الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ YAHIA FARÈS DEMÉDÉA
Département de Génie des Procédés et Environnement
Laboratoire de Biomatériaux et Phénomènes de Transport (LBMPT)

THESE DE DOCTORAT

Présentée par :

M^me^. Lamia GHERABA

Ingénieur en Génie des Procédés : Université de Médéa, Algérie
Magister en Génie des Procédés : Université de Médéa, Algérie

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

Thème

Modélisation Dynamique des Paramètres de Pollution et Traitement Biologique par un Modèle Neuro-Fuzzy.

Soutenue le, 08/07/2020 devant le jury composé de :

M^me^ H. CHERIFI	Professeur (UYFM de Médéa)	Présidente
M^me^ R. RIHANI	Professeur (USTHB d'Alger)	Examinatrice
M. H. BOUTOUMI	Professeur (USDB1 de Blida)	Examinateur
M. A. ABDELLAH ELHADJ	MCA (USDB1 de Blida)	Examinateur
M. S. HANINI	Professeur (UYFM de Médéa)	Directeur de thèse
M. O. BENKORTBI	Professeur (UYFM de Médéa)	Co-directeur de thèse

Dédicaces

***A** mes chers parents,*

***A** mon mari,*

***A** mes enfants : Khadidja, Younes, Mayassine et Djawad*

***A** ma sœur et mes frères,*

***A** mes nièces et neveux,*

***A** toute ma famille,*

***A** tous ceux qui m'ont soutenu de près ou de loin.*

Remerciements et Gratitude

Au terme de ce travail, je tiens à exprimer ma reconnaissance envers tous ceux qui m'ont prodigué leur aide et leur soutien.

Mes premiers et sincères remerciements vont tout d'abord à mon directeur de thèse Monsieur le Professeur Salah HANINI qui m'a accordé sa confiance sans faille tout au long de cette thèse. Sa compétence, son sérieux et son implication dans ce travail m'ont beaucoup soutenu. Qu'il trouve ici l'expression de ma reconnaissance et de ma profonde gratitude.

Je tiens à exprimer mes plus vifs remerciements à mon Co-rapporteur Professeur Othmane BENKORTBI, pour ses aides, ses accompagnements et ses éclairages techniques.

Toute ma gratitude et tout mon respect s'adressent à Monsieur Mabrouk HAMADACHE pour son aide précieuse dans ce travail, pour sa générosité scientifique et pour sa disponibilité. A aucun moment, il n'a cessé de suivre et de participer passionnément à la progression de cette thèse, je ne peux que le remercier du fond du cœur.

Ce serait de l'ingratitude si j'ose oublier l'aide précieuse du Docteur Latifa KHAOUANE et du Docteur Chérif Si MOUSSA. À leur aide multiforme et leurs conseils qui étaient spontanés. Pour leur disponibilité, je ne peux que les remercier du fond du cœur.

Je suis très reconnaissant envers les membres du jury : Professeur ***H.CHERIFI*** *(UYF de Médéa), Professeur* ***R.REHANI*** *(USTHB d'Alger), Professeur* ***H.BOUTOUMI*** *(USDB1 de Blida) et Docteur* ***A. ABDELLAH ELHADJ*** *(USDB1 de Blida) qui me font le plaisir d'examiner ce travail et de participer à ma soutenance de thèse.*

Merci à tous ceux qui m'ont aidé de près ou de loin durant ces années de recherche et d'études.

Résumé

Modélisation dynamique des paramètres de pollution et traitement biologique par un modèle Neuro-Fuzzy

L'objectif de ce travail consiste en l'étude du fonctionnement hydro-climatique du barrage Ghrib dans le but d'établir une relation entre le climat et l'eau de surface. La méthodologie entreprise au cours de cette recherche a été basée sur l'analyse physico-chimique mensuelle des eaux de surface du barrage et l'analyse climatique de l'environnement sur une période allant de 2003 à 2015 avec une base de 156 points de données expérimentaux.

Trois modèles ont été appliqués à savoir : ANN-MLP (réseau de neurones de type perceptron multicouche), ANFIS (Système d'inférence neuro-flou adaptatif) et SVM (Machine à vecteurs de support) pour prédire les paramètres climatiques : la température minimale de l'air ($T_{air,min}$), la température maximale de l'air ($T_{air,max}$), l'humidité relative (HR) et la durée d'ensoleillement (DE) à partir des paramètres physico-chimiques de l'eau de surface de barrage Ghrib : la température de l'eau (T_{eau}), la conductivité électrique (CE), la turbidité (Turb), l'oxygène de saturation (O_2sat), le pH, la demande chimique en oxygène (DCO), l'ion d'ammonium (NH_4^+), l'ion de nitrate (NO_3^-), la matière organique (MO) et les résidus secs (RS). Une deuxième contribution est apportée dans le sens inverse, c'est-à-dire la prédiction des paramètres physico-chimiques de l'eau de surface de barrage Ghrib à partir des paramètres climatiques.

Concernant le réseau direct, les résultats obtenus montrent que les performances corrélatives de ces méthodes sont très acceptables et encourageantes, avec un avantage du modèle SVM. Les coefficients de corrélation R et les erreurs relatives absolues moyennes (ERAM) pour le modèle SVM sont de 0.9945 et une erreur relative absolue moyenne tend vers l'infini pour $T_{air,min}$, 0.9979 et 0.77 pour $T_{air,max}$, 0.9877 et 0.96 pour HR et 0.9869 et 3,07 pour DE.

De même pour le réseau inverse, le modèle SVM présente une meilleure performance par rapport à celui d'ANFIS et de ANN-MLP avec un coefficient de corrélation R compris entre 0.8967 et 0.9948 et une erreur relative absolue moyenne (ERAM) compris entre 2.50 et 30.11 sur l'ensemble des paramètres physico-chimiques de l'eau de surface de barrage Ghrib.

Mots clés **:** Paramètres climatiques, Paramètres physico-chimiques de l'eau, Barrage Ghrib, Réseaux de neurones artificiels de type perceptron multicouches, Système d'inférence neuro-flou adaptatif, Machine à vecteurs de support.

ملخص

النمذجة الديناميكية لعوامل التلوث والمعالجة البيولوجية باستخدام نموذج عصبى ضبابي

الهدف من هذا العمل هو دراسة الأداء الهيدروميناخي لسد غريب من أجل تحليل العلاقة بين المناخ والمياه السطحية. استندت المنهجية المتبعة خلال هذا البحث إلى التحليل الفيزيائي الكيميائي الشهري للمياه السطحية للسد والتحليل المناخي للبيئة خلال الفترة من الممتدة من 2003 إلى 2015 مع قاعدة بيانات تتكون من 156 نقطة.

تم تطبيق ثلاث نماذج و هي : الشبكة العصبية من نوع متعدد الطبقات ANN-MLP، نظام الاستدلال العصبي الضبابي التكيفي ANFIS و آلة ناقلات الدعم SVM للتنبؤ بالعوامل المناخية : درجة الحرارة الدنيا و القصوى، الرطوبة النسبية و مدة الإشعاع باستعمال العوامل الفيزيوكيميائية للمياه السطحية لسد غريب : درجة حرارة الماء، الناقلية الكهربائية (CE) ، التعكر (Turb) ، تشبع الأكسجين (O_2sat) ، الرقم الهيدروجيني ، الطلب الكيميائي على الأكسجين (COD) ، الأمونيوم (NH_4^+) ، النترات (NO_3-) ، المواد العضوية (MO) والمخلفات الجافة (RS). تم تقديم مساهمة ثانية في الاتجاه المعاكس ، أي التنبؤ بالعوامل الفيزيائية الكيميائية لمياه سطح سد غريب من العوامل المناخية.

فيما يتعلق بالشبكة المباشرة، توضح النتائج التي تم الحصول عليها أن الأداء المترابط لهذه الطرق جد مقبول و مشجع، مع ميزة نموذج آلة ناقلات الدعم. معاملات الارتباط R و متوسط الانحرافات النسبية المطلقة (ERAM) لنموذج آلة ناقلات الدعم هي 0.9945 و متوسط الخطأ النسبي المطلق يميل إلى ما لا نهاية لدرجة الحرارة الدنيا، 0.9979 و 0.77 لدرجة الحرارة القصوى، 0.9877 و 0.96 للرطوبة النسبية و 0.9869 و 3.07 لمدة الإشعاع.

على غرار الشبكة العكسية، يقدم نموذج آلة ناقلات الدعم أداءً أفضل مقارنةً بأداء ANFIS و ANN-MLP مع معامل ارتباط R بين 0.8967 و 0.9948 ومتوسط انحراف نسبي مطلق (ERAM) بين 2.50 و 30.11 على جميع العوامل الفيزيوكيميائية للمياه السطحية لسد غريب.

الكلمات المفاتيح : العوامل المناخية ، العوامل الفيزيوكيميائية للمياه، سد غريب، الشبكات العصبية الاصطناعية متعددة الطبقات، نظام الاستدلال العصبي الضبابي التكيفي، آلة ناقلات الدعم.

Summary

Dynamic modeling of pollution parameters and biological treatment by a Neuro-Fuzzy model

The objective of this work is to study the hydro-climatic functioning of the Ghrib dam in order to establish a relationship between climate and surface water. The methodology undertaken during this research was based on the monthly physico-chemical analysis of the surface water of the dam and the climatic analysis of the environment over a period from 2003 to 2015 with a database of 156 experimental points.

Three models have been applied, namely: ANN-MLP (neural network of multilayer perceptron type), ANFIS (Adaptive neuro-fuzzy inference system) and SVM (Support vector machine) to predict climatic parameters: minimum air temperature ($T_{air,min}$), maximum air temperature ($T_{air,max}$), relative humidity (RH) and sunshine duration (SD) from the physico-chemical parameters of the surface water Ghrib dam: water temperature (T_{water}), electrical conductivity (EC), turbidity (Turb), oxygen saturation (O_2sat), pH, chemical oxygen demand (COD), ammonium ion (NH_4^+), nitrate ion (NO_3^-), organic matter (OM) and dry residues (DR). A second contribution is made in the opposite direction, that is to say the prediction of the physico-chemical parameters of the Ghrib dam surface water from climatic parameters.

Concerning the direct network, the results obtained show that the correlative performances of these methods are very acceptable and encouraging, with an advantage of the SVM model. The correlation coefficients R and the mean absolute relative errors (MARE) for the SVM model are 0.9945 and an mean absolute relative error tends to infinity for $T_{air,min}$, 0.9979 and 0.77 for $T_{air,max}$, 0.9877 and 0.96 for HR and 0.9869 and 3.07 for SD.

Similarly for the reverse network, the SVM model presents a better performance compared to that of ANFIS and ANN-MLP with a correlation coefficient R between 0.8967 and 0.9948 and an mean absolute relative errors (MARE) between 2.50 and 30.11 on all the physico-chemical parameters of Ghrib dam surface water.

Keywords: Climatic parameters, Physico-chemical parameters of water, Ghrib dam, Multilayer perceptron type artificial neural networks, Adaptive neuro-fuzzy inference system, Support vector machine.

Table des matières

Table des matières i
Abréviations iii
Liste des tableaux iv
Liste des figures v

Introduction générale 1

Chapitre I : Etat de l'art sur l'utilisation des approches d'intelligence artificielle pour la modélisation des paramètres de pollution de l'eau et des paramètres climatiques

I.1. Relation eau-climat 4
I.1.1. Généralités sur les eaux de surfaces 4
I.1.2. Caractéristiques des barrages 4
I.1.3. Paramètres physico-chimiques et biologiques des eaux de surfaces 5
I.1.3.1. Température de l'eau 5
I.1.3.2. pH 5
I.1.3.3. Conductivité électrique 6
I.1.3.4. Matières en suspension (MES) 6
I.1.3.5. Résidus secs (RS) 6
I.1.3.6. Oxygène dissous (OD) et % de saturation en oxygène 6
I.1.3.7. Charge en matières organiques : demande chimique en oxygène (DCO) et demande biochimique en oxygène (DBO) 7
I.1.3.8. Eléments nutritifs 8
I.1.4. Généralités sur le climat 9
I.1.4.1. Paramètres climatiques 9
I.1.4.1.1. Température de l'air 9
I.1.4.1.2. Humidité relative 9
I.1.4.1.3. Précipitations 10
I.1.4.1.4. Radiation solaire 11
I.1.4.1.5. Vent 11
I.2. Changement climatique et cycle hydrologique 11
I.2.1. Etat de l'art sur la prévision des variables météorologiques 12
1.2.1.1. Prédiction de la température de l'air 13
1.2.1.2. Prédiction de l'humidité relative (HR) 18
1.2.1.3. Prédiction de la durée d'ensoleillement (DE) 20
1.2.1.4. Prédiction des précipitations 22
I.2.2. Etat de l'art sur la prévision des variables de pollution de l'eau 24
I.2.2.1. Prédiction de la température de l'eau 25
I.2.2.2. Prédiction de la teneur en oxygène dissous (OD) 26
I.3. Conclusion 30

Chapitre II : Techniques d'intelligence artificielle pour la modélisation

II.1. Introduction 31
II.2. Modélisation par réseaux de neurones artificiels 31
II.2.1. Neurone formel 32
II.2.2. Architecture d'un réseau de neurones 32
II.2.3. Perceptron multicouche 33
II.2.4. Optimisation de la structure du réseau 34
II.2.5. Apprentissage 34
II.2.6. Méthodologie de synthèse 35
II.3. Modélisation par les systèmes neuro-flous 36
II.3.1. La logique floue 36
II.3.2. Système Neuro-Floue 37
II.3.3. Système d'Inférence Neuro-Floue Adaptatif (ANFIS) 37
II.3.4. Apprentissage et validation du modèle ANFIS 39
II.4. Modélisation par machines à vecteurs supports (SVM) 40
II.4.1. SVM pour la régression linéaire 41
II.4.2. SVM pour la régression non linéaire 44
II.5. Conclusion 45

Chapitre III : Modélisation des paramètres climatiques et des paramètres physico-chimiques de l'eau : Résultats et Discussion

III.1. Présentation du site étudié 46
III.2. Collecte d'une base de données 46
III.3. Développement du modèle 50
III.4. Prédiction des paramètres climatiques à partir des paramètres physico-chimiques de d'eau 50
III.4.1. Modélisation par le modèle ANN-MLP 50
III.4.1.1. Résultats de corrélation par le modèle ANN-MLP 56
III.4.2. Modélisation par le modèle ANFIS 59
III.4.2.1. Résultats de corrélation par le modèle ANFIS 59
III.4.3. Modélisation par le modèle SVM 62
III.4.3.1. Résultats de corrélation par le modèle SVM 63
III.4.4. Comparaison des trois Modèles ANN-MLP, ANFIS et SVM 65
III.5. Prédiction des paramètres physico-chimiques de l'eau à partir des paramètres climatiques (réseau inverse) 71
III.5.1. Collecte de la base de données 71
III.5.2. Modélisation par le modèle ANN-MLP 71
III.5.3. Comparaison entre les modèles de prédiction ANN-MLP, ANFIS et SVM (Réseau inverse) 77
III.6. Conclusion 81

Conclusion générale et perspectives 82

Références bibliographiques

Abréviations

A	Année
AG	Algorithmes Génétiques
ANFIS	Système d'Inférence Neuro Fuzzy Adaptatif
ANN	Réseau de Neurones Artificiels (Artificial Neural Network)
ANRH	Agence Nationale des Ressources Hydrauliques
CE	Conductivité électrique
DCO	Demande chimique en oxygène
DE	Durée d'ensoleillement
ER	Erreur relative
ERAM	Erreur relative absolue moyenne
HR	Humidité relative
IA	Intelligence artificielle
LF	Logique Floue
M	Mois
MF	Fonction d'appartenance (Membership Function)
MLP	Perceptron Multi Couches (Multilayer Perceptron)
MLR	Régression Linear Multiple (Multiple Linear Regression)
MO	Matière organique
EQM	Erreur quadratique moyenne
NH_4^+	Ammonium
NO_3^-	Nitrate
O_{2sat}	Saturation en Oxygène
ONM	Office National de la Météo
pH	Potentiel d'Hydrogène
RBF	(Radial Basis Function)
REQM	Racine de l'Erreur quadratique moyenne
RS	résidus secs
SVM	Machine à Vecteurs de Support (Support Vector Machine)
SVR	Régression à Vecteurs de Support (Support Vector Regression)
$T_{air, max}$	Température maximale de l'air
$T_{air, min}$	Température minimale de l'air
T_{eau}	Température de l'eau
Turb	Turbidité
VIF	Facteur d'inflation de la variance (Variance Inflation Factor)

Liste des tableaux

Chapitre I		
Tableau I.1	Caractéristiques des eaux de surface	5
Chapitre III		
Tableau III.1	Résumé statistique des variables retenues	47
Tableau III.2	Matrice des coefficients de corrélation des paramètres d'entrée pour $T_{air,min}$	48
Tableau III.3	Matrice des coefficients de corrélation des paramètres d'entrée pour $T_{air,max}$	48
Tableau III.4	Matrice des coefficients de corrélation des paramètres d'entrée pour HR	48
Tableau III.5	Matrice des coefficients de corrélation des paramètres d'entrée pour DE	49
Tableau III.6	Valeurs du coefficient VIF des paramètres d'entrée	49
Tableau III.7	Architecture des réseaux de neurones optimisés	53
Tableau III.8	Structure des réseaux de neurones optimisés	56
Tableau III.9	Valeurs des paramètres statistiques pour les ensembles d'apprentissage et de test	58
Tableau III.10	Les paramètres optimisés du modèle SVM	63
Tableau III.11	Comparaison des critères statistiques des trois modèles	68
Tableau III.12	Structure des réseaux de neurones optimisés (Réseau inverse)	72
Tableau III.13	Comparaison des critères statistiques des trois modèles (Réseau inverse)	77

Liste des figures

Chapitre I		
Figure I.1	Nombre de travaux inhérents à la température de l'air publiés durant les années 2000 à 2020	15
Figure I.2	Nombre de travaux inhérents à l'humidité relative publiés durant les années 2000 à 2020	18
Figure I.3	Nombre de travaux inhérents au rayonnement solaire publiés durant les années 2000 à 2020	21
Figure I.4	Nombre de travaux inhérents à la température de l'eau publiés durant les années 2000 à 2020	25
Figure I.5	Nombre de travaux inhérents à la teneur en oxygène dissous publiés durant les années 2000 à 2020	27
Chapitre II		
Figure II.1	Schématisation d'un neurone formel	32
Figure II.2	Perceptron multicouche	34
Figure II.3	Organigramme de la synthèse d'un modèle neuronal	35
Figure II.4	Système d'inférence floue	36
Figure II.5	Principe du système neuro-flou	37
Figure II.6	L'Architecture de l'ANFIS	38
Figure II.7	Les paramètres de la fonction de perte pour un SVM linéaire	43
Figure II.8	Architecture d'une machine à vecteurs support	44
Chapitre III		
Figure III.1	Présentation géographique du barrage Ghrib	46
Figure III.2	Capture d'écran des logiciels STATISTICA 10	51
Figure III.3	Architecture du modèle ANN-MLP	51
Figure III.4	Contribution des paramètres d'entrée dans les modèles ANN-MLP	52
Figure III.5	Prédiction des paramètres climatiques : $T_{air,min}$, $T_{air,max}$, HR et DE par le modèle MLP-ANN avec 12 entrées	54
Figure III.6	Paramètres climatiques observés et prédites ($T_{air,min}$, $T_{air,max}$, HR et DE) pour l'année 2016	53
Figure III.7	Prédiction des paramètres climatiques : $T_{air,min}$, $T_{air,max}$, HR et DE par le modèle MLP-ANN avec 7 entrées	57
Figure III.8	L'ER du modèle ANN-MLP des paramètres : $T_{air,min}$, $T_{air,max}$, HR et DE	58
Figure III.9	Logiciel de MATLAB R2018b	59
Figure III.10	Prédiction des paramètres climatiques : $T_{air,min}$, $T_{air,max}$, HR et DE par	61

	le modèle ANFIS	
Figure III.11	L'ER du modèle ANFIS des paramètres : $T_{air,min}$, $T_{air,max}$, HR et DE	61
Figure III.12	Prédiction des paramètres climatiques : $T_{air,min}$, $T_{air,max}$, HR et DE par le modèle SVM	64
Figure III.13	L'ER du modèle SVM des paramètres : $T_{air,min}$, $T_{air,max}$, HR et DE	64
Figure III.14	Prédiction de quelques paramètres physico-chimiques de l'eau : T_{eau}, CE, Turb, RS, O_2sat, NH_4^+, DCO par le modèle ANN-MLP	74
Figure III.15	Prédiction de quelques paramètres physico-chimiques de l'eau : T_{eau}, CE, Turb, RS, O_2sat, NH_4^+, DCO par le modèle ANFIS	75
Figure III.16	Prédiction de quelques paramètres physico-chimiques de l'eau : T_{eau}, CE, Turb, RS, O_2sat, NH_4^+, DCO par le modèle SVM	77

INTRODUCTION GENERALE

Introduction Générale

La connaissance du climat d'une région et des modalités climatiques qui la caractérisent est une tâche primordiale en ce sens qu'elle permet des prédictions sur les paramètres climatiques dans le futur. Ainsi, on dénombre plus de cinquante paramètres climatiques vitaux qui sont la clé pour des observations climatiques durables. Parmi ces paramètres, on cite la durée d'ensoleillement, l'humidité, la vitesse du vent, la pression atmosphérique, la température de l'air et celle des eaux, la couverture nuageuse, les précipitations, l'humidité du sol, etc. La température de l'air, la durée d'ensoleillement et l'humidité relative sont parmi les paramètres climatiques les plus importants et les plus influents. A titre d'exemple, la connaissance de la variation de la température de l'air est de la plus haute importance en agriculture car des changements extrêmes de la température de l'air peuvent endommager les plantes et les animaux. Par ailleurs, la connaissance de la durée d'ensoleillement à la surface de la terre est d'une importance majeure non seulement du point de vue climatologique mais aussi agro météorologique, ingénierie, architecture des gains de chaleur des bâtiments ainsi que d'autres études appliquées en sciences de l'environnement. L'humidité, quant à elle, joue un rôle très important dans la formation des différents phénomènes météorologiques comme la pluie, les inondations, la neige et les orages. Mise à part son influence sur la qualité des cultures, l'humidité joue aussi un rôle vital dans le processus de séchage de certains produits agricoles.

Malheureusement, la mesure de ces paramètres climatiques nécessite le plus souvent l'existence d'un réseau de stations météorologiques et d'appareillage de mesure très performant. Cependant, ce n'est pas le cas pour beaucoup de pays en raison des coûts et des exigences de maintenance et d'étalonnage de l'équipement de mesure. Pour pallier à cet impondérable, le développement de méthodes alternatives pour prédire les paramètres climatiques est un domaine de recherche largement exploré. Cette voie est imposée pour plusieurs raisons : considérations économiques et réduction des contraintes de temps. Les techniques d'intelligences artificielles, telles que les réseaux de neuronaux artificiels (ANN), les algorithmes génétiques (AG) et la théorie de la logique floue (LF) sont des méthodes alternatives de plus en plus utilisées dans les événements météorologiques.

La relation Climat - Eau est évidente, et les interrelations entre le système climatique et le cycle de l'eau sont nombreuses et d'une redoutable complexité. La montée des températures affecte simultanément différentes composantes des systèmes hydrologiques : la quantité des précipitations, mais aussi leur intensité et leur fréquence, la fonte de la neige et des glaces, l'augmentation de la présence de vapeur d'eau dans l'atmosphère, l'évapotranspiration, la teneur en eau du sol et de la végétation, la force des ruissellements et le débit des cours d'eau, etc. Tous ces phénomènes interagissent les uns avec les autres.

L'évolution des concentrations de CO_2 affecte l'évapotranspiration de la végétation et donc l'augmentation potentielle de la perte d'eau. Des sols plus secs peuvent entraîner des sécheresses plus longues, ce qui modifiera également la couverture végétale, affectant ainsi le captage du ruissellement en réponse aux précipitations. De plus, la recharge en eau souterraine sera également affectée, ce qui modifiera la quantité de percolation. Ce changement modifiera par conséquent la contribution du débit de base des eaux souterraines.

En plus de la quantité d'eau, la qualité va également changer. Les précipitations extrêmes entraîneront une augmentation des eaux de ruissellement, entraînant ainsi l'évacuation des polluants des zones urbaines et des exploitations agricoles jusqu'aux plans d'eau récepteurs. Cela posera également un risque physique pour les infrastructures d'approvisionnement en eau telles que les barrages et les systèmes d'approvisionnement en eau et de traitement des eaux usées en raison de la surcharge de contaminants et des difficultés de traitement. D'autre part, une réduction des débits et des volumes d'eau entraînerait une augmentation de la concentration en polluants.

L'objectif de cette étude a deux axes, le premier axe consiste à prédire les paramètres climatiques (la température minimale et maximale de l'air, la durée d'ensoleillement et l'humidité relative) en utilisant dix paramètres physicochimiques de l'eau du barrage Ghrib. Les données recueillies en 2003-2015 ont été exploitées pour établir les modèles de prédiction en utilisant la technique de réseaux de neurones artificiels (ANN).

La performance et la robustesse des modèles RNA ont été interprétées sur la base de certains critères statistiques. Par la suite les résultats obtenus par la technique ANN ont été comparés avec d'autres modèles basés sur la technique de système d'inférence neuro-flou adaptatif (ANFIS) et la technique de machines à vecteurs support SVM.

Le deuxième axe va dans le sens inverse, c'est-à-dire l'utilisation des paramètres climatiques pour la prédiction des paramètres physicochimiques de l'eau du barrage Ghrib.

Pour atteindre l'objectif assigné à cette thèse, trois chapitres ont été développés. La présente introduction générale a permis de situer le contexte et les motivations de cette recherche.

Le premier chapitre est subdivisé en deux parties. La première partie est consacrée à une revue critique de la littérature relative aux travaux de modélisation des différents paramètres climatiques, et la deuxième partie est consacrée à une revue critique de la littérature relative aux travaux de modélisation des différents paramètres physico-chimiques de l'eau par les différentes méthodes d'intelligences computationnelles.

Des généralités sur les différentes méthodes d'intelligence computationnelle utilisée dans le présent travail sont présentées dans le second chapitre.

Les résultats de contributions réalisées par différentes approches dans ce travail, les techniques d'intelligences computationnelles (réseaux de neurones artificiels (ANN), système d'inférence neuro-flou adaptatif (ANFIS) et les machines à vecteurs support (SVM), pour la modélisation des paramètres climatiques, ainsi que pour la modélisation des paramètres physico-chimiques de l'eau sont présentés dans le troisième chapitre.

Les conclusions tirées de ce travail ainsi que les études qui peuvent être envisagées dans le futur sont présentées à la fin de cette thèse.

CHAPITRE I

ETAT DE L'ART SUR L'UTILISATION DES APPROCHES D'INTELLIGENCE ARTIFICIELLE POUR LA MODELISATION DES PARAMETRES DE POLLUTION DE L'EAU ET DES PARAMETRES CLIMATIQUES

CHAPITRE I

Etat de l'art sur l'utilisation des approches d'intelligence artificielle pour la modélisation des paramètres de l'eau et des paramètres climatiques

I.1. Relation eau-climat

I.1.1. Généralités sur les eaux de surfaces

Les eaux de surfaces englobent toutes les eaux circulantes ou stockées à la surface des continents (rivières, lacs, étangs, barrages, …). Elles ont pour origine soit des nappes profondes dont l'émergence constitue une source de ruisseau ou de rivière, soit les eaux de ruissellement. Ces eaux se rassemblent en cours d'eau, et sont caractérisées par une surface de contact eau atmosphère toujours en mouvement et une vitesse de circulation appréciable. Elles sont caractérisées par une surface d'échange eau-atmosphère quasiment immobile, une profondeur qui peut être importante et un temps de séjour appréciable. La composition chimique des eaux de surface dépend de la nature des terrains traversés par ces eaux durant leurs parcours dans l'ensemble des bassins versants (Champiat et Larpent, 1988).

I.1.2. Caractéristiques des barrages

Un barrage est une étendue d'eau, plus ou moins profonde, dont la rétention des eaux est due à l'existence d'une contre-pente qui peut se remplir par ruissellement (ou par des apports d'eaux de sources), et qui ne subit pas l'influence des mers et océans du globe. La composition des eaux tient à la nature géologique des terrains du bassin versant et à celle du fond de la retenue (tableau I.1). Elle subit des variations grâce :

- au régime des vents qui modifie la concentration en CO_2 (par agitation de l'eau),
- aux pluies qui amènent surtout des matières en suspension (MES),
- à la température qui règle les cycles de la circulation de la masse d'eau.

La surface des barrages est en contact avec l'atmosphère et possède une température sensiblement identique. Les parties les plus profondes sont occupées par de l'eau à une densité supérieure à celle de la surface. Cette densité peut atteindre le maximum au fond, pour une température de 4°C.

Tableau I.1 : Caractéristiques des eaux de surface (Degrement, 1989).

Caractéristiques des eaux de surface	
Température	Variable suivant les saisons
Turbidité, MES (vraie ou colloïdale)	Variables, parfois élevées
Couleur	Liée surtout aux MES (argile, algue, etc.) sauf dans les eaux très douces ou acides (acide humique)
Minéralisation globale	Variable en fonction des terrains, des rejets, et des précipitations
Fer et manganèse divalents à l'état dissous	Généralement absents sauf en profondeur des pièces d'eau en état d'eutrophisation
CO_2 agressif	Généralement absent
O_2 dissous	Le plus souvent, au voisinage de la saturation. Absent dans le cas d'eaux très polluées
H_2S	Généralement absent
NH_4^+	Présent dans les eaux très polluées
NO_3^-	Peu abondant, en général
Micropolluants minéraux et organiques	Les sources des micropolluants sont les effluents industriels, agricoles, domestiques et urbains, souvent présents en faibles doses.
Éléments vivants	Bactéries (dont certaines sont pathogènes), virus, plancton (animal et végétal)
Solvants chlorés	Rarement présents
Caractère eutrophe	Fréquent, accentué par les températures élevées

I.1.3. Paramètres physico-chimiques et biologiques des eaux de surfaces

I.1.3.1. Température de l'eau

La température de l'eau, est un facteur écologique qui entraîne d'importantes répercutions écologiques (Leynaud, 1968). Elle agit sur la densité, la viscosité, la solubilité des gaz dans l'eau, la dissociation des sels dissous, de même que sur les réactions chimiques et biochimiques, le développement et la croissance des organismes vivants dans l'eau et particulièrement les microorganismes (W.H.O, 1987). La valeur de ce paramètre est influencée par la température ambiante.

I.1.3.2. pH

Le pH correspond, pour une solution donnée, à la concentration d'ions d'hydrogène et exprime l'acidité ou l'alcalinité d'une eau. L'échelle des pH s'étend de 0 (très acide) à 14 (très alcalin) ; la valeur médiane 7 correspond à une solution neutre à 25°C. Le pH d'une eau naturelle peut varier de 4 à 10 en fonction de la nature acide ou basique des terrains traversés (Beaudry, 1984). Des pH faibles (eaux acides) augmentent notamment le risque de présence

de métaux sous une forme ionique plus toxique. Des pH élevés augmentent les concentrations d'ammoniac toxique pour les poissons.

I.1.3.3. Conductivité électrique

La conductivité électrique (CE) est une expression numérique de la capacité d'une solution à conduire le courant électrique. La plupart des sels minéraux en solution sont de bons conducteurs. Par contre, les composés organiques sont de mauvais conducteurs. La conductivité électrique standard s'exprime généralement en milli-siemens par cm (mS/cm) à 20°C. La CE d'une eau naturelle est comprise entre 50 et 1500 µS/cm (Beaudry et Henry, 1984).

I.1.3.4. Matières en suspension (MES)

Les matières en suspension comprennent toutes les matières minérales ou organiques qui ne se solubilisent pas dans l'eau. Elles incluent les argiles, les sables, les limons, les matières organiques et minérales de faible dimension, le plancton et autres micro-organismes aquatiques. La quantité de matières en suspension varie notamment selon les saisons et le régime d'écoulement des eaux, la nature des rejets, etc. Ces matières affectent la transparence de l'eau et diminuent la pénétration de la lumière et, par suite, la photosynthèse. Elle peut également gêner la respiration des poissons. Par ailleurs, la matière en suspension peut contenir des quantités élevées de matières toxiques (métaux, pesticides, huiles minérales, hydrocarbures aromatiques polycycliques, …). La matière en suspension est exprimée en mg/l (Marcel, 1995).

I.1.3.5. Résidus secs (RS)

C'est la mesure de la quantité de matières minérales dissoute dans l'eau. Elle donne une indication du degré de minéralisation de l'eau. Les résidus secs sont obtenus après évaporation de l'eau à 180°C (Marcel, 1995). Les résidus secs donnent une information sur la teneur en substances dissoutes non volatiles. Suivant le domaine d'origine de l'eau, cette teneur peut varier de moins de 100 mg/l (eaux provenant de massifs cristallins) à plus de 3000 mg/l.

I.1.3.6. Oxygène dissous (OD) et % de saturation en oxygène

L'oxygène est l'un des paramètres particulièrement utiles pour l'eau et constitue un excellent indicateur de sa qualité. C'est un des paramètres les plus sensibles à la pollution. Sa valeur nous renseigne sur le degré de pollution et par conséquent sur le degré de l'auto-épuration d'un cours d'eau.

L'oxygène dissous dans les eaux de surface provient essentiellement de l'atmosphère et de l'activité photosynthétique des algues et des plantes aquatiques (Beaupoil et Bornens,

1997). La concentration en oxygène dissous varie de manière journalière et saisonnière car elle dépend de nombreux facteurs tels que la pression partielle en oxygène de l'atmosphère, la température de l'eau, la salinité, la pénétration de la lumière, l'agitation de l'eau et la disponibilité en nutriments. Cette concentration en oxygène dissous est également fonction de la vitesse d'appauvrissement du milieu en oxygène par l'activité des organismes aquatiques et les processus d'oxydation et de décomposition de la matière organique présente dans l'eau (Beaupoil et Bornens, 1997). Globalement, plus la concentration en oxygène dissous (OD) est faible, plus l'aptitude de l'eau à absorber la pollution est grande.

I.1.3.7. Charge en matières organiques : demande chimique en oxygène (DCO) et demande biochimique en oxygène (DBO)

Deux méthodes permettant d'évaluer la quantité en matière organique présente dans l'eau sont généralement utilisées : la demande chimique en oxygène (DCO) et la demande biochimique en oxygène (DBO). Ces deux méthodes se basent sur la différence entre la teneur en oxygène dissous initiale et la teneur en oxygène dissous finale après oxydation de la matière organique présente dans un échantillon d'eau.

La demande chimique en oxygène (DCO) correspond à la quantité d'oxygène nécessaire pour la dégradation de la matière organique présente dans l'eau. Elle permet de mesurer la teneur en matières organiques totales, y compris celles qui ne sont pas dégradables par les bactéries. Il s'agit donc d'un paramètre important permettant de caractériser la pollution globale d'une eau par des composés organiques.

La demande biochimique en oxygène (DBO) représente la quantité d'oxygène utilisée par les bactéries pour décomposer partiellement ou pour oxyder totalement les matières biochimiques oxydables présentes dans l'eau et qui constituent leur source de carbone (graisses, hydrates de carbone, tensioactifs, etc.).

La dégradation complète de la matière organique peut être relativement longue (plusieurs semaines). D'autre part, l'oxydation des dérivés ammoniacaux et des nitrites en nitrates (nitrification) absorbe également de l'oxygène. Cette nitrification, dans les eaux naturelles, ne débute qu'au bout d'une dizaine de jours. Pour ces deux raisons, on mesure la DBO en 5 jours, ou DBO_5, c'est-à dire la quantité d'oxygène consommée pendant ce laps de temps, pour l'oxydation partielle des matières organiques biodégradables sous l'action des microorganismes ; elle s'exprime en milligrammes d'oxygène par litre (mg O_2/l) (Beaudry et Henry, 1984 ; Marcel, 1995 ; Degrement, 2005). Le rapport DCO/DBO_5 permet d'évaluer le

caractère biodégradable de la matière organique (le rapport à peu varier entre 1,5 et 2) **(**Marcel, 1995).

I.1.3.8. Eléments nutritifs

On trouve l'azote à l'état naturel dans l'air, l'eau et le sol. L'élément azote est un élément indispensable à tous les organismes vivants, il est constamment recyclé dans l'environnement grâce à divers processus : fixation, assimilation, ammonisation, nitrification, dénitrification, etc. C'est un cycle qui transforme chimiquement l'azote en nitrate, nitrite, ammoniac ou en substance organique.

Les concentrations en nitrites (NO_2^-), nitrates (NO_3^-), ammoniac (NH_3), ammonium (NH_4^+) et azote (N) sont dès lors des paramètres importants pour le suivi de la qualité des eaux de surface. L'azote "Kjeldahl" représente l'azote organique (ex. acides aminés, urée) et l'azote ammoniacal. Quant à l'azote "total", il correspond à la somme de l'azote organique, de l'azote ammoniacal, des nitrites et des nitrates.

Les nitrates (NO_3^-) sont des ions naturels présents partout dans l'environnement. Ils sont le produit de l'oxydation de l'azote (qui représente 78% de l'atmosphère) par les microorganismes dans les plantes. Le nitrate est la forme oxydée de l'azote qui est la plus stable, mais il peut être réduit en nitrite, modérément réactif, par action microbienne. Le nitrate, étant un sel très soluble dans l'eau et très mobile dans le sol, pénètre dans la nappe phréatique lorsque sa quantité dépasse ce que les plantes peuvent utiliser.

Les nitrites (NO_2^-) constituent le stade intermédiaire entre les ions ammonium (NH_4^+) et les nitrates. Peu stable en rivière, on ne les rencontre que lorsqu'il existe un déséquilibre au niveau de l'oxygénation ou de la flore bactérienne de la rivière (Méthodes d'analyse et d'appréciation des cours d'eau, 2010).

On trouve l'azote ammoniacal (NH_3 et NH_4^+) dans les eaux polluées à cause de la biodégradation des protéines des acides aminés de l'urée ; sa présence indique une pollution récente, il est graduellement oxydé en nitrites, puis en nitrates (Méthodes d'analyse et d'appréciation des cours d'eau, 2010). La présence d'ammonium en quantité importante est l'indice d'une contamination par des rejets d'origine humaine ou industrielle. Les urines humaines ou animales contiennent en effet de grandes quantités qui se transforment rapidement en ammoniaque. Ce paramètre est souvent utilisé comme traceur des eaux usées domestiques (Méthodes d'analyse et d'appréciation des cours d'eau, 2010).

I.1.4. Généralités sur le climat

Selon l'Organisation Météorologique Mondiale (OMM), le climat est défini comme "la synthèse des conditions météorologiques dans une région donnée, caractérisée par des statistiques à long terme (valeurs moyennes, variances, probabilités de valeurs extrêmes, etc.) des éléments météorologiques (températures, précipitations, vents) dans cette région". Les régions continentales du globe connaissent un climat qui dépend de leur latitude et de la distance qui les sépare des principaux plans d'eau de la terre.

I.1.4.1. Paramètres climatiques

Les stations de mesure météorologique se composent généralement d'un mat sur lequel des capteurs sont installés. Ces derniers sont reliés à un boitier qui enregistre, stocke et envoie généralement les mesures via le réseau mobile à une base de données. Les stations mesurent généralement les paramètres environnementaux suivants : la température de l'air, l'humidité relative, la pression atmosphérique, les précipitations, la vitesse et la direction du vent, l'enneigement et les radiations solaires. Ces données sont très utiles. Elles permettent par exemple de quantifier le cycle de l'eau, de prévoir le temps, de simuler des processus environnementaux complexes et de prédire les dangers naturels.

I.1.4.1.1. Température de l'air

La température est considérée comme une grandeur physique liée à la notion immédiate de chaud et du froid. La température est la manifestation, à l'échelle macroscopique, du mouvement des atomes et molécules. L'unité internationale de la température est le kelvin (°K).

Tous les appareils mesurant la température fonctionnent sur un principe similaire. Un élément de référence possède des caractéristiques connues qui varient selon la température. Pour illustrer ce propos, prenons le cas du thermomètre à mercure. Lorsque la température est élevée, le mercure se dilate alors qu'il se contracte lors de basses températures.

La course de la terre autour du soleil et son inclinaison influence grandement la température et plus particulièrement sa variation au cours du temps et de l'espace. En conséquence, au fur et à mesure de la progression de la Terre sur son orbite autour du Soleil, l'orientation des rayons solaires varie au cours de l'année selon la latitude. Ainsi, la latitude, la période de l'année et l'altitude influencent la température.

I.1.4.1.2. Humidité relative

L'humidité représente la quantité de vapeur d'eau présente dans l'air, sans compter l'eau liquide et la glace. On doit distinguer l'humidité relative et l'humidité absolue. L'humidité

relative joue un rôle dans la formation du brouillard, de la rosée et des nuages. En général, quand on parle de mesure d'humidité, on fait allusion à l'humidité relative exprimée en %. L'humidité relative de l'air correspond au rapport de la pression partielle de vapeur d'eau contenue dans l'air sur la pression de vapeur saturante à la même température et pression. Ce rapport changera si on change la température ou la pression, ce qui rend sa mesure complexe.

L'humidité relative est donc une mesure du rapport entre la teneur en vapeur d'eau de l'air et sa capacité maximale à en contenir dans ces conditions. L'humidité relative est souvent appelée degré hygrométrique. Elle suit une échelle allant de 0 à 100 %. Un air saturé en vapeur d'eau a une humidité relative de 100 % ; un air très sec, une humidité de 10 à 20 %. La pression de vapeur saturante et l'humidité relative dépendent de la température, plus la température de l'air est élevée, plus il peut contenir de vapeur d'eau.

On définit l'humidité absolue comme le rapport de la masse de vapeur d'eau, généralement en "g" sur le volume d'air humide en m^3 à la pression et la température considérées. On peut aussi la définir comme le produit de l'humidité relative par l'humidité absolue de saturation. La mesure de l'humidité relative peut se faire grâce à deux types d'instrument : l'hygromètre et le psychromètre. L'humidité relative est étroitement liée à d'autres grandeurs physiques, telles que la température et la pression. Il est difficile de quantifier ces variations, son évolution et ses tendances.

I.1.4.1.3. Précipitations

Les précipitations sont les eaux qui tombent sur la surface de la terre, tant sous forme liquide (bruine, pluie, averse) que sous forme solide (neige, grésil, grêle) ou déposée (rosée, gelée blanche, givre, ...). Quelle que soit la forme de la précipitation, on mesure la quantité d'eau tombée durant un certain laps de temps. On l'exprime généralement en millimètres (mm). L'intensité de la pluie est la hauteur d'eau précipitée par unité de temps (généralement en mm/h).

Le pluviomètre recueille la quantité d'eau totale précipitée à l'intérieur d'un récipient calibré dans un intervalle de temps séparant deux relevés. Le pluviographe, quant à lui, mesure les hauteurs d'eau ainsi que leur répartition dans le temps, c'est-à-dire les intensités des précipitations. Les radars météorologiques envoient des ondes radios qui sont réfléchies par les pluies présentes dans l'atmosphère. Les informations reçues en "écho", permettent de localiser les précipitations en temps réel jusqu'à 200 km de distance et de mesurer leur intensité, puisque le signal réfléchi est proportionnel à la taille des précipitations.

Chacun de ces deux systèmes (pluviomètre et radar) a des spécificités propres (prix, type d'information, résolution spatiale, …). Ainsi, par exemple, un pluviomètre fournit une information ponctuelle alors que le radar renseigne sur la répartition spatiale de la pluie, il produit une surface de pluie. C'est pourquoi ces systèmes sont complémentaires.

I.1.4.1.4. Radiation solaire

Le rayonnement solaire est une onde, provenant du soleil, fournissant de l'énergie. Son unité est le Watts par mètre carré (W/m^2). En fonction de longueur d'ondes observées et du lieu d'observation (terre ou espace), il existe différents appareils de mesure. L'héliographe fut inventé par Campbell vers 1853, puis modifié par Stokes en 1879. L'héliographe est une sphère de verre constituant une lentille convergente qui concentre les rayons du soleil sur une bande de mesure (bande de carton, cellules photovoltaïque) permettant d'estimer la durée d'insolation. Un grand nombre d'appareils, aux dénominations très variées, mesurent les différentes composantes du rayonnement solaire intéressant la météorologie. Le pyranomètre mesure le rayonnement solaire global, provenant du Soleil et de la totalité de la voûte céleste. Le pyrhéliomètre, ou actinomètre, mesure le rayonnement solaire direct, provenant seulement du disque solaire.

I.1.4.1.5. Vent

Le vent est le paramètre physique représentatif des mouvements de l'air. Il naît de la différence de pression. Le vent se déplace des hautes pressions (anticyclones) vers les basses pressions (dépressions). La direction et la vitesse du vent sont des grandeurs mesurables dont la connaissance est nécessaire à l'étude de la dynamique des masses d'air. La direction indique d'où souffle le vent. Ses unités sont soit les points cardinaux (N, S, E, O) ou les degrés centigrades. La vitesse s'exprime soit en mètre par seconde (m/s), soit en kilomètre par heure (km/h) soit en nœud (1 nœud correspond à une distance de 1 mille nautique parcouru en 1 heure, soit 1,852 km/h). 1 m/s équivaut à 3,6 km/h et environ 2 nœuds.

Pour mesurer la vitesse du vent, on utilise un anémomètre. Cet instrument se compose de 3 petites coupelles hémisphériques, disposées à l'extrémité de bras métalliques espacés de 120°. Un compteur relié à l'axe de rotation permet de lire la vitesse et de l'enregistrer.

I.2. Changement climatique et cycle hydrologique

Le système climatique est défini comme un système complexe formé de cinq composantes majeures : l'atmosphère, l'hydrosphère (océans, rivières, lacs, eaux souterraines, ...), la cryosphère (neige, glaciers, banquise, ...), les surfaces continentales et labiosphère (êtres vivants), et des interactions entre celles-ci. Les interactions entre les différentes

composantes ont lieu sous forme de transferts de masse (eau, carbone, sédiments) et d'énergie (rayonnement, flux de chaleur). De par ces échanges, les différentes composantes du système climatique sont dépendantes les unes des autres.

En d'autres termes, la modification des caractéristiques de l'une d'entre elles aura nécessairement un effet sur au moins une autre de ces composantes.

Bien que le couplage entre le système climatique et le cycle de l'eau soit évident, les impacts hydrologiques du changement climatique restent relativement difficiles à caractériser. Cette difficulté provient de la diversité des processus mis en jeu : physiques, biologiques ou encore chimiques, ainsi que des différentes échelles spatiales et temporelles à prendre en compte. La variabilité naturelle est notamment beaucoup plus importante pour les précipitations que pour les températures, et la détection d'une tendance est donc plus difficile à mettre en évidence sur les premières. Néanmoins, grâce à la densification des réseaux de mesures hydrométéorologiques depuis le milieu du XX^{ème}, certains changements sont d'ores et déjà détectés dans plusieurs régions de la planète. D'autre part, les différentes simulations climatiques sur l'évolution du climat à l'échelle planétaire nous donnent un aperçu des changements auxquels s'attendre pour le futur.

I.2.1. Etat de l'art sur la prévision des variables météorologiques

La prévision météorologique est une tâche complexe et souvent difficile pour prédire l'état atmosphérique d'un lieu donné pour une période donnée dans le futur. Aussi, la compréhension du climat d'un site donné et l'utilisation intelligente des variables climatiques qui en résultent est essentielle pour la gestion de l'énergie dans les bâtiments, pour des applications telles que les installations photovoltaïques, pour des applications agricoles, pour le contrôle de l'environnement des serres et la gestion de l'irrigation, etc. Actuellement, on dénombre plus de cinquante variables climatiques vitales, dites essentielles, qui sont la clé pour des observations climatiques durables. Elles ont été choisies en raison de leur importance mais aussi parce qu'elles sont techniquement et économiquement accessibles à l'observation systématique. Ces variables climatiques essentielles sont réparties selon les trois domaines classiques : atmosphère (température de l'air, précipitation, pression, vitesse et direction du vent, humidité, propriétés des nuages, composition en dioxyde de carbone, méthane et autres gaz à effet de serre, aérosols ainsi que leurs précurseurs), océan (température de surface de la mer, salinité superficielle, niveau de la mer, état de la mer, glaces de mer, courants de surface, couleur de l'océan, pression partielle de dioxyde de carbone, acidité, phytoplancton, salinité, nutriments, teneur en oxygène dissous, acidité) et

surfaces (débit des rivières, utilisation de l'eau, eau souterraine, lacs, couverture neigeuse, glaciers, couverture des sols, durée d'ensoleillement, biomasse aérienne, carbone dans les sols, perturbation induite par les feux, humidité des sols). Notons que la plupart de ces variables sont dépendantes les unes des autres.

En raison de leurs répercussions immédiates et durables sur le milieu naturel et sur l'homme, les questions de changement et de variabilité climatiques sont placées depuis quelques temps au centre des préoccupations des scientifiques. Par conséquent, la prévision de paramètres météorologiques tels que la température de l'air, la vitesse du vent, l'humidité relative, les précipitations, etc. est d'une très grande utilité. Cette prévision est conduite moyennant l'utilisation de techniques d'apprentissage statistique, comme les réseaux de neurones artificiels (RNA) et les systèmes d'inférence neuro-flou. En effet, ces techniques ont été utilisées avec succès dans divers domaines de la science et de l'ingénierie en raison de leur capacité à modéliser des systèmes linéaires et non linéaires sans avoir à formuler des hypothèses implicites dans les approches statistiques conventionnelles. Cette revue de la littérature a porté sur la prédiction de quelques variables climatiques par utilisation des techniques d'apprentissage statistique.

1.2.1.1. Prédiction de la température de l'air

L'un des paramètres météorologiques les plus importants est la température de l'air. Cette dernière est l'un des facteurs les plus importants qui influencent la croissance, le développement et le rendement des cultures, entraînant un nombre incalculable de réactions biochimiques. C'est également l'une des variables explicatives les plus efficaces d'estimation de l'évapotranspiration, laquelle est nécessaire aux politiques de planification de l'irrigation.
La prévision de la température de l'air est préoccupante dans les domaines de l'environnement, de l'industrie et de l'agriculture. Aussi, cette prévision a attiré l'attention des chercheurs ces dernières années, car elle a de nombreuses applications différentes dans des domaines tels que l'industrie, l'agriculture ou l'énergie. Certaines de ces applications sont le dimensionnement dans les systèmes de climatisation des bâtiments, la conception de systèmes à énergie solaire, le calcul des températures dans les serres afin d'éviter les pertes de récolte, la prévision d'aléas naturels tels que incendies, la neige, le grésil, etc.

A titre d'exemple, connaître la variation de la température ambiante est important en agriculture car des changements extrêmes de la température de l'air peuvent causer des dommages aux plantes et aux animaux. Cette prévision est également utile pour connaître la probabilité de tornade et d'inondation dans une zone donnée. Dans le domaine de l'énergie, la

température de l'air est un bon indicateur de la consommation d'énergie maximale, car elle peut être utilisée pour prévoir l'utilisation massive de systèmes de chauffage en hiver ou de climatiseurs en été.

Comme les systèmes d'observation météorologique sont coûteux et leur installation complexe, des modèles de prévision de la température de l'air sont de ce fait nécessaires. Au cours des dernières années, différentes approches de calcul ont été appliquées dans différents domaines aux fins de prédiction de la température de l'air (Figure I.1). La majorité de ces approches ont utilisé des techniques informatiques neuronales, rapides et fournissant des résultats précis. Par exemple, plusieurs travaux récents ont appliqué des perceptrons multicouches (MLP) à des fins de prévision de la température, dans différents scénarios ou pays. D'autres types de réseaux de neurones, tels que la fonction de base radiale (RBF), les réseaux de régression généralisés ainsi que des approches statistiques ont également été appliqués à la prévision de la température de l'air (Tasaduqq et Rehman Bubshait, 2002 ; Smith et al., 2007 ; Smith et al., 2009 ; Ustaoglu et al., 2008 ; Boi, 2004 ; Huld et al., 2006 ; Dombayc et al., 2009).

Smith et al. (2007) ont élaboré un projet dont le but était d'explorer l'application de réseaux de neurones artificiels (RNA) pour la prévision de la température de l'air pendant toute l'année. Le modèle d'ANN a été développé à l'aide de données météorologiques détaillées recueillies par le réseau de surveillance automatisée de l'environnement de la Géorgie (AEMN). Le modèle a été en mesure de fournir des prévisions tout au long de l'année, avec une erreur absolue moyenne (MAE) des modèles annuels qui était inférieure pendant les mois d'hiver à celle des modèles résultant de l'application de modèles spécifiques à l'hiver précédemment développés.

Dans une autre étude (Chronopoulos et al., 2008) des modèles de réseaux de neurones artificiels (RNA) sont développés pour estimer les valeurs de données météorologiques dans les zones où les stations météorologiques sont rares dans le sud de la Grèce. Un modèle d'interpolation plus traditionnel (modèle de régression multiple, MLR) est également utilisé pour comparer les résultats et les performances du modèle. L'évaluation des résultats du modèle ANN a montré que des estimations de la température de l'air passables à très bonnes pouvaient être obtenues en fonction du nombre de stations météorologiques utilisées comme stations de référence. En outre, les performances du modèle ANN ont été meilleures que celles du modèle MLR : les valeurs d'erreur absolue moyenne se situaient entre 0,82 et 1,72 °C et entre 0,90 et 1,81 °C pour les modèles ANN et MLR, respectivement. Les auteurs ont

conclu que ces modèles peuvent donc être adoptés comme une composante importante de diverses études de modélisation et de gestion environnementales.

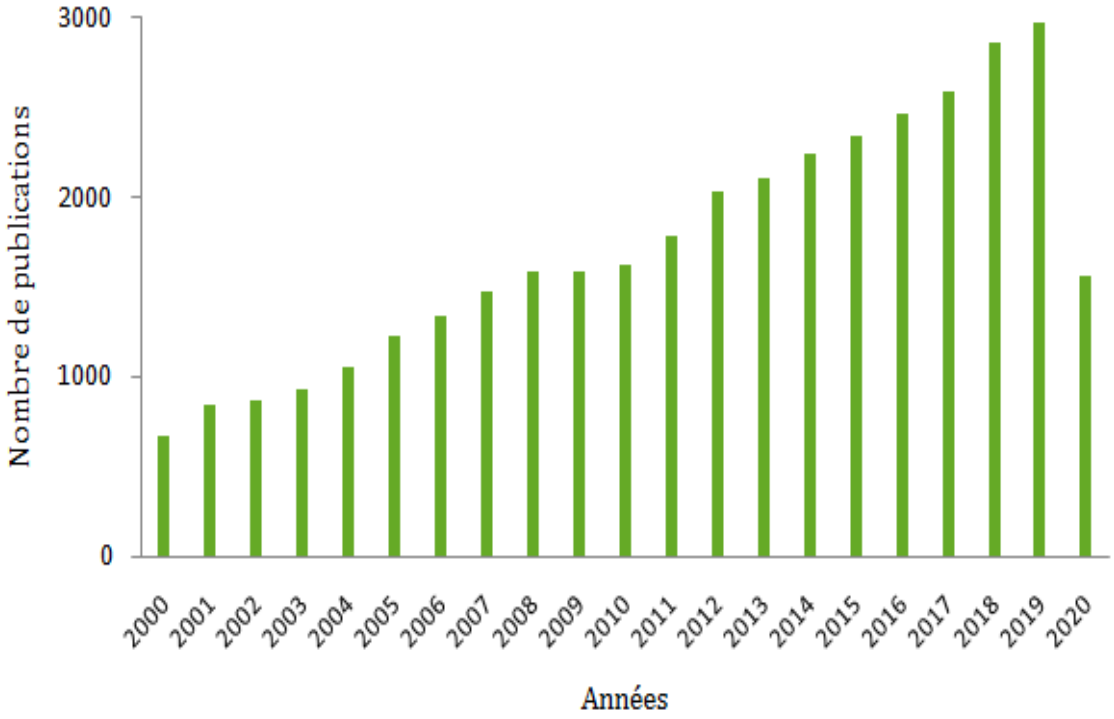


Figure I.1 : Nombre de travaux inhérents à la température de l'air publiés durant les années 2000 à 2020 (site PubMed consulté le 20/06/2020).

Bilgili et Sahin (2009) ont utilisé des réseaux de neurones artificiels pour prédire la température et les précipitations mensuelles à long terme en tout point de la Turquie, sur la base de l'utilisation des données de stations de mesure voisines. À cette fin, les données météorologiques mesurées par le service météorologique de Turquie entre 1975 et 2006 à partir de 76 stations de mesure ont été utilisées comme données d'apprentissage (59 stations) et de test (17 stations). Quatre neurones recevant des signaux d'entrée de latitude, longitude, altitude et mois ont été utilisés dans la couche d'entrée du réseau. Deux neurones, qui produisent des signaux de sortie correspondants de la température et des précipitations mensuelles à long terme, ont été utilisés dans la couche de sortie du réseau. Les valeurs déterminées par le modèle de réseau de neurones artificiels ont été comparées aux données réelles. Les erreurs obtenues dans ce modèle se situent bien dans les limites acceptables. Au cours de la même année, Dombayc et Golcu (2009) ont prédit les températures ambiantes moyennes quotidiennes en utilisant un modèle ANN à Denizli, dans le sud-ouest de la Turquie. Ils ont utilisé les données météorologiques des années 2003-2005 et 2006 comme données d'apprentissage et de test, respectivement. Le modèle RNA sélectionné comprend 3

entrées, 6 neurones cachés et 1 sortie. L'entraînement du réseau a été réalisé à l'aide de l'algorithme de propagation de Levenberg Marquardt (LM). Les résultats montrent que l'approche RNA est un modèle fiable de prévision de la température ambiante.

Dans une étude élaborée par Paniagua-Tineo et al. (2011), une approche par algorithme de régression vectorielle de support (SVR) pour prédire la température maximale quotidienne en Europe a été présentée. Les données réelles mesurées dans les stations météorologiques européennes pendant 10 ans (température minimale, précipitations, pression atmosphérique au niveau de la mer et humidité relative) ont été utilisées et il a été montré comment l'approche SVR permet d'obtenir une prévision précise de la température maximale. Une comparaison avec d'autres méthodes neuronales, basée sur des tests statistiques, a montré que le SVR se comporte mieux qu'un perceptron multicouche dans ce problème de prédiction. Durant la même année, Afzali et al. (2012) ont utilisé l'approche ANN pour prédire la température minimale, moyenne et maximale de l'air ambiant en développant les modèles ANN pour un jour et un mois à l'avance. Les données de température continues sur 44 ans pour l'étude de cas de la ville de Kerman, en Iran, ont été utilisées dans les phases d'apprentissage, de validation et de test des modèles ANN, les valeurs des modèles ANN ont été élaborés selon deux réseaux : Feed Forward et Elman. Les valeurs des variables R, MSE et MAE dans les deux réseaux ont montré que l'approche ANN est un modèle souhaitable pour la prévision de la température de l'air ambiant, tandis que les résultats de la température moyenne du jour et de la température maximale d'un mois sont plus précis avec le réseau Elman.

Par ailleurs, dans une étude menée par Kisi et Shiri (2013), les capacités du système d'inférence neuro-fuzzy adaptatif (ANFIS) et des réseaux de neurones artificiels (RNA) ont été évaluées afin de prévoir les valeurs mensuelles à long terme de la température de l'air dans 30 stations météorologiques d'Iran. Par conséquent, les valeurs de composante de périodicité, latitude, longitude et altitude de la station ont été introduites en tant que variable d'entrée. Le modèle ANN a été généralement plus performant que le modèle ANFIS au cours de la période de test. Pour le modèle ANN, les valeurs maximales et minimales des coefficients de détermination ont été établies à 0,995 et 0,921 alors que celles du modèle ANFIS sont respectivement de 0,999 et 0,876.

Récemment, les moyennes des températures mensuelles maximales, minimales et moyennes ont été estimées en fonction des coordonnées géographiques et du nombre de mois en Turquie par Cobaner et al. (2014) moyennant l'utilisation d'un système d'inférence neuro-fuzzy adaptatif (ANFIS) et des modèles de régression linéaire multiple (MLR). Les données

mensuelles mesurées dans 275 stations enregistrant des enregistrements d'au moins 20 ans sont utilisés pour élaborer les modèles. La latitude, la longitude et l'altitude de l'emplacement ainsi que le numéro du mois sont utilisés comme variables d'entrée. Les valeurs observées sont comparées à celles prédites par les modèles ANN, ANFIS et MLR en évaluant leurs erreurs. Le modèle ANFIS s'avère donc être le meilleur.

Par ailleurs, afin d'établir un modèle mathématique empirique pour prédire la température mensuelle de l'air dans la région de Meknès au Maroc, Ben El Houari et al. (2015) ont développés deux types de réseaux de neurones artificiels (RNA) : le perceptron multicouche (MLP) et le réseau récurrent Elman. La base de données utilisée contient l'historique des paramètres météorologiques mensuels enregistrés dans la région de Meknès entre 1996 et 2013. Ces paramètres sont la pression atmosphérique, l'humidité, les précipitations, la visibilité, la vitesse du vent, la vitesse maximale du vent et le point de rosée en tant que paramètres d'entrée et la température de l'air en tant que paramètre de sortie. Les résultats obtenus, basés sur les coefficients de corrélation (R), l'erreur quadratique moyenne (MSE) et la somme des erreurs carrées (SSE) démontrent que la prévision de la température de l'air est optimale et plus efficace avec le modèle MLP, l'algorithme étant celui de Levenberg-Marquard avec l'architecture [7-4-1] et la fonction Tansig dans la couche cachée et la fonction Purelin dans la couche de sortie.

Durant la même année, Un système d'inférence neuro-fuzzy adaptatif est utilisé par Daneshmand et al. (2015) pour modéliser et prévoir la température minimale mensuelle de la ville de Mashhad, au nord-est de l'Iran. La température minimale quotidienne de la station météorologique de la ville de Mashhad recueillie sur une période statistique de 42 ans (1966-2008) a été utilisée. Un algorithme de formation hybride est utilisé pour former le système. Selon les résultats de la simulation, un coefficient de corrélation de 0,987 entre les valeurs observées et les valeurs prédites, ainsi que des écarts en pourcentage absolus moyens de 27,6% indiquent une estimation acceptable du modèle.

Très récemment, Yamamoto et al. (2017) ont développé une méthode d'étalonnage basée sur l'apprentissage machine pour la mesure de la température de l'air par un capteur à faible coût. Un réseau neuronal artificiel (RNA) a été utilisé pour équilibrer l'effet de multiples facteurs environnementaux sur les mesures. Les données ont été collectées sur 305 jours, à trois endroits différents au Japon. L'application de leur méthode montre une amélioration moyenne de l'erreur absolue moyenne (EAM) de 1,62 à 0,67.

Pour leur part, Onafuye et al. (2018) ont mis au point et validé un modèle ANFIS pour la prédiction de la température de l'air, la pression atmosphérique et l'humidité relative à Calabar, dans l'État de Cross River au Nigéria. Les résultats ont révélé que la prévision obtenue par ANFIS avait une précision de prévision via l'erreur de pourcentage absolu supérieure à 90 % pour la température, la pression, l'humidité relative et la réfractivité. Le modèle aide les concepteurs de réseaux sans fil et les autres parties prenantes ayant besoin du profil vertical des paramètres radio climatiques primaires à obtenir les données de profil vertical basées sur la valeur du paramètre mesuré en surface ne nécessitant pas la stabilisation des équipements de radiosondage dans l'atmosphère.

1.2.1.2. Prédiction de l'humidité relative (HR)

Divers domaines, tels que l'hydrologie, l'écologie, l'agriculture et la médecine, utilisent souvent l'humidité, mais sous la forme d'humidité relative. En hydrologie, les prévisions d'humidité peuvent être appliquées efficacement dans la prise de décision et l'utilisation optimale des ressources en eau.

Bien que de très nombreuses recherches sur l'humidité aient été menées dans les divers domaines (Figure I.2), peu d'études ont été consacrées dans la littérature à la prédiction de l'humidité relative par les techniques d'apprentissage statistique comme les réseaux de neurones ou les méthodes de régression.

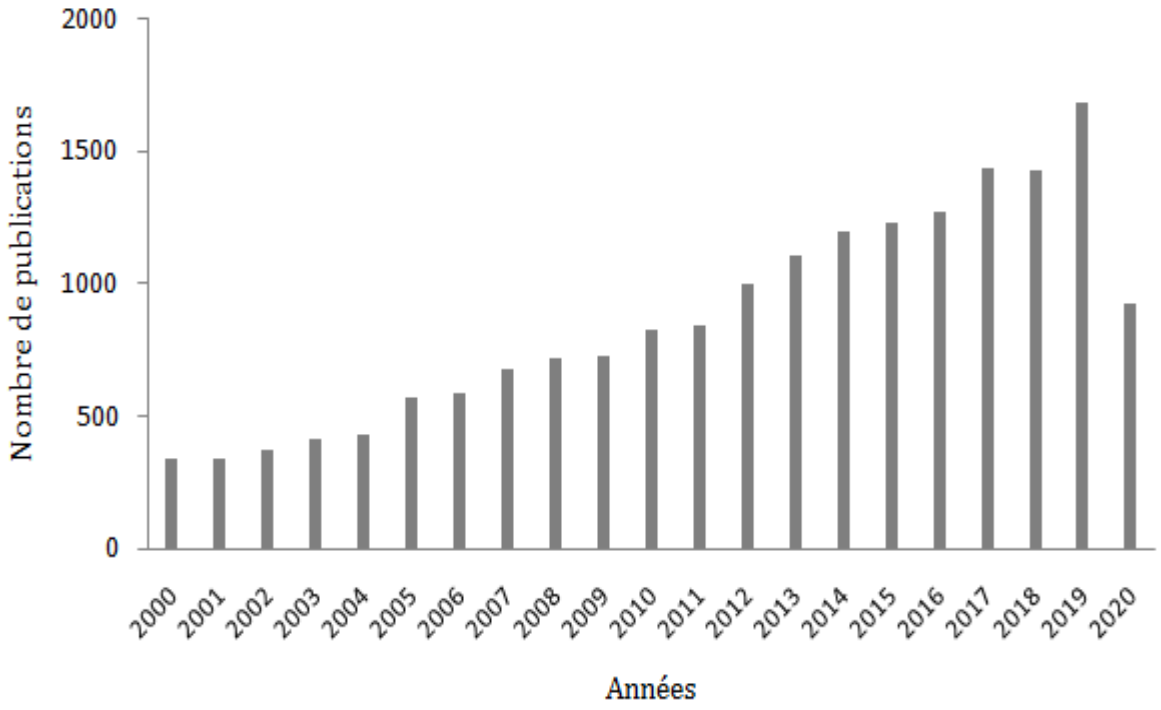


Figure I.2 : Nombre de travaux inhérents à l'humidité relative publiés durant les années 2000 à 2020 (site PubMed consulté le 20/06/2020).

Parishwad et al. (1998) ont prédit les valeurs horaires mensuelles moyennes de l'humidité relative pour un site en Inde en utilisant des valeurs horaires mensuelles moyennes de ce paramètre et ce sur la base de l'analyse de régression par les moindres carrés. Ils ont conclu que la méthode développée peut être utilisée pour prédire l'humidité relative à différents endroits de l'Inde. Lu et Viljanen (2008) ont conçu quatre modèles de réseaux neuronaux développés par auto régression non linéaire, avec un modèle d'entrée externe et un algorithme génétique, afin de prévoir la température intérieure et l'humidité relative. Les coefficients de détermination sont de 0,994 pour l'humidité relative.

Récemment, Yasar et al. (2011) ont présenté une étude dont le but est d'estimer les valeurs mensuelles d'humidité relative moyenne (HRM) dans la région égéenne de la Turquie à l'aide des paramètres topographiques et météorologiques fondés sur l'approche du réseau neuronal artificiel (RNA). Les valeurs mensuelles de HRM ont été calculées à partir des mesures effectuées dans les stations d'observation météorologique établies dans les provinces d'Izmir, Mugla, Aydin, Denizli, Usak, Manisa, Kutahya et Afyonkarahisar entre 2000 et 2006. La latitude, la longitude, l'altitude, les précipitations et les mois de l'année étaient utilisés dans la couche d'entrée du réseau RNA. Le modèle RNA a été développé à l'aide du logiciel MATLAB, et d'après les auteurs, Il semblerait que les valeurs obtenues se situaient dans les limites d'erreur acceptables. Ils ont conclu que la détermination des valeurs d'humidité relative est possible en tout point cible de la région où la mesure ne peut être effectuée.

Par ailleurs, un système d'inférence neuro-fuzzy adaptatif (ANFIS) pour extraire des profils de température et d'humidité jusqu'à 10 km au-dessus de la station tropicale Gadanki (Inde) a été élaboré par Ramesh et al. (2015). La comparaison des biais, de l'erreur absolue moyenne (EMA), de la RMSE et de l'erreur en pourcentage absolu moyen symétrique (SMAPE) des profils de température et d'humidité relative (HR) récupérés à l'aide de l'ANN et d'ANFIS a également indiqué que les profils récupérés à l'aide d'ANFIS sont meilleurs, comparés à ceux de la technique ANN.

Dans une étude basée sur la base de données en nuage des paramètres de l'industrie, un modèle de prévision amélioré basé sur les réseaux de neurones à rétro propagation (BP) a été établi par Shi et al. (2018) pour prévoir l'humidité relative (HR) toutes les 10 minutes et de 6 à 72 heures à l'avance. Les résultats expérimentaux ont montré que les prédictions de l'humidité relative du modèle présentent de fortes corrélations avec les données réelles, le coefficient de

détermination étant de 0.9778. Le modèle développé peut être utilisé pour la prévision et le contrôle de l'humidité relative dans la production industrielle.

1.2.1.3. Prédiction de la durée d'ensoleillement (DE)

La durée pendant laquelle la surface du sol reçoit le rayonnement solaire direct est dite durée d'ensoleillement. La connaissance de cette variable revêt une importance majeure étant donné qu'elle est largement utilisée dans les études de bilan énergétique atmosphérique, l'analyse des charges thermiques sur les bâtiments, la recherche sur le climat et l'évaluation des ressources agricoles. Par ailleurs, les ressources énergétiques mondiales étant limitées. Il est devenu impératif maintenant de diversifier les moyens de production d'énergie en utilisant de nouvelles sources d'énergie propres et renouvelables, telles que l'énergie éolienne, l'énergie solaire photovoltaïque et l'énergie solaire. L'importance de cette dernière augmente car la demande en énergie augmente chaque jour. De ce fait, de nombreuses études sont en cours et continueront de l'être à l'avenir afin de bénéficier de l'énergie solaire inestimable et propre.

Estimer avec précision et analyser la durée d'ensoleillement (DE) dans le temps et dans l'espace est devenu alors une nécessité impérative pour les chercheurs travaillant en météorologie, en hydrologie, en agriculture, etc. Bien que la DE puisse être mesurée presque dans toutes les stations météorologiques, les réseaux de stations météorologiques sont encore insuffisants et/ou limités pour certaines régions du monde en raison de contraintes géographiques et parfois financières. Afin de surmonter ce problème, les techniques d'estimation sont améliorées pour les zones où les mesures de la DE ne sont pas disponibles ou fiables, en particulier pour les régions éloignées et inaccessibles dépourvues de stations météorologiques.

Pour ce faire, différentes études ont été menées par utilisation d'une multitude de méthodes empiriques. C'est ainsi, qu'au fil des ans, de nombreux modèles ont été proposés pour l'estimation de la radiation solaire, l'estimation de sa variation dans le temps et dans l'espace, mais aussi pour définir sa relation avec les variables géographiques et climatologiques (Krivoshein et al., 2020 ; Bailek et al., 2020 ; Berrizbeitia et al., 2020 ; Ihaddadene et al., 2019 ; Benatiallah et al., 2019 ; Adeniji et al., 2019 ; Alsamamra, 2019 ; Hai et al., 2019 ; Wu et al., 2016 ; Morf, 2014 ; Yildirim et al., 2013 ; Robaa, 2008 ; Sahin, 2007 ; El-Metwally, 2005 ; Rangarajan et al., 1984 ; Yin, 1999). A titre illustratif, sur la figure II.3 est reporté le nombre d'articles consacrés aux prévisions du rayonnement solaire

durant la période s'étalant de l'année 2000 au 20 juin 2020, date de consultation de la banque PubMed.

Ces trente dernières années, les réseaux de neurones artificiels (RNA) ont été largement utilisés pour résoudre de nombreux problèmes dans de nombreux domaines. Lors de notre recherche bibliographique, on a remarqué que de nombreuses études basées sur les RNA existent dans la littérature pour l'estimation du rayonnement solaire. Cependant, il n'y a que peu d'études qui ont été dédiées à l'estimation de la durée d'ensoleillement.

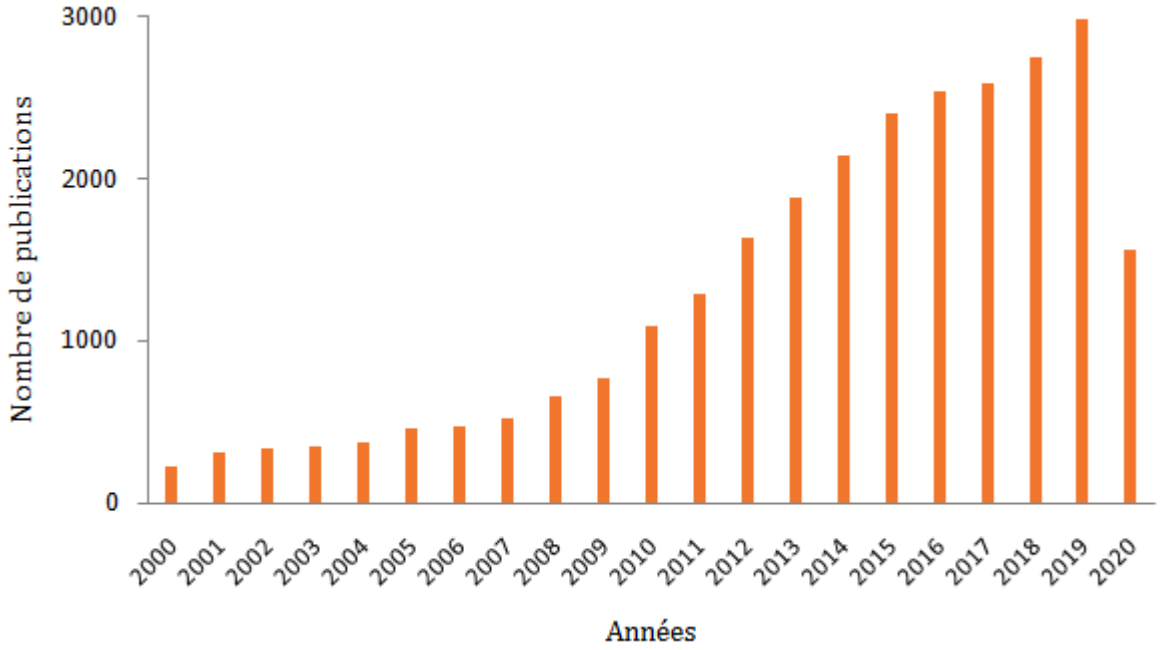


Figure I.3 : Nombre de travaux inhérents au rayonnement solaire publiés durant les années 2000 à 2020 (site PubMed consulté le 20/06/2020).

Des cartes de contour pour les heures de soleil et les ratios de soleil pour Oman ont été générées dans une étude réalisée par Jervase et al. (2003). Les données permettant de générer ces cartes ont été obtenues à l'aide d'un modèle de réseau neuronal RBF. Ce modèle estime les heures d'ensoleillement et les ratios pour un point donné en fonction de sa latitude, de sa longitude, de son altitude et du mois de l'année. Les données de 25 emplacements ont été utilisées pour tracer les cartes de contour. Ces cartes fournissent une référence nécessaire pour la répartition spatiale de l'ensoleillement et des ratios d'ensoleillement sur une base mensuelle pour l'ensemble d'Oman, à partir de laquelle des estimations d'ensoleillement pour les emplacements sans appareils de mesure coûteux peuvent être établies. Trois années plus tard, Matzarakis et Katsoulis (2006) ont présenté une étude se rapportant à la distribution temporelle et spatiale des heures d'ensoleillement et de la durée relative d'ensoleillement sur

la Grèce. Les ensembles de données utilisés pour cette étude ont été obtenus des archives du service météorologique national hellénique (HNMS). En général, une bonne relation entre les valeurs d'ensoleillement estimées et observées a été obtenue.

Mohandes et Rehman (2013) ont proposé deux algorithmes d'apprentissage automatique, à savoir, l'algorithme PSO (Particle Swarm Optimization) et SVM (Support Vector Machine) permettant d'estimer la durée d'ensoleillement dans n'importe quelle région d'Arabie saoudite à l'aide de paramètres (utilisés comme paramètres d'entrée) faciles à obtenir, notamment : la longueur de jour maximale possible, le rayonnement solaire extraterrestre à cet endroit, la latitude, la longitude, l'altitude et le numéro du mois. Les pourcentages minimaux absolus d'erreur (MAPE) minimaux de 2,3% et 2,7% ont été obtenus à la station Al-Madina en utilisant les méthodes PSO et SVM, tandis que les valeurs maximales respectives de 22,9% et 16,7% ont été observées à la station Al-Numas. Les auteurs conclurent que les résultats indiquent la viabilité des méthodes proposées pour estimer la durée d'ensoleillement en Arabie saoudite, mais aussi que le système devrait fonctionner de manière similaire dans d'autres pays à environnement similaire.

Pour leur part, Kandirmaz et al (2014) ont présenté une approche par réseau de neurones artificiels (RNA) permettant d'estimer les valeurs journalières moyennes mensuelles de la durée d'ensoleillement pour la Turquie. Trois modèles ANN différents, à savoir GRNN, MLP et RBF, ont été utilisés dans les processus d'estimation. Une variable climatique (couverture nuageuse) et deux variables géographiques (durée du jour et mois) ont été utilisés comme paramètres d'entrée afin d'obtenir la durée d'ensoleillement moyenne mensuelle comme sortie. Les jeux de données de 34 stations réparties dans toute la Turquie ont été divisés en deux parties. La première partie couvrant 21 ans (1980-2000) a été utilisée pour la formation et la seconde partie couvrant les six dernières années (2001-2006) a été utilisée pour les tests. Les indicateurs statistiques ont montré que les modèles GRNN et MLP produisaient de meilleurs résultats que le modèle RBF et pouvaient être utilisés en toute sécurité pour l'estimation de l'écart-type moyen mensuel.

1.2.1.4 Prédiction des précipitations

Dans le contexte actuel du changement climatique, les précipitations sont devenues l'un des paramètres météorologiques les plus importants. La prévision de cette variable est primordiale pour différents secteurs tels que l'agriculture et la gestion des ressources en eau, de surcroit dans les pays arides et semi-arides. Par ailleurs, les précipitations constituent un apport en eau et le principal facteur de la plupart des processus écologiques du système de

végétation (Hilker et al. 2014 ; Zhao et al. 2014 ; Pravalie et al. 2014 ; Farajzadeh et Alizadeh, 2018 ; Sarmadi et Shokoohi, 2015 ; Darand et Mansouri-Daneshvar, 2014 ; Raziei, 2017). De plus, les précipitations atmosphériques constituent la principale source d'humidité du sol et sont un indicateur important qui peut être utilisé pour caractériser directement la sécheresse climatique.

Ces dernières années, plusieurs prévisions météorologiques numériques ont été proposées pour la prévision météorologique des précipitations moyennant l'utilisation de techniques telles que le réseau de neurones artificiels (RNA) ou la logique floue. Aussi, de nombreuses études ont été menées sur les prévisions de précipitations dans de nombreux pays tels que le Brésil (Ramirez et al., 2005), le Mexique (Tereza, 1999) ou la Turquie (Saplioglu et al., 2010). Par ailleurs, de nombreux modèles de RNA ont été mis au point pour la prévision des précipitations, tels que le réseau neuronal à couches multiples (Luk et al., 2001), le réseau neuronal à couches multiples à perceptron (Ramirez et al., 2005 ; Ben El Houari et al., 2014 ; Venkata Ramana et al., 2013). Cependant, pour la prévision des précipitations, le réseau de neurones MLP (Multi Layer Perceptron) est le plus largement utilisé.

Un système d'inférence neuro-fuzzy adaptatif (ANFIS) pour développer un modèle de prévision des précipitations a été proposé par Bushara et Abraham (2015). Les données météorologiques mensuelles recueillies par le Bureau central de la statistique du Soudan de 2000 à 2012 pour 24 stations météorologiques réparties dans le pays ont été utilisées. Les auteurs ont construit plusieurs modèles ANFIS en utilisant différents types de fonctions d'appartenance, différentes méthodes d'optimisation et différents ratios d'ensembles de données pour la formation et les tests. Les résultats montrent que le modèle neuro-fuzzy d'ANFIS est capable de capturer le comportement dynamique des données de précipitations et qu'il produit des résultats satisfaisants. Il peut donc être utile pour la prévision des précipitations à long terme.

Sojitra et al. (2015) ont réalisé des études pour développer des modèles de prévision des précipitations. Le système d'inférence neuro-fuzzy adaptatif (ANFIS) a été utilisé pour développer les modèles de précipitations de la ville d'Udaipur. Deux ensembles de données ont été préparés en utilisant les paramètres météorologiques sur 35 ans, à savoir la température de bulbe humide, la température moyenne, l'humidité relative et l'évaporation du jour précédent et de la semaine moyenne mobile précédente. Les fonctions d'appartenance gaussienne et généralisée de Bell ont été utilisées pour préparer des modèles. Les indices de performance statistique et hydrologique d'ANFIS ont donné de meilleures performances

parmi les quatre modèles développés. L'étude a montré que l'analyse de sensibilité révélait que la température de bulbe humide était le paramètre le plus sensible, suivie de la température moyenne, de l'humidité relative et de l'évaporation.

Dans une étude très récente, un modèle hybride basé sur le système adaptatif Neuro Fuzzy Inférence - algorithme génétique (ANFIS-GA) a été proposé par Hanefi (2019) pour estimer la quantité de précipitations régionales. Le but de l'étude est d'estimer la quantité de pluie annuelle, en assurant une gestion efficace des ressources en eau et en permettant certaines évaluations et préparations en fonction des changements climatiques possibles. Le modèle d'estimation a été développé en codant dans le programme de progiciel MATLAB.

Pour le développement du modèle, 3650 données météorologiques de 2008 à 2018 concernant la ville de Bâle (Suisse) ont été utilisées. Les données réelles ont été testées à la fois sur le réseau de neurones artificiels (ANN) et sur le modèle hybride ANFIS-GA. Les résultats obtenus ont montré que la valeur R d'apprentissage du modèle ANFIS-GA proposé était de 0,9920, la valeur R testée de 0,9840 et le taux d'erreur de 0,0011. Les auteurs ont conclu que les performances prédictives du modèle sont élevées et que le niveau d'erreur faible, ce qui fait que des approches hybrides telles qu'ANFIS-GA peuvent être facilement utilisées pour prédire des événements météorologiques.

I.2.2. Etat de l'art sur la prévision des paramètres de l'eau

Le changement climatique a un effet en chaîne sur le cycle hydrologique. Les températures plus élevées modifieront la quantité de précipitations, qui seront converties en neige et affecteront le moment et la proportion de la fonte des glaces. L'augmentation de l'évaporation conduira également à des sols plus secs. L'évolution des concentrations de CO_2 affectera également l'évapotranspiration de la végétation et donc l'augmentation potentielle de la perte d'eau. Des sols plus secs peuvent entraîner des sécheresses plus longues, ce qui modifiera également la couverture végétale, affectant ainsi le captage du ruissellement en réponse aux précipitations. De plus, la recharge en eau souterraine sera également affectée, ce qui modifiera la quantité de percolation. Ce changement modifiera par conséquent la contribution du débit de base des eaux souterraines.

En plus de la quantité d'eau, la qualité va également changer. Les précipitations extrêmes entraîneront une augmentation des eaux de ruissellement, entraînant ainsi l'évacuation des polluants des zones urbaines et des exploitations agricoles jusqu'aux plans d'eau récepteurs. Cela posera également un risque physique pour les infrastructures d'approvisionnement en eau telles que les barrages et les systèmes d'approvisionnement en

eau et de traitement des eaux usées en raison de la surcharge de contaminants et des difficultés de traitement. D'autre part, une réduction des débits et des volumes d'eau entraînerait une augmentation de la concentration en polluants. Cette revue de la littérature a porté sur la prédiction de quelques variables hydrologiques par utilisation des techniques d'apprentissage statistique.

I.2.2.1. Prédiction de la température de l'eau

La température de l'eau est l'un des facteurs les plus importants qui déterminent la santé générale des poissons et des autres organismes aquatiques dans les études sur les écosystèmes aquatiques. Il est vital pour les processus chimiques et influe sur de nombreux processus biologiques, tels que la croissance, le comportement et la mortalité des organismes aquatiques (Casas - Mulet et al 2016, Westhoff et Rosenberger 2016). Les variations de la température de l'eau des cours d'eau peuvent avoir des effets positifs ou négatifs sur les organismes. Outre le fait qu'elle contrôle plusieurs processus physiques et biochimiques dans les rivières, tels que les processus de réoxygénation et le métabolisme dépendant de la température des plantes et des animaux aquatiques (Rajwakuligiewicz et al. 2015 ; Sandersfeld et al. 2017), elle a également une influence sur la solvabilité de l'oxygène. Par conséquent, la surveillance et la prévision des changements néfastes potentiels de la température de l'eau sont très importantes en raison de son rôle écologique important dans les systèmes aquatiques. Cette importance est justement mise en évidence par le nombre important de travaux scientifiques publiés chaque année (figure I.4).

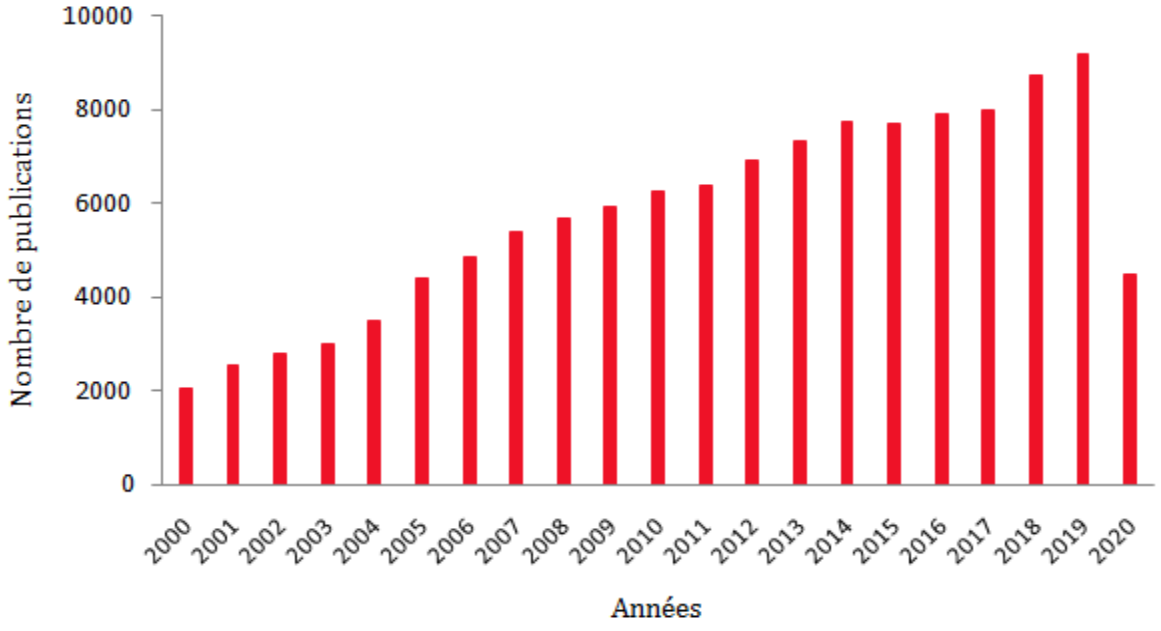


Figure I.4 : Nombre de travaux inhérents à la température de l'eau publiés durant les années 2000 à 2020 (site PubMed consulté le 20/06/2020).

Cet intérêt a favorisé le développement de modèles statistiques, qui ont été largement utilisés dans les prévisions de la température de l'eau en raison de leur simplicité et de la nécessité de disposer de données minimales. Parmi ces modèles mis en œuvre avec succès pour, on retrouve des modèles de régression linéaire (Krider et al. 2013) et non linéaire (Vliet et al. 2012), des modèles de régression stochastique (Rabi et al. 2015) et des modèles hybrides statistiques-physiques (Gallice et al., 2015 ; Toffolon et Piccolroaz, 2015 ; Piccolroaz et al., 2016). De plus, les modèles d'apprentissage statistique, ont été de plus en plus utilisés pour prévoir la température de l'eau ces dernières années, les modèles de réseaux de neurones artificiels (RNA) étant les approches les plus largement utilisées (Hadzima Nyarko et al., 2014 ; DeWeber et Wagner, 2014 ; Piotrowski et al., 2014, 2015 ; Rabi et al., 2015 ; Temizyurek et Dadasercelik, 2018 ; Zhu et al 2018a et 2018b).

Très récemment, un modèle de prévision informatique souple pour le réflecteur de Scheffler a été conçu par Deep et Prakash (2018) pour prévoir la température de l'eau obtenue par absorption de l'énergie solaire. Dans ce modèle, un système d'interface neuro-fuzzy adaptatif (ANFIS) basé sur le premier ordre du modèle Takagi-Sugeno est utilisé pour prédire le comportement du système. Les paramètres d'entrée utilisés dans le modèle sont l'intensité du rayonnement, la température ambiante et le temps d'expérimentation. Le paramètre de sortie est la température de l'eau générée. Le modèle donne une excellente capacité de prédiction pour le réflecteur de Scheffler.

I.2.2.2. Prédiction de la teneur en oxygène dissous (OD)

En égard aux diverses activités humaines, la qualité des eaux de surface est l'un des problèmes majeurs aujourd'hui en raison de ses effets sur la santé humaine et les écosystèmes aquatiques. La qualité d'une masse d'eau est généralement décrite par des ensembles de variables physiques, chimiques et biologiques interdépendantes. Les contaminants des eaux (rejets de substances dangereuses issues des processus industriels et des eaux usées urbaines, pollution accidentelle par des substances dangereuses, pollution diffuse provenant de zones agricoles) ont une incidence directe ou indirecte sur les paramètres clés de la qualité des eaux tels que la teneur en oxygène dissous (OD), la température, le pH, la conductivité, la transparence et la viscosité. Parmi ces paramètres, la quantité en OD des eaux est l'un des paramètres importants utilisé fréquemment et en continu pour déterminer la qualité de l'eau des rivières et est l'un des paramètres clés de l'analyse de la qualité de l'eau, aussi bien pour la vie aquatique que pour d'autres utilisations de l'eau.

De nombreux travaux traitant de la teneur en OD ont été réalisés au niveau international. La revue de la littérature accomplie au moyen de la base de données "PubMed" est représentée sur la figure I.5 ci-dessous. A la lecture des résultats de cette recherche, il ressort que plus de 1500 articles scientifiques ont été publiés durant la période 2000-2020 le 20 juin 2020 (date de la consultation du site).

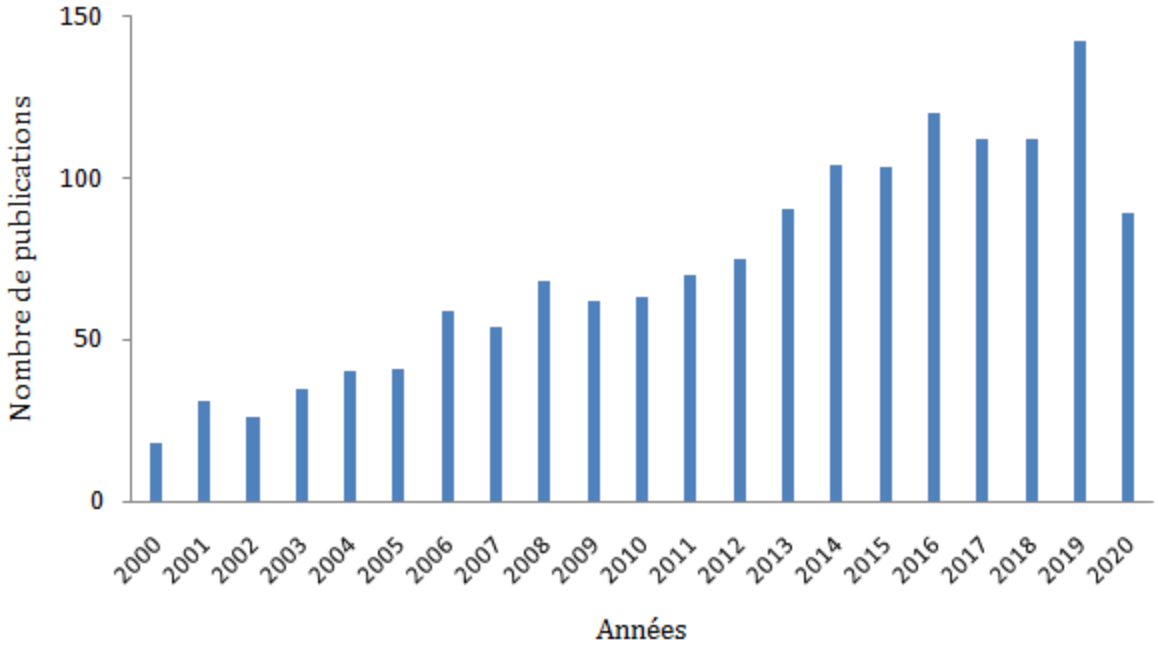


Figure I.5 : Nombre de travaux inhérents à la teneur en oxygène dissous publiés durant les années 2000 à 2020 (site PubMed consulté le 20/06/2020).

L'obtention de données sur la qualité de l'eau par une surveillance régulière est un processus fastidieux qui nécessite du personnel qualifié et des ressources stables. Pour faciliter les procédures longues et coûteuses d'acquisition de données, diverses applications de différentes approches pour quantifier les niveaux d'OD ont été utilisées. La modélisation de la qualité de l'eau est de plus en plus reconnue comme un outil utile pour acquérir des informations précieuses pour une gestion optimale de la qualité de l'eau. Etant donné qu'un grand nombre de facteurs, tels que la température, le débit d'eau, la sédimentation, la réactivation, la décomposition, la nitrification et la photosynthèse, affectent l'OD dans l'eau, une relation non linéaire entre les variables d'entrée et l'OD est à prévoir. De nombreuses approches informatiques et statistiques, comme les réseaux de neurones artificiels (RNA) ont été appliquées pour prédire la qualité de l'eau.

Ces dernières années, des progrès considérables ont été réalisés dans l'élaboration de modèles RNA pour la prévision du contenu en OD de diverses rivières. Certains exemples

d'application des RNA pour la modélisation d'OD incluent des modèles prédictifs développés pour la rivière Melen, Turquie (Samandar, 2010), Bow River, Canada (He et al., 2011), Foundation Creek in Colorado, États-Unis (Ay et Kisi, 2012 ; Ay et Kisi, 2017), le Danube, dans le nord de la Serbie (Antanasijevic et al., 2013 ; Antanasijevic et al., 2014) et la rivière Upper Klamath dans l'Oregon, aux États-Unis (Heddam, 2014). Dans ces articles, les auteurs ont testé diverses architectures, appliquées à différentes périodes, ainsi qu'à l'aide de différentes représentations de données.

A titre d'exemple, dans une étude entreprise par Ozgur et Murat (2012), la méthode du réseau neuronal à base radiale (RBNN) et du système d'inférence neuro-fuzzy (ANFIS) ont été développés pour estimer la concentration en oxygène dissous en utilisant diverses combinaisons de variables journalières d'entrée qui sont le pH, la décharge, la température et la conductivité électrique mesurées par US Geological Survey (USGS). Les données de la station de jaugeage des ruisseaux de Fountain Creek, qui couvrent 9 années de données quotidiennes entre 1990 et 1998, ont été utilisées dans l'étude. Les résultats des modèles RBNN et ANFIS ont été comparés. Le modèle RBNN s'est révélé meilleur que le modèle ANFIS pour l'estimation de la concentration en oxygène dissous.

Najah et al. (2013) ont développé un modèle pour évaluer la précision et les performances du système d'inférence neuro-fuzzy (ANFIS) pour la prévision des concentrations en oxygène dissous (OD). Le modèle a été utilisé pour analyser les données historiques générées par le biais d'une surveillance continue des paramètres de qualité de l'eau à plusieurs stations sur la rivière Johor. Quatre paramètres de qualité de l'eau ont été sélectionnés pour la modélisation ANFIS, notamment la température, le pH, la concentration en nitrates (NO_3) et la concentration en azote ammoniacal. Une analyse de sensibilité a été réalisée pour évaluer les effets des paramètres d'entrée. La température était le paramètre ayant le moins d'effet, alors que le pH apportait la plus faible contribution au modèle proposé.

Pour évaluer les performances du modèle, trois indices statistiques ont été utilisés : le coefficient de détermination (R^2), l'erreur de prédiction absolue moyenne et le coefficient de corrélation. La performance du modèle ANFIS a été comparée à un modèle de réseau neuronal artificiel. Le modèle ANFIS s'est avéré capable de fournir une plus grande précision, en particulier dans le cas d'événements extrêmes.

Des modèles basés sur le système neuro-fuzzy (ANFIS), le perceptron multicouche (MLP) et le réseau de base radial (RBF) ont été développés par Emamgholizadeh et al. (2013) pour la prédiction de l'oxygène dissous (DO), de la demande biochimique en oxygène (DBO)

et de la demande chimique en oxygène (DCO) de la rivière Karoon (Iran). Neuf variables d'entrée, mesurées dans l'eau de rivière, ont été utilisées : la conductivité électrique (CE), le pH, la concentration en calcium, magnésium, sodium, phosphates, nitrates et nitrites ainsi que la turbidité. La performance de ces modèles a été évaluée par le coefficient de détermination R^2, l'erreur fondamentale moyenne et l'erreur absolue moyenne. Les résultats ont montré que les valeurs calculées de l'OD, de la DBO et de la DCO utilisant à la fois le réseau de neurones artificiels et les modèles ANFIS étaient en parfait accord avec leurs valeurs mesurées respectives dans l'eau de rivière. Le modèle MLP était également acceptable pour prédire les variables de la qualité de l'eau. Enfin, une analyse de sensibilité a montré que le phosphate était le paramètre le plus efficace pour l'OD, la DBO et la DCO.

Des données mesurées quotidiennement sur les solides en suspension dans l'eau, le pH, la température, les rejets et la concentration en DCO en amont du système de station d'épuration municipale de la province d'Adapazari en Turquie ont été utilisées comme entrée par Ay et Kisi (2014) pour l'élaboration de modèles de prédiction de la demande chimique en oxygène (DCO). Trois techniques furent utilisées : ANFIS, MLP/ANN et k-means clustering. L'erreur quadratique moyenne, l'erreur absolue moyenne, l'erreur relative absolue moyenne et les statistiques du coefficient de détermination ont été utilisées pour l'évaluation des performances de chaque modèle. Il a été constaté que la méthode k-means-MLP donnait de meilleurs résultats que les autres techniques d'estimation de la DCO. Les auteurs ont suggéré que la technique k-means combiné au MLP pourrait être utilisée comme outil de modélisation de la concentration quotidienne de DCO.

Très récemment, diverses architectures de réseaux de neurones, telles que les réseaux de neurones à feed-forward formés à l'algorithme de rétrodiffusion, les réseaux de neurones à régression générale et la fonction de base radiale ont été principalement utilisés pour la simulation du contenu en OD pour une seule station de surveillance (Raheli et al., 2017 ; Ay et Kişi, 2017 ; Heddam et Kisi, 2017 ; Keshtegar et Heddam, 2017 ; Olyaie et al., 2017 ; Csábrági et al., 2017), alors que les études sur plusieurs sites étaient assez limitées (Ji et al., 2017 ; Siljic Tomic et al., 2018).

Des techniques de régression multilinéaire (MLR), de réseau neuronal artificiel (RNA) et de système d'inférence neuro-flou adaptatif (ANFIS) ont été développées par Abba et al. (2017) pour prédire la concentration en oxygène dissous en aval de la ville d'Agra, en utilisant des données d'entrée mensuelles suivantes : le pH, la demande biologique en oxygène (DBO) et la température de l'eau à trois endroits différents. La performance a été évaluée à l'aide du

coefficient de détermination et de l'erreur quadratique moyenne (RMSE). Les résultats ont montré que l'ANN et l'ANFIS peuvent tous deux être appliqués à la modélisation de la concentration en DO dans la ville d'Agra et indique également que le modèle ANN est légèrement meilleur qu'ANFIS. Cependant, l'ANN et l'ANFIS ont montré une supériorité considérable par rapport à la technique MLR.

I.3. Conclusion

A la lumière de cette revue bibliographique, il ressort que certaines variables hydro-climatiques revêtent une importance capitale dans plusieurs secteurs, à l'instar de l'hydrologie, l'écologie, l'agriculture, la médecine, les écosystèmes aquatiques, la gestion des ressources en eau, les études environnementales, la recherche sur le climat. Cette importance se perçoit au vu des nombreuses études scientifiques qui sont publiées chaque année.

La prévision de ces variables est une tâche plus que primordiale. Pour l'estimation ou le calcul de ces variables, Il ressort également que les réseaux de stations météorologiques sont encore insuffisants et/ou limités pour certaines régions du monde en raison de contraintes géographiques et financières. A cet égard, un intérêt continu est accordé aux modèles mathématiques de prévision de ces variables. Aussi, beaucoup de travaux inhérents aux prévisions des variables hydro-climatiques par l'utilisation de méthodes empiriques ont été passés en revue dans cet état de l'art. Cependant, l'utilisation de modèles dérivés à partir des réseaux de neurones s'avère insignifiante, particulièrement pour la prédiction de l'humidité relative, la température de l'air et la durée d'ensoleillement.

CHAPITRE II

TECHNIQUES D'INTELLIGENCE ARTIFICIELLE
POUR LA MODELISATION

CHAPITRE II

Techniques d'Intelligence Artificielle pour la Modélisation

II.1. Introduction

L'intelligence artificielle (I.A) au sens le plus large, signifierait l'aptitude d'une machine à effectuer le même type de fonctions et/ou de raisonnement qui caractérisent la pensée humaine (Kalogirou, 2003). Les techniques de l'intelligence artificielle (I.A) sont de plus en plus utilisées comme une alternative à d'autres approches classiques ou comme des composants dans des systèmes intégrés. Elles ont été utilisées pour résoudre des problèmes pratiques complexes dans divers domaines. Les techniques d'I.A ont les caractéristiques suivantes : apprentissage à partir d'exemples, tolérances vis-à-vis des pannes dans le sens où elles sont capables de traiter des données bruitées et incomplètes, capacité de traiter des problèmes non-linéaires, et une fois constituées, elles peuvent effectuer la prévision et la généralisation à grande vitesse (Mellit et al., 2009).

Les systèmes à base d'I.A sont développés et déployées dans de nombreuses applications, principalement en raison de leur raisonnement symbolique, de leur flexibilité et des capacités d'explication. L'I.A a été utilisée et appliquée dans différents secteurs, tels que l'ingénierie, l'économie, la médecine, l'armée, etc. Elle a également été appliquée pour la modélisation, l'identification, l'optimisation, la prévision, et le contrôle de systèmes complexes et/ou non-linéaires.

II.2. Modélisation par réseaux de neurones artificiels

Les réseaux de neurones artificiels sont amplement admis comme technologie présentant une manière alternative d'aborder des problèmes complexes et mal définis et sont susceptibles d'apporter des solutions efficaces et élégantes. Ils sont utilisés dans de nombreuses applications dans divers domaines. Ils sont particulièrement utilises dans la modélisation de systèmes et dans le contrôle de processus.

Un premier modèle neuronal fut élaboré en 1943 par Mc Culloch et Pitts (Mc Culloch et Pitts, 1943), se basant sur une analogie quasi directe avec le neurone biologique.

L'idée principale des réseaux de neurones est la suivante : On se donne une unité simple, un neurone (appelé neurone formel), qui est capable de réaliser des calculs élémentaires. On relie ensuite un certain nombre de ces unités sous forme de réseau capable

de résoudre un problème plus complexe. La structure, la taille du réseau et le mécanisme d'apprentissage utilisé dépendent du type et de la complexité de l'application en question.

II.2.1. Neurone formel

Le neurone formel est l'élément essentiel d'un réseau de neurones. Il peut être considéré comme un opérateur recevant un nombre variable d'entrées (x_1, ..., x_n), du milieu extérieur ou d'autres neurones, chacune de ces entrées étant pondérée par un poids dit poids synaptique (w_1, ..., w_n), et fournissant une sortie seulement quand la somme dépasse un certain seuil interne. La sortie est une fonction non linéaire f d'une combinaison linéaire p des entrées (x_i). Mathématiquement, ceci peut être modélisé par les équations suivantes :

$$p = w_0 + \sum_{i=1}^{n}(w_i x_i) \qquad \text{(II.1)}$$

$$y = f(p) = f(w_0 + \sum_{i=1}^{n}(w_i x_i)) \qquad \text{(II.2)}$$

Précisons que le poids w_0 est affecté à une entrée constante appelé "biais". f est appelée fonction d'activation, et est appliquée à ce potentiel p. Plusieurs types de fonction d'activation sont utilisés, servant à limiter l'amplitude du signal de sortie du neurone et reproduire l'effet de seuil observé sur les neurones biologiques.

Cependant, certaines fonctions sont les plus utilisées, telles que : la fonction sigmoïde, la fonction arctangente et la fonction tangente hyperbolique.

Conceptuellement le fonctionnement du neurone formel peut être représenté par le modèle de la figure (II.1).

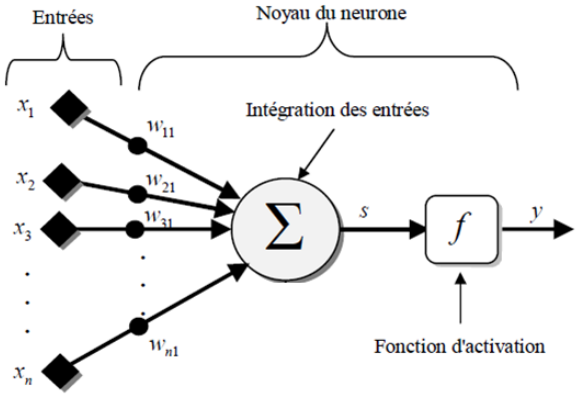


Figure II.1 : Schématisation d'un neurone formel.

II.2.2. Architecture d'un réseau de neurones

Un réseau de neurones est un maillage de plusieurs neurones, généralement organisé en couches. Ce réseau dispose de deux couches principales une d'entrée et une autre de sortie reliées par le biais de couches de connexions modifiables dites cachées.

On peut distinguer deux types de réseaux de neurones : les réseaux non bouclés et les réseaux bouclés.

- Réseaux non bouclés : l'information dans ces réseaux se propage des entrées vers les sorties, sans aucune rétroaction. Chaque neurone dans une couche est relié par des connexions avec tous les neurones de la couche suivante et de la couche précédente sans qu'il n'y ait de connexions entre neurones d'une même couche.

Les réseaux de neurones non bouclés sont des outils statiques, utilisés principalement pour effectuer des tâches d'approximation de fonctions non linéaires, de modélisation de processus statiques non linéaires.

- Réseaux bouclés : les réseaux de neurones bouclés ont des connexions récurrentes qui ramènent l'information en arrière par rapport au sens de propagation défini dans un réseau non bouclé. Un réseau de neurones bouclé est donc un système dynamique, régi par des équations différentielles ; comme l'immense majorité des applications sont réalisées par des programmes d'ordinateurs, on se place dans le cadre des systèmes à temps discret, où les équations différentielles sont remplacées par des équations aux différences.

Dans notre étude, nous avons adopté le modèle le plus couramment utilisé, le Perceptron Multi- couches, illustré par la figure II.2 et dont le principe est exposé dans ce qui suit.

II.2.3. Perceptron multicouche

Le perceptron multicouche (Multi layer perceptron MLP) est un réseau de neurones non bouclé caractérisé par une structure bien particulière (Hornik et al., 1989). La figure II.2 illustre la structure de ce type de réseau. Ses neurones sont organisés en couches successives où les informations circulent dans un seul sens, de la couche d'entrée vers la couche de sortie. Les neurones d'une même couche ne sont pas interconnectés. Un neurone ne peut envoyer son résultat qu'à un neurone situé dans une couche postérieure à la sienne. Le réseau est formé de plusieurs couches, qui s'appellent, par convention :

- couche d'entrée : cette couche représente toujours une couche associée aux entrées du système, composée des cellules d'entrée qui correspondent aux *k* variables d'entrée.
- couche(s) cachée(s) : chaque couche est composée de p neurones, dont la fonction d'activation est de la famille des courbes en "S" bornées. Ces neurones n'ont aucun lien avec l'extérieur et sont appelés neurones cachés (Rivals, 1995).
- couche de sortie : elle est constituée par la valeur de sortie de la combinaison linéaire des fonctions sigmoïdes de la couche cachée.

Les neurones sont reliés entre eux par des connexions pondérées. Ce sont les poids de ces connexions qui gouvernent le fonctionnement du réseau et programment une application de l'espace des entrées vers l'espace des sorties, à l'aide de la transformation non linéaire (Parizeau, 2004).

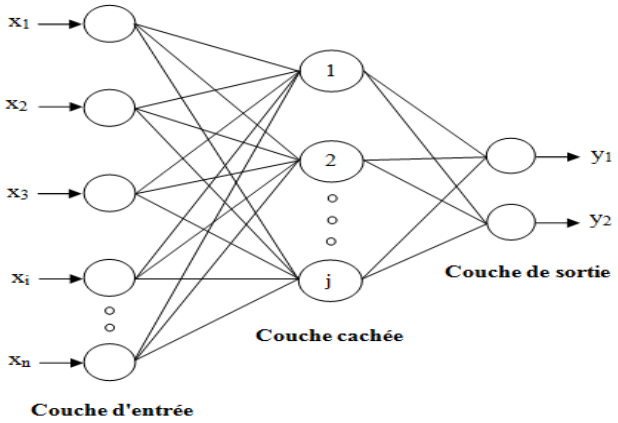


Figure II.2 : Perceptron multicouche.

II.2.4. Optimisation de la structure du réseau

La détermination de l'architecture optimale est basée sur diverses méthodes empiriques mettant en jeu plusieurs tests statistiques da validation. Comme nous l'avons mentionné, nous avons retenu le Perceptron Multicouches comme base du modèle. Nous structurerons ce réseau en précisant le nombre de couches et de neurones cachés pour que le réseau soit en mesure de reproduire ce qui est déterministe dans les données. La performance du modèle de PMC est aussi validée par d'autres paramètres statistiques des phases d'apprentissage et du test et qui correspondent au nombre optimal de neurones de la couche cachée.

II.2.5. Apprentissage

Une fois l'architecture choisie, elle doit subir une phase d'apprentissage. Cette dernière consiste à calculer les poids w_i de telle manière que les sorties du réseau de neurones soient, pour les exemples utilisés lors de l'apprentissage, aussi proches que possibles des sorties désirées. Celles-ci correspondent à la valeur de la fonction que l'on veut approcher ou à la sortie du processus que l'on veut modéliser (Dreyfus, 2008). L'apprentissage correspond donc à la phase du développement du réseau durant laquelle les coefficients de la régression sont modifiés jusqu'à l'obtention du comportement désiré. Il existe deux types d'apprentissage, supervisé ou non supervisé.

- L'apprentissage est dit supervisé lorsque les exemples sont constitués des couples de valeurs du type : (valeur d'entrée, valeur de sortie désirée). Tout le problème de l'apprentissage supervisé consiste, étant donné un ensemble d'apprentissage de n couples

(entrée-sortie associée), à déterminer le vecteur des poids w_i d'un réseau capable de mettre ces informations en correspondance.

- L'apprentissage est qualifié de non supervisé lorsque seules les valeurs d'entrée sont disponibles. Dans ce cas, les exemples présentés à l'entrée provoquent une auto-adaptation du réseau afin de produire des valeurs de sortie qui soient proches des réponses à des valeurs d'entrée similaires.

II.2.6. Méthodologie de synthèse

La synthèse des réseaux de neurones artificiels est illustrée sur l'organigramme présenté sur la figure II.3. Afin d'établir une base de données fiables et représentatives quatre étapes sont suivies :

- Collection de base de données,
- Sélection des entrées et la sortie du modèle neuronal,
- Préparation de base de données,
- Format de données.

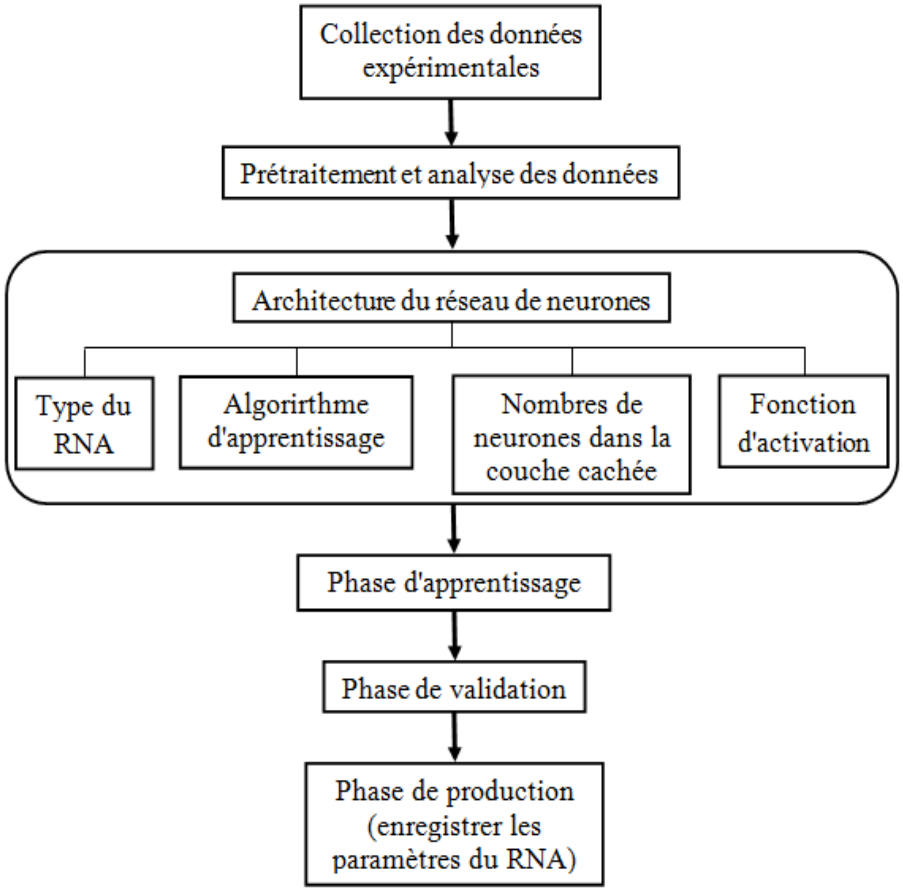


Figure II.3 : Organigramme de la synthèse d'un modèle neuronal.

II.3. Modélisation par les systèmes neuro-flous

II.3.1. La logique floue

La logique floue a été développée par Zadeh (1965) qui a proposé de modéliser un système complexe par un raisonnement "approximatif" basé sur des variables linguistiques et des sous-ensembles flous (Zadeh, 1971). Un sous-ensemble flou A est défini sur un domaine physique appelé univers de discours U, et par une fonction d'appartenance F(x) qui associe, à chaque élément x de U, le degré de vérité (d'appartenance) $f_{(A)}$ à X compris entre l'intervalle 0 et 1, soit, et cela contrairement à la logique classique où le degré d'appartenance ne peut prendre que deux valeurs (0 ou 1).

Les systèmes flous s'appuient sur une représentation de la connaissance sous forme de règles "Si … Alors" qui permettent de représenter les relations entre les variables d'entrée et de sortie dont l'expression générique est de la forme :

Si (Antécédent) **Alors** (Conséquent) (II.3)

Si (X est A) **Alors** (Y est B) (II.4)

La modélisation d'un système entrée/sortie par la logique floue passe par trois étapes essentielles illustrées sur la figure II.4 :

- La fuzzification des variables d'entrée, qui consiste à transformer les entrées numériques disponibles en parties floues. Il est alors possible d'associer à des variables des coefficients d'appartenance à des sous-ensembles flous prenant des valeurs dans l'intervalle [0, 1].

- L'inférence floue, composée par la base de règles et par la base de données. La combinaison des entrées avec les règles floues permet de tirer des conclusions.

- La défuzzification qui est l'opération inverse de la fuzzification. Elle convertit les parties floues relatives aux sorties du mécanisme d'inférence en sorties numériques.

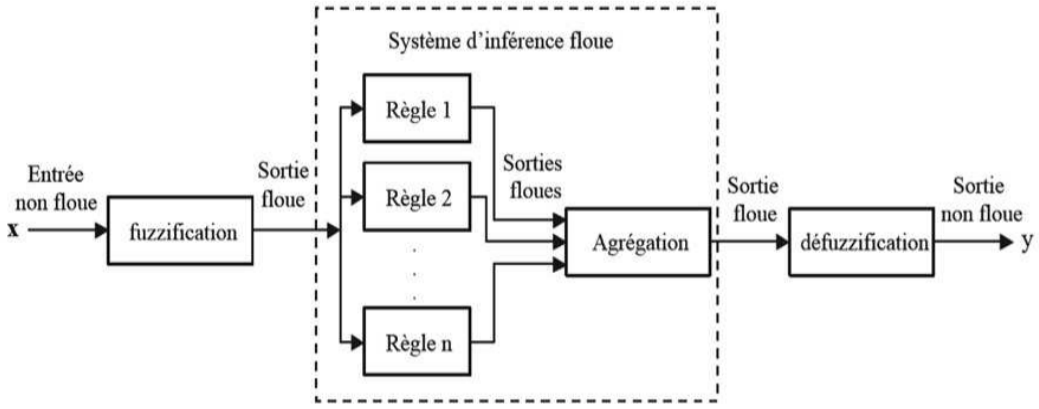


Figure II.4 : Système d'inférence floue.

II.3.2. Système Neuro-Floue :

L'utilisation conjointe des méthodes neuronales et floues dans des modèles hybrides permet de tirer des avantages, principalement, des capacités d'apprentissage des réseaux de neurones, ainsi que la lisibilité et la souplesse de la logique floue. La figure II.5 résume le principe du système neuro-flou qui représente l'intersection entre la logique floue et les réseaux de neurones.

Le principal type d'association entre réseaux de neurones et systèmes flous est le cas où un système d'inférence flou est mis sous la forme d'un réseau multicouche (Buckley et Hayashi, 1994), dans lequel les poids correspondent aux paramètres du système, l'architecture du réseau dépendant du type de règles et des méthodes d'inférence, d'agrégation et de défuzzification choisies. Nous nous limiterons, dans la suite, à la présentation du modèle ANFIS, lequel est le plus utilisé dans ce domaine.

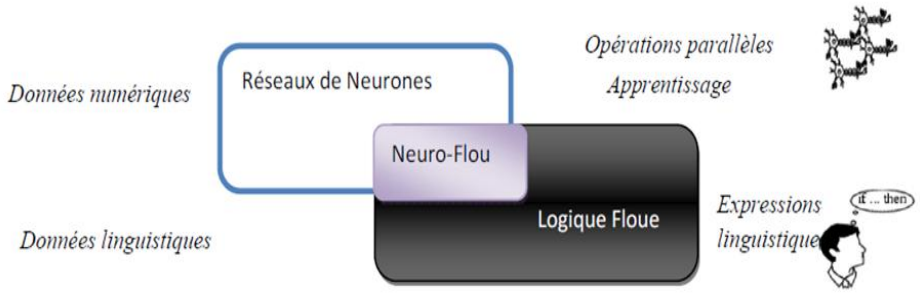


Figure II.5 : Principe du système neuro-flou.

II.3.3. Système d'Inférence Neuro-Floue Adaptatif (ANFIS) :

Le modèle ANFIS (Adaptif Neuro Fuzzy Inference System) développé par Jang (1993) est un approximateur universel (Jang et al., 1997). ANFIS est une technique qui incorpore les concepts de la logique floue dans les réseaux de neurones. Ce modèle simule la relation entre l'entrée et la sortie d'un processus à travers un apprentissage hybride pour déterminer la distribution optimale des fonctions d'appartenances. Il est basé sur les règles floues "Si … Alors" de Takagi et Sugeno (Takagi et Sugeno, 1985). L'architecture équivalente du modèle comporte cinq couches, chacune comportant plusieurs nœuds, comme le montre la figure II.6. Les nœuds carrés (adaptatifs) représentent des fonctions avec des paramètres à apprendre, alors que les nœuds circulaires représentent des opérations fixes.

Pour décrire le principe de l'ANFIS, considérons le système d'inférence floue qui contient deux entrées *x* et *y* et une seule sortie *z* donné par le schéma de la figure II.6.

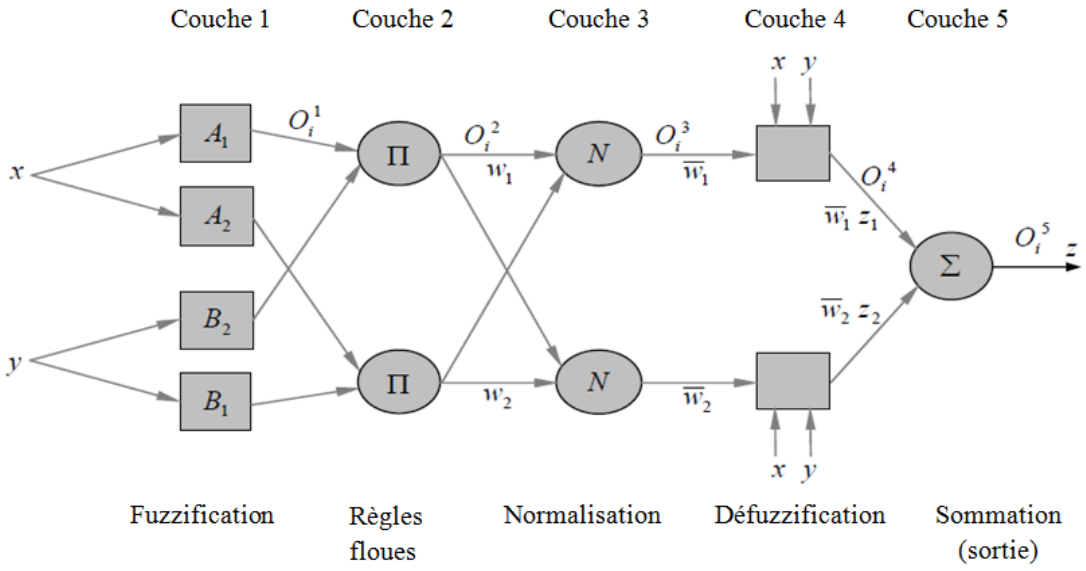


Figure II.6 : L'Architecture de l'ANFIS.

Chaque variable d'entrée est décrite par deux termes linguistiques : A_1 et A_2 pour la variable *x*, B_1 et B_2 pour la variable *y*, respectivement, d'où une base de règle "Si … Alors" décrite par deux règles floues R1 et R2 :

R1 : Si *x* est A_1 et *y* est B_1 alors $z = f_1(x,y)$ (II.5)

R2 : Si *x* est A_2 et *y* est B_2 alors $z = f_2(x,y)$ (II.6)

$f_1(x,y) = p_1x + q_1y + r_1$ (II.7)

$f_2(x,y) = p_2x + q_2y + r_2$ (II.8)

Où p_i, q_i, r_i correspondent aux paramètres de la partie conclusion à ajuster durant l'apprentissage.

- **Couche 1 (Fuzzification) :** Cette couche permet la fuzzification **d**es entrées *x* et *y*. Chaque neurone dans cette couche correspond à une variable linguistique. Les entrées *x* et *y* sont passées par la fuzzification en utilisant des fonctions d'appartenances des variables linguistiques A_i et B_i, données par :

$$O_i^1 = \mu_{A_i}(x) \quad \text{(II.9)}$$

$$O_i^1 = \mu_{B_i}(y) \quad \text{(II.10)}$$

μ_{A_i} représente la fonction d'appartenance de A_i et il spécifie le degré d'appartenance avec lequel *x* satisfait, généralement est de forme triangulaire, trapézoïdale ou Gaussienne). Par exemple, la fonction d'appartenance Gaussienne est définie par :

$$\mu_{A_i}(x) = \exp\left[-\frac{(x-c_i)}{\sigma_i}\right] \quad \text{(II.11)}$$

Avec c_i et σ_i qui sont des paramètres qui font référence aux paramètres prémisses. Les valeurs de ces derniers changent selon diverses exhibitions de la fonction d'appartenance.

- Couche 2 (Règles) : chaque nœud de cette couche reçoit les sorties des nœuds de fuzzification et calcule leur activation. Elles sont obtenues par une simple multiplication des entrées dans chaque cellule.

$$O_i^2 = w_i = \mu_{A_i}(x) \times \mu_{B_i}(y) \qquad \text{(II.12)}$$

- Couche 3 (Normalisation) : c'est la couche de normalisation dans laquelle chaque nœud calcule le degré d'appartenance normalisé à une règle floue donnée. Le résultat obtenu représente la participation de chaque règle floue au résultat final. Cette couche renvoie des sorties normalisées de défuzzification selon la relation :

$$O_i^3 = \overline{w}_i = \frac{w_i}{w_1+w_2} \quad , i = 1,2 \qquad \text{(II.13)}$$

- Couche 4 (Défuzzification) : Chaque nœud *i* de cette couche est un nœud carré adaptatif qui correspond à l'entrée initiale pondérée par le degré d'appartenance normalisé de la règle floue.

$$O_i^4 = \overline{w}_i z_i = \overline{w}_i(p_i x + q_i y + r_i) \qquad \text{(II.14)}$$

Où : $\overline{w}_i$ sont les sorties normalisées de la couche 3, et (p_i, q_i, r_i) sont les paramètres de sortie de la règle *i*. C'est la phase de défuzzification.

- Couche 5 (Sommation) : composée d'un seul nœud fixe circulaire qui reçoit la somme des sorties de tous les nœuds de défuzzification, et fournit la sortie du modèle ANFIS. La valeur approximative de la fonction désirée est exprimée par :

$$O_i^5 = \sum_i \overline{w}\, z_i = \frac{\sum_{i=1}^{2} w_i z_i}{\sum_{i=1}^{2} w_i} \qquad \text{(II.15)}$$

Ceci est, en général, la façon dont le vecteur d'entrée est acheminé à travers le réseau couche par couche. Nous considérons maintenant comment l'ANFIS apprend la prémisse et les paramètres des conséquents pour les fonctions d'appartenance et les règles.

II.3.4. Apprentissage et validation du modèle ANFIS

L'ajustement des paramètres de l'ANFIS est réalisé lors de la phase d'apprentissage. Pour cela, un ensemble de données associant séquences d'entrées et de sorties est nécessaire. Pour la réalisation de cette phase, l'algorithme d'apprentissage hybride est utilisé. Cet algorithme est une association de la méthode de descente de gradient de l'erreur et de la méthode d'estimation des moindres carrés. La méthode de descente de gradient de l'erreur

permet d'ajuster les prémisses alors que la méthode LSM (Least Square Method) ajuste les paramètres linéaires (conséquents ou conclusions).

L'apprentissage se fait de façon itérative jusqu'à ce que le nombre de cycles d'apprentissage soit atteint ou jusqu'à ce que l'erreur moyenne entre la valeur de sortie désirée et générée par l'ANFIS atteigne une valeur prédéterminée. Cette phase dépend donc de la qualité de l'ensemble des données au sens où cet ensemble doit représenter au mieux les différents comportements attendus (Wang et Mendel, 1992a et 1992b).

Le modèle ANFIS permet de s'affranchir de l'effet "boîte noire" reproché aux réseaux de neurones classiques, d'associer la connaissance dysfonctionnelle disponible sous la forme de règles floues et de conserver une capacité d'apprentissage issue des réseaux de neurones. Une des plus importantes étapes pour la génération de la structure des réseaux neuro flous ANFIS est l'établissement des règles d'inférence floues. En utilisant un mécanisme d'inférence, les règles sont définies comme combinaisons des fonctions d'appartenance des différentes variables d'entrée. Les variables d'entrée sont divisées en un nombre limité de valeurs linguistiques (étiquette), chacune caractérisée par une fonction d'appartenance (et leurs combinaisons mènent à beaucoup de règles d'inférences floues).

La validation permet de juger l'aptitude du modèle à reproduire les variables modélisées. Plusieurs critères ont été choisis. Dans notre cas, nous nous sommes basés sur le coefficient de détermination (R^2) et la racine de l'erreur quadratique moyenne (REQM).

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i^{exp} - y_i^{cal})^2}{\sum_{i=1}^{N}(y_i^{exp} - \overline{y^{exp}})^2} \qquad \text{(II.16)}$$

$$REQM = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i^{exp} - y_i^{cal})^2} \qquad \text{(II.17)}$$

Avec : y_i^{exp} et y_i^{cal} correspondent respectivement aux valeurs observées et calculées. $\overline{y^{exp}}$ est la moyenne des valeurs observées.

N représente le nombre de valeurs utilisées. Le modèle est bien optimisé si la valeur du REQM est proche de zéro.

II.4. Modélisation par machines à vecteurs supports (SVM)

Les Machines à Vecteurs Supports (Support Vector Machines en anglais : SVM) souvent traduit par l'appellation de Séparateur à Vaste Marge (SVM) sont une classe d'algorithmes d'apprentissage initialement définis pour la discrimination ; c'est à dire la

prévision d'une variable qualitative initialement binaire. Ils ont été ensuite généralisés à la prévision d'une variable quantitative.

Cette approche découle directement des travaux de Vapnik en théorie de l'apprentissage à partir de 1995. Elle s'est focalisée sur les propriétés de généralisation (ou prévision) d'un modèle en contrôlant sa complexité (Vapnik, 1995). Les SVM ont rapidement été adoptés pour leur capacité à travailler avec des données de grandes dimensions, avec un faible nombre de paramètres, leurs garanties théoriques, et leurs bons résultats en pratique. Les SVM ont été appliqués à de très nombreux domaines (bio-informatique, recherche d'information, vision par ordinateur, finances, …etc.).

Cette technique est une méthode de classification à deux classes qui tente de séparer les exemples positifs des exemples négatifs dans l'ensemble des exemples. La méthode cherche alors l'hyperplan qui sépare les exemples positifs des exemples négatifs, en garantissant que la marge entre le plus proche des positifs et des négatifs soit maximale.

II.4.1. SVM pour la régression linéaire

Le principe de base de la régression par Machines à Vecteurs est basé sur la minimisation d'une fonction de perte pénalisées (Vapnik, 1995), la pénalisation étant l'inverse du carré de la marge, et la fonction de perte pouvant être en régression la fonction "ε–insensible".

Cette fonction *f(x)* qui a au plus une déviation ε par rapport aux exemples d'apprentissage (*xi*, *yi*), pour *i* = 1, …, *N*, et qui est la plus plate possible. Cela revient à ne pas considérer les erreurs inférieures à ε et à interdire celles supérieures à ε (Smola, 2004). Maximiser la platitude de la fonction permet de minimiser la complexité du modèle qui influe sur ses performances en généralisation. En effet, la théorie de l'apprentissage permet de borner l'erreur de généralisation par une somme de deux termes : l'un dépendant de la complexité du modèle et l'autre dépendant de l'erreur sur les données d'apprentissage (Cristani, 2000, Lauer, 2006). Les méthodes SVMs sont basées sur le contrôle de la complexité du modèle lors de l'apprentissage.

Dans ce qui suit, on va présenter l'algorithme SVM pour un modèle linéaire :

La fonction linéaire *f(x)*est de la forme :

$$f(x) = \langle w, x_i \rangle + b \qquad \text{(II.18)}$$

où $w \in \mathbb{R}^{p}$ est un vecteur des paramètres (ou poids), orthogonal au plan, $x \in \mathbb{R}^{n}$ le vecteur d'entrée$\langle w, x_i \rangle$ désigne le produit scalaire entre *w* et $\boldsymbol{x_i}$ et *b* une constante à déterminer.

Pour assurer la platitude de la fonction *f(x)*, la norme des poids est minimisée (contrainte sur la dérivée de la fonction *f(x)*). Le problème revient donc à minimiser cette norme en garantissant que les erreurs sont inférieures à ε et peut s'écrire :

$$\begin{cases} \min \frac{1}{2}\|w\|^2 \\ avec\ |y_i - \langle w, x_i \rangle - b| \leq \varepsilon,\ i = 1, \dots, N \end{cases} \qquad \text{(II.19)}$$

L'objectif de cette formulation n'est pas de minimiser l'erreur d'apprentissage comme dans les réseaux de neurones ou la plupart des algorithmes de régression, mais d'assurer la platitude de la fonction. La minimisation de l'erreur n'apparaît que sous forme de contraintes (aucune erreur n'est autorisée). Cette description du problème considère donc qu'une fonction linéaire *f(x)* qui approxime tous les exemples avec une précision ε existe.

En présence de valeurs aberrantes, il est plus important d'autoriser certaines erreurs. De manière analogue à la fonction de perte "marge souple" (soft margin) (Bennett, 1992) introduite dans les machines SV par Cortes et Vapni, on peut introduire des variables de relâchement (*slack variables*) $\boldsymbol{\xi_i, \xi_i^*}$ d'optimisation pour faire face aux contraintes du problème d'optimisation ci-dessus. Nous arrivons donc à la formulation suivante :

$$\mathbf{min}\ \frac{1}{2}\|\boldsymbol{w}\|^2 + \boldsymbol{C}\sum_{i=1}^{N}(\boldsymbol{\xi_i} + \boldsymbol{\xi_i^*}) \qquad \text{(II.20)}$$

Sous les contraintes :

$$\begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, \dots, N \end{cases}$$

où $\boldsymbol{\xi_i et \xi_i^*}$ représentent respectivement les erreurs ($|\boldsymbol{y_i - f(x)}|$) positives et négatives. La constante C > 0 détermine le compromis entre la platitude de *f(x)*et la quantité d'erreurs autorisées (les écarts supérieurs à ε autorisés). Cela correspond à traiter une fonction de perte dite ε-insensible $|\boldsymbol{\xi}|_{\boldsymbol{\varepsilon}}$décrite par :

$$|\xi|_\varepsilon = \begin{cases} 0 & si\ |\xi_i| \leq \varepsilon \\ |\xi_i| - \varepsilon & autrement \end{cases} \qquad \text{(II.21)}$$

Les paramètres de la fonction de perte ε-insensible sont présentés sur la figure II.7. Seuls les points en dehors de la région ombrée contribuent au coût dans la mesure où les écarts sont pénalisés de manière linéaire.

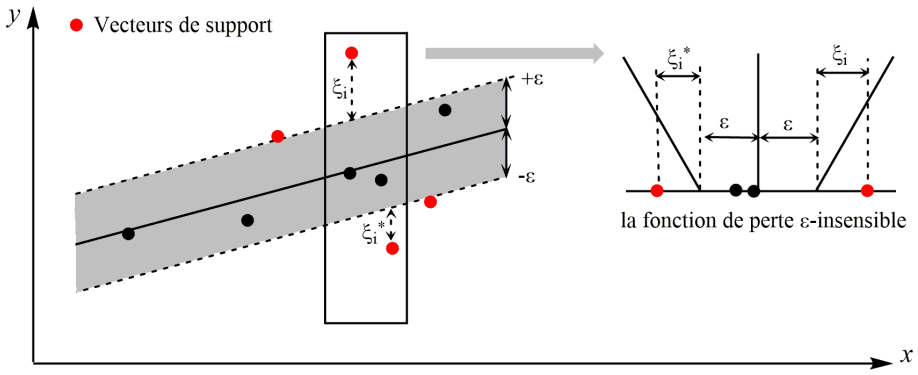


Figure II.7 : Les paramètres de la fonction de perte pour un SVM linéaire.

Il s'avère que, dans la plupart des cas, le problème d'optimisation (équation II.20) peut être résolu plus facilement dans sa formulation duale. De plus, la formulation duale fournit la clé pour étendre SV aux fonctions non linéaires. Par conséquent, nous utiliserons une méthode de dualisation standard utilisant des multiplicateurs de Lagrange. Le problème (équation II.20) se résout en minimisant le Lagrangien L donné par :

$$L = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N}(\eta_i\xi_i + \eta_i^*\xi_i^*) - \sum_{i=1}^{N}\alpha_i(\varepsilon + \xi_i - y_i + \langle w, x_i\rangle + b) - \sum_{i=1}^{N}\alpha_i^*(\varepsilon + \xi_i^* - y_i - \langle w, x_i\rangle - b) \quad \text{(II.22)}$$

où η_i, η_i^*, α_i, α_i^* représentent les multiplicateurs de Lagrange. La solution de la minimisation du lagrangien $\boldsymbol{L}$ peut s'obtenir en annulant le gradient du lagrangien par rapport aux variable primales $(\boldsymbol{w}, \boldsymbol{b}, \boldsymbol{\xi_i}, \boldsymbol{\xi_i^*})$:

$$\partial L_b = \sum_{i=1}^{N}(\alpha_i^* - \alpha_i) = 0 \quad \text{(II.23)}$$

$$\partial L_w = w - \sum_{i=1}^{N}(\alpha_i^* - \alpha_i)x_i = 0 \quad \text{(II.24)}$$

$$\partial L_{\xi_i} = C - \alpha_i - \eta_i = 0 \quad \text{(II.25)}$$

$$\partial L_{\xi_i^*} = C - \alpha_i^* - \eta_i^* = 0 \quad \text{(II.26)}$$

Les variables η_i et η_i^* peuvent être éliminés selon (équations II.25 et II.26) pour avoir le Lagrangien qui doit être maximisé :

$$L_D = -\frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j\rangle - \varepsilon\sum_{i=1}^{N}(\alpha_i - \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) \quad \text{(II.27)}$$

Le modèle final s'écrit :

$$f(x) = \sum_{i,j=1}^{N}(\alpha_i - \alpha_i^*)\langle x_i, x\rangle + b \quad \text{(II.28)}$$

Le paramètre *b* peut se calculer grâce aux conditions de Karush-Kuhn-Tucker (KKT).

On peut remarquer que seuls les exemples avec des α_i et α_i^* non nuls sont nécessaires pour implémenter le modèle (équation II.27). Ces exemples sont appelés Vecteurs de Support.

II.4.2. SVM pour la régression non-linéaire

Même lorsque les populations à discriminer sont relativement bien séparées, la séparation peut ne pas être linéaire. On cherche alors, au moyen d'une transformation non linéaire Φ, à passer de l'espace d'origine dans un espace de dimension plus grande doté d'un produit scalaire (espace de Hilbert) et où existe une séparation linéaire (figure II.8).Le problème devient linéaire et le modèle est donné par :

$$f(x) = \sum_{i,j=1}^{N} (\alpha_i - \alpha_i^*) k\langle x_i, x\rangle + b \qquad \text{(II.29)}$$

où, $\Phi(x_i)$ est la projection de x_i dans l'espace étendu $\mathcal{F}$.

La quantité $k\langle x_i, x\rangle = \langle \Phi(x_i), \Phi(x)\rangle$est appelée fonction noyau ou noyau (*kernel function* ou *kernel*), qui doit vérifier les conditions de Mercer (Mercer, 1909) qui se résument à vérifier que la fonction *k* correspond bien à un produit scalaire définissant une norme dans un espace étendu $\mathcal{F}$. Dans l'algorithme, c'est le noyau *k* et non Φ qui est choisi, et en s'y prenant bien, on peut calculer $k\langle x_i, x\rangle$ sans faire apparaitre Φ. Les calculs sont alors fait dans l'espace de départ, et deviennent beaucoup plus simples et plus rapides ; pour cette raison, on parle de machine à noyau ou « *kernel machine* ».

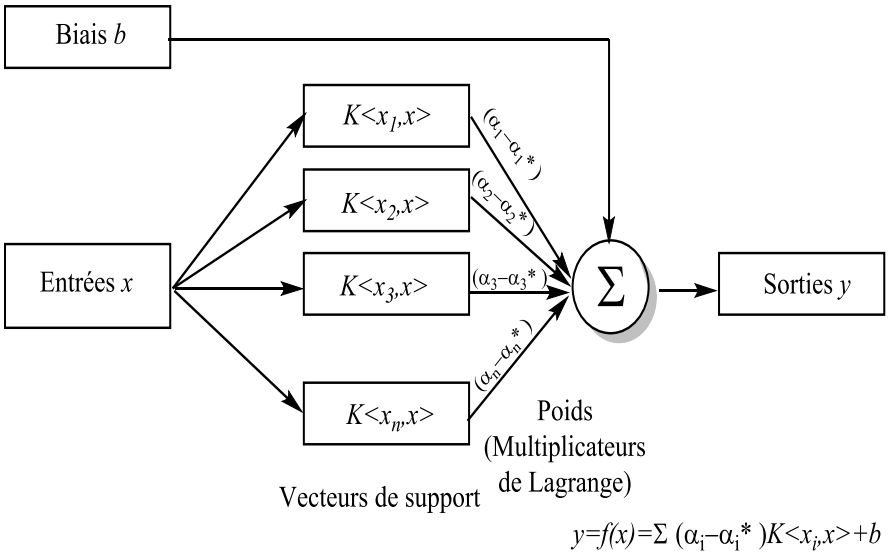


Figure II.8 : Architecture d'une machine à vecteurs support.

Les noyaux les plus couramment utilisés pour les SVMs sont :

- Linéaire: $k(x, x') = \langle x, x' \rangle$; le plus simple, utilisé de grands volumes de données, notamment en « *textmining* » ;
- Polynomial: $k(x, x') = (\gamma \langle x, x' \rangle + c)^d$; où d est le degré du polynôme, utilisé en traitement de l'image ;
- Sigmoïdal: $k(x, x') = tanh(\gamma \langle x, x' \rangle + c)$; souvent utilisée pour le perceptron multicouches ;
- Gaussien: $k(x, x') = exp\left(\frac{-\|x-x'\|^2}{2\sigma^2}\right)$; l'une des plus couramment employés avec le noyau radial gaussien ;
- Radial gaussien (RBF): $k(x, x') = exp(-\gamma \|x - x'\|^2)$.

II.5. Conclusion

Nous avons présenté dans ce chapitre une description générale des réseaux de neurones, de la théorie floue, et Neuro-floue. Nous nous sommes intéressés principalement aux systèmes ANFIS. L'utilisation conjointe des méthodes neuronales et floues permet de tirer l'avantage de leurs qualités, entre autres et essentiellement, les capacités d'apprentissage des réseaux de neurones ainsi que la lisibilité et la souplesse des systèmes flous. L'architecture du système neuro-flou adopté dans ce chapitre montre des bonnes performances dans les simulations des systèmes non linéaires.

Nous avons introduit aussi les machines à vecteur support (SVM), leur fondement théorique et leurs différentes variantes. Il est tiré que les SVMs sont en avance, en les comparant aux autres méthodes d'apprentissage, dans plusieurs points : Les SVMs permettent de traiter plusieurs problèmes de datamining : classification, régression, clustering, ...etc. Les SVMs permettent de traiter les données numériques et symboliques, ce qui les a favorisées dans plusieurs applications complexes. Les capacités de généralisation et la simplicité d'entrainement des SVMs sont bien au-delà des autres méthodes. Les SVMs sont très efficaces sur les données à nombre élevé d'attributs, même avec peu d'exemples. Elles n'imposent aucune limite sur le nombre d'attributs sauf les limites imposées par le hardware.

CHAPITRE III

MODELISATION DES PARAMETRES CLIMATIQUES ET DES PARAMETRES PHYSICO-CHIMIQUES DE L'EAU :
RESULTATS ET DISCUSSION

CHAPITRE III

Modélisation des paramètres climatiques et des paramètres physico-chimiques de l'eau : Résultats et Discussion

III.1. Présentation du site étudié

Barrage Ghrib (N36°9'55'', E2°33'23''), a été mis en service en 1939, il est implanté sur Oued Cheliff, à 155 km d'Alger, dans la commune d'Oued Chorfa (wilaya d'Ain Defla). C'est une importante réserve d'eau destinée à l'irrigation du périmètre du haut Cheliff, aux transferts vers le barrage Bouroumi et à l'alimentation en eau potable de la ville de Médéa, de Berrouaghia et de la commune d'Oued Chorfa.

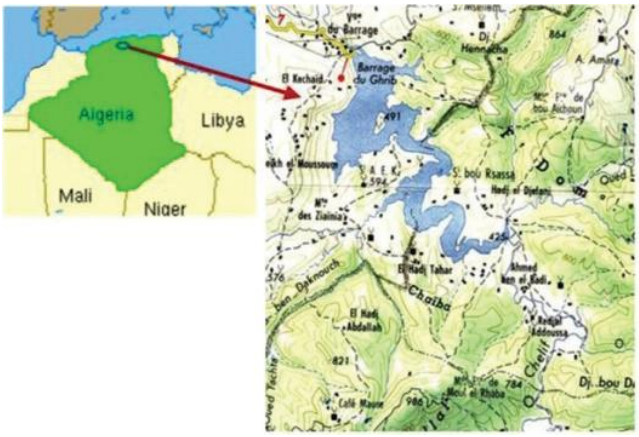


Figure III.1 : Présentation géographique du barrage Ghrib.

III.2. Collecte d'une base de données

L'objectif de cette étape est de rassembler un nombre de données suffisant pour construire une base représentative de données, qui servira à l'apprentissage et au test du réseau de neurones.

Cette base de données constitue l'entrée du réseau de neurones, et par conséquent c'est elle qui détermine à la fois la taille du réseau (et donc le temps de simulation), et les performances du système (capacité et généralisation).

Pour notre application, nous avons utilisé les données météorologiques mensuelles fournis par l'ONM (Office National de la Météo), à savoir, (la température minimale de l'air

($T_{air,\ min}$ /°C), la température maximale de l'air ($T_{air,\ max}$ /°C), l'humidité relative (HR /%) et la durée d'ensoleillement (DE /h).

Parallèlement, les données physico-chimiques mensuelles de l'eau du barrage Ghrib sont fournis par l'ANRH (Agence Nationale des Ressources Hydrauliques) sont : l'année (A), le mois (M), la température de l'eau (T_{eau}), la saturation en oxygène (O_{2sat}), la demande chimique en oxygène (DCO), le pH, la conductivité électrique (CE), l'ion d'ammonium (NH_4^+), l'ion de nitrate (NO_3^-), la turbidité (Turb), la matière organique (MO) et les résidus secs (RS). Toutes ces données ont été mesurées sur une période de 13 ans (2003 à 2015). A cet effet nous avons pu construire, après prétraitement, une base de données composées de 156 exemples.

Les valeurs minimales et maximales des paramètres météorologiques et des paramètres physico-chimiques de l'eau sont présentées dans le tableau III.1.

Tableau III.1 : Résumé statistique des variables retenues.

Catégorie	Variables	Symbole	Minimum	Maximum
Paramètres Climatiques	Température minimale de l'air (°C)	$T_{air,min}$	-3.6	22.3
	Température maximale de l'air (°C)	$T_{air,max}$	5.7	39.7
	Humidité relative (%)	HR	38	89
	Durée d'ensoleillement (h)	DE	91	379
Paramètres physico-chimiques de l'eau	Température de l'eau (°C)	T_{eau}	8.2	35.5
	Saturation en oxygène (%)	O_2 sat	31.5	135.5
	Demande chimique en oxygène (mg/l)	DCO	5	148
	Potentiel d'hydrogène	pH	6.8	8.4
	Conductivité électrique (µS/cm)	CE	1125	4083
	Ion d'ammonium (mg/l)	NH_4^+	0	0.84
	Ion de nitrate (mg/l)	NO_3^-	0	31
	Turbidité (NTU)	Turb	0.53	11
	Matière organique (mg/l)	MO	2	18
	Résidus secs (mg/l)	RS	411	2996

La matrice a été établie afin d'étudier la corrélation entre les 12 paramètres d'entrée sélectionnées. La valeur du coefficient de corrélation de chaque paire des paramètres d'entrée a été examinée pour les quatre paramètres climatiques (Tableau III.2 à III.5). Sur la base des résultats obtenus, on remarque que la valeur du coefficient de corrélation la plus élevée de chaque paire des paramètres d'entrée choisis est de 0,682 (donc < 0.75), ce qui signifie que les

paramètres physicochimiques de l'eau retenus comme des variables d'entrée sont indépendants.

Tableau III.2 : Matrice des coefficients de corrélation des paramètres d'entrée pour $T_{air,min}$.

	A	M	T_{eau}	O_2sat	DCO	pH	CE	NH_4^+	NO_3^-	Turb	MO	RS
A	1	−0.005	−0.098	−0.260	−0.268	0.270	0.080	−0.263	−0.078	0.208	0.017	−0.119
M		1	0.362	0.002	0.046	0.128	0.075	−0.127	−0.115	0.005	0.115	0.126
T_{eau}			1	0.315	0.106	0.223	0.017	−0.203	−0.056	−0.146	0.065	0.090
O_2sat				1	0.045	−0.005	0.039	0.005	−0.033	−0.023	0.219	0.117
DCO					1	−0.151	−0.278	0.014	−0.042	−0.112	−0.160	−0.136
pH						1	−0.008	−0.192	−0.025	−0.018	0.066	−0.047
CE							1	0.115	−0.311	−0.095	0.215	0.677
NH_4^+								1	−0.122	−0.004	−0.001	0.148
NO_3^-									1	0.160	−0.056	−0.310
Turb										1	0.034	−0.118
MO											1	0.231
RS												1

Tableau III.3 : Matrice des coefficients de corrélation des paramètres d'entrée pour $T_{air,max}$.

	A	M	T_{eau}	O_2sat	DCO	pH	CE	NH_4^+	NO_3^-	Turb	MO	RS
A	1	0.011	−0.080	−0.258	−0.269	0.268	0.084	−0.274	−0.082	0.210	0.027	−0.108
M		1	0.359	−0.002	0.046	0.132	0.088	−0.132	−0.126	0.007	0.120	0.134
T_{eau}			1	0.312	0.106	0.229	0.029	−0.205	−0.065	−0.145	0.066	0.095
O_2sat				1	0.045	−0.004	0.039	0.008	−0.032	−0.023	0.216	0.114
DCO					1	−0.150	−0.274	0.013	−0.042	−0.111	−0.159	−0.134
pH						1	−0.008	−0.194	−0.025	−0.018	0.068	−0.044
CE							1	0.099	−0.320	−0.091	0.227	0.682
NH_4^+								1	−0.109	−0.006	−0.009	0.134
NO_3^-									1	0.157	−0.067	−0.321
Turb										1	0.036	−0.114
MO											1	0.240
RS												1

Tableau III.4 : Matrice des coefficients de corrélation des paramètres d'entrée pour HR.

	A	M	T_{eau}	O_2sat	DCO	pH	CE	NH_4^+	NO_3^-	Turb	MO	RS
A	1	0.011	−0.085	−0.256	−0.264	0.263	0.089	−0.266	−0.086	0.212	0.018	−0.109
M		1	0.356	−0.007	0.039	0.144	0.084	−0.139	−0.123	0.004	0.130	0.135
T_{eau}			1	0.309	0.101	0.238	0.023	−0.211	−0.061	−0.147	0.076	0.096
O_2sat				1	0.042	0.001	0.037	0.005	−0.031	−0.025	0.220	0.115
DCO					1	−0.146	−0.275	0.012	−0.042	−0.113	−0.156	−0.134
pH						1	−0.005	−0.191	−0.027	−0.016	0.061	−0.044
CE							1	0.100	−0.321	−0.091	0.226	0.681
NH_4^+								1	−0.109	−0.007	−0.009	0.133
NO_3^-									1	0.157	−0.066	−0.320
Turb										1	0.037	−0.114
MO											1	0.241
RS												1

Tableau III.5 : Matrice des coefficients de corrélation des paramètres d'entrée pour DE.

	A	**M**	**T_{eau}**	**O_2sat**	**DCO**	**pH**	**CE**	**NH_4^+**	**NO_3^-**	**Turb**	**MO**	**RS**
A	1	−0.005	−0.101	−0.262	−0.263	0.260	0.083	−0.265	−0.082	0.208	0.016	−0.110
M		1	0.362	0.001	0.050	0.120	0.081	−0.132	−0.121	0.006	0.118	0.135
T_{eau}			1	0.314	0.112	0.214	0.020	−0.205	−0.059	−0.146	0.064	0.096
O_2sat				1	0.045	−0.006	0.037	0.007	−0.031	−0.024	0.217	0.115
DCO					1	−0.142	−0.271	0.010	−0.045	−0.110	−0.155	−0.134
pH						1	−0.018	−0.188	−0.019	−0.022	0.054	−0.046
CE							1	0.103	−0.319	−0.093	0.221	0.682
NH_4^+								1	−0.111	−0.005	−0.007	0.133
NO_3^-									1	0.158	−0.063	−0.320
Turb										1	0.034	−0.114
MO											1	0.241
RS												1

Les facteurs de variation d'inflation (VIF) ont été calculés afin d'évaluer la multi-colinéarité entre les douze paramètres d'entrées choisis. Selon Jaiswal et al., 2004, les valeurs des VIF peuvent être classer comme suit : si VIF est égal à 1, implique l'inexistence d'inter-corrélation pour chaque variable ; si VIF est situé entre 1 et 5, la sélection des paramètres est acceptable et si VIF est supérieur à 10, le modèle associé aux paramètres sélectionnés est instable et une nouvelle sélection s'avère nécessaire. A partir du tableau III.6, on constate que toutes les variables d'entrées ont des valeurs VIF inférieures à 3, ce qui indique que la sélection effectuée est bonne et que ces variables ne sont pas corrélées entre elles. La valeur du facteur d'inflation de la variance est calculée par le logiciel XLSTAT.

Tableau III.6 : Valeurs du coefficient VIF des paramètres d'entrée.

	$T_{air,min}$	**$T_{air,max}$**	**HR**	**DE**
A	1.506	1.506	1.503	1.501
M	1.230	1.229	1.232	1.233
T_{eau}	1.450	1.454	1.449	1.451
O_2sat	1.281	1.282	1.277	1.280
DCO	1.246	1.245	1.238	1.237
pH	1.194	1.193	1.195	1.179
CE	2.992	2.940	2.993	2.994
NH_4^+	1.205	1.205	1.204	1.202
NO_3^-	1.227	1.219	1.228	1.227
Turb	1.132	1.134	1.133	1.132
MO	1.153	1.148	1.156	1.150
RS	2.930	2.881	2.932	2.942

III.3. Développement du modèle

Le but de cette étude est de construire des modèles statistiquement robustes en interpolation et en extrapolation. L'ensemble de données a été divisé en deux ensembles, d'apprentissage et de test. L'ensemble d'apprentissage a été dédié au développement des modèles, tandis que l'ensemble de test a été laissé pour tester l'optimalité et la capacité de généralisation des modèles développés.

Pour le développement des modèles, trois méthodes statistiques ont été utilisées : le réseau de neurones de type perceptron multicouche (MLP), le système d'inférence flou à base de réseaux de neurones adaptatifs (ANFIS) et les machines à vecteurs supports SVM.

Le présent chapitre a trait à de la prédiction selon deux axes. Le premier axe, consiste en l'étude de la prédiction des paramètres climatiques à partir des paramètres physico-chimiques de l'eau. Le deuxième axe a été consacré à la possibilité de la prédiction en sens inverse, à savoir la prédiction des paramètres physico-chimiques de l'eau à partir des paramètres climatiques.

III.4. Prédiction des paramètres climatiques à partir des paramètres physico-chimiques de l'eau

III.4.1. Modélisation par le modèle ANN-MLP

La structure ANN-MLP adoptée pour cette application, comprend une couche d'entrée, une couche intermédiaire ou cachée et une couche de sortie. Chaque couche peut avoir un certain nombre de neurones, qui sont connectés linéairement par des poids aux neurones des couches voisines.

La démarche adoptée consiste à faire varier le nombre des neurones dans la couche cachée et à retenir l'architecture qui présente les meilleures performances qui permet de donner le minimum de l'erreur sur la base de test, afin d'éviter le problème de sur-apprentissage.

Le logiciel STATISTICA (STATISTICA 10.0, Tulsa, StatSoft Inc., OK, USA) a été utilisé pour le calcul de ANN-MLP.

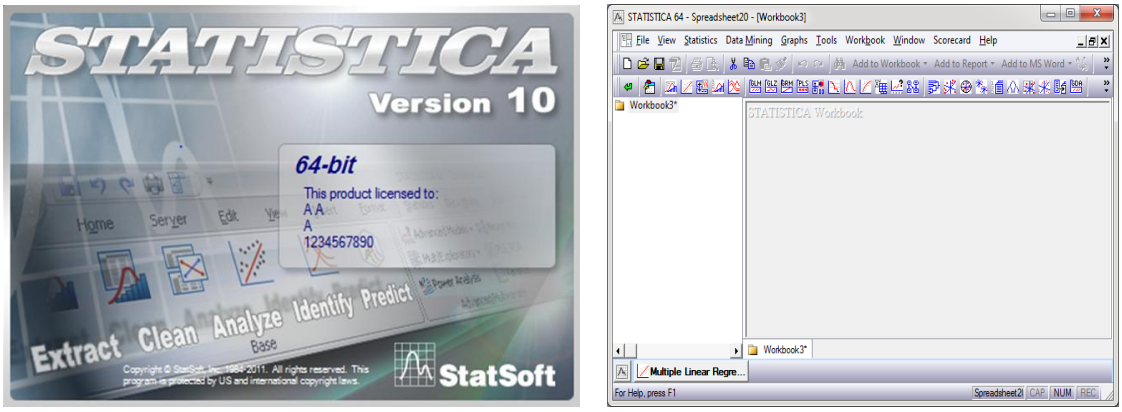


Figure III.2 : Capture d'écran des logiciels STATISTICA 10.

Cette partie a été consacrée à la prédiction des paramètres climatiques en utilisant toutes les variables physico-chimiques de l'eau disponibles de plus, l'année et le mois pour représenter le vecteur d'entrée du modèle. Le vecteur de sortie représente les valeurs prévues des variables climatiques groupés en deux séries ($T_{air,min}$ et $T_{air,max}$) et (HR et DE) (Figure III.3).

Le jeu de données a été divisé en deux ensembles : un ensemble d'apprentissage (80% de la base de données totale) a été utilisé pour générer le modèle et un ensemble de test (20% de la base de données totale) a été utilisé pour évaluer le modèle obtenu.

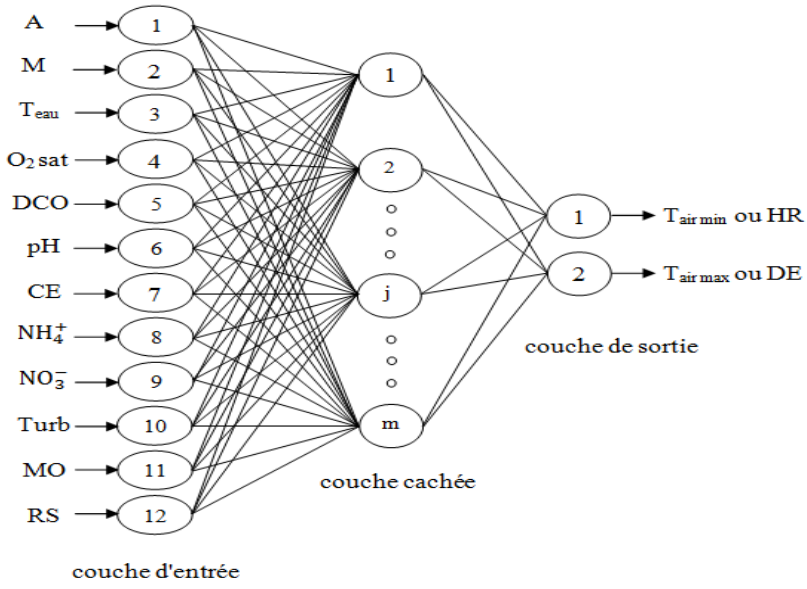


Figure III.3 : Architecture du modèle ANN-MLP.

Une analyse de sensibilité utilisant la méthode proposée par (Garson, 1991), a été également réalisée pour déterminer la contribution de chaque variable dans la prédiction des quatre paramètres climatiques individuellement. La contribution de chacun des paramètres

d'entrée dans les modèles MLP obtenus est reproduite sur la figure III.4. Il est à noter que les trois variables d'entrée à savoir : l'Année, le Mois et la Température de l'eau représentent presque 50% de la contribution à la prédiction des quatre paramètres climatiques. La contribution relative des autres variables d'entrée est différente dans les quatre modèles.

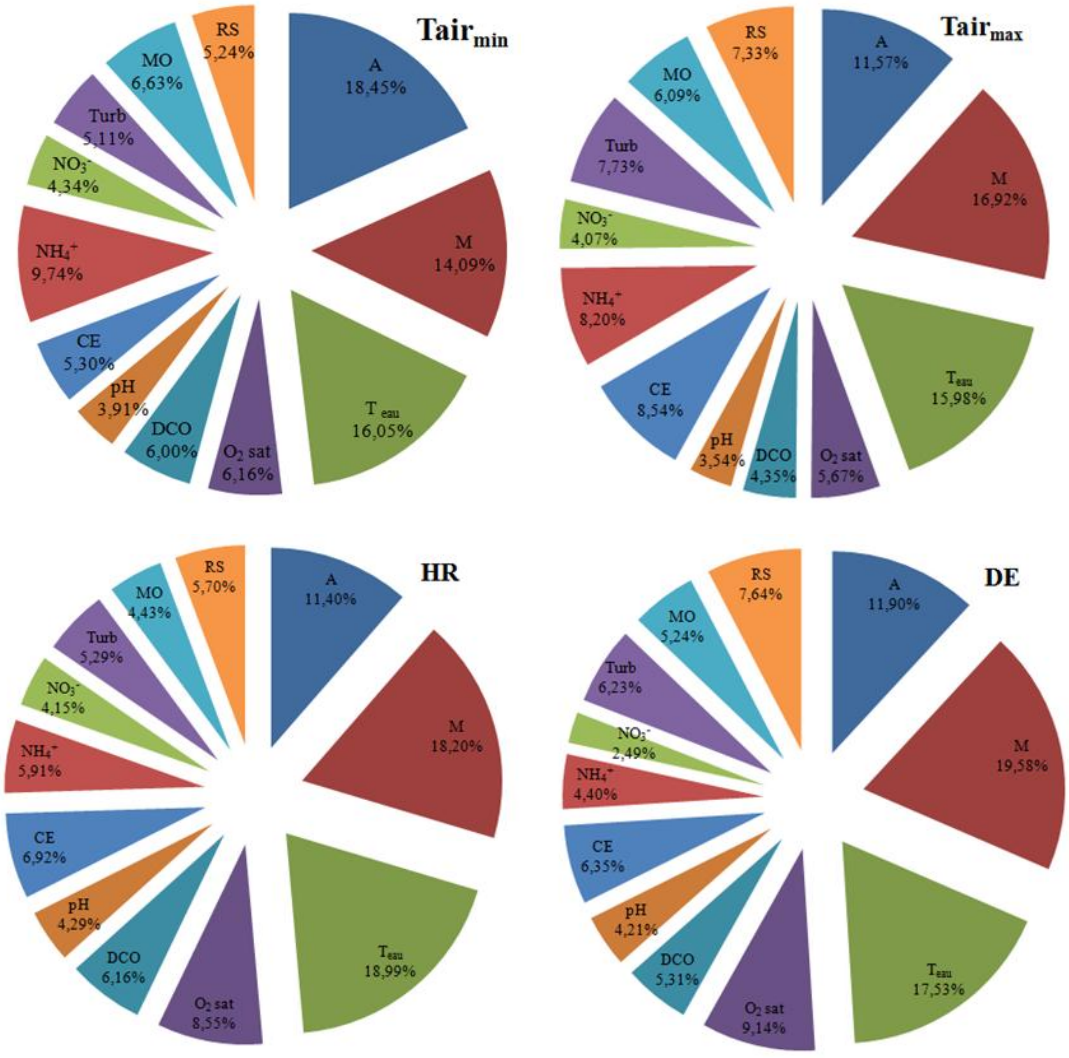


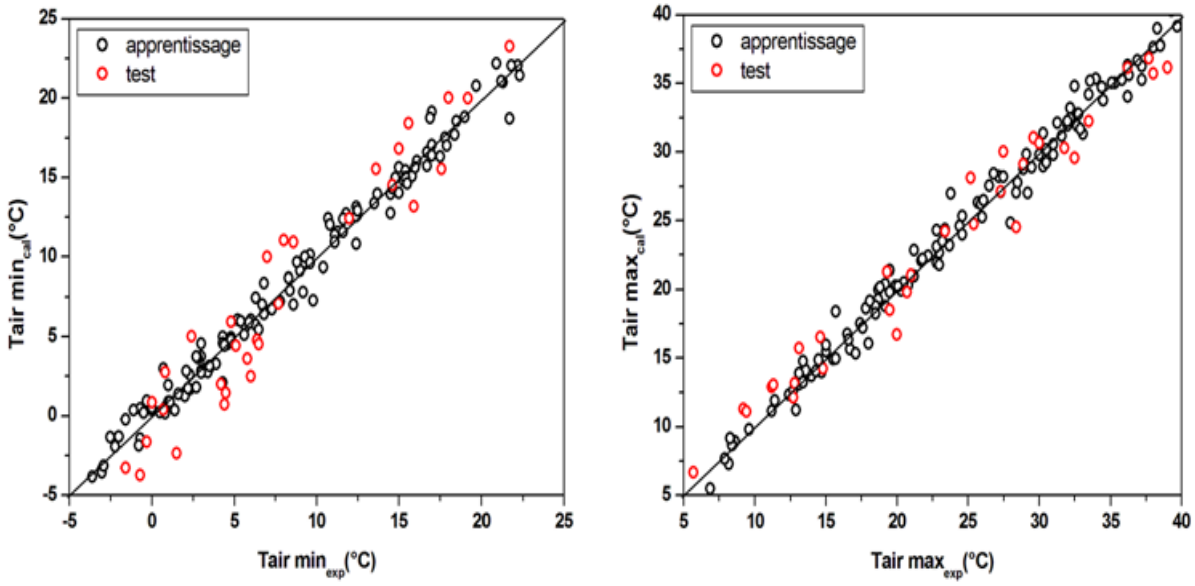
Figure III.4 : Importance relative des paramètres d'entrée dans les modèles ANN-MLP.

Les critères REQM, ERAM et R^2 ont été utilisés pour évaluer la précision des ensembles d'apprentissage et de test. Le meilleur modèle a été choisi sur la base des valeurs minimales de la racine de l'erreur quadratique moyenne REQM et de l'erreur relative absolue moyenne ERAM et de la valeur maximale du coefficient de détermination R^2 pour les quatre paramètres climatiques. Les résultats de l'optimisation des réseaux de neurones sont présentés dans le tableau III.7.

Tableau III.7 : Architecture des réseaux de neurones optimisés.

		$T_{air,min}$	$T_{air,max}$	HR	DE
Algorithme d'apprentissage		BFGS	BFGS	BFGS	BFGS
Neurones cachés		4	4	7	7
Fonction d'activation	**couche cachée**	Tangente hyperbolique	Tangente hyperbolique	Sigmoïde	Sigmoïde
	couche de sortie	Sigmoïde	Sigmoïde	Linéaire	Linéaire
Apprentissage	**N**	125	125	125	125
	R^2	0.9941	0.9966	0.9633	0.9409
	REQM	0.8786	1.0009	3.6980	19.2645
	ERAM	///	0.0355	0.0438	0.0677
Test	**N**	31	31	31	31
	R^2	0.9731	0.9691	0.9076	0.8678
	REQM	1.2872	2.3520	6.7654	36.4913
	ERAM	///	0.0622	0.0974	0.1471
Global	**R^2**	0.9909	0.9811	0.9264	0.9248
	REQM	1.3874	1.5613	5.1211	23.7063
	ERAM	///	0.0592	0.0544	0.0835

Les valeurs prédites des quatre paramètres climatiques $T_{air,min}$, $T_{air,max}$, HR et DE dans les ensembles d'apprentissage et de test à l'aide du modèle ANN-MLP en fonction des valeurs expérimentales sont présentées sur la figure III.5.



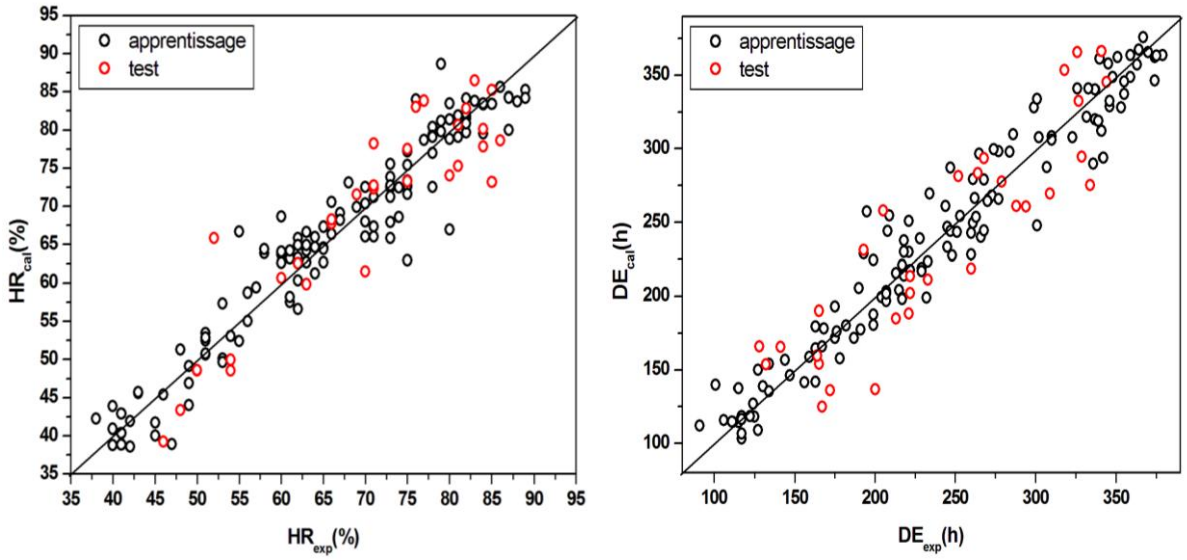


Figure III.5 : Prédiction des paramètres climatiques : $T_{air,min}$, $T_{air,max}$, HR et DE par le modèle ANN-MLP avec 12 entrées.

Les valeurs prédites de $T_{air,min}$, $T_{air,max}$, HR et DE par la régression neuronale ANN-MLP corroborent avec les valeurs expérimentales. Par ailleurs, les modèles présentent des valeurs des coefficients de corrélation R^2 de 0,9909, 0.9911, 0.9764 et 0.9798 pour $T_{air,min}$, $T_{air,max}$, HR et DE respectivement (tableau III.7). D'autre part, les erreurs relatives pour les quatre paramètres climatiques étudiés ont été examinées à l'aide des résidus normalisés par rapport aux valeurs prédites et observées. Les caractéristiques statistiques obtenues dans cette étude, indique que le modèle ANN-MLP à un bon pouvoir prédictif des paramètres climatiques ($T_{air,min}$, $T_{air,max}$, HR et DE).

Les modèles ANN-MLP ont été testés pour prédire les paramètres climatiques ($T_{air,min}$, $T_{air,max}$, HR et DE) pendant l' année 2016. Cette prédiction a été faite en utilisant le logiciel Statistica. Les quatre paramètres climatiques ont été calculés et emportés pour les comparer avec les valeurs expérimentales. Les résultats sont présentés sur la figure III.6. Il a été constaté que les valeurs mensuelles prévues de ces paramètres climatiques sont très proches des valeurs mesurées.

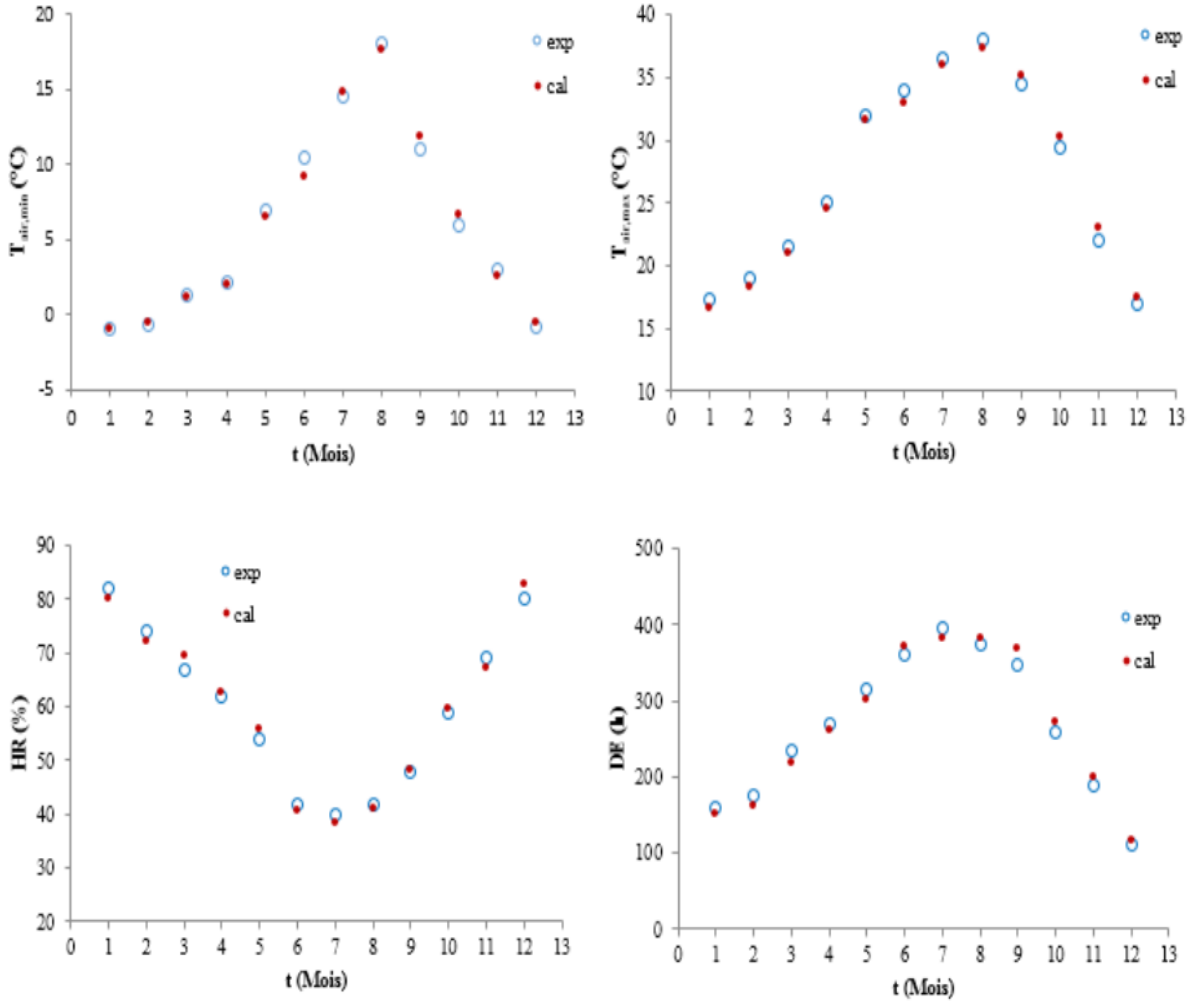


Figure III.6 : Paramètres climatiques observés et prédites ($T_{air,min}$, $T_{air,max}$, HR et DE) pour l'année 2016.

Les résultats de la modélisation des paramètres climatiques par trois méthodes d'analyse : les réseaux de neurones de type perceptron multicouche (MLP), le système d'inférence flou à base de réseaux de neurones adaptatifs (ANFIS) et les machines à vecteurs supports SVM (en anglais Support Vector Machine) ont été présentés dans cette étude et ont été utilisés pour comparer les performances et la qualité de la prédiction des trois modèles obtenus *(ANN-MLP, ANFIS et SVM).*

L'approche ANN-MLP a été réalisée à l'aide du logiciel STATISTICA (STATISTICA 10.0, StatSoft, Inc., Tulsa, OK, USA). Les approches ANFIS et SVM ont, quant à elles, été réalisées à l'aide du logiciel MATLAB R2018b.

Une nouvelle base de données a été élaborée pour être compatible avec les trois modèles étudiés (ANN-MLP, ANFIS et SVM). Lors de cette étape, sept paramètres ont été sélectionné

pour construire le vecteur d'entrée à savoir : l'année (A), le mois (M), la température de l'eau (T_{eau}), la saturation en oxygène (O_{2sat}), l'ion d'ammonium (NH_4^+), la conductivité électrique (CE) et les résidus secs (RS).

La base de données contenant 156 paramètres d'entrées a été répartie comme suit ; un ensemble de 80% destiné à l'apprentissage et un ensemble de 20% de la base globale pour le test. L'ensemble d'apprentissage avec 125 valeurs est dédié au développement des modèles. Par contre, l'ensemble de test, comprenant 31 valeurs qui n'ont pas été utilisés pour le développement des modèles, a été laissé pour la validation.

III.4.1.1. Résultats de corrélation par le modèle ANN-MLP

Il n'y a pas de principes théoriques rigoureux pour choisir la topologie du réseau appropriée. Ainsi, l'utilisation d'une régression neuronale nécessite l'optimisation de l'architecture du réseau neuronal. Dans cette étude, le réseau a été formé à l'aide de l'algorithme des méthodes quasi-Newton BFGS. Pour optimiser le nombre de neurones dans la couche cachée, plusieurs calculs ont été effectués avec différents nombres de neurones cachés (de 1 à 14). Un neurone de sortie a été utilisé pour représenter indépendamment les quatre paramètres climatiques ($T_{air,min}$, $T_{air,max}$, HR et DE). La couche d'entrée est formée de 7 variables sélectionnés auparavant. L'ensemble de 156 données a été divisé en deux lots : un ensemble d'apprentissage (80%) et un ensemble de test (20%). Le tableau III.8 récapitule les différents éléments structurels des réseaux de neurones optimaux.

Tableau III.8 : Structure des réseaux de neurones optimisés.

		$T_{air,min}$	$T_{air,max}$	HR	DE
Algorithme d'apprentissage		BFGS	BFGS	BFGS	BFGS
Neurones cachés		5	4	4	4
Fonction d'activation	**couche cachée**	Tangente hyperbolique	Tangente hyperbolique	Exponentiel	Tangente hyperbolique
	couche de sortie	Purelin	Purelin	Purelin	Purelin

Les résultats de la modélisation par le modèle ANN-MLP des quatre paramètres climatiques sont présentés sous forme de nuage de points qui montrent la qualité de concordance entre les valeurs calculées par ANN-MLP et les valeurs expérimentales. Cette qualité est visualisée à travers l'alignement des points sur la première bissectrice pour l'ensemble des données et le coefficient de corrélation R (figure III.7) ci-dessous pour $T_{air,min}$, $T_{air,max}$, HR et DE pour les ensembles d'apprentissage et du test respectivement.

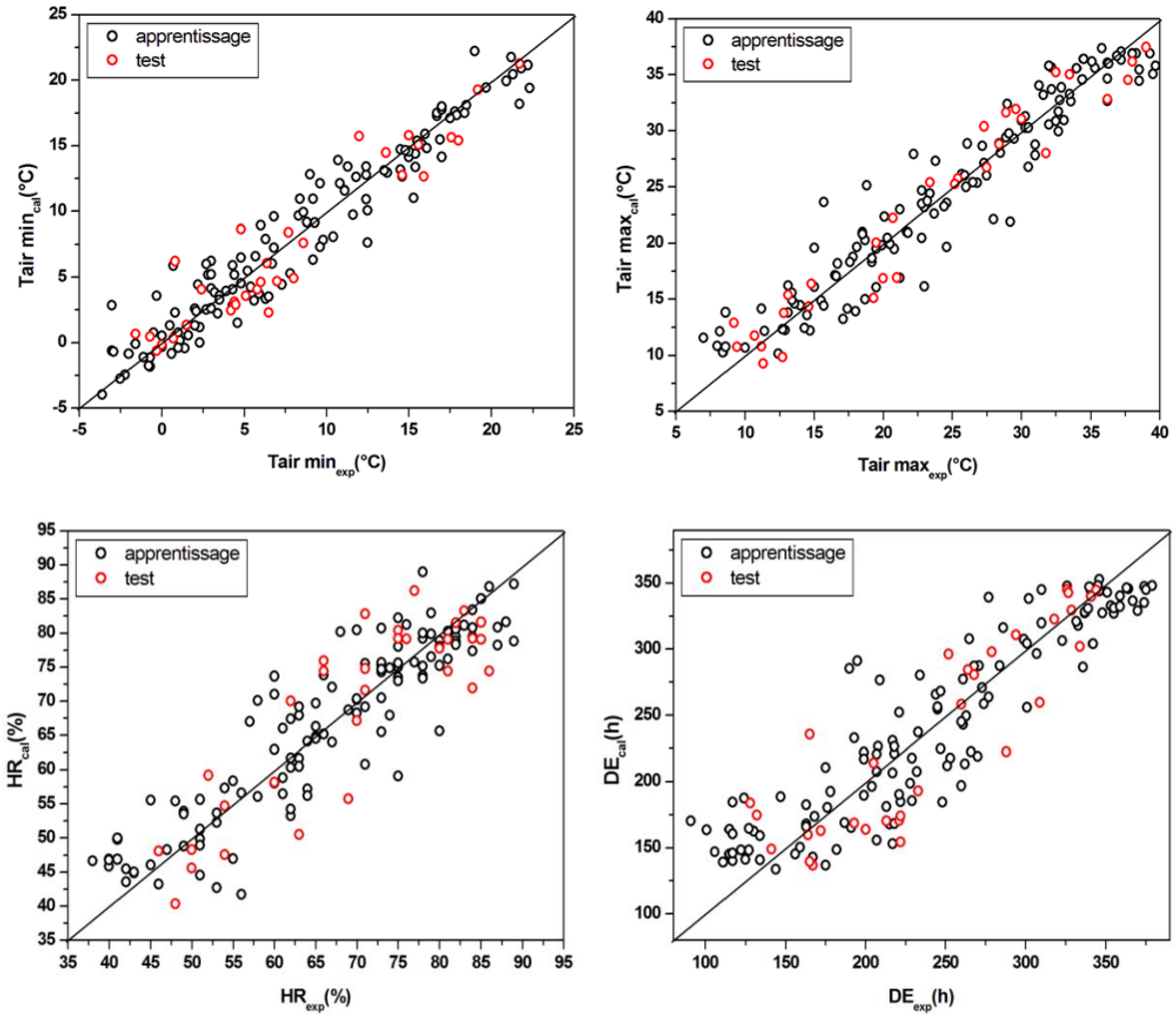


Figure III.7 : Prédiction des paramètres climatiques : $T_{air,min}$, $T_{air,max}$, HR et DE par le modèle ANN-MLP avec 7 entrées.

D'après la figure III.7, une bonne corrélation entre les valeurs prédites par le modèle ANN-MLP et les valeurs expérimentales a été constatée pour les quatre paramètres climatiques notamment pour $T_{air,min}$ et $T_{air,max}$.

Les performances des technique d'intelligence computationnelles ont été évaluées en termes d'erreur relative absolue moyenne (ERAM), du coefficient de détermination R^2 ainsi en terme du pourcentage de l'erreur relative (*ER%*), en utilisant l'équation suivante :

$$ER(\%) = 100 * \frac{y_i^{exp} - y_i^{cal}}{y_i^{exp}} \qquad \text{(III.1)}$$

La figure III.8 montre la variation des erreurs relatives en % pour 156 points expérimentaux des quatre paramètres climatiques.

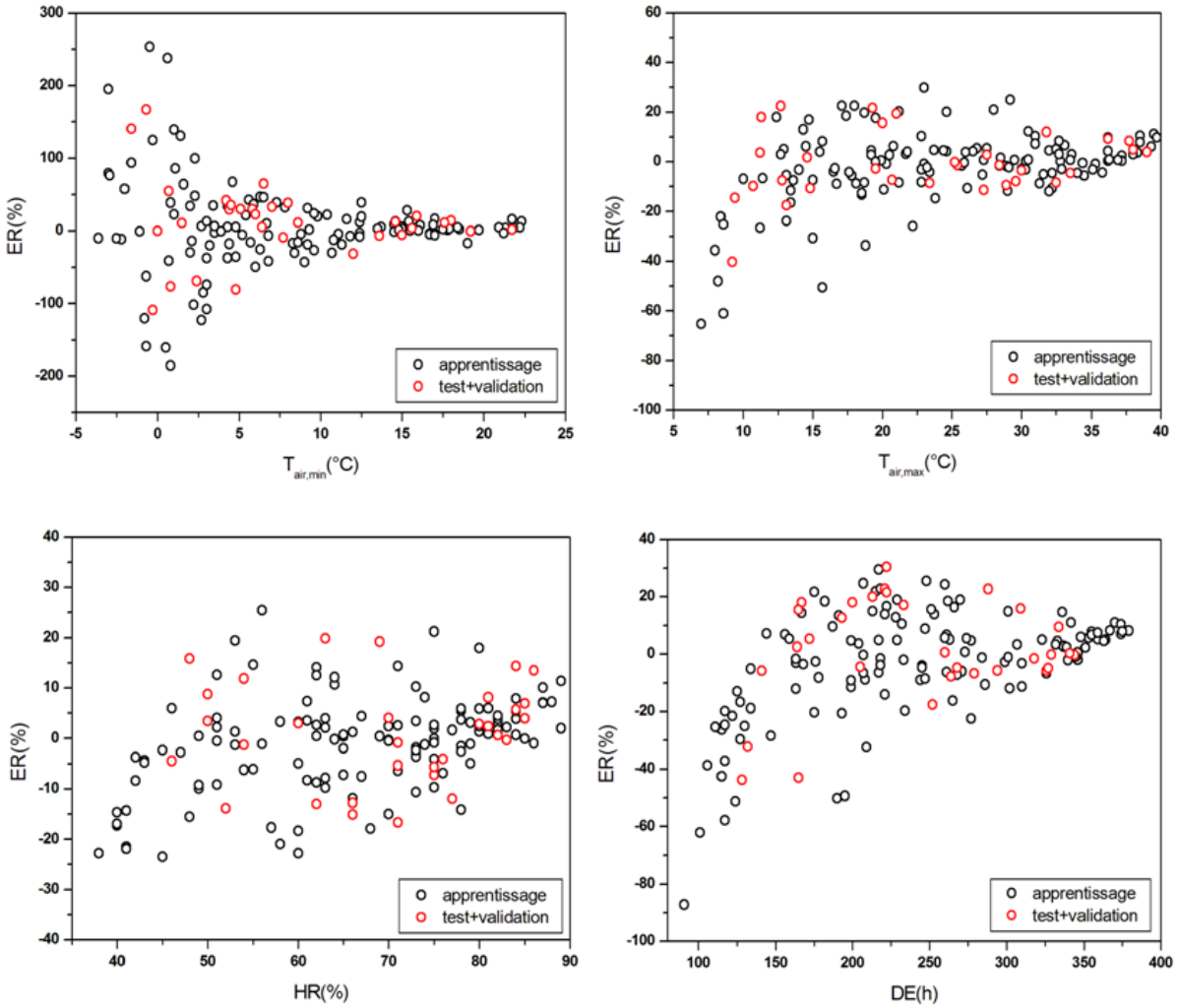


Figure III.8 : L'ER du modèle ANN-MLP des paramètres : $T_{air,min}$, $T_{air,max}$, HR et DE.

Les valeurs des paramètres statistiques pour les deux ensembles (apprentissage et test) sont regroupées sur le tableau III.9

Tableau III.9 : Valeurs des paramètres statistiques pour les ensembles d'apprentissage et de test.

	$T_{air,min}$		$T_{air,max}$		HR		DE	
Paramètres statistiques	**Apprentissage**	**Test**	**Apprentissage**	**Test**	**Apprentissage**	**Test**	**Apprentissage**	**Test**
N	125	31	125	31	125	31	125	31
R	0.9637	0.9461	0.9571	0.9686	0.9109	0.8638	0.9085	0.8792
R^2	0.9286	0.8918	0.9160	0.9380	0.8297	0.6990	0.8250	0.72
ERAM	///	///	0.0988	0.9001	1.0702	3.0828	3.1365	10.1346

D'après le tableau III.9, le modèle non linéaire des réseaux de neurones (modèle MLP) a donné des résultats satisfaisants avec des coefficients de corrélation globaux R^2 élevés ($R^2(T_{air,min})$=0.9421, $R^2(T_{air,max})$=0.9250, R^2(HR)=0.8090, R^2(DE)=0.8104). En outre, d'après la figure III.8, le modèle ANN-MLP n'a pas montré d'erreur proportionnelle et systématique,

étant donné que la distribution des erreurs des deux côtés de la ligne médiane est aléatoire, précisément pour HR et DE.

III.4.2 Modélisation par le modèle ANFIS

Le matériel utilisé dans cette partie est le logiciel MATLAB R2018b. MATLAB est un puissant outil de calcul numérique, de programmation et de visualisation graphique. Son nom signifie *matrix laboratory*, c'est à dire un environnement interactif de travail avec des matrices. La facilité de développement des applications dans son langage à fait en sorte qu'il est devenu pratiquement le standard dans son domaine. Actuellement, on trouve des nombreuses boîtes à outils (Toolboxes) qui contiennent des fonctions spécialisées permettant d'utiliser l'environnement MATLAB pour résoudre des classes spécifiques de problèmes.

Avec ses fonctions spécialisées, MATLAB peut être aussi considéré comme un langage de programmation adapté pour les problèmes scientifiques. MATLAB est un interpréteur : les instructions sont interprétées et exécutées ligne par ligne, que l'on appelle un algorithme.

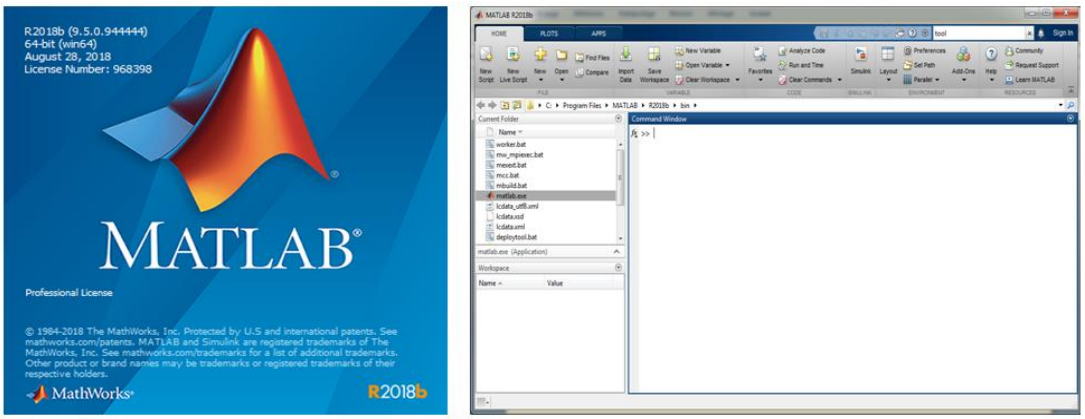


Figure III.9 : Logiciel de MATLAB R2018b.

III.4.2.1 Résultats de corrélation par le modèle ANFIS

L'architecture de l'ANFIS (Adaptive Network Based Fuzzy Inference System) c'est un système d'inférence adaptatif neuro-fuzzy qui consiste à utiliser un réseau de neurones de type MLP à 5 couches pour lequel chaque couche correspond à la réalisation d'une étape d'un système d'inférence floue de type Takagi Sugeno. (Voir section II.3.3)

L'implémentation MATLAB de la régression en utilisant ANFIS est la fonction *anfis* (version R2018b).

Syntaxe :

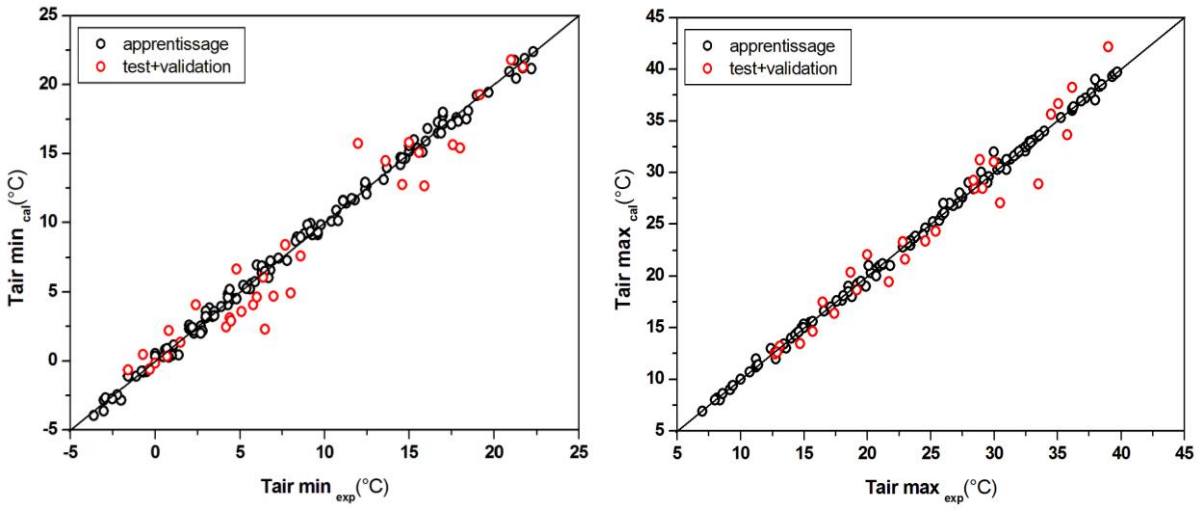
```
fis = anfis(trainingData)
fis = anfis(trainingData,options)
[fis,trainError] = anfis(___)
[fis,trainError,stepSize] = anfis(___)
[fis,trainError,stepSize,chkFIS,chkError] = anfis(trainingData,options)
```

Les fonctions d'appartenance (Membership Functions MF) gaussmf et linear ont été utilisé dans la modélisation pour l'entrée et la sortie respectivement. L'ANFIS a été formé selon une méthode prédéfinie dans la boite à outils de la logique floue MATLAB, appelé méthode hybride.

Pour appliquer la méthode ANFIS, la base de données est composée de sept entrées (colonnes) est 156 points expérimentaux (lignes), Les 156 points de données ont été subdivisés en ensembles de données d'apprentissage, de validation et de test. Les données d'apprentissage, comprenant 80% des points de données, ont été utilisées pour construire et former la structure du modèle mis en œuvre, 10% des points ont été utilisés pour la validation des paramètres du modèle et les 10% restants ont été utilisés pour tester l'estimation de données.

Les résultats de la modélisation ANFIS des quatre paramètres climatiques ($T_{air,min}$, $T_{air,max}$, HR et DE) sont présentés sous forme de digrammes de dispersion (y_{cal}) en fonction de (y_{exp}), comme illustré sur la figure III.10.

La figure III.11 montre une autre présentation de résultat dont la variation des ER% pour 156 points expérimentaux.



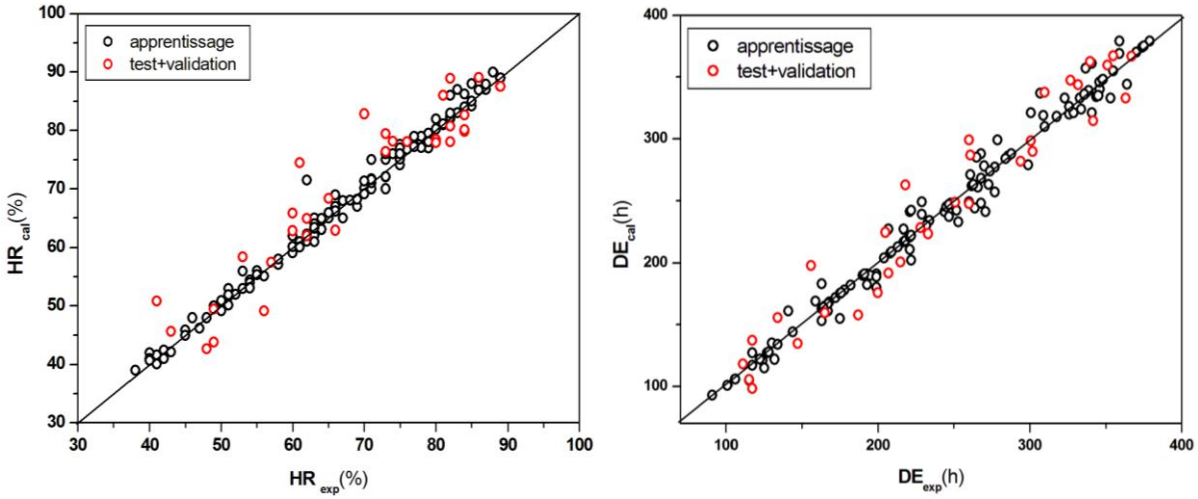


Figure III.10 : Prédiction des paramètres climatiques : $T_{air,min}$, $T_{air,max}$, HR et DE par le modèle ANFIS.

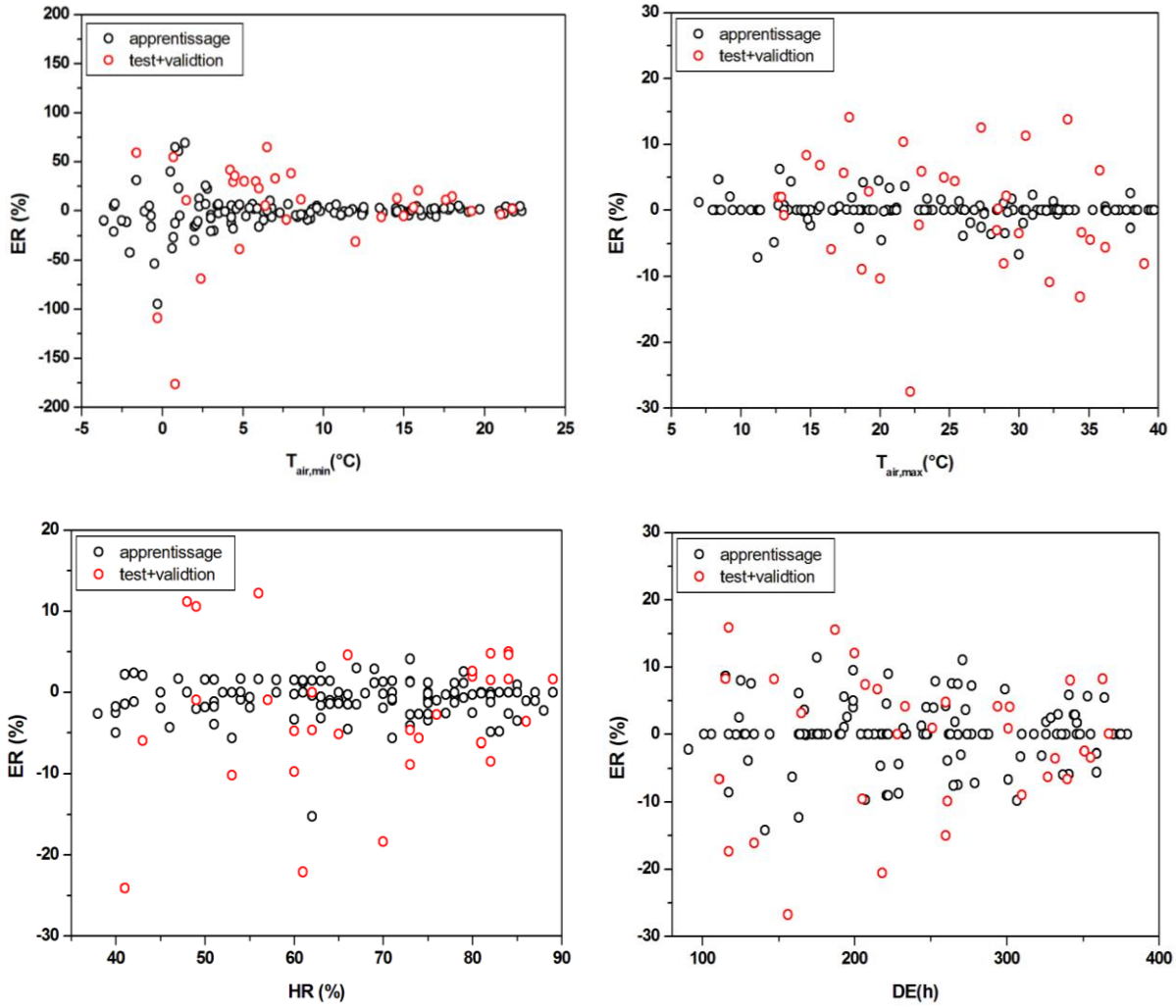


Figure III.11 : L'ER du modèle ANFIS des paramètres : $T_{air,min}$, $T_{air,max}$, HR et DE.

D'après la figure III.10, une corrélation étroite entre les valeurs prédites par le modèle ANFIS et les valeurs expérimentales a été constatée pour l'ensemble d'apprentissage. A partir du tableau III.11, le modèle ANFIS a donné des excellents résultats pour l'ensemble d'apprentissage avec des coefficients de corrélation R^2 élevés, tandis que les coefficients globaux ont des valeurs inferieures. Comme le montre la figure III.11, le modèle a montré des erreurs proportionnelles et systématiques pour l'ensemble d'apprentissage, tandis que pour l'ensemble de test, la distribution des points des deux côtés de la ligne médiane (ligne zéro) est aléatoire.

III.4.3. Modélisation par le modèle SVM

L'algorithme des machines à vecteurs supports SVM (en anglais Support Vector Machine) comprend trois paramètres à optimiser : le paramètre de capacité C, ε de la fonction de perte ε-insensible et le type de fonction du noyau. Tout d'abord, la fonction du noyau devrait être décidée, ce qui détermine la distribution de l'échantillon dans l'espace de mappage. Parmi les fonctions d'activation utilisées en modélisation SVM (la fonction linéaire, la fonction gaussienne, la fonction rbf, et la fonction polynomial), la fonction de base radiales (Radial Basis Function : RBF) a été choisi dans cette étude en raison de leur bonne performance générale.

Deuxièment, le γ de la fonction du noyau affecte grandement le nombre de vecteurs de support, ce qui a une relation étroite avec les performances du SVM et le temps d'apprentissage. Trop de vecteurs de support pourraient produire un sur-apprentissage et augmenter le temps d'entraînement. De plus, γ contrôle l'amplitude de la fonction RBF.

Le paramètre de ε-insensible empêche tout l'ensemble d'apprentissage de répondre aux conditions aux limites et permet donc la possibilité de rareté dans la solution de formulation double. La valeur optimale de ε dépend du type de bruit présent dans les données, qui est généralement inconnu.

Finalement, l'effet du paramètre de capacité C a été testé. Il contrôle le compromis entre la maximisation de la marge et la minimisation de l'erreur d'entraînement. Si "C" est trop petit, le stress lié à l'ajustement des données d'entraînement sera insuffisant. Si "C" est trop grand, l'algorithme sur-adaptera les données d'apprentissage.

L'implémentation MATLAB de la régression en utilisant SVM est la fonction *fitrsvm* (version R2018b).

Syntaxe :

```
Mdl = fitrsvm(Tbl,ResponseVarName)
Mdl = fitrsvm(Tbl,formula)
Mdl = fitrsvm(Tbl,Y)
Mdl = fitrsvm(X,Y)
Mdl = fitrsvm(___,Name,Value)
```

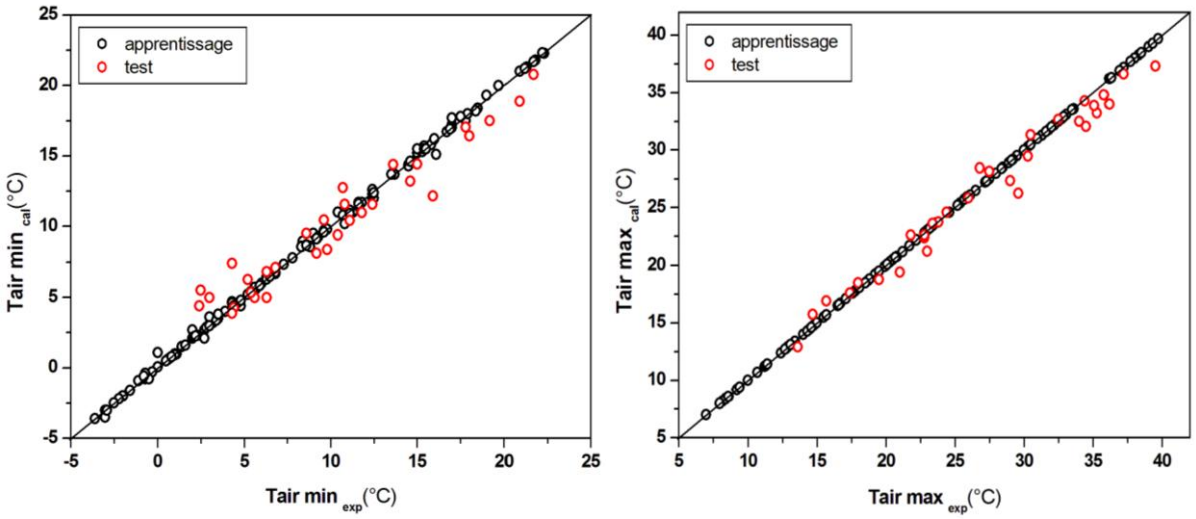
III.4.3.1. Résultats de corrélation par le modèle SVM

Les paramètres optimisés du modèle SVM, la fonction noyau (Kernel function), le paramètre de planéité **C**, la fonction de perte **ε**, et le paramètre du noyau **γ**, sont représentés dans le tableau III.10.

Tableau III.10 : Les paramètres optimisés du modèle SVM.

	$T_{air,min}$	$T_{air,max}$	**HR**	**DE**
C	157	34	130	100
ε	0.023	0.021	0.018	0.020
γ	1.21	0.96	1.84	1.37
Fonction noyau	Rbf	rbf	rbf	rbf

Les résultats de la modélisation par SVM des quatre paramètres climatiques ($T_{air,min}$, $T_{air,max}$, HR et DE) sont présentés sous forme de deux graphes, Un digrammes de dispersion illustre les valeurs calculées par SVM et les valeurs expérimentales (figure III.12) ci-dessous. Et un deuxième graphe qui représente la variation des ER% pour 156 points expérimentaux, (figure III.13). Les paramètres statistiques trouvés pour ce modèle sont reportés sur le tableau III.11.



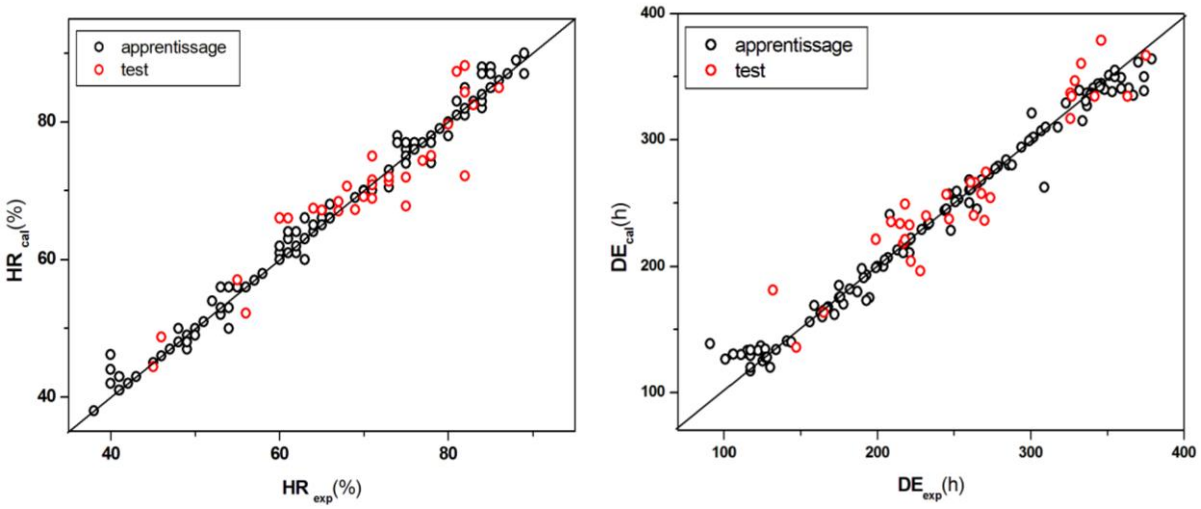


Figure III.12 : Prédiction des paramètres climatiques : $T_{air,min}$, $T_{air,max}$, HR et DE par le modèle SVM.

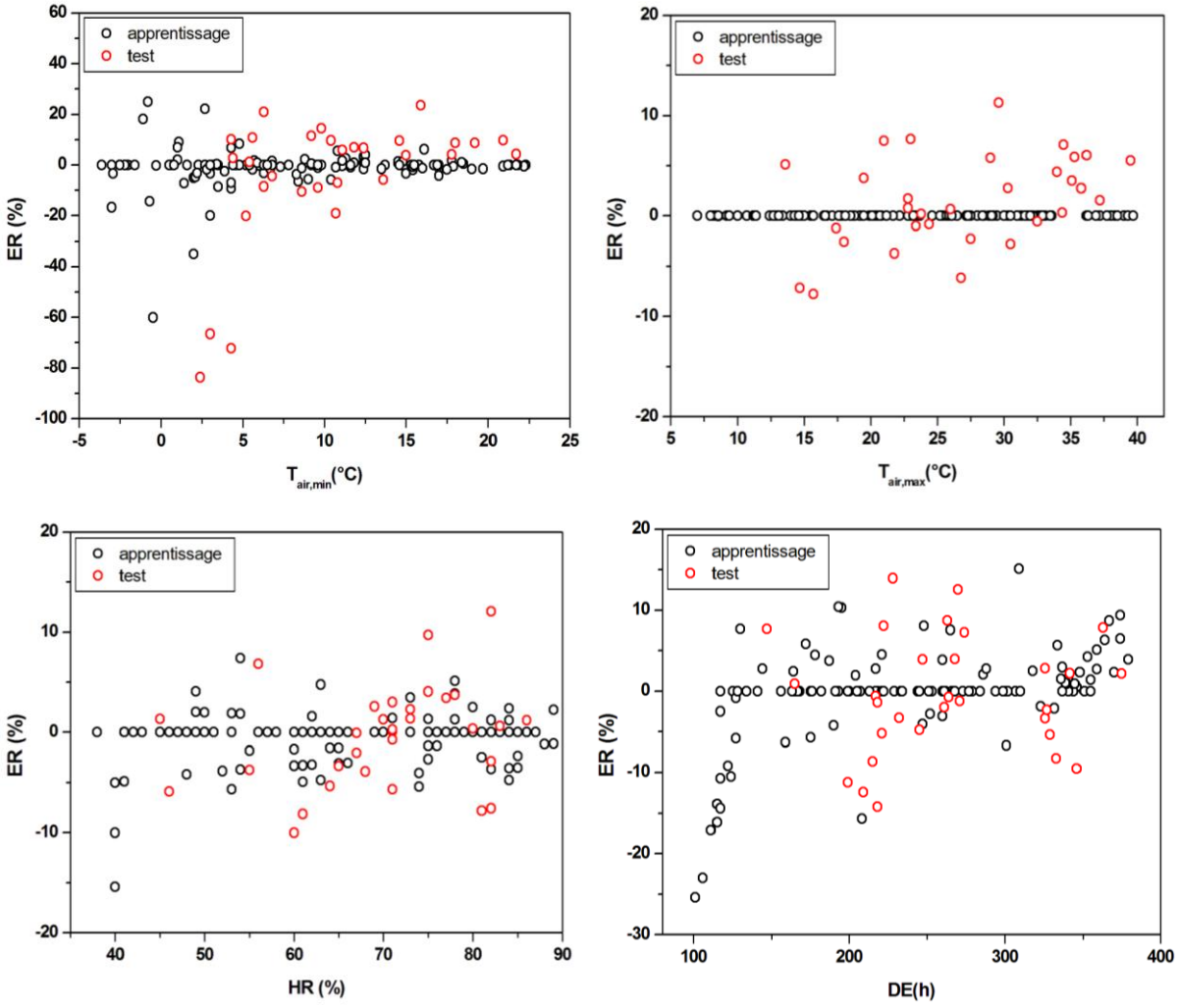


Figure III.13 : L'ER du modèle SVM des paramètres : $T_{air,min}$, $T_{air,max}$, HR et DE.

D'après la figure III.12, une bonne concordance entre les valeurs prédites par le modèle SVM et les valeurs expérimentales a été constatée pour l'ensemble d'apprentissage, comme on peut le remarquer à travers la dispersion serrée des points sur la première bissectrice. A partir du tableau III.11, le modèle SVM a donné de bons résultats avec des coefficients de corrélation R élevés pour l'ensemble d'apprentissage, tandis que pour l'ensemble de test et les coefficients globaux ont des valeurs inferieures, et ce pour les quatre paramètres climatiques. Comme le montre la figure III.13, le modèle a montré des erreurs proportionnelles et systématiques pour l'ensemble d'apprentissage, tandis que pour l'ensemble de test, la distribution des points des deux côtés de la ligne médiane (ligne zéro) est aléatoire.

III.4.4. Comparaison des trois Modèles ANN-MLP, ANFIS et SVM

La fiabilité et la précision des modèles étudiés reposent sur divers critères statistiques : Erreur quadratique moyenne (EQM), Racine de l'erreur quadratique moyenne (REQM), Erreur relative absolue moyenne (ERAM), Coefficient de détermination (R^2), Coefficient de corrélation (R), Facteur de précision (A_f) et Facteur de biais (B_f) qui sont données par les formules suivantes (Soleimani et al., 2017) :

$$EQM = \frac{1}{N}\sum_{i=1}^{N}(y_i^{exp} - y_i^{cal})^2 \quad \text{(III.2)}$$

$$REQM = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i^{exp} - y_i^{cal})^2} \quad \text{(III.3)}$$

$$ERAM = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i^{exp} - y_i^{cal}}{y_i^{exp}}\right| \quad \text{(III.4)}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i^{exp} - y_i^{cal})^2}{\sum_{i=1}^{N}(y_i^{exp} - \overline{y^{exp}})^2} \quad \text{(III.5)}$$

$$R = \left(\left[\sum_{i=1}^{N}(y_i^{exp} - \overline{y^{exp}}) \times (y_i^{cal} - \overline{y^{cal}})\right] \Big/ \left[\sqrt{\sum_{i=1}^{N}(y_i^{exp} - \overline{y^{exp}})^2 \times \sqrt{(y_i^{cal} - \overline{y^{cal}})^2}}\right]\right) \quad \text{(III.6)}$$

- Facteur de biais B_f et facteur de précision A_f :

Ross (1996) a proposé deux indices différents pour la détermination de la performance du modèle. Le facteur de biais (B_f) est une mesure de la concordance globale entre les valeurs prédites et les valeurs observes. Il indiquera si, et dans quelle mesure, les prévisions se situent au-dessus ou au-dessous de la ligne d'équivalence. Une pondération égale est donnée à la sur-prédiction, où la valeur prédite est supérieure à la valeur observée, et a la sous-prédiction. Le B_f est montré dans l'équation (III.7)

$$B_f = 10^{\left(\frac{\Sigma_{i=1}^{N} log\frac{y_i^{cal}}{y_i^{exp}}}{N}\right)} \qquad \text{(III.7)}$$

Un accord parfait entre les valeurs observées et les valeurs prédites donnerait un B_f de 1. Une valeur supérieure à 1 indiquerait que, en moyenne, les prédictions étaient plus grandes que les valeurs observées et représenteraient un modèle à sécurité intrinsèque ; par exemple un B_f de 1,2 indiquerait que les prévisions étaient 20% plus grandes que les valeurs observées.

Une valeur inférieure à 1 indiquerait un modèle à sécurité intrinsèque où, en moyenne, les prévisions étaient plus courtes que celles observées. Le B_f est maintenant largement utilisé comme mesure de la performance du modèle et constitue un critère commun pour la comparaison de différents modèles. Dans le calcul de B_f, les prédictions sur (+) et sous (-) peuvent s'annuler. Ils ne montrent donc pas les erreurs absolues dans le modèle. Ceci peut être réalisé en utilisant le facteur de précision (A_f). Ceci est basé sur une équation similaire à III.7 mais ignore si la différence entre la valeur prédite et la valeur observée est positive ou négative :

$$A_f = 10^{\left(\frac{\Sigma_{i=1}^{N} \left| log\frac{y_i^{cal}}{y_i^{exp}} \right|}{N}\right)} \qquad \text{(III.8)}$$

Toutes les différences étant désormais positives, la valeur de A_f sera toujours égale (si l'accord est parfait) ou supérieure à un. Une précision de 1,4 signifie qu'en moyenne, la valeur prédite est différente de 40% (plus petite ou plus grande) de la valeur observée.

Où N est le nombre de points de données (paires d'entrée / sortie), y_i^{exp} est la valeur des ensembles de données expérimentales au point d'échantillonnage i, y_i^{cal} est la $i^{ème}$ valeur des données prédites par le modèle. La moyenne des données expérimentales et prédites sont définies respectivement comme suit :

$$\overline{y^{exp}} = \frac{1}{N}\Sigma_{i=1}^{N} y_i^{exp} \qquad \text{(III.9)}$$

$$\overline{y^{cal}} = \frac{1}{N}\Sigma_{i=1}^{N} y_i^{cal} \qquad \text{(III.10)}$$

En outre, les critères suggérés par les chercheurs (Golbraikh et Tropsha, 2002) et (Pratim, 2009) ont été utilisés pour assurer une plus grande validité du modèle ANN. Ces critères statistiques sont donnés comme suit :

$$k = \frac{\sum y_i^{exp} y_i^{cal}}{\sum \left(y_i^{cal}\right)^2} \quad \text{(III.11)}$$

$$k' = \frac{\sum y_i^{exp} y_i^{cal}}{\sum \left(y_i^{exp}\right)^2} \quad \text{(III.12)}$$

Avec

$$\mathbf{0.85} \leq \boldsymbol{k} \leq \mathbf{1.15}$$

$$\mathbf{0.85} \leq \boldsymbol{k'} \leq \mathbf{1.15}$$

$$R_0^2 = 1 - \frac{\sum \left(y_i^{cal} - k y_i^{cal}\right)^2}{\sum \left(y_i^{cal} - \overline{y^{cal}}\right)^2} \quad \text{(III.13)}$$

$$R_0'^2 = 1 - \frac{\sum \left(y_i^{exp} - k' y_i^{exp}\right)^2}{\sum \left(y_i^{exp} - \overline{y^{exp}}\right)^2} \quad \text{(III.14)}$$

$$M = \frac{R^2 - R_0^2}{R^2} \leq 0.1 \quad \text{(III.15)}$$

$$N = \frac{R^2 - R_0'^2}{R^2} \leq 0.1 \quad \text{(III.16)}$$

$$R_m^2 = R^2 \times \left(1 - \sqrt{|R^2 - R_0^2|}\right) \geq 0.5 \quad \text{(III.17)}$$

- Définitions des métriques Q^2 : (Todeschini et al, 2016)

Dans ce paragraphe, les métriques Q^2 analysées sont brièvement présentées. Les trois premiers indicateurs (Q_{F1}^2, Q_{F2}^2 et Q_{F3}^2) sont définis comme suit :

$$Q_{F1}^2 = 1 - \frac{\sum_{i=1}^{n_{out}} (y_i - \hat{y}_{i/i})^2}{\sum_{i=1}^{n_{out}} (y_i - \bar{y}_{TR})^2} \quad \text{(III.18)}$$

$$Q_{F2}^2 = 1 - \frac{\sum_{i=1}^{n_{out}} (y_i - \hat{y}_{i/i})^2}{\sum_{i=1}^{n_{out}} (y_i - \bar{y}_{out})^2} \quad \text{(III.19)}$$

$$Q_{F3}^2 = 1 - \frac{\sum_{i=1}^{n_{out}} \left(y_i - \hat{y}_{i/i}\right)^2 / n_{out}}{\sum_{i=1}^{n_{TR}} (y_i - \bar{y}_{TR})^2 / n_{RT}} \quad \text{(III.20)}$$

Où y_i est la réponse expérimentale du i^{ème} valeur, $\hat{y}_{i/i}$ est la réponse prédite lorsque la i^{ème} valeur n'est pas dans l'ensemble d'apprentissage, n_{TR} et n_{out} sont le nombre des valeurs d'apprentissage et de test respectivement, $\bar{y}_{TR}$ est la valeur moyenne des réponses expérimentales de l'ensemble d'apprentissage et $\bar{y}_{out}$ est la valeur moyenne des réponses expérimentales des valeurs de test externes.

- Le coefficient de corrélation de concordance (CCC) :

Le coefficient de corrélation de concordance (CCC) a été proposé pour évaluer la concordance de deux essais d'un test ou d'un instrument en termes de validité et de répétabilité de la méthode. Plus tard, le coefficient de corrélation de concordance CCC a été introduit pour évaluer la capacité de prédiction des modèles de régression. En particulier, le coefficient de corrélation de concordance, appelé ici Q^2_{ccc} est défini comme suit :

$$Q^2_{ccc} = \frac{2\cdot\sum_{i=1}^{n_{out}}(y_i-\bar{y}_{exp})\cdot(\hat{y}_{i/i}-\bar{y}_{cal})}{\sum_{i=1}^{n_{out}}(y_i-\bar{y}_{exp})^2+\sum_{i=1}^{n_{out}}(\hat{y}_{i/i}-\bar{y}_{cal})^2+n_{out}\cdot(\bar{y}_{exp}-\bar{y}_{cal})^2} \qquad \text{(III.21)}$$

Où $\bar{y}_{exp}$ et $\bar{y}_{cal}$ sont les moyennes des réponses expérimentales et prédites de l'ensemble externe respectivement.

En supposant que les réponses expérimentales de l'ensemble externe sont désignées par X et que les réponses prédites par Y, Q^2_{ccc} peut être exprimé comme suit :

$$Q^2_{ccc} = \frac{2\cdot S_{XY}}{S_{XX}+S_{YY}+(\bar{X}-\bar{Y})^2} \qquad \text{(III.22)}$$

Où S_{XY}, S_{XX} et S_{YY} sont la covariance entre X et Y et leurs variances, respectivement.

Pour comparer à la fois les performances et la qualité de la prédiction des trois modèles développés dans ce travail (ANN-MLP, ANFIS et SVM), une évaluation statistique est donnée dans le tableau III.11. Les coefficients statistiques de la validation pour les trois modèles sont tous acceptables et satisfaisants. En conséquence, ces modèles sont robustes. Aussi, les trois modèles se distinguent par un bon pouvoir de prédiction.

On peut noter aussi que les points dans les trois modèles sont presque alignés sur les lignes d'ajustement alors que la précision augmente graduellement de ANN-MLP à SVM.

Tableau III. 11 : Comparaison des critères statistiques des trois modèles.

	$T_{air,min}$			$T_{air,max}$			HR			DE		
Modèles	**ANN-MLP**	**ANFIS**	**SVM**	**ANN-MLP**	**ANFIS**	**SVM**	**ANN-MLP**	**ANFIS**	**SVM**	**ANN-MLP**	**ANFIS**	**SVM**
EQM_train	3,4710	0.1824	0.0662	6,4319	0.1521	5.1442e-08	32,0740	2.4741	2.4976	198,9917	116.1143	104.5880
EQM _test	4,5993	3.1339	2.2776	5,2959	5.0456	1.7375	46,8692	27.7649	12.9617	527,8663	436.9412	368.8567
EQM _global	3,6952	0.7878	**0.5056**	6,2061	1.2297	**0.3453**	35,0141	7.6620	**4.5770**	224,6014	181.9250	**157.1029**
REQM _train	1,8630	0.4271	0.2572	2,5361	0.3900	2.2681e-04	5,6633	1.5729	1.5804	33,1510	10.7756	11.6012
REQM _test	2,1446	1.7703	1.5092	2,3012	2.0325	1.3181	6,8461	5.2692	3.6002	35,0409	20.9031	19.9714
REQM_global	1,9223	0.8876	**0.7111**	2,4912	1.1089	**0.5876**	5,9172	2.7680	**2.1394**	33,5350	13.4880	**11.6786**
ERAM_train	///	///	///	0,0988	0.0885	9.2239e-	1,0702	1.0163	0.2870	3,1365	2.9310	2.3700

						04						
ERAM_test	///	///	1.3951	3,9001	3.1504	2.8774	5,0828	4.6920	3.9098	10,1346	8.2973	5.8983
ERAM_global	///	///	///	0,9910	0.8702	**0.7713**	1,7270	1.3575	**0.9645**	6,1361	4.0318	**3.0711**
R^2_train	0,9286	0.9963	0.9986	0,9160	0.9982	1.0000	0,8296	0.9862	0.9876	0,8250	0.9758	0.9788
R^2_test	0,8918	0.9324	0.9242	0,9380	0.9089	0.9666	0,6990	0.8568	0.8724	0,7281	0.9443	0.8959
R^2_global	0,9421	0.9837	**0.9889**	0,9210	0.9844	**0.9956**	0,8090	0.9582	**0.9750**	0,8104	0.9593	**0.9695**
R_train	0,9637	0.9981	0.9993	0,9571	0.9991	1.0000	0,9109	0.9938	0.9940	0,9085	0.9900	0.9910
R_test	0,9461	0.9695	0.9689	0,9686	0.9628	0.9880	0,8638	0.9343	0.9356	0,8792	0.9661	0.9682
R_ global	0,9604	0.9920	**0.9945**	0,9597	0.9925	**0.9979**	0,9008	0.9806	**0.9877**	0,9005	0.9850	**0.9869**
k_train	1,0073	0.9983	0.9975	1,0028	0.9985	1.0000	0,9985	0.9927	0.9958	1,0002	0.9996	1.0095
k _ test	1,0210	1.0293	1.0341	1,0003	0.9829	1.0237	1,0041	0.9771	0.9967	1,0081	0.9841	0.9891
k _ global	1,0097	1.0042	1.0052	1,0023	0.9951	1.0053	0,9997	0.9893	0.9960	1,0017	0.9963	1.0049
k'_train	0,9585	1.0001	1.0019	0,9876	1.0013	1.0000	0,9944	1.0069	1.0027	0,9827	0.9986	0.9885
k' _ test	0,9368	0.9459	0.9517	0,9910	1.0097	0.9752	0,9865	1.0182	1.0007	0,9725	1.0096	1.0055
k' _ global	0,9582	0.9891	0.9907	0,9882	1.0031	0.9942	0,9927	1.0093	1.0030	0,9807	1.0009	0.9923
A_f _ train	///	///	///	1.0025	1.0089	1.0000	1.0188	1.0161	1.0147	1.0025	1.0297	1.0329
A_f _ test	///	///	1.0491	1.0087	1.0738	1.0402	1.0111	1.0472	1.0398	1.0004	1.0450	1.0295
A_f _ global	///	///	///	0.9866	1.0219	1.0079	1.0144	1.0264	1.0197	1.0155	1.0408	1.0400
B_f_train	///	///	///	1.0002	0.9999	1.0000	0.9657	1.0070	1.0050	0.9854	0.9978	1.0046
B_f _ test	///	///	1.0456	1.2040	0.9967	0.9834	0.9788	1.0216	1.0029	0.9724	1.0075	1.0117
B_f _ global	///	///	///	0.9854	0.9992	0.9967	0.9877	1.0100	1.0046	0.9711	0.9997	1.0061
R_0^2_train	1.0000	1.0000	1.0000	0.9999	1.0000	1.0000	0.9852	0.9986	0.9996	0.9988	1.0000	0.9990
R_0^2_test	0.9922	0.9983	0.9933	0.9922	0.9971	0.9907	0.9628	0.9863	0.9995	0.9852	0.9976	0.9977
R_0^2_global	0.9998	1.0000	0.9999	0.9985	0.9998	0.9998	0.9861	0.9971	0.9996	0.9987	0.9999	0.9997
$R_0'^2$_train	1.0000	1.0000	1.0000	0.9999	1.0000	1.0000	0.9855	0.9988	0.9997	0.9982	1.0000	0.9987
$R_0'^2$_test	0.9952	0.9927	0.9895	0.9822	0.9989	0.9908	0.9944	0.9916	1.0000	0.9920	0.9991	0.9994
$R_0'^2$_global	0.9984	0.9997	0.9998	0.9888	0.9999	0.9997	0.9866	0.9978	0.9998	0.9922	1.0000	0.9994
M_train	-0.0056	-0.0038	-0.0013	-0.0082	-0.0018	-6.1625e-04	-0.0224	-0.0126	-0.0122	-0.0104	-0.0206	-0.0206
M_test	-0.0605	-0.0703	-0.0747	-0.0997	-0.0970	-0.0249	-0.1822	-0.1512	-0.1456	-0.0421	-0.0677	-0.1136
M_global	-0.0185	-0.0166	-0.0111	-0.0244	-0.0157	-0.0042	-0.0524	-0.0405	-0.0252	-0.0302	-0.0315	-0.0323
N_train	-0.0042	-0.0038	-0.0014	-0.0102	-0.0018	-6.1626e-04	-0.0135	-0.0127	-0.0123	-0.0189	-0.0206	-0.0203
N_test	-0.6032	-0.0647	-0.0706	-0.1128	-0.0990	-0.0250	-0.0162	-0.1574	-0.1462	-0.0425	-0.0693	-0.1155
N_global	-0.0180	-0.0163	-0.0110	-0.0742	-0.0158	-0.0041	-0.0522	-0.0413	-0.0254	-0.0301	-0.0316	-0.0319
R_m^2_train	0.9214	0.9354	0.9620	0.9247	0.9557	1.0000	-0.8466	0.8765	0.8792	0.8235	0.8405	0.8397
R_m^2_test	0.6235	0.6936	0.6813	0.5241	0.6390	0.8166	0.5244	0.5484	0.5615	0.6255	0.6993	0.6102
R_m^2_global	0.8045	0.8581	0.8852	0.8022	0.8620	0.9315	-0.7022	0.7694	0.8223	0.8022	0.8000	0.7972
Q^2_{F1}_train	0.9624	0.9963	0.9986	0.9684	0.9982	1.0000	0.9622	0.9862	0.9876	0.9688	0.9798	0.9788
Q^2_{F1}_test	0.9022	0.9324	0.9336	0.9088	0.9109	0.9731	0.8422	0.8597	0.8890	0.8922	0.9344	0.9088
Q^2_{F1}_global	0.9425	0.9837	0.9890	0.9421	0.9844	0.9956	0.9214	0.9583	0.9751	0.9246	0.9693	0.9686
Q^2_{F2}_train	0.9657	0.9963	0.9987	0.9802	0.9982	1.0000	0.9621	0.9865	0.9884	0.9428	0.9798	0.9804
Q^2_{F2}_test	0.9079	0.9324	0.9242	0.8890	0.9089	0.9666	0.8233	0.8568	0.8724	0.9054	0.9343	0.8959
Q^2_{F2}_global	0.9425	0.9837	0.9889	0.9644	0.9844	0.9956	0.9365	0.9582	0.9750	0.9438	0.9693	0.9685
Q^2_{F3}_train	0.9628	0.9963	0.9986	0.9647	0.9982	1.0000	0.9621	0.9862	0.9876	0.9644	0.9798	0.9788
Q^2_{F3}_test	0.9132	0.9357	0.9532	0.9085	0.9352	0.9790	0.8211	0.8457	0.9354	0.9028	0.9239	0.9371

Q^2_{F3}_global	0.9622	0.9838	0.9896	0.9624	0.9852	0.9958	0.9214	0.9574	0.9772	0.9244	0.9683	0.9705
Q^2_{CCC}_train	0.9655	0.9981	0.9993	0.9755	0.9991	1.000	0.9524	0.9932	0.9937	0.9438	0.9900	0.9887
Q^2_{CCC}_test	0.9214	0.9662	0.9560	0.9622	0.9586	0.9821	0.9471	0.9291	0.9353	0.9276	0.9681	0.9474
Q^2_{CCC}_global	0.9522	0.9918	0.9944	0.9844	0.9923	0.9978	0.9255	0.9794	0.9874	0.9620	0.9849	0.9833

Les paramètres présentés dans le tableau III.11 sont définis pour justifier la validité et la performance des modèles des méthodes ANN-MLP, ANFIS et SVM et indiquer si ces modèles peuvent être appliqués pour estimer les paramètres climatiques à partir des paramètres physico-chimiques de l'eau de surface.

La concordance globale entre les valeurs estimées et les valeurs réelles est indiquée par la valeur B_f. Un B_f de 1 signifie un accord complet. Une valeur supérieure à 1 indiquerait que, en moyenne, les estimations sont supérieures aux valeurs réelles et correspondrait à un modèle acceptable. Par exemple, un B_f de 1,002 montrerait que les estimations étaient 0.2% plus grandes que les valeurs réelles. Une valeur inférieure à 1 indiquerait un modèle à sécurité intrinsèque. Pour les modèles proposés B_f globale est au voisinage de 1.

Les valeurs de A_f globale seront toujours égales à (s'il y a accord parfait) ou supérieure à 1 puisque tous les écarts sont positifs. Par exemple la valeur A_f de 1,0042 indique qu'en moyenne, la valeur estimée est différente de 0,4% (plus grande ou plus petite) de la valeur réelle.

Les prédictions par les modèles des méthodes d'intelligence computationnelle sont idéales si ERAM, R^2, R, A_f et B_f sont proches de 0, 1, 1, 1 et 1, respectivement, avec une supériorité de la performance du modèle SVM.

Le tableau III.11 indique également les valeurs k, k', R^2_0, R'^2_0, M, N, R^2_m ainsi que Q^2 pour tous les ensembles de données, y compris l'apprentissage, le test et le globale, et la validité du modèle de différentes méthodes est clairement démontrée selon les critères d'acceptation (III.11), (III.12), (III.15), (III.16) et (III.17).

Toutes ces constatations ont permis d'affirmer que les modèles obtenus permettent l'établissement, dans une large mesure, d'une relation non linéaire entre les paramètres climatiques ($T_{air,min}$, $T_{air,max}$, HR et DE) et les paramètres physico-chimiques de l'eau de surface de barrage Ghrib. Cependant, une amélioration substantielle des paramètres statistiques pour le modèle SVM peut être notée. Par conséquent, nous pouvons conclure que le modèle SVM semble légèrement meilleur, tant en ce qui concerne la robustesse que le pouvoir de prédiction.

III.5. Prédiction des paramètres physico-chimiques de l'eau à partir des paramètres climatiques (Réseau inverse)

Dans cette partie de cette thèse et d'après les travaux cités dans la section I.2.2 pour la prédiction des paramètres physicochimiques et biologiques de l'eau, ainsi la relation étroite qui existe entre le système climatique et le cycle de l'eau, on a tenté de rechercher dans le sens inverse des modèles de prédiction des paramètres physicochimiques de l'eau à partir des paramètres climatiques et ce par les trois techniques (ANN-MLP, ANFIS, SVM) utilisées dans la première partie (réseau direct).

III.5.1. Collecte de la base de données

Pour la présente application du réseau inverse, nous avons utilisé les données météorologiques mensuelles (la température minimale et maximale de l'air ($T_{air,\ min}$, $T_{air,\ max}$), l'humidité relative (HR) et la durée d'ensoleillement (DE) plus l'Année et le Mois comme des paramètres d'entrée pour les trois techniques ont été utilisés.

Comme paramètres de sortie, les données physico-chimiques mensuelles de l'eau de surface de barrage Ghrib (la température de l'eau (T_{eau}), la saturation en oxygène (O_{2sat}), la demande chimique en oxygène (DCO), le pH, la conductivité électrique (CE), l'ion d'ammonium (NH_4^+), l'ion de nitrate (NO_3^-), la turbidité (Turb), la matière organique (MO), les résidus secs (RS) et la demande chimique en oxygène (DCO)) ont été choisis séparément. La base globale de données est composée de 156 exemples.

Le développement des modèles de cette seconde partie a été mené de la même manière que celle explicitée au paragraphe III.4 de la première partie de notre étude. Les données ont été réparties en deux ensembles : 80% pour l'apprentissage et 20% pour le test.

III.5.2. Modélisation par le modèle ANN-MLP

La modélisation non linéaire au moyen d'un perceptron multicouche (ANN-MLP) a été également établie pour le réseau inverse. Un neurone de sortie a été utilisé pour représenter individuellement les paramètres physico-chimiques de l'eau de surface de barrage Ghrib. L'algorithme d'apprentissage BFGS a été utilisé pour le développement du réseau.

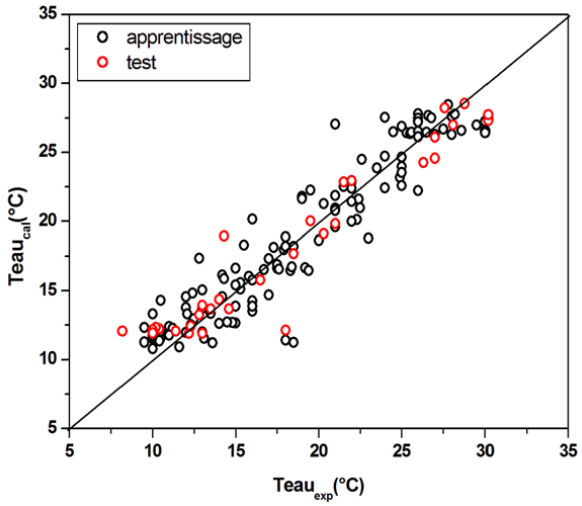
Pour optimiser le nombre de neurones dans la couche cachée, plusieurs calculs ont été effectués (de 1 à 14). Le réseau inverse se composait de 6 entrées et d'une seule sortie. Sur le tableau III.12 sont reportés les résultats de l'optimisation du réseau inverse.

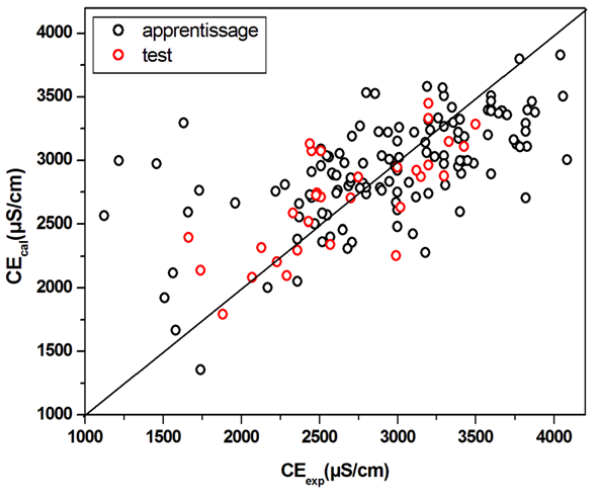
Tableau III.12 : Structure des réseaux de neurones optimisés (Réseau inverse).

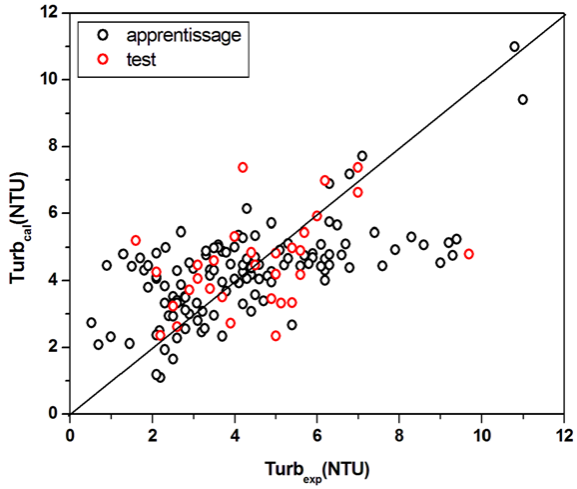
		T_{eau}	CE	Turb	RS	O_2sat	NH_4^+	DCO
Algorithme d'apprentissage		BFGS	BFGS	BFGS	BFGS	BFGS	BFGS	BFGS
Neurones cachés		4	9	11	8	4	8	4
Fonction d'activation	**couche cachée**	Tangente hyperbolique	Sigmoïde	Tangente hyperbolique	Tangente hyperbolique	Sigmoïde	Tangente hyperbolique	Tangente hyperbolique
	couche de sortie	Linéaire	Linéaire	Linéaire	Linéaire	Linéaire	Linéaire	Linéaire

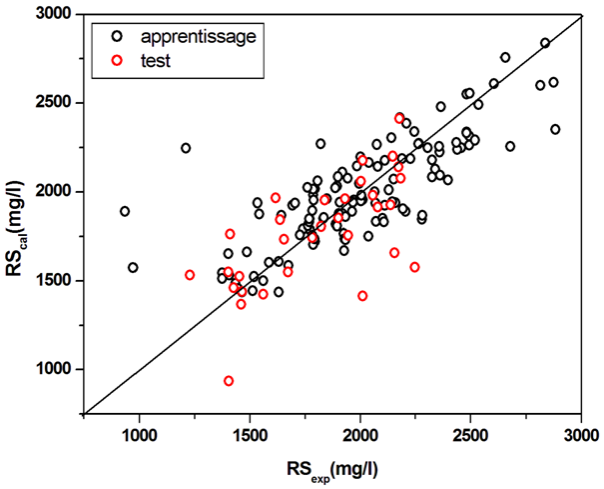
Les critères ERAM et R^2 ont été utilisés pour évaluer la précision des ensembles d'apprentissage et de test. Le meilleur modèle a été choisi pour une ERAM minimale et un coefficient de détermination R^2 maximale (Tableau III.13).

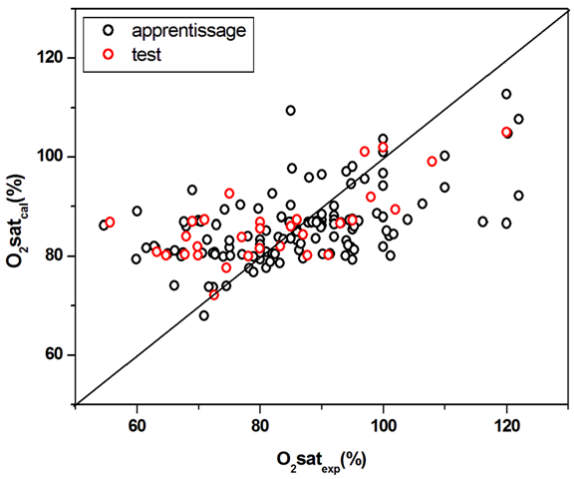
Les valeurs prédites de quelques paramètres physicochimiques de l'eau dans les ensembles d'apprentissage et de test, ont été tracées en fonction de leurs valeurs expérimentales sur la figure III.14. Les autres paramètres physicochimiques (pH, MO, NH_3^-) n'apparaissaient pas au tableau III.13, car les valeurs expérimentales ne sont pas en bonne corrélation avec les valeurs prédites.

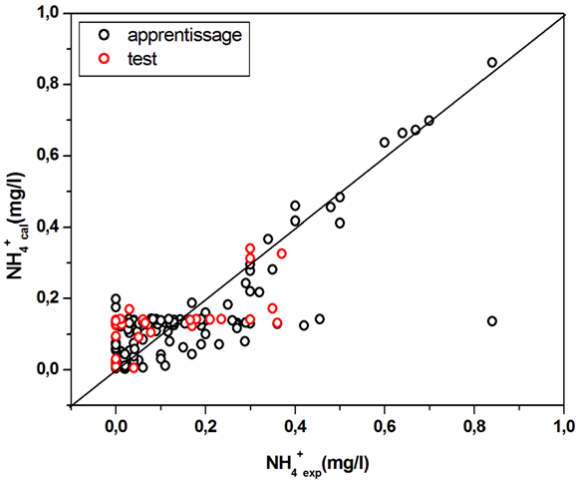

apprentissage
test
Teau$_{cal}$(°C)
Teau$_{exp}$(°C)


apprentissage
test
CE$_{cal}$(µS/cm)
CE$_{exp}$(µS/cm)


apprentissage
test
Turb$_{cal}$(NTU)
Turb$_{exp}$(NTU)


apprentissage
test
RS$_{cal}$(mg/l)
RS$_{exp}$(mg/l)


apprentissage
test
O$_2$sat$_{cal}$(%)
O$_2$sat$_{exp}$(%)


apprentissage
test
NH$_4^+$$_{cal}$(mg/l)
NH$_4^+$$_{exp}$(mg/l)

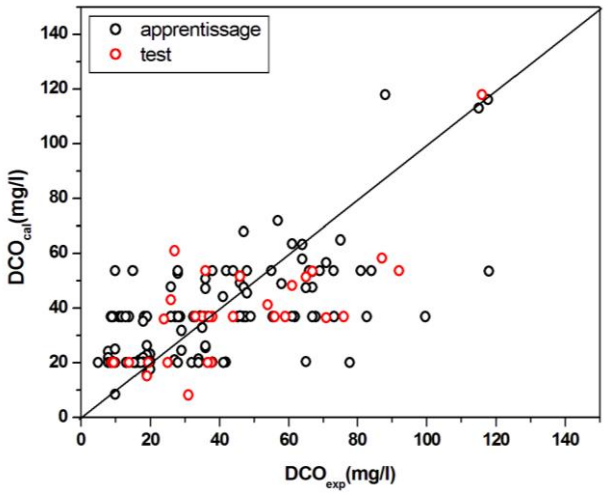
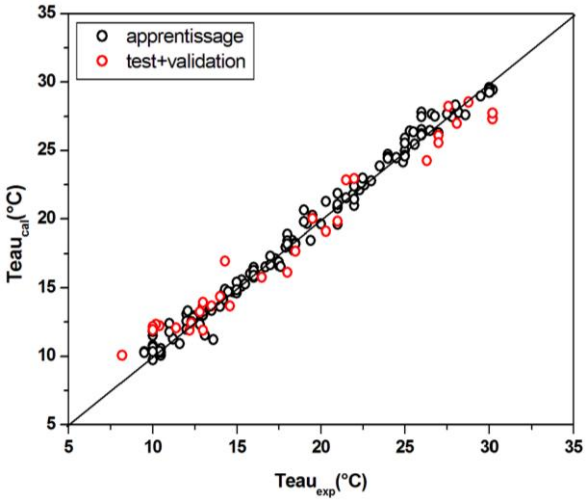
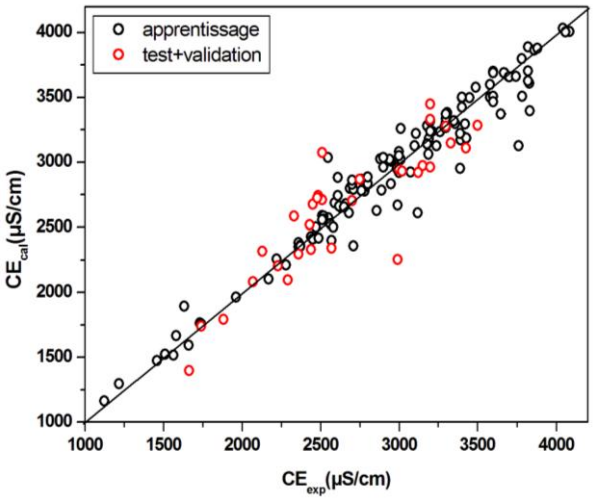


Figure III.14 : Prédiction de quelques paramètres physico-chimiques de l'eau : T_{eau}, CE, Turb, RS, O_2sat, NH_4^+, DCO par le modèle ANN-MLP.

Une corrélation étroite entre les valeurs prédites par le modèle ANN-MLP et les valeurs observées a été trouvée pour la température de l'eau. Tandis que pour les autres paramètres physicochimiques de l'eau, une concordance moyenne entre les valeurs prédites et les valeurs expérimentales a été observée. Les valeurs des ERAM et des coefficients R^2 cités au tableau III.13 confirment ce résultat.

La modélisation par les deux techniques ANFIS et SVM a été également établie pour le réseau inverse (figure III.15 et figure III.16).





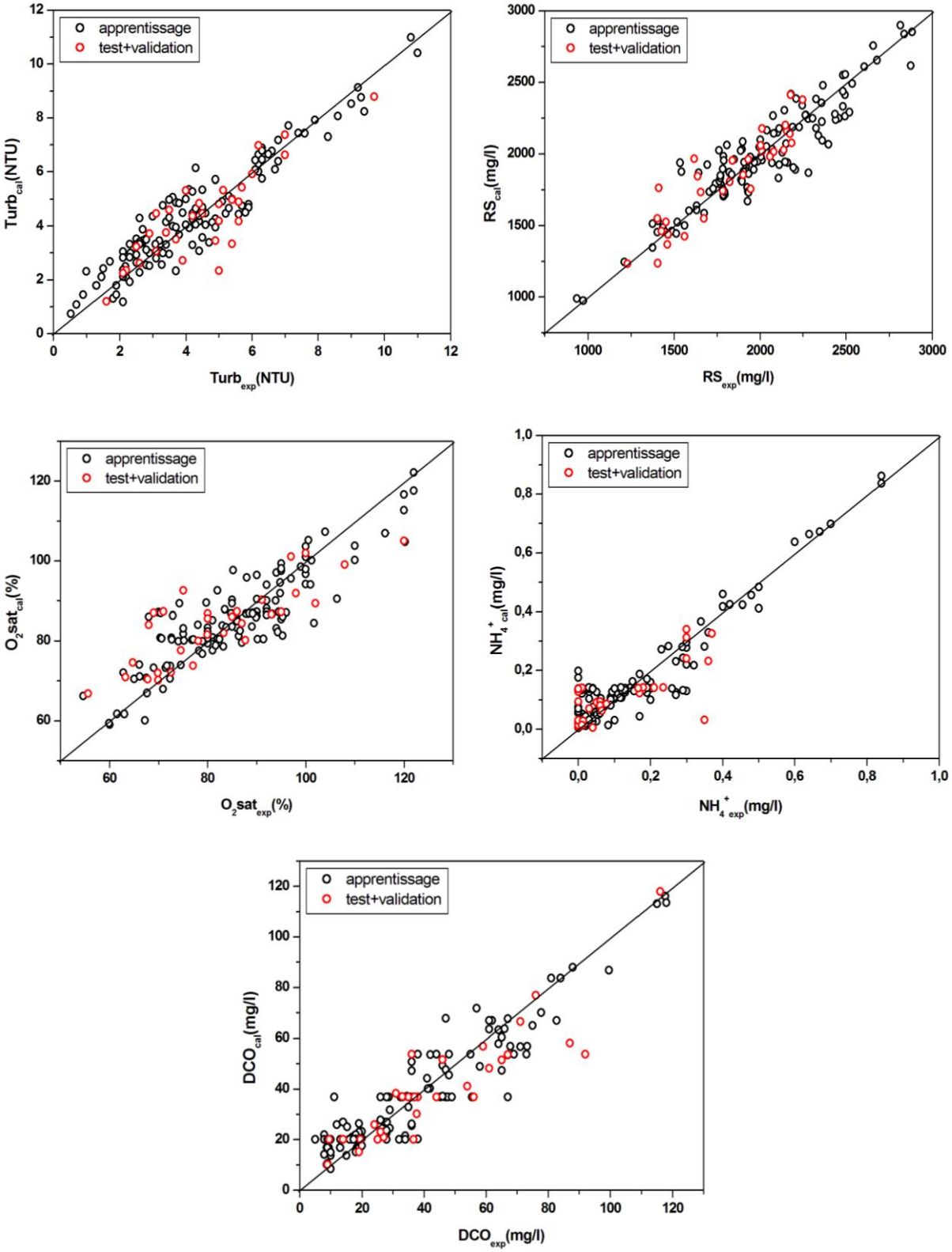
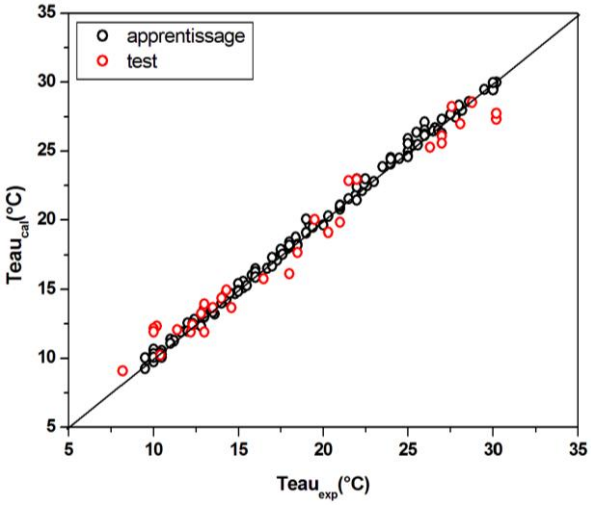
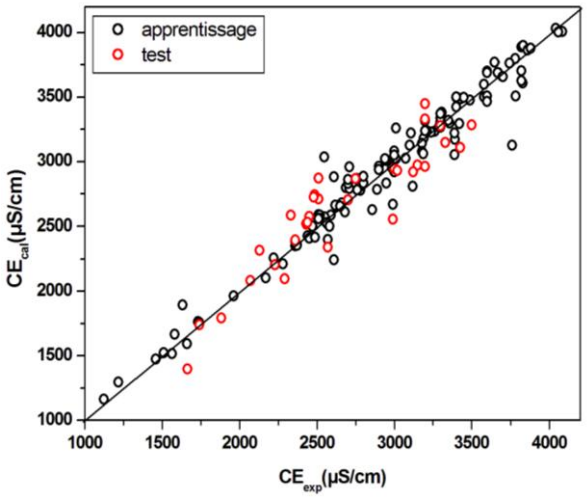
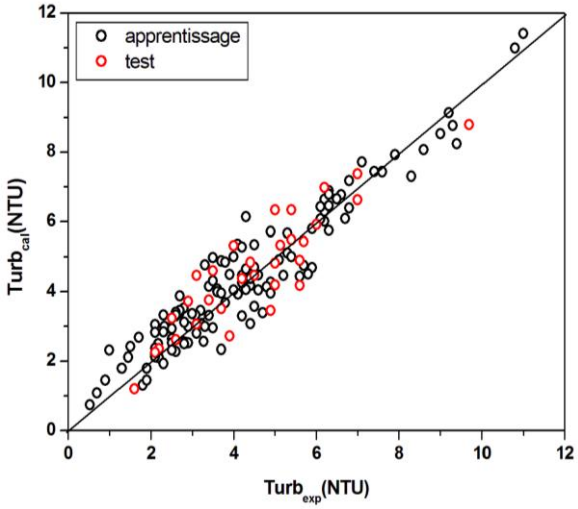
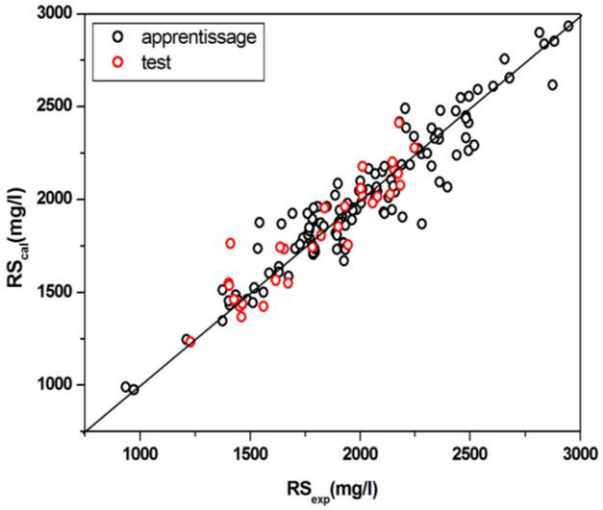


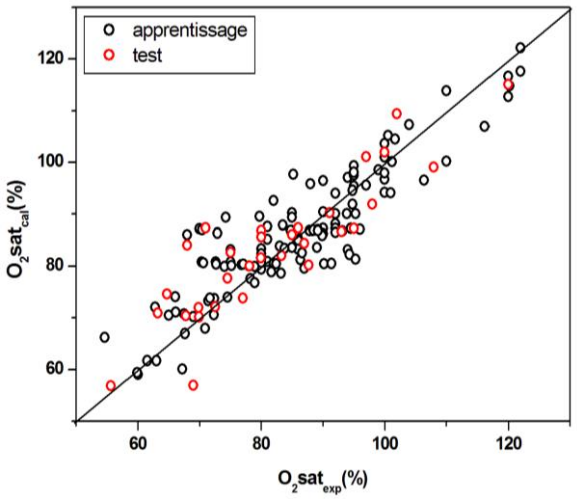
Figure III.15 : Prédiction de quelques paramètres physico-chimiques de l'eau : T_{eau}, CE, Turb, RS, O_2sat, NH_4^+, DCO par le modèle ANFIS.

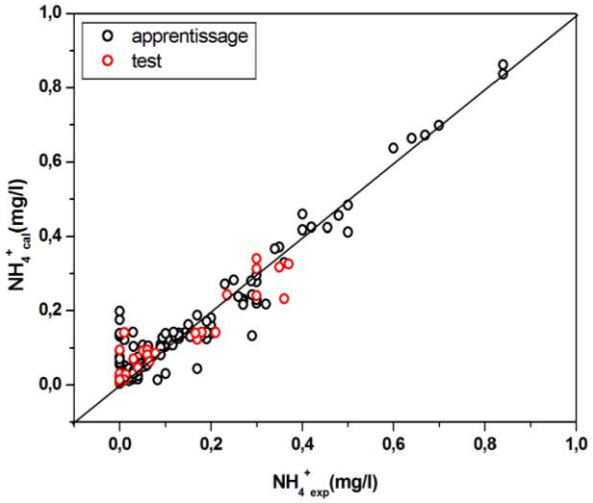

apprentissage
test
Teau$_{cal}$(°C)
Teau$_{exp}$(°C)


apprentissage
test
CE$_{cal}$(µS/cm)
CE$_{exp}$(µS/cm)


apprentissage
test
Turb$_{cal}$(NTU)
Turb$_{exp}$(NTU)


apprentissage
test
RS$_{cal}$(mg/l)
RS$_{exp}$(mg/l)


apprentissage
test
O$_2$sat$_{cal}$(%)
O$_2$sat$_{exp}$(%)


apprentissage
test
NH$_4^+$$_{cal}$(mg/l)
NH$_4^+$$_{exp}$(mg/l)

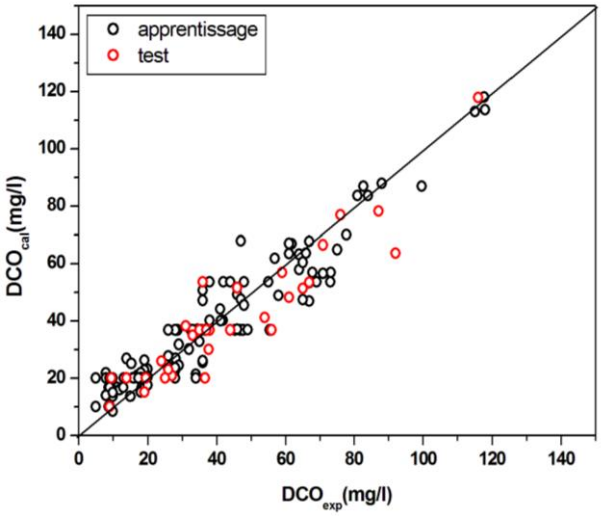


Figure III.16 : Prédiction de quelques paramètres physico-chimiques de l'eau : T_{eau}, CE, Turb, RS, O_2sat, NH_4^+, DCO par le modèle SVM.

III.5.3. Comparaison entre les modèles de prédiction ANN-MLP, ANFIS et SVM (Réseau inverse)

Pour statuer sur le modèle le plus performant d'entre ceux obtenus pour la prédiction des paramètres physico-chimiques, quelques paramètres statistiques y afférents sont regroupés dans le tableau III.13.

Tableau III. 13 : Comparaison des critères statistiques des trois modèles (Réseau inverse).

Paramètres physico-chimiques de l'eau	T_{eau}			CE			O_2sat			RS		
	ANN-MLP	ANFIS	SVM	ANN-MLP	ANFIS	SVM	ANN-MLP	ANFIS	SVM	ANN-MLP	ANFIS	SVM
ERAM_train	9.2546	3.0125	1.5375	14.9402	3.3685	2.5247	10.6953	6.0459	5.6393	9.0010	5.2430	4.5070
ERAM_test	9.9773	7.9030	6.4234	10.7585	7.0011	4.1098	12.1665	9.1854	6.5838	10.3974	6.1515	5.0309
ERAM_global	9.3982	4.0564	**2.5084**	14.1092	4.0904	**2.8088**	10.9876	6.4958	**5.8270**	9.2785	5.4235	**4.6111**
R^2_train	0.8841	0.9870	0.9965	0.3841	0.9043	0.9228	0.2954	0.7679	0.8035	0.6658	0.8724	0.9012
R^2_test	0.9152	0.9565	0.9671	0.5350	0.7686	0.7835	0.3633	0.5376	0.7818	0.2860	0.7856	0.8519
R^2_global	0.8922	0.9795	**0.9892**	0.4251	0.8933	**0.9022**	0.3177	0.7454	**0.8013**	0.6354	0.8678	**0.8998**
R_train	0.9403	0.9936	0.9984	0.6207	0.9423	0.9555	0.5436	0.8773	0.8977	0.8160	0.9348	0.9499
R_test	0.9608	0.9839	0.9874	0.7406	0.8846	0.8925	0.6858	0.7385	0.8883	0.6594	0.9028	0.9276
R_ global	0.9446	0.9901	**0.9948**	0.6529	0.9213	**0.9339**	0.5681	0.8662	**0.8967**	0.8001	0.9321	**0.9490**
k_train	0.9985	0.9956	0.9950	1.0015	1.0063	1.0092	1.0005	1.0001	0.9941	0.9998	1.0054	1.0037

k _ test	1.0124	1.0130	1.0173	0.9842	1.0041	1.0024	0.9568	0.9903	0.9856	1.0131	0.9824	0.9911
k _ global	1.0011	0.9988	0.9991	0.9985	1.0059	1.0019	0.9918	0.9965	0.9925	1.0020	1.0014	1.0016
k'_train	0.9911	1.0033	1.0047	0.9718	0.9914	0.9926	0.9818	0.9941	1.0009	0.9881	0.9901	0.9928
k' _ test	0.9773	0.9819	0.9792	1.0003	0.9881	0.9882	1.0277	0.9964	1.0084	0.9686	1.0124	1.0051
k' _ global	0.9885	0.9993	0.9999	0.9765	0.9909	0.9901	0.9903	0.9968	1.0023	0.9849	0.9938	0.9948
A_f _ train	1.0963	1.0312	1.0154	1.1449	1.0346	1.0446	1.1093	1.0614	1.0568	1.0884	1.0539	1.0463
A_f _ test	1.1010	1.0793	1.0649	1.1090	1.0733	1.0373	1.1191	1.0939	1.0664	1.1146	1.0618	1.0505
A_f _ global	1.0972	1.0406	1.0250	1.1377	1.0422	1.0242	1.1113	1.0656	1.0587	1.0935	1.0555	1.0471
B_f_train	1.0067	1.0070	1.0061	1.0138	0.9963	0.9983	1.0092	1.0047	1.0098	1.0078	0.9957	0.9966
B_f _ test	1.0173	1.0183	1.0063	1.0201	0.9906	0.9916	1.0602	1.0194	1.0174	0.9748	1.0129	1.0071
B_f _ global	1.0088	1.0093	1.0061	1.0151	0.9951	0.9901	1.0191	1.0098	1.0113	1.0011	0.9991	0.9987
R_0^2_train	1.0000	0.9998	0.9997	0.9999	0.9990	0.9980	1.0000	1.0000	0.9982	1.0000	0.9991	0.9996
R_0^2_test	0.9984	0.9983	0.9972	0.9892	0.9995	09985	0.7393	0.9929	0.9911	0.9940	0.9891	0.9970
R_0^2_global	1.0000	1.0000	1.0000	0.9999	0.9991	0.9990	0.9906	0.9993	0.9972	0.9999	0.9999	0.9999
$R_0'^2$_train	0.9991	0.9999	0.9998	0.9817	0.9983	0.9966	0.9868	0.9986	1.0000	0.9961	0.9973	0.9986
$R_0'^2$_test	0.9959	0.9974	0.9966	1.0000	0.9959	0.9962	0.9739	0.9995	0.9976	0.9628	0.9942	0.9990
$R_0'^2$_global	0.9986	1.0000	1.0000	0.9873	0.9981	0.9984	0.9964	0.9996	0.9998	0.9937	0.9989	0.9993
M_train	-0.1311	-0.0129	-0.0033	-1.6033	-0.0579	-0.0529	-2.3857	-0.3023	-0.2423	-0.5020	-0.1453	-0.1092
M_test	-0.0909	-0.0438	-0.0311	-0.8489	-0.3005	-0.3065	-1.0348	-0.8468	-0.2676	-2.4758	-0.2590	-0.1703
M_global	-0.1208	-0.0209	-0.0109	-1.3520	-0.0821	-0.0841	-2.1180	-0.3406	-0.2445	-0.5735	-0.1522	-0.1113
N_train	-0.1302	-0.0130	-0.0033	-1.5559	-0.0571	-0.0581	-2.3411	-0.3005	-0.2445	-0.4961	-0.1431	-0.1080
N_test	-0.0881	-0.0428	-0.0305	-0.8691	-0.2958	-0.2978	-1.6805	-0.8593	-0.2760	-2.3664	-0.2654	-0.1726
N_global	-0.1193	-0.0209	-0.0109	-1.3223	-0.0810	-0.0861	-2.1363	-0.3411	-0.2477	-0.5638	-0.1510	-0.1105
R_m^2_train	0.5831	0.8757	0.9396	0.0827	0.7235	0.7805	0.4474	0.3979	0.4490	0.2809	0.5619	0.6185
R_m^2_test	0.6512	0.7608	0.7993	0.1745	0.3992	0.3912	0.2405	0.2749	0.4242	0.4453	0.4312	0.5275
R_m^2_global	0.5993	0.8394	0.9965	0.1028	0.6691	0.6891	0.4571	0.3698	0.4466	0.2518	0.5524	0.6151
Q^2_{F1}_train	0.8841	0.9870	0.9965	0.3841	0.9443	0.9843	0.2954	0.7679	0.8035	0.6658	0.8724	0.9012
Q^2_{F1}_test	0.9169	0.9573	0.9677	0.6508	0.8262	0.8362	0.4051	0.5554	0.7962	0.5169	0.8550	0.8998
Q^2_{F1}_global	0.8923	0.9795	0.9892	0.4300	0.9240	0.9420	0.3197	0.7461	0.8019	0.6396	0.8693	0.9009
Q^2_{F2}_train	0.8870	0.9874	0.9966	0.4905	0.9540	0.9630	0.3447	0.7775	0.8173	0.7386	0.9002	0.9227
Q^2_{F2}_test	0.9152	0.9565	0.9671	0.5350	0.7686	0.7866	0.3633	0.5376	0.7818	0.2860	0.7856	0.8519
Q^2_{F2}_global	0.8922	0.9795	0.9892	0.4251	0.9233	0.9353	0.3177	0.7454	0.8013	0.6354	0.8678	0.8998
Q^2_{F3}_train	0.8841	0.9870	0.9965	0.3841	0.9443	0.9423	0.2954	0.7679	0.8035	0.6658	0.8724	0.9012
Q^2_{F3}_test	0.8870	0.9420	0.9562	0.7069	0.8541	0.8621	0.3175	0.5006	0.7661	0.5836	0.8750	0.9136

Q^2_{F3}_global	0.8846	0.9781	0.9885	0.7782	0.9264	0.9614	0.2998	0.7387	0.7961	0.6495	0.8729	0.9037
Q^2_{CCC}_train	0.9393	0.9935	0.9982	0.5680	0.9509	0.9819	0.4498	0.8633	0.8885	0.7996	0.9333	0.9492
Q^2_{CCC}_test	0.9519	0.9754	0.9819	0.7257	0.8837	0.8885	0.5243	0.6843	0.8801	0.6529	0.8985	0.9267
Q^2_{CCC}_global	0.9425	0.9894	0.9945	0.6090	0.9403	0.9723	0.4655	0.8453	0.8879	0.7915	0.9313	0.9485

suite (tableau III.13)

Paramètres physico-chimiques de l'eau	**Turb**			**NH_4^+**			**DCO**		
	ANN-MLP	**ANFIS**	**SVM**	**ANN-MLP**	**ANFIS**	**SVM**	**ANN-MLP**	**ANFIS**	**SVM**
ERAM_train	42.9741	19.4926	17.2460	///	///	///	64.8376	37.3272	32.6355
ERAM_test	31.1566	16.0128	14.2895	///	///	///	41.3891	21.2331	19.9457
ERAM_global	40.6258	18.8011	**16.6585**	///	///	///	60.1779	34.1291	**30.1138**
R^2_train	0.4081	0.8730	0.8965	0.6332	0.8706	0.9126	0.4663	0.8668	0.8958
R^2_test	0.3088	0.6992	0.7988	0.4075	0.5805	0.8659	0.5094	0.7855	0.8549
R^2_global	0.3648	0.8496	**0.8834**	0.6115	0.8396	**0.9082**	0.4821	0.8525	**0.8892**
R_train	0.6389	0.9363	0.9480	0.7960	0.9348	0.9569	0.6842	0.9336	0.9492
R_test	0.4475	0.8495	0.8980	0.6608	0.7622	0.9384	0.7470	0.9033	0.9349
R_ global	0.6071	0.9223	**0.9411**	0.7823	0.9175	**0.9547**	0.6960	0.9258	**0.9452**
k_train	1.0007	0.9783	0.9812	0.9924	0.9756	0.9794	0.9820	1.0090	1.0078
k _ test	0.9719	1.0195	0.9769	0.8925	1.0169	1.0409	1.1056	1.0970	1.0743
k _ global	0.9945	0.9861	0.9803	0.9806	0.9792	0.9855	1.0087	1.0279	1.0226
k'_train	0.8823	0.9970	0.9987	0.7894	0.9473	0.9687	0.8375	0.9472	0.9579
k' _ test	0.9137	0.9451	0.9990	0.7448	0.7434	0.8871	0.7980	0.8682	0.9006
k' _ global	0.8888	0.9866	0.9988	0.7843	0.9242	0.9595	0.8278	0.9278	0.9438
A_f _ train	1.4034	1.1961	1.1728	///	///	///	1.5925	1.3336	1.2964
A_f _ test	1.3185	1.1849	1.1507	///	///	///	1.4988	1.2389	1.2182
A_f _ global	1.3861	1.1939	1.1684	///	///	///	1.5734	1.3142	1.2805
B_f_train	1.0785	1.0566	1.0486	///	///	///	1.1825	1.1303	1.1255
B_f _ test	1.0195	0.9612	1.0165	///	///	///	0.9318	0.9414	0.9539
B_f _ global	1.0665	1.0369	1.0422	///	///	///	1.1278	1.0900	1.0891
R_0^2_train	1.0000	0.9972	0.9980	0.9999	0.9989	0.9992	0.9984	0.9997	0.9998
R_0^2_test	0.9904	0.9969	0.9953	0.9587	0.9994	0.9964	0.9485	0.9602	0.9776
R_0^2_global	0.9997	0.9988	0.9976	0.9992	0.9992	0.9996	0.9996	0.9970	0.9981
$R_0'^2$_train	0.9300	1.0000	1.0000	0.9249	0.9953	0.9983	0.9207	0.9916	0.9947

$R_0'^2$_test	0.9391	0.9754	1.0000	0.8881	0.8869	0.9781	0.8408	0.9323	0.9614
$R_0'^2$_global	0.9323	0.9990	1.0000	0.9218	0.9903	0.9972	0.9070	0.9837	0.9901
M_train	-1.4503	-0.1422	-0.1132	-0.5792	-0.1473	-0.0949	-1.1411	-0.1534	-0.1160
M_test	-1.9684	-0.4257	-0.2459	-1.3527	-0.7215	-0.1507	-0.8618	-0.2225	-0.1435
M_global	-1.7406	-0.1756	-0.1293	-0.6340	-0.1900	-0.1007	-1.0736	-0.1695	-0.1225
N_train	-1.2787	-0.1454	-0.1155	-0.4607	-0.1432	-0.0940	-0.9745	-0.1441	-0.1104
N_test	-1.9176	-0.3950	-0.2518	-1.1794	-0.5279	-0.1295	-0.6506	-0.1869	-0.1246
N_global	-1.5559	-0.1758	-0.1320	-0.5075	-0.1795	-0.0981	-0.8814	-0.1538	-0.1135
R_m^2_train	0.0941	0.5655	0.6109	0.2497	0.5588	0.6439	0.1262	0.5507	0.6070
R_m^2_test	0.0036	0.3177	0.4447	0.1050	0.2048	0.5532	0.1719	0.4571	0.5555
R_m^2_global	0.0741	0.5215	0.5849	0.2308	0.5043	0.6336	0.1353	0.5285	0.5957
Q^2_{F1}_train	0.4081	0.8730	0.8965	0.6332	0.8706	0.9126	0.4663	0.8668	0.8958
Q^2_{F1}_test	0.1031	0.7072	0.8041	0.4628	0.6197	0.8785	0.5484	0.8025	0.8664
Q^2_{F1}_global	0.3652	0.8497	0.8835	0.6124	0.8400	0.9084	0.4838	0.8530	0.8896
Q^2_{F2}_train	0.4182	0.8752	0.8982	0.6514	0.8770	0.9169	0.5089	0.8774	0.9041
Q^2_{F2}_test	0.0788	0.6992	0.7988	0.4075	0.5805	0.8659	0.5094	0.7855	0.8549
Q^2_{F2}_global	0.3648	0.8496	0.8834	0.6115	0.8396	0.9082	0.4821	0.8525	0.8892
Q^2_{F3}_train	0.4081	0.8730	0.8965	0.6332	0.8706	0.9126	0.4663	0.8668	0.8958
Q^2_{F3}_test	0.4084	0.8068	0.8708	0.6992	0.7870	0.9319	0.5060	0.7840	0.8539
Q^2_{F3}_global	0.4082	0.8599	0.8914	0.6463	0.8540	0.9164	0.4742	0.8503	0.8875
Q^2_{CCC}_train	0.5821	0.9318	0.9455	0.7796	0.9304	0.9537	0.6500	0.9240	0.9414
Q^2_{CCC}_test	0.4359	0.8446	0.8967	0.5856	0.7297	0.9202	0.7040	0.8780	0.9204
Q^2_{CCC}_global	0.5611	0.9194	0.9388	0.7618	0.9124	0.9509	0.6628	0.9152	0.9376

Les données du tableau III.13 montrent clairement que la techniques SVM pour la prédiction est plus fiable que les techniques ANN-MLP et ANFIS et ce par les valeurs de l'ERAM qui sont les plus petites et par les valeurs de R^2 qui sont les plus élevées comparées aux deux autres techniques. Néanmoins, les trois techniques n'ont pas donné une bonne prédiction des paramètres physico-chimiques, notamment les paramètres indicateurs de la pollution de l'eau, à partir des paramètres climatiques hormis la température de l'eau qui a donné des bonnes valeurs de l'ERAM et de R^2 suivi par les paramètres : O_2sat, RS et NH_4^+ qui ont donné des résultats acceptables. Les valeurs k, k', R^2_0, R'^2_0, M, N, R^2_m ainsi que Q^2 pour tous les ensembles de données, y compris l'apprentissage, le test et le globale, soutiennent les valeurs de l'ERAM et de R^2.

III.6. Conclusion

L'étude comparative des techniques d'intelligence artificielles sur une base de données élaborée dans le cadre de ce travail (156 points expérimentaux) montre que les données des paramètres climatiques sont mieux corrélées par le modèle SVM proposé. En général les techniques de calculs établissent une corrélation assez satisfaisante des données du réseau direct (Prédiction des paramètres climatiques à partir des paramètres physico-chimiques de l'eau). Toutefois, lorsqu'on compare la performance de corrélation des techniques ANN-MLP, ANFIS et SVM pour le réseau inverse (Prédiction des paramètres physico-chimiques de l'eau à partir des paramètres climatiques), les modèles proposés présentent une diminution de performance de corrélation, ce fait est confirmé par les valeurs des divers critères statistiques. De même que pour le réseau direct, la meilleure corrélation est donnée par la technique de SVM, suivie par la technique d'ANFIS puis ANN-MLP.

CONCLUSION GENERALE ET PERSPECTIVES

CONCLUSION GENERALE ET PERSPECTIVES

L'objectif de ce travail consiste en l'étude du fonctionnement hydro-climatique du barrage Ghrib de la région d'Ain Defla-Algérie dans le but d'établir une relation entre la variabilité du climat et celle de la ressource en eau de surface par utilisation de trois approches statistiques ont été développées à savoir : les réseaux de neurones artificiels de type perceptron multicouches (ANN-MLP), le système d'inférence adaptatif neuro-fuzzy (ANFIS) et les machines à vecteurs supports (SVM).

Notre contribution comporte deux axes. Le premier a trait d'une part à l'établissement des trois réseaux, Il consiste à la prédiction des paramètres climatiques (les températures maximale et minimale de l'air, l'humidité relative et la durée d'ensoleillement) à partir de douze variables indépendantes (année, mois, température de l'eau, saturation en oxygène, demande chimique en oxygène, pH, conductivité électrique, ion ammonium, ion nitrate, turbidité, matières sèches et résidus secs).

D'autre part, pour la comparaison des trois modèles proposés, le vecteur d'entrée a été constitué de sept variables indépendantes (année, mois, température de l'eau, saturation en oxygène, conductivité électrique, ion ammonium et résidus secs).

Une base de données constituée de 156 valeurs de paramètres climatiques et physico-chimiques de l'eau du barrage Ghrib, mesurées mensuellement sur une période allant de 2003 à 2015 a été utilisée pour tester les performances des modèles établis.

Les données mensuelles concernant l'année 2016 ont été utilisées pour tester le pouvoir prédictif du modèle ANN-MLP.

Pour le réseau direct, les modèles proposés ont montré une bonne (ANN-MLP) à une très bonne corrélation (ANFIS et SVM) des données expérimentales avec celles prédites notamment pour les températures maximale et minimale de l'air. En termes de coefficients statistiques, les trois modèles sont tous acceptables et satisfaisants, ils se distinguent par un excellent pouvoir prédictif ce qui démontre leurs robustesses. De plus, la précision et l'exactitude des modèles ont été vérifiées avec les valeurs des coefficients Q^2_{F1}, Q^2_{F2}, Q^2_{F3} et Q^2_{CCC}.

La comparaison entre ces trois modèles établis a montré une amélioration substantielle des paramètres statistiques pour le modèle SVM et ANFIS. Par conséquent, le modèle SVM semble légèrement meilleur que le modèle ANFIS. En générale les trois techniques utilisées

ont montré de bonnes valeurs des critères statistiques : R, R^2, R_0^2, $R_0'^2$, R_m^2 et des faibles valeurs de l'ERAM pour l'ensemble d'apprentissage et de test.

Le second axe a été consacré à l'élaboration des modèles de prédiction des paramètres physico-chimiques de l'eau de surface à partir des paramètres climatiques (réseau inverse). Cette contribution a montré que l'utilisation des paramètres climatiques (les températures maximale et minimale de l'air, l'humidité relative et la durée d'ensoleillement) pour la prédiction des paramètres physico-chimiques de l'eau de surface a donné une performance de corrélation acceptable pour les paramètres (T_{eau}, NH_4^+, RS, Turb), et ce sur la base des valeurs des critères statistiques : R, R^2, R_0^2, $R_0'^2$, R_m^2, Q^2_{F1}, Q^2_{F2}, Q^2_{F3}, Q^2_{CCC} et des valeurs de l'ERAM. La technique SVM a donné la meilleure performance suivie par la technique ANFIS.

Cette étude confirme la capacité des techniques ANN-MLP, ANFIS et SVM, dans une large mesure, à prédire les paramètres climatiques avec précision, et les données prédites peuvent donc être utilisées en absence de mesures. Les résultats obtenus indiquent que la modélisation par ces méthodes parait prometteuse pour l'évaluation du potentiel des ressources climatiques ou des ressources en eau de surface dans les zones isolées ou dépourvues de stations de mesure. Il est évident que les trois techniques (ANN-MLP, ANFIS et SVM) sont plus rapides et moins coûteuses. Ces techniques présentent donc d'importants avantages économiques pour un pays comme l'Algérie.

Malgré la bonne précision des modèles développés dans cette thèse, en particulier le modèle SVM, les techniques de modélisation utilisées sont toutes corrélatives et par conséquent nécessitent une large base de données.

Généralement, les modèles ne s'étendent pas bien en dehors de la gamme de données, comme perspectives on suggère :

- L'élargissement de la base de données, basée essentiellement sur une analyse à l'échelle mensuelle.
- Une poursuite logique de ce travail consisterait donc à aborder l'échelle journalière tant du point de vue de la régionalisation des scénarios climatiques que de la modélisation hydrologique.

Ce travail pourrait être élargi sur d'autres barrages ou bassins ou des fleuves au niveau de l'Algérie et même sur les zones côtières. D'autre part, ces données peuvent être confrontées à d'autres études dans de nombreux domaines scientifiques, tels que l'écologie, la qualité de l'eau et les pratiques agricoles afin de détecter des coévolutions qui peuvent traduire une modification des systèmes environnementaux et socio-économiques.

- Modéliser d'autres propriétés climatiques comme la vitesse du vent, les précipitations ou l'évaporation, et modéliser d'autres propriétés de l'eau comme les métaux lourds, à titre d'exemple, qui sont significatives pour leurs applications en science de l'environnement.
- Utiliser d'autres méthodes d'apprentissage telles que les algorithmes évolutifs et hybrides afin de développer d'autres modèles capables de corrélation et de prévision.

REFERENCES BIBLIOGRAPHIQUES

Références bibliographiques

Abba, S. I. Hadi, S. J. and Abdullahi, J. (2017). River water modeling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques. Procedia Computer Science, 120, 75-82.

Adeniji, N.O. Akinpelu, J.A. Adeola, S.O. Adeniji, J.O. (2019). Estimation of Global Solar Radiation, Sunshine Hour Distribution and Clearness Index in Enugu, Nigeria. Journal of Applied Sciences and Environmental Management, 23(2), 345-349.

Afzali, M. Afzali, A. and Zahedi, G. (2012). The potential of artificial neural network technique in daily and monthly ambient air temperature prediction. International Journal of Environmental Science and Development. 3 (1), 33-38.

Alsamamra, H. (2019). Estimation of Global Solar Radiation from Temperature Extremes: A Case Study of Hebron City, Palestine. Journal of Energy and Natural Resources, 8(1), 1-5.

Antanasijević, D. Pocajt, V. Perić-Grujić, A. and Ristić, M. (2014). Modelling of dissolved oxygen in the Danube River using artificial neural networks and Monte Carlo Simulation uncertainty analysis. Journal of Hydrology, 519, 1895-1907.

Antanasijević, D. Pocajt, V. Povrenović, D. Perić-Grujić, A. and Ristić, M. (2013). Modeling of dissolved oxygen content using artificial neural networks: Danube River, North Serbia, case study. Environmental Science and Pollution Research, 20(12), 9006-9013.

Ay, M. and Kişi, O. (2017) Estimation of dissolved oxygen by using neural networks and neuro fuzzy computing techniques. Korean Society of Civil Engineers, 21(5) :1631-1639.

Ay, M. and Kisi, O. (2014). Modeling of chemical oxygen demand by using ANNs, ANFIS and k-means clustering techniques. Journal of Hydrology, 511, 279-289.

Ay, M. Kisi, O. (2012). Modeling of dissolved oxygen concentration using different neural network techniques in Foundation Creek, El Paso County, Colorado. Environmental engineering,138, 654-662.

Bailek, N. Bouchouicha, K. A.Abdel-Hadi, Y. El-Shimy, M. Slimani, A. Jamile, B. Djaafari, A. (2020). Developing a new model for predicting global solar radiation on a horizontal surface located in southwest region of Algeria. NRIAG Journal of Astronomy and Geophysics, 9(1), 341–349.

Beaudry, T. (1984). Traitement des eaux. Ed. Le Griffon d'argile INC, Canada, 231 p.

Beaudry, T. Henry, T.M. (1984). Chimie des eaux. Ed. Les griffons d'argiles, 1ère édition INC, Canada. 537 p.

Beaupoil, C. Bornens P. (1997). Oxygène dissous et toxicité de l'ammoniaque en zones estuariennes : seuils d'acceptabilité. Rapport de synthèse, Biotecmer – AELB, 49 p.

Benatiallah, D. Benatiallah, A. Bouchouicha, K. and Nasri, B. (2019). Estimation of clear sky global solar radiation in Algeria. AIMS Energy, 7(6), 710-727.

Ben El Houari, M. Zegaoui, O. Abdallaoui, A. (2014). Development of mathematical models to Forecasting the Monthly precipitation. American Journal of Engineering Research, 03 (11), 38-45.

Ben El Houari, M. Zegaoui, O. Abdallaoui, A. (2015). Development of an Artificial Neural Network model to predict the monthly air temperature in the region of Meknes (Morocco). International Research Journal of Computer Science, 2 (11).

Bennett, K.P. Mangasarian, O.L. (1992). Robust linear programming discrimination of two linearly inseparable sets. Optimization methods and software, 1(1), 23-34.

Berrizbeitia, S.E. Gago, E.J. and Muneer, T. (2020). Empirical models for the estimation of solar sky-diffuse radiation. A review and experimental analysis. Energies, 13, 701.

Bilgili, M. and Sahin, B. (2009). Prediction of Long-term Monthly Temperature and Rainfall in Turkey. Energy Sources, Part A : Recovery, Utilization and Environmental Effects, 32(1), 60-71.

Boi, P. (2004). A statistical method for forecasting extreme daily temperatures using ECMWF 2-m temperatures and ground station measurements. Meteorological Applications, 11, 245-51.

Bushara, N. and Abraham, O. (2015). Using adaptive neuro-fuzzy inference system (ANFIS) to improve the long-term rainfall forecasting. Journal of Network and Innovative Computing, 3, 146-158.

Casas-Mulet, R. Saltveit, S.J. and Alfredsen, K.T. (2016). Hydrological and thermal effects of hydropeaking on early life stages of salmonids : a modeling approach for implementing mitigation strategies. Science of the Total Environment, 573, 1660-1672

Champiat, D. et Larpent, J.P. (1988). Biologie Des Eaux. Méthodes et Techniques. Ed. Masson. 375 p.

Chronopoulos, K. I. Tsiros, I. X. Dimopoulos, I. F. and Alvertos, N. (2008). An application of artificial neural network models to estimate air temperature data in areas with sparse network of meteorological stations. Journal of Environmental Science and Health, Part A, 43(14), 1752-1757.

Cobaner, M. Citakoglu, H. Kisi, O. and Haktanir, T. (2014). Estimation of mean monthly air temperatures in Turkey. Computers and Electronics in Agriculture, 109, 71-79.

Cristiani, N. Taylor, S.J. (2000). An introduction to support vector machines.

Csabragi, A. Molnar, S. Tanos, P. and Kovacs, J. (2017). 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.

Daneshmand, H. Tavousi, T. Khosravi, M. andTavakoli, S. (2015). Modeling minimum temperature using adaptive neuro-fuzzy inference system based on spectral analysis of climate indices : A case study in Iran. Journal of the Saudi Society of Agricultural Sciences, 14(1), 33-40.

Darand, M. Mansouri-Daneshvar, M.R. (2014). Regionalization of precipitation regimes in Iran using principal component analysis and hierarchical clustering analysis. Environmental Processes, 1 (4), 517-532.

Degremont, (2005). Mémento technique de l'eau, Lavoisier SAS -Lexique technique de l'eau, Paris, 2eme édition.

Deep, H. and Prakash O. (2018). ANFIS model for prediction of water temperature for Scheffler solar reflector Carbon. Science and Technology, 10(1), 27-33.

Degremont. (1989). Mémento technique de l'eau, 9ème édition. Ed. Lavoisier, Paris, Tome 1.

Deweber, J.T. and Wagner, T. (2014). A regional neural network ensemble for predicting mean daily river water temperature. Journal of Hydrology, 517, 187-200

Dombayc, O.A. and Golcu, M. (2009). Daily means ambient temperature prediction using artificial neural network method: A case study of Turkey. Renewable Energy, 24, 1158-1161.

El-Metwally, M. (2005). Sunshine and global solar radiation estimation at different sites in Egypt. Journal of Atmospheric and Solar Terrestrial Physics, 67(14), 1331-1342.

Emamgholizadeh, S. Kashi, H. Marofpoor, I. and Zalaghi, E. (2013). Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. International Journal of Environmental Science and Technology, 11(3), 645-656.

Farajzadeh, J. Alizadeh, F. (2018). A hybrid linear–nonlinear approach to predict the monthly rainfall over the Urmia Lake watershed using wavelet-SARIMAX-LSSVM conjugated model. Journal of Hydroinformatics, 20 (1), 246-262.

Gallice, A. Schaefli, B. Lehning, M. Parlange, M.B. and Huwald, H. (2015). Stream temperature prediction in ungauged basins: review of recent approaches and description of a new physics-derived statistical model. Hydrology and Earth System Sciences, 19, 3727-3753.

Garson, G.D. (1991). Interpreting neural-network connection weights. Artificial Intelligence Expert, 6(4), 46-51.

Golbraikh, A. Tropsha, A. (2002). Beware of q2, J. Mol. Graphics Modelling. 20, 269-276.

Gunawardhana, L.N. Al-Rawas, G.A. and Kazama, S. (2017). An alternative method for predicting relative humidity for climate change studies. Meteorological Applications, 24(4), 551-559.

Hadzima-Nyarko, M. Rabi, A. and Sperac, M. (2014). Implementation of artificial neural networks in modeling the water-air temperature relation-ship of the river Drava. Water Resources Management, 28, 1379-1394.

Hai, T. Sharafati, A. Mohammed, A. Q.Salih, S. C.Deo, R. Al-ansari, N. and Mundher yaseen, Z. (2019). Global Solar Radiation Estimation and Climatic Variability Analysis Using Extreme Learning Machine Based Predictive Model. IEEE Access, 8, 12026-12042.

Hanefi, M. (2019). A hybrid ANFIS-GA approach for estimation of regional rainfall amount. Gazi University Journal of Science, 32(1), 145-162.

He, J. Chu, A. Ryan, M.C. Valeo, C. and Zaitlin, B. (2011). Abiotic influences on dissolved oxygen in a riverine environment. Ecological engineering, 37, 1804-1814.

Heddam, S. Kisi, O. (2017). Extreme learning machines: a new approach for modeling dissolved oxygen (DO) concentration with and without water quality variables as predictors. Environmental Science and Pollution Research, 24(20), 16702-16724.

Heddam, S. (2014). Generalized regression neural network-based approach for modelling hourly dissolved oxygen concentration in the Upper Klamath River, Oregon, USA. Environmental Technology, 35, 1650-1657.

Hilker, T. Lyapustin, A.I. Tucker, C.J. Hall, F.G. Myneni, R.B. Wang, Y. et al. (2014). Vegetation dynamics and rainfall sensitivity of the Amazon. Proceedings of the National Academy of Sciences, 111(45), 16041-16046.

Huld, T.A. Suri, M. Dunlop, E. and Micale, F. (2006). Estimating average daytime and daily temperature profiles within Europe. Environmental Modeling and Software, 21, 1650-1661.

Ihaddadene, R. Ihaddadene, N. Ould Ahmedou Bemba Jed, M.H. and De Souza, A. (2019). Daily global solar radiation estimation based on air temperature: case of study south of Algeria. E3S Web of Conferences 80, 01002.

Jervase, J. Al-Lawati, A. and Dorvlo, A.S. (2003). Contour maps for sunshine ratio for Oman using radial basis function generated data. Renewable Energy, 28(3), 487-497.

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. Environmental Science and Pollution Research, 24(19), 16062-16076.

Kandirmaz, H.M. Kaba, K. and Avci, M. (2014). Estimation of monthly sunshine duration in Turkey using artificial neural networks. International Journal of Photoenergy, 1-9.

Keshtegar, B. Heddam, S. (2017). Modeling daily dissolved oxygen concentration using modified response surface method and artificial neural network : a comparative study. Neural Computing and Applications, 30, 2995-3006.

Kisi, O. and Shiri, J. (2013). Prediction of long-term monthly air temperature using geographical inputs. International Journal of Climatology, 34(1), 179-186.

Krider, L.A. Magner, J.A. Perry, J. Vondracek, B. Ferrington, L.C. (2013). Air-water temperature relationships in the trout streams of southeastern minnesota's carbonate-sandstone landscape. Journal of the American Water Resources Association, 49, 896-907.

Krivoshein, Y.O. Tolstykh, A.V. Tsvetkov, N.A. and Khutornoy, A.N. (2020). Mathematical model for calculating solar radiation on horizontal and inclined surfaces for the conditions of Yakutsk. Conf. Series : Earth and Environmental Science, 408, 012002.

Lauer, F. Bloch, G. (2006). Méthodes SVM pour l'identification. In Journées Identification et Modélisation Expérimentale (JIME 2006).

Leynaud, G. (1968). Les pollutions thermiques, influence de la température sur la vie aquatique. B.T.I. Ministère de l'agriculture, 224-881.

Lu, T. and Viljanen, M. (2008). Prediction of indoor temperature and relative humidity using neural network models: Model comparison. Neural Computing and Applications, 18(4), 345- 357.

Luk, K.C. Ball, J. and Sharma, A. (2001). An application of artificial neural networks for rainfall forecasting. Mathematical and Computer modelling, 33(6), 683-693.

Marcel, D. (1995). Chimie des oxydants et traitement des eaux. L'université de Poitiers (E.S.I.P). 230 p.

Matzarakis, A. P. and Katsoulis, V.D. (2005). Sunshine duration hours over the Greek region. Theoretical and Applied Climatology, 83(1-4), 107-120.

Mercer, J. (1909). Functions of positive and negative type, and their connection the theory of integral equations. Philosophical Transactions of the Royal Society A, 209(441-458), 415-446.

Mohandes, M.A. and Rehman, S. (2013). Estimation of sunshine duration in Saudi Arabia. Journal of Renewable and Sustainable Energy, 5(3), 033128.

Molina, A. Falvey, M. and Rondanelli, R. (2017). A solar radiation database for Chile. Scientific Reports, 7(1).

Morf, H. (2014). Sunshine and cloud cover prediction based on Markov processes. Solar Energy, 110, 615-626.

Najah, A. El-Shafie, A. Karim, O.A. and El-Shafie, A.H. (2013). Performance of ANFIS versus MLP-NN dissolved oxygen prediction models in water quality monitoring. Environmental Science and Pollution Research, 21(3), 1658-1670.

Olyaie, E. ZareAbyaneh, H. and DanandehMehr, A. (2017). A comparative analysis among computational intelligence techniques for dissolved oxygen prediction in Delaware River. Geoscience Frontiers, 8, 517-527.

Onafuye, O.O. Kalu, C. Ozuomba, S. (2018). Adaptive Neuro-Fuzzy Inference System (ANFIS) model for prediction of vertical profile of radio climatic parameters. European Journal of Engineering and Technology, 6 (5).

Ozgur, K. and Murat, A.Y. (2012). Comparison of ANN and ANFIS techniques in modeling dissolved oxygen. Sixteenth International Water Technology Conference, IWTC-16. Istanbul, Turkey.

Paniagua-Tineo, A. Salcedo-Sanz, S. Casanova-Mateo, C. Ortiz-García, E.G. Cony,M.A. Hernández-Martín, E. (2011). Prediction of daily maximum temperature using a support vector regression algorithm. Renewable Energy, 36(11), 3054-3060.

Parishwad, G.V. Bhardwaj, R. K. and Nema, V.K. (1998). Prediction of monthly-mean hourly relative humidity, ambient temperature, and wind velocity for India. Renewable Energy, 13(3), 363-380.

Piccolroaz, S. Calamita, E. Majone, B. Gallice, A. Siviglia, A. Toffolon, M. (2016). Prediction of river water temperature: a comparison between a new family of hybrid models and statistical approaches. Hydrological Processes, 30, 3901-3917.

Pratim Roy, P. Paul, S. Mitra, I. Roy, K. (2009). On two novel parameters for validation of predictive QSAR models. Molecules, 14, 1660-1701.

Pravalie, R. Sîrodoev, I. Peptenatu, D. (2014). Detecting climate change effects on forest ecosystems in southwestern Romania using Landsat TM NDVI data. Journal of Geographical Sciences, 24(5), 815-832.

Rabi, A. Hadzima-Nyarko, M. Sperac, M. (2015). Modelling river temperature from air temperature in the river Drava (Croatia). Hydrological Sciences Journal, 60, 1490-1507.

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. Environmental Earth Sciences, 76(503).

Rajwakuligiewic, A. Bialik, R.J. Rowinski, P.M. (2015). Dissolved oxygen and water temperature dynamics in lowland rivers over various timescales. journal of hydrology and hydromechanics, 63,353-363.

Ramesh, K. Kesarkar, A.P. Bhate, J. VenkatRatnam, M. and Jayaraman, A. (2015). Adaptive neuro-fuzzy inference system for temperature and humidity profile retrieval from

microwave radiometer observations. Atmospheric Measurement Technique, 8(1), 369-384.

Ramirez, M.C.V. Ferreira, N.J. and Velho, H.F. (2005). Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region. Journal of Hydrology, 301, 146-162.

Rangarajan, S. Swaminathan, M. S. and Mani, A. (1984). Computation of solar radiation from observations of cloud cover. Solar Energy, 32(4), 553-556.

Raziei, T. (2017). A precipitation regionalization and regime for Iran based on multivariate analysis. Theor. Appl. Climatol. 1-20.

Robaa, S. (2008). Evaluation of sunshine duration from cloud data in Egypt. Energy, 33(5), 785-795.

Ross, T. (1996). Indices for performance evaluation of predictive models in food microbiology. Journal of Applied Microbiology, 81, 501-508.

Sahin, A.D. (2007). A new formulation for solar irradiation and sunshine duration estimation. International Journal of Energy Research, 31(2), 109-118.

Samandar, A. (2010). Ranking water quality variables using feature selection algorithms to improve generalization capability of artificial neural networks. Scientific Research and Essays, 5, 1254-1259.

Sandersfeld, T. Mark, F.C. Knust, R. (2017). Temperature-dependent metabolism in Antarctic fish: do habitat temperature conditions affect thermal tolerance ranges? Polar Biology, 40, 1-9.

Saplioglu, K. Cimeny, M. and Akman, B. (2010). Daily precipitation prediction by artificial neural networks. Proceedings of the 4th International Scientific Conference on Water Observation and Information System for Decision Support Water Observation and Information System for Balkan Countries. Ohrid, Republic of Macedonia.

Sarmadi, F. Shokoohi, A.R. (2015). Regionalizing precipitation in Iran using GPCC gridded data via multivariate analysis and L-moment methods. Theoretical and Applied Climatology, 122, 121-128.

Shi, X. Lu, W. Zhao, Y. and Qin, P. (2018). Prediction of Indoor Temperature and Relative Humidity Based on Cloud Database by Using an Improved BP Neural Network in Chongqing. IEEE Access, 6, 30559–30566.

SiljicTomic, A. Antanasijevic, D. Ristic, M. Peric-Grujic, A. Pocajt, V. (2018). A linear and non-linear polynomial neural network modeling of dissolved oxygen content in surface water: inter and extrapolation performance with inputs significance analysis. Science of the Total Environment, 610-611, 1038-1046.

Smith, B.A. Hoogenboom, G. McClendon, R.W. (2007). Improving air temperature prediction with artificial neural networks. International Journal of Computational Intelligence, 3(3), 179-186.

Smith, B.A. Hoogenboom, G. McClendon, R.W. (2009). Artificial neural networks for automated year-round temperature prediction. Computers and Electronics in Agriculture, 68(1), 52-61.

Smola, A.J. Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and computing, 14(3), 199-222.

Sojitra, M.A. Purohit, R.C. Pandya, P.A. (2015). Comparative study of daily rainfall forecasting models using adaptive-neuro fuzzy inference system (ANFIS). Current World Environment, 10, 529-536.

Soleimani, R. Saeedi Dehaghani, A.H. Shoushtari, N.A. Yaghoubi, P. Bahadori, A. (2018). Toward an intelligent approach for predicting surface tension of binary mixtures containing ionic liquids. Korean Journal of Chemical Engineering, 35(7), 1556-1569.

Tasaduqq, S. Rehman Bubshait, K. (2002). Application of neural networks for the prediction of hourly mean surface temperatures in Saudi Arabia. Renewable Energy, 25, 545-554.

Temizyurek, M. and Dadaser-Celik, F. (2018). Modelling the effects of meteorological parameters on water temperature using artificial neural networks. Water Science and Technology, 77(6), 1724-1733.

Tereza, C. (1997). Downscaling large-scale circulation to local winter rainfall in north–eastern Mexico. International Journal of Climatology, 17, 1069-1082.

Todeschini, R. Ballabio, D. Grisoni, F. (2016). Beware of unreliable Q2. A comparative study of regression metrics for predictivity assessment of QSAR models. Journal of Chemical Information and Modeling, 56, 10, 1905-1913.

Toffolon, M. Piccolroaz, S. (2015). A hybrid model for river water temperature as a function of air temperature and discharge. Environmental Research Letters, 10, 114011.

Ustaoglu, B. Cigizoglu, H.K. Karaca, M. (2008). Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. Meteorological Applications, 15, 431-45.

Vapnik, V. Cortes, C. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

Vapnik, V.N. (1995). The Nature of Statistical Learning Theory. Springer New York.

VenkataRamana, R. Krishna, B. Kumar, S.R. and Pandey, N.G. (2013). Monthly Rainfall Prediction Using Wavelet Neural Network Analysis. Water Resources Management, 27(10), 3697-3711.

Vliet, M.T.H.V. Yearsley, J.R. Franssen, W.H.P. Ludwig, F. Haddeland, I. Lettenmaier, D.P. and Kabat, P. (2012). Coupled daily stream flow and water temperature modeling in large river basins. Hydrology and Earth System Sciences, 16, 4303-4321.

W.H.O. (1987). Global pollution and health resultas of related environmental monitoring. Global Environment Monitoring system, WHO, UNEP.

Westhoff, J.T. Rosenberger, A.E. (2016). A global review of freshwater cray-fish temperature tolerance, preference, and optimal growth. Fish Biology and Fisheries, 26, 329-349.

Wu, B. Liu, S. Zhu, W. Yu, M. Yan, N. and Xing, Q. (2016). A Method to Estimate Sunshine Duration Using Cloud Classification Data from a Geostationary Meteorological Satellite (FY-2D) over the Heihe River Basin. Sensors, 16(11), 1859.

Yamamoto, K. Togami, T. Yamaguchi, N. and Ninomiya, S. (2017). Machine Learning-Based Calibration of Low-Cost Air Temperature Sensors Using Environmental Data. Sensors, 17(6), 1290.

Yasar, A. Simsek, E. Bilgili, M. Yucel, A. and Ilhan, I. (2011). Estimation of relative humidity based on artificial neural network approach in the Aegean Region of Turkey. Meteorology and Atmospheric Physics, 115(1-2), 81-87.

Yildirim, U. Yilmaz I.O. and Akinoglu B.G. (2013). Trend analysis of 41 years of sunshine duration data for Turkey. Turkish Journal of Engineering and Environmental Sciences, 37, 286-305.

Yin, X. (1999). Bright sunshine duration in relation to precipitation, air temperature and geographic location. Theoretical and Applied Climatology, 64(1-2), 61-68.

Zhao, Z. Gao, J. Wang, Y. Liu, J, Li, S. (2014). Exploring spatially variable relationships between NDVI and climatic factors in a transition zone using geographically weighted regression. Theoretical and Applied Climatology, 120(3-4), 507-519.

Zhu, S. Nyarko, E.K. Nyarko, M.H. (2018a). Modelling daily water temperature from air temperature for the Missouri River. Peer J, 6, 4894.

Zhu, S. Heddam, S. Nyarko, E.K. Hadzima-Nyarko, M. Piccolroaz, S. and Wu, S. (2018b). Modeling daily water temperature for rivers: comparison between adaptive neuro-fuzzy inference systems and artificial neural networks models. Environmental Science and Pollution Research, 26, 402-420.