** An introduction to fuzzy control, Springer-Verlag.
- Eamus D., Hatton T., Cook P. & Colvin C., 2006.** *Ecohydrology: Vegetation Function, Water and Resource Management*, 1, CSIRO Publishing, Collingwood, VIC Australia.
- Einax J.W., Aulinger A., Timpling W.V., Prange A., 1999.** Quantitative description of element concentrations in longitudinal river profiles by multiway PLS models. *Fres J Anal Chem* 363:655–661
- Elshemy M. M., 2010.** Water quality modeling of large reservoirs in semi-arid regions under climate change –Example Lake Nasser (Egypt). Thèse doctorat à l'Université Technique Carolo Wilhelmina à Brunswick, Allemagne.
- Emamgholizadeh S., Kashi H., Marofpoor I & Zalaghi E., 2014.** Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *Int. J. Environ. Sci. Technol*, 11(3), 645–656.
- Eric G., 2000.** Un parcours dans l'étude des phénomènes extrêmes en hydrologie. Mémoire d'habilitation à diriger des recherches, 269p.
- Félix-Faure J., Chanudet V., Walter C., Dorioz J. M., Baudoin J.M., et al., 2017.** Evolution des sols ennoyés sous les retenues de barrage : Influence sur l'écologie des plans d'eau et la dynamique des gaz à effet de serre. *Etude et Gestion des Sols*, Association française pour l'étude des sols, 24 (1), pp.45-58.
- Forsberg C., Ryding S.O., Claesson A. and Forsberg A., 1978.** Water chemical analysis and/or algal assay-Sewage effluent and polluted lake water studies. *Mitt. Int. Ver. Limnol.* 21, pp. 353-263.
- French, M., Recknagel F., 1994.** Modeling algal blooms in freshwaters using artificial neural networks. In: Zanetti, P. (Ed.), *Computer Techniques in Environmental Studies V*, vol. II: Environment Systems. Computational Mechanics Publications, Boston, pp. 87–94.
- French, M., Recknagel, F., Jarrett, G.L., 1998.** Scaling issues in artificial neural network modelling and forecasting of algal bloom dynamics. In: Abt, S.R., Young-Pezeshk, J., Watson, C.C. (Eds.), *Proceedings of the International Water Resources Engineering Conference*, ASCE, August 3/7, 1998, Memphis, Tennessee, 1, pp. 891-896.
- Fusco J., 2016.** Analyse des dynamiques spatio-temporelles des systèmes de peuplement dans un contexte d'incertitude : Application à l'archéologie spatiale. Géographie. Université Nice Sophia Antipolis.

- Gazzaz N.M., Yusoff M.K., Aris A.Z., Juahir H & Ramli M.F., 2012.** Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine Pollution Bulletin*, 64, 2409–2420.
- Gharibi H., Mahvi A.H., Nabizadeh R., Arabalibeik H., Yunesian M., Sowlat, M.H., 2012.** A novel approach in water quality assessment based on fuzzy logic, *Journal of Environmental Management* 112, 87-95.
- Glibert P.M., 2007.** Eutrophication and Harmful Algal Blooms: A Complex Global Issue, Examples from the Arabian Seas including Kuwait Bay, and an Introduction to the Global Ecology and Oceanography of Harmful Algal Blooms (GEOHAB) Programme. *International Journal of Oceans and Oceanography*, ISSN 0973-2667 Volume 2 Number 1 (2007), pp. 157–169.
- Ha H., Stenstrom M.K., 2003.** Identification of land use with water quality data in stormwater using a neural network. *Water Research* 37 (17), 4222–4230.
- Haggard B. E., Moore P. A., Daniel T. C., Edward D.R., 1999.** Trophic conditions and gradients of the Headwater reaches of Beaver Lake, Arkansas. *Proceedings of the Oklahoma Academy of Science*, 79, 73-84.
- Hamaidi M.S., Hamaidi F., Zoubiri A., Benouaklil F., Dhan Y., 2009.** Etude de la Dynamique des Populations Phytoplanctoniques et Résultats Préliminaires sur les Blooms Toxiques à Cyanobactéries dans le Barrage de Ghrib (Ain Defla-Algérie). *European Journal of Scientific Research*, 32, 369-380.
- Hamlet A., 2014.** Contribution à la gestion des ressources en eau des bassins versants de l'Ouest Algérien à l'aide d'un système informatisé. Thèse doctorat à l'université d'Oran. 243p.
- Han H.G., Chen Q., Qiao JF., 2011.** An efficient self-organizing RBF neural network for water quality prediction. *Neural Networks*, 24, 717–725.
- Harper D., 1992.** Eutrophication of Freshwaters: Principles, problems and restoration. St Edmundsbury Press, Bury St Edmunds, Suffolk. ISBN 978-94-010-5366-2.
- Hebb D.O., 1949.** *The organization of behaviour*. New York: J. Wiley.
- Heddam S., Bermad A, Dechemi N., 2011.** ANFIS-based modelling for coagulant dosage in drinking water treatment plant: a case study. *Environ Monit Assess*, 184:1953–1971.
- Heddam S., 2006.** Contribution à la modélisation de la qualité des eaux par les réseaux de neurones. Thèse magister, INA. EL-Harrach, 95p.
- Heddam S., 2017.** Fuzzy Neural Network (EFuNN) for Modelling Dissolved Oxygen Concentration (DO). *Intelligence Systems in Environmental Management. Theory and Applications*, 231-253.

- Heddam S., Hilal L & Samir F., 2016.** Predicting Effluent Biochemical Oxygen Demand in a Wastewater Treatment Plant Using Generalized Regression Neural Network Based Approach: A Comparative Study. *Environmental Processes*, 3(1), 153–165.
- Heiskary S. and Markus H., 2001.** Establishing Relationships among Nutrient Concentrations, Phytoplankton Abundance, and Biochemical Oxygen Demand in Minnesota, USA, Rivers. *Lake and Reservoir Management*, 17 (4), 251-262.
- Hosseini S M et al., 2012.** Numerical and Meta-Modeling of Nitrate Transport Reduced by Nano-Fe/Cu Particles in Packed Sand Column. *Transp Porous Med* 94:149–174.
- Howarth R.W and Marino R., 2006.** Nitrogen as the limiting nutrient for eutrophication in coastal marine ecosystems: Evolving views over three decades. *Limnol. Oceanogr.*, 51, 364-376.
- Hull V., Parrella L., Falcucci M., 2008.** Modeling dissolved oxygen dynamics in coastal lagoons. *Ecol Model*, 2, 468–480.
- Huo S., He Z., Su Jing, Xi B., Zhu C., 2013.** Using artificial neural network models for eutrophication prediction. *Procedia Environmental Sciences*. 18, 310 – 316.
- Jang J.S.R., 1993.** ANFIS: adaptive network based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics* 23(3):665–685.
- Jang J.S.R., Sun C.T., Mizutani E., 1997.** Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence. Prentice-Hall International (UK) Limited, London.
- Ji X., Shang X., Dahlgren R.A & Zhang M., 2017.** Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: a case study of Wen-Rui Tang River, China. *Environ Sci Pollut Res*, 24(19), 16062–16076.
- Ji Z.G ., 2005.,** Water quality models: chemical principles. in *Water Encyclopedia*. Water quality and resource development. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Ji Z.G. 2008.** Hydrodynamics and water quality: modeling rivers, lakes, and estuaries, 1st ed. John Wiley & Sons, Inc., Hoboken, New Jersey, ISBN 978-0-470-13543-3 (cloth), p 247.
- Jia H.F., Zhang Y.S., Guo Y., 2010.** The development of a multi-species algal ecodynamic model for urban surface water systems and its application. *Ecological Modelling* 221, 1831–1838.
- Jørgensen S.E and Bendoriccho G., 2001.** Fundamentals of Ecological Modelling, 3nd ed. Elsevier, Amsterdam, ISBN: 0-080-44028-2, pp. 520–527.
- Jorgensen S.E., Heinz Loffler., Walter Rast & Milan Straskraba. 2005.,** Lake and Reservoir Management, Volume 54. ISBN: 9780080535340.

- Jørgensen S.E., Tundisi J.G., and Tundisi T.M., 2013.** Handbook of Inland Aquatic Ecosystem Management. CRC Press Taylor & Francis Group 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742.
- Jung N.C. et Price R.K., 2005.** Applicability of Data-Driven Modelling For Reservoir Water Quality Management. XXXI IAHR Congress, 1, pp 174-175.
- Jung N.C., 2009.** Eco-hydraulic Modelling of Eutrophication for Reservoir Management. Thèse doctorat. CRC Press/Balkema PO Box 447, 2300 AK Leiden, The Netherlands.
- Kaneko N & Phu S.T.P., 2014.** Research on the Correlation Between Chlorophyll-a and Organic Matter BOD, COD, Phosphorus, and Total Nitrogen in Stagnant Lake Basins. *Sustainable Living with Environmental Risks*, DOI 10.1007/978-4-431-54804-1_15.
- Karimipour F., Delavar M.R. et Rezayan H., 2005.** Neighborhood Analysis in Water Pollution Estimation. *International Society for Environmental Information Sciences, Environmental Informatics Archives*. 3, 232-238.
- Karul C., Soyupak S., Cilesiz A. F., Akbay N., Germen E., 2000.** Case studies on the use of neural networks in eutrophication modeling, *Ecolog. Model.*, 134, 145-152.
- Karul C., Soyupak S., Yurteri C., 1999.** Neural network models as a management tool in lakes. *Hydrobiologia*. 409, 139-144.
- Keshtegar B & Heddami S., 2017.** Modeling daily dissolved oxygen concentration using modified response surface method and artificial neural network: a comparative study. *Neural Comput & Applic*, 1–12. <https://doi.org/10.1007/s00521-017-2917-8>.
- Kishan M., Chilukuri K M., Sanjay R., 1996.** Elements of Artificial Neural Networks. 337p.
- Kisi O., Karahan M.E., & Sen Z., 2006.** River suspended sediment modeling using a fuzzy logic approach. *Hydrological Processes*; 20, 4351–4362.
- Kordon A.K., 2010.** Applying Computational Intelligence: How to Create Value, Springer-Verlag Berlin Heidelberg, ISBN: 978-3-540-69910-1. DOI: 10.1007/978-3-540-69913-2.
- Kotti I.P., Sylaios G.K & Tsihrintzis V.A., 2013.** Fuzzy logic models for BOD removal prediction in free-water surface constructed wetlands. *Ecological Engineering*, 51, 66– 74.
- Kovacic Z., Bogdan S., 2006.** “Fuzzy controller design: theory and applications” Taylor & Francis.
- Kristensen P., Jensen P. and Jeppesen E., 1990.** Eutrophication Models for Lakes. Research Report C9. DEPA. Copenhagen, 120 pp.
- Kuo J., Hsieh M., Lung W., She N., 2007.** Using artificial neural network for reservoir eutrophication prediction. *Ecol. Model.* 200, 171–177.

- Kuo J.T., Lung W.S., Yang C.P., Liu W.C., Yang M.D. and Tang T.S., 2006.** "Eutrophication modelling of reservoirs in Taiwan", *Journal of Environmental Modelling and Software*, 21(6), 829-844.
- Kurtulus B., Razack M., 2010.** Modeling daily discharge responses of a large karstic aquifer using soft computing methods: artificial neural network and neuro-fuzzy. *J Hydrol*, 381, 101–111.
- Laanemets J., Lilover M.J., Raudsepp U., Autio R., Vahtera E., Lips I., Lips U., 2006.** A fuzzy logic model to describe the cyanobacteria *Nodularia spumigena* blooms in Gulf of Finland, Baltic Sea. *Hydrobiologia*, 554, 31–45.
- Ladlani I., Houichi L., Djemili L., Heddami S., Belouz K., 2013.** Estimation of Daily Reference Evapotranspiration (ET₀) in the North of Algeria Using Adaptive Neuro-Fuzzy Inference System (ANFIS) and Multiple Linear Regression (MLR) Models: A Comparative Study. *Arab J Sci Eng*, 39, 5959–5969.
- Lau H.C.W., Wong T.T. and Ning A., 2001.** "Incorporating Machine Intelligence in a Parameter-Based Control System: A Neural-Fuzzy Approach." *Artif. Intel. Eng.* 15, 253–264.
- Lee J.J., McEnery M.E., Röttger R., and Reimer C.W., 1980.** The isolation, culture and identification of endosymbiotic diatoms from *Heterostegina depressa* D'Orbigny and *Amphistegina lessonii* D'Orbigny (Larger Foraminifera) from Hawaii. *Botanica mar.* 23, pp. 297-301.
- Lee S.J., Jang M.H., Kim H.S., Yoon B.D. and Oh H.M., 2000.** Variation of microcystin content of *Microcystis aeruginosa* relative to medium N:P ratio and growth stage. *Journal of Applied Microbiology*, 89, 323-329.
- Lee, J.H.W., Huang, Y., Dickman, M., Jayawardena, A.W., 2003.** Neural Network Modelling of Coastal Algal Blooms. *Ecological Modelling*, 159, 179–201.
- Lefkir A, 2009.** Modélisation du transport solide par les modèles neuroflous. Thèse doctorat ENP d'Alger, El-Harrach.
- Legates DR., McCabe GJ., 1999.** Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35 (1): 233–241.
- Lermontov A., Yokoyama L., Lermontov M., Augusta M., Machado S., 2009.** River quality analysis using fuzzy water quality index: Ribeira do Iguape river watershed, Brazil. *Ecol. Indic.* 9, 1188-1197.
- Libin C., Zhifeng Y., Haifei L., 2016.** Assessing the eutrophication risk of the Danjiangkou Reservoir based on the EFDC model. *Ecol. Eng.* 96, 117–127.

- Lin W.S., Tsai C.H. and Liu J.S., 2001.** “Robust Neuro-Fuzzy Control of Multivariable Systems by Tuning Consequent Membership Functions.” *Fuzzy Sets Syst.* 124, 181–195.
- Lobrecht A.H., Dibike Y.B., Solomatine D.P., 2002.** Applications of Neural Networks and Fuzzy Logic to Integrated Water Management . Project Report , IHE, Delft, Netherland.
- Maier H.R., Dandy G.C., Burch1 M.D., 1998.** Use of artificial neural networks for modelling cyanobacteria *Anabaena* spp. in the River Murray, South Australia, *Ecol. Model.* 105, 257–272.
- Mamdani E.H & Assilian S., 1975.** An experiment in linguistic synthesis with a fuzzy logic controller. *Int J Man Mach Stud*, 7(1), 1–13.
- Mamdani E.H.1974.** Application of fuzzy algorithms for simple dynamic plant. *Pro. IEE.*, 121(12), 1585-1588.
- Marsili-Libelli S., 2004.** Fuzzy pattern recognition of circadian cycles in ecosystems, *Ecological Modelling*, 174, 67-84.
- Martin T ET Hargan H.B., 1996.** Neural network design. 734p.
- McCulloch W., & Pitts, W., 1943.** A logical calculus of ideas imminent in nervous activity. In *Bull. of Mathematical Biophysics*. Colorado Springs: The Dentan Printing Co.
- Millie F.D., Weckman R.G., Young A.W., Ivey E.J., Carrick J .H., Fahnenstiel G. L., 2012.** Modeling microalgal abundance with artificial neural networks: Demonstration of a heuristic ‘Grey-Box’ to deconvolve and quantify environmental influences. *Environmental Modelling & Software*, 38, 27-39.
- Minns, A.W., Mynett, A.E., Chen, Q., Boogaard, H.F.P., van den., 2000.** A cellular automata approach to ecological modelling. In: Odgaard, A.J., (Ed.), *Proceedings of Hydroinformatics Conference*, Iowa, USA.
- Minsky M. & Papert S., 1968.** *Perceptrons*. Cambridge, MA: MIT Press.
- Mohammadpour R., Shaharuddin S & Zakaria N.A. et al., 2016.** Prediction of water quality index in free surface constructed wetlands. *Environmental Earth Sciences*, 75, 139. <https://doi.org/10.1007/s12665-015-4905-6>.
- Moore L. and Thornton K., 1988.** *The Lake and Reservoir Restoration Guidance Manual*. First Edition. EPA 440/5-88-002, US EPA, Office of Wetlands, Oceans, and Watersheds, Assessment & Watershed Protection Division, WH-553, 401 M Street, S.W., Washington, DC 20460.
- Moses S. A., Letha J., and Joseph S., 2011.** Environmental status of a tropical lake. *Environmental Monitoring and Assessment* 180:427-449.

- Najah A., El-Shafie A., Karim O.A & El-Shafie A.H., 2014.** Performance of ANFIS versus MLP-NN dissolved oxygen prediction models in water quality monitoring. *Environ. Sci. Pollut. Res*, 21 (3), 1658–1670.
- Nayak P.C., Sudheer K.P., & Ramasastry K. S., 2005.** Fuzzy computing based rainfall–runoff model for real time flood forecasting. *Hydrol. Process*, 19(4) ,955–968.
- Nguyen H. T., Prasad N. R., Walker C. L., Walker E. A., 2003.** “A first course in fuzzy and neural control” Chapman & Hall/CRC, United States of America, New York.
- Noori R., Safavi S., Shahrokni S A N., 2013.** A reduced-order adaptive neuro-fuzzy inference system model as a software sensor for rapid estimation of five-day biochemical oxygen demand. *Journal of Hydrology*, 495, 175–185.
- Noori R., Yeh HD., Abbasi M., Kachoosangi FT & Moazami S., 2015.** Uncertainty analysis of support vector machine for online prediction of five-day biochemical oxygen demand. *J Hydrol*, 527(6), 833–843.
- Ocampo-Duque W., Ferre-Huguet N., Domingo J.L., Schuhmacher M., 2006.** Assessing water quality in rivers with fuzzy inference systems: a case study. *Environ. Int.* 32, 733-742.
- Odum E.P., 1971.** *Fundamentals of ecology*. 3rd ed. Saunders, Philadelphia and London.
- Olszewski, T., Ryniecki, A., Boniecki, P., 2008.** Neural network development for automatic identification of the endpoint of drying barley in bulk. *Journal of Research and Applications in Agricultural Engineering* 53 (1), 26–31.
- Overbeck J., 1988.** Ecosystem Concepts. In *Guidelines of Lake Management*, Vol. 1, Principles of Lake Management, S.E. Jorgensen and R.A. Vollenweider (Eds.). ILEC, UNEP.
- Ozgun K., Singh P.K., 2016.** Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *J. Hydrol.* 534, 104–112.
- Ozkan O., Ozdemir O., Azgin ST., 2009.** Prediction of Biochemical Oxygen Demand in a Wastewater Treatment Plant by Artificial Neural Networks. *Asian Journal of Chemistry* 21(6):4821-4830.
- Palani S., Liong S.Y., Tkalic P., 2008.** An ANN application for water quality forecasting. *Marine Pollution Bulletin*. 56, 1586–1597.
- Paquet P., 1997.** L’utilisation des réseaux de neurones artificiels en finance. Université d’Orléans, 26 p.
- Pastor-Bárcenas O., Soria-Olivas E., Martín-Guerrero J.D., Camps-Valls G., Carrasco-Rodríguez J.L., del Valle-Tascón S., 2005.** Unbiased sensitivity analysis and pruning

techniques in neural networks for surface ozone modelling. *Ecological Modelling*, 182 (2), 149–158.

Pereira G.C., Evsukoff A., Ebecken N.F.F., 2009. Fuzzy modelling of chlorophyll production in a Brazilian upwelling system, *Ecological Modelling*, 220, 1506–1512.

Perez E., Booth F.A., Martin D.F., 2000. Modelling organism–organism interactions as a means of predicting Florida red tide management potential. *Journal of Environment Science and Health, Part A Environmental Science*, 35, 219–227.

Phoslock., 2008. Blue-Green Algae Management in Aquaculture. Phoslock Water Solutions Ltd.

Pinay G., Gascuel C., Ménesguen A., Souchon Y., Le Moal M., Levain A., Moatar F., Pannard A., Souchu P., 2017. *L'eutrophisation : manifestations, causes, conséquences et prédictibilité*. Synthèse de l'Expertise scientifique collective CNRS - Ifremer - INRA - Irstea (France), 148 pages.

Price R. K., 2002. Hydroinformatics – Lecture1. *Proceedings of RBM 2002: Advanced study course on river basin modeling for flood risk mitigation*, 7-18 October 2002, The University of Birmingham, UK, pp5.1-5.19.

Quijano J.C., Zhu Z., Morales V., et al 2017. Three-dimensional model to capture the fate and transport of combined sewer overflow discharges: a case study in the Chicago Area Waterway System. *Sci. Total Environ.* **576**, 362–373.

Raheli B., Aalami M. T., El-Shafie A., Ghorbani M.A & Deo R. C., 2017. Uncertainty assessment of the multilayer perceptron (MLP) neural network model with implementation of the novel hybrid MLP-FFA method for prediction of biochemical oxygen demand and dissolved oxygen: a case study of Langat River. *Environ Earth Sci* ,76, 503. <https://doi.org/10.1007/s12665-017-6842-z>.

Rajai M.E., 2002. Modélisation cognitive et prédictive du comportement dynamique de l'activité électrique du coeur par approches neuronales. Thèse doctorat, Institut national des sciences appliquées de Lyon, 180p.

Recknagel F., Bobbin J., Whigham P et Wilson, H., 2002. Comparative application of artificial neural networks and genetic algorithms for multivariate time-series modelling of algal blooms in freshwater lakes. *Journal of Hydroinformatics*/04.2/2002.

Recknagel F., French M., Harkonen P., Yabunaka KI., 1997. Artificial neural network approach for modelling and prediction of algae blooms. *Ecol. model* 96, 11-28.

Regieg M.F., 2002. Etude du phénomène d'eutrophisation dans le barrage de Keddara. Thèse magistère, Ecole nationale supérieure d'hydraulique, Blida, Algérie.

- Remini B., 2005.** Problématique de l'eau en Algérie. 162p, ONPU. ISBN: 9961.0.0884.7.
- Richard, P. A. and McQuisten, R., 1968.** Glossary of Selected Hydrologic Terms. Laramie: Water Resources Institute, University of Wyoming, pp. 1–53.
- Robertson A., 1997.** *Limiting Nutrient Workshop*. LWRRDC Occasional Paper 7/99. School of Science and Technology, Charles Sturt University, Wagga NSW 2678.
- Romain B., 2003.** Modèles Impulsionnels de Réseaux de Neurones Biologiques. Thèse Doctorat, Paris pp 03-10
- Rosenblatt F., 1958.** The perceptron: A probabilistic model for information storage and organisation in the brain. *Psychological Review*, Vol 65, pp. 386-408.
- Rumelhart D.E., 1986.** Learning internal representations by error propagation. *Parallel Distributed Processing Explorations in the Microstructure of Cognition*. Eds. D. E. Rumelhart and J.L. McClelland, Cambridge, MA, MIT Press, Bradford Books, vol. 1, pp. 318-362.
- Sahu M., Mahapatra S.S., Sahu H.B. et al., 2011.** Prediction of Water Quality Index Using Neuro Fuzzy Inference. *System Water Qual Expo Health*, 3, 175–191.
- Salacinska K., El Serafy G.Y., Los F.J., Blauw A., 2010.** Sensitivity analysis of the two dimensional application of the Generic Ecological Model (GEM) to algal bloom prediction in the North Sea. *Ecological Modelling*, 221, 178–190.
- Sattar A. M. A., Gharabaghi B., Sabouri F et Thompson A.M., 2017.** Urban Stormwater Thermal Gene Expression Models for Protection of Sensitive Receiving Streams. *Hydrological Processes*, 31(13), 2330–2348. DOI: 10.1002/hyp.11170.
- Scardi M., Harding L.W., 1999.** Developing an empirical model of phytoplankton primary production: a neural network case study. *Ecological Modelling* 120, 213–223.
- Scholten M.C. Th., Foekema E.M., Van Dokkum H.P., Kaag N.H.B.M. and Jak R.G. 2005.** *Eutrophication Management and Ecotoxicology*. ISBN 3-540-22210-3 Springer Berlin Heidelberg New York.
- Sen Z., 2010.** *Fuzzy logic and hydrological modeling*. Boca Raton, FL: CRC Press.
- Shapiro J., 1970.** A Statement on Phosphorus. *Journal (Water Pollution Control Federation)*, Vol. 42, No. 5, Part I (May, 1970), pp. 772-775.
- Sharma V., Negi S.C., Rudra R.P., Yang S., 2003.** Neural networks for predicting nitrate-nitrogen in drainage water. *Agricultural water management*, 63, 169-183.
- Shen, Q., Chouchoulas, A., 2001.** “FuREAP: A Fuzzy-Rough Estimator of Algae Populations.” *Artif. Intell. Eng.* 15, 13–24.

- Shiri J & Kişi O., 2011.** Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. *Computers & Geosciences*, 37, 1692–1701.
- Šiljić A., Antanasijević D., Perić-Grujić A., Ristić M & Pocajt V., 2015.** Artificial neural network modelling of biological oxygen demand in rivers at the national level with input selection based on Monte Carlo simulations. *Environmental Science and Pollution Research*, 22(6), 4230–4241.
- Šiljić Tomić A.N., Antanasijević D.Z., Ristić M.Đ. et al., 2016.** Modeling the BOD of Danube River in Serbia using spatial, temporal, and input variables optimized artificial neural network models. *Environ Monit Assess*, 188(5):300.
- Silvey J.K., Henley D.E. and Wyatt J.T., 1972.** Planktonic blue-green algae; growth and odorproduction studies. *J. Am. Waterw. Assoc.* 64, pp. 35-39.
- Singh KP., Basant A., Malik A., Jain G., 2009.** Artificial neural network modeling of the river water quality -A case study. *Ecol Model*, 220, 888–895.
- Sivanandam S.N., Sumathi S., & Deepa S.N., 2007.** Introduction of Fuzzy Logic using MATLAB. *Springer-Verlag*, Berlin Heidelberg.
- Smith K.A & Gupta J.N.D., 2001.** Neural networks in business: techniques and applications. Idea Group Publishing, 1331 E. Chocolate Avenue Hershey PA, USA, pp 193.
- Smith V. H., 2003.** Eutrophication of freshwater and coastal marine ecosystems a global problem. *Environmental Science and Pollution Research* 10: 126–139.
- Soyupak S et Chen D.G., 2004.** Fuzzy logic model to estimate seasonal pseudo steady state chlorophyll-a concentrations in reservoirs.” *Environmental Modeling and Assessment*. 9, 51-59.
- Sridharan N. and Lee G.F., 1977.** Algal nutrient limitation in Lake Ontario and tributary waters. *Water Res.* 11, pp. 849-858.
- Sumathi S & Surekha P., 2010.** Computational Intelligence paradigms: Theory and Applications using MATLAB®. Taylor and Francis Group, LLC, ISBN: 978-1-4398-0902-0.
- Sungwon K., Hungo SK., 2008.** Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modelling. *Journal of Hydrology* 351:299– 317.
- Sylaios K.G., Gitsakis N., Koutroumanidis T., Tsihrintzis, A.V., 2008.** CHLfuzzy: a spreadsheet tool for the fuzzy modeling of chlorophyll concentrations in coastal lagoons. *Hydrobiologia*, 610, 99–112.

- Thomann R.V. and Fitzpatrick J.F., 1982.** Calibration and verification of a mathematical model of the eutrophication of the Potomac Estuary; report by Hydroqual, Inc., Mahwah, N J, to DES, Dist. Col. Washington D.C. 80.
- Thomas R., Meybeck M., Beim A., 1996.** Lakes. In: Chapman D.V. (Editor), Water quality assessments: a guide to the use of biota, sediments and water in environmental monitoring. Spon, London [u.a.].
- Tilman D., 1982.** *Resource competition and community structure*. Princeton University Press, Princeton, New Jersey, USA.
- Toprak Z. F., Eris E., Agiralioglu N., Cigizoglu H. K., Yilmaz L., Aksoy H., Coskun H. G., Andic G., & Alganci U., 2009.** Modeling Monthly Mean Flow in a Poorly Gauged Basin by Fuzzy Logic. *Clean*, 37 (7), 555 – 564.
- Tsakiris G., Alexakis D., 2012.** Water quality models. A review. *European Water*, 37, 33-46.
- Veatch J.O., and Humphrys C.R., 1966.** Water and Water Use Terminology. Kaukauna: Thomas Printing & Publishing, pp. 1–75.
- Von Neumann J., 1956.** *The Computer and the brain*. Yale University Press.
- Walsh J.J., Penta B., Dieterie D.A., Bissett W.P., 2001.** Predictive ecological modeling of harmful algal blooms. *Human and Ecological Risk Assessment*, 7, 1369–1383.
- Wang X., Jia J., Su T., et al 2018.** A fusion water quality soft sensing method based on WASP model and its application in water eutrophication evaluation. *J. Chem*, 1–14.
- Wei M., Bai B., Sung AH., Liu Q., Wang J., Cather ME., 2007.** Predicting injection profiles using ANFIS. *Information Sciences* 177:4445–4461.
- Wen G., Li X., Qiu X., Cheng Y., Sun Y., and Huang T., 2016.** Water Pollution and Water Quality Control of Selected Chinese Reservoir Basins.
- Wheater H., Sorooshian S., Sharma K. D., 2007.** *Hydrological Modelling in Arid and Semi-Arid Areas*. ISBN-13 978-0-511-37710-5.
- Whitehead P., Howard A., Arulmani C., 1997.** Modelling algal growth and transport in rivers: a comparison of time series analysis, dynamic bass balance and neural network techniques. *Hydrobiologia*, 349, 39–46.
- Widrow B., Hoff M., 1960.** In IRE Western Electric Show and Convention Record, pp. 96-104.
- Witzel K.P. and Overbeck H.J., 1979.** Heterotrophic Nitrification by *Arthrobacter* sp. (Strain 9006) as Influenced by Different Cultural Conditions, Growth State and Acetate Metabolism. *Arch. Microbiol.* 122, 137-143.

- Wu G and Xu Z., 2011.** Prediction of algal blooming using EFDC model: Case study in the Daoxiang Lake. *Ecological Modelling*, 222,1245–1252.
- Xi L., Xuyong L., 2018.** Using the EFDC model to evaluate the risks of eutrophication in an urban constructed pond from different water supply strategies. *Ecol. Modell.* **372**, 1–11.
- Xiang S., Liu Z., Ma L., 2006.** Study of multivariate linear regression analysis model for ground water quality prediction. *Guizhou Science*, 24(1):60–62.
- Xiao-long W., Yong-long LU., Gui-zhen HE., Jing-yi HAN., Tie-yu W., 2007.** Exploration of relationships between phytoplankton biomass and related environmental variables using multivariate statistic analysis in a eutrophic shallow lake: A 5-year study. *Journal of Environmental Sciences*, 19, 920-927.
- Xu S., Zhang, K., Zhao, S., 2015.** Prediction methods analysis of surface water quality based on the MIKE 21 FM numerical model. *Environ. Sci. Technol.* **38**, 386–390.
- Yabukana K., Hosomi M., Murakami M., 1997.** Novel application of backpropagation artificial neural network formulated to predict algal bloom. *Water Sci. Technol*, 36 (5), 89–99.
- Yan H., Zou Z., Wang H., 2010.** Adaptive neuro-fuzzy inference system for classification of water quality status. *Journal of Environmental Sciences*, 22(12), 1891–1896.
- Zadeh L.A. 1965.** “Fuzzy sets” *Information Control*, 8, Pp. 94–102.
- Zadeh L.A., 1999.** From Computing with Numbers to Computing with Words. From manipulation of measurements to manipulation of perceptions. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* (Volume: 46, Issue: 1)
- Zheng C. and Gordon B., 1995.** *Applied Contaminant Transport Modelling*. Van Nostrand Reinhold, New York.
- Zhuravlev Y.N & Avetisov V.A., 2006.** The definition of life in the context of its origin. *Biogeosciences*, 3, 281–291, 2006.