

# ***Contribution à la modélisation de la qualité des eaux***

**approche par les réseaux de neurones, les systèmes  
neuroflous et les méthodes de classification :  
Application à la station de traitement des eaux  
potables de BOUDOUAOU-ALGERIE**

## **PRESENTEE ET SOUTENUE PAR : HEDDAM SALIM**

Directeur de thèse : Mr. N. DECHEMI Professeur :  
Ecole Nationale Polytechnique d'Alger  
Année universitaire : 2011-2012

Devant le jury composé de : Président : Mr. A. AIDAOUI Professeur : Ecole Nationale Supérieure  
Agronomique d'Alger Examineurs : Mr. T. HARTANI Professeur : Ecole Nationale Supérieure  
Agronomique d'Alger Mme. B. TOUAIBIA : Professeur : Ecole Nationale Supérieure d'Hydraulique-  
Blida Mme. D. SOUAG-GAMANE : M. de conférences A : Université Science et technique Houari  
Boumediene-Alger



# Table des matières

Dédicace . .	7
Remerciements . .	8
تريڤدم لاص خلم . .	9
RESUME . .	10
ABSTRACT . .	11
Liste des Publications & Communications . .	12
Introduction Générale . .	13
Chapitre I Procèdes Generaux de Traitement Des Eaux . .	17
I.1. Introduction . .	17
I.2. Qualité de la ressource . .	18
I.3. Méthodes générales de traitement . .	19
I.3.1. Généralités . .	19
I.3.2. Classement des traitements . .	19
I.4. Procédés généraux . .	20
I.4.1. Stockage d'eau brute . .	20
I.4.2. Dégrossissage . .	21
I.4.3. Prétraitement au chlore . .	21
I.4.4. Traitement physique : Aération . .	21
I.4.5. Pré oxydation . .	22
I.4.6. La clarification . .	23
I.4.7. Coagulation floculation . .	23
I.4.8. La décantation flottation . .	23
I.4.9. Filtration . .	26
I.4.10. Désinfection . .	29
I.5. Conclusion . .	31
Chapitre II Coagulation & Floculation . .	32
II.1. Introduction . .	32
II.2. Coagulation . .	32
II.2.1. Principes de base . .	32
II.2.2. Les particules mises en jeu : les colloïdes . .	33
II.2.3. Phénomène de la coagulation . .	35
II.2.4. Chimie de la coagulation . .	37
II.2.5. Caractéristiques des coagulants . .	39
II.2.6. Les essais Jar-Test . .	40
II.2.7. Coagulants organiques . .	42
II.2.8. Adjuvants . .	42
II.3. Floculation . .	43
II.3.1. Phénomène de floculation . .	43
II.3.2. Nature des floculants . .	43
II.3.3. Processus de la floculation . .	43

II.3.4. Phénomène de base de la floculation . . .	44
II.4. Choix du coagulant . . .	45
II.5. Modélisation de la dose de coagulant : état de l'art. . .	46
II.6. Conclusion . . .	50
Chapitre III Modélisation linéaire de la dose de coagulant . . .	51
III.1. Introduction . . .	51
III.2. L'analyse en composantes principales ACP . . .	52
III.2.1. Généralités . . .	52
III.2.2. Principes de l'ACP . . .	52
III.2.3. Les données de l'ACP et leurs caractéristiques . . .	53
III.2.4. Définitions algébrique . . .	56
III.2.5. Définition géométrique . . .	58
III.2.6. Formulation mathématique de l'ACP . . .	58
III.2.7. Procédé d'application de l'ACP . . .	58
III.3. Résultats obtenus par l'analyse en composantes principales ACP . . .	60
III.3.1. Présentation de la station d'étude . . .	60
III.3.2. Analyse statistique . . .	62
III.3.3. Résultats de l'application de l'ACP . . .	66
III.4. Détermination de l'équation de prédiction par régression linéaire multiple . . .	70
III.4.1. Construction des modèles de régression linéaire multiple . . .	73
III.4.2. Résultats et interprétations . . .	74
III.5. Conclusion et recommandations . . .	77
Chapitre IV Modèles A Base De Réseaux De Neurones Artificiels . . .	78
IV.1. Introduction . . .	78
IV.2. Les fondements biologiques . . .	78
IV.3. le neurone formel . . .	79
IV.4. type de réseaux de neurones . . .	81
IV.4.1. réseaux de neurones non bouclés . . .	81
IV.4.2. réseaux de neurones bouclés . . .	81
IV.5. structure des réseaux de neurones . . .	82
IV.5.1. Le perceptron et l'Adaline . . .	82
IV.5.2. Le perceptron multicouches MLP . . .	83
IV.6. Apprentissage des réseaux de neurones artificiels . . .	85
IV.6.1. Apprentissage supervisé . . .	85
IV.6.2. Apprentissage non supervisé . . .	85
IV.6.3. Apprentissage par renforcement . . .	85
IV.7. la rétro propagation du gradient . . .	86
IV.7. 1. Généralités . . .	86
IV.7. 2. Fonctions coût . . .	86
IV.7. 3. Choix d'initialisation des poids . . .	86
IV.7. 4. Formulation mathématique . . .	87
IV.7. 5. Variantes de l'algorithme . . .	91

IV.8. Propriété fondamentale des réseaux de neurones non bouclés . . .	93
IV.8.1. L'approximation universelle . . .	93
IV.8.2. La parcimonie . . .	93
IV.9. Les réseaux de neurones à fonctions de base radiales (rbfNN) . . .	94
IV.9.1. Généralités . . .	94
IV.9.2. architectures . . .	95
IV.9.3. apprentissage des réseaux RBFNN . . .	96
IV.10. Réseaux de neurones à fonction de régression générale (GRNN) . . .	97
IV.11. Résultats de la modélisation par les réseaux de neurones . . .	100
IV.11.1. Résultats obtenus par le perceptron multicouches MLPNN . . .	101
IV.11.2. Résultats obtenus par le réseau RBFNN . . .	111
IV.11.3. Résultats obtenus par le réseau à régression générale GRNN . . .	113
IV.11.4. Comparaison des résultats obtenus par les réseaux de neurones . . .	116
IV.12. Conclusions . . .	117
<b>Chapitre V Modèles A Base De Systèmes Neuroflous . . .</b>	<b>119</b>
V.1. introduction . . .	119
v.2. historique de la logique floue . . .	120
V.3. La Logique floue . . .	120
V.4. Éléments de théorie des sous-ensembles flous . . .	121
V.4.1. Définition et éléments caractéristiques . . .	121
V.4.2. Raisonnement à partir des sous-ensembles flous . . .	129
V.4.3. Système d'inférence flou (SIF) . . .	132
V.4.4. Types de Système d'inférence flou . . .	136
V.4.5. Caractéristiques d'un Sif . . .	137
V.4.6. Méthodes d'optimisation des SIF . . .	137
V.4.7. Théorème d'approximation universelle . . .	138
V.5. systèmes hybrides neuroflous . . .	138
V.5.1. Système neuroflous type Mamdani. . . .	139
V.5.2. Système Neuroflou type Takagi Sugeno. . . .	140
V.6. le modèle utilisé: Adaptive neuro-Fuzzy Inference System (ANFIS) . . .	140
V.6.1. Apprentissage du modèle ANFIS . . .	143
V.6.2. Mise en évidence . . .	144
V.7. Résultats de la modélisation par le modèle ANFIS . . .	146
V.7.1. Méthodologie . . .	146
V.7.2. Application de la méthode <i>GIRD PARTITION</i> . . .	147
V.7.3. Application de la méthode du Groupage Soustractive . . .	165
V.7.4. Comparaison des résultats obtenus par le modèle ANFIS . . .	174
V.8. Conclusion . . .	175
<b>Chapitre VI Application Des Technique De Classification . . .</b>	<b>178</b>
VI.1. Introduction . . .	178
VI.2. Formalisation mathématique d'un problème de classification . . .	179
VI.3. Les cartes auto organisatrices de KOHONEN . . .	179

VI.3.1. Historique . .	179
VI.3.2. Définition . .	180
VI.3.3. Principe de base . .	181
VI.3.4. Architecture du Réseau . .	181
VI.3.5. Règle de transmission . .	182
VI.3.6. Formalisme mathématique . .	184
VI.4. Analyse des données à l'aide de la carte de Kohonen . .	186
VI.4.1. Visualisation de la topologie obtenue et phase de classification . .	187
VI.5. Conclusion . .	192
Conclusion Générale . .	193
Références Bibliographiques . .	197

## Dédicace

*A tous ceux qui m'aiment A ma Famille A ma femme A ma fille TAKOUA A mes Enseignants  
A mes étudiants A mes amis A l'âme de mon père; son souci était de voir ce travail aboutir A  
eux tous je dédie ce travail*

## Remerciements

Cette thèse de doctorat a été dirigée par Monsieur **Noureddine DECHEMI**, Professeur à l'Ecole National Polytechnique d'Alger (ENP). Ce moment est pour moi, l'occasion de vous témoigner à nouveau ma plus grande estime et ma reconnaissance. Monsieur **DECHEMI** a toujours fait preuve d'une infinie patience. Sa présence discrète mais au combien efficace a été des plus précieuses et sa sempiternelle disponibilité inestimable. Monsieur, il m'est difficile de vous exprimer toute ma reconnaissance, alors tout simplement je vous dis merci.

Je suis très reconnaissant aussi à M. professeur **AIDAOUI Abdellah** qui a bien voulu accepter la présidence du jury, il mérite tous les remerciements.

Je tiens à remercier Monsieur **Tarik HARTANI**, Professeur à l'Ecole Nationale Supérieur Agronomique d'Alger (ENSA), et Madame **Benina TOUAIBIA**, Professeur à l'Ecole nationale Supérieure de l'Hydraulique de Blida, pour l'honneur qu'ils m'ont fait en acceptant d'examiner ce travail et pour leur participation à ce jury, malgré leurs multiples occupations.

Merci également à Madame **Doudja SOUAG-GAMANE**, Maitre de Conférences à l'Université Houari Boumediene à Alger, qui a accepté de participer à mon jury de thèse.

J'exprime mes vifs remerciements aux personnels du laboratoire de la station de traitement des eaux de **BOUDOUAOU (SEAAL)** qui ont toujours étaient disponible pour me faire profiter de leur connaissances en traitement des eaux, et pour les nombreuses orientations qu'ils m'ont donné durant la réalisation de ce travail; à eux et tous le personnel de la station de **BOUDOUAOU**, je présente mes chaleureux remerciements.

Enfin, j'exprime un grand merci à toutes les personnes faisant vivre la science.



## قرنكذمل اص خلم

تهدف هذه الدراسة إلى إيجاد علاقة رياضية ما بين خصائص الماء الخام الغير معالج عند مدخل محطة معالجة المياه و ما بين نسبة مادة سلفات الألمنيوم التي تستعمل خلال مرحلة التختير. عملية النمذجة الرياضية تمت بواسطة الطريقة العلمية المعروفة باسم الشبكات العصبية الاصطناعية.

### **RESEAUX DE NEURONES ARTIFICIELS ET LES SYSTEMES NEURONES FLOUES**

قمنا أولا بتحضير و إعداد بنك معلومات استنادا إلى أرشيف محطة معالجة المياه الخاصة بمدينة الجزائر العاصمة (محطة بوداوي) ثم قمنا بمعالجة المعلومات بواسطة الطريقة الإحصائية المعروفة بالمعالجة عن طريق المقومات الأساسية وأخيرا قمنا بضبط ثوابت النموذج الرياضي المستعمل. لقد أثبتت هذه الدراسة إن الشبكات العصبية الاصطناعية جد فعالة في ميدان معالجة المياه و يجب الاعتماد عليها بجدية من أجل إدخال التآليه في تسيير محطات لمعالجة المياه.

## RESUME

Les usines de production d'eau potable se composent d'un certain nombre de procédés très complexes faisant appel à des procédures physico - chimiques et biologiques; le passage de l'eau dans un tel système provoque une modification notable de sa qualité, celle-ci varie d'une façon significative dans sa composition organique et physico-chimique. Les niveaux de variation se développent dans une large gamme de variables descriptives, en particulier, la turbidité, la température, la conductivité, etc., et dépendent donc des caractéristiques de l'eau et en particulier du mode et du degré d'interaction avec les composantes extérieurs caractérisant l'installation de production à savoir les réactions chimiques. Le grand nombre de variables, ainsi que leur nature dynamique complique grandement le suivi de l'évolution de la qualité de l'eau dans une usine de production.

Parmi toutes les étapes de la chaîne de traitement des eaux nous nous sommes intéressés à la phase de coagulation. C'est un procédé qui fait appelle à des phénomènes très complexes, son efficacité dépend en grand partie de la qualité de l'eau à traiter et principalement de la dose du coagulant à injecter, cette dernière peut être (selon les unités de production) soit à base de sels d'aluminium ou de fer.

Afin d'évaluer les conditions optimales de coagulation, des essais dits de «Jar-Test» (JT) sont conduits à l'échelle de laboratoire. Ceux-ci menés dans une large gamme de conditions d'opérations permettent, de déterminer le type de coagulant et son dosage. Ce type d'approche (Jar-Test) à l'inconvénient de nécessiter un temps de réponse relativement long, de plus, elle ne permet pas de suivre finement l'évolution de la qualité de l'eau brute. On voit ici tout l'intérêt de disposer d'un contrôle automatique et efficace de ce procédé pour un meilleur rendement de traitement et une réduction des coûts d'exploitation.

Dans cette thèse, on propose une nouvelle méthode de prédiction de la dose du coagulant (CO) en fonction de six variables descriptives caractérisant l'eau brute à l'entrée de la station de traitement des eaux potables de Boudouaou (Alger) à savoir la température (TE), le pH, la turbidité (TU), la conductivité (CON), l'oxygène dissous (OX) et les ultraviolet (UV<sub>254</sub>) ; basée sur la techniques des réseaux de neurones artificiels et les systèmes neuroflous. Pour le modèle à base de réseau de neurones nous avons appliqué trois types d'architecture à savoir le perceptron multicouches (MLPNN), les réseaux de neurones à fonctions radiale de base (RBFNN) ainsi que les réseaux de neurones à fonction de régression générale (GRNN). Pour le modèle neuroflou nous avons appliqué le modèle universel appelé ANFIS (Adaptive Neurofuzzy Inference System). Afin d'évaluer les performances des modèles utilisés, les résultats obtenus sont comparés à ceux d'un modèle à base de régression linéaire multiple (RLM). L'analyse en composantes principales (ACP) a été appliquée afin de déceler d'éventuels effets prépondérants des différents variables pris en considération. Les résultats obtenus sont d'un grand intérêt, d'une part la dose du coagulant varie en fonction des six variables caractérisant l'eau brute, ce qui confirme la complexité du processus mise en jeu, et il n'est y à aucune variable prépondérante, et d'autres part le modèle ANFIS a montré sa capacité à modéliser des processus complexes et non linéaires.

**Mots clé :** *Traitement; Modélisation ; Coagulation; Réseaux de neurones ; logique floue; ANFIS ; MLPNN ; RBFNN ; GRNN ; ACP ; RLM ; Apprentissage ; Validation ; Test.*

## ABSTRACT

Drinking water treatment plants have to provide good water quality and at the same time low operational costs. Owing to various physical, chemical and biological interactions drinking water treatment processes are often difficult to handle and reliable predictions for the course of processes are difficult to obtain. A conventional treatment process employing coagulation/flocculation–sedimentation–filtration is a commonly used treatment method. The coagulation process is one of the most important stages in water treatment plant, which involves many complex physical and chemical phenomena. Moreover, coagulant dosing rate is non-linearly correlated to raw water characteristics such as turbidity, conductivity, pH, temperature, etc. As such, coagulation reaction is hard or even impossible to control satisfactorily by conventional methods. Traditionally, jar tests are used to determine the optimum coagulant dose. However, this is expensive and time-consuming and does not enable responses to changes in raw water quality in real time. An optimal modeling can be used to overcome these limitations. In this paper, artificial neural network (ANN) and Adaptive Network- Based Fuzzy Inference System (ANFIS) models were used to model poly aluminum chloride (PAC) dosing of water treatment plant of Algeria.

The data used for the determination of the models include six (6) input variables describing the raw water characteristics (temperature, pH, turbidity, conductivity, dissolved oxygen and the ultraviolet absorption). As the selection and the optimization of the input number for a model requires specific analysis of the database used to design an adequate modelling of the process, a principal components analysis (PCA) has been conducted to reveal the contribution of each input-variable for understanding the phenomena subject to investigation, and even the determination of the coagulant dose for injection.

The coagulant models having various input structures are trained and tested to investigate the applicability of the used methods. To get more accurate evaluation of the ANFIS models results, the best fit model structures are also tested by artificial neural networks (ANN) and multiple linear regression (MLR) methods. The results of three methods are compared, and it is observed that the ANFIS is preferable and can be applied successfully because it provides high accuracy and reliability for coagulant dosage modelling.

**Keywords** : *Water treatment; Coagulant; Artificial neural network; neuro fuzzy systems; modelling; ANFIS; MLPNN; RBFNN; GRNN; MLR, PCA .*

## Liste des Publications & Communications

1. Heddami S., Dechemi N., (2008). Modeling of optimal coagulant dose using artificial neural network, application to water treatment plant of Boudouaou (Algeria). EJSR Vol 20, issue 3 pp 652–663.
2. Heddami S., Bermad A., Dechemi N., (2010). Modélisation de la dose de coagulant par les Réseaux de Neurones Artificiels (ANN) application à la station de traitement des eaux potables de Boudouaou (Algérie). 1er Colloque international de l'eau CIEAU 2010. Université Badji Mokhtar.
3. Heddami S., Bermad A., Dechemi N., (2011a). Modelling of Coagulant Dosage by Neuro Fuzzy Technique Application to Boudouaou Drinking Water Treatment Plant (Algeria). 4 ème colloque international « Ressources en eau et Développement Durable CIRED » février 2011/Alger.
4. Heddami S., Bermad A., Dechemi N. (2011b). ANFIS-based modelling for coagulant dosage in drinking water treatment plant: a case study. Environmental Monitoring and Assessment (EMAS). 184:1953–1971. Doi: 10.1007/s10661-011-2091-x.
5. Heddami S., Bermad A., Dechemi N. (2011c). Applications of Radial Basis Function and Generalized Regression Neural Networks for Modelling of Coagulant Dosage in a Drinking Water Treatment: A Comparative Study. ASCE Journal of Environmental Engineering. Vol. 137, No. 12, pp. 1209-1214 .Doi: 10.1061/ (ASCE) EE.1943-7870.0000435
6. Heddami S., Bermad A., Dechemi N. (2011d). Modélisation de la dose de coagulant par les systèmes à base d'inférence floue (ANFIS) application à la station de traitement des eaux de Boudouaou (Algérie). (RSE) Revue des sciences de l'eau. Revue des Sciences de L'eau 25(1) (2012) 1-19.
7. Heddami S., Bermad A., Dechemi N. (2011e). Le contrôle automatique du procédé de coagulation méthodes actuelles et perspectives d'amélioration : cas de la station de traitement des eaux potables de BOUDOUAOU Alger. Séminaire National sur l'eau, l'environnement et la biodiversité. Centre Universitaire Abbas Laghrour KHENCHELA Novembre 2011.
8. Heddami S., Bermad A., Dechemi N. (2012a) A Comparative Study of Artificial Intelligence Techniques for Coagulant Dosage Modelling: Case Study of Drinking Water Treatment Plant of Algeria. Troisième forum de l'eau JERBA le 26,27 et 28 mars 2012. Institut Supérieur des Sciences et Techniques des Eaux de Gabès & Tunisia.
9. Heddami S., Bermad A., Dechemi N. (2012b). A New Approach to Estimate Coagulant Dosage in Water Treatment Plant of Algeria by Adaptive Neuro-Fuzzy Inference System (ANFIS): A Comparative Study. Troisième forum de l'eau JERBA le 26,27 et 28 mars 2012. Institut Supérieur des Sciences et Techniques des Eaux de Gabès & Tunisia.

---

# Introduction Générale

L'évolution de la vie a été possible grâce à la présence de l'eau. Cependant, bien qu'existant en quantité abondante, seule une petite portion de cette eau est réellement disponible pour la consommation humaine ou animale. Les eaux salées des mers et des océans représentent près de 97% du volume total d'eau de notre planète, et une grande partie des 3% restants se trouve bloquée sous forme de neige, de glaciers ou d'eaux souterraines profondes. La quantité d'eau douce sous forme liquide représente environ 0,6% du volume total d'eau, dont seulement 0,014 % est facilement exploitable en tant qu'eaux de surface ou nappes phréatiques peu profondes. Cette ressource n'est d'ailleurs pas répartie équitablement sur la terre, presque la moitié des réserves d'eau de surface se situe en Amérique du Sud et un autre quart en Asie.

La source primaire d'eau douce provient des précipitations. Le total mondial des précipitations est d'environ 500000 km<sup>3</sup> par an (OMM, 1994), dont seul un cinquième de cette quantité retombe sur les terres. Environ 65 % des eaux de pluies continentales s'évaporent et retournent vers l'atmosphère. Le reste s'écoule dans les rivières et forme les lacs en surface ou s'écoule dans le sol, où elles se stockent éventuellement dans les nappes phréatiques. Ces réserves d'eau peuvent prendre plusieurs centaines, voire des milliers d'années pour se constituer.

Les eaux de surface : Qu'elles soient sous forme de lacs, de réservoirs d'eau ou de rivières, les eaux de surface sont facilement accessibles, mais leur qualité peut ne pas être appropriée ni à la consommation humaine, ni à l'agriculture, ni à l'industrie. Cela est dû au fait que l'eau de surface peut être polluée plus directement par des matières organiques décomposées d'origines humaine, animale et végétale ou par des matières minérales nocives en provenance de l'industrie, des mines et de l'agriculture. La pollution bactériologique (qui peut contenir des germes pathogènes) est souvent favorisée par les conditions particulières (humidité, éléments nutritifs, chaleur et énergie solaire) qui se retrouvent dans les eaux de surface. En plus, l'humanité a pollué et continue à polluer certaines réserves d'eau douce à cause du manque de planification et des pratiques irresponsables contribuant ainsi au déficit global en eau potable. Cette situation oblige le recours aux systèmes de traitement des eaux destinés à la consommation humaine, dans le but de rendre l'eau le plus potable possible.

Depuis les origines du traitement de l'eau destinée à la consommation, qui remontent au 19<sup>e</sup> siècle, les chercheurs travaillent sans cesse à développer des procédés et des stratégies visant à améliorer la qualité de l'eau potable distribuée à la population. Ainsi, au fil des ans, les recherches ont permis de mettre au point divers types de traitement de plus en plus efficaces et d'établir des normes relatives à la qualité de l'eau potable. De plus, ces recherches ont aussi permis d'identifier de nouvelles problématiques obligeant sans cesse à modifier les chaînes de traitement des usines de production d'eau potable en plus de revoir la réglementation sur l'eau potable. Deux objectifs principaux sont ciblés dans le développement de nouveaux systèmes de traitement, un objectif de qualité et un objectif de coût, avec une préoccupation générale de souplesse de procédé et de fiabilité.

Les gestionnaires des stations de traitement d'eau potable font face à deux défis importants. En premier lieu l'eau doit être distribuée en quantité suffisante et de manière continue afin de répondre aux besoins des consommateurs (résidentiels, commerciaux et industriels) ainsi qu'aux urgences (incendies). En second lieu. Cette eau doit en tout temps satisfaire aux normes de potabilité (normes chimiques et microbiologiques, qualité organoleptique) en vigueur.

L'atteinte du premier objectif implique la conception de stations de traitement relativement complexe, largement dimensionné et incluant un système étendu de mécanismes physico-chimique. Au passage de l'eau dans un tel système, la qualité de l'eau obtenue à l'usine peut être modifiée tant au niveau physique, chimique que microbiologique. La nature et l'étendue de ces changements dépendent des caractéristiques de l'eau à traiter (variables descriptives de l'eau brute) et du degré d'interaction avec les paramètres internes (produit chimiques injectés, temps de séjour, etc.). Le grand nombre de variables ainsi que leur nature dynamique complexifie grandement l'analyse de l'évolution de la qualité de l'eau dans une station de traitement des eaux potable.

En raison du grand nombre de variables, ainsi que la complexité des phénomènes biologiques, physiques et chimiques impliqués dans les procédés de traitement de l'eau potable, il est souvent très difficile de quantifier au préalable les interactions et les relations qui existent entre les entrées (variables) et les sorties (paramètres de qualité) de ces procédés. Parmi toutes les étapes de la chaîne de traitement des eaux, le procédé de coagulation-floculation est l'une des étapes les plus importantes, celles-ci mettent en œuvre des mécanismes fort complexes qui dépendent d'un nombre important de variables. La consommation d'agent coagulant fait de cette étape de traitement, l'opération la plus coûteuse dans la chaîne de traitement. La dose de coagulant est la variable principale utilisée pour conduire une unité de coagulation. Les coagulants peuvent être de nature organique ou inorganique. Les coagulants inorganiques les plus couramment utilisés sont les sels ferriques et l'alun. Récemment d'autres polymères inorganiques ont été introduits, soit les Polyaluminium-sulfates (PAS), les Polyaluminium -chlorures (PAC) et les Polyaluminium-silico-sulfates (PASS).

Dans une chaîne de traitement conventionnelle, la coagulation consiste à déstabiliser les particules, tandis que la floculation consiste à faire agglomérer les particules déstabilisées. Les particules présentes dans les eaux naturelles sont chargées négativement et sont entourées d'un nuage ionique. Cette charge électrostatique, ainsi que la couche ionique diffuse, constituent une barrière à l'agglomération de particules à cause de la répulsion électrostatique et parce que les forces attractives ne peuvent agir. Les principaux facteurs influençant l'efficacité de la coagulation – floculation sont la dose de coagulant, le pH, la turbidité initiale et la température de l'eau ainsi que certains d'autres variables descriptives de l'eau brute.

Lors de la coagulation, il faut chercher : (i) La maximisation de la déstabilisation des particules et des colloïdes organiques pour faciliter leur agglomération et leur enlèvement subséquent, par un procédé de séparation solide-liquide; (ii) La minimisation de la concentration en coagulant résiduel; (iii) La minimisation de la production des boues; (iv) La minimisation des coûts d'opération. La coagulation optimale correspond au dosage de coagulant et au pH qui assurent l'atteinte de tous ces objectifs.

Afin d'évaluer les conditions optimales de coagulation et de floculation, des essais dits de «Jar-Test» (JT) sont conduits à l'échelle de laboratoire. Pour des filières de traitement physico-chimiques conventionnelles, des JT menés dans une large gamme de conditions d'opérations permettent de déterminer le type de coagulant, son dosage, le pH et les



conditions d'agitation qui maximisent la réduction de la turbidité, et offre le PH optimum. Ce type d'approche (essai Jar-Test) a l'inconvénient de nécessiter un temps de réponse relativement long. En effet, on ne modifie la dose de coagulant qu'une fois un événement apparu, de plus, elle ne permet pas de suivre finement l'évolution de la qualité de l'eau brute. Étant donné l'importance d'une bonne coagulation dans une chaîne de traitement conventionnel, il est important d'avoir un contrôle très précis sur la dose de coagulant, on voit ici tout l'intérêt de disposer d'un contrôle automatique et efficace de ce procédé pour une meilleure efficacité de traitement et une réduction des coûts d'exploitation. L'objectif étant de déterminer la dose à appliquer en fonction des paramètres de l'eau brute observés.

L'information permettant le suivi et le contrôle automatique du procédé de coagulation est donnée par les mesures des variables de ce processus; pour cela on trouve dans la littérature que toutes les techniques proposées pour l'automatisation du phénomène incluent les variables descriptives de l'eau brute ainsi que la dose de coagulant. Afin de mieux comprendre le processus, une approche naturelle consiste à essayer d'en reproduire le comportement. Une telle démarche implique l'établissement d'un modèle du système étudié. Selon la connaissance des phénomènes physiques mis en jeu dont on dispose, plusieurs possibilités de modélisation peuvent se présenter. Les modèles de connaissance reposent sur une très bonne connaissance des lois physiques qui régissent le procédé étudié. A l'opposé, les modèles de comportement sont purement mathématiques et ne nécessitent aucune connaissance a priori sur la physique du système considéré. Le principe consiste donc à un développement d'un modèle mathématique reliant la dose de coagulant optimum aux différentes variables descriptives de l'eau brute.

Quelle que soit leur nature (physique, mathématique ou composition des deux), les modèles ont pour la plus grande partie un point commun : ils sont caractérisés par un certain nombre de variables que l'on nomme communément paramètres. De la valeur de ces paramètres dépend en grande partie la qualité du modèle et sa capacité à représenter le comportement du système réel sur un domaine de fonctionnement spécifique. Une fois définie la forme du modèle la plus apte à représenter correctement le fonctionnement d'un système réel, l'identification des paramètres du modèle consiste à calculer leur valeur optimale. Ces valeurs paramétriques optimales permettent au modèle de représenter fidèlement le système sur le domaine de fonctionnement choisi. En l'absence de connaissance du système, le choix de modélisation se porte naturellement sur un modèle de comportement et l'identification paramétrique ne peut être effectuée qu'à l'aide des seules données disponibles : des signaux d'entrée judicieusement choisis (variables d'entrée du modèle ou variables explicatives) et la réponse du système à ces signaux d'entrée (variable de sortie ou variable à expliquer).

Ces dernières années, des méthodes issues du "Soft Computing" ont été largement utilisées pour la modélisation des systèmes complexes. Dans ce sens les réseaux de neurones sont particulièrement exploités. Bien que l'apport de ces modèles par rapport aux modèles classiques linéaires à base de régression semble mitigé dans certains contextes, leur capacité à modéliser des relations non-linéaires et leur faculté d'apprentissage et d'adaptation en font des outils de modélisation très performants. Les systèmes d'inférence flous, pouvant modéliser les connaissances humaines et étant facilement interprétables, sont également employés dans le domaine de la modélisation: ils sont notamment très intéressants pour quantifier l'influence des variables explicatives utilisés pour la construction des modèles. Enfin, la combinaison de plusieurs techniques du "Soft Computing" (réseaux de neurones, logique floue) permet de réaliser des systèmes adaptatifs capables d'apprendre des relations complexes en milieu incertain. Cet ensemble

de caractéristiques nous a amené à traiter les contraintes spécifiques au domaine de traitement des eaux et en particulier à la modélisation de la dose de coagulant, en appliquant les techniques du "Soft Computing" à notre système.

L'objet de cette thèse est donc de développer un modèle mathématique pour l'optimisation de la dose de coagulant (CO) au niveau de la station de traitement des eaux de Boudouaou alimentant en eau potable la ville d'Alger, en fonction des variables descriptives caractérisant l'eau brut à savoir la température (TE), le pH, la turbidité (TU), la conductivité (CON), l'oxygène dissous (OX) et les ultraviolet (UV<sub>254</sub>). Notre approche sera basée d'une part, sur les concepts de la théorie de la logique floue et d'autre part, sur la modélisation à base de réseaux de neurones artificiels.

Cette thèse est organisée en six chapitres :

- Le premier chapitre présente les principes fondamentaux de traitement des eaux.
- Dans le deuxième chapitre est détaillé le procédé sur lequel a porté plus spécifiquement notre étude : le procédé de coagulation floculation. On s'intéresse plus particulièrement à l'aspect physique et chimique du procédé, et un état de l'art de la modélisation de la dose de coagulant est présenté.
- Le troisième chapitre traite l'aspect modélisation linéaire de la dose de coagulant, dans ce contexte l'analyse en composantes principales a été appliquée dans le but de détecter la présence des variables prépondérantes dans la détermination de la dose de coagulant.
- Le quatrième chapitre, propose une investigation systématique des réseaux de neurones artificiels, leurs principes, les différentes architectures, ainsi que leur application pour la modélisation de la dose de coagulant.
- Les notions de base de la logique floue et des systèmes neuroflous sont rappelées dans le cinquième chapitre afin de mieux comprendre l'intérêt d'utiliser la logique floue au sein d'un modèle mathématique. Notre travail traite plus spécifiquement du système ANFIS (Adaptive Neuro-Fuzzy Inference System). Le système ANFIS est une classe de réseau adaptatif. Il peut être vu comme un réseau de neurones non bouclé pour lequel chaque couche est un composant d'un système neuroflou et, à ce titre, c'est un "approximateur" universel.
- Dans le sixième chapitre nous présentons, les résultats de l'application des techniques de classification à notre station d'étude. Les cartes auto-organisatrices de Kohonen, qui appartiennent à la catégorie des réseaux de neurones à « apprentissage non supervisé », semblent particulièrement adaptées pour atteindre des objectifs de classification.
- Une conclusion générale reprenant les principaux résultats auxquels a abouti ce travail de recherche finalisera ce document.



# Chapitre I Procèdes Generaux de Traitement Des Eaux

## I.1. Introduction

Une installation de traitement d'eau destinée à la consommation doit produire, en toutes circonstances, une eau répondant aux normes (fixées dans chaque pays) alors que l'eau brute peut avoir, avant traitement, des caractéristiques variables, tout particulièrement si cette eau est une eau de surface,(Degrément, 2005). Pour faire face à ces variations, la conception de l'installation peut faire appel à plusieurs procédés qui vont être développés au niveau de ce chapitre.

Le stock d'eau douce utilisable comme ressource est engagé dans un mouvement cyclique permanent d'évaporation et de précipitations avec une partie intermédiaire, de longueur très variable, de cheminement à la surface ou au sein de la terre. Les terres sont donc lessivées systématiquement et les eaux brutes disponibles comprennent, en quantités et proportions variables suivant les lieux, tout le spectre des composants naturels de la croûte terrestre, des produits de la vie végétale et animale ainsi que des constituants des rejets de l'activité humaine. Toutes ces substances, comme les produits de leurs dégradations ou transformations ultérieures dans le milieu, se trouvent soit en suspension, soit sous forme colloïdale, soit encore en solution. Les teneurs de ces paramètres dans le milieu naturel sont souvent incompatibles avec la réglementation sanitaire et la perception des consommateurs.

Le traitement des eaux pour la distribution a pour objet :

- De réduire ces teneurs pour les rendre inférieures à des valeurs (ou pour les amener à l'intérieur d'une fourchette de valeurs) fixées par des normes qui sont en fait des critères de qualité ;
- D'assurer une sécurité parfaite par désinfection.

La détermination du traitement demande une étude analytique préalable tendant à la connaissance, d'une part de la nature et de la quantité des corps indésirables à éliminer et, d'autre part, de l'environnement physico-chimique qui peut influencer sur les méthodes à utiliser pour cette élimination(Godart, 2008).

La figure I.1 schématise les principales chaînes de traitement que l'on peut rencontrer, et qui doivent être complétées par des traitements spécifiques éventuellement rendus nécessaires par la présence d'un composé bien déterminé et indésirable (fluor, nitrate, calcium...)(Degrément, 2005).La ligne n° 1 concerne une eau claire et sans pollution, dont le seul traitement peut être une désinfection destinée à assurer la qualité microbiologique. La ligne n° 2 concerne une eau dont les seules impuretés sont des matières en suspension (MES). Une simple filtration avant désinfection est suffisante. Lorsque l'eau contient une faible quantité de colloïdes, ou présente une couleur plus importante, une coagulation sur filtre permet de résoudre le problème (ligne n° 3). Si la quantité de coagulant nécessaire à l'élimination des colloïdes ou de la couleur devient trop élevée le volume de flot formé devient très important, ce qui provoquerait un colmatage rapide du filtre et des lavages trop

fréquents; il est indispensable de prévoir un étage de séparation du flot formé (décantation ou flottation) avant filtration (ligne n° 4).

Le flot formé après l'addition de coagulant permet de bien clarifier l'eau. Ce flot a également des propriétés adsorbantes et certains polluants dissous peuvent s'adsorber à sa surface, mais si la concentration en matière organique (MO) polluantes est trop importante, il peut être indispensable de mettre en place des traitements complémentaires: oxydation (ligne n° 6) ou adsorption (ligne n° 5) qui peuvent se combiner avec l'un ou l'autre des traitements de clarification.

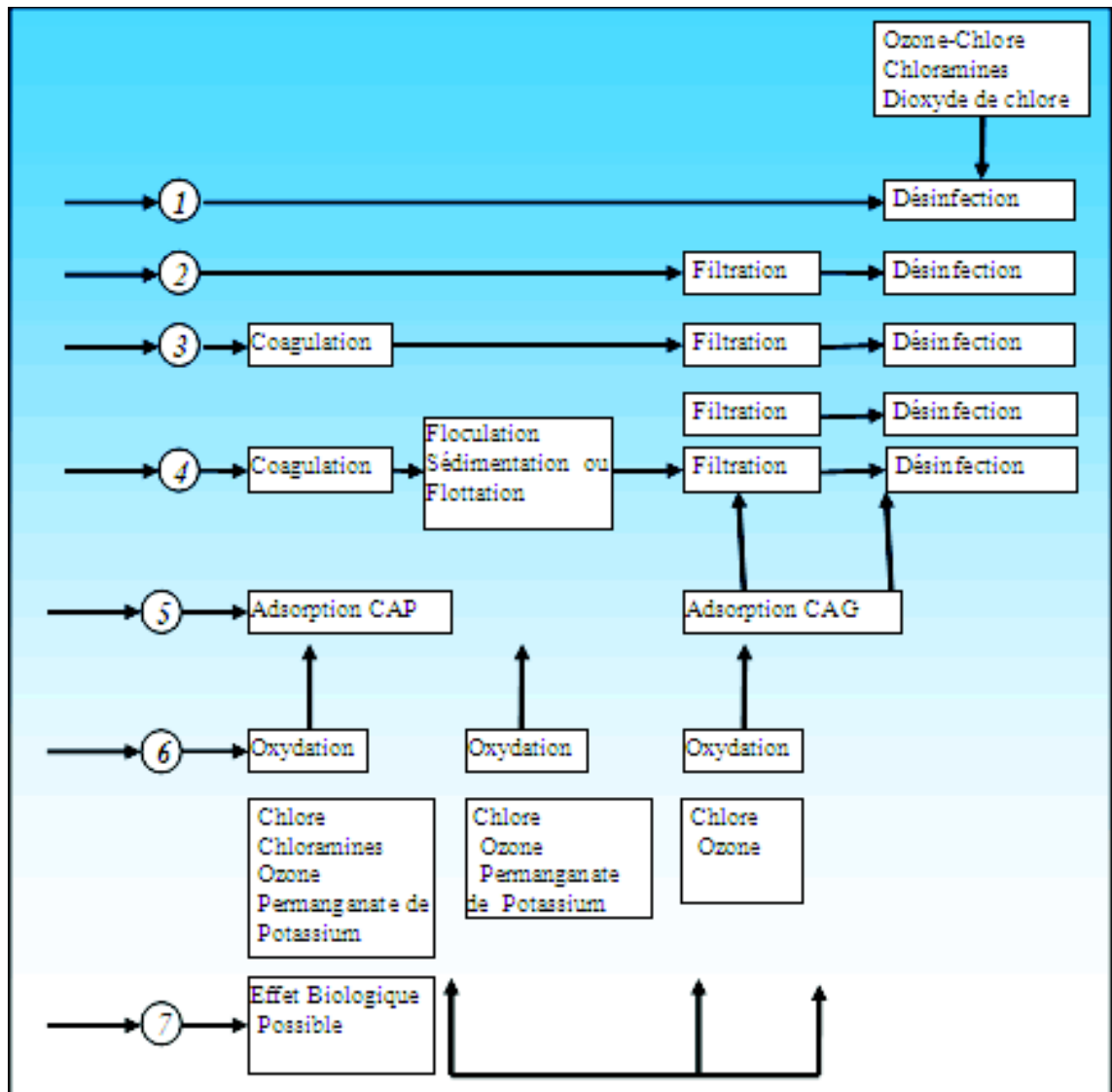


Figure I.1 Schéma des principaux traitements (Degrément, 2005)

## I.2. Qualité de la ressource

Il s'agit de cerner l'identité de la matière première soumise au traitement. Les eaux souterraines en un lieu donné ont des caractéristiques constantes ou ne variant que lentement ; cela n'est pas le cas pour les eaux stockées (lacs, barrages-réservoirs) ni surtout pour les eaux de surface courantes (cours d'eau). Dans ces deux derniers cas, des études longues (portant sur au moins un cycle hydrologique) sont indispensables pour optimiser la filière de traitement. L'étude préalable, en dehors de son aspect analytique, doit comporter simultanément des essais techniques permettant d'apprécier l'efficacité des processus unitaires envisageables.

### I.3. Méthodes générales de traitement

#### I.3.1. Généralités

---

La qualité des eaux brutes étant supposée connue à la suite des campagnes analytiques et des tests de laboratoire, il s'agira, de ramener, en permanence, les valeurs des paramètres de qualité au-dessous de la concentration maximale admissible, ou à l'intérieur des fourchettes fixées par la réglementation. La valeur des paramètres de départ (des eaux brutes) étant variable, surtout dans le cas des eaux superficielles, on pourra se fixer une hiérarchie des exigences : il serait inadmissible d'accepter volontairement la présence, ne serait-ce qu'une fois, de facteurs de maladies hydriques aiguës. Le dispositif de traitement doit assurer une correction respectant parfaitement la réglementation en période normale. Il doit de plus être conçu pour être renforcé en période de crise (statistiquement peu fréquente) ; certaines exigences pouvant dans ce cas, et pour de courtes périodes, être légèrement moins grandes (sauf pour les dangers aigus) (Godart, 2008).

#### I.3.2. Classement des traitements

---

On procède souvent au classement des traitements en processus unitaires en fonction de l'objectif principal visé par l'intervention de chacun d'eux, pris isolément : traitement de clarification, traitement de désinfection, etc. Il est maintenant bien établi que, chaque processus unitaire de traitement influera sur un grand nombre de paramètres de qualité et que les effets des différents processus unitaires utilisés dans une filière ne doivent pas être considérés séparément. La filière de traitement constitue un ensemble où chaque sous-ensemble amont intervient dans le rendement et les effets de tous les sous-ensembles situés en aval, cela implique en retour que les caractéristiques d'un sous-ensemble amont doivent être déterminées en tenant compte de tout ce qui se trouve en aval, il peut donc y avoir une véritable itération pour obtenir l'optimisation recherchée (Godart, 2008).

Le but principal de la filière est généralement l'élimination totale ou partielle de certains corps existant dans l'eau brute de façon à ramener les paramètres aux niveaux fixés par la réglementation. Quelquefois, il s'agira au contraire d'ajouter quelques corps, au départ en quantité insuffisante (reminéralisation d'une eau trop douce). En dehors de ce dernier cas, il s'agira, par conséquent, de concevoir une suite de processus conduisant :

- A une extraction, c'est-à-dire à un transfert de masse solide liquide, liquide liquide ou gaz liquide : dégrillage, tamisage, sédimentation, filtration, séparation par

membranes, flottation, adsorption, dégazage, échange ionique, transfert ionique, centrifugation ;

- A une modification d'état ou de structure: précipitation de colloïdes, transformations des ions en sels solides ou en gaz, désagrégation de molécules en molécules plus simples, formation de molécules plus complexes, destruction de la vie (micro-organismes), inactivation (virus).

## I.4. Procédés généraux

Que la ressource soit superficielle (rivière, lac naturel ou artificiel) ou souterraine, l'eau brute qu'elle fournit à l'utilisateur présentera dans le cas le plus général l'ensemble des caractéristiques défavorables suivantes :

- *Critères organoleptiques* : turbidité, couleur, goût, odeur ;
- *Critères chimiques*, correspondant à des constituants :
  - 1) Soit *naturellement présents dans l'eau* : sels minéraux (exemple : dureté, sulfates, chlorures...), fer, manganèse, ammonium, fluorures, arsenic, matières organiques (notamment les substances humiques responsables de la couleur)...
  - 2) Soit *apportés par la pollution* : micropolluants minéraux (métaux lourds, nitrates) ou organiques (pesticides, hydrocarbures, phénols, détergents...) ;
- *Critères biologiques* : il s'agit surtout des germes pathogènes apportés par la pollution fécale, mais aussi des organismes dont l'eau est l'habitat naturel (micro algues planctoniques ou phytoplancton, micro invertébrés ou zooplancton, bactéries de l'environnement).

Pour éliminer ou corriger tous ces paramètres défavorables, on dispose d'un certain nombre d'outils qui doivent d'abord être passés en revue : décanteurs ; flottateurs ; lits de matériaux granulaires à effet filtrant, adsorbant et/ou neutralisant ; réacteurs de floculation ou d'oxydation ; échangeurs d'ions ; membranes, etc. Ces outils seront combinés dans des filières plus ou moins complexes, suivant l'industrie considérée et l'usage de l'eau (Mouchet, 2008).

### I.4.1. Stockage d'eau brute

---

Le stockage d'eau brute présente un intérêt dans le cas de sécheresse prolongée (baisse du débit des rivières, avec souvent altération de la qualité de l'eau). Le volume d'eau brute emmagasinée doit permettre de répondre à la demande en eau de la période la plus longue envisagée. Le stockage présente également un intérêt lorsqu'il y a une pollution accidentelle qui détériore la qualité de l'eau brute de façon inacceptable pour l'installation de traitement. On peut arrêter le pompage en rivière pour utiliser l'eau stockée, en évitant ainsi l'arrêt de la production d'eau traitée. Le volume du stockage est alors déterminé en fonction des risques de pollution en amont de la prise d'eau, et de la durée maximale envisagée pour l'arrêt du pompage direct d'eau brute.

### I.4.2. Dégrossissage

---

Une eau, avant d'être traitée, doit être débarrassée de la plus grande quantité possible d'éléments dont la nature et la dimension constitueraient une gêne pour les traitements ultérieurs. Pour cela, on effectue des prétraitements de l'eau de surface (C.I.D.F, 1995).

Suivant la nature des eaux prélevées, le premier traitement possible est un dégrossissage, ayant pour but d'éliminer les matières de grandes dimensions susceptibles de gêner la mise en œuvre des autres traitements (Degrément, 2005), il peut comporter :

- 1) Un dégrillage ;
- 2) Un tamisage, aussi appelé macrotamisage, il est nécessaire si l'eau charrie des herbes, des feuilles, des débris plastiques... Le nettoyage doit être automatique. L'absence de cet appareil, par souci d'économie, est la cause de perturbations de bon nombre d'usines, surtout quand elles comportent un pompage;
- 3) Un dessablage, qui peut être placé suivant les conditions de prise, soit avant, soit après l'éventuel tamisage. Un tel dessablage est indispensable quand les appareils de traitement qui suivent sont susceptibles d'être fortement affectés par des quantités notables de sable ;
- 4) Un microtamisage, si la quantité de plancton est limitée et si aucune décantation n'est prévue ultérieurement. Étant donné l'action limitée des micro-tamis, il est rare que ce traitement soit bien adapté aux stations modernes,
- 5) Un déshuilage de surface ;
- 6) Un débouage, nécessaire quand la quantité de MES de l'eau brute à éliminer (limons, argile...), dépasse la capacité de concentration et d'extraction des décanteurs situés en aval.

### I.4.3. Prétraitement au chlore

---

La protection des conduites d'eau brute peut nécessiter un prétraitement par le chlore ou par l'un de ses composés (eau de javel, hypochlorite, dioxyde de chlore). Lorsqu'une eau riche en matière organique (MO) et en plancton doit être amenée à la station principale par une conduite de grande longueur, il est indispensable de prévoir au départ un traitement par un oxydant. Des bactéries ferrugineuses ou sulfato-réductrices peuvent attaquer le fer des conduites métalliques, on voit alors s'élever la teneur en fer de l'eau, surtout lors de périodes d'arrêt du traitement; une chloration permet de réduire ces inconvénients (Degrément, 2005).

### I.4.4. Traitement physique : Aération

---

Cette aération est nécessaire si l'eau présente un défaut d'oxygène; l'aération permet alors :

- L'oxydation des ions ferreux ;
- L'augmentation de la teneur en oxygène, afin de rendre l'eau agréable à boire, de lutter contre l'anaérobiose et d'éviter la corrosion de conduites métalliques en permettant la formation de la couche protectrice.

L'aération peut également être nécessaire si l'eau contient des gaz en excès :

- H<sub>2</sub>S, qui donne un goût très désagréable et qui s'élimine aisément par une simple aération atmosphérique ;

- Oxygène, quand l'eau est en état de sursaturation et que son dégagement est susceptible de troubler le fonctionnement des décanteurs, dont le floc a tendance à remonter en surface, et des filtres qui subissent un faux encrassement par dégazage au sein de la masse filtrante (embolie gazeuse) ;
- Dioxyde de carbone (CO<sub>2</sub>), rendant une eau agressive; son élimination par aération doit s'effectuer à la pression atmosphérique. Elle doit être plus ou moins poussée en fonction de la minéralisation de l'eau; en effet, il peut être nécessaire de ne procéder qu'à une élimination partielle du gaz carbonique, la portion restante pouvant servir à accroître la minéralisation de l'eau par réaction sur produits neutralisants (Degrément, 2005).

### I.4.5. Pré oxydation

---

On peut utiliser à cet effet le chlore et ses dérivés, l'ozone ou le permanganate de potassium.

#### I.4.5.1. Pré oxydation par le chlore

La préchloration avant clarification s'est surtout développée dans les années 60, ses avantages sont bien connus :

- Amélioration de la floculation, par action sur les matières organiques (MO) adsorbées et sur les MES. Il en résulte une augmentation du coefficient de cohésion de la boue, permettant des vitesses de décantation plus élevées ;
- Élimination de la plupart des algues et autres organismes (zooplancton, phytoplancton) susceptibles de se développer dans les décanteurs et les filtres, ce qui facilite l'exploitation (propreté des décanteurs, durée de vie des filtres). On évite ainsi la possibilité de fermentations anaérobies éventuelles ;
- Diminution de la couleur (lorsqu'elle est due à des matières humiques),
- Élimination de l'ammonium (si le taux appliqué est supérieur au point critique) et des nitrites ;
- Diminution éventuelle du taux de traitement en coagulant ;
- Élimination partielle des mauvais goûts ;
- Stabilisation de l'absorption de chlore en fonction du temps ;
- Plus grande sûreté dans la désinfection de l'eau distribuée.

#### I.4.5.2. Pré oxydation par les Chloramines

Si l'eau brute ne contient pas d'ammonium, on peut envisager d'injecter dans l'eau des Chloramines préalablement produites par action du chlore sur de l'ammoniaque ou du sulfate d'ammonium.

#### I.4.5.3. Pré oxydation par le dioxyde de chlore

Cette technique s'est développée momentanément pour tenter de remplacer le chlore en pré oxydation. En effet, le dioxyde de chlore, s'il ne permet pas d'oxyder l'ammonium, ne conduit pas non plus à la formation d'halo-formes. Mais la réaction d'oxydoréduction du dioxyde avec les MO libère des ions ClO<sub>2</sub> qu'il est nécessaire d'éliminer par la suite. L'usage de dioxyde de chlore en pré oxydation a donc tendance à décroître.

#### I.4.5.4. Pré oxydation par l'ozone

---

Utilisée avant filtration, cette technique peut être mise en œuvre pour l'élimination du fer et du manganèse. Elle peut également être employée pour l'amélioration de la clarification.

### **I.4.6. La clarification**

---

La clarification est l'élimination des matières en suspension, des matières colloïdales et des macromolécules susceptibles de communiquer à l'eau une turbidité ou une couleur indésirables. Il s'agit de matières organiques aussi bien que minérales, et les organismes vivants du phytoplancton (algues) et du zooplancton entrent également dans cette définition.

On peut distinguer deux catégories de procédés (Mouchet, 2008): les traitements physiques et les traitements physico-chimiques.

#### **I.4.6.1. Traitements physiques**

Ils sont appliqués lorsqu'il s'agit d'une simple séparation mécanique solide liquide, sans introduction de réactifs :

- Dégrillage (si possible, grille à nettoyage automatique) ;
- Tamisage (vide de maille de l'ordre de 1 mm) ;
- Micro tamisage (vide de maille de l'ordre de la dizaine de micromètres) ;
- Dessablage ;
- Débouage ;
- Déshuilage.

#### **I.4.6.2. Traitements physico-chimiques**

Ils sont nécessaires pour clarifier les eaux contenant des colloïdes, c'est-à-dire la quasi-totalité des eaux superficielles, certaines eaux de sources issues de terrains karstiques, certaines eaux profondes contenant des acides humiques en pseudo solution colloïdale, etc. La clarification se subdivisera en trois étapes, (Degrément, 2005):

- *A) Coagulation* : déstabilisation des particules colloïdales et formation de micro micelles constituées d'hydroxyde métallique et de colloïdes ;
- *B) Flocculation* : rassemblement des micros micelles sous forme séparable de la phase aqueuse ;
- *C) Séparation* des phases solide et liquide :

Décantation ou flocculation → filtration

### **I.4.7. Coagulation flocculation**

---

Dans le cadre de ce travail de recherche nous avons consacré le chapitre 2 pour traiter en détail la coagulation flocculation.

### **I.4.8. La décantation flottation**

---

#### **I.4.8.1. Décantation**

La décantation procédé qu'on utilise dans pratiquement toutes les usines d'épuration et de traitement des eaux, a pour but d'éliminer les particules en suspension dont la densité est



supérieure à celle de l'eau. Ces particules sont en général des particules de floc ou des particules résultant de la précipitation qui a lieu lors des traitements d'adoucissement ou d'élimination du fer et du manganèse. Les particules s'accumulent au fond du bassin de décantation d'où on les extrait périodiquement, l'eau clarifiée, située près de la surface est dirigée vers l'unité de filtration(Desjardins, 1997).

On peut distinguer trois types d'appareils :

1- Ceux où la séparation est acquise par chute des agglomérats jusqu'au contact d'une surface fixe ;

2- Ceux où la séparation s'obtient par contact avec un lit fluidisé de boues préformées ;

3- Et, dans les appareillages les plus récents, ceux où la décantation est accélérée par un lest de micro sable (Godart, 2008).

Une fois le floc formé, il faut réaliser la séparation solide-liquide qui permettra d'obtenir l'eau clarifiée d'un côté et les boues de l'autre. On utilise en général l'action de la pesanteur sur les particules en suspension et l'on recueille ensuite l'eau claire à la partie supérieure de l'appareil, alors qu'on soutire les boues sédimentées dans la partie inférieure : c'est la *décantation*. Dans certains cas particuliers, où le floc est léger, on réalise une *flottation* et la situation est alors inversée : les boues sont évacuées sous forme d'écume à la partie supérieure et c'est le sous nageant, recueilli par ramifications perforées ou cloison siphonide, qui donnera l'eau traitée.

### **1.4.8.2. Problèmes de décantation**

Les plus fréquents peuvent être :

1. Présence de sable: un dessablage préliminaire évitera l'envasement du décanteur ;
2. Fortes pointes de matières en suspension: suivant le cas, on choisira un appareil statique raclé de préférence à un décanteur accéléré, ou bien on placera un débourbeur (pré décanteur spécialement conçu à cet effet, précédé ou non d'une injection de coagulant) en amont d'un appareil accéléré à contact de boues ;
3. Variations brutales de température: du fait des courants de convection qu'elles engendrent, elles provoquent des départs de boues si l'eau n'est pas parfaitement répartie dans l'appareil (dans une usine, il faut donc éviter d'utiliser l'eau brute comme échangeur de chaleur avant une phase de décantation, surtout si cet échange thermique présente des paramètres variables) ;
4. Abondance d'algues: les bulles gazeuses qu'elles émettent provoquent des remontées de boues ; une préchloration, une pré-ozonation et/ou un traitement simultané au charbon actif en poudre entravent ces phénomènes (Mouchet, 2008).

### **1.4.8.3. Types de décantation**

Selon la concentration en solides et la nature des particules (densité et forme), on distingue quatre types de décantation (Desjardins, 1997): la décantation des particules discrètes, la décantation des particules floculante, la décantation freinée et la décantation en compression de boue.

· TYPE I : décantation de particules discrètes

Ce type de décantation (dite discrète ou individuelle) est caractérisé par le fait que les particules conservent leurs propriétés physiques initiales (forme, dimension et densité) au cours de leur chute. La vitesse de chute est indépendante de la concentration en solides;



c'est pourquoi les lois de Newton et de Stokes s'appliquent ici, on retrouve ce type de décantation dans les dessableurs.

- TYPE II : décantation de particules floculante

Ce type de décantation est caractérisé par l'agglomération des particules au cours de leur chute. Les propriétés physiques de ces particules (forme, dimension, densité et vitesse de chute) sont donc modifiées pendant le processus, on retrouve ce type de décantation dans les décanteurs des usines de traitement des eaux et les décanteurs secondaires des usines d'épuration.

- TYPE III : décantation freinée

Ce type de décantation est caractérisé par une concentration élevée de particules, ce qui entraîne la formation d'une couche de particules et par conséquent l'apparition d'une démarcation nette entre les solides décantés et le liquide surnageant, on retrouve ce type de décantation dans la partie profonde des décanteurs.

- TYPE IV : décantation en compression de boue

Dans ce type de décantation, les particules entrent en contact les unes avec les autres et reposent sur les couches inférieures. Dans cette zone, le phénomène de consolidation est relativement lent, on retrouve ce type de décantation dans les épaisseurs de boue par gravité.

### **I.4.8.4. Flottation**

La flottation est en quelque sorte une décantation à l'envers. Le procédé consiste à alléger à l'aide de très fines bulles d'air un floc naturellement léger et fragile. Le floc dont la masse volumique est alors plus faible que celle de l'eau, est entraîné vers la surface de l'ouvrage où il s'accumule, l'eau clarifiée est collectée en partie basse du flottateur, les boues de surface sont raclées vers une goulotte d'évacuation (Godart, 2008).

La flottation est à recommander pour :

- Les eaux peu chargées en MES (eaux de lac, de barrage, par exemple),
- Les eaux produisant un flot léger, décantant mal,
- Les eaux riches en plancton qui, en été, par suite de l'activité chlorophyllienne des algues, ont tendance à être saturées en oxygène, ce qui contrarie la décantation et provoque une remontée du flot, dans le cas de l'utilisation de la décantation, la flottation possède, en outre, les avantages suivants:
- Grande souplesse d'emploi: le démarrage de l'installation est très rapide, pratiquement instantané.
- Économie de réactifs de coagulation on observe une économie qui peut aller jusqu'à 30 % par rapport aux doses de coagulant utilisées en décantation (Degrément, 2005).
- Certaines eaux (lacs, barrages en particulier) sont relativement claires, mais colorées ou riches en algues : elles réclament alors de fortes doses de coagulant sans que le floc puisse être alourdi par des matières initialement en suspension dans l'eau brute ; sur de telles eaux, la décantation est donc souvent difficile (à moins de jouer sur les forces de cohésion qui s'exercent au sein d'un appareil à lit de boues comme le Pulsator).

Il existe divers types de flottateurs fonctionnant, suivant les modèles, à des vitesses de séparation de l'ordre de  $7 \text{ à } 10 \text{ m}^3 \cdot \text{h}^{-1} \cdot \text{m}^{-2}$ . Pour traiter les eaux naturelles, les plus

courants sont rectangulaires et forment un ensemble compact avec les flocculateurs et les filtres. En cas de besoin, il est possible de choisir un appareil circulaire avec raclage de fond pour répondre aussi à des périodes de turbidité notable, il est également possible de construire des flottateurs filtres combinés, intégrant la filtration au-dessous de la zone de flottation.

### I.4.9. Filtration

---

#### I.4.9.1. Généralités

Dans la recherche d'une clarification, la filtration est le processus de finition par excellence. Parallèlement à ses propriétés de fixation des particules en suspension, la filtration peut agir par adsorption et être le siège de processus biologiques. Un filtre est composé d'un lit de matières granuleuses qui est traversé par l'eau à traiter à des vitesses variables de  $2 \text{ m}^3 \cdot \text{m}^{-2} \cdot \text{j}^{-1}$  à  $20 \text{ m}^3 \cdot \text{m}^{-2} \cdot \text{h}^{-1}$  ( $480 \text{ m}^3 \cdot \text{m}^{-2} \cdot \text{j}^{-1}$ ) ; la hauteur du lit granuleux, la granulométrie et la nature du matériau sont adaptées aux conditions d'emploi (Godart, 2008).

Que l'eau soit coagulée ou non, décantée (ou « flottée ») ou non, il faut toujours la filtrer pour que l'élimination des matières insolubles soit aussi complète que possible (c'est-à-dire pour que la turbidité soit ramenée au voisinage de zéro). La filtration est également indispensable en déferrisation, en déminéralisation, en polissage de décarbonatation, en aval des réacteurs biologiques d'élimination de l'ammonium ou des nitrates, etc.

C'est surtout la *filtration en profondeur*, sur *matériau granulaire*, qui constitue le cœur de la majorité des stations de préparation d'eau industrielle ; cependant, on trouve aussi divers procédés où l'épaisseur du milieu filtrant est négligeable ou très faible devant sa surface et où les impuretés sont retenues sous forme d'un *gâteau* ne pénétrant pratiquement pas dans ce milieu :

- *Filtration sur support*  $t$  : filtre à cartouches, à bougies, à précouche ;
- *Filtration sur membranes de microfiltration et d'ultrafiltration.*

La filtration est donc une opération cyclique, alternant (ou combinant) des phases (ou cycles) de production par filtration et des phases de lavage. Ce dernier doit intervenir : soit lorsque la perte de charge atteint une valeur fixée par construction ou par sécurité ; soit lorsque la qualité du filtrat commence à se détériorer (par exemple, augmentation de la turbidité : on dit alors que le filtre « crève ») (Mouchet, 2008).

#### I.4.9.2. Buts de la filtration

La filtration est un procédé physique destiné à clarifier un liquide qui contient des matières solides en suspension en le faisant passer à travers un milieu poreux. Les solides en suspension ainsi retenus par le milieu poreux s'y accumulent : il faut donc nettoyer ce milieu de façon continue ou de façon intermittente. La filtration habituellement précédée des traitements de coagulation, de floculation et de décantation, permet d'obtenir une bonne élimination des bactéries, de la couleur, de la turbidité et indirectement de certains goûts et odeurs (Desjardins, 1997).

#### I.4.9.3. Matériaux des filtres

Selon le type de filtre adopté, on a recourt à divers matériaux filtrants on utilise ainsi :

- A) Des tissus de fibres des toiles métalliques ou des pierres poreuses à interstices très fins; Ces matériaux retiennent les solides en surface, on les utilise rarement pour traiter des quantités d'eau importantes.
- B) Des granulats libre qui n'adhèrent pas les unes aux autres, ces matériaux sont insolubles et inattaquables par le liquide filtré ou par les solides qui s'y déposent. La filtration a lieu soit en surface, soit en profondeur, selon les caractéristiques granulométriques du matériau filtrant et selon la grosseur et la cohésion des solides en suspension, le sable, l'antracite et l'ilménite sont couramment utilisés dans les usines de traitement des eaux.

### **I.4.9.4.Types de filtres**

Pour le traitement des eaux potables, on utilise principalement :

- Des filtres à sable rapides ;
- Des filtres à sable lents ;
- Des filtres sous pression ;
- Des filtres à terre diatomée.

En pratique, les filtres à sable rapides sont les plus utilisés. Les filtres à sable lent, de construction et de fonctionnement simples, nécessitent de grandes superficies. Les filtres sous pression et les filtres a terre diatomée servent surtout à traiter les eaux de piscine ou sont utilisés dans de petites unités de traitement préfabriquées(Desjardins, 1997).

### **I.4.9.5.Caractéristiques des matériaux filtrants**

Les principales caractéristiques d'un matériau filtrant sont : le diamètre effectif, le coefficient d'uniformité, la densité relative, la masse unitaire sèche (maximale et minimale) et la porosité (maximale et minimale). Il existe d'autres caractéristiques beaucoup plus difficiles à mesurer, comme la forme des grains et la surface spécifique(Desjardins, 1997).

### **I.4.9.6.Écoulement de l'eau dans un filtre à sable**

La perte de charge à travers un milieu filtrant est une caractéristique importante de la filtration,de l'ordre de 0.5 à 0.75 m lorsque le filtre est propre, elle augmente au fur et à mesure que le filtre s'encrasse. L'équation de Carman-Kozeny ou l'équation de Rose permettent de calculer la perte de charge dans un filtre propre lorsque l'écoulement a lieu par gravité. Nous considérons ici trois cas différents, selon que le milieu filtrant est :

- A) Composé de particules de même diamètre.
- B)Homogène et composé de particules de diamètre varié.
- C) Stratifié.

### **I.4.9.7.Filtre à sable rapide**

Le filtre à sable rapide est le type de filtre le plus utilisé dans le traitement des eaux de consommation. Le matériau filtrant est maintenu en place par gravité et l'écoulement de l'eau à lieu de haut en bas. Lorsque le milieu filtrant est encrassé, on lave le filtre en inversant le sens de l'écoulement de l'eau : le milieu filtrant est alors en expansion et les particules d'impuretés beaucoup moins dense que les grains de sable, sont décollées et évacuées vers l'égout à l'aide des goulottes de lavage. Les principaux éléments d'un filtre rapide sont : le fond de filtre, gravité de support et le milieu filtrant.

- **Fond de filtre** : Le fond de filtre est la structure qui sépare le milieu filtrant de l'eau filtrée. Il doit donc être suffisamment solide pour supporter le milieu filtrant (1m de sable et de gravier) et l'eau située au-dessus du milieu filtrant (2m d'eau), il permet en outre :
  - De collecter et d'évacuer les eaux filtrées ;
  - De distribuer uniformément l'eau de lavage.
- **Gravier de support** : La gravité de support situé immédiatement au-dessus du fond de filtre, permet :
  - de retenir le sable du milieu filtrant.
  - D'améliorer la répartition de l'eau de lavage dans le filtre.

L'épaisseur et les caractéristiques granulométriques de la couche de gravier de support dépendent des caractéristiques du fond de filtre de celle du milieu filtrant

### I.4.9.8. Filtration lente

C'est la première application historique, On cherche à y reproduire les phénomènes spontanés de la filtration naturelle par le sol. Les filtres lents doivent subir une période préalable de maturation au cours de laquelle se forme une membrane superficielle, dite membrane biologique, où se développent des colonies de micro-organismes. L'action du filtre lent est double, on y obtient une clarification mécanique ainsi qu'une purification par dégradation biologique accompagnée d'une adsorption par le complexe biologique de la membrane. Les filtres lents (quelques  $m^3 / m^2 / j^1$ ) donnent d'excellents résultats dans les cas où la pollution est naturelle (non toxique pour les micro-organismes) et la quantité de matières en suspension faible (Godart, 2008).

### I.4.9.9. Filtration rapide

La filtration à travers un milieu granuleux (de l'ordre de quelques  $m^3 / (m^2 \cdot h)$ ) présente un double phénomène de blocage dimensionnel et d'adhésion à la surface des grains du lit filtrant, qui est très différent de l'arrêt provoqué par un tamis, par exemple, où le vide des mailles est la caractéristique essentielle. Dans la filtration sur grains, le pouvoir d'arrêt est ainsi commandé par la surface totale des grains du lit qui elle-même, pour un lit donné, est une fonction inverse de la dimension caractéristique du grain moyen. Le phénomène de filtration est varié en raison du colmatage inhérent à l'accumulation des impuretés sur la surface des grains qui réduit la dimension des canalicules ; il peut se produire également une détérioration du filtrat en raison de l'érosion des boues arrêtées, consécutive à l'accroissement de vitesse dans les canalicules.

### I.4.9.10. Types de filtres américains et européens

Cette latitude théorique dans le choix des caractéristiques d'un filtre est tempérée par l'expérience. En fait, les eaux admises sur les filtres, qu'elles soient brutes et légèrement coagulées ou au contraire prétraitées par coagulation et sédimentation, restent peu chargées en matières en suspension et, au fil du temps, deux types de filtres ouverts se sont trouvés définis :

- 1) Type américain de granulométrie fine (0.5 mm) et faible épaisseur (0.5 m) ;
- 2) Type européen de granulométrie moyenne (0.8 à 1.5 mm) et d'épaisseur plus élevée (0.9 à 1.2 m).

Pour ces deux types, la vitesse de filtration optimale est de l'ordre de 5 à 7 m/h. La séquence de lavage doit être adaptée au type de filtre, au matériau filtrant, et à la nature de l'eau à traiter, le lavage doit décolmater à cœur, sans entraîner de perte de matériau, (Godart, 2008).

### I.4.9.11. Paramètres fondamentaux de la filtration en profondeur

Il existe une grande variété de filtres, que l'on peut classer selon les paramètres énoncés dans ce qui suit.

- **C omposition du lit filtrant :** à savoir la nature des matériaux ; la granulométrie et la hauteur de couche. Le sable représente le matériau traditionnel et encore le plus utilisé, d'autres matériaux, combinés ou non avec le sable, peuvent entrer dans la composition d'un filtre. La granulométrie du ou des matériaux (x) filtrant(s) est caractérisée par les paramètres suivants : *taille effective* et le *coefficient d'uniformité*. La hauteur de couche d'un lit filtrant se situe entre 0.8 et 2 m en général ; le plus souvent, elle est de l'ordre de 1 m.
- **C aractéristiques de la filtration :** les sens de la filtration ; la vitesse de filtration ; lavage du filtre et le cycle de filtration. Suivant le cas, le sens de filtration se fait : de haut en bas (classique) ; de bas en haut ; dans les deux sens avec un bifiltre (distribution simultanément à la partie supérieure et à la partie inférieure ; reprise de l'eau par un collecteur au milieu de la masse filtrante); par filtration horizontale (type de filtre plus rustique). La vitesse de filtration s'exprime en  $m^3/h/m^2$ , suivant le domaine d'application, on peut appliquer des vitesses de filtration comprises entre 5 et  $50 m^3/h/m^2$ . Pour le lavage des filtres La *perte de charge* maximale à laquelle il faut laver le filtre est, en général, 1 à 2 m de colonne d'eau pour les filtres gravitaires, 0.5 à 0.8 bars (5 à 8 m de colonne d'eau) pour les filtres sous pression. La durée des cycles de filtration entre deux lavages dépend de la nature et de la concentration des matières en suspension (MES) arrivant sur le filtre, de la vitesse de filtration, des caractéristiques du lit filtrant, de la température de l'eau, de la perte de charge admissible et de la garantie donnée pour la qualité du filtrat.

## I.4.10. Désinfection

---

### I.4.10.1. Introduction

La désinfection est un traitement qui permet de détruire ou d'éliminer les microorganismes susceptibles de transmettre des maladies ; ce traitement n'inclut pas nécessairement la stérilisation, qui est la destruction de tous les organismes vivants dans un milieu donné.

On peut procéder à la désinfection en ajoutant à l'eau une certaine quantité d'un produit chimique doté de propriétés germicides. Les produits chimiques les plus utilisées sont: le chlore, le dioxyde de chlore, l'ozone, le brome, l'iode et le permanganate de potassium. On peut également désinfecter l'eau grâce à des moyens physiques : ébullition, ultrasons, ultraviolets ou rayons gamma (Desjardins, 1997).

### I.4.10.2. Critères permettant de choisir le désinfectant

Tous les procédés et les produits de désinfection n'étant pas équivalents, il faut choisir le procédé le plus approprié, compte tenu de certaines conditions particulières

(caractéristiques et usages de l'eau, type de microorganismes à éliminer, etc.) et sachant qu'un désinfectant ou un procédé de désinfection doit:

- Etre non toxique pour les humains ou les animaux ;
- Etre toxique, à de faibles concentrations pour les microorganismes ;
- Etre soluble dans l'eau ;
- Former avec l'eau une solution homogène ;
- Etre efficace aux températures normales de l'eau de consommation (de 0 à 25 °C) ;
- Etre stable afin de favoriser le maintien d'une certaine concentration résiduelle pendant une longue période;
- Ne pas réagir avec la matière organique autre que celle des microorganismes ;
- Ne pas détériorer les métaux ni endommager les vêtements lors de la lessive ;
- Eliminer les odeurs ;
- Exister en grande quantité, et être vendu à un prix abordable ;
- Etre facile à manipuler, et ne faire courir aucun danger aux opérateurs ;
- Permettre une mesure aisée de sa concentration, et un bon contrôle de celle-ci.

On effectue le plus souvent la désinfection à l'aide de chlore (80 % de la désinfection dans le monde), car ce désinfectant présente plusieurs des avantages énoncés ci-dessus. Cependant, l'addition de ce produit peut entraîner des effets secondaires indésirables qui, dans certains cas, obligent à utiliser d'autres désinfectants. Ainsi, le chlore réagit avec la matière organique de l'eau, ce qui peut parfois entraîner la formation de substances cancérigènes (trihalométhanes) ou d'odeurs désagréables (chlorophénols). Pas ailleurs, le chlore n'est pas suffisamment puissant pour éliminer complètement certains microorganismes très résistants comme les virus et les protozoaires. Afin de pallier à ces carences, on utilise le dioxyde de chlore ou l'ozone, ces désinfectants, beaucoup plus puissant que le chlore, ont toutefois l'inconvénient d'être instable (par exemple, l'ozone réagissant très vite dans l'eau, on ne peut maintenir une concentration résiduelle pendant longtemps), c'est pourquoi on doit les produire à l'usine de traitement des eaux.

### **1.4.10.3. Principes généraux de la désinfection**

Le taux de destruction des microorganismes par un désinfectant est fonction de plusieurs variables : puissance et concentration de désinfectant, temps de contact, nombre de microorganisme à éliminer, type de microorganisme, température de l'eau, pH de l'eau et concentration de matières organiques dans l'eau. Il n'existe actuellement aucune loi mathématique connue qui permette de calculer le taux de destruction des microorganismes en fonction de ces variables. Cependant, sur la base de mesures effectuées en laboratoire, on peut énoncer certaines règles particulières, avant d'appliquer ces règles, il faut toutefois procéder à des vérifications en laboratoire concernant l'eau à traiter afin de déterminer quelle règle s'applique et d'évaluer les constantes requises (Desjardins, 1997)..

### **1.4.10.4. Désinfection par le chlore**

Les produits chimiques les plus utilisés pour obtenir une désinfection des eaux par le chlore sont : le chlore gazeux ( $\text{Cl}_2$ ), les hypochlorites de sodium ( $\text{NaOCl}$ ), les hypochlorites de calcium ( $\text{Ca}(\text{OCl})_2$ ), les monochloramines ( $\text{NH}_2\text{Cl}$ ) et le dioxyde de chlore ( $\text{ClO}_2$ ), parmi ces produits c'est le chlore gazeux auquel on recourt le plus pour la désinfection des eaux



potables. Dans les petites usines de traitement, on utilise souvent des hypochlorites car ces produits sont faciles à manipuler, et font courir moins de danger aux opérateurs (Desjardins, 1997).

#### **I.4.10.5. Désinfection par le dioxyde de chlore**

Jusqu'en 1974, on utilisait le dioxyde de chlore principalement à cause de son pouvoir oxydant élevé, grâce auquel on pouvait éliminer les goûts, les odeurs, le fer et le manganèse présents dans l'eau. Pour une désinfection continue, on utilisait plutôt le chlore, moins coûteux. Or, à partir de 1974, on a remarqué que, dans certains cas, le chlore réagissait avec la matière organique pour former des substances cancérigènes (trihalométhanes), ce qui n'est pas le cas avec le dioxyde de chlore. C'est pourquoi on utilise de plus en plus le dioxyde de chlore et le chlore. Ainsi, on traite d'abord avec du dioxyde de chlore les eaux qui contiennent beaucoup de matières organiques, ce qui permet de désinfecter et d'oxyder cette dernière, ainsi que d'éliminer une bonne partie des goûts et des odeurs, et ce sans générer de produits cancérigènes. Ensuite, lorsqu'une fraction importante de la matière organique est oxydée ou éliminée, on effectue une post-chloration qui permet de maintenir une concentration résiduelle de chlore dans l'eau distribuée.

#### **I.4.10.6. Désinfection par l'ozone**

L'ozone est un gaz instable composé de molécules d'oxygène triatomique  $O_2$ . On produit l'ozone grâce à l'émission d'un effluve électrique sous haute tension dans une atmosphère contenant de l'air sec. Dans la production de l'ozone, les coûts imputables à l'énergie électrique utilisée, et à l'entretien sont très importants. En Amérique du Nord, jusqu'à récemment, grâce à la qualité relativement élevée des eaux brutes, on a pu obtenir une eau traitée de bonne qualité en n'utilisant que des produits de désinfection moins coûteux que l'ozone (chlore, hypochlorites, etc.). Toutefois, depuis le début des années 1970, la baisse de la qualité des eaux brutes, causée par les déversements de plus en plus importants d'eaux usées municipales, industrielle et agricoles, incite les spécialistes à préconiser l'utilisation de l'ozone à cause des propriétés oxydantes et désinfectantes (Desjardins, 1997).

## **I.5. Conclusion**

À travers ce premier chapitre nous avons présentés la filière de traitement la plus utilisée, on procédait souvent au classement des traitements en processus unitaires en fonction de l'objectif principal visé par l'intervention de chacun d'eux, pris isolément : traitement de clarification, traitement de désinfection, etc.

L'objet de la filière est généralement l'élimination totale ou partielle de certains corps existant dans l'eau brute de façon à ramener les paramètres au niveau fixés par la réglementation. Les processus pourront être de nature physique, physico-chimique ou biologique. Le chapitre suivant fera l'objet d'une étude détaillée de l'étape la plus importante et qui est l'objet de notre travail de recherche : la coagulation floculation.

# Chapitre II Coagulation & Floculation

## II.1. Introduction

Avant d'appliquer tout autre traitement, y compris ceux de désinfection qui sont communs à toutes les distributions publiques d'eau de consommation humaine, il faut rendre l'eau apte à recevoir efficacement ces traitements. Cela suppose, en particulier, une eau débarrassée de ses impuretés macroscopiques ou microscopiques : c'est le domaine de la coagulation floculation.

La coagulation est l'une des étapes les plus importantes dans le traitement des eaux de surface. La difficulté principale est de déterminer la quantité optimale de réactif à injecter en fonction des caractéristiques de l'eau brute. Un mauvais contrôle de ce procédé peut entraîner une augmentation importante des coûts de fonctionnement et le non-respect des objectifs de qualité en sortie. Cette opération a également une grande influence sur les opérations de décantation et de filtration ultérieures. En revanche, un contrôle efficace peut réduire les coûts de main d'œuvre et de réactifs et améliorer la conformité de la qualité de l'eau traitée.

Le contrôle de cette opération est donc essentiel pour trois raisons : la maîtrise de la qualité de l'eau traitée en sortie (diminution de la turbidité), le contrôle du coagulant résiduel en sortie (réglementation de plus en plus stricte de la présence de coagulant résiduel dans l'eau traitée), et la diminution des contraintes et des coûts de fonctionnement (Hernandez, 2006).

Dans ce chapitre, on va étudier l'aspect physico chimique de la coagulation et floculation, la méthode de détermination de la dose de coagulant, et on termine par un état de l'art sur la modélisation de la dose de coagulant et les méthodes qu'on propose dans le cadre de cette thèse pour estimer efficacement la dose de coagulant.

## II.2. Coagulation

### II.2.1. Principes de base

---

Le mot coagulation vient du latin coagulare qui signifie agglomérer (Masschelein, 1999). La couleur et la turbidité d'une eau de surface sont dues à la présence de particules de très faible diamètre (les colloïdes), leur élimination ne peut se baser sur la simple décantation. En effet, leur vitesse de sédimentation est extrêmement faible. La coagulation et la floculation sont les processus qui permettent l'élimination des colloïdes, la coagulation consiste à les déstabiliser, il s'agit de neutraliser leurs charges électrostatiques de répulsion pour permettre leur rencontre, la floculation rend compte de leur agglomération en agrégats éliminés par décantation et /ou filtration (Cardot, 1999).

L'opération de coagulation-floculation a donc pour but la croissance des particules (qui sont essentiellement colloïdales) par déstabilisation des particules en suspension puis



formation de flocons par absorption et agrégation (Valiron ,1989). Les flocons ainsi formés seront décantés et filtrés par la suite.

La difficulté de séparation des colloïdes présents dans une eau par simple décantation ou filtration provient du fait que ces particules, qui sont d'origine minérale (argiles), végétale (acides humiques ou fulviques, algues), animale (zooplancton), humaine (suspensions graisseuses, rejets ERU) ou industrielle (eaux résiduaires diverses), se trouvent dans l'eau à l'état de particules chargées électriquement (toujours négativement dans les eaux naturelles). La répulsion mutuelle exercée par ces particules les empêche de s'agglutiner sous l'effet des forces d'attraction naturelle (appelées à cette échelle forces de Van der Waals) et d'acquérir ainsi une dimension suffisante pour devenir décantables ou filtrables (Figure II.1), on dit alors que la suspension est stable (Mouchet, 2008).

La coagulation consiste à ajouter à l'eau un électrolyte permettant de neutraliser les charges négatives qui sont à l'origine du maintien en suspension stable (Lind, 1995), on utilise généralement des sels d'un métal trivalent,  $Fe^{3+}$  ou  $Al^{3+}$  (Figure II.1).

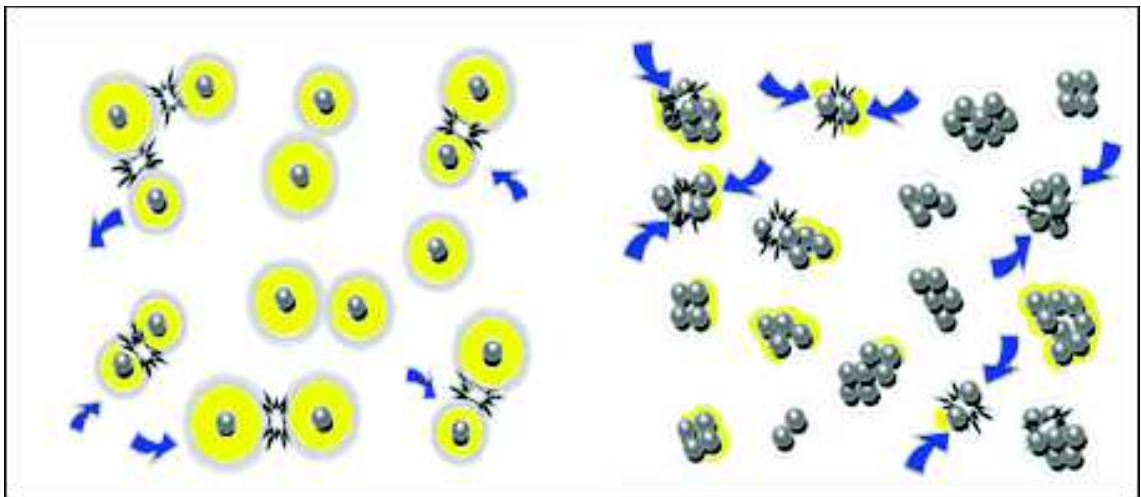


Figure II.1. Eau brute additionnée de coagulant

## II.2.2. Les particules mises en jeu : les colloïdes

### II.2.2.1. Généralités

Les matières existantes dans l'eau peuvent se présenter sous les trois états suivants:

- Etat de suspension qui regroupe les plus grosses particules ;
- Etat colloïdal ;
- Etat dissous des sels minéraux et des molécules organiques (Cardot ,1999).

Cette classification résulte de la taille des particules, les colloïdes présentent un diamètre compris entre 1  $\mu m$  et 1 nm, ils possèdent deux autres caractéristiques très importantes, leur rapport surface/volume leur confère des propriétés d'adsorption des ions présents dans l'eau. Ce phénomène explique en partie l'existence de charges électriques à leur surface. Ces charges, souvent négatives, engendrent des forces de répulsion inter colloïdales. Celles-ci permettent d'expliquer la grande stabilité de ces particules mise en solution. Leur densité, qui peut atteindre 2,5, ne modifie pas la valeur de leur vitesse de décantation.

L'origine des colloïdes est très diverse. On peut citer l'érosion des sols, la dissolution des substances minérales, la décomposition des matières organiques, le déversement des eaux résiduaires urbaines et industrielles ainsi que les déchets agricoles (Cardot, 1999).

### II.2.2.2. Les types de colloïdes

Les macromolécules organiques (protéines, sucres) constituent les particules hydrophiles, elles présentent des groupements électro-négatifs de type R-NH<sub>2</sub>, R-OH, R-COOH, Ces radicaux créent des liaisons hydrogènes avec les molécules d'eau. Ces liaisons sont essentiellement de nature électrostatique entre les atomes d'hydrogène de charge positive partielle des molécules d'eau et les éléments électro-négatifs (N,O) portant une charge négative due au doublet d'électrons libres, l'énergie unitaire mise en jeu est de l'ordre de quelques kcal.moles<sup>-1</sup>, soit dix fois l'énergie des forces d'attraction de Van Der Waals.

Cette couche d'hydratation s'oppose au rapprochement des particules colloïdales et constitue un facteur de stabilisation, il est à noter que les colloïdes de nature hydrophile sont responsables de la coloration des eaux (Cardot, 1999).

Les colloïdes hydrophobes sont en général de nature minérale, ils n'ont aucune affinité pour les molécules, la formation d'une couche d'hydratation est impossible, leur surface présente des charges négatives qui engendrent une répulsion mutuelle empêchant toute agglomération. Ce type de particules est en partie responsable de la turbidité, peu de colloïdes sont exclusivement hydrophiles ou hydrophobes, il s'agit en fait d'un degré d'hydratation plus au moins important autorisé par la présence de groupements fonctionnels de type hydrophile.

### II.2.2.3. Structure des colloïdes

Dans une eau de surface, les colloïdes portent des charges électriques négatives situées à leur surface, l'ensemble constitue la charge primaire de la particule qui a pour origine l'ionisation de groupements chimiques tels qu'OH, COOH et NH<sub>2</sub> et l'adsorption d'anions présents dans l'eau. Ces charges négatives attirent les ions positifs en solution dans l'eau. Ceux-ci sont étroitement collés au colloïde et forment la couche liée ou de STERN.

Cette couronne accompagne la particule lors de ces déplacements dans l'élément liquide, la couche de STERN attire à son tour des anions accompagnés d'une faible quantité de cations.

Cette nouvelle formation comporte une alternance d'anions et de cations. C'est la couche diffuse ou couche de GOUY (Figure II.2). Il y a donc formation d'une double couche ionique, présente au niveau de tout colloïde (Cardot, 1999). Il se crée entre ces deux couches un potentiel électrostatique ou de NERNST qui est fonction de la température. Ce potentiel s'exprime en fonction de la quantité de charges positives et négatives présentes au niveau de la double couche.

La couche entre la couche de STERN et celle de GOUY, définit le plan de cisaillement. Il caractérise le plan entre la partie de la solution solidaire du colloïde (couche liée) et celle se déplaçant indépendamment ou avec un certain retard. La valeur de potentiel électrostatique au niveau du plan de cisaillement, est le potentiel zêta noté  $\xi$ . Il caractérise le signe de la charge primaire, les colloïdes étant chargés négativement, ce potentiel est négatif, les particules de potentiel zêta négatif se repoussent très fortement inhibant toute agglomération, elles peuvent rester très longtemps en suspension quelque soit leur densité. La valeur de  $\xi$  est de -30 à -35 mV dans les eaux naturelles.

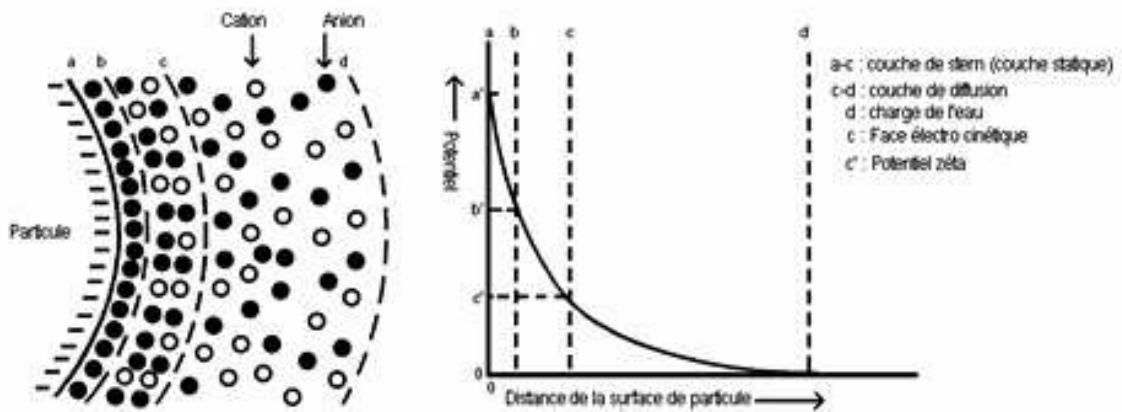


Figure II.2. Structure de la double couche

## II.2.3. Phénomène de la coagulation

### II.2.3.1. Théorie

Il apparaît évident que la stratégie d'élimination des colloïdes passe par l'annulation du potentiel zêta. En effet, la neutralisation des charges primaires permet d'annuler les forces de répulsion. On considère deux particules colloïdales A et B de diamètre  $D$  et distantes d'une valeur  $d$ . Entre ces deux particules, il existe des forces de répulsion électrostatiques et des forces d'attraction intermoléculaires (Cardot, 1999).

Les charges primaires négatives de A et B engendrent des forces de répulsion pour des distances inférieures à leur propre diamètre, l'énergie potentielle de répulsion est notée  $W_r$ .

Les forces intermoléculaires d'attraction sont celles de Van Der Waals, elles englobent trois types d'interaction :

- Les interactions dipôle-dipôle dues aux moments dipolaires des particules ;
- L'effet d'induction entre particule polaire et polarisable ;
- L'effet de dispersion dû aux moments dipolaires instantanés des colloïdes.

Il est à noter que l'énergie potentielle de ces interactions moléculaires est unitairement très faible, de l'ordre de quelques dixièmes de  $\text{kcal.moles}^{-1}$ . En effet, cette énergie, notée  $W_a$ , varie selon l'inverse de la distance à la puissance 7, l'énergie potentielle totale ou énergie globale interparticulaire est notée  $W_t$ .

On peut donc écrire  $W_t = W_a + W_r$ . Si  $W_t$  est positive, alors l'énergie de répulsion est prépondérante. Dans le cas contraire, l'énergie d'attraction permet l'adhésion des deux particules.

Pour permettre l'adhésion des particules, il faut vaincre la barrière énergétique de répulsion. Deux stratégies sont théoriquement possibles: soit augmenter l'énergie cinétique des particules (pour avoir  $E_c$  supérieure à  $W_t$ ), ou d'abaisser cette barrière en réduisant ou en annulant les forces de répulsion électrostatiques. Dans le premier cas, il faut augmenter l'agitation des particules et donc le mouvement brownien par élévation de température, cette méthode est inadaptée pour les volumes d'eau à traiter (Cardot, 1999).

La réduction des forces de répulsion se conçoit avec l'annulation du potentiel zêta à l'origine de l'énergie de répulsion. Il faut annuler la charge primaire superficielle de chaque colloïde (Figure II.3 et II.4). Ce procédé consiste en une coagulation chimique due à l'action d'ions de charge inverse de celle des particules (Cardot ,1999). La coagulation est optimale pour une introduction de charges électriques positives correspondant à l'annulation du *potentiel zêta*, potentiel que l'on mesure en observant la vitesse de déplacement des particules colloïdales sous l'action d'un champ électrique dans un appareil d'électrophorèse dit zêta mètre(Mouchet, 2008). Les solutions possibles pour annuler le potentiel zêta, sont nombreuses :

- Ajustement du pH atteindre le point isoélectrique des groupements fonctionnels portés par les colloïdes ;
- Augmentation de la salinité pour obtenir une compression de la couche diffuse ;
- Neutralisation chimique de la charge primaire par des cations polyvalents ou polymères ;
- Piégeage dans des précipités de type  $Al(OH)_3$  ou  $Fe(OH)_3$  (sweep coagulation) ;
- Adsorption et pontage par des polymères chargés à longue chaîne.

Les deux premières possibilités ne sont pas applicables au niveau industriel, on n'ose imaginer les flots d'acide ou de base, les montagnes de sels nécessaires au traitement de millions de mètres cubes d'eau. Néanmoins, la limpidité de l'eau de mer est due à sa forte salinité qui permet la déstabilisation de la décantation des colloïdes. Les mécanismes d'action des coagulants s'interprètent par la neutralisation de la charge primaire suivie par le piégeage dans les précipités (Cardot ,1999).

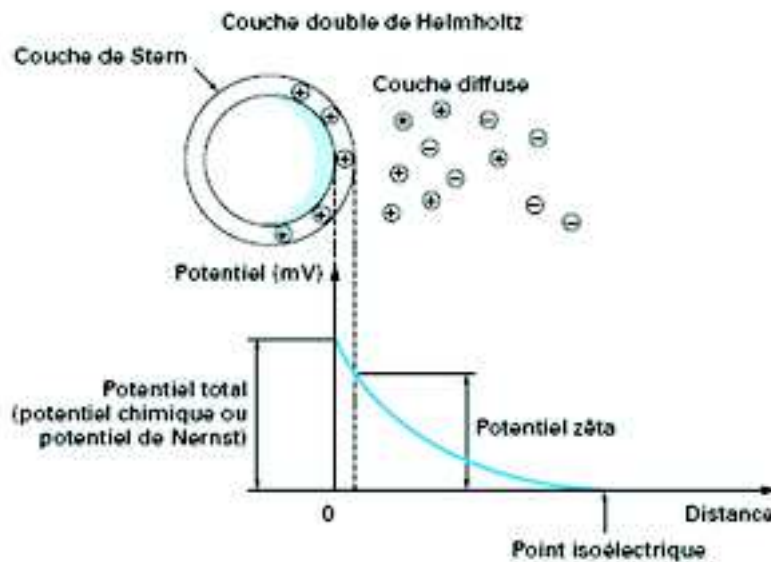


Figure II.3 Évolution du potentiel électrostatif

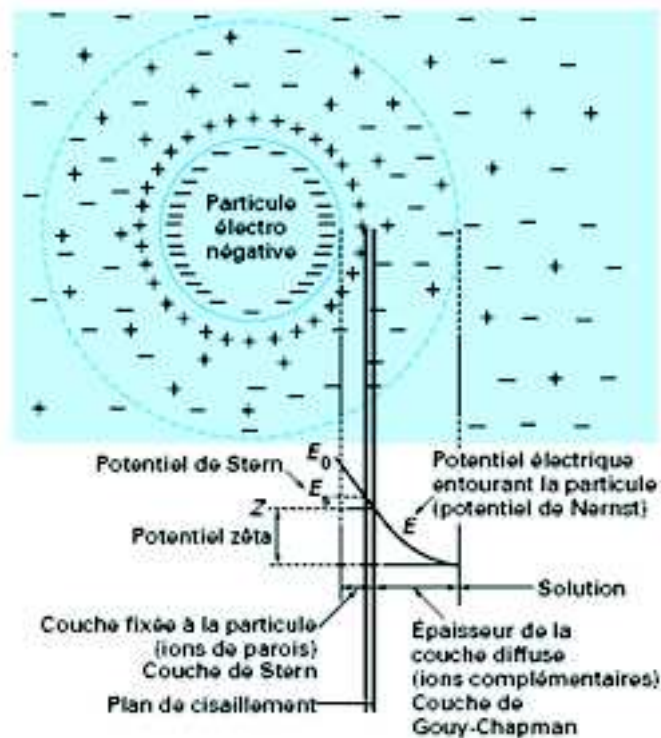


Figure II.4 Distribution ionique autour d'une particule colloïdale électro-négative

## II.2.4. Chimie de la coagulation

Il est important de connaître l'influence de l'emploi d'un coagulant sur les caractéristiques physico-chimiques de l'eau traitée, notamment lorsque le taux de traitement est élevé, quant la clarification est suivie d'une déminéralisation par échange d'ions (Mouchet, 2008).

### II.2.4.1. Réactifs utilisés

L'efficacité des coagulants est directement liée à la valence des cations utilisés, les ions monovalents sont moins efficaces que les bi ou trivalents. La règle de Schulze-Hardy précise l'efficacité des ions en fonctions de leur valence, la référence étant le  $\text{Na}^+$ , le  $\text{Ca}^{2+}$  d'une part et les ions trivalents  $\text{Al}^{3+}$  et  $\text{Fe}^{3+}$  d'autre part, présentent respectivement une efficacité 7 à 200, et 500 à 1000 fois plus importante (Cardot, 1999 ; Lind, 1995).

La mise en solution d'un coagulant se déroule en deux étapes, le cas du sulfate d'alumine est très significatif, les réactions peuvent être représentées de la façon suivant :

- 1) L'étape 1 est une phase d'hydrolyse, des intermédiaires polychargés positifs se forment. Ces composés assez fugaces présentent un atome d'aluminium dont le nombre d'oxydation est très grand. Les formes Al IV, V et VII sont rencontrées. Conformément à la règle de Schulze-Hardy, ces intermédiaires polychargés positifs sont très efficaces pour neutraliser la charge primaire négative des colloïdes, il s'agit de la véritable forme coagulant qui déstabilise les particules chargées négativement (Cardot, 1999). L'étape 1 dépend de la température et nécessite un pH compatible



avec l'existence de ces intermédiaires polychargés, le temps de formation de ces composés est de l'ordre de 0,1 seconde.

2) L'étape 2 permet la formation du précipité  $\text{Al}(\text{OH})_3$ , cette réaction dépend de l'agitation du milieu, ce précipité est l'élément qui assure le pontage et la coalescence entre les colloïdes déstabilisés: c'est la forme floculante. Tout coagulant présente successivement les deux formes actives, coagulante et floculante, le maintien de cette dernière dépend du pH du milieu, cette notion de pH permet de définir les zones optimales de coagulation floculation.

La gamme optimale de pH pour l'utilisation du sulfate d'alumine est conditionnée par l'obtention de la persistance de la forme précipitée  $\text{Al}(\text{OH})_3$ . Il s'agit ici de la fourchette 5.8 à 7.2 (Cardot, 1999) entre ces deux valeurs, le précipité est prédominant. De plus, la forme dissoute  $\text{Al}^{3+}$  est inexistante, toute fuite de cet ion au niveau d'une filière de traitement est problématique. L'ajout d'un adjuvant de floculation pour obtenir un floc cohérent est recommandé dans le cas de fuite chronique après vérification du pH.

Le chlorure ferrique possède une fourchette plus large comprise entre 5,5 et 8,3, toute fuite de  $\text{Fe}^{3+}$  peut entraîner une coloration rougeâtre de l'eau et une corrosion sur le réseau de distribution du fait de l'existence des ferrobactéries et des phénomènes d'oxydoréduction.

L'utilisation de formes polymérisées permet de s'affranchir de la première étape. Leur formule chimique est identique aux intermédiaires polychargés. Leur prix est évidemment beaucoup plus élevé, en contrepartie, les difficultés de coagulation en hiver sont résolues (Cardot, 1999).

Une autre particularité de l'addition de coagulant dans l'eau est à examiner, tout coagulant mis en solution consomme de l'alcalinité, cette perte de TAC (titre alcalimétrique complet) est fonction de la nature et la concentration du composé mis en jeu (Tableau II.1). La consommation de TAC par le coagulant et la production de  $\text{CO}_2$  sont les éléments importants de cette réaction, cette perte d'alcalinité se traduit par une chute de pH, donc par une acidification de l'eau, sauf si en même temps, on a un dégazage du  $\text{CO}_2$ . Le précipité d'hydroxy d'alumine se forme au dépend de la dureté temporaire ( $\text{Ca}(\text{HCO}_3)_2$ ) qui se transforme en dureté permanente sous forme de sulfate de calcium, ce composé peut précipiter et donner le gypse.

L'acidification de l'eau qui en découle peut perturber la réaction d'hydrolyse du coagulant, la correction de pH par addition de soude, de chaux ou de calcaire dans le cas d'eaux acides et d'acide sulfurique ou de  $\text{CO}_2$  pour les eaux incrustantes est quelquefois nécessaire, l'acidité est à vérifier avant et après la coagulation, il s'agit du problème de l'équilibre calco-carbonique.

**Tableau II.1 Perte d'alcalinité et production de  $\text{CO}_2$  de quelques coagulants**

Nom commercial	Perte d'alcalinité en g de CaO	Perte de TAC En °F	Production de Co <sub>2</sub>
Sulfate d'alumine	25	4.5	40
WAC	5.4	1	8.6
Aqualenc	5	9	8
Chlorure ferrique	21.1	3.8	33.3
Clairtan	20.1	3.6	31.6

Les Réactifs coagulants les plus généralement utilisés (Godart, 2008):

1. Sulfate d'alumine (ou d'aluminium) :  $Al_2(SO_4)_3 \cdot 18H_2O$

Présentation :

- Cristallisé (poudre, grains, noisettes, pains, etc.): 17-18 %  $Al_2O_3$  ;
- Solution à 600 ou 720 g · L<sup>-1</sup> environ de sulfate d'alumine cristallisé.

1. Chlorure ferrique :  $FeCl_3$

Présentation :

- Sublimé ou cristallisé ( $FeCl_3$  ou  $FeCl_3 \cdot 6H_2O$ ) ;
- Solution à 600 g · L<sup>-1</sup> de  $FeCl_3$  environ.

1) WAC (de l'anglais Water aluminium chloride), Aqualenc, PCBA (liquides) ou PACI (**poudre**) : ce sont diverses formes de polychlorure basique d'aluminium ; ces réactifs sont moins acidifiants que le sulfate d'aluminium ou le chlorure ferrique et donnent souvent (mais pas toujours) naissance à un floc mieux décantables que celui obtenu à partir de sulfate d'aluminium.

2) Autres réactifs parfois mis en œuvre :

- **Sulfate ferreux** :  $FeSO_4 \cdot 7H_2O$  (présentation : en poudre).

Les taux de traitement les plus couramment appliqués sont récapitulés dans le tableau II.2.

Tableau II.2 Taux de traitement appliqué

Taux de traitement usuels (g/m <sup>3</sup> )			
Traitement	Sulfate d'aluminium cristallisé	Chlorure ferrique	
		Pur ( $FeCl_3$ )	Solution (41 %)
Flocculation sur filtre	3 à 10	1.5 à 4	3.5 à 10
Eau peu chargées	15 à 30	6 à 12	15 à 30
Eau de turbidité moyenne	30 à 60	12 à 25	30 à 60
Eau très turbides	60 à 150	25 à 60	60 à 150
Eau très colorées	100 à 250	40 à 100	100 à 250
Eaux riches en plancton	60 à 150	25 à 60	60 à 150

### II.2.5. Caractéristiques des coagulants

Les coagulants utilisés pour le traitement des eaux destinées à l'alimentation humaine doivent :

- Être peu coûteux ;

- Être totalement inoffensifs par eux-mêmes et par les produits qu'ils forment ;
- Pouvoir être mis en œuvre aisément ;
- Pouvoir être dispersés sans difficultés ;
- Avoir un pouvoir destabilisant, et rassemblant vis-à-vis des colloïdes hydrophiles ;
- Avoir, par eux-mêmes ou leurs produits, un pouvoir agglomérant (collant) vis-à-vis des autres particules en suspension, et de préférence adsorbant vis-à-vis des matières organiques ;
- Permettre la formation d'agglomérats lourds suffisamment cohérents pour résister sans être détruits à la turbulence résultant de l'écoulement des eaux dans les ouvrages de traitement (Godart, 2008).

Dans tous les cas, le coagulant, qui doit être très rapidement dispersé dans l'eau à traiter, réagit avec les bicarbonates, contenus dans l'eau (alcalinité) pour former un floculat d'hydroxydes métalliques de structure complexe résultant de leur tendance à la polymérisation. L'alcalinité de l'eau doit, par conséquent, être suffisante pour provoquer la décomposition du coagulant, dans le cas contraire, on ajoute simultanément de la chaux ou, plus rarement, du carbonate de sodium (Godart, 2008).

## II.2.6. Les essais Jar-Test

En laboratoire, le choix par comparaison entre les coagulants et la détermination du taux de traitement optimal se fait au moyen de l'essai dit de « Jar Test » ou essai en flacons. Cet essai s'effectue à l'aide d'un appareillage comprenant cinq (ou plus) récipients d'un litre environ (Jars) soumis, simultanément, à une agitation de rotation identique (Figure II.5). Cette agitation est d'abord rapide, de façon à disperser les réactifs testés au sein de l'eau étudiée, puis lente (40 à 20 tr/min), de façon à favoriser la coalescence des flocons formés.

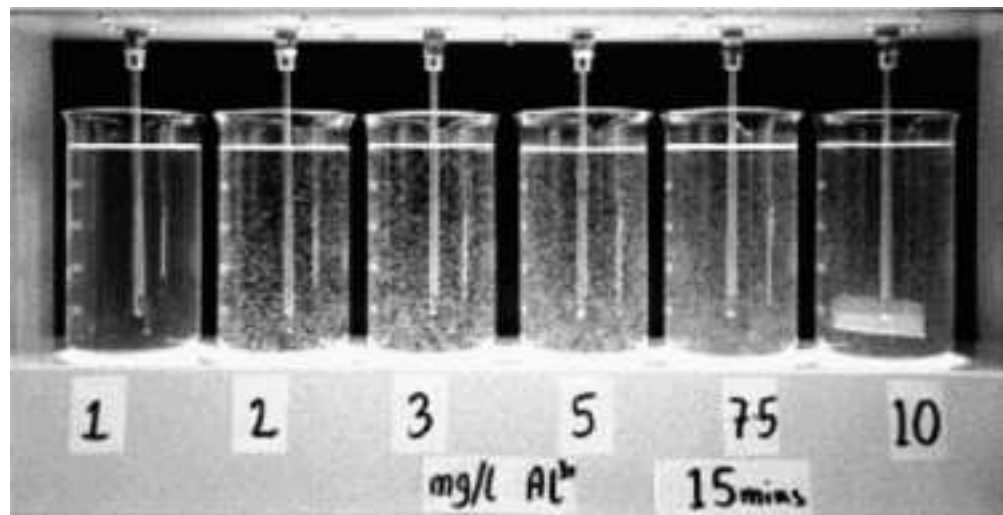


Figure II.5 Jar Test

L'agitation terminée (15 à 20 min), on laisse les matières en suspension se déposer et l'on note le volume et l'aspect des flocons et des boues ainsi que la turbidité du liquide surnageant, en l'état et filtré sur papier. Il faut procéder à une température voisine de celle de l'eau qui sera traitée en station. Dans un premier essai, on recherche la dose optimale pour un coagulant donné (dosage échelonné croissant d'un récipient à l'autre), puis l'on compare l'efficacité de coagulants et adjuvants divers, tout en ajustant le pH optimal. La notion de pH optimal de coagulation floculation est particulièrement importante ; pour être dans la zone



des précipités d'hydroxydes métalliques, des pH acides sont requis, entre 4.5 et 6.5 selon les coagulants (Figures II.6 ; II.7). Le coagulant lui-même abaisse le pH proportionnellement à la dose introduite. Cela ne suffit pas toujours et il faut alors acidifier préalablement. Avec des eaux naturellement très acides, ce peut être le contraire, tout cela doit être compatible avec l'alcalinité suffisante (déjà évoquée précédemment) pour permettre la décomposition du coagulant.

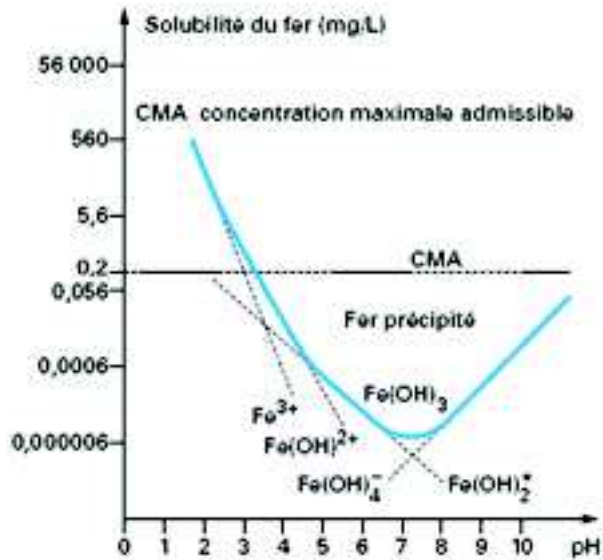


Figure II.6. Diagramme de solubilité de fer

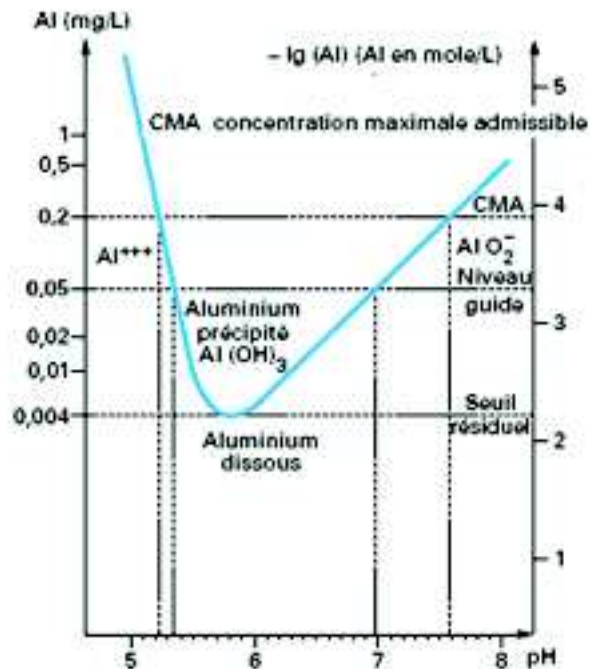


Figure II.7. Diagramme de solubilité d'aluminium

L'essai se prête aussi à l'étude de l'effet d'addition de microsable utilisé pour la décantation lestée (Lind, 1994a ; 1994b).

## II.2.7. Coagulants organiques

Dans certains pays, on remplace tout ou partie du coagulant minéral par un poly électrolyte cationique de synthèse (mélamine formaldéhyde, EPI-DMA, POLYDAD-MAC) ; ces réactifs de masse molaire relativement peu élevée (de l'ordre de 105) sont présentés en solution aqueuse et dilués en ligne juste avant injection dans l'eau à traiter ; leur taux de traitement usuel, exprimé en produit commercial, est généralement compris entre 5 et 15 g/m<sup>3</sup> ; leur avantage est, entre autres, de diminuer la production de boues et de ne pas modifier le pH ni la salinité de l'eau (Figure II.8); et ne peuvent pas être utilisés dans toutes les industries (Mouchet, 2008).

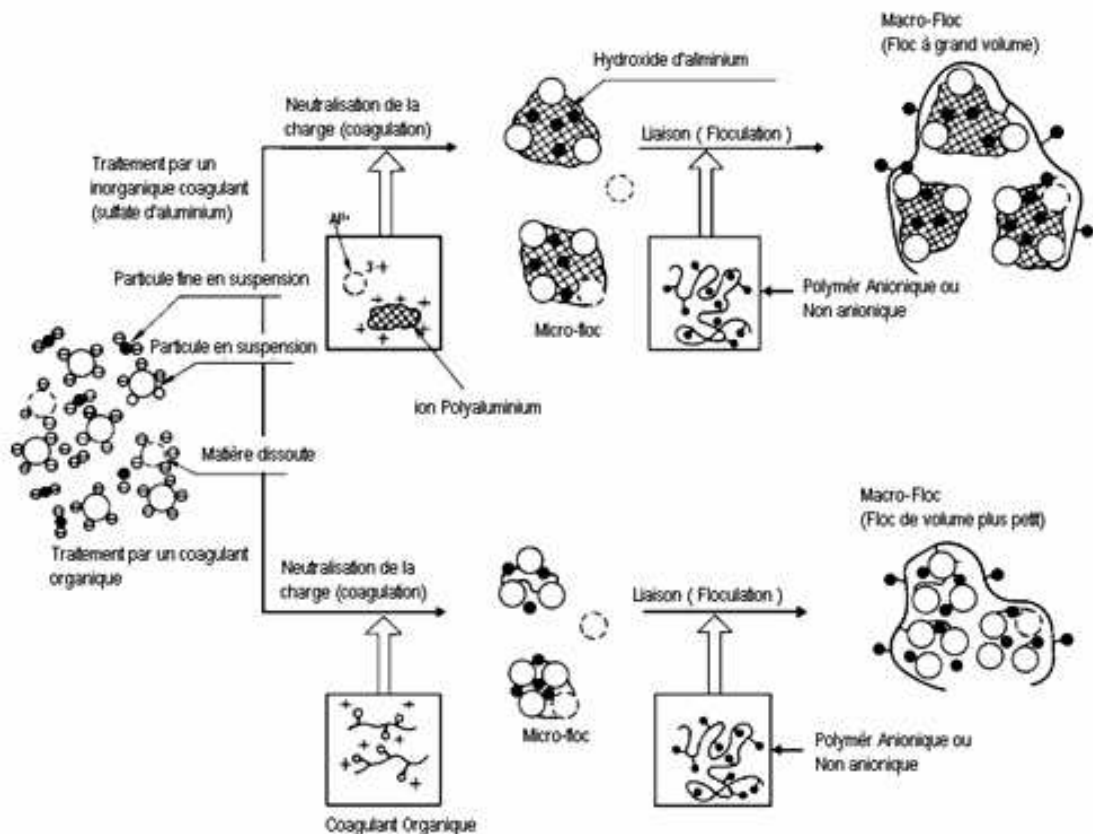


Figure II.8 Comparaison entre un coagulant organique et non organique

## II.2.8. Adjuvants

Pour améliorer l'effet des coagulants vis-à-vis de l'agglomération des particules en suspension, ou encore pour améliorer, d'une part, la résistance mécanique des agglomérats obtenus et pour, d'autre part, accroître leur pouvoir d'adhésion au sable des filtres ou au microsable fluidisé, on utilise des adjuvants à dose faible. Ces adjuvants jouent également un rôle pour améliorer la floculation : on les appelle tantôt adjuvants de coagulation, tantôt adjuvants de floculation. Le plus ancien est minéral : c'est la silice activée (silicate de sodium neutralisé par addition d'un corps fournissant des ions acides, comme CO<sub>2</sub> ; la structure du produit est celle d'un poly silicate, dont le comportement peut être assimilé à celui d'un poly électrolyte anionique). Les plus récents et les plus efficaces sont des poly électrolytes organiques formant de longues chaînes qui jouent le rôle d'armature au sein

des agglomérats et qui, par ailleurs, rendent plus adhésives les surfaces des grains des filtres (Godart, 2008).

## II.3. Flocculation

### II.3.1. Phénomène de flocculation

---

Après avoir été déstabilisées par le coagulant, les particules colloïdales s'agglomèrent lorsqu'elles entrent en contact, c'est la flocculation. Le floc ainsi formé, peut décanter, flotter ou filtrer (coagulation sur filtre), suivant le procédé de rétention le plus adapté choisi.

Toutes les techniques de flocculation appliquent l'un et/ou l'autre de ces principes, on distingue la flocculation péri-cinétique due à l'agitation désordonnée des particules et la flocculation ortho-cinétique provoquée par les mouvements de l'eau, cette flocculation est fonction du gradient de vitesse conféré au milieu par agitation, elle est responsable en grande partie du phénomène de flocculation (Cardot, 1999).

### II.3.2. Nature des flocculants

---

Les flocculants sont des composés agissant par pontage, il s'agit de polymères à haut poids moléculaire présentant des groupes réactionnels de charge inverse de celle de la suspension à traiter, on distingue des flocculants naturels et synthétiques.

Les polymères synthétiques se répartissent en polyanions, polycations, non ioniques et amphotères, ce sont des polymères de type polycrylamide ou polyacrylate. Toutes ces molécules se définissent par leur poids moléculaire (PM). Un PM élevé est caractéristique de polymères à longue chaîne, les polymères chargés, ou polyélectrolytes, se définissent également par leur densité de charge (DC), de nombreuses charges électriques neutralisantes procurent une forte DC (Cardot, 1999).

### II.3.3. Processus de la flocculation

---

Les polyanioniques agissent par pontage en association avec les cations bivalents du milieu tels que les ions  $\text{Ca}^{2+}$  et  $\text{Mg}^{2+}$ , dans ce cas, le titre hydrotimétrique de l'eau est essentiel, on ne l'ajoute qu'en fin de flocculation lorsque les microflocs sont formés (Figure II.9). Les polycationiques agissent directement par leurs charges positives, elles neutralisent les charges primaires des colloïdes qui ont échappé au coagulant, il faut les ajouter juste après la phase d'agitation rapide (coagulation).

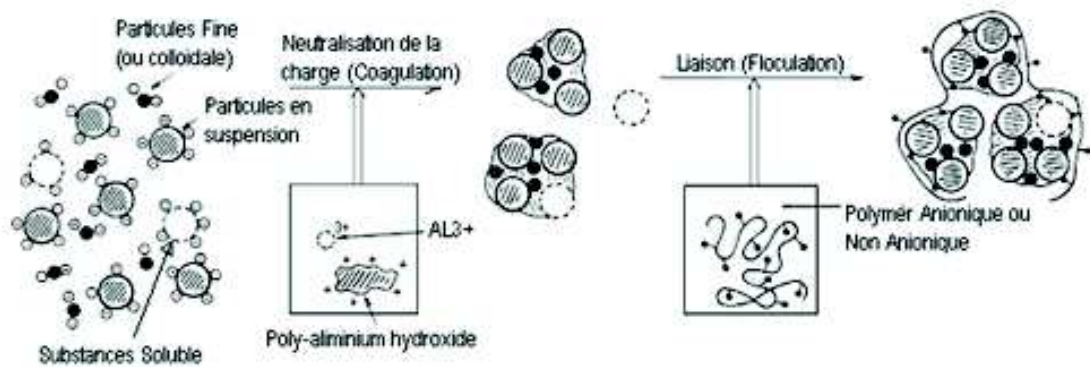


Figure II.9. Processus de floculation

On considère que l'adsorption de la chaîne de polymère sur la surface du colloïde est solide quand au moins 50% de cette surface est occupée par 70% de la longueur des molécules de floculant, les 30% restants forment des boucles et des queues qui réalisent le pontage inter-colloïdes. Pour respecter cette condition, la chaîne de polymère doit présenter un poids moléculaire (PM) élevé et une densité de charge (DC) faible, dans le cas contraire, les molécules de floculant sont trop courtes et trop riches en charges, en conséquence, 100% du polymère s'adsorbe à plat sans boucles ni queues. Le pontage inter colloïdes est impossible (Cardot, 1999).

L'action des alginates a été particulièrement étudiée, ils établissent des ponts avec les colloïdes déstabilisés grâce à leurs acides  $\alpha$ -L-guloniques et  $\beta$ -D-mannuroniques, l'effet de pontage est renforcé par les ions  $\text{Ca}^{2+}$  qui permettent la formation d'un réseau maillé.

### II.3.4. Phénomène de base de la floculation

Pour favoriser la dispersion des réactifs dans l'eau à traiter et la formation consécutive de gros agglomérats, on introduit les coagulants et adjuvants en des endroits où règne une forte turbulence (chute, remous, agitation mécanique rapide, etc.) ; on parle de flash mixing ou de flash mélange et la durée de séjour ne doit pas y excéder deux minutes ; la notion d'énergie de dispersion est très importante (Godart, 2008).

On introduit ensuite l'eau dans un ouvrage où règne une turbulence faible mais non négligeable. Les micro-flocons formés au stade de l'agitation rapide voient leur dimension croître considérablement par effet de pontage notamment. En fait la floculation est un cas particulier de coagulation par adsorption ; on distingue usuellement la floculation péricinétique (mouvement brownien des particules) et la floculation ortho-cinétique (liée au déplacement du fluide). L'ouvrage où se produit l'opération s'appelle un floculateur ; il existe des floculateurs statiques, du type à chicanes (section fixe ou croissante) et des floculateurs dynamiques, les plus fréquents actuellement, à pales ou à équipages tournants, à un ou deux compartiments verticaux ou horizontaux, Les quatre notions de base pour une bonne floculation sont (Godart, 2008) :

- Le brassage ne doit pas avoir d'effet destructif ;
- Le temps de floculation qui doit se situer entre 10 et 30 minutes ;
- Le gradient d'écoulement :

$$G = \sqrt{\frac{P}{\mu}} \quad (II.1)$$

Avec  $\mu$  viscosité dynamique,

P puissance dissipée par unité de volume (frottements) ;

- Une totale absence de courts circuits hydrauliques.

La cohésion du floc peut être augmentée de deux façons :

- En introduisant dans la masse de flocons une partie des flocons constituant les boues en excès extraites lors de la décantation ;
- En maintenant une masse de flocons en expansion, que l'eau chargée de flocons, plus ou moins développés, traverse de bas en haut (Godart, 2008).

## II.4. Choix du coagulant

Les critères de choix d'un coagulant sont nombreux (Tableau II.3), son efficacité à réduire la couleur, la turbidité et les matières organiques d'une eau est essentielle, on ne peut, néanmoins, négliger son prix et les effets secondaires comme les fuites en  $Fe^{3+}$  et  $Al^{3+}$ .

Il faut savoir que le chlorure ferrique contient du manganèse sous forme de  $Mn^{2+}$  à raison de 0,057 g pour 100 g de coagulant, Le traitement d'eau chargée en matières organiques exige du chlorure ferrique.

Les coagulants à base de sels d'aluminium permettent d'obtenir une eau traitée très faible en turbidité, la vérification de la dose de coagulation à mettre en œuvre passe par le JAR-TEST.

Nom commercial	Masse molaire	Densité	Forme	Formule	Apport en $M_2O_3$	Remarques
<b>SELS ALUMINIUM</b>						
Sulfate d'alumine	666	-	solide	$Al_2(SO_4)_3 \cdot 18 H_2O$	17 à 18	Poudre irritante
	645	1.45	Liquide		8.2	Liquide corrosif
Polychlorosulfate Basique D'aluminium		1.2	Liquide	$Al_n(OH)_m(SO_4^{2-})_k Cl_{3n-m-2k}$	10	Réactif acide dénommé WAC
<b>SELS DE FER</b>						
Chlorure ferrique	162.5	1.45	Liquide	$FeCl_3$	41 % Fe	Réactif acide
Sulfate ferreux	278	-	Poudre	$FeSO_4 \cdot 7H_2O$	20 % Fe	
Chlorosulfate ferrique	187	1.5	liquide	$FeClSO_4$	14 % Fe	Réactif acide

Tableau II.3 Caractéristiques de quelques coagulants

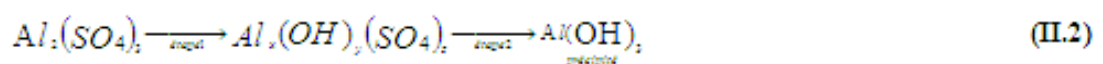
## II.5. Modélisation de la dose de coagulant : état de l'art.

En raison de la complexité des phénomènes biologiques, physiques et chimiques mis en jeu dans les procédés impliqués dans les unités de production de l'eau potable, il est souvent très difficile de quantifier les interactions et les relations qui existent entre les entrées et les sorties des procédés. Les modèles des procédés, lorsqu'ils existent, sont souvent spécifiques à un site et sont incapables de traiter simultanément des variations continues sur plus d'une ou deux variables clés du procédé (Baxter et al., 2002). Différents travaux de recherche ont été réalisés, la plupart concerne des études d'optimisation, de commande et d'estimation des paramètres (Villa et al., 2003 ; Demotier et al., 2003). Ils sont basés, explicitement ou implicitement sur un modèle mathématique qui est exprimé généralement sous la forme d'équations différentielles ou aux différences.

Deux principaux objectifs sont ciblés lors de la gestion et l'exploitation d'une station de traitement des eaux, un objectif de qualité et un objectif de coût. L'atteinte de l'objectif de qualité, exige la production d'eau en quantité suffisante, de façon continue et avec une qualité répondant aux normes de potabilité (normes chimiques, microbiologiques..., etc.) en vigueur de plus en plus sévères, cela implique la conception d'une station de traitement relativement complexe largement dimensionnée et incluant un système étendu d'équipement mécanique et hydromécanique. Au passage de l'eau dans un tel système, la qualité de l'eau obtenue se trouve modifiée tant au niveau physique, chimique ou microbiologique. La nature et l'étendue de ces changements dépendent des caractéristiques de l'eau à l'entrée de la station et du degré d'interaction entre les différentes composantes du processus mis en jeu.

Parmi toutes les étapes de la chaîne de traitement des eaux, le procédé de coagulation floculation est l'une des étapes les plus importantes, plus de 90 % des stations sont concernées celle-ci met en œuvre des mécanismes fort complexe qui dépendent d'un nombre important de variables. Dans une chaîne de traitement conventionnelle, la coagulation consiste à déstabiliser les particules en suspension, tandis que la floculation consiste à faire agglomérer les particules déstabilisées, puis formation de flocons par absorption et agrégation (Amirtharajah et O'melia, 1990). Les particules présentes dans les eaux naturelles sont chargées négativement et sont entourées d'un nuage ionique, cette charge électrostatique ainsi que la couche ionique diffuse, constituent une barrière à l'agglomération des particules parce que les forces attractives ne peuvent agir, (O'melia 1972).

Les produits chimiques impliqués dans le processus de coagulation sont connus comme coagulant, ces coagulants sont d'origine minérale à base d'aluminium ( $Al_2(SO_4)_3 \cdot 18H_2O$ ) (Dempsey et al., 1985) ou de sel de fer ( $FeCl_3$ ) (Lefebvre et Legube, 1993), des polymères naturels ou des polymères de synthèses. Dans le cas de notre étude le coagulant utilisé au niveau de la station de traitement des eaux de BOUDOUAOU est le sulfate d'aluminium, La mise en solution se déroule en deux étapes, (Cardot, 1999) :



L'étape 1 est une phase d'hydrolyse, tandis qu'au niveau de la phase 2 il y a formation du précipité  $Al(OH)_3$ .



L'ajout du coagulant dans l'eau a les effets suivants : (i) réduction de la charge hydrostatique par leur adsorption à la surface des particules ;(ii) réduction de la charge diffuse ; les principaux facteurs influençant l'efficacité de la coagulation sont la dose du coagulant, le pH (Amirtharajah et mills, 1982, Stumm etmorgan, 1962),la turbidité initiale (Edwards et Amirtharajah, 1985) et la température de l'eau ; cette dernière est un facteur très important pouvant affecter le bon déroulement du processus de coagulation (mohtadi et rao, 1973),Ces auteurs notent que la température influence largement, l'efficacité du traitement à l'usine. D'autres variables caractérisant l'eau brute influent considérablement sur le processus de coagulation à savoir, la conductivité de l'eau, l'absorbance à 254 nm (UV<sub>254</sub>) ainsi que l'oxygène dissous (OD) (Lind, 1994a et 1994b). L'absorbance à 254 nm exprime la capacité de l'eau à absorber un rayonnement UV<sub>254</sub> à une longueur d'onde de 254 nanomètres. Cette mesure permet au professionnel de s'assurer que la désinfection aux ultraviolets est possible (Weishaar et al., 2003).

Lors de la coagulation, il faut chercher : (i) la maximisation de la déstabilisation des particules et des colloïdes organiques pour faciliter leur agglomération et leur enlèvement subséquent, par un procédé de séparation solide liquide ; (ii)la minimisation de concentration en coagulant résiduel; (iii)éventuellement la minimisation des coûts de l'opération, la coagulation optimale correspond au dosage du coagulant et au PH qui assure l'atteinte de tous ces objectifs (Edzwald et Tobiason, 1999).

Afin d'évaluer les conditions optimales de coagulation et de flocculation, des essais de «Jar-test» (JT) sont conduits à l'échelle de laboratoire (Bombaugh et al., 1967; Brodart et al., 1989). Le dispositif expérimental consiste en une série de béchers (6) de 1 litre qui contiennent les échantillons d'eau brut à tester. La température des échantillons est d'abord ajustée à la valeur choisie. Chaque bécher est muni d'un agitateur à pales, la vitesse d'agitation est variable mais elle est la même pour tous les béchers. On injecte des doses variables de coagulant, dans les béchers. Un essai de JT comprend trois (3) phases :(i) Coagulation, ajout de coagulant; de courte durée et forte agitation;(ii) Flocculation, ajout de flocculant; durée moyenne avec agitation lente;(iii)Sédimentation, durée relativement longue et aucune agitation (Baxter et al., 1999; Joo et al., 2000; Yu et al., 2000). Pour des filières de traitement physico-chimiques conventionnelles des JT menés dans une large gamme de conditions d'opération permettent de déterminer le type de coagulant, son dosage, le pH et les conditions d'agitation qui maximisent la réduction de la turbidité et la réduction de la couleur, et offre le PH optimum (Amirtharajah et O'melia, 1990).

Ce type d'approche a le désavantage d'avoir un temps de retard relativement long. En effet, on ne modifie la dose de coagulant qu'une fois un événement apparu. De plus, elle ne permet pas de suivre finement l'évolution de la qualité de l'eau brute, par exemple, si l'eau brute devient plus « facile à traiter » l'opérateur ne le verra pas forcément et donc ne modifiera pas la dose de coagulant d'où un coût d'exploitation plus élevé que nécessaire et une économie non réalisée. On voit ici tout l'intérêt de disposer d'un contrôle automatique et efficace de ce procédé pour une meilleure efficacité de traitement et une réduction des coûts d'exploitation (Valentin, 2000).

Il s'avère donc nécessaire de s'assurer en permanence de la conduite optimale du procédé de coagulation au niveau d'une station de traitement des eaux, la moindre défaillance sur les processus impliqués dans le procédé lui même est néfaste et influe directement sur le rendement d'utilisation de la station de traitement. L'information permettant le suivi et le contrôle du procédé est donnée par les mesures des variables de ce processus; pour cela on trouve dans la littérature que toutes les techniques proposées pour l'automatisation du phénomène incluent les variables descriptives de l'eau brute ainsi



que la dose de coagulant (Zhang et Stanley, 1999). Automatiser peut avoir des objectifs diverses, les plus fréquents sont d'augmenter les performances du système de production, de garantir la qualité du produit, de diminuer les coûts de production et d'améliorer la sécurité de l'installation industrielle et de son environnement (Boillereaux et Flaus, 2003).

Deux catégories de technique d'automatisation du procédé de coagulation sont utilisées : (i) une technique basée sur la qualité de la coagulation ;(ii) et une deuxième technique basée sur les variables descriptives de l'eau brute, (Bernazeau et al., 1990).

La première technique d'automatisation basée sur la qualité de coagulation, procède à l'ajustement automatique de la dose de coagulant avec l'eau brute. Cette technique nécessite l'installation d'un capteur appelé Streaming Current Detector (SCD), ce dernier évalue en continu l'addition d'un coagulant à l'eau brute et renvoie un signal directement proportionnel aux résultats de cette coagulation. L'inconvénient de ce système est en premier lieu son coût d'exploitation élevé, et que son efficacité était limitée pour des PH de l'eau brute élevée (PH  $\square$  8),(dentel, 1995 ; dentel et al., 1989).

La deuxième approche utilisée pour l'automatisation du procédé de coagulation est basée sur la relation dose de coagulant-variables descriptives de la qualité de l'eau brute ; le principe consiste dans le développement d'un modèle reliant la dose de coagulant optimum à ces différentes variables descriptives, cette technique permet de s'affranchir des coûts d'exploitation élevés du SCD et de son efficacité limitée pour certains types de qualités des eaux. Les techniques statistiques de modélisation ont déjà fait l'objet de nombreuses applications dans ce domaine. Les approches proposées reposent le plus souvent sur l'utilisation des modèles de régression linéaire ou stochastique.

Bazer-bachi et al.(1990), ont proposé deux modèles mathématiques basés sur des équations polynomiales, reliant la dose optimale du coagulant, le sulfate d'aluminium, à la qualité de l'eau brute. Le premier modèle applicable aux eaux dont la turbidité est inférieure à 20 NTU tient compte de quatre variables caractéristiques de l'eau brute qui sont : la turbidité, la résistivité, la teneur en matière organique et la température. Le deuxième modèle utilisable pendant les périodes de crue intègre un cinquième descripteur: la nature de la suspension minérale. D'autres chercheurs (Critchley et al., 1990, Ellis et al., 1991; Ratnaweera et Blom, 1995; Girou et al., 1992; van leeuwen et al., 1999a) ont proposé des modèles linéaires, reliant la dose de coagulant aux variables descriptives de la qualité l'eau. Ces derniers modèles impliquent l'utilisation de différents types de données incluant les variables descriptives de la qualité de l'eau brute. Le modèle de Girou et al. (1992) est basé sur la concentration en ions de calcium, ions de bicarbonate, ions de sulfate, la turbidité initiale, la température et le PH. Les données utilisées dans le modèle développé par, Ratnaweera et Blom, (1995)sont le débit de la rivière, le temps de sédimentation, la température, la turbidité, le PH et la conductivité. Alors que le modèle proposé par Critchley et al. (1990) inclue la couleur, le débit de la rivière, PH, Conductivité et la température.

La variation du nombre et du type de paramètre impliqué dans l'élaboration des modèles mathématiques pour l'estimation de la dose de coagulant peut être due à : (i) Certains paramètres caractérisant la qualité de l'eau brute des eaux de rivières qui ne varient pas considérablement et peuvent être considérés comme des constantes dans le développement des modèles, (ii) Certaines variables d'eau brute exigent des doses de coagulant beaucoup plus importantes que d'autres; dans ce sens les paramètres à influence mineure ne peuvent pas être inclus dans les modèles,(iii) et enfin la qualité de l'eau demandée qui conditionne les taux de rabattement pour chaque variable. La modélisation de la dose du coagulant s'appuyait, jusqu'à il y a une vingtaine d'années, presque exclusivement sur les modèles linéaires, cependant et malgré leur apport dans

le développement des systèmes de traitement des eaux, et compte tenu du nombre de phénomènes physiques qui interviennent, ce type de modèles restent insuffisant pour représenter la nature complexe qui gouverne les processus mis en jeu.

Le progrès important réalisé au cours des dernières années dans le domaine de l'intelligence artificielle, a permis de réduire les difficultés et de s'affranchir des limitations des modèles linéaires ; les réseaux de neurones artificiels et la logique floue font parties de ces techniques. Ces derniers peuvent être utilisés pour essayer de fournir des modèles de procédés de traitement avec deux objectifs principaux : une meilleure efficacité du traitement et une réduction des coûts d'exploitation. En effet, les réseaux de neurones ne nécessitent pas de connaissance explicite sur le procédé à modéliser, le réseau de neurones va élaborer lui-même ces connaissances à partir d'exemples qu'on lui présentera. Il faut donc passer par une phase d'apprentissage pour ajuster les différents paramètres du réseau, par la suite le réseau est capable d'analyser de nouveaux cas (Valentin, 2000).

Dans ce sens de nombreux modèles basés sur la technique des réseaux de neurones artificiels ont été mis au point, ces modèles incluent : (i) un système global basé sur les réseaux de neurones supervisé et non supervisé, utilisant les cartes topologiques de kohonen pour la validation et la reconstruction des données, et un perceptron multicouches pour l'estimation de la dose du coagulant (Valentin et al., 1999a et 1999b), ce système à été intégré dans le cadre de la construction d'un capteur logiciel, (ii) d'autres modèles ont été proposés (Gagnon et al., 1997; Evan et Enoch, 1998; Yu et al., 2000; Adgar et al., 1995, 2000; Nahm et al., 1996; Sperring et al., 1992; Maier et dandy, 1996a; Maier et al., 2004; Mirsepassi et al., 1995, 1996, 1997; Böhme et al., 1999); ils expriment la dose de coagulant à injecter en fonction des différentes variables descriptives caractérisant l'eau brute à l'entrée de la station de traitement des eaux. D'autres travaux de recherches (Baxter et al., 2001a, Baxter et al., 2001b; Baxter et al., 2002; Lamrini et al., 2005a, 2005b; Cox et al., 1994, 1996, 2003; Hernandez, 2006) ont montré l'importance des réseaux de neurones comme outils pour l'élaboration des modèles mathématiques pour des fins d'automatisation et de supervision des procédés impliqués dans les stations de traitements des eaux.

Récemment d'autres chercheurs ont montrés l'importance des méthodes basées sur les techniques à base d'intelligences artificielles pour la modélisation de la dose de coagulant aux niveaux des stations de traitements des eaux (Chen et Hou, 2006; Heddam et Dechemi, 2008; Park et al., 2008; Robenson et al., 2009; Heddam et al., 2010; Heddam et al., 2011a, 2011b, 2011c; Wu et Lo, 2008, 2010).

Dans le cadre de notre travail de recherche on s'intéresse particulièrement à la modélisation de la dose de coagulant par les techniques d'intelligences artificielles. Dans ce sens trois types de modèles ont été développées : en premier lieu un modèle à base de régression linéaire multiple (RLM) à été élaborer ; un deuxième type de modèle à base de réseaux de neurones artificiels à savoir : le perceptron multicouches (MLPNN), les réseaux de neurones à fonction radiales de base (RBFNN), et enfin les réseaux de neurones dites générale régression neural network (GRNN), ces trois types de réseaux de neurones ont été appliquer et comparer les uns par rapport aux autres. Le troisième type de modèle utilisé dans le cadre de cette thèse est le modèle à base de logique floue appelé ANFIS (Système d'inférence flou organisé en réseau adaptatif on bien Adaptive Neuro-Fuzzy Inference System), qui combine la logique floue et les réseaux de neurones artificielles.

## II.6. Conclusion

Dans ce deuxième chapitre nous avons décrit les différentes composantes de la coagulation floculation. Nous avons détaillé, plus particulièrement, l'aspect physico chimique qui fait appel à des réactions chimiques très complexes. Nous avons décrit les différents paramètres physicochimiques influençant le bon fonctionnement du procédé de coagulation.

Le chapitre suivant est consacré à la modélisation linéaire de la dose de coagulant. Nous verrons en détail les différentes étapes à suivre pour le développement d'un tel type de modèle, et les différents variables descriptives à prendre en considération pour la construction du modèle.

# Chapitre III Modélisation linéaire de la dose de coagulant

## III.1. Introduction

L'analyse linéaire de la régression, appelée aussi plus simplement régression linéaire (*RL*), est l'une des méthodes statistiques les plus utilisées dans les sciences appliquées. Son objectif est double : Elle consiste tout d'abord à décrire les relations entre une variable privilégiée, appelée variable à expliquer (ou dépendante), et plusieurs variables jouant un même rôle par rapport à la première, appelées variables explicatives (ou indépendantes). Elle permet aussi d'effectuer des prévisions de la variable à expliquer en fonction des variables explicatives.

Les liaisons entre les variables explicatives exercent une influence très importante sur l'efficacité de la méthode, quel que soit l'objectif visé.

La régression linéaire multiple (*RLM*) est la méthode la plus simple de modélisation. C'est une généralisation du modèle de régression simple lorsque les variables explicatives sont en nombre fini. Elle consiste à rechercher une équation linéaire par rapport à ses paramètres, reliant la variable à modéliser  $Y$  (variable expliquée) au vecteur d'entrées  $\mathbf{X} = \{x_j, j = 1 \dots N\}$  ou (variables explicatives). On note  $N$  le nombre d'individus statistiques considérés,  $y_i$  la  $i$ -ème observation de la variable  $Y$  et  $x_{ij}$  celle de la Variable  $X_j$ . L'équation linéaire recherchée est de la forme :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_j x_{ij} \quad (\text{III.1})$$

Les paramètres  $\beta_j$  obtenus sont appelés coefficients de régression partielles ; chacun d'eux mesure l'effet de la variable explicative  $X_j$  concernée sur la propriété modélisée lorsque les autres variables explicatives sont maintenues constantes ( $i$  exprime l'individu, tandis que  $j$  exprime la variable), la régression linéaire est facile à mettre en œuvre, et les coefficients  $\beta_j$  obtenus peuvent être interprétés : ils mesurent l'influence de chacune des variables sur la grandeur étudiée. Cependant, il est souvent nécessaire d'avoir recours à des modèles de plus grande complexité.

Pour chaque élément  $i$  de la base d'apprentissage, le résidu est défini comme la différence entre la valeur de la grandeur à modéliser pour cet élément  $i$  et l'estimation du modèle :

$$R_i = y_i^d - y_i^e \quad (\text{III.2})$$

Avec  $y_i^c$  désigne la valeur observée de l'individu  $i$ , tandis que  $y_i^d$  désigne la valeur calculée par le modèle.

L'objectif de ce travail de recherche est de développer un modèle mathématique qui permet d'estimer la dose de coagulant en fonction des variables descriptives caractérisant l'eau brute à l'entrée de la station.

Le premier modèle exploité dans le cadre de notre travail est basé sur la régression linéaire multiple (MLR). Comme nous l'avons déjà rappelé, un grand nombre de descripteurs différents sont collectés pour la modélisation de la dose de coagulant (six variables), car les facteurs déterminants du processus étudié ne sont a priori pas connus. Cependant, les descripteurs envisagés n'ont pas tous une influence significative sur la grandeur modélisée, et les variables ne sont pas toujours mutuellement indépendantes. De plus, le nombre de descripteurs, c'est-à-dire la dimension du vecteur d'entrée, détermine la dimension du vecteur des paramètres à ajuster. Si cette dimension est trop importante par rapport au nombre d'exemples de la base d'apprentissage, le modèle risque d'être sur ajusté à ces exemples, et incapable de prédire la grandeur modélisée sur de nouvelles observations.

Il est donc nécessaire d'étudier la possibilité de réduire la dimensionnalité des variables d'entrée. Plusieurs approches sont possibles pour résoudre ce problème ; dans notre cas nous avons utilisé la technique de l'analyse en composantes principales (ACP).

## III.2. L'analyse en composantes principales ACP

### III.2.1. Généralités

---

L'analyse en composantes principales (ACP) est une technique descriptive permettant d'étudier les relations qui existent entre des variables descriptives, sans tenir compte, a priori, d'une quelconque structure, ni des variables, ni des individus (Palm, 1998), les domaines d'application de cette méthode sont très variés, et de nombreux exemples sont proposés, notamment, par Jackson (1991) et Press (1972).

L'ACP est une méthode d'analyse des données, lorsqu'un nombre très réduit de composantes suffit à concentrer le maximum d'informations, cette méthode, a le mérite de faciliter la visualisation et l'analyse préliminaire des données (Oukhellou, 1997).

L'utilisation de l'analyse en composantes principales pour l'exploitation des données remonte au début du siècle dernier. Elle est principalement issue des travaux de psychomètres américains (Pearson, 1901 ; Spearman, 1904 ; Hotelling, 1933 ; Thurstone, 1947).

### III.2.2. Principes de l'ACP

---

L'ACP (Saporta, 1990) a deux objectifs majeurs, le premier est de déterminer des représentations graphiques planes, appelées cartes, qui permettent d'analyser au mieux les proximités et les écarts entre les différentes variables. En particulier, elles doivent permettre d'isoler les individus "atypiques" et de regrouper les individus "semblables".

Le deuxième objectif de l'ACP est de fournir des "résumés linéaires" des variables, c'est à dire de remplacer les variables initiales par des combinaisons linéaires de celles-ci, ces nouvelles variables, appelées composantes principales ou facteurs, ont des propriétés qu'il est important de présenter pour faciliter l'interprétation de l'analyse.

### III.2.3. Les données de l'ACP et leurs caractéristiques

#### III.2.3.1. tableau des données

L'analyse en composantes principales (ACP) décrit un tableau [X] de données quantitatives de n lignes et p colonnes. Les lignes correspondent aux individus et les colonnes correspondent aux variables observées (X= [individus / variables]).

Le tableau des données brutes a la forme suivante :

$$X = \begin{matrix} & \begin{matrix} 1 & 2 & j & p \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ n \end{matrix} & \begin{bmatrix} x_1^1 & x_2^1 & x_j^1 & x_p^1 \\ x_1^2 & x_2^2 & x_j^2 & x_p^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_1^n & x_2^n & x_j^n & x_p^n \end{bmatrix} \end{matrix} \quad (\text{III.3})$$

$x_i^j$  Valeur prise par la variable n° j sur l'ième individu

Chaque variable est associée à une colonne de X, elle correspond à la liste des n valeurs quelle prend sur les n individus :

$$x^j = (x_1^j \ x_2^j \ x_3^j \ x_n^j) = \begin{bmatrix} x_1^j \\ \vdots \\ x_n^j \end{bmatrix} \quad (\text{III.4})$$

Chaque individu est associé à une ligne de X, elle correspond à la liste des p valeurs qu'elle prend sur les p variables :

$$x_i = (x_i^1 \ x_i^2 \ x_i^3 \ x_i^p) = [x_i^1 \ \dots \ x_i^p] \quad (\text{III.5})$$

Pour avoir une image de l'ensemble des unités, on se place dans un espace affine en choisissant comme origine un vecteur particulier de  $X_p$ , par exemple le vecteur dont toutes les coordonnées sont nulles. Alors, chaque unité sera représentée par un point dans cet espace. L'ensemble des points qui représentent les unités est appelé traditionnellement "nuage des individus".

En faisant de même dans  $R_n$  (espace de variables), chaque variable pourra être représentée par un point de l'espace affine correspondant, l'ensemble des points qui représentent les variables est appelé "nuage des variables".

Ces espaces étant de dimension supérieure en général à 2 et même 3, on ne peut visualiser ces représentations. L'idée générale des méthodes factorielles est de trouver un système d'axes et de plans tels que les projections de ces nuages de points sur ces axes et ces plans permettent de reconstituer les positions des points les uns par rapport aux autres, c'est-à-dire avoir des images les moins déformées possible (Duby & Robin, 2006).

### III.2.3.2. la matrice des poids

Les poids sont des nombres positifs de somme 1, ils sont regroupés dans une matrice diagonale D de taille n :

$$D = \begin{bmatrix} p_1 & : & : & : & 0 \\ : & p_2 & : & : & : \\ : & : & . & : & : \\ : & : & : & : & : \\ 0 & : & : & : & p_n \end{bmatrix} \quad (\text{III.6})$$

Pour l'ACP on choisit de donner le même poids  $\frac{1}{n}$  à tous les individus.

### III.2.3.3. le centre de gravité

Le vecteur g des moyennes arithmétique de chaque variable :

$$|g = (\bar{x}^1 \quad \bar{x}^2 \quad \bar{x}^3 \quad \bar{x}^j \quad \bar{x}^p) \quad (\text{III.7})$$

Avec  $\bar{x}^j$  correspondent à la moyenne de la variable  $x^j$

Le tableau Y tel que :

$$y_i^j = x_i^j - \bar{x}^j \quad (\text{III.8})$$

est le tableau centré associé à X.

Le centre de gravité G du nuage des individus est alors le point dont les coordonnées sont les valeurs moyennes des variables :



$$G = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_i^1 \\ \frac{1}{n} \sum_{i=1}^n x_i^j \\ \frac{1}{n} \sum_{i=1}^n x_i^p \end{bmatrix} = \begin{bmatrix} \bar{x}^1 \\ \bar{x}^j \\ \bar{x}^p \end{bmatrix} \quad (\text{III.9})$$

Le vecteur des coordonnées centrées de la variable  $x^j$  est :

$$x_i^j = \begin{bmatrix} x_i^j - \bar{x}^j \\ x_i^j - \bar{x}^j \\ x_i^j - \bar{x}^j \end{bmatrix} \quad (\text{III.10})$$

Celui des coordonnées centrées de l'individu  $x^i$  est :

$$x_i^i = [x_i^1 - \bar{x}^1 \quad \dots \quad x_i^j - \bar{x}^j \quad \dots \quad x_i^p - \bar{x}^p] \quad (\text{III.11})$$

#### III.2.3.4. Rôle et choix de la métrique

Chaque individu étant un point défini par p coordonnées est considéré comme un élément d'un espace vectoriel F appelé nuage ou espace des individus. L'ensemble des n individus est alors un nuage de point dans F et g en est le centre de gravité (Saporta, 1990).

$$d^2 = (x_1^1 - x_2^1) + (x_1^2 - x_2^2) + \dots + (x_1^p - x_2^p) \quad (\text{III.12})$$

$$d^2 = \sum_{j=1}^p (x_i^j - x_j^j)^2 \quad (\text{III.13})$$

On utilisera donc la forme générale suivante : la distance entre deux individus  $e_i$  et  $e_j$  est définie par la forme quadratique :

$$d^2(e_i, e_j) = (e_i - e_j)M(e_i - e_j) \quad (\text{III.14})$$

$$\langle e_i, e_j \rangle = e_i' M e_j \quad (\text{III.15})$$

$$M = D_{1/S_j^2} = \begin{bmatrix} 1/S_1^2 & & & 0 \\ & 1/S_2^2 & & \\ & & \vdots & \\ 0 & & & 1/S_p^2 \end{bmatrix} \quad (\text{III.16})$$

Ce qui revient à diviser chaque individu par son écart type : entre autres avantages, la distance entre deux individus ne dépend plus des unités de mesures puisque les nombres  $x_i^j / S_j$  sont sans dimension, ce qui est très utile lorsque les variables ne s'expriment pas avec les mêmes unités.

### III.2.3.5. inertie totale du nuage des points

On appelle inertie totale du nuage de points la moyenne pondérée des carrés des distances des points au centre de gravité :

$$I_G = \sum_{i=1}^n p_i (e_i - G)' M (e_i - G) = \sum_i p_i \|e_i - G\|^2 \quad (\text{III.17})$$

L'inertie en un point a quelconque est définie par :

$$I_a = \sum_i p_i (e_i - a)' M (e_i - a) \quad (\text{III.18})$$

L'inertie totale est intéressante, car c'est une mesure de la dispersion du nuage des individus par rapport à son centre de gravité. Si cet inertie est grande, cela signifie que le nuage est très dispersé, tandis que si elle est petite, alors le nuage est très concentré sur son centre de gravité (Duby & Robin, 2006), on note aussi que l'inertie est égale à la somme des variances des p variables ; l'inertie totale est égale donc au nombre de variables, et ne dépend pas de leurs valeurs.

### III.2.4. Définitions algébrique

---

On est en présence d'un tableau *individus variables*, X (n, p), de n individus et p variables.

Le tableau X se présente comme suit :

$$[X] = \begin{bmatrix} X_{11} & X_{12} & X_{1j} & X_{1p} \\ X_{21} & X_{22} & X_{2j} & X_{2p} \\ X_{31} & X_{32} & X_{3j} & X_{3p} \\ X_{N1} & X_{N2} & X_{N3} & X_{Np} \end{bmatrix} \quad (\text{III.19})$$

Pour chaque variable on calcule la moyenne et l'écart type

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{III.20})$$

La variance est donnée par l'équation suivante :

$$\text{var}(x_i) = \sum_i p_i (x_i^j - \bar{x}_i)^2 \quad (\text{III.21})$$

L'écart type est la racine carrée de la variance :

$$S_j = \left[ \frac{1}{N} \sum_{i=1}^N (x_i^j - \bar{x}_j)^2 \right]^{1/2} \quad (\text{III.22})$$

La covariance de deux variables X, Y est donnée par :

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (\text{III.23})$$

Le coefficient de corrélation entre deux variables est :

$$r_{xy} = \frac{s_{xy}}{s_x s_y} \quad (\text{III.24})$$

Calculer les composantes principales notées  $C_j$  revient à déterminer N relations linéaires entre les variables  $X_j$  :

$$Z = \begin{cases} Z_1 = a_{10} + a_{11} X_1 + \dots + a_{1j} X_j + \dots + a_{1N} X_N \\ Z_2 = a_{20} + a_{21} X_1 + \dots + a_{2j} X_j + \dots + a_{2N} X_N \\ \dots \\ Z_j = a_{j0} + a_{j1} X_1 + \dots + a_{jj} X_j + \dots + a_{jN} X_N \\ \dots \\ Z_n = a_{n0} + a_{n1} X_1 + \dots + a_{nj} X_j + \dots + a_{nN} X_N \end{cases} \quad (\text{III.25})$$

$Z_j$  : j<sup>ème</sup> Composante Principale ;  
 $X_j$  : Vecteur variable initiale ;  
 $a_{ji}$  : Coefficient du système.

### III.2.5. Définition géométrique

### III.2.6. Formulation mathématique de l'ACP

Le but est de trouver le meilleur modèle réduit à n variables synthétiques qui représente au mieux l'ensemble des informations de la matrice. Une fois générée la matrice de variance – covariance entre les variables initiales, il suffit de rechercher les vecteurs propres associés aux valeurs propres de plus forts modules de cette matrice, pour déterminer les composantes principales de ce modèle optimal.

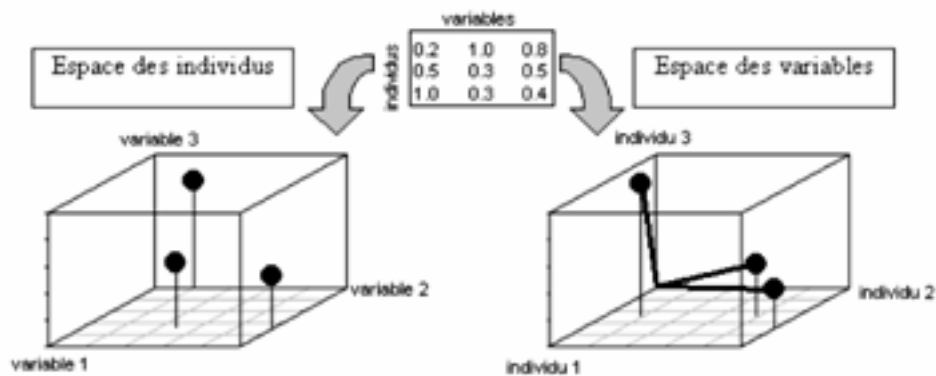


Figure III.1 Représentation géométrique des individus et variables.

Les étapes de L'ACP sont les suivantes :

1. Normalisation des données pour être indépendant des unités des P paramètres;
2. Calcul de la matrice de corrélation;
3. Recherche des éléments propres de la matrice de corrélation (C), qui donnent les axes principaux;
4. Représentation des individus dans le nouvel espace (en ne considérant que les valeurs propres expliquant une variance cumulée suffisante).

### III.2.7. Procédé d'application de l'ACP

### III.2.7.1. Calcul de la matrice de variance covariance

La matrice des covariances, notée  $[R]$  est la base de l'ACP, elle est obtenue en appliquant la relation suivante :

La matrice des covariances, notée  $[R]$  est la base de l'ACP, elle est obtenue en appliquant la relation suivante :

$$[R] = (1/n)[X]^t \cdot [M] \cdot [X] \quad (\text{III.26})$$

$[R]$  : Matrice de covariance de dimension  $(p \times p)$  ;

$[X]$  : Matrice de données ;

$[X]^t$  : Matrice transposée de  $[X]$  ;

$[M]$  : Métrique.

### III.2.7.2. Calcul des valeurs et vecteurs propres

Un scalaire  $\lambda$  est une valeur propre de la matrice  $A$ , si et seulement si, l'une des conditions équivalentes suivantes est vérifiée :

- Le rang de la matrice  $A - \lambda I$  est strictement inférieur à  $n$  ;
- Le déterminant de la matrice  $A - \lambda I$  est nul  $\text{Det}(A - \lambda I) = 0$ .

On dit que  $V$  est un vecteur propre de  $A$  associé à la valeur propre  $\lambda$  si  $V$  est un vecteur non nul et  $AV = \lambda V \leftrightarrow (A - \lambda I)V = 0$ .

Nous utiliserons la méthode de la puissance, pour la détermination des vecteurs et valeurs propres. Celle-ci permet de déterminer la plus grande valeur propre (en valeur absolue) de la matrice  $A$  et le vecteur propre associé. Partant d'un vecteur initial arbitraire  $x^{(0)}$  dont une composante est unitaire, on construit un vecteur intermédiaire  $y^{(1)}$  tel que  $A \cdot x^{(0)} = y^{(1)}$ , puis on construit  $x^{(1)}$  en "normalisant"  $y^{(1)}$  de manière à ce que la composante choisie soit à nouveau égale à l'unité.

Le polynôme  $|A - \lambda I| = 0$  ; est appelé polynôme caractéristique, donc, on aura par exemple la matrice  $A$  :

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad (\text{III.27})$$

En introduisant la notion de vecteur et valeur propres on aura

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{23} & a_{33} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \lambda \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad (\text{III.28})$$

Donc en introduisant la matrice unité on aura :

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{23} & a_{33} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \lambda \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad (\text{III.29})$$

$$\left[ \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{23} & a_{33} \end{pmatrix} - \lambda \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right] \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad (\text{III.30})$$

Finalement

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{23} & a_{33} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad (\text{III.31})$$

## III.3. Résultats obtenus par l'analyse en composantes principales ACP

### III.3.1. Présentation de la station d'étude

Avant l'année 1987, l'alimentation en eau potable du Grand Alger était assurée par les champs captants du Mazafran, de Baraki et du Hamiz. La quantité des eaux souterraines n'arrivaient plus à couvrir les besoins toujours grandissants des consommateurs. Pour combler le déficit en eau potable de la région algéroise, un nouveau système d'approvisionnement a été mis en service à partir de 1987, et qui consistait en la mobilisation des eaux superficielles des oueds Isser, Boudouaou et Berraba.

Le système, appelé SPIK (Système de Production Isser Keddara) se compose des ouvrages suivants (Figure III.2) :

- Barrage Béni Amrane, capacité 16 millions de m<sup>3</sup>.
- Barrage Keddara, capacité 145 millions de m<sup>3</sup>.
- Station de traitement, capacité 6.25m<sup>3</sup>/s.

- Station de pompage Béni Amrane, 7 m<sup>3</sup>/s
- Station de pompage Gué de Constantine, 2.2 m<sup>3</sup>/s
- Station de pompage Garidi, 1.2 m<sup>3</sup>/s.
- Station de pompage Tafourah, 2.2 m<sup>3</sup>/s.
- Réservoir Gué de Constantine, 25000 m<sup>3</sup>.
- Réservoir Garidi, 20000 m<sup>3</sup>.
- Réseau d'adduction et de distribution, 250 km,
- Diamètres de 500 mm à 2000 mm.



Figure III.2 Schéma générale du système SPIK et localisation de la station de traitement.

La station de traitement des eaux potables a été mise en service en 1987, cette station se situe à environ 7 km du barrage de Keddara, entre la ville de Boudouaou et d'Ouled Moussa. Elle occupe une superficie de 17 hectares, et fait partie du Système de Production Isser Keddara (SPIK). Elle traite les eaux des barrages de Béni Amrane, de Keddara et du Hamiz et alimente une population estimée à 4.000.000 d'habitants, avec une capacité de traitement de 540.000 m<sup>3</sup>/j (SEAAL, 2008).

Cette station de traitement compte : (i) un ouvrage d'arrivée et de mélange, (ii) une étape de clarification assurée par le procédé de coagulation-floculation grâce à des décanteurs de type «PULSATOR» lamellaire à lit de boue, utilisant le sulfate d'aluminium comme coagulant, (iii) des filtres type «AQUAZUR V», et enfin l'eau est stockée dans deux réservoirs de capacité totale 2 x 50 000 m<sup>3</sup> (Figure III.3) , avant qu'elle ne soit pompée vers la ville d'Alger (SEAAL, 2008).



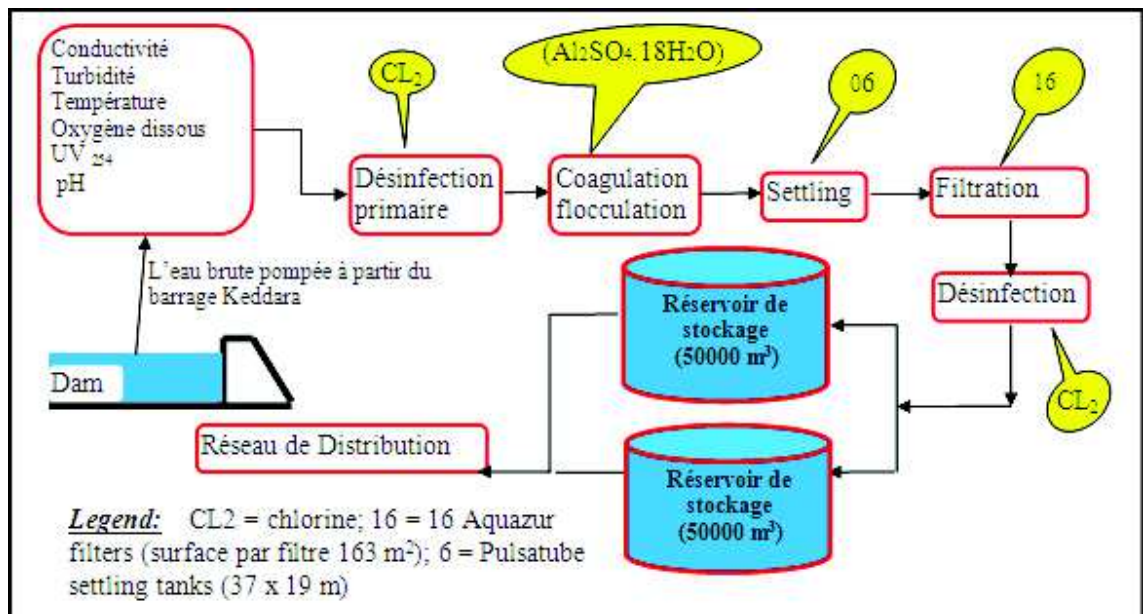


Figure III.3 Représentation de la station de traitement des eaux de Boudouaou.

### III.3.2. Analyse statistique

L'objectif de notre travail est la modélisation de la dose du coagulant (CO) en fonction des variables descriptives caractérisant l'eau brute à l'entrée de la station. Nous disposons pour cela de six variables : la température (TE), le pH, la turbidité (TU), la conductivité électrique de l'eau (CON), l'oxygène dissous (OX) et les ultraviolets (UV<sub>254</sub>). Ces six variables sont mesurées deux fois par jour, parallèlement la dose de coagulant est déterminée par les essais Jar-Test effectués au laboratoire. Les caractéristiques statistiques des variables retenues sont présentées dans le tableau III.1 ; avec  $x_{min}$  la valeur minimale,  $x_{max}$  la valeur maximale,  $x_{moy}$  la valeur moyenne,  $s_x$  l'écart type et  $C_{vx}$  la variance. Les évolutions des variations annuelles des différentes variables sont présentées dans les figures (III.4, 5, 6, 7, 8,9 et 10).

Tableau III.1 Statistiques des variables utilisées

Variabes	Unité	$x_{min}$	$x_{max}$	$x_{moy}$	$s_x$	$C_{vx}$
Turbidité	NTU	0.44	32.40	16.53	4.54	20.62
pH	/	7.23	8.60	7.76	0.25	0.06
Oxygène dissous	(mg/l)	0.14	13.20	4.72	2.98	8.89
Conductivité électrique	μS	668.00	1432.00	1009.66	122.26	14929.22
Température	°C	10.20	26.20	22.77	3.49	12.19
Les UV <sub>254</sub>	(do/m)	0.013	0.98	0.11	0.052	0.0027
Coagulant	(mg/l)	10.00	35.00	22.77	7.46	55.58

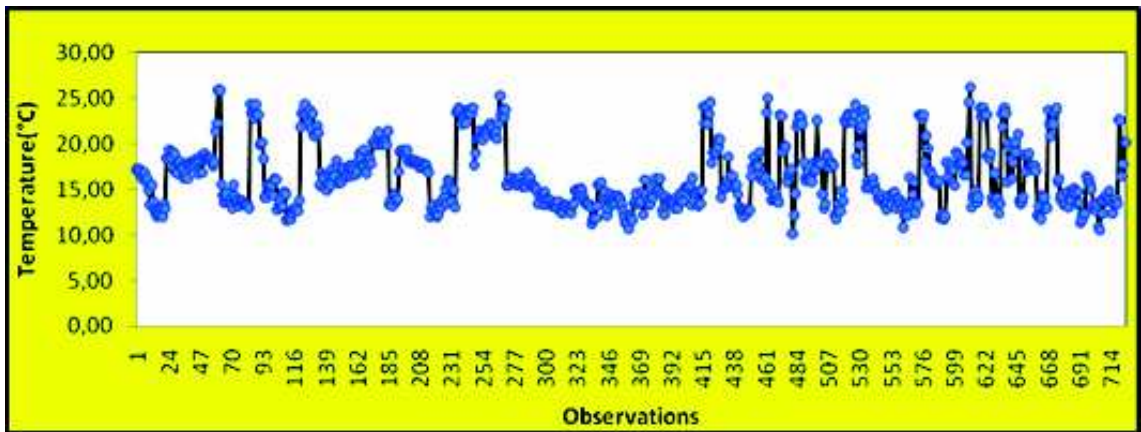


Figure III.4 Variation annuelle de la température

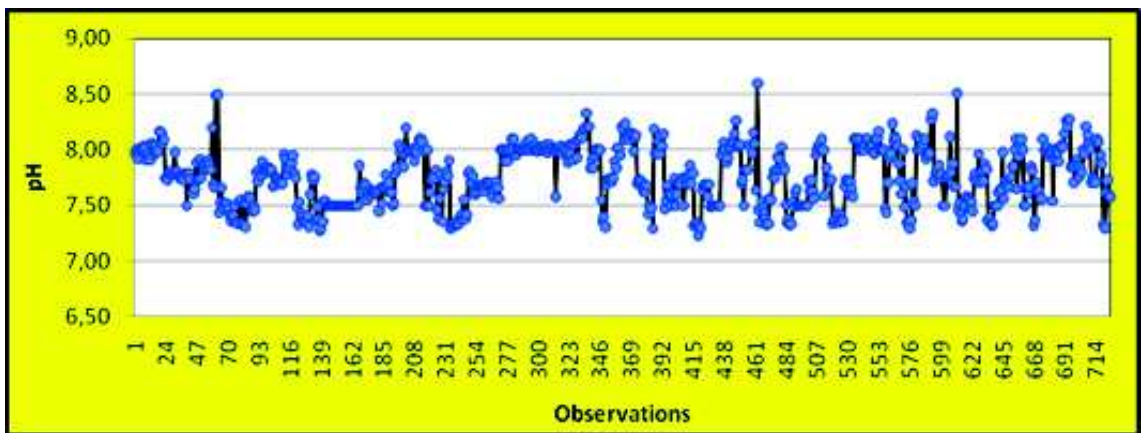


Figure III.5 Variation annuelle du pH

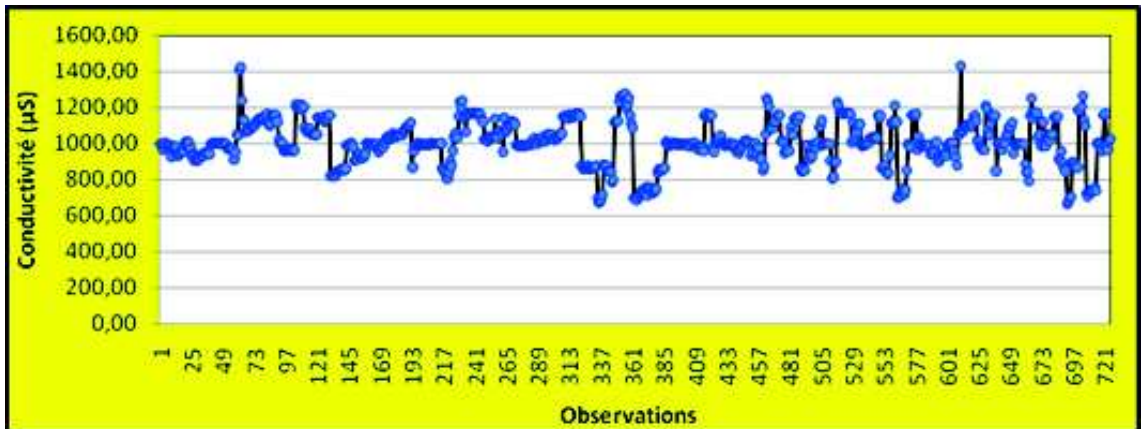


Figure III.6 Variation annuelle de la conductivité électrique

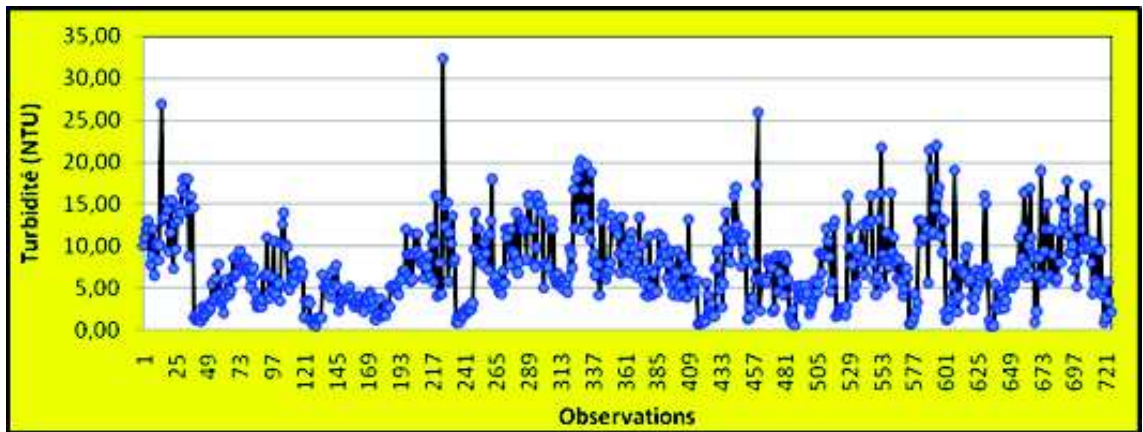


Figure III.7 Variation annuelle de la turbidité

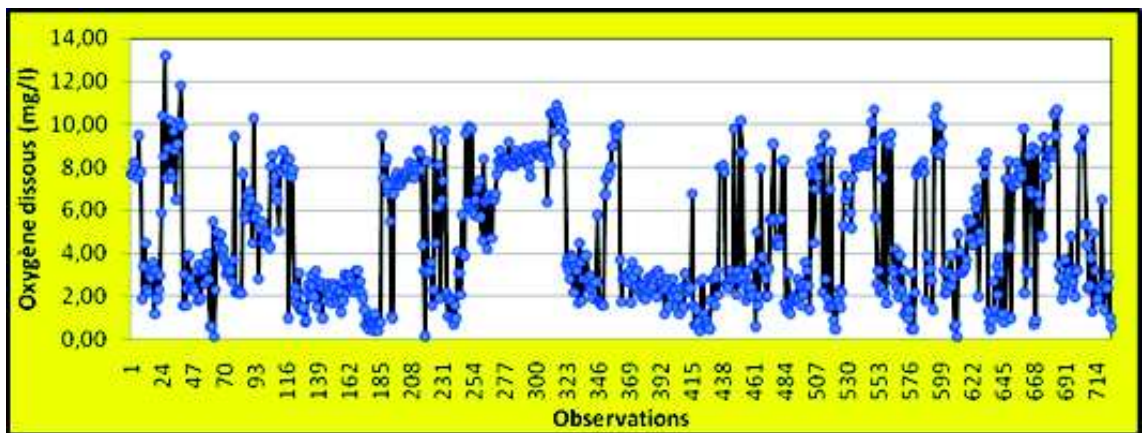


Figure III.8 Variation annuelle de l'oxygène dissous

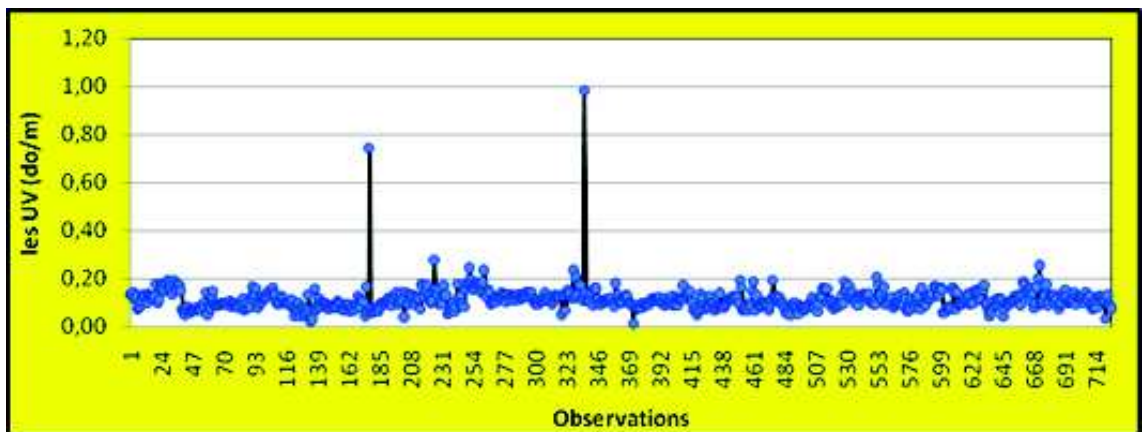


Figure III.9 Variation annuelle des UV<sub>254</sub>



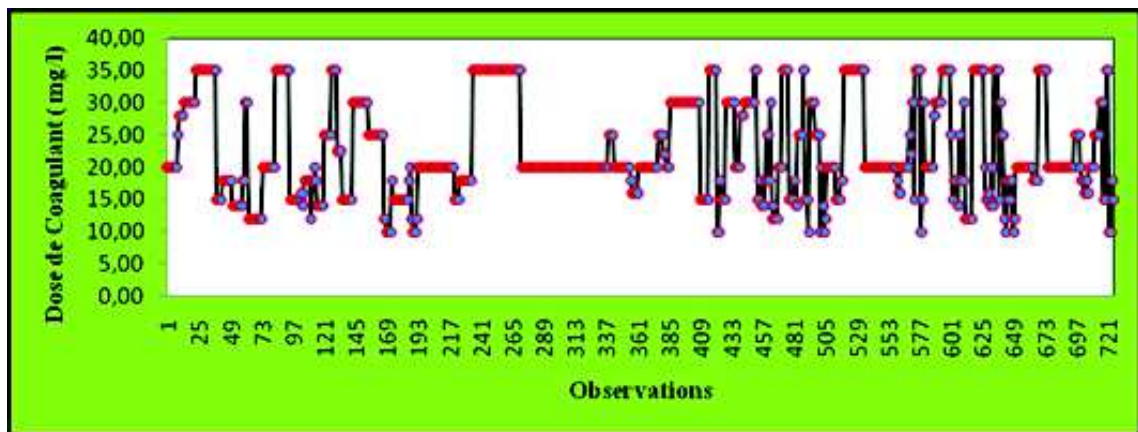


Figure III.10 Variation annuelle de la dose de coagulant

En premier lieu et compte tenu de la taille de la base de données utilisée ; nous avons procédé à la construction de la matrice de corrélation linéaire et le calcul du coefficient de corrélation (CC) (Tableau III.2), ce dernier concerne la relation entre les variables considérées deux à deux, il va nous permettre de déceler toutes les variables explicatives potentiellement intéressantes pour la construction de l'équation de régression linéaire multiple (RLM). Les résultats obtenus sont donnés dans le Tableau III.2.

Tableau III.2 Matrice de Corrélation

	CO	TE	pH	CON	TU	OX	UV <sub>254</sub>
CO	1.00						
TE	0.4917	1.00					
pH	-0.1989	-0.4334	1.00				
CON	0.1347	0.3336	-0.3350	1.00			
TU	-0.0032	-0.4893	0.5144	-0.2318	1.00		
OX	-0.0396	-0.2178	0.3779	0.2071	0.4358	1.00	
UV <sub>254</sub>	0.0791	-0.1244	0.2375	-0.1243	0.4451	0.2441	1.00

On remarque des faibles corrélations positives (en gras dans le Tableau III.2) entre certaines variables, notamment entre la dose de coagulant et la température, et la dose de coagulant et la conductivité. Ces coefficients sont de 0.491 (pour la « température ») et 0.134 (pour la « conductivité »). Il est à noter également que les deux variables température et conductivité sont corrélées entre elles avec un coefficient de 0.333. Ce phénomène est également observable pour les variables turbidité, oxygène dissous et les UV<sub>254</sub> ; ces dernières étant elles-mêmes corrélées à la variable « pH ». Enfin, une faible corrélation positive existe entre la turbidité et les UV<sub>254</sub> ainsi que l'oxygène dissous avec un coefficient de Spearman s'élevant à 0.445 et 0.435 respectivement. Ces résultats nous montrent d'une part, qu'il existe une certaine redondance d'information, comme on pouvait le pressentir, et d'autre part aucune variable ne présente une forte corrélation avec la dose de coagulant. Une analyse multidimensionnelle aura pour objectif l'étude simultanée des variables.

L'analyse en composantes principales (ACP) nous permet d'obtenir une vue d'ensemble sur les données, à savoir de déterminer s'il existe des sous-populations d'individus et comment sont reliées les variables prises simultanément. Nous conservons, pour la suite de l'analyse, les composantes principales qui représentent 90 % de la variance totale. L'analyse en composantes principales (ACP) (Lebart et al., 2006) permet également

de résumer les variables d'origine, par un plus petit nombre de variables qui correspondent à une combinaison linéaire des variables de base, en vue de les intégrer par la suite dans un modèle de régression linéaire multiple.

Dans ce qui suit nous présenterons les résultats détaillés issus de l'analyse en composantes principales, ainsi qu'une justification sur le choix du nombre de variables retenue pour construire le modèle à base de régression linéaire multiple (RLM)

### III.3.3. Résultats de l'application de l'ACP

L'ACP produit des combinaisons linéaires non-corrélées entre elles (CP) à partir des variables originelles. Les combinaisons linéaires (composantes principales) sont ordonnées en fonction de leur variance. Ainsi, les premières montrent l'essentiel de la variation (variance) présente dans les variables originelles. Autrement dit, les premières composantes principales ont les plus grandes corrélations multiples au carré ( $R^2$ ) avec chacune des variables originelles. Les CP peuvent être extraites à partir de la matrice de corrélation, ou de la matrice variance-covariance des variables originelles (Tableaux III.2 et III.3).

Tableau III.3 Matrice de variance covariance

	CO	TE	pH	CON	TU	OX	UV <sub>254</sub>
CO	55.659						
TE	12.820	12.213					
pH	-0.378	-0.386	0.065				
CON	122.848	142.539	-10.428	14949.844			
TU	-0.108	-7.772	0.595	-128.811	20.656		
OX	-0.883	-2.272	0.287	75.564	5.911	8.907	
UV <sub>254</sub>	0.031	-0.023	0.003	-0.791	0.105	0.038	0.003

Les vecteurs de pondération et les variances associés à chaque composante principale sont en fait les vecteurs propres et valeurs propres de la matrice utilisée (Tableaux III.4). Pour les besoins présents, la matrice de corrélation sert à extraire les CP.

Tableau III.4 Matrice des vecteurs et valeurs propres

Matrices des vecteurs propres							
N°	U1	U2	U3	U4	U5	U6	U7
1	-0.214	-0.550	0.461	0.417	-0.304	-0.028	-0.415
2	-0.462	-0.336	0.171	0.109	0.513	0.046	0.606
3	0.479	0.007	0.019	0.458	0.550	-0.487	-0.146
4	-0.264	-0.411	-0.627	-0.184	-0.110	-0.563	-0.083
5	0.497	-0.247	0.116	0.057	-0.490	-0.198	0.629
6	0.323	-0.448	-0.485	0.192	0.130	0.635	-0.057
7	0.299	-0.394	0.342	-0.729	0.271	-0.021	-0.184
Matrice des valeurs propres							
$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$	$\lambda_6$	$\lambda_7$	Somme
2.627	1.427	1.091	0.694	0.520	0.369	0.273	7.00

La première valeur propre représente la variance expliquée par la première composante principale ( $\lambda_1 = 2.627$ ), elle correspond à 37.52 % de la variance totale (Tableaux III.4 et III.5), et représente donc être l'axe prédominant. Il est expliqué par les variables pH, TU et les UV<sub>254</sub> (Tableau III.6). Il existe une corrélation positive entre ces variables et la première composante principale. Nous tiendrons compte par la suite de cette observation pour éviter ainsi la redondance d'information une fois intégré. Algébriquement, la première composante principale,  $y_1$ , est une combinaison linéaire de  $x_1, x_2, \dots, x_p$  tel que  $CP1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p$  et la variance de  $CP1$  est maximisée sous contrainte que la somme des pondérations au carré soit égale à un :

$$\sum_{i=1}^p a_{1i}^2 = 1 \quad (\text{III.32})$$

Dans notre cas nous avons :

$$CP_1 = -0.348*CO - 0.749*TE + 0.777*pH - 0.428*CON + 0.826*TU + 0.523*OX + 0.425*UV \quad (\text{III.33})$$

L'ACP établit le vecteur optimal de pondérations ( $a_{11}, a_{12}, \dots, a_{1p}$ ) et la variance associée de  $CP1$  qui est dénotée par  $\lambda_1$ .

Tableau III.5 Contribution et cumul des composantes principales %

Contributions des composantes principales							
CP <sub>1</sub>	CP <sub>2</sub>	CP <sub>3</sub>	CP <sub>4</sub>	CP <sub>5</sub>	CP <sub>6</sub>	CP <sub>7</sub>	Somme
37.52	20.38	15.58	9.91	7.42	5.27	3.9	100
Cumul de contributions des composantes principales							
CP <sub>1</sub>	CP <sub>2</sub>	CP <sub>3</sub>	CP <sub>4</sub>	CP <sub>5</sub>	CP <sub>6</sub>	CP <sub>7</sub>	Somme
37.52	57.90	73.48	83.89	90.81	96.08	99.98	100

La deuxième composante principale,  $CP_2$ , établit un deuxième vecteur de pondération ( $a_{21}, a_{22}, \dots, a_{2p}$ ) tel que la variance de  $CP_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p$  soit maximisée sous les contraintes que la composante principale est non corrélée avec la première composante principale et que :

$$CP_2 = -0.657*CO - 0.401*TE + 0.009*pH - 0.490*CON - 0.295*TU - 0.535*OX - 0.470*UV \quad (\text{III.34})$$

La composante principale  $CP2$  a la deuxième plus grande valeur propre ( $\lambda_2 = 1.427$ ). Elle représente 20.38 % de la variance totale et est construite autour de la variable OX avec un coefficient de corrélation de 0.53 en valeur absolue. Les deux composantes  $CP1$  et  $CP2$  expliquent 57.90 % de la variance totale. La figure (III. 11) représente le cercle de corrélation ( $CP1/CP2$ ).

Tableau III.6 Matrice de corrélation composante principales-variable d'origine

	CP1	CP2	CP3	CP4	CP5	CP6	CP7
CO	-0.348	-0.657	0.482	0.348	-0.219	-0.017	-0.216
TE	-0.749	-0.401	0.179	0.091	0.370	0.028	0.316
pH	0.777	0.009	0.020	0.381	0.396	-0.296	-0.076
CON	-0.428	-0.490	-0.654	-0.153	-0.079	-0.342	-0.043
TU	0.806	-0.295	0.121	0.047	-0.354	-0.120	0.328
OX	0.523	-0.535	-0.506	0.160	0.093	0.386	-0.030
UV <sub>254</sub>	0.485	-0.470	0.357	-0.607	0.196	-0.013	-0.096

La troisième composante principale, représente 15.58 % de la variance totale, est construite autour de la variable CON avec un coefficient de corrélation de 0.654, ainsi que la température avec un coefficient de corrélation de 0.179, ces deux dernières variables n'ont pas été prise en considération par les deux premières composantes principales, la variable UV<sub>254</sub> avec un coefficient de corrélation de 0.357, déjà prise en considération par la première composante principale. La figure (III.12) représente le cercle de corrélation (CP1/CP3), les trois composantes expliquent 73.48 % de la variance totale.

$$CP_3 = -0.482*CO+0.179*TE+0.020*pH-0.654*CON+0.121*TU-0.506*OX+0.357*UV \quad (III.35)$$

Le processus d'extraction des composantes principales se poursuit jusqu'à ce qu'il y ait autant de composantes principales que de variables. Les statistiques intéressantes issues d'une ACP sont les vecteurs de pondération des variables  $\mathbf{a} = (a_{j1}, a_{j2}, \dots, a_{jp})$ , associés à chaque composante principale (Tableau III.6), et leur variance,  $\lambda_j$  (Tableau III.4). Tous les vecteurs de pondération sont orthogonaux entre eux. On remarque que les cinq premières composantes principales, expliquent 90 % de la variance totale. Le portrait des pondérations des variables originelles sert à interpréter chaque composante principale alors que la variance associée indique quel pourcentage de la variance totale de l'ensemble des variables originelles chaque composante principale représente. L'importance relative des éléments dans le vecteur de pondération indique la contribution de chaque variable originelle à la variance de la composante principale ou, inversement, la part relative des variations dans les variables originelles expliquées par les composantes principales. Les corrélations des variables avec une composante principale sont proportionnelles aux éléments du vecteur de pondération associé.



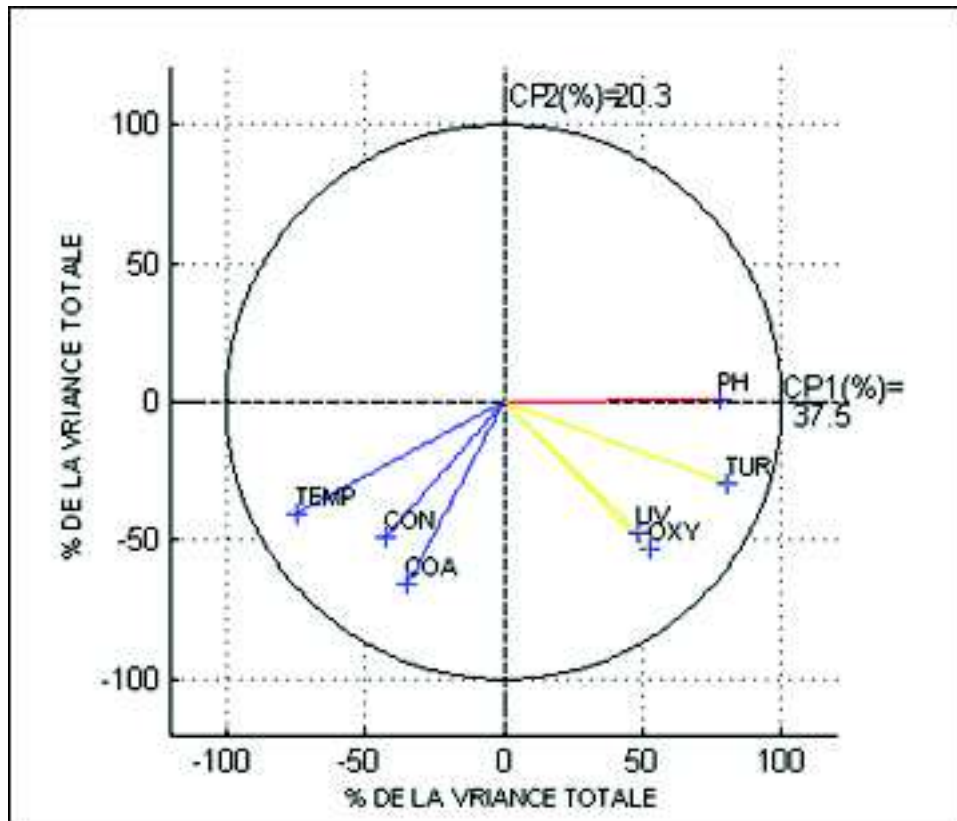


Figure III.11 Cercle de corrélation CP1/CP2

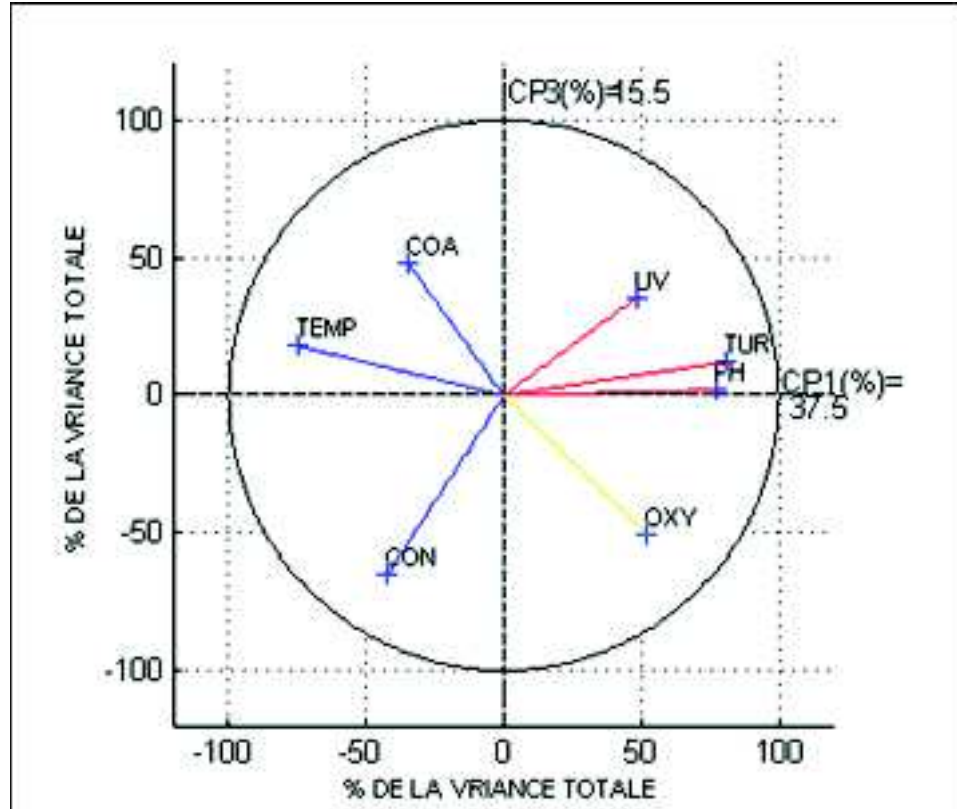


Figure III.12 Cercle de corrélation CP1/CP3

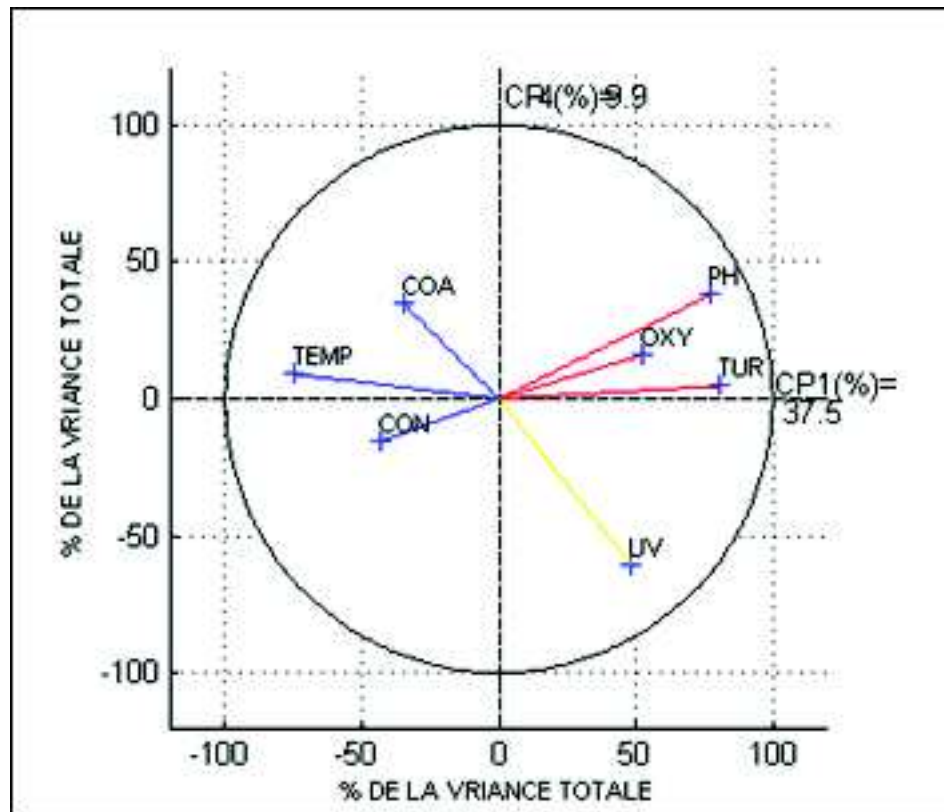


Figure III.13 Cercle de corrélation CP1/CP4

### III.4. Détermination de l'équation de prédiction par régression linéaire multiple

Étant donné que l'analyse en composantes principales n'a pas permis de réduire significativement le nombre de variables d'origine, par un plus petit nombre, une tentative de comparaison entre différentes combinaisons des variables d'entrée a été conduite. Plusieurs modèles ont été testés. Nous avons élaboré cinq (5) variantes de modèles (Tableau III.7); à savoir la variante V2 à deux variables d'entrées (15 modèles), V3 à trois variables d'entrées (18 modèles), V4 à quatre variables d'entrées (13 modèles), V5 à cinq variables d'entrée (6 modèles) et la variante V6 à six variables d'entrée (1 modèle). En totalité 53 modèles représentant cinq variantes ont été testés (Tableau III.8) et le meilleur modèle de chaque variante a été retenu, à savoir le modèle M2 utilisant TE et CON; le modèle M3 avec TE, CON et pH; le modèle M4 avec TE, CON, pH et TU; le modèle M5 avec TE, CON, pH, TU et OX et le modèle M6 avec TE, CON, pH, TU, OX et l'absorbance à 254 nm (UV<sub>254</sub>) comme entrées. Pour les cinq modèles cités la dose de coagulant représente la sortie du modèle (CO) (Heddami et Dechemi, 2008; Heddami et al., 2011d).

Tableau III.7 Structures des modèles utilisés

Variantes	Nombre de Modèles testés	Modèles retenus	Variables d'entrées						Variable de sortie
			TE	CON	pH	TU	OX	UV <sub>254</sub>	
V2	15	M02	I	I	O	O	O	O	CO
V3	18	M03	I	I	I	O	O	O	CO
V4	13	M04	I	I	I	I	O	O	CO
V5	6	M05	I	I	I	I	I	O	CO
V6	1	M06	I	I	I	I	I	I	CO

I : variable inclus, O: variable exclus.

Tableau III.8 Structure générale des variantes testées

Modèles M06	Température	Conductivité	pH	Turbidité	Oxygène Dissous	UV <sub>254</sub>
01	I	I	I	I	I	I
Modèles M05	Température	Conductivité	pH	Turbidité	Oxygène Dissous	UV <sub>254</sub>
01	I	I	I	I	I	O
02	I	I	I	I	O	I
03	I	I	I	O	I	I
04	I	I	O	I	I	I
05	I	O	I	I	I	I
06	O	I	I	I	I	I
Modèles M04	Température	Conductivité	pH	Turbidité	Oxygène Dissous	UV <sub>254</sub>
01	I	I	I	I	O	O
02	I	I	I	O	I	O
03	I	I	I	O	O	I

Tableau III.8 Structure générale des variantes testées (suite)

## Contribution à la modélisation de la qualité des eaux

04	I	I	O	I	I	O
05	I	I	O	I	O	I
06	I	I	O	O	I	I
07	I	O	I	I	I	O
08	I	O	I	O	I	I
09	I	O	O	I	I	I
10	O	I	I	I	I	O
11	O	I	I	O	I	I
12	O	I	I	O	I	I
13	O	O	I	I	I	I
<i>Modèles M03</i>	Température	Conductivité	pH	Turbidité	Oxygène Dissous	UV254
01	I	I	I	O	O	O
02	I	I	O	I	O	O
03	I	I	O	O	I	O
04	I	I	O	O	O	I
05	I	O	I	I	O	O
06	I	O	I	O	I	O
07	I	O	I	O	O	I
08	I	O	O	I	I	O
09	I	O	O	I	O	I
10	I	O	O	O	I	I
11	O	I	I	I	O	O
12	O	I	I	O	I	O
13	O	I	I	O	O	I
14	O	I	O	I	I	O
15	O	I	O	I	O	I
16	O	O	I	I	I	O
17	O	O	I	O	I	I
18	O	O	O	I	I	I
<i>Modèles M02</i>	Température	Conductivité	pH	Turbidité	Oxygène Dissous	UV254
01	I	I	O	O	O	O
02	I	O	I	O	O	O
03	I	O	O	I	O	O
04	I	O	O	O	I	O
05	I	O	O	O	O	I
06	O	I	I	O	O	O
07	O	I	O	I	O	O
08	O	I	O	O	I	O
09	O	I	O	O	O	I
10	O	O	I	I	O	O
11	O	O	I	O	I	O
12	O	O	I	O	O	I
13	O	O	O	I	I	O
14	O	O	O	I	O	I
15	O	O	O	O	I	I

I : variable inclus, O : variable exclus

### III.4.1. Construction des modèles de régression linéaire multiple

La modélisation à base de régression linéaire multiple consiste à chercher généralement une relation entre les variables indépendantes (ou explicatives), et les variables dépendantes (ou à expliquer). Il peut y avoir une simple variable explicative (régression simple) ou comme dans notre cas plusieurs variables (régression multiple) ; dans ce cas la formule de prédiction prend la forme générale représentée par la formule (III.1).

La construction d'un modèle par la régression linéaire multiple (*RLM*) passe par deux étapes : estimation des coefficients du modèle appelés coefficients de régression partielle, et la validation des résultats de la régression. Il existe de nombreux critères de validation des modèles utilisés dans les simulations et/ou dans les prévisions. Ces critères sont classés en deux groupes : un groupe des indicateurs graphiques et un groupe des mesures numériques. Du premier groupe, nous avons choisi deux critères, le premier concerne la comparaison entre les valeurs observées et celles prédites (en anglais Scatterplots), et le deuxième, le tracé des courbes de variation des doses prédites et celle des valeurs observés. Du deuxième groupe, nous avons sélectionné trois critères de mesure de l'erreur numérique: la Racine Carrée de l'Erreur Quadratique Moyenne (RMSE) (III.36), le Coefficient de Corrélation (CC) (III.37) et la moyenne des erreurs (MAE) (III.38)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - P_i)^2} \quad (III.36)$$

$$CC = \frac{\frac{1}{N} \sum (O_i - O_m)(P_i - P_m)}{\sqrt{\frac{1}{N} \sum_{i=1}^n (O_i - O_m)^2} \sqrt{\frac{1}{N} \sum_{i=1}^n (P_i - P_m)^2}} \quad (III.37)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |O_i - P_i| \quad (III.38)$$

Avec :  $O_i$  : Valeur de la dose de coagulant observée ;

$P_i$  : Valeur de la dose de coagulant calculée par le modèle ;

$N$  : Nombre total des observations (nombre d'individus) ;

$O_m$  : Moyenne de valeurs observées de la dose de coagulant ;

$P_m$  : Moyenne de valeurs calculées de la dose de coagulant.

La RMSE donne une indication quantitative sur l'erreur de simulation obtenue pendant la phase de modélisation. Elle mesure la déviation de prévision et/ou de simulation de la valeur réelle mesurée. Le Coefficient de Corrélation représente la proportion de variation de la dose de coagulant qui a été observée ou qui a été calculée par la régression linéaire. Les valeurs idéales pour RMSE et MAE sont 0, pour un Coefficient de Corrélation c'est l'unité.

### III.4.2. Résultats et interprétations

La base de données est scindée en trois parties : 60 % pour la partie calage (apprentissage), cette dernière doit être suffisamment représentative; 20 % pour la partie validation, et 20 % pour la partie test afin de juger la qualité prédictive du modèle. Il est à signaler que la répartition de la base de données se fait d'une façon aléatoire (Tableau III.9, III.10 et III.11).

**Tableau III.9 Statistiques des variables utilisées partie apprentissage**

Variables	Unité	x min	x max	x moy	s <sub>x</sub>	C <sub>vx</sub>
Turbidité	NTU	0.5	32.40	7.45	4.48	20.06
pH	/	7.23	8.50	7.75	0.24	0.0612
Oxygène dissous	(mg/l)	0.14	13.20	7.70	2.99	8.97
Conductivité électrique	μS	676	1428	1010.68	120.44	14473.80
Température	°C	10.70	25.90	16.51	3.42	11.66
Les UV <sub>254</sub>	(do/m)	0.0135	0.98	0.1139	0.062	0.0038
Coagulant	(mg/l)	10	35	22.73	7.35	54.03

**Tableau III.10 Statistiques des variables utilisées partie validation**

Variables	Unité	x min	x max	x moy	s <sub>x</sub>	C <sub>vx</sub>
Turbidité	NTU	0.54	26	7.23	4.46	19.77
pH	/	0.500	10.70	4.56	2.94	8.60
Oxygène dissous	(mg/l)	7.30	8.60	7.777	0.27	0.076
Conductivité électrique	μS	699	1252	1013.24	120.07	14317.76
Température	°C	10.20	25	16.68	3.61	12.98
Les UV <sub>254</sub>	(do/m)	0.04	0.204	0.1067	0.030	0.0009
Coagulant	(mg/l)	10	35	23.04	7.65	58.14

**Tableau III.11 Statistiques des variables utilisées partie test**

Variables	Unité	x min	x max	x moy	s <sub>x</sub>	C <sub>vx</sub>
Turbidité	NTU	0.44	22	8.27	4.76	22.51
pH	/	7.30	8.51	7.79	0.2453	0.0617
Oxygène dissous	(mg/l)	0.1430	10.80	4.96	2.98	8.87
Conductivité électrique	μS	668	1432	1003.02	130.24	16846.70
Température	°C	10.600	26.20	16.44	3.61	12.95
Les UV <sub>254</sub>	(do/m)	0.0310	0.255	0.1135	0.0325	0.0010
Coagulant	(mg/l)	10	35	22.61	7.61	57.56

Les techniques statistiques fonctionnent bien, quand chacune des variables d'entrée a le même poids relatif. Cependant dans notre cas d'étude les différentes variables sont de nature physique différente, caractérisée par des unités différentes, ce qui nous oblige à les normaliser. La normalisation vise à amener la plage d'évolution des valeurs entre 0 et 1.

Pour notre cas d'étude, nous avons normalisé les données en utilisant la formule suivante :

$$x_{ni,k} = \frac{x_{i,k} - m_k}{\sigma_k} \quad (\text{III.39})$$

Avec :  $x_{ni,k}$ : Valeur normalisé de la variable  $k$  pour l'individu  $i$  ;

$m_k$ : Moyenne de la variable  $k$  ;

$\sigma_k$ : Ecart type de la variable  $k$

Dans un premier temps, nous avons utilisé uniquement deux entrées à savoir la température et la conductivité électrique de l'eau conformément au (Tableau III.7). Pendant toutes les phases de calcul, nous avons été intéressés par (a) l'observation de la comparaison des graphiques issus de la validation et l'apprentissage du modèle, (b) l'évolution des résultats de la phase de test et (c) la racine carrée de l'erreur quadratique moyenne (RMSE) de la phase de test. Dans un deuxième temps, nous avons modifié le nombre de variable d'entrée du modèle en intégrant à chaque fois une nouvelle variable, et nous contrôlons à chaque fois (a) , (b) et (c) ; cette situation nous a permis de tester cinq variantes différentes (Tableau III.7). Les variables d'entrée varient de deux à six selon la variante. Dans chacune des variantes proposées, nous calculons les performances numériques (Tableau III.12) et comparons les performances graphiques. Le Tableau (III.13) illustre les coefficients de régression pour chaque variante.

**Tableau III.12 Comparaison des résultats de la modélisation par RLM**

Paramètres statistiques		Modèles				
		RLM-M02	RLM-M03	RLM-M04	RLM-M05	RLM-M06
Apprentissage	CC	0.50	0.50	0.57	0.57	0.57
	RMSE	6.33	6.33	5.99	5.99	5.99
	MAE	5.08	5.06	4.97	4.97	4.97
Validation	CC	0.47	0.47	0.54	0.54	0.54
	RMSE	6.72	6.73	6.38	6.39	6.39
	MAE	5.45	5.44	5.42	5.42	5.42
Test	CC	0.46	0.47	0.56	0.56	0.56
	RMSE	6.72	6.71	6.25	6.25	6.25
	MAE	5.41	5.38	5.24	5.24	5.24

On remarque d'après le Tableau (III.12), que les résultats obtenus par le modèle à base de régression linéaire multiple (RLM), sont très médiocres durant les trois phases de calcul à savoir la phase d'apprentissage, de validation et de test ; quelque soit le nombre de variables d'entrée utilisée, le coefficient de corrélation ne dépasse pas 0.57 , tandis que le RMSE avoisine 6.25 en phase de test pour les deux modèles à cinq et à six entrée. Malgré que les deux modèles à cinq et à six entrées présente des performances meilleures que les autres modèles (Figure III.14 et 15) et sont presque similaire de point de vue performances numériques , les résultats obtenus par la régression linéaire multiple sont loin d'être accepté, cela reflète clairement la complexité du problème et la forte non linéarité de la relation dose de coagulant, en fonction des variables descriptives de l'eau brute.

**Tableau III.13 Coefficients de régression pour les différentes variantes**



Modèle	Coefficients de Régression						
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$
RLM-M02	-0,0297	0.4954	-0.0084				
RLM-M03	-0.0296	0.5001	-0.0141	-0.0219			
RLM-M04	-0.0235	0.6049	-0.1454	-0.0224	0.3380		
RLM-M05	-0.0238	0.6036	-0.1405	-0.0159	0.3435	-0.0162	
RLM-M06	-0.0239	0.6020	-0.1408	-0.0150	0.3392	-0.0167	0.0087

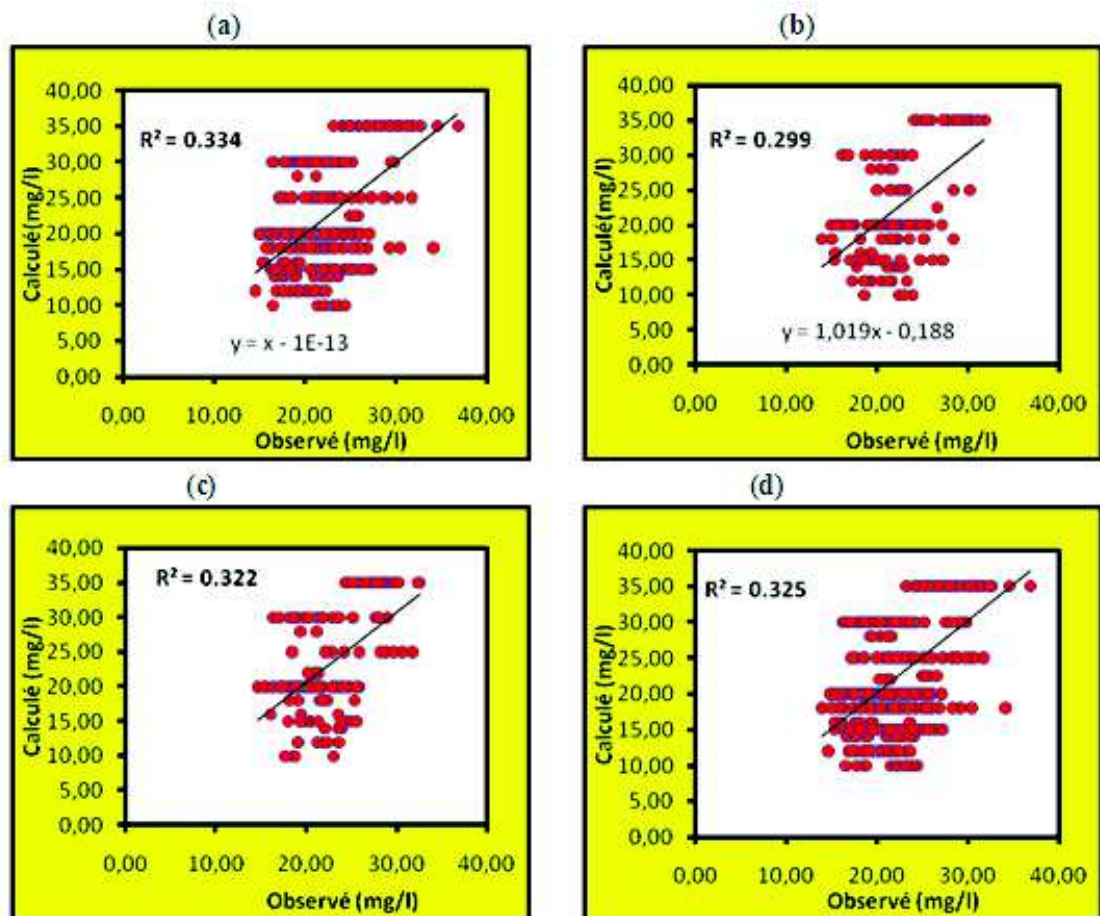


Figure III.14 Comparaison des résultats du modèle à base de régression linéaire multiple RLM (a) apprentissage, (b) validation, (c) test et (d) l'ensemble des données

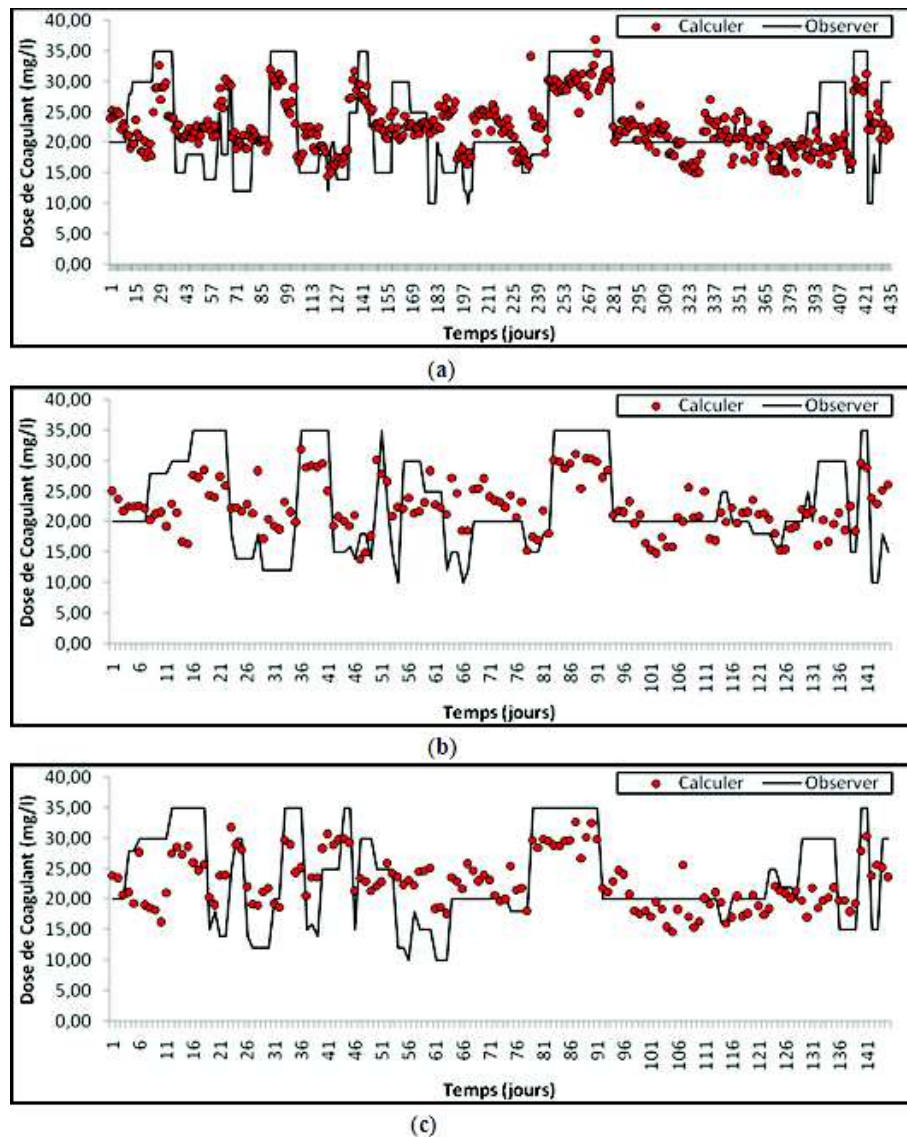


Figure III.15 Modélisation de la dose de coagulant par le modèle à base de régression linéaire multiple RLM (a) partie apprentissage, (b) partie validation et (c) partie test

### III.5. Conclusion et recommandations

Les résultats obtenus par le modèle à base de régression linéaire multiple (RLM), confirme la forte non linéarité de la relation dose de coagulant variables descriptives de l'eau brute à l'entrée de la station, à savoir la température (TE), le pH, la turbidité (TU), la conductivité électrique de l'eau (CON), l'oxygène dissous (OX) et les ultraviolets (UV<sub>254</sub>).

Malgré les différents modèles testés, aucune amélioration des résultats n'a été obtenue, cette situation nous a amené à nous orientés vers les modèles non linéaires en particuliers les réseaux de neurones, et les systèmes à base de logique floues.

# Chapitre IV Modèles A Base De Réseaux De Neurones Artificiels

## IV.1. Introduction

Les réseaux de neurones constituent une famille de fonctions non linéaires paramétrées, utilisées dans de nombreux domaines (physique, chimie, biologie, finance, etc.), notamment pour la modélisation de processus complexe. Les réseaux de neurones sont de bons outils pour modéliser des relations entre plusieurs variables, à partir d'observations expérimentales, sous réserve que celles-ci soient suffisamment nombreuses et représentatives.

Ce chapitre est consacré aux réseaux de neurones artificiels, en particulier les réseaux non bouclés à propagation en avant, une catégorie de modèles non linéaire utilisé dans divers domaines de recherche en particulier dans le domaine de l'ingénierie de l'eau (Maier et Dandy, 1996b; Gavin et al., 2005a, 2005b).

Du fait de leurs propriétés, les réseaux de neurones présentent un intérêt pour modéliser la dose de coagulant à utiliser en phase de coagulation-floculation dans une station de traitement d'eau potable: ils nécessitent relativement peu de données et n'exigent pas de connaissance préalable sur la forme de la relation susceptible d'exister entre la variable à expliquer et les variables explicatives. Ainsi, une telle approche est envisagée en complément de l'approche statistique classique avec deux objectifs : modéliser les données expérimentales à l'aide d'une relation non linéaire, et évaluer l'importance relative des différentes variables explicatives.

Dans un premier temps nous commençons par l'introduction de quelques propriétés des réseaux de neurones en particulier les architectures classiques rencontrées dans la littérature. Trois types de réseaux de neurones ont été retenus dans le cadre de ce travail :

1. Le perceptron multicouche (MLPNN)
2. Les réseaux de neurones à fonction de base radiale (RBFNN)
3. Les réseaux de neurones à fonction de régression générale (General Regression Neural Networks GRNN)

## IV.2. Les fondements biologiques

Le neurone est une cellule spécialisée dans le traitement et la transmission de l'information, ce qui est reflété par sa morphologie très particulière (Figure IV.1). Il se compose d'un corps cellulaire, le soma, et d'expansions : les dendrites d'une part, et l'axone d'autre part. Les dendrites forment des ramifications qui rentrent en contact avec d'autres neurones, typiquement de l'ordre de 10000, et dont le rôle est de recevoir des informations, électriques ou chimiques. L'axone est un prolongement de la cellule, typiquement long de quelques millimètres, qui conduit un signal électrique jusqu'à son arborisation terminale, où il peut

alors entrer en contact avec les dendrites d'autres neurones, la jonction axone dendrite est appelée synapse (Romain, 2003).

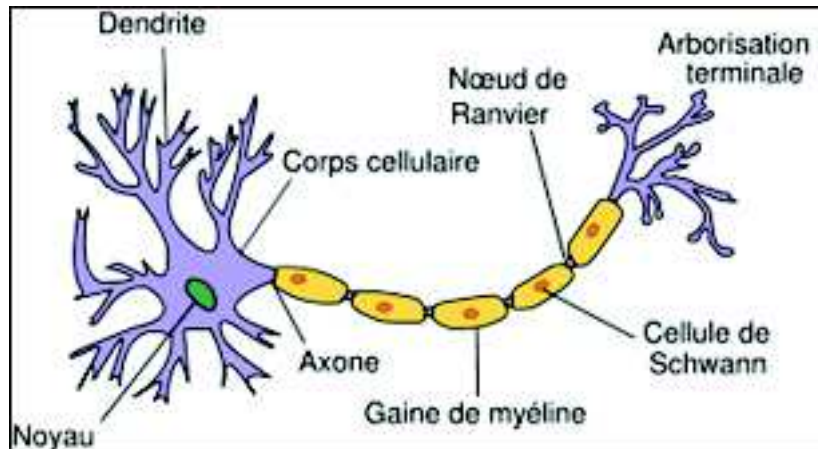


Figure IV.1 Représentation du neurone biologique

Les neurones se distinguent des autres cellules par leurs fonctions, en effet, des signaux se propagent entre les neurones, chaque neurone intègre dans son soma les signaux qu'il reçoit des autres neurones par l'intermédiaire des dendrites, l'activité du neurone est définie comme une pondération des signaux reçus, si cette activité est suffisante, il engendre un signal qu'il transmet à son tour aux autres neurones par son axone (Gaudier, 1999).

### IV.3. le neurone formel

L'étude biologique du système nerveux, a permis le passage des observations neurophysiologiques au neurone formel (Lippmann, 1987), ce dernier a été proposé par McCulloch et Pitts (1943), le modèle mathématique de neurone proposé a été repris par Rosenblatt pour définir le premier réseau de neurones artificiels, le Perceptron (Rosenblatt, 1962).

Le neurone formel est un modèle mathématique simplifié du neurone biologique. Il réalise une somme pondérée des potentiels d'action qui lui parviennent, chacun de ces potentiels est une valeur numérique qui représente l'état du neurone qui l'a émis, ces valeurs étant pondérées par des poids synaptiques qui sont définis lors d'un processus d'apprentissage puis s'active suivant la valeur de cette sommation pondérée, cette étape de sommation correspond à la collecte et à l'intégration de l'information.

Donc par comparaison au neurone biologique, on peut dire que la fonction de la synapse biologique (Figure IV.2) est modélisée par un poids de connexion modifiable qui est associé à chacune des connexions et le signal électrique dû à la dépolarisation du neurone est représenté par un niveau d'activation.

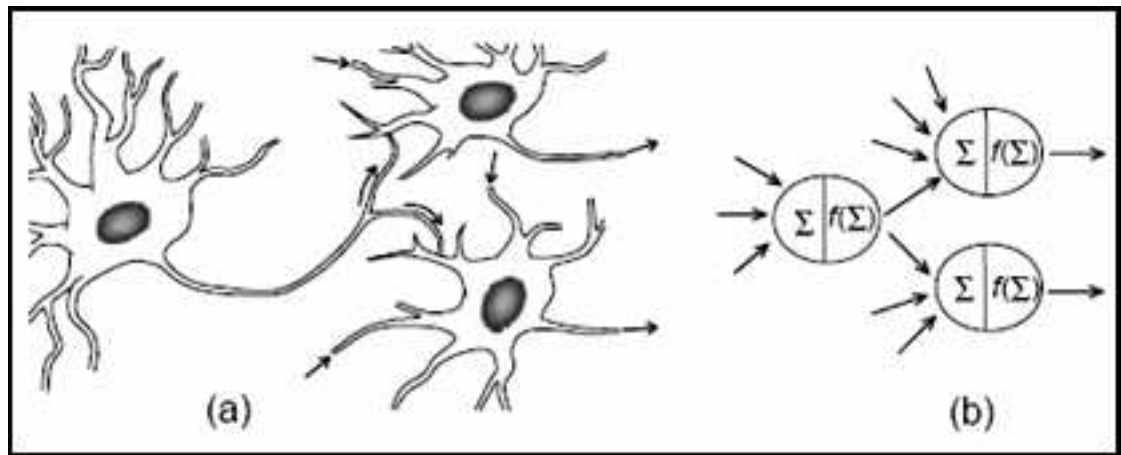


Figure IV.2 Communication entre les neurones (a) réseau de neurones Biologique, (b) modèle de réseau de neurone

Si cette somme dépasse un certain seuil, le neurone est activé et transmet une réponse (sous forme de potentiel d'action) dont la valeur est celle de son activation, si le neurone n'est pas activé, il ne transmet rien (Figure IV.3).

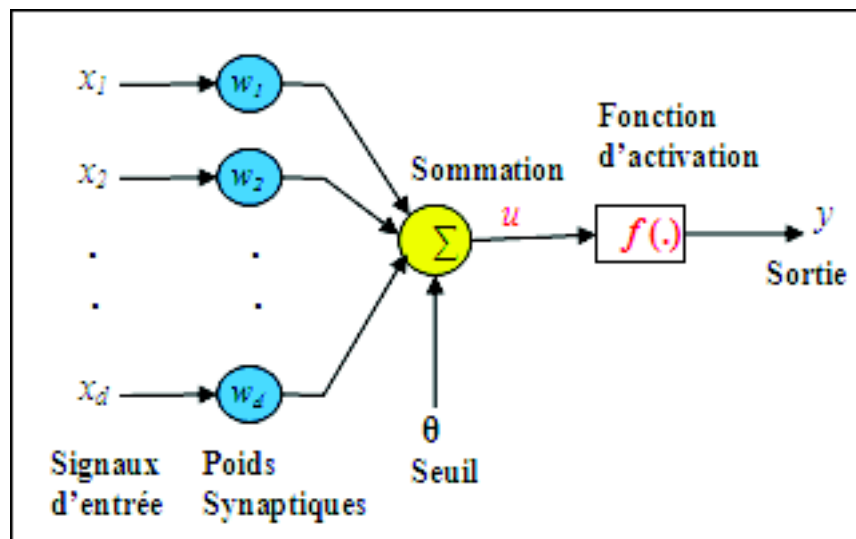


Figure IV.3 Le neurone formel

Le neurone reçoit les entrées  $x_1, x_2, x_3 \dots x_n$ , ou encore les excitations.  $w_1, w_2, w_3$  sont appelés les coefficients synaptiques ou poids synaptiques.

Le potentiel  $p$  est donné par  $A_i = x_1 * w_1 + x_2 * w_2 + x_3 * w_3 \dots + x_n * w_n$ . La fonction de sortie est  $y_i = f(A_i)$  ou  $f()$  est la fonction de transfert. Le neurone créé par Mc Culloch et Pitts était un neurone utilisant une fonction binaire (Figure IV.4) :

Si  $p > a$  alors  $S=1$  sinon  $S = -1$



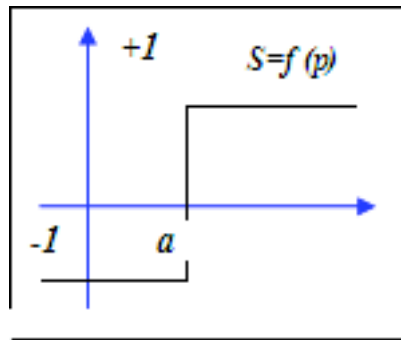


figure Iv.4 Fonction de transfert binaire

Les variables sur lesquelles opère le neurone sont habituellement désignées sous le terme d'entrées du neurone, et la valeur de la fonction sous celui de sortie ; il est commode (Dreyfus, 2004) de représenter graphiquement un neurone comme indiqué sur la figure (IV.2). Cette représentation est le reflet de l'inspiration biologique qui a été à l'origine de la première vague d'intérêt pour les neurones formels, dans les années 1940 à 1970 (Minsky et Papert, 1969).

Cette représentation simplifiée est à la base de la plupart des réseaux de neurones développés par la suite (Widrow et Hoff., 1960; Fukushima, 1975; Grossberg, 1976; Kohonen, 1982; Hopfield, 1982).

## IV.4.type de réseaux de neurones

Un neurone réalise simplement une fonction non linéaire, paramétrée, de ses variables d'entrée. L'intérêt des neurones réside dans les propriétés qui résultent de leur association en réseaux, c'est-à-dire de la *composition* des fonctions non linéaires réalisées par chacun des neurones. On distingue deux types de réseaux de neurones : les réseaux non bouclés et les réseaux bouclés (Dreyfus, 2004).

### IV.4.1. réseaux de neurones non bouclés

Un réseau de neurones non bouclé (Feedforward) est donc représenté graphiquement par un ensemble de neurones connectés entre eux, l'information circulant des entrées vers les sorties sans retour en arrière, il réalise une (ou plusieurs) fonctions de ses entrées, par composition des fonctions réalisées par chacun des neurones.

### IV.4.2. réseaux de neurones bouclés

Les réseaux bouclés ou récurrents, disposent d'un graphe de connexions *cyclique* : lorsqu'on se déplace dans le réseau en suivant le sens des connexions, il est possible de trouver au moins un chemin qui revient à son point de départ (un tel chemin est désigné sous le terme de « cycle »). La sortie d'un neurone du réseau peut donc être fonction d'elle-même; cela n'est évidemment concevable que si la notion de *temps* est explicitement prise en considération.

## IV.5. structure des réseaux de neurones

Les réseaux à couches sont les modèles connexionnistes les plus couramment utilisés, leur architecture, organisée en couches successives, comprend une couche d'entrée, une couche de sortie et une ou plusieurs couches intermédiaires appelées couches cachées, car elles ne sont pas vues de l'extérieur. Chaque couche est composée d'un certain nombre de neurones. Les connexions sont établies entre les neurones appartenant à des couches successives mais les neurones d'une même couche ne peuvent pas communiquer entre eux dans le cas des réseaux à couches.

### IV.5.1. Le perceptron et l'Adaline

---

Les premiers modèles de réseaux de neurones sont nés d'un compromis entre les principes théoriques énoncés par Mc Culloch et Pitts et les capacités des machines et des algorithmes de l'époque. Les premières architectures utilisées étaient donc monocellulaires, basées sur une cellule booléenne dont la fonction de décision était la fonction seuil (Herault, 1994). Il s'agit du Perceptron (Rosenblatt, 1958), qui est la forme la plus simple de réseau de neurones, ce réseau, capable d'apprendre à différencier des formes simples, et à calculer certaines fonctions logiques, est inspiré du système visuel, et de l'ADALINE (pour «ADAPtive LInear NETwork»)(Widrow et Hoff., 1960).

Ces deux réseaux ont été conçus pour des tâches de classification: la réponse de l'unique neurone, fonction des pondérations associées à chacun des paramètres d'entrée détermine l'appartenance à l'une ou l'autre des classes.

L'algorithme d'apprentissage utilisé pour le Perceptron est le suivant : en cas d'erreur en sortie le vecteur d'entrée correspondant est ajouté ou retranché au vecteur des poids synaptiques. L'ADALINE, par contre, utilise une règle plus complète car l'erreur est calculée non plus à partir de l'état binaire de la cellule mais à partir de la stimulation que la cellule aurait due recevoir en entrée (Figure IV.5). De plus cette méthode (dite « Apprentissage de Widrow-Hoff » ou « règle  $\Delta$  » «*Règle de Delta*») permet une réduction du gradient d'erreur à partir de la méthode des moindres carrés. L'erreur quadratique (E) à minimiser s'écrit comme suit :

$$E = (\Psi - \Psi_{désiré})^2 \text{ (erreur multivaluée dans le cas de l'Adaline)} \quad (IV.1)$$

$$E = (Y - Y_{désiré}) \text{ (erreur binaire dans le cas du perceptron)} \quad (IV.2)$$

On note : potentiel désiré  $\Psi_{désiré}$  pour l'Adaline, sortie désirée pour le perceptron  $Y_{désiré}$

La principale limitation de ces deux réseaux provient directement de leur structure : en 1968 Minsky et Papert ont montré, (Minsky et Papert, 1969), que les modèles de réseaux à une couche de connexions et à fonction d'activation linéaire sont incapables de distinguer deux classes si celles-ci ne sont pas linéairement séparables.



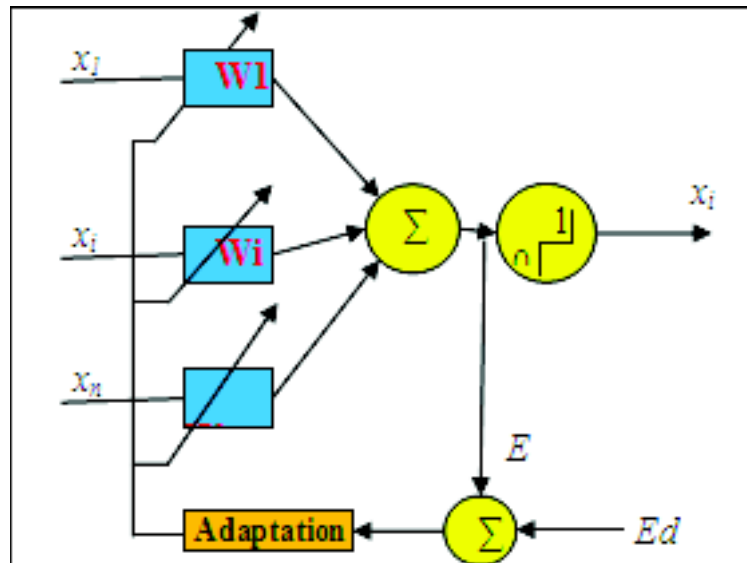


Figure IV.5 L'ADALINE de Widrow-Hoff

Une solution à cette limitation est les réseaux de neurones considérés comme des approximateurs universels dont les plus étudiés et appliqués en général sont les perceptrons multicouches (Bishop, 1995).

### IV.5.2. Le perceptron multicouches MLP

Les perceptrons multicouches sont les réseaux à la base des méthodes connexionnistes. Ils sont, en effet, les plus employés et les plus étudiés (Lippmann, 1987). Une abréviation anglaise est utilisée dans la littérature pour le nommer : *MLP* pour *Multi Layer Perceptrons* (Le Cun, 1987 ; Bishop, 1995 ; Haykin, 1994).

Un perceptron multicouche est composé de plusieurs couches de neurones et de connexions (Figure IV.6), ce nombre est au moins égal à deux, signifiant ainsi que le réseau possède deux couches de poids connexionnistes, une couche de sortie et une couche cachée.

Le nombre de couches cachées détermine la complexité des frontières des différents sous-espaces que le réseau pourra représenter (lippmann, 1987). La complexité de l'approximation est également déterminée par le nombre de neurones de chaque couche puisque ce nombre détermine le nombre maximal d'informations que le réseau peut extraire du signal traité (Makhoul et al., 1989 ; Murata et al., 1992). La couche d'entrée, correspondant le plus souvent à un vecteur de données, n'est pas véritablement considérée comme appartenant au réseau (Buniet, 1997).

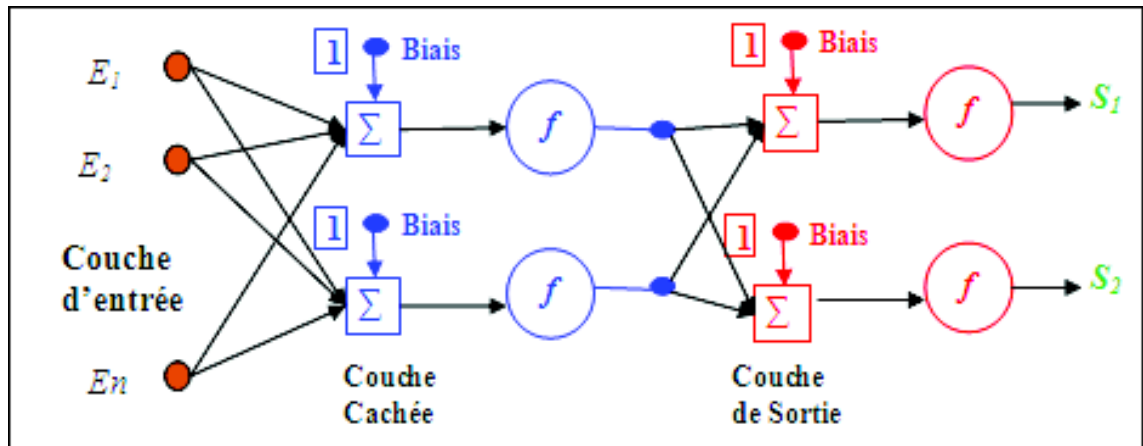


Figure IV.6 L'architecture générale du perceptron multicouche

Un neurone d'un perceptron multicouches se caractérise par trois concepts: son état, ses connexions avec d'autres neurones, et sa fonction d'activation.

$$s_i = f_i(A_i) = f_i\left(\sum_{j=1}^{n_i} w_{ij} s_j\right) + W_0 \quad (\text{IV.3})$$

Avec :

- S: L'ensemble des états possibles des neurones ;
- $S_i$ : L'état du neurone  $i$ , où  $S_i \in S$  ;
- $f_i$ : Fonction d'activation associée au neurone  $i$  ;
- $A_i$ : Activité du neurone  $i$  ;
- $W_{ij}$ : Poids de la connexion entre les neurones  $j$  et  $i$  ;
- $W_0$ : Le biais

La fonction  $f$  est la fonction d'activation du neurone, il existe plusieurs types de fonctions d'activation parmi lesquels :

La fonction  $f$  est la fonction d'activation du neurone, il existe plusieurs types de fonctions d'activation parmi lesquels :

- La fonction identité :  $f(A) = A$
- La fonction sigmoïde :  $f(A) = \tanh(A)$ . C'est une fonction bornée à valeurs réelles comprises entre -1 et +1 :

$$S_i = \frac{1}{1 + \exp(-A_i)} \quad (\text{IV.4})$$

La construction d'un réseau de neurones multicouches consiste à choisir le nombre de neurones et leurs fonctions de transfert, les connexions déterminées par les neurones, et à fixer les poids de ces derniers. L'attrait de cette approche réside dans le fait qu'en ne changeant que les poids des connexions, on peut utiliser une vaste gamme de fonctions différentes, et qu'il existe une méthode générique pour régler les poids des connexions, appelée règle d'apprentissage.

## IV.6. Apprentissage des réseaux de neurones artificiels

Dans le cas d'un cerveau humain, l'utilisation correcte des neurones nécessite au moins une phase d'apprentissage qui permettra d'avoir le fonctionnement escompté. Cet apprentissage consiste à jouer sur les synapses qui relient les neurones. De la même manière, les réseaux de neurones artificiels ont besoin d'apprendre. Il faudra la aussi agir sur les synapses, pour cela, on devra régler les valeurs des poids de chaque connexion entre neurones ainsi que sur les seuils (ou biais), c'est cette phase que l'on appelle l'apprentissage d'un réseau de neurones.

L'apprentissage donc est le processus d'adaptation des paramètres d'un système pour donner une réponse désirée à une entrée ou stimulation quelconque, l'apprentissage des poids connexionnistes est la partie la plus importante de la définition d'un réseau connexionniste. C'est cet apprentissage qui permet de mettre en adéquation l'architecture du réseau avec la tâche qui lui est présentée.

On distingue trois types d'apprentissages :

- L'apprentissage supervisé ;
- L'apprentissage non supervisé ;
- L'apprentissage par renforcement.

### IV.6.1. Apprentissage supervisé

La méthode classique pour l'apprentissage supervisé (supervised learning) consiste à se procurer un ensemble d'exemples, c'est-à-dire un ensemble fini de couples de vecteurs  $(x_j, y_j)$ . Dans de tels couples,  $x_j$  désigne l'entrée du réseau, et  $y_j$  la sortie désirée pour cette entrée. On écrit alors la fonction calculée par le réseau sous une forme paramétrique :  $f(x, w)$  désigne la sortie du réseau quand on lui présente en entrée le vecteur  $x$ , et qu'il utilise les poids synaptiques contenus dans la matrice  $w$ . Enfin, on se donne une distance sur l'espace vectoriel de sortie, c'est à dire un moyen de mesurer l'erreur commise en un point par le réseau. Si cette distance est notée  $d$ , on cherche alors à trouver la valeur de  $w$  qui minimise l'erreur totale commise par le réseau, c'est à dire la somme des distances entre les sorties obtenues et les sorties désirées, c'est à dire la somme des  $d(f(x_j; w); y_j)$ . Cette erreur est une fonction de  $w$  et on peut utiliser les techniques classiques d'optimisation de fonctions pour trouver son minimum (Lemaire, 1999).

### IV.6.2. Apprentissage non supervisé

Dans l'apprentissage *non supervisé*, il n'y a pas de notion de sortie désirée, on dispose seulement d'un nombre fini de données d'apprentissage, constituées "d'entrées", sans qu'aucun label n'y soit rattaché. Après initialisation aléatoire des poids, comme précédemment, une forme est présentée en entrée du réseau, l'apprentissage entre alors dans une phase de compétition, c'est la cellule dont le potentiel d'activation est le plus fort en fonction de l'entrée qui est choisie comme vainqueur (Buniet, 1997)

### IV.6.3. Apprentissage par renforcement

L'apprentissage renforcé est une technique similaire à l'apprentissage supervisé à la différence qu'au lieu de fournir des résultats désirés au réseau (Coulbaly et al., 1999) on lui accorde plutôt un grade (ou score) qui est une mesure du degré de performance du réseau après quelques itérations, les algorithmes utilisant la procédure d'apprentissage renforcé sont surtout utilisés dans le domaine des systèmes de contrôle (White et Sofge, 1992; Sutton, 1992)

## IV.7. la rétro propagation du gradient

### IV.7. 1. Généralités

---

Le problème d'apprentissage se résume donc à minimiser une fonction coût analogue à celles présentées par l'équation (IV.5) ou (IV.6). Cette minimisation s'effectue à l'aide de méthodes de descente du gradient. L'algorithme le plus connu, l'algorithme de rétropropagation du gradient (Le Cun, 1985 ; Rumelhart et al., 1986), un algorithme très simple et devenu célèbre, est basé sur l'idée suivante: au départ, les poids  $\{w_{ij}\}$  sont initialisés à des valeurs aléatoires, le but de la méthode d'apprentissage est de les faire évoluer de façon à ce que le réseau soit capable, étant donné un vecteur d'entrée, de calculer le bon vecteur en sortie.

### IV.7. 2. Fonctions coût

---

Dans le contexte des réseaux de neurones, l'apprentissage se fait de façon itérative, à chaque itération, on modifie les poids des synapses et les biais du réseau au moyen d'une technique de descente de gradient, on définit un critère à minimiser que l'on appelle le coût.

La fonction coût la plus fréquemment utilisée est la différence de l'erreur quadratique (MSE), cette dernière mesure l'écart entre les sorties effectives du réseau de neurones ( $s$ ) et les sorties désirées ou cibles ( $d$ ).

$$E_p = \frac{1}{2} \sum_k (d_k - s_k)^2 \quad (\text{erreur sur l'exemple } p) \quad (\text{IV.5})$$

$$E = \sum_{p=1}^P E_p \quad (\text{erreur totale sur la base d'apprentissage}) \quad (\text{IV.6})$$

### IV.7. 3. Choix d'initialisation des poids

---

Le choix initial des poids d'un réseau influence significativement le processus d'entraînement, les poids doivent être choisis aléatoirement. Dans le cas où les poids ont des valeurs numériques élevées, les sigmoïdes vont se saturer et engendrer des gradients faibles qui ralentiront l'apprentissage initial, généralement les poids doivent être choisis dans l'intervalle  $[0, 1]$  ou  $[-1, 1]$ .

#### IV.7. 4. Formulation mathématique

Supposant un réseau de neurones de type perceptron multicouches avec trois couches complètement interconnectée, avec (l) nombre de neurones dans la couche d'entrée, (m) nombre de neurones dans la couche cachée et (n) nombre de neurones dans la couche de sortie. A chacune de ces couches est attachée une matrice de poids [W],  $w_{ij}$  les poids ou paramètres du réseau entre la couche d'entrée et la couche cachée, d'une manière analogue,  $w_{jk}$  représente les poids entre la couche cachée et la couche de sortie.

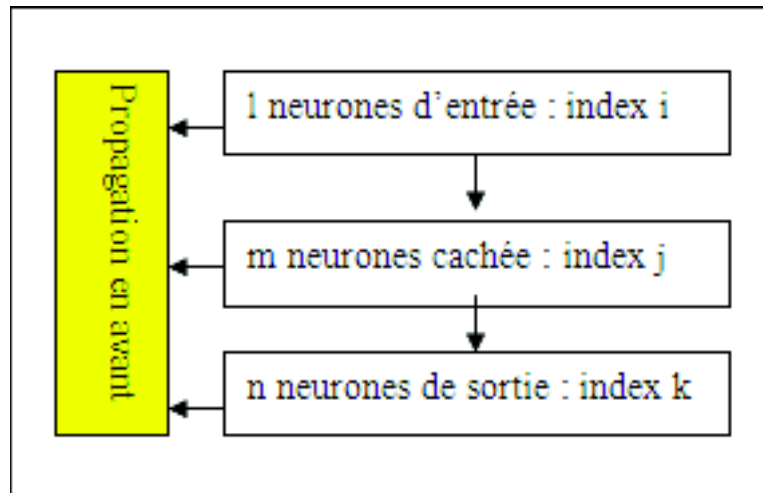


figure Iv.7: Propagation en avant de l'information dans un réseau de neurones à trois couches

Dans ce type de réseau, chaque neurone fonctionne comme un perceptron simple, c'est à dire qu'il effectue une transformation non-linéaire sur la somme pondérée de ses entrées (y compris un poids supplémentaire représentant le biais  $b_0$ ) :

$$X_j = \sum_{i=1}^l w_{ij} Y_i \quad (IV.7)$$

$$Y_j = f(X_j) \quad (IV.8)$$

$$X_k = \sum_{j=1}^m w_{jk} Y_j \quad (IV.9)$$

$$Y_k = f(X_k) \quad (IV.10)$$

$Y_j$  représente la  $j$ ème variable d'entrée (variables explicatives),  $f$  est appelée fonction d'activation.

Le problème pour un réseau à une ou plusieurs couches cachées, est d'évaluer l'erreur imputable au poids d'une connexion vers un neurone caché, la solution proposée par plusieurs auteurs (Werbos, 1974 ; Parker, 1985 ; Rumelhart et al., 1986 ; Le Cun ,1987) est

de "rétro-propager" l'erreur proportionnellement aux poids, la rétro propagation comprend donc deux phases :

- Présentation d'un exemple et propagation des activations vers la sortie (Figure IV.7);
- Calcul de l'erreur en sortie et propagation vers l'entrée (Figure IV.8).

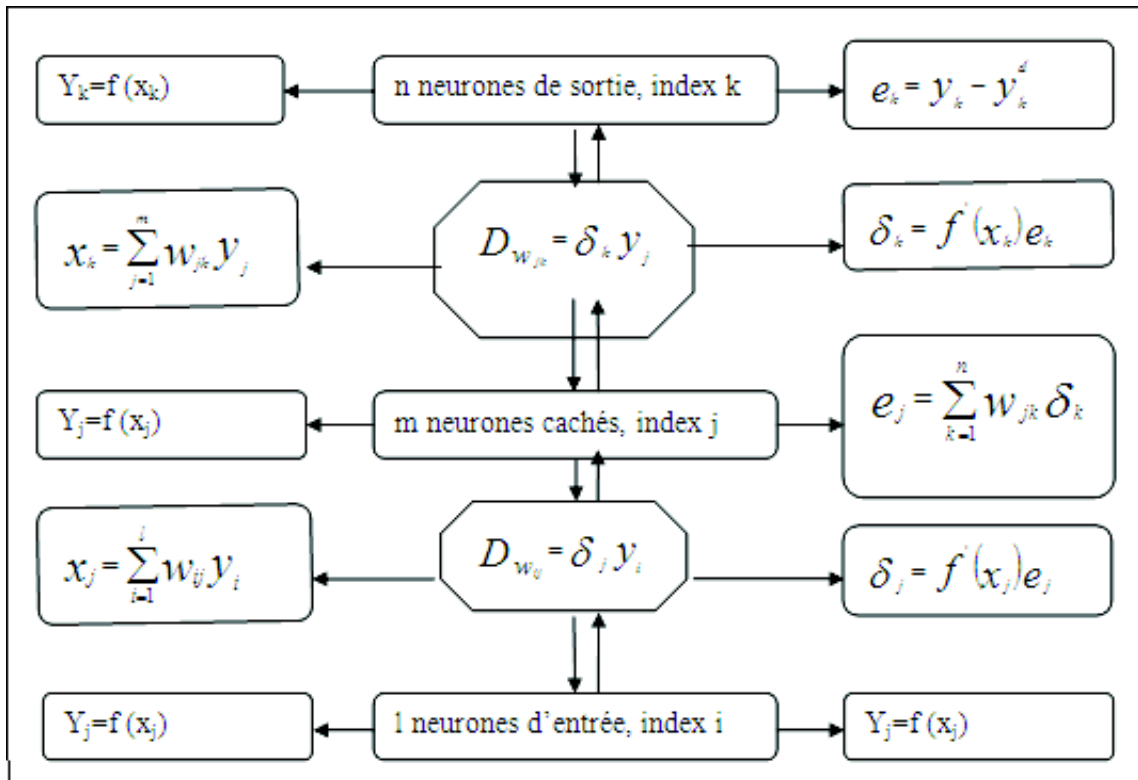


figure Iv.8 Rétro propagation en arrière de l'information dans un réseau de neurones à trois couches

L'algorithme de rétro propagation peut prendre deux formes : la première, dite rétro propagation continue ou stochastique (*on line*), consiste à modifier les poids après chaque présentation (Wilson et Martinez, 2003). Au contraire, la rétropropagation périodique ou par lots (*batch back propagation*) n'ajuste les poids qu'au bout d'un certain nombre de présentations (une époque), généralement égal au nombre d'exemples de la base d'apprentissage (Le Cun et al., 1990, 1998a, 1998b). Sous cette forme, on évite l'interférence entre les modifications successives des poids, en calculant le gradient sur la somme des erreurs obtenues pour chaque exemple du lot, on effectue une sorte de moyenne, cependant, si la base d'apprentissage est très importante, la redondance des informations rend la procédure périodique moins efficace que la procédure continue.

Supposant T le nombre de vecteurs disponibles pour l'apprentissage du réseau de neurones, après la présentation d'un vecteur t (t comporte les variables descriptives de 1 à j) au réseau, nous obtenons à la sortie du réseau un signal d'erreur représentant le cumul des erreurs de tous les neurones de sortie :

$$E(t) = \frac{1}{2} \sum_{k=1}^n (y_k(t) - y_k^d(t))^2 \quad (IV.11)$$

Avec :  $y_k(t)$  et  $y_k^d(t)$  correspondent respectivement à la valeur calculée et celle désirée de la variable à expliquer (ou à modéliser), par conséquent on obtient l'erreur moyenne d'un vecteur arbitraire  $t$  :

$$E_{moy} = \frac{1}{T} \sum_{t=1}^T E(t) \quad (IV.12)$$

L'algorithme de rétropropagation du gradient est un algorithme itératif; les poids sont modifiés à chaque étape selon la règle suivante:

$$w_i^j(t) = w_i^j(t-1) + \Delta(w_i^j) \quad (IV.13)$$

Les poids à l'itération  $t$  correspondent aux poids à l'itération  $(t-1)$  plus une correction dépendant du signal d'erreur, cette particularité justifie le nom de l'algorithme.

#### IV.7.4.1. De la couche de sortie vers la couche cachée

L'objectif principal de l'algorithme de la rétropropagation du gradient est la minimisation de l'erreur moyenne  $E_{moy}$ , par l'ajustement des poids (paramètres) du réseau, pour accomplir ceci on doit tout d'abord calculer les dérivées partielles de l'équation d'erreur (IV.12) par rapport à  $w_{jk}$ , on obtient :

$$\frac{\partial E_{moy}}{\partial w_{jk}} = \frac{1}{T} \sum_{t=1}^T (y_k - y_k^d) \frac{\partial y_k}{\partial x_k} \frac{\partial x_k}{\partial w_{jk}} \quad (IV.14)$$

$$\frac{\partial E_{moy}}{\partial w_{jk}} = \frac{1}{T} \sum_{t=1}^T (y_k - y_k^d) f'(x_k) y_j \quad (IV.15)$$

On trouve :

$$\delta_k = (y_k - y_k^d) f'(x_k) \quad (IV.16)$$

On remplace  $\delta_k$  dans la formule (IV.15) on aura :

$$\frac{\partial E_{moy}}{\partial w_{jk}} = \frac{1}{T} \sum_{t=1}^T \delta_k y_j \quad (IV.17)$$

Dans ce cas les poids du réseau sont modifiés de la façon suivante :

Donc



$$\Delta w_{jk} = -\eta \frac{\partial E_{msy}}{\partial w_{jk}} \quad (\text{IV.18})$$

$$\Delta w_{jk} = -\eta \frac{1}{T} \sum_{i=1}^T \delta_k y_j \quad (\text{IV.19})$$

Où est une constante positive appelée : le de gradient.

#### IV.7.4.2. De la couche cachée vers la couche d'entrée

Du moment que le réseau est complètement interconnecté, un ensemble de k:j dérivées partielles existe, ces dérivées (équation IV.19) constituent le cumul du gradient d'erreur de la couche cachée vers la couche de sortie, d'une façon analogue les dérivées partielles de l'équation (IV.12) par rapport à  $w_{ij}$  s'écrivent :

$$\frac{\partial E_{msy}}{\partial w_{ij}} = \frac{1}{T} \sum_{i=1}^T \sum_{k=1}^n (y_k - y_k^d) \frac{\partial y_k}{\partial x_k} \frac{\partial x_k}{\partial y_j} \frac{\partial y_j}{\partial x_j} \frac{\partial x_j}{\partial y_{ij}} \quad (\text{IV.20})$$

$$\frac{\partial E_{msy}}{\partial w_{ij}} = \frac{1}{T} \sum_{i=1}^T \sum_{k=1}^n (y_k - y_k^d) f'(x_k) w_{jk} f'(x_j) y_i \quad (\text{IV.21})$$

On introduit le terme du gradient ( $\delta_k$ ) dans la formule (IV.21), on trouve :

$$\frac{\partial E_{msy}}{\partial w_{ij}} = \frac{1}{T} \sum_{i=1}^T \sum_{k=1}^n \delta_k w_{jk} f'(x_j) y_i \quad (\text{IV.22})$$

Pour simplification on définit un autre terme de gradient

$$\delta_j = f'(x_j) \sum_{k=1}^n w_{jk} \delta_k \quad (\text{IV.23})$$

On remplace  $\delta_j$  dans la formule (IV.22), on aura :

$$\frac{\partial E_{msy}}{\partial w_{ij}} = \frac{1}{T} \sum_{i=1}^T \delta_j y_i \quad (\text{IV.24})$$

Dans ce cas, les poids du réseau sont modifiés de la façon suivante :

$$\Delta w_{ij} = -\eta \frac{\partial E_{msy}}{\partial w_{ij}} \quad (\text{IV.25})$$

Donc

$$\Delta w_{ij} = -\eta \frac{1}{T} \sum_{i=1}^T \delta_j y_i \quad (\text{IV.26})$$

Cet algorithme, possède de nombreuses variantes (Riedmiller et Braun ,1993).

Pour accélérer la convergence de la méthode du gradient, Plaut et al., (1986) proposent d'ajouter un terme d'inertie dans la formule de mise à jour des poids :

$$\Delta w_{ij}(t) = -\eta \frac{1}{T} \sum_{i=1}^T \delta_j y_i + \mu \Delta w_{ij}(t-1) \quad (\text{IV.27})$$

$$w_{ij}(t) = w_{ij}(t-1) + (-\eta \frac{1}{T} \sum_{i=1}^T \delta_j y_i + \mu \Delta w_{ij}(t-1)) \quad (\text{IV.28})$$

Le scalaire  $\mu$  est appelé moment (*momentum*), et prend ses valeurs dans  $[0, 1]$  (généralement  $\mu = 0.1$ ).

Le calcul des erreurs pour une couche donnée fait intervenir les erreurs de la couche suivante, d'où le nom de rétropropagation de l'erreur, il y a exactement autant d'erreurs locales qu'il y a de neurones dans le réseau, l'expression pour la mise à jour des poids prend alors la forme usuelle.

La rétropropagation implique l'entraînement du réseau sur des exemples pour lequel les valeurs correctes de sortie sont connues, et procède de manière itérative par un mécanisme (essai erreur correction) à partir d'un ensemble de poids aléatoires de départ. A chaque itération, le réseau ajuste les poids dans la direction qui réduit l'erreur (c'est à dire l'écart entre la sortie du perceptron multicouches (MLP), et les comportements observés de la fonction que l'on cherche à reproduire).

Pour cela, le signal d'erreur est rétropropagé à travers le réseau comme illustré par la (Figure IV.8), la propagation de ces informations se fait dans le sens inverse du calcul de la réponse à un motif d'entrée donné, et les changements dans les poids sont effectués entre chacune des couches, les poids convergent alors graduellement vers leur valeur optimale et un nombre important d'itérations est nécessaire pour un traitement complet.

## IV.7. 5. Variantes de l'algorithme

Pour déterminer une solution  $W^*$  qui minimise la fonction coût, la méthode précédente utilise uniquement l'information donnée par la dérivée du premier ordre de la fonction d'erreur. Or il est connu en optimisation classique que l'utilisation des informations données par les dérivées d'ordre supérieur peut accélérer le processus de recherche d'une solution optimale (tran, 1999).

Les méthodes de gradient peuvent être réparties en deux catégories: les méthodes du premier ordre, qui n'utilise que le gradient de la fonction et dont nous venons de détailler plus haut avec l'algorithme de rétropropagation du gradient, et les méthodes du second ordre, qui généralisent la descente du gradient au deuxième degré de la fonction d'erreur, ce sont des méthodes itératives de descente du gradient qui consistent à remplacer la fonction de coût par son approximation quadratique au voisinage de point courant :

$$Q(s) = E(w_k) + g(w_k)^T s + \frac{1}{2} s^T G(w_k) s \quad (\text{IV.29})$$

E : désigne la fonction de coût.

$$w^{k+1} - w^k = S \quad (\text{IV.30})$$

Avec :

$$g(w^k) = \nabla_w J \quad \text{Gradient} \quad (\text{IV.31})$$

$$G(w_k) = \nabla_w^2 J \quad \text{Hessien} \quad (\text{IV.32})$$

w désigne la matrice de l'ensemble des poids des couches du réseau, et la variable k désigne la k<sup>ème</sup> itération.

On peut citer par exemple les méthodes de Newton, de quasi-Newton de « Broyden-Fletcher-Goldfarb-Shanno » ou (BFGS), du gradient conjugué et de Levenberg-Maquardt.

La première catégorie de méthodes est la plus simple mais aussi la moins performante, notamment par la lenteur de sa convergence, une amélioration sensible de la vitesse de convergence est apportée par l'utilisation des méthodes du second ordre.

#### IV.7. 5.1. algorithme de Newton

La méthode de Newton consiste à calculer  $w_{k+1}$  de manière à minimiser Q(s) de l'équation (iv.29) , la formule itérative se déduit telle que :

$$w_{k+1} = w_k - [G(w_k)]^{-1} g(w_k) \quad (\text{IV.33})$$

Le minimum existe, si le Hessien G est défini positif, la méthode de Newton nécessite le calcul du vecteur gradient et de l'inverse du Hessien de la fonction de coût.

Dans ce cas :

$$p_k = -G_x^{-1} g_k \quad \text{et} \quad \alpha_k = 1 \quad (\text{IV.34})$$

Pour améliorer les propriétés de convergence de la méthode de Newton, de nombreuses méthodes vont porter sur la reformulation de l'équation (IV.34) :

$$w_{k+1} - w_k = s_k \quad \text{et} \quad s_k = \alpha_k \cdot p_k \quad (\text{IV.35})$$

#### IV.7. 5.2. algorithme de quasi Newton

Le principe des méthodes quasi-newtoniennes consiste en une généralisation de la méthode de Newton de l'équation (IV.34) :

$$P_* = -H_*^{-1} g_* \quad (\text{IV.36})$$

Le Hessien (ou le gradient) est approximé par une méthode itérative.

### IV.7.5.3. algorithme de Levenberg-Maquardt

La méthode de Levenberg-Maquardt (Levenberg, 1944 ; Maquardt, 1963) est un standard pour l'optimisation de l'erreur quadratique due à ses propriétés de convergence rapide et de robustesse, elle s'appuie sur les techniques des moindres carrés non linéaires et de l'algorithme de Gauss-Newton à voisinage restreint.

## IV.8. Propriété fondamentale des réseaux de neurones non bouclés

### IV.8.1. L'approximation universelle

Toute fonction bornée suffisamment régulière peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire (Hornik et al., 1989 ; Hornik et al., 1990 ; Hornik, 1991 ; Cybenko, 1989 ; Funahashi, 1989).

### IV.8.2. La parcimonie

Dans la pratique, le nombre de fonctions nécessaires pour réaliser une approximation est un critère important dans le choix d'un approximateur. Le concepteur de modèle doit toujours faire en sorte que le nombre de paramètres ajustables soit le plus faible possible, on dit que l'on cherche l'approximation la plus *parcimonieuse* (Dreyfus, 2004).

Barron, (1993) et Hornik (1994) montrent que si l'approximation dépend des paramètres ajustables de manière non linéaire, elle est plus parcimonieuse que si elle dépend linéairement des paramètres.

Plus précisément, on montre que le nombre de paramètres, pour une précision donnée, croît exponentiellement avec le nombre de variables dans le cas des approximateurs linéaires par rapport à leurs paramètres, alors qu'il croît linéairement avec ce nombre pour les approximateurs non linéaires par rapport à leurs paramètres. La parcimonie est donc d'autant plus importante que le nombre d'entrées du modèle est grand : pour un modèle à 1 ou 2 entrées, on peut utiliser indifféremment un modèle linéaire par rapport à ses paramètres (polynôme par exemple) ou un modèle non linéaire par rapport à ses paramètres (réseau de neurones par exemple) (Dreyfus, 2004).

## IV.9. Les réseaux de neurones à fonctions de base radiales (rbfNN)

### IV.9.1. Généralités

Les réseaux de neurones à fonctions de base radiales (rbfNN) : RBFNN (Radial Basis Functions) sont des réseaux de type *feedforward*, avec une seule couche cachée (Figure IV.9). Ils forment une classe particulière de réseaux de neurones. Ils trouvent leur origine dans des techniques d'interpolation d'un ensemble de points dans un espace multidimensionnel.

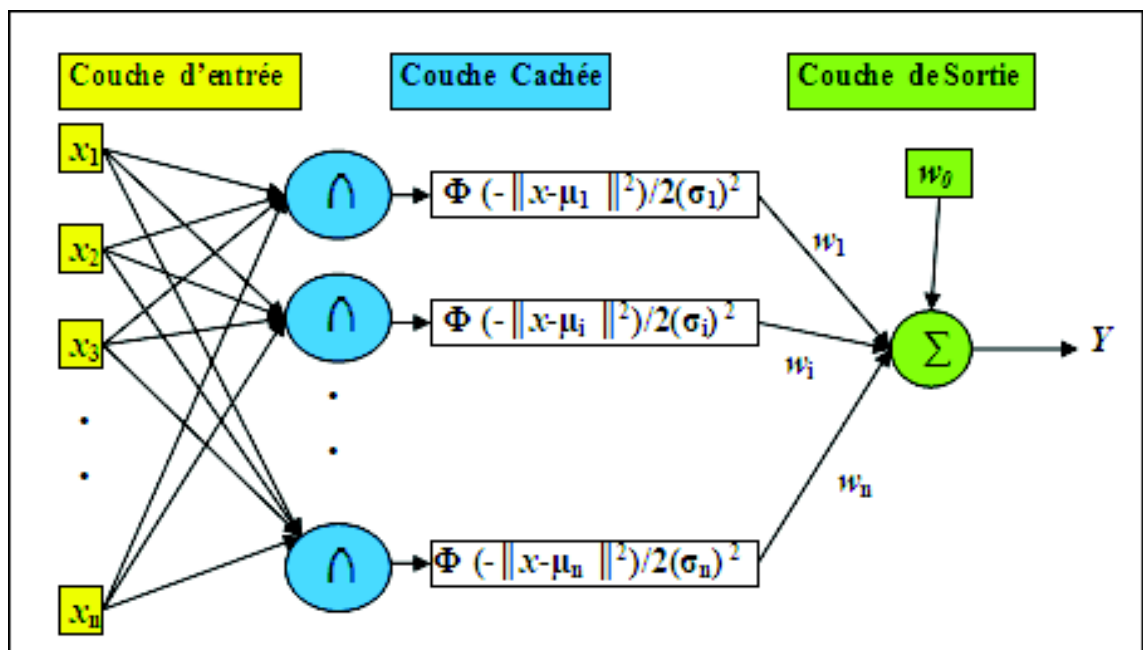


Figure IV.9 Réseau de neurones à fonctions radiales RBFNN.

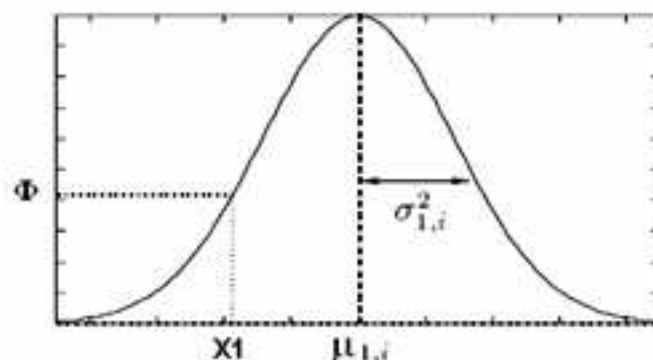


Figure IV.10 Fonction d'activation d'un neurone cachée possédant une seule entrée.

L'utilisation de ces réseaux remonte aux années 1970 (Hardy, 1971 ; Agteberg, 1974) pour résoudre des problèmes d'interpolation multi variables, les bases théoriques de ces réseaux ont été ensuite approfondies par (Powell, 1987 ; Poggio *et al.*, 1989) et (Moody et Darken., 1989).

D'autres travaux se sont succédés où l'application des réseaux de neurones à fonctions de base radiales (rbFNN) a été élargie à d'autres domaines, à savoir la prédiction de l'évolution des systèmes dynamiques (Broomhead et Lowe., 1988) et la classification. La particularité de ces réseaux réside dans le fait qu'ils sont capables de fournir une représentation locale de l'espace grâce à des fonctions de base radiales  $\phi (\|.\|)$  dont l'influence est restreinte à une certaine zone de cet espace ( $\|.\|$  représente la norme euclidienne)(zemouri, 2003, zemouri et al., 2003).

La propriété d'approximateurs universels pour ces réseaux a été prouvée pour des gaussiennes radiales (Hartman et Kowalski, 1990), et plus généralement pour des réseaux de neurones à fonctions de base radiales RBFNN (Park et Sandberg, 1991).

### IV.9.2. architectures

Ces réseaux sont généralement composés de deux couches, une couche cachée de neurones RBF (dont la fonction de transfert est donnée par l'Équation IV.37) et une couche de sortie constituée de perceptrons. Leur architecture est semblable à celle des perceptrons multicouches, toutefois, les réseaux RBFNN utilisent une approche par noyaux, un noyau est une fonction définie dans l'espace des entrées, qui délimite une région restreinte appelé *zone d'activité ou champ d'action*, l'ensemble de l'espace d'entrée est recouvert par les différentes zones d'activités des noyaux, qui se recouvrent. Le problème à résoudre est divisé en un ensemble de sous problèmes, chacun d'eux étant résolu dans l'un des noyaux.

Les fonctions d'activation également appelées fonctions de base sont radiales (généralement gaussiennes(Figure IV.10)), dont les paramètres ( $\mu_i$  et  $\sigma$ ) sont à déterminer lors de l'apprentissage, en effet, leur sortie ne dépend que de la distance entre l'entrée et un point particulier de l'espace d'entrée : le centre de la fonction

$$\Phi (x, \mu_i, \sigma_i) = \exp \left( \frac{-\|x - \mu_i\|^2}{2\sigma_i^2} \right) \quad (IV.37)$$

Les sorties réalisent une combinaison linéaire des fonctions radiales :

$$Y_k = \sum_{j=1}^J w_{jk} \Phi (x, \mu_j, \sigma_j) + w_{0k} \quad (IV.38)$$

Où  $w_{kj}$  est le poids de la connection entre l'unité cachée  $\Phi$  et la sortie  $y_k$ , les poids de connection sont des paramètres à adapter (en plus des paramètres des  $y_j$ ) lors de la phase d'apprentissage, afin que les sorties du réseau soient proches des sorties désirées.  $\sigma$  est l'écart type de la gaussienne et  $\mu_j$  est le vecteur des coordonnées de son centre.

Enfin, la configuration d'un réseau de neurones RBFNN optimal est difficile. Lors du processus d'apprentissage du réseau, deux stratégies sont possibles, la première consiste à modifier simultanément tous les paramètres du réseau (les coordonnées des centres des fonctions radiales, leur écart-type et les poids), par descente de gradient, cependant, les dynamiques de convergence des fonctions radiales et des poids sont différentes, et les poids convergent plus rapidement que les autres paramètres.

L'apprentissage conduit très souvent à un minimum local, une autre méthode consiste à optimiser séparément les paramètres de la couche cachée, par apprentissage non supervisé, et les poids entre la couche cachée et la couche de sortie, par descente de gradient.

### IV.9.3. apprentissage des réseaux RBFNN

---

L'apprentissage d'un modèle RBFNN consiste à déterminer son architecture (le nombre  $N$  de fonctions radiales) et à fixer les valeurs des paramètres.

L'apprentissage d'un réseau RBFNN est de type *supervisé* : on dispose d'un ensemble d'apprentissage constitué de  $l$  couples (vecteur d'entrée, valeur cible) :

$$(x_1, y_1), \dots, (x_n, y_n) \quad x_i \in \mathbb{R}^d, y_i \in \mathbb{R} \quad (\text{IV.39})$$

Et du coût associé à chaque exemple :

$$E_i = \frac{1}{2} (y_i - F(x_i))^2 \quad (\text{IV.40})$$

Une caractéristique intéressante des modèles RBFNN est que l'on peut diviser les paramètres en trois groupes : les centres  $\mu$ , les largeurs  $\sigma$  et les poids  $\lambda$ . L'interprétation de chaque groupe permet de proposer un algorithme d'apprentissage, simple et performant (Moody et Darken, 1989)

#### IV.9.3.1. Apprentissage par descente de gradient

L'apprentissage décrit dans la section précédente consiste à optimiser les paramètres du modèle RBFNN par descente de gradient, comme on le fait pour d'autres modèles connexionnistes. Il faut pour cela calculer les dérivées du coût (éventuellement régularisé) par rapport aux différents paramètres.

Pour une fonction gaussienne :

$$\Phi_j(x_i, \mu_j, \sigma_j) = \exp\left(\frac{-\|x_i - \mu_j\|^2}{2\sigma_j^2}\right) \quad (\text{IV.41})$$

Et un coût :

$$E_i = \frac{1}{2} (y_i - y(x_i))^2 \quad (\text{IV.42})$$

Les dérivées partielles s'écrivent :



$$\frac{\partial E_i}{\partial w_j} = -w_j \left( y_i - \sum_j w_j \Phi_j \right) \Phi_j \quad (\text{IV.43})$$

$$\Delta w_j = -\eta \frac{\partial E_i}{\partial w_j} \quad (\text{IV.44})$$

$$\frac{\partial E_i}{\partial \sigma_j} = -w_j \frac{\|x - \mu_j\|^2}{\sigma_j^3} \left( y_i - \sum_j w_j \Phi_j \right) \Phi_j \quad (\text{IV.45})$$

$$\Delta \sigma_j = -\eta \frac{\partial E_i}{\partial \sigma_j} \quad (\text{IV.46})$$

$$\frac{\partial E_i}{\partial \mu_j^k} = -w_j \frac{x_i^k - \mu_j^k}{\sigma_j^2} \left( y_i - \sum_j w_j \Phi_j \right) \Phi_j \quad (\text{IV.47})$$

$$\Delta \mu_j^k = -\eta \frac{\partial E_i}{\partial \mu_j^k} \quad (\text{IV.48})$$

A partir de ces équations, on peut mettre en œuvre un algorithme d'apprentissage standard de minimisation de l'erreur, en version *batch* (calcul de l'erreur sur l'ensemble des exemples avant mise à jour des paramètres).

## IV.10. Réseaux de neurones à fonction de régression générale (GRNN)

Le réseau de neurones, dite General Régression Neural Network (GRNN), a été proposé par Donald F. Specht (Specht, 1991), il se classe dans la catégorie des réseaux de neurones probabilistes. Ce type de réseau de neurones comme d'autres réseaux probabilistes nécessitent uniquement une partie de la base de données, habituellement exigée par les réseaux de neurones utilisant la rétropropagation du gradient comme algorithme d'apprentissage (Specht, 1991). Un réseau de neurones probabiliste à une capacité de converger très rapide à la fonction optimum avec un échantillon assez réduit, ceci fait des réseaux GRNN un outil très utile pour effectuer des prévisions et des simulations.

Le réseau de neurone GRNN est un type de réseau bien adapté à l'interpolation. Il est basé sur l'évaluation d'une fonction de densité de probabilité d'une variable aléatoire de vecteur, [X], et d'une variable aléatoire scalaire, [Y]. Si les fonctions de densité de probabilité commune de ces variables sont connues, alors la fonction de densité conditionnelle de probabilité et la valeur prévue peuvent être calculées. Selon Specht (1991), la valeur prévue de Y pour un X indiqué est présentée dans l'équation de régression générale suivante :

$$E[Y / X] = \frac{\int_{-\infty}^{+\infty} Y f(X, Y) dY}{\int_{-\infty}^{+\infty} f(X, Y) dY} \quad (IV.49)$$

$E[X/Y]$  = moyenne conditionnelle de Y sur X  
 $f(X, Y)$  = fonction de densité de probabilité

La fonction de densité de probabilité,  $f(X, Y)$ , est estimée à partir des observations d'échantillon de X et de Y. Pour une évaluation non paramétrique de  $f(X, Y)$ , l'évaluation de Parzen (Parzen, 1962), est employée par le réseau GRNN, elle est définie par l'équation suivante pour les observations des couples,  $X_i$  et  $Y_i$  du vecteur scalaire X et Y :

$$f'(X, Y) = \frac{1}{(2\pi)^{\frac{p+1}{2}} \sigma^{p+1}} \cdot \frac{1}{n} \sum_{i=1}^n f_x f_y \quad (IV.50)$$

Où

$$f_x = \exp \left[ -\frac{(X - X_i)^T (X - X_i)}{2\sigma^2} \right] \quad (IV.51)$$

Et

$$f_y = \exp \left[ -\frac{(Y - Y_i)^2}{2\sigma^2} \right] \quad (IV.52)$$

$n$  : Nombre des observations ;

$p$  : Dimension du vecteur X ;

$\sigma$  = paramètre (appelé en anglais SPREAD).

Une estimation de la valeur désirée de Y à n'importe quel X donné, est dérivée dans l'équation (IV.53) en combinant les équations (IV.49) et (IV.50) et en effectuant l'intégration après avoir d'abord échangé l'opération d'intégration et d'addition.

$$\hat{Y}(X) = \frac{\sum_{i=1}^n Y_i \exp\left(-\frac{D_i^2}{2\sigma^2}\right)}{\sum_{i=1}^n \exp\left(-\frac{D_i^2}{2\sigma^2}\right)} \quad (\text{IV.53})$$

Où, la fonction scalaire  $D_i^2$  est donnée par :

$$D_i^2 = (X - X_i)^T (X - X_i) \quad (\text{IV.54})$$

$$D(X, X_i) = \sum_{k=1}^n \left(\frac{X_k - X_{ik}}{\sigma}\right)^2, \quad (\text{IV.55})$$

L'algorithme principal du modèle GRNN est exprimé par les Equations (IV.53), (IV.54) et (IV.55). L'évaluation  $\hat{Y}(X)$  est une moyenne pondérée de tous les échantillons observés,  $Y_i$ , où chaque échantillon est pondéré d'une façon exponentielle selon la distance euclidienne, des  $D_i$ , de chaque  $X_i$ . Cette pondération appropriée est expliquée par le rapport inversement proportionnel entre l'expression  $\exp\left(-\frac{D_i^2}{2\sigma^2}\right)$  et les  $D_i$ , c'est-à-dire, comme augmentations de  $D_i$ , correspond à une diminution de  $\exp\left(-\frac{D_i^2}{2\sigma^2}\right)$  et vice-versa. Une valeur optimale pour le paramètre,  $\sigma$ , est déterminée par des méthodes itératives. Elle doit être plus grande que 0 et peut généralement s'étendre de 0.01 à 1.

La figure IV.11 représente l'architecture détaillée du réseau de neurone GRNN, dont l'algorithme est représenté par les équations (IV.53) et (IV.54). La distance euclidienne,  $D_i$ , est calculée par les liens entre la couche d'entrée et la première couche cachée.

Basé sur les échantillons observés,  $X_i$ , et le paramètre,  $\sigma$ , l'expression,  $\exp\left(-\frac{D_i^2}{2\sigma^2}\right)$

est calculée. Un nœud dans la deuxième couche cachée prend la somme des valeurs exponentielles de tous les échantillons (le nœud A). Dans les autres nœuds de cette même couche, les produits, des valeurs exponentielles et la valeur correspondante  $Y_i$  observée sont calculés. Le nœud dans la troisième couche cachée calcule la somme de toutes les valeurs des produits (le nœud B), qui est alors transmise au nœud de production ou de rendement, où le rapport entre les B et A est calculé.

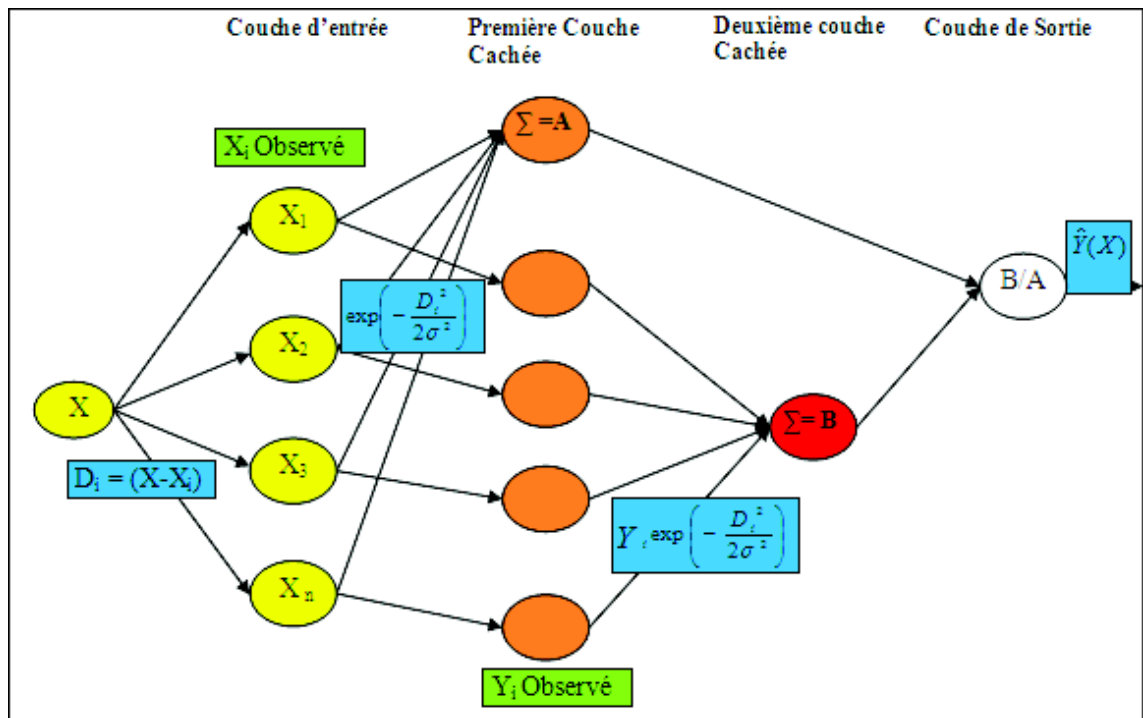


Figure IV.11. Structure détaillé du réseau de neurones GRNN

## IV.11. Résultats de la modélisation par les réseaux de neurones

Nous présentons ci après les résultats obtenus avec les diverses architectures (MLPNN, RBFNN, GRNN). Toutes les architectures ont été entraînées et validées sur les mêmes ensembles de données tels que décrits dans le chapitre III. Nous rappelons qu'il y avait une base de données pour la modélisation de la dose de coagulant, qui contient 726 individus, et caractérisée par six variables descriptives, cette base a été divisée en un ensemble d'apprentissage, un ensemble de validation et un ensemble de test. On a utilisé 60% des points pour faire l'entraînement et les points restants pour la validation et le test.

Pour chaque architecture, nous avons fait varier les paramètres de conception pour observer l'effet sur les performances de la modélisation. Dans le cadre de ce travail, nous avons étudié l'impact du nombre de neurones dans la couche cachée pour le perceptron multicouche (MLPNN). C'est le paramètre par excellence qui permet de sélectionner la complexité intrinsèque du modèle implémenté par ce type d'architecture de réseaux de neurones.

Pour les réseaux à fonctions de base radiales, nous avons varié le nombre de neurones dans l'unique couche cachée.

Pour les réseaux de type General Regression Neural Network GRNN, nous avons varié l'unique paramètre clé qui est le SPEARD.

Pour comparer les architectures entre elles, il est nécessaire de définir un critère d'erreur ou de performance de modélisation qui est couramment appelé fonction coût.

L'entraînement proprement dit des réseaux de neurones nécessite également de faire un tel choix. Généralement, la fonction coût la plus utilisée pour faire l'apprentissage est la somme quadratique des termes d'erreur. ("*Mean Square Error MSE*").

### IV.11.1. Résultats obtenus par le perceptron multicouches MLPNN

---

#### IV.11.1.1. La mise en œuvre

La mise en œuvre d'un réseau de neurones pour résoudre un problème particulier nécessite plusieurs étapes successives :

- Etape 1 : choix de l'architecture du réseau : en premier lieu, il faut choisir l'architecture du réseau, c'est à dire le type de réseau, les entrées externes, le nombre de neurones cachés, la ou les sorties et les agencements des neurones entre eux. Cette étape influence considérablement la qualité du résultat. En effet, le nombre de poids ajustables est l'un des paramètres clés (Pagotto, 1999). De nombreux auteurs considèrent qu'il n'y a pas de règles pour définir l'architecture optimale du réseau et que seule l'expérimentation le permet (Livingstone et al., 1997).
- Etape 2 : apprentissage : il faut ensuite calculer les valeurs des poids synaptiques, à partir des exemples fournis, de telle façon que le réseau reproduise au mieux la ou les sorties. Il fait appel à des *algorithmes d'optimisation* qui cherchent à minimiser par des méthodes non linéaires une fonction coût qui mesure l'écart entre les réponses réelles du réseau et ses réponses désirées. Cette optimisation se fait de manière itérative, en modifiant les poids en fonction du gradient de la fonction coût. Le système le plus communément utilisé est l'algorithme de *rétropropagation de l'erreur* (Livingstone et al., 1997).
- Etape 3 : la validation : pour terminer, il faut tester la qualité du réseau obtenu en lui présentant des exemples qui ne font pas partie de l'ensemble d'apprentissage. Ce test de la capacité de généralisation du réseau est le seul véritable test de la performance du réseau (Livingstone et al., 1997) qui pourra ensuite être utilisé en prédiction. Au préalable à ces trois étapes, comme pour toute méthode statistique, une vérification des données ainsi qu'un prétraitement éventuel (normalisation) sont souvent nécessaires (Stein, 1999).

#### IV.11.1.2. L'utilisation des perceptrons multicouches

Le logiciel utilisé, Matlab 2008Rb, est un logiciel de la société MathWorks (USA) qui permet de développer et d'exécuter des applications utilisant la technique des réseaux de neurones formels. La fonction d'activation choisie pour les neurones de la couche cachée est la fonction sigmoïde, et celle pour le neurone de sortie est la fonction linéaire.

L'algorithme de minimisation retenu est celui de Levenberg- Marquardt. Un jeu de validation dont la dimension a été fixée à 20 % de l'ensemble des données est utilisé afin d'éviter le sur-apprentissage. Les données du jeu sont choisies non groupées et sont tirées au hasard dans l'ensemble de données. L'apprentissage est arrêté lorsque le coût sur l'ensemble de validation commence à croître. La qualité de l'apprentissage est vérifiée en comparant les valeurs calculées par le réseau à celles désirées.

$$EQM = \frac{1}{N} * \sum_{i=1}^N (D_{C_i} - D_{O_i})^2 \quad (IV.56)$$

EQM : Erreur quadratique moyenne ;  
N : Nombre de données ;  
D<sub>C<sub>i</sub></sub> : Valeur calculée de la sortie i ;  
D<sub>O<sub>i</sub></sub> : Valeur observée de la sortie i.

Des graphiques représentant la sortie calculée en fonction de la sortie désirée (diagrammes Scatterplots) permettent de visualiser la qualité des ajustements. La qualité des apprentissages reflète alors la capacité descriptive des réseaux de neurones. Pour étudier les capacités de généralisation des réseaux, une phase de test est effectuée sur des données non utilisées ni pour l'apprentissage, ni pour la validation.

#### IV.11.1.3. Généralisation et critère d'arrêt pour l'apprentissage

Dans un processus d'apprentissage le réseau de neurones est construit en minimisant, par exemple, une fonction coût sur un ensemble fini d'exemples, l'ensemble d'apprentissage. Cependant, le plus important est la faculté de généraliser la représentation construite par le réseau à toutes les données, y compris celles n'appartenant pas à l'ensemble d'apprentissage. Une manière d'évaluer cette faculté consiste à mesurer les performances du réseau de neurones sur des données représentatives du problème non apprises. Il s'agit d'une évaluation de l'erreur de généralisation (Lemaire, 1999). La différence entre l'erreur d'apprentissage et l'erreur de généralisation représente une mesure de la qualité de l'apprentissage effectué (Vapnik, 1982; Vapnik, 1995; Boucheron, 1992).

L'erreur de généralisation dépend avant tout de trois paramètres : le nombre d'exemples utilisés pour l'apprentissage, la complexité du problème sous-jacent et l'architecture du réseau. On constate qu'une amélioration de la généralisation peut être vue sous deux aspects imbriqués (Vapnik et Chervonenkis, 1971; Vapnik, 1982):

- Si la taille du réseau est fixée, quel est l'ensemble d'exemples d'apprentissage qui donnera la meilleure généralisation?
- Si le nombre des exemples est fixé, comment choisir le réseau pour avoir la meilleure généralisation?

Il est difficile de répondre à chacune de ces questions de manière individuelle tant la réponse à l'une nécessite de se pencher sur l'autre. Il est nécessaire ici d'introduire la notion de capacité : si on considère le réseau de neurones comme un système permettant de choisir une fonction parmi un ensemble déterminé par la structure du réseau, alors la capacité du système représente le nombre d'exemples que le réseau peut apprendre correctement à tout coup. Plus un système peut approximer de fonctions différentes, plus sa capacité est élevée en général plus le nombre de poids est élevé, plus la capacité augmente (Lemaire, 1999).

L'une des méthodes qui permet de contrôler la capacité de généralisation d'un réseau de neurones consiste à arrêter l'apprentissage "à temps". De ce fait différents critères permettant de décider quand stopper l'algorithme d'apprentissage ont été développés :

1. Quand l'erreur d'apprentissage a atteint un seuil fixe ;

2. Après un nombre fixe de cycles d'apprentissage ;
3. Quand une estimation de l'erreur de généralisation est minimum.

Les deux principaux pièges que l'on peut rencontrer lors de l'utilisation d'un réseau de neurones sont l'existence de *minimas locaux* dans la fonction d'erreur et un possible *sur apprentissages* (En anglais *overfitting*) (Heddam et al., 2010). Le premier écueil est évité en commençant l'algorithme d'optimisation des poids  $w_i$  en diverses positions de l'espace des variables d'entrée (Figure IV.12). Le sur apprentissage est un problème complexe qui apparaît lorsque le modèle apprend des détails « bruités » lors de la phase d'apprentissage. Un modèle « bien appris » est un modèle qui passe au plus près des points expérimentaux. Un modèle « sur appris » est un modèle qui passe exactement par les points expérimentaux, mais qui oscille entre ceux-ci (Figure IV.13).

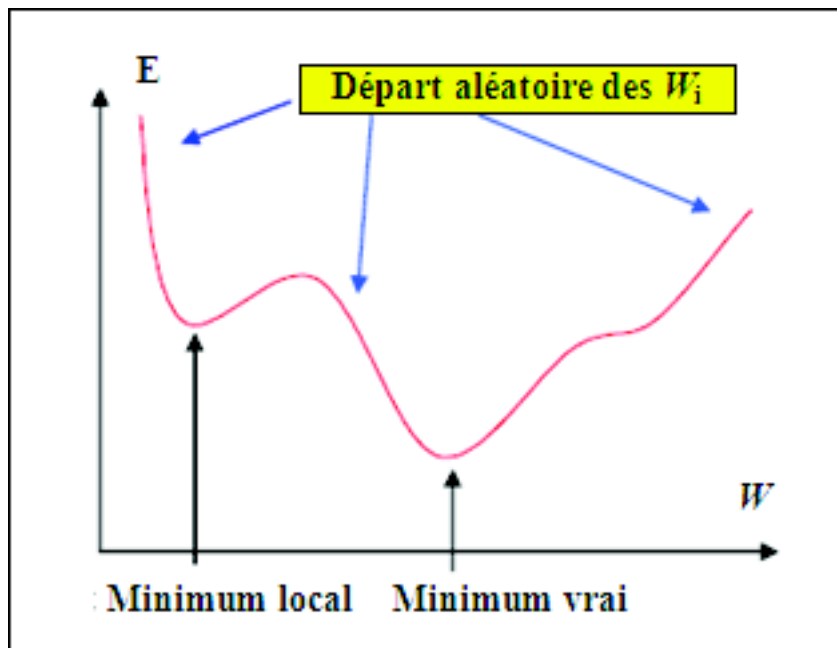


Figure IV.12 Algorithme de rétropropagation avec départ aléatoire des  $W_i$



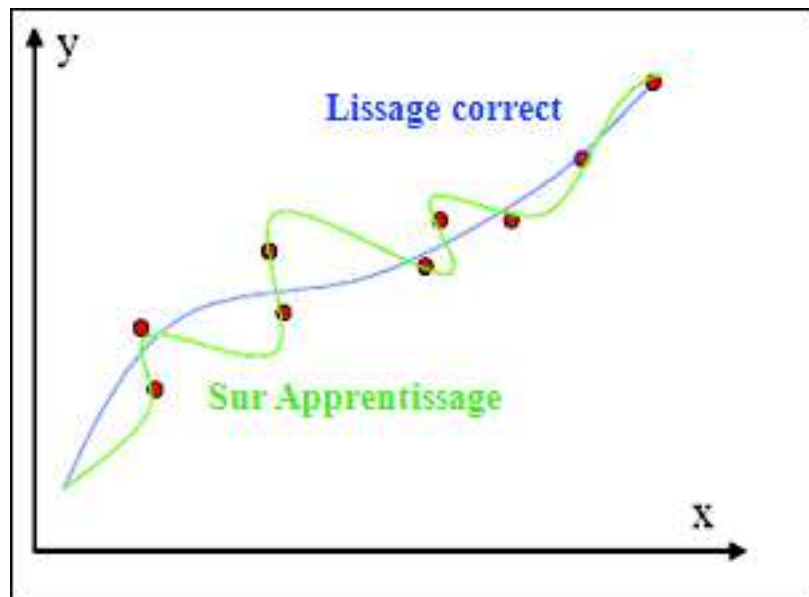


Figure IV.13. Ajustement d'une courbe à un nuage de points

Ce phénomène apparaît lorsqu'il y a trop de poids ajustables (et donc de neurones cachés) ou quand l'optimisation de ces poids est poussée trop loin. De tels modèles ont de très faibles performances en termes de prévision. Dans ce travail, afin d'éviter ce genre de phénomènes, les données sont partitionnées en trois sets de données : un set d'apprentissage, un set de validation, et un set de test (Figure IV.14). L'optimisation des poids ( $w_i$ ), qui est calculée uniquement sur le set d'apprentissage, conduit à diminuer la fonction d'erreur sur le jeu d'apprentissage. A chaque pas de modification des poids, on présente le modèle au jeu de données de validation. L'erreur est calculée pour ces données. Tant que cette erreur sur le jeu de validation diminue, on poursuit la modification des poids. Si l'erreur sur le set de validation augmente, c'est le signe d'un sur apprentissage. On arrête donc le processus et on retient les valeurs de poids finaux au minimum de l'erreur sur le test. Cette optimisation est stoppée, en accord avec la technique appelée « Early-Stopping », lorsqu'une fonction d'erreur est minimum sur le set de données de validation (Figure IV.15).

<b>Base de données de départ : 726 individus</b>		
<b>Apprentissage : 435 individus</b>	<b>Validation : 135 individus</b>	<b>Test : 135 individus</b>
Construction du modèle : 80% de la base de Données		Validation du modèle 20 % de la base de Données

Figure IV.14 Répartition de la base de données pour la construction du modèle

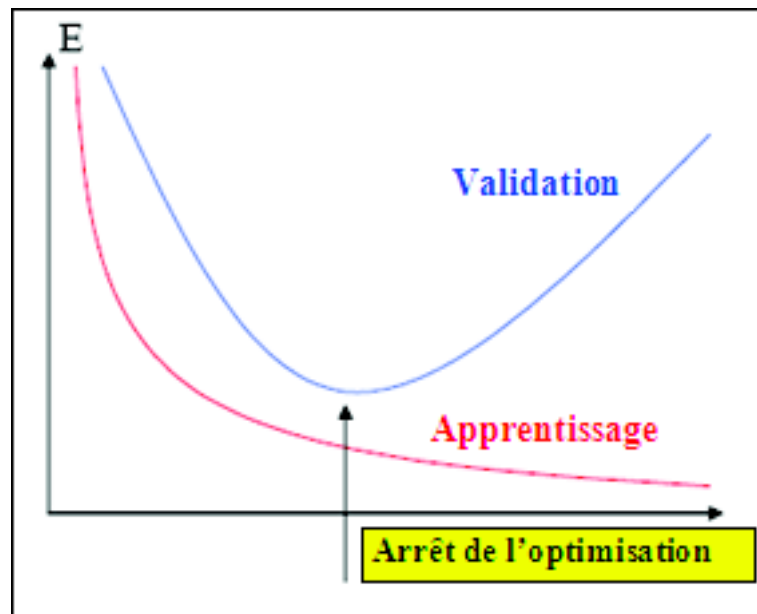


Figure IV.15 L'early-stopping

#### IV.11.1.4. Résultats et interprétations

L'utilisation d'un modèle neuronale passe à la fois par le choix des variables d'entrée mais aussi par l'optimisation de l'architecture du réseau de neurones lui-même. Ceci signifie qu'il faut trouver à la fois le type de fonction d'activation (de transfert) qui convient au problème et le nombre de neurones qui constitue la couche cachée. L'optimisation à la fois du type de fonction non linéaire et de la géométrie du réseau de neurones (nombre de neurones dans la couche cachée) a été faite après de nombreux essais empiriques.

Lors de l'apprentissage d'un modèle neuronal, nous procéderons donc à un grand nombre d'apprentissages successifs, correspondant à des initialisations différentes des paramètres du réseau. De plus, afin d'accélérer le début de l'apprentissage, il est recommandé de choisir des valeurs initiales de ces paramètres suffisamment proches de zéro pour que les sigmoïdes des neurones cachés soient dans leur zone linéaire.

Comme déjà expliqué au chapitre III, nous avons testé cinq variantes ayant des configurations différentes (Tableau III.8) et (Figure IV.16), pour chaque modèle nous avons retenu la meilleure configuration, c'est-à-dire le nombre optimum de neurones dans l'unique couche cachée qui donne les meilleures performances (RMSE, MAE, CC), sur la partie test.

Comme nous pouvons le remarquer sur la Figure (IV. 16), dans la première variante nous avons utilisé deux variables d'entrée: la température (*TE*) et la conductivité (*CON*).

Dans la deuxième variante, nous intégrons une troisième variable qui est le pH.

Dans la troisième et la quatrième variante, nous introduisons respectivement la turbidité (*TU*) et l'oxygène dissous (*OX*).

Enfin et dans la cinquième variante, qui représente le modèle retenu (Figure IV.17) pour la modélisation de la dose de coagulant, nous incluent les six variables descriptives telles qu'ils sont présentés au chapitre III.

Dans ce qui suit, nous présenterons les résultats obtenus par la meilleure configuration de chaque modèle. Le Tableau (IV.1) présente les caractéristiques de chaque meilleure

modèle à savoir le nombre d'entrée, le nombre de neurones dans la couche cachée ainsi que le nombre de paramètres optimisé pour le dit réseau de neurones.

Tableau IV.1. Caractéristiques des meilleures variantes

Modèle	Nombre d'entrée	Nombre de neurones cachés	Nombre de paramètres
M02-MLPNN	02	7	29
M03-MLPNN	03	19	96
M04-MLPNN	04	15	91
M05-MLPNN	05	7	50
M06-MLPNN	06	13	105

Nous avons construit des réseaux de neurones avec 1 à 20 unités dans la couche cachée. Chaque réseau a ainsi été entraîné avec la fonction non linéaire *sigmoïde*. Parmi les différentes configurations testées, nous avons retenu celle produisant le meilleur coefficient de corrélation (CC) (sur le set de données de test (i.e. les données qui n'ont pas été utilisées pendant la phase d'apprentissage et de validation). Le Tableau (IV.2) montre les caractéristiques des meilleurs réseaux de neurones construits et retenu en fonction du nombre de neurones dans la couche cachée, pour les trois phases, l'apprentissage, la validation et le test, les résultats détaillés pour le modèle retenu (M06-MLPNN) sont présentées dans le Tableau (IV.3).

Le Tableau (IV.2) présente une comparaison entre les performances obtenues par les cinq modèles testés. La Figure IV.18 a-d présente les courbes Scatterplots, pour le modèle retenu (M06-MLPNN). En effet, on constate que en intégrant à chaque fois une nouvelle variable à l'entrée du réseau de neurone, on constate une nette amélioration des performances du réseau, et le meilleur modèle est obtenu en intégrant les six variables explicatives, à savoir la température, le pH, la conductivité, la turbidité, les UV<sub>254</sub> et l'oxygène dissous. Le modèle converge bien et on enregistre un bon rapport entre les valeurs observées et calculées de la dose de coagulant (Figure IV.19 a-c) avec une RMSE acceptable (3.47), une erreur moyenne absolue de 2.49 et un coefficient de corrélation de 0.90. Le modèle utilise 13 neurones dans l'unique couche cachée, et bien sur un seul neurone dans la couche de sortie qui correspond à la dose de coagulant, ce qui nous donne d'après la formule (IV.57), un nombre de 105 paramètres à optimiser.

$$COD = f_2 \left[ W_2 \right] f_1 \left[ W_1 \right] \begin{bmatrix} TE \\ pH \\ CO \\ TU \\ OD \\ UV \end{bmatrix} + [b_1] + [b_2] \quad (IV.57)$$

Avec

$W_1$  : Matrice des poids entre la couche d'entrée et la couche cachée ;

$W_2$  : Matrice des poids entre la couche cachée et la couche de sortie ;

$b_1$  : Matrice des biais de la couche cachée ;

$b_2$  : Matrice des biais de la couche de sortie.

$[6 \text{ (Entrées)} * 13 \text{ (neurones cachée)}] + 13 \text{ (biais)} + [13 * 1 \text{ (neurones de sortie)}] + 1 \text{ (biais)}$   
 = 105.

Les paramètres finaux sont présentés dans le Tableau (IV.4).

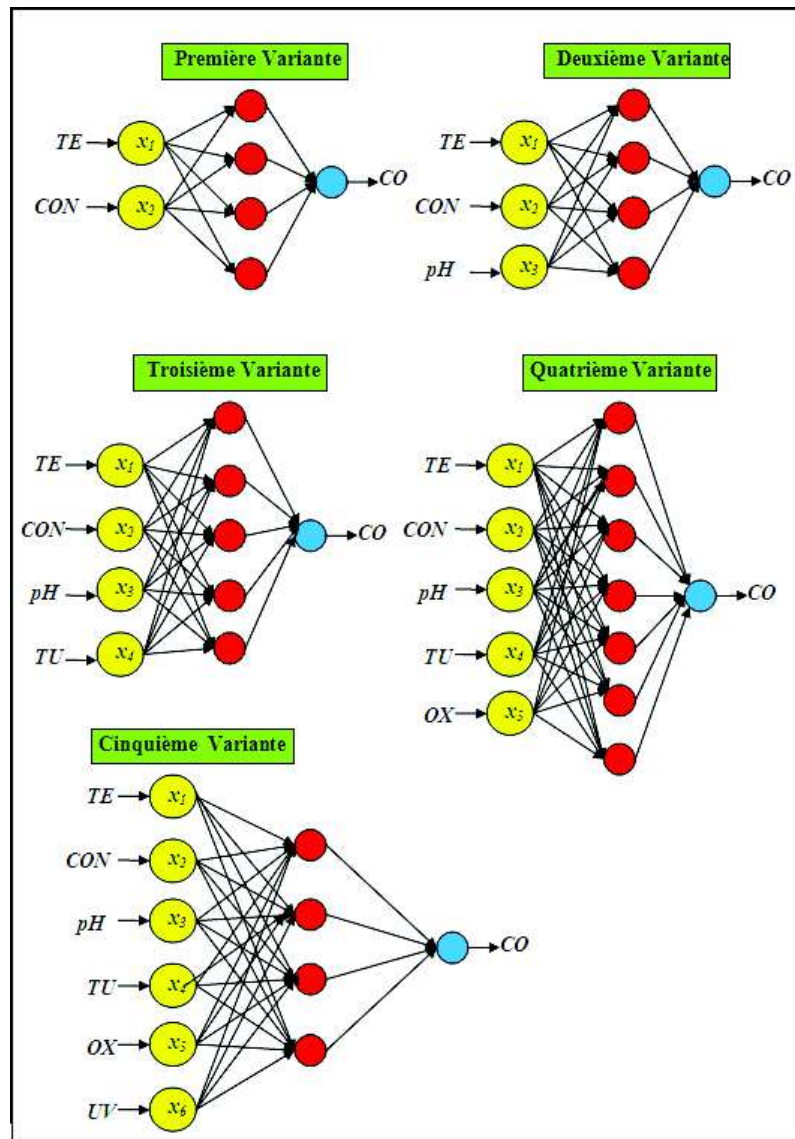


figure Iv.16 Variantes étudiées

Tableau IV.2. Performances des meilleures modèles du réseau MLPNN

Nbr de neurones	Apprentissage			Validation			Test		
	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE	CC
M02-MLPNN	5.88	4.51	<b>0.60</b>	5.55	4.25	<b>0.65</b>	5.73	4.36	<b>0.67</b>
M03-MLPNN	5.06	3.66	<b>0.73</b>	4.70	3.58	<b>0.76</b>	5.05	3.66	<b>0.75</b>
M04-MLPNN	4.54	3.55	<b>0.79</b>	5.31	4.11	<b>0.68</b>	4.68	3.70	<b>0.79</b>
M05-MLPNN	3.87	2.93	<b>0.85</b>	3.88	2.93	<b>0.84</b>	3.61	2.74	<b>0.88</b>
M06-MLPNN	<b>2.65</b>	<b>2.03</b>	<b>0.93</b>	<b>3.62</b>	<b>2.54</b>	<b>0.87</b>	<b>3.47</b>	<b>2.49</b>	<b>0.90</b>

Tableau IV.3. Résultats obtenus par le réseau MLPNN

## Contribution à la modélisation de la qualité des eaux

Nbr de neurones	Apprentissage			Validation			Test		
	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE	CC
01	5.18	3.67	0.71	4.93	3.28	0.75	5.65	4.20	0.65
02	4.43	3.47	0.81	5.37	4.24	0.68	4.58	3.56	0.76
03	4.09	3.20	0.84	4.32	3.28	0.80	5.17	4.00	0.70
04	4.25	3.14	0.82	4.40	3.16	0.80	4.58	3.42	0.77
05	3.79	2.94	0.86	4.74	3.13	0.75	3.67	2.75	0.86
06	3.71	2.79	0.86	4.23	3.20	0.84	4.30	3.23	0.78
07	4.18	3.15	0.82	4.70	3.76	0.79	5.01	3.86	0.74
08	2.68	3.60	0.87	4.54	3.42	0.80	4.50	3.39	0.79
09	4.02	3.02	0.84	4.23	3.29	0.83	4.12	3.07	0.81
10	3.27	2.49	0.89	3.58	2.77	0.88	3.84	3.00	0.84
11	2.97	2.19	0.91	3.96	3.00	0.85	4.73	3.13	0.81
12	4.10	3.13	0.82	4.73	3.70	0.79	4.71	3.42	0.78
13	2.65	2.03	0.93	3.62	2.54	0.87	3.47	2.49	0.90
14	2.91	2.06	0.92	4.49	3.14	0.80	3.84	2.77	0.85
15	2.49	1.82	0.94	3.35	2.52	0.89	3.54	2.44	0.88
16	2.89	1.96	0.92	4.03	2.74	0.82	3.98	2.77	0.86
17	3.10	2.23	0.90	4.01	2.89	0.85	4.14	2.99	0.84
18	2.94	2.11	0.91	4.32	3.23	0.82	4.59	2.92	0.79
19	2.38	1.75	0.94	3.10	2.27	0.89	3.39	2.66	0.85
20	3.35	2.54	0.89	3.68	2.92	0.88	3.83	2.82	0.82

Tableau IV.4 Paramètres finaux du modèle retenu

Neurone N°	Variables d'entrée du modèle M06-MLPNN						Biais	CO	Biais
	TE	pH	CON	TU	OX	UV254			
01	-0.7351	-0.2144	1.1192	-0.1211	0.7456	-0.1587	-1.4797	3.7295	2.5579
02	0.0655	1.1199	1.3037	-0.2445	0.4491	0.1119	-2.8737	2.1782	
03	-2.9305	2.1825	0.2317	-0.1897	-2.5130	-0.6259	-0.9857	2.5707	
04	-0.6488	-0.5772	0.4742	-0.1228	-2.3344	0.8347	-0.3087	3.3769	
05	0.6056	0.4107	-2.6636	0.2281	0.4290	-0.1454	-1.4855	3.3571	
06	0.3042	-0.0269	-2.6711	-0.2646	0.8510	1.0249	-1.3689	-0.0199	
07	0.6489	0.6548	1.6148	0.1632	0.0242	-1.2934	-1.2829	-0.4432	
08	-1.6129	-2.3803	1.6612	-2.8799	-2.3087	0.5966	-0.4392	-1.8704	
09	2.2465	0.0354	-2.3095	-0.2948	2.4695	1.2301	0.6400	-1.3836	
10	-0.3919	1.2871	2.5958	-0.3382	0.1530	0.0830	2.2881	3.8241	
11	-1.3526	-0.8132	2.2960	0.1541	-0.8607	0.8890	-1.0157	-3.0101	
12	-1.3032	3.1494	2.4505	0.5679	-1.6324	-2.9364	0.4802	-0.9113	
13	-1.6501	-1.7136	1.5138	-0.2848	1.6736	0.9944	-0.5514	-2.8799	

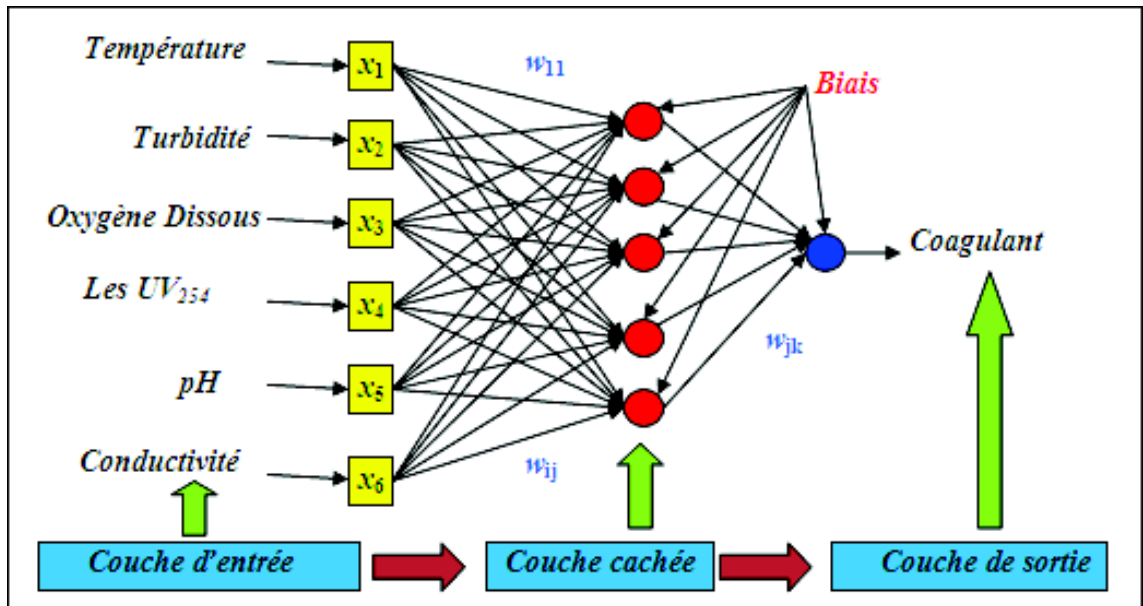


Figure IV.17 : Structure du modèle neuronal final retenu



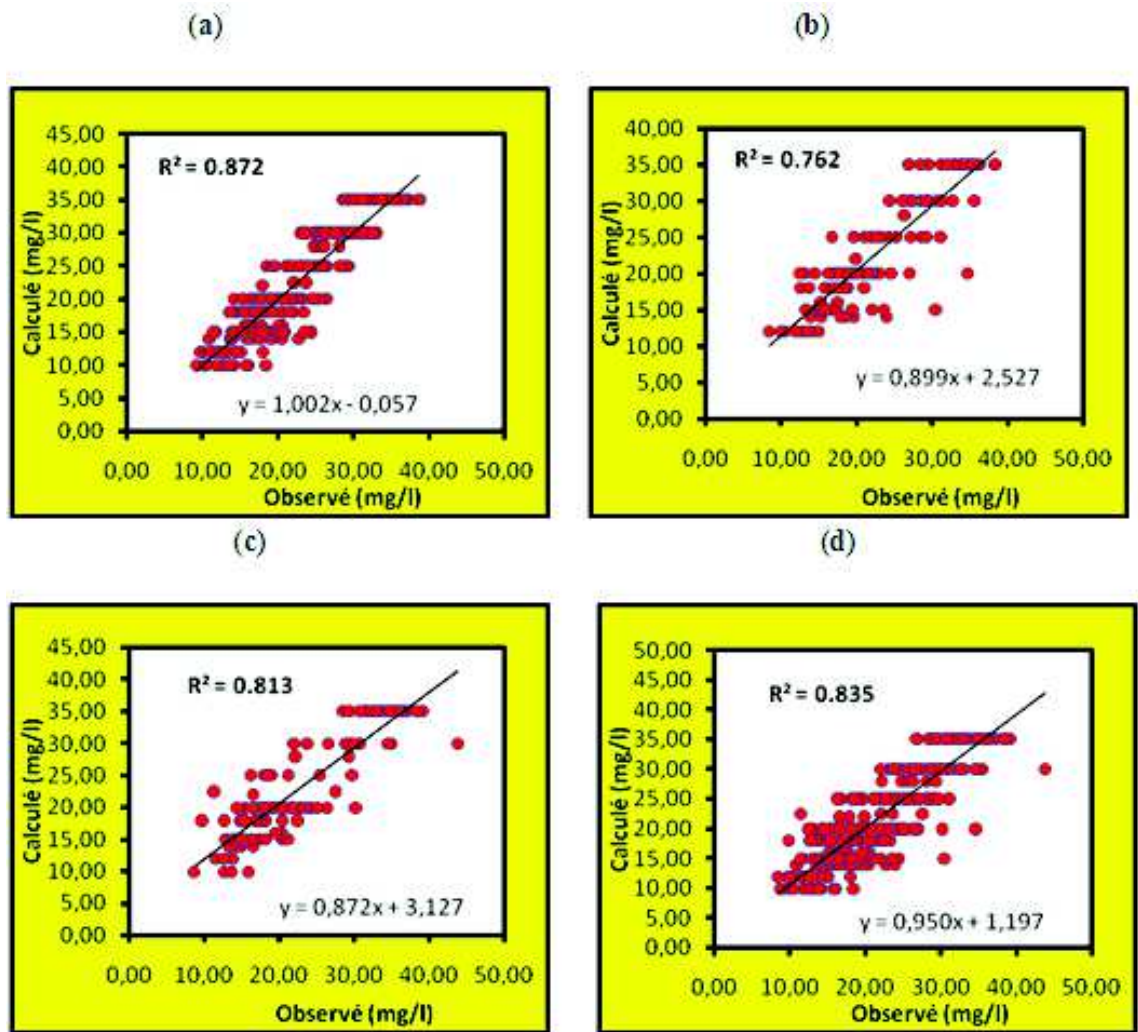


Figure IV.18 Comparaison des résultats du modèle à base de réseau de neurones MLPNN (a) apprentissage, (b) validation, (c) test et (d) l'ensemble des données



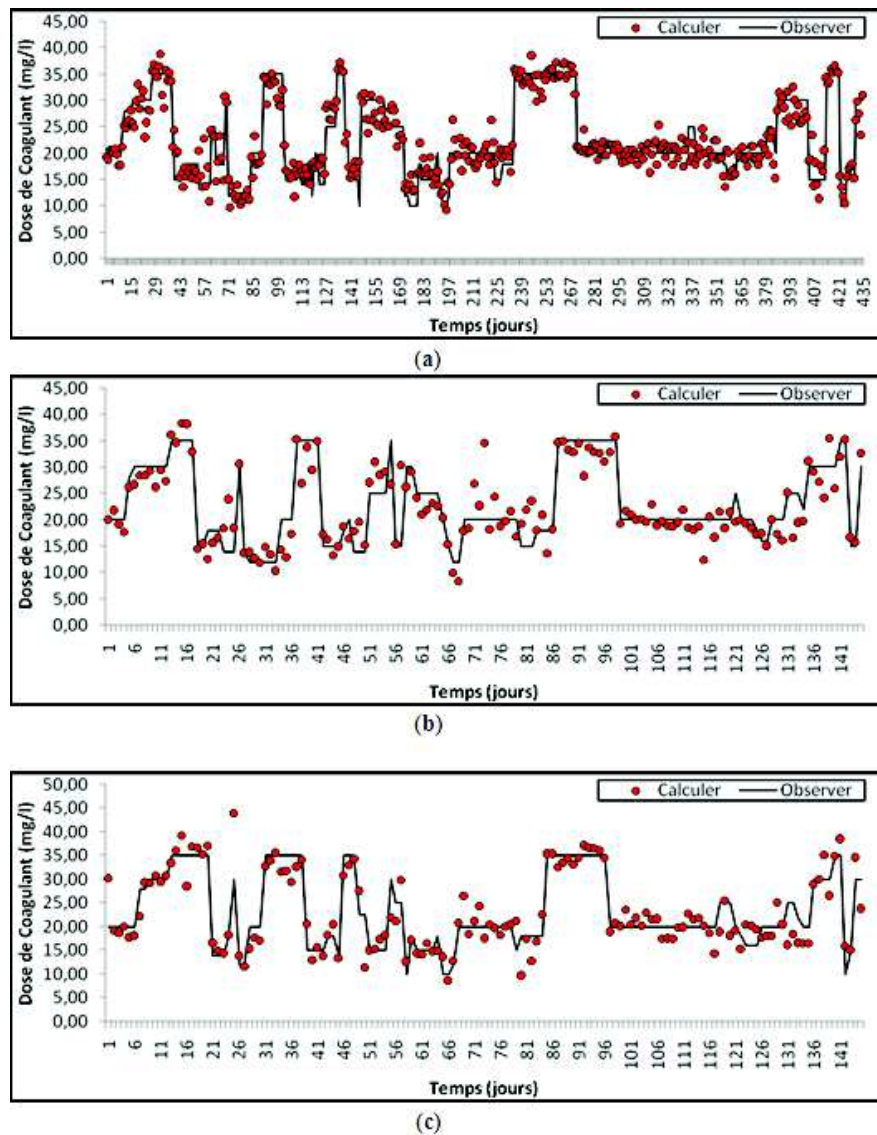


Figure IV.19 Modélisation de la dose de coagulant par le modèle à base de réseau de neurones MLPNN (a) partie apprentissage, (b) partie validation et (c) partie test.

#### IV.11.2. Résultats obtenus par le réseau RBFNN

Le Tableau (IV.5) présente une comparaison entre les performances obtenues par les cinq variantes du réseau RBFNN. La Figure (IV.20 a-d) présente les courbes Scatterplots, pour le modèle retenu (M06-RBFNN). En effet, on constate qu'en intégrant à chaque fois une nouvelle variable à l'entrée du réseau de neurone, on constate une nette amélioration des performances du modèle, au niveau des phases de validation et de test, cependant pour la phase d'apprentissage les résultats sont pratiquement similaires. On constate aussi que les deux modèles M02 et M03 présentent des résultats médiocres pour les deux phases de validation et de test, là où on enregistre un coefficient de corrélation faible ( $0.45 \div 0.65$ ), et une RMSE importante de l'ordre de ( $6.04 \div 8.04$ ) en phase de validation et elle atteint les 9.15 en phase de test pour le modèle à deux entrées. Les performances du modèle RBFNN commencent à s'améliorer nettement à compter de la variante M04 (utilisant quatre entrées)

et le meilleur modèle est obtenu en incluant les six variables, à savoir la température, le pH, la conductivité, la turbidité, les UV<sub>254</sub> et l'oxygène dissous. Le modèle converge bien et on enregistre un bon rapport entre les valeurs observées et calculées de la dose de coagulant (Figure IV.21 a-c), avec une RMSE acceptable (3.35), une erreur moyenne absolue de (2.46) et un coefficient de corrélation de 0.87, en phase de test.

Tableau IV.5. Performances des meilleurs modèles du réseau RBFNN

Modèle	Apprentissage			Validation			Test		
	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE	CC
M02-RBFNN	2.35	1.38	<b>0.94</b>	8.04	5.53	<b>0.45</b>	9.15	4.91	<b>0.29</b>
M03-RBFNN	2.34	1.59	<b>0.94</b>	6.07	3.82	<b>0.65</b>	4.87	2.86	<b>0.75</b>
M04-RBFNN	2.34	1.71	<b>0.94</b>	6.43	4.31	<b>0.69</b>	4.97	3.37	<b>0.77</b>
M05-RBFNN	2.35	1.81	<b>0.94</b>	4.09	2.75	<b>0.85</b>	3.61	2.68	<b>0.86</b>
M06-RBFNN	2.35	1.74	<b>0.94</b>	3.87	2.71	<b>0.86</b>	3.35	2.46	<b>0.87</b>

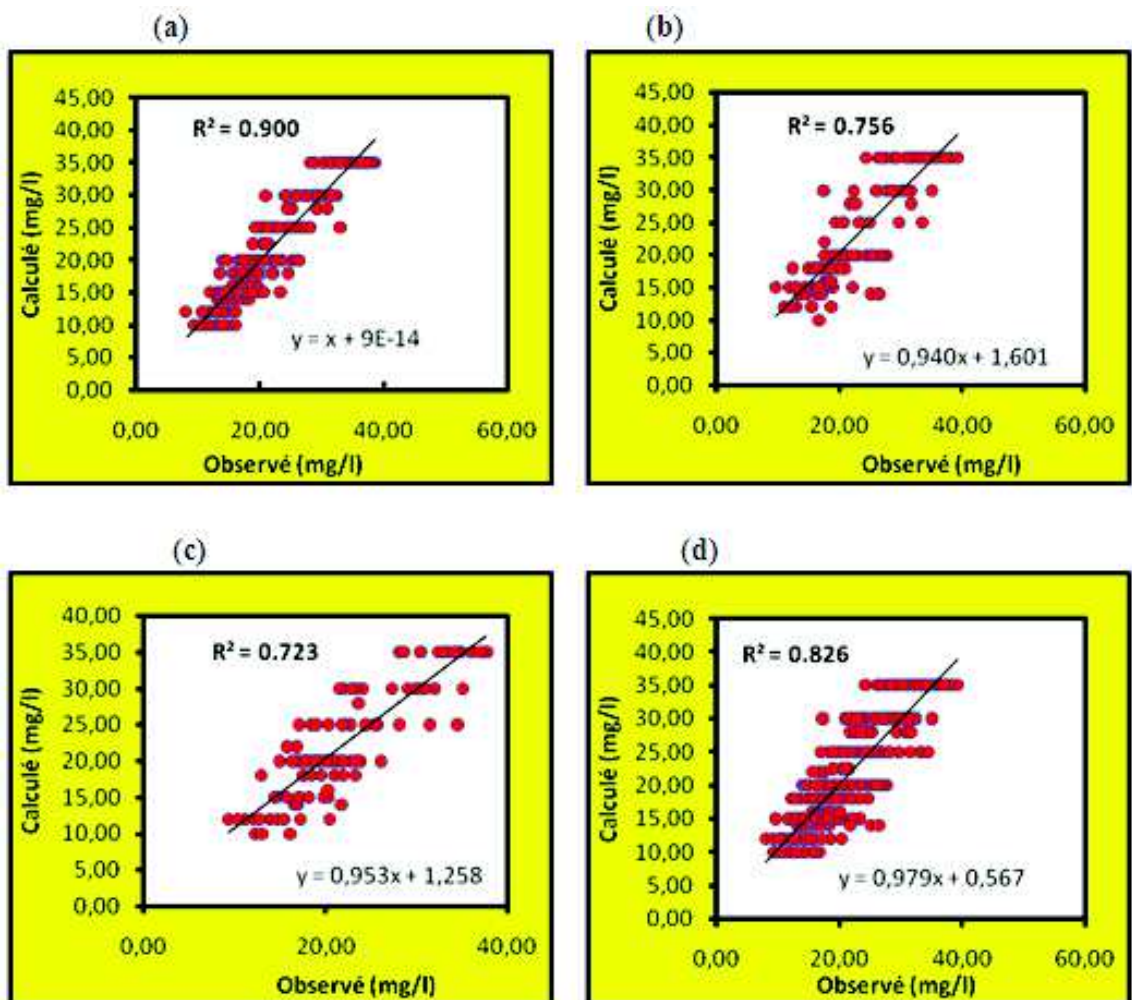
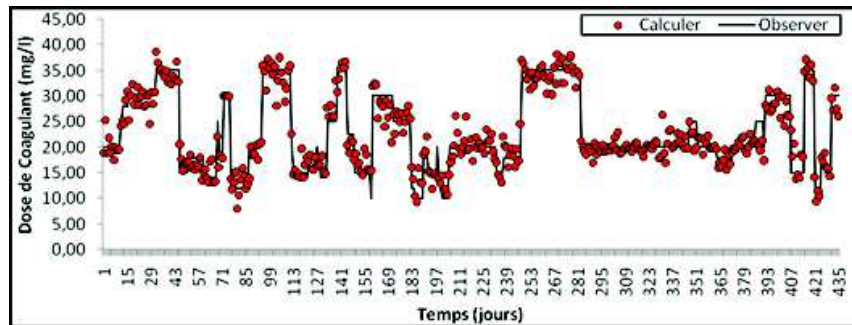
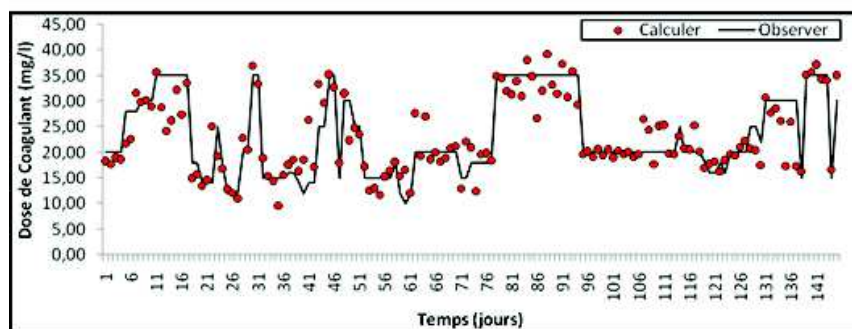


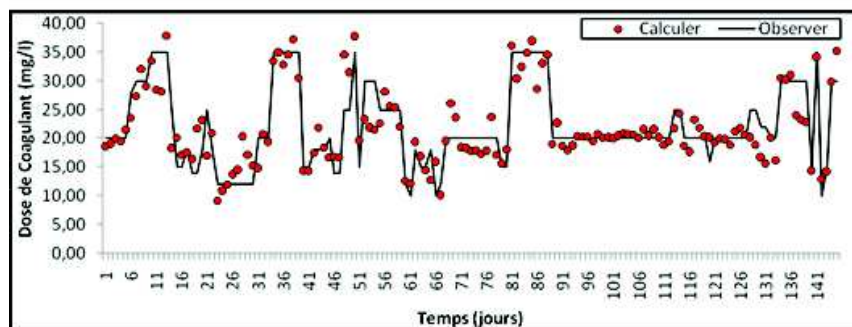
Figure IV.20 Comparaison des résultats du modèle à base de réseau de neurones RBFNN(a) apprentissage, (b) validation, (c) test et (d) l'ensemble des données



(a)



(b)



(c)

Figure IV.21 Modélisation de la dose de coagulant par le modèle à base de réseau de neurones RBFNN(a) partie apprentissage, (b) partie validation et (c) partie test.

### IV.11.3. Résultats obtenus par le réseau à régression générale GRNN

Tout comme les réseaux MLPNN et RBFNN, le réseau de type General Regression Neural Network (GRNN) doit être entraîné avant d'être utilisé. Pour l'entraîner, il faut fixer un unique paramètre : « SPREAD ». Comme sur la couche cachée, il y a autant de neurones que d'observations, le réseau GRNN assigne un neurone à chaque observation (individu). Ce neurone possède une zone d'influence, dont le rayon est fonction du paramètre « SPREAD », autour de l'observation correspondante. Plus le paramètre « SPREAD » est faible, plus le neurone mémorise bien l'observation qui lui a été assignée (Heddam et al., 2011c). Le réseau GRNN donne alors une excellente précision lorsque l'observation à l'entrée est une des observations apprises (bonne capacité à mémoriser). Mais, plus le réseau

GRNN mémorise bien, moins il généralise c'est-à-dire qu'il donne une faible précision quant l'observation à l'entrée n'a pas été apprise, un compromis doit donc être fait en fixant la valeur du paramètre « SPREAD ».

L'algorithme utilisé pour l'apprentissage du réseau GRNN est beaucoup plus simple que celui utilisé pour l'apprentissage du réseau MLPNN; il s'ensuit donc un temps d'apprentissage beaucoup moins long pour le réseau GRNN. Celui-ci dépend aussi du nombre de données d'apprentissage. La précision donnée par le réseau GRNN, contrairement au réseau MLP, n'est pas fonction des valeurs initiales de ses poids et biais.

Tableau IV.6. Performances des meilleures modèles du réseau GRNN

Modèle	Apprentissage			Validation			Test		
	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE	CC
M02-GRNN	3.97	2.66	<b>0.85</b>	5.05	3.59	<b>0.73</b>	5.67	3.95	<b>0.62</b>
M03-GRNN	1.84	0.69	<b>0.97</b>	5.09	2.49	<b>0.73</b>	4.64	2.36	<b>0.77</b>
M04-GRNN	0.31	0.07	<b>0.99</b>	3.36	1.33	<b>0.89</b>	3.08	1.36	<b>0.90</b>
M05-GRNN	2.58	1.70	<b>0.94</b>	3.20	2.08	<b>0.90</b>	3.43	2.14	<b>0.87</b>
M06-GRNN	0.13	0.01	<b>0.99</b>	2.92	1.14	<b>0.92</b>	3.11	1.29	<b>0.90</b>

Le Tableau (IV.6) présente une comparaison entre les performances obtenues par les cinq modèles GRNN testés. La Figure (IV.22 a-d) présente les courbes Scatterplots, pour le modèle retenu (M06-GRNN). En intégrant à chaque fois une nouvelle variable à l'entrée du réseau de neurone, on constate une nette amélioration des performances du modèle. Les performances du modèle GRNN commencent à s'améliorer nettement à compter de la variante M04, utilisant quatre entrées, ainsi le meilleur modèle est obtenu en travaillant avec les six variables, à savoir la température, le pH, la conductivité, la turbidité, les UV<sub>254</sub> et l'oxygène dissous. Le modèle converge bien, et on enregistre un bon rapport entre les valeurs observées et calculées de la dose de coagulant (Figure IV.23 a-c), avec une RMSE acceptable (3.11), une erreur moyenne absolue de (1.29), et un coefficient de corrélation de (0.90), en phase de test.



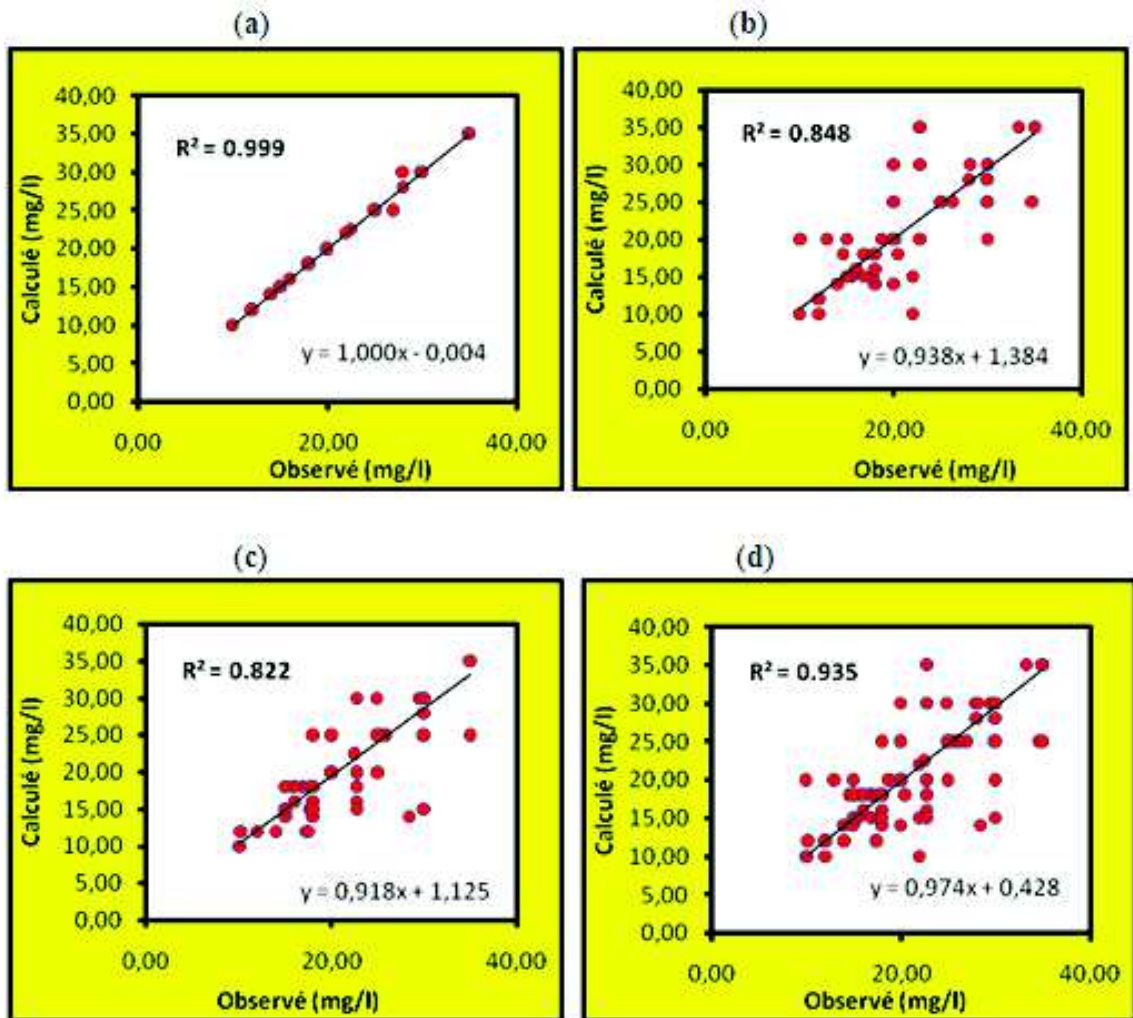


Figure IV.22 Comparaison des résultats du modèle à base de réseau de neurones GRNN: (a) apprentissage, (b) validation, (c) test et (d) l'ensemble des données

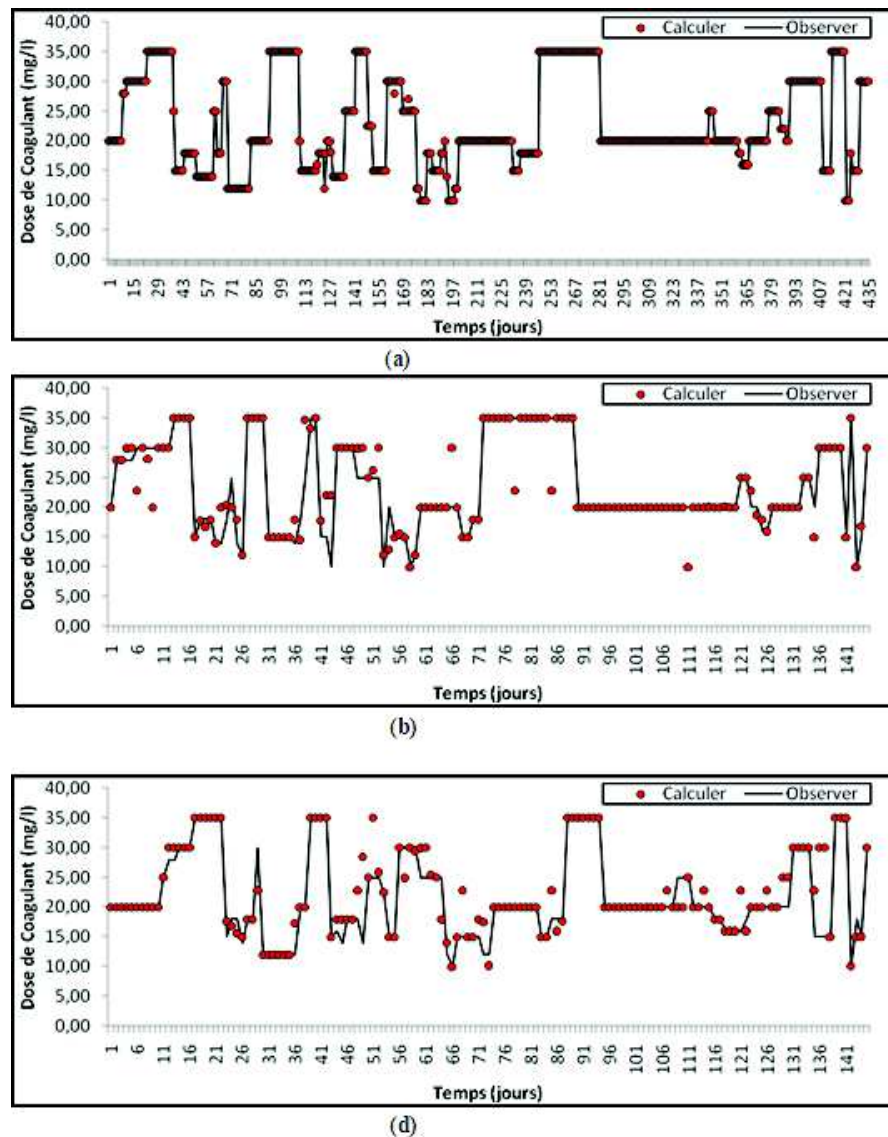


Figure IV.23 Modélisation de la dose de coagulant par le modèle à base de réseau de neurones GRNN : (a) partie apprentissage, (b) partie validation et (c) partie test.

#### IV.11.4. Comparaison des résultats obtenus par les réseaux de neurones

Pour chaque type de réseau, l'architecture donnant la meilleure performance telle que mesurée sur l'ensemble de test a été retenue. Une comparaison globale des réseaux de neurones se trouve résumée au tableau (IV.7). Dans le but d'établir une ligne de base de comparaison, le tableau comporte les résultats obtenus par les trois réseaux de neurone à savoir les réseaux MLPNN, RBFNN, GRNN. Bien qu'employant un algorithme d'apprentissage simple, le réseau de neurones type GRNN, donne les meilleurs résultats, en, particulier en phase d'apprentissage et de validation, en phase de test, on peut remarquer que les deux modèles GRNN et MLPNN, dépassent légèrement le modèle RBFNN. On peut conclure en fin de comparaison que les trois types de réseaux de neurones

présentent un outil efficace pour la modélisation de la dose de coagulant en intégrant les six variables explicatives pris en considération.

Tableau IV.7. Comparaison des résultats obtenus par les réseaux de neurones

Modèle	Apprentissage			Validation			Test		
	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE	CC
M06-MLPNN	2.65	2.03	<b>0.93</b>	3.62	2.54	<b>0.87</b>	3.47	2.49	<b>0.90</b>
M06-RBFNN	2.35	1.74	<b>0.94</b>	3.87	2.71	<b>0.86</b>	3.35	2.46	<b>0.87</b>
M06-GRNN	0.13	0.01	<b>0.99</b>	2.92	1.14	<b>0.92</b>	3.11	1.29	<b>0.90</b>

## IV.12. Conclusions

Dans ce chapitre, nous avons introduit les définitions essentielles relatives aux réseaux de neurones. Nous avons fait une distinction entre les différents types de réseaux de neurones non bouclés, qui réalisent des fonctions non linéaires, à savoir les réseaux de types MLPNN, GRNN, RBFNN.

Nous avons aussi mis l'accent sur l'utilisation des réseaux de neurones comme outils de modélisation par apprentissage. Ces derniers permettent d'ajuster des fonctions non linéaires à des ensembles de données. Comme toute méthode qui s'appuie sur des techniques statistiques, l'utilisation de réseaux de neurones nécessite que l'on dispose de données suffisamment nombreuses, représentatives et de qualité (bien mesurées).

Ensuite, nous avons présenté les concepts fondamentaux de la modélisation à l'aide de réseaux de neurones, et une méthodologie complète de conception.

Enfin, nous avons décrit les applications des réseaux de neurones en traitement des eaux plus précisément la modélisation de la dose de coagulant. Du point de vue des modèles neuronales, nous avons développé une architecture originale qui consiste à élaborer un réseau pour chaque type d'architecture, puis à comparer la capacité prédictive des trois architectures développées. Les entrées des modèles sont les variables descriptives caractérisant l'eau brute à l'entrée de la station de traitement des eaux et la sortie des modèles était la dose de coagulant.

On peut observer deux faits notables, d'abord, les réseaux ont plus de facilité à prédire les données d'apprentissage que les données de validation. C'est un résultat somme toute évident. Une deuxième observation qui ressort systématiquement est la recherche des meilleures performances des réseaux sur l'ensemble de test.

Dans le cas du tableau (IV.7) on remarque que le réseau GRNN donne de meilleurs résultats, il est suivi par le réseau MLPNN pour les trois phases ; apprentissage, validation et test. On peut aussi noter qu'il y a plus de différences entre les modèles neuronaux dans les données concernant la phase de validation, et de test que pour les données de phase d'apprentissage (calage). De plus, les modèles neuronaux font généralement moins d'erreur que la régression linéaire; peut être marginalement pour les ensembles d'apprentissage mais la différence est plus prononcée dans les ensembles de validation. Ceci donne une certaine indication de l'utilité des modèles neuronaux par rapport aux techniques linéaires. Les ensembles d'apprentissage présentent comparativement un comportement moins démarqué, la plupart des architectures apprennent avec une performance comparable pour la base d'apprentissage.



Dans le but d'améliorer les résultats obtenus par les réseaux de neurones, nous proposons dans le chapitre suivant, une nouvelle approche de modélisation de la dose de coagulant basée sur les systèmes neuroflous, elle s'appuie sur l'incorporation de la technique de la logique floue dans un réseau de neurone. Nous nous sommes intéressés plus particulièrement dans la section suivante à une classe bien particulière : c'est le modèle ANFIS «Adaptive neuro-Fuzzy Inference System ».

---

# Chapitre V Modèles A Base De Systèmes Neuroflous

## V.1. introduction

Le travail effectué dans ce chapitre traite globalement de la spécification, et du développement d'un système neuroflous capable de prendre en compte l'information disponible à l'entrée de la station de traitement des eaux, résumé par les variables descriptives de l'eau brute pour la construction d'un modèle capable d'estimer la dose de coagulant.

Donc, l'intérêt est de construire un système prédictif qui s'appuie sur l'intégration des réseaux de neurones et des systèmes d'inférence floue (SIF) à cause de leur complémentarité. Les SIF exploitent des règles linguistiques du type SI-ALORS traduisant une connaissance sur l'état d'un système. Cependant, un SIF n'est pas à même d'apprendre : les règles doivent être formulées, ce qui est parfois difficile. Ainsi, une solution consiste à utiliser les propriétés des réseaux de neurones afin "d'apprendre" la structure floue et d'en adapter les paramètres en conséquence.

Très généralement, les réseaux neuroflous (NF) remplacent les différentes couches cachées des réseaux de neurones par des règles floues (règles linguistiques). Ils utilisent ensuite des algorithmes d'apprentissage pour définir et optimiser ces paramètres. Enfin, les systèmes neuroflous sont très prometteurs, en étant à l'intersection des réseaux de neurones et de la logique floue, les réseaux neuroflous tirent avantages des deux méthodes. Les réseaux de neurones constituent déjà un outil performant, un réseau neuroflous permet de déterminer automatiquement les paramètres du système.

Ce chapitre est structuré en quatre parties :

1. La première est consacrée à la description des principes de base de la logique floue généralement, et des systèmes d'inférence floue plus précisément ;
2. La deuxième est consacrée à la description des systèmes neuroflous, en précisant qu'ils sont de deux types «Mamdani» et «Takagi\_Sugeno», ce dernier est utilisé dans le cadre de ce travail de recherche;
3. Dans la troisième partie, le système ANFIS est détaillé du point de vue fonctionnement, des algorithmes d'apprentissage qu'il nécessite, des avantages et des inconvénients d'implémentation dans différentes applications ;
4. Enfin, dans la quatrième partie, nous étudions les étapes permettant la construction d'un système de type ANFIS. Les éléments théoriques nécessaires à cette analyse sont décrits et une nouvelle "technique" pour la construction de la base de règle est proposée en conséquence, il s'agit de la technique du GROUPAGE SOUSTRACTIF «GS», cette dernière contrairement à la technique GRID PARTITION «GP», donne des résultats meilleurs. L'influence de celle-ci sur les performances du réseau est discutée, nous synthétiserons l'ensemble dans une conclusion générale, et donnons quelques perspectives de travail.

## v.2. historique de la logique floue

Les prémisses de la logique floue, visant à traiter la notion d'incertitude, datent des années 30, il faudra cependant attendre que Zadeh (Zadeh, 1965) introduise le concept de sous-ensembles flous, en 1965, pour assister aux premières grandes avancées dans le domaine. Par la suite, en 1974, Mamdani (Mamdani, 1975) introduisait la commande floue pour la régulation de processus industriel. Enfin, dans les années 80, la commande floue connaît un essor considérable au Japon, notamment grâce aux travaux de Sugeno pour se répandre ensuite dans le monde entier (faucher, 2006).

La logique floue représente une alternative à la logique dite «conventionnelle» introduite par Aristote, où un élément doit absolument appartenir ou non à un ensemble. En contre partie, la logique floue, permet d'affirmer qu'un élément peut appartenir à un ensemble avec un certain degré d'appartenance (connaissance imparfaites). Il existe différentes théories décrivant la modélisation et le raisonnement à partir des connaissances imparfaites (Ciofolo, 2005), les principales sont les probabilités, les fonctions de croyances (Shafer, 1976), et les sous-ensembles flous (Zadeh, 1965) associés à la théorie des possibilités (Zadeh, 1965 ; Zadeh, 1973 ; Zadeh, 1978 ; Dubois et Prade, 1980 ; Dubois et Prade, 1988).

## V.3. La Logique floue

La notion de logique floue permet d'étendre la notion de logique classique, associée aux variables booléennes ne prenant que deux valeurs 0 et 1, il est alors possible d'associer à des variables des coefficients d'appartenance à des sous-ensembles flous prenant des valeurs dans l'intervalle  $[0, 1]$  et quantifiant l'incertitude sur la variable (faucher, 2006). Un événement certain pour la variable se traduira par un coefficient d'appartenance au sous-ensemble flou, c'est-à-dire à la propriété, égal à 1 alors que la valeur sera inférieure à 1 en présence d'incertitudes, il vient alors qu'une valeur nulle pour un coefficient d'appartenance indique que la possibilité d'appartenance au sous-ensemble sélectionné de la variable représentant la grandeur concernée est complètement rejetée.

L'univers de discours d'une variable donnée, c'est-à-dire son domaine de variation, peut alors être divisé en plusieurs sous-ensembles au moyen de fonctions d'appartenance (FA), comme illustré par la (figure V.1) avec des fonctions d'appartenance triangulaires.

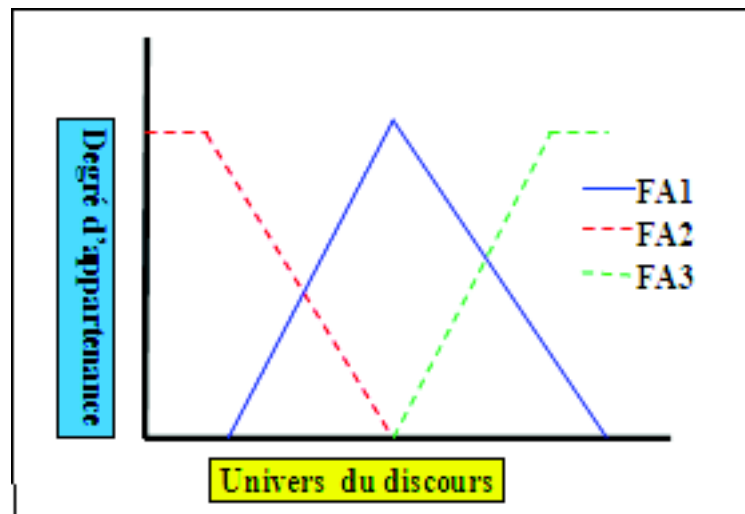


Figure V.1 Exemple de fonctions d'appartenance

Pour chaque valeur de la variable considérée (faucher, 2006), des degrés d'appartenance à chacun des sous-ensembles flous vont être définis, les fonctions d'appartenance FA1, FA2, FA3 se rapportent alors respectivement au sous ensemble flou SF1, SF2, SF3, la problématique sera donc de choisir le nombre, et le type de sous-ensembles flous pour chacune des variables devant être traitée.

## V.4. Éléments de théorie des sous-ensembles flous

Dans cette partie, nous allons exposer, de manière synthétique les éléments de la théorie des sous-ensembles flous, et les notions nécessaires à une bonne compréhension d'un raisonnement flou.

### V.4.1. Définition et éléments caractéristiques

#### V.4.1.1. Sous-ensembles classique

En logique binaire (classique ou booléenne), un sous-ensemble classique A de S est défini par sa fonction caractéristique, appelée fonction d'appartenance et notée  $\mu$ :

$$\mu_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases} \quad (V.1)$$

Par exemple, on peut donner la fonction caractéristique du sous-ensemble A des doses de coagulant compris entre 20 et 40 mg/litre :

$$\mu_A(x) = \begin{cases} 1 & \text{si } 20 \leq x \leq 40 \\ 0 & \text{si non} \end{cases} \quad (V.2)$$

La figure (V.2) montre le graphe de cette fonction caractéristique, il est binaire, définit précisément l'appartenance à un sous-ensemble classique et peut être exprimée dans le contexte de l'algèbre booléenne en faisant correspondre le vrai booléen (1) avec l'intervalle [20÷40] et le faux booléen (0) avec l'union d'intervalles [0÷20 [U] 40÷1 [.

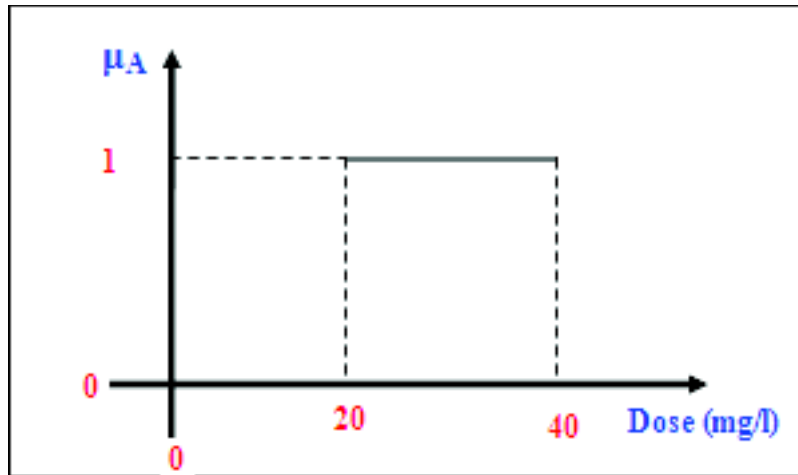


Figure V.2 Fonction caractéristique du sous-ensemble classique A.

On peut définir le degré d'appartenance de la variable dose de coagulant à l'ensemble (faible) comme le (degré de vérité) de la proposition (la dose de coagulant est faible).

Dans ce cas la dose de coagulant peut être :

- Faible :  $\mu_{faible}(dose) = 1, \mu_{moyenne}(dose) = 0, \mu_{elevee}(dose) = 0$
- Moyenne :  $\mu_{faible}(dose) = 0, \mu_{moyenne}(dose) = 1, \mu_{elevee}(dose) = 0$
- Élevée :  $\mu_{faible}(dose) = 0, \mu_{moyenne}(dose) = 0, \mu_{elevee}(dose) = 1$

Elle ne peut pas prendre deux qualificatifs à la fois (figure V.3).

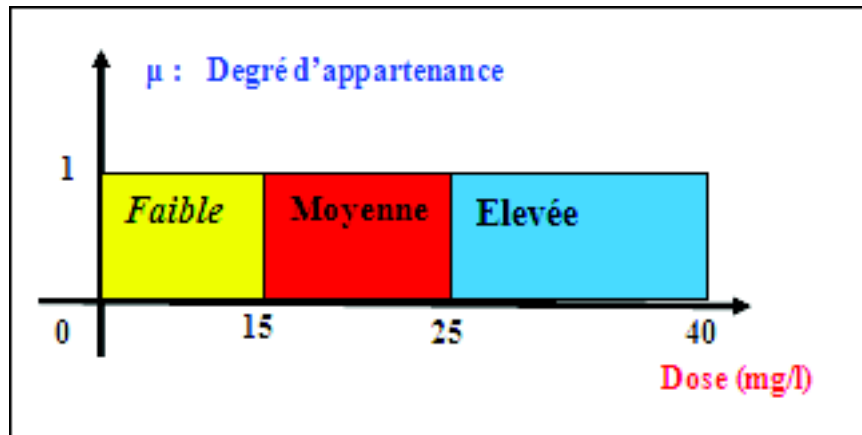


Figure V.3 Exemple d'ensembles considérés en logique booléenne

#### V.4.1.2. Sous-ensembles flous

La théorie des sous-ensembles flous permet de s'affranchir du caractère binaire ou booléen de l'appartenance à un ensemble, en introduisant une notion d'appartenance graduelle. Ainsi, on peut définir un sous-ensemble flou A de S par sa fonction d'appartenance  $\mu_A$ , qui associe à tout élément x de S une valeur réelle  $\mu_A(x)$  dans l'intervalle  $[0;1]$ , indiquant le degré d'appartenance de x à A (Ciofalo, 2005).

Considérons par exemple l'ensemble B des doses de coagulant avoisinant les 30 mg/litre, le graphe d'une fonction d'appartenance  $\mu_B$  possible pour B est représenté par la figure ci après (figure V.4).

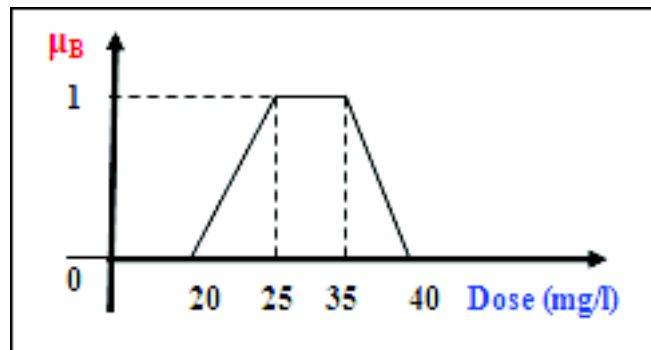


Figure V.4 Fonction caractéristique du sous-ensemble flou B

Donc en logique floue, le degré d'appartenance devient une fonction qui peut prendre une valeur réelle comprise entre 0 et 1 inclus.  $\mu_{moyenne}$  (dose), par exemple, permet de quantifier le fait que la dose de coagulant puisse être considérée comme moyenne, dans ce cas, la dose de coagulant peut être considérée, à la fois, comme faible avec un degré d'appartenance de 0,2 et comme moyenne avec un degré d'appartenance de 0,8 (figure V.5) :

$$\mu_{faible}(dose) = 0.2, \mu_{moyenne}(dose) = 0.8, \mu_{elevée}(dose) = 0$$

Figure V.5 Exemple d'ensembles considérés en logique floue.



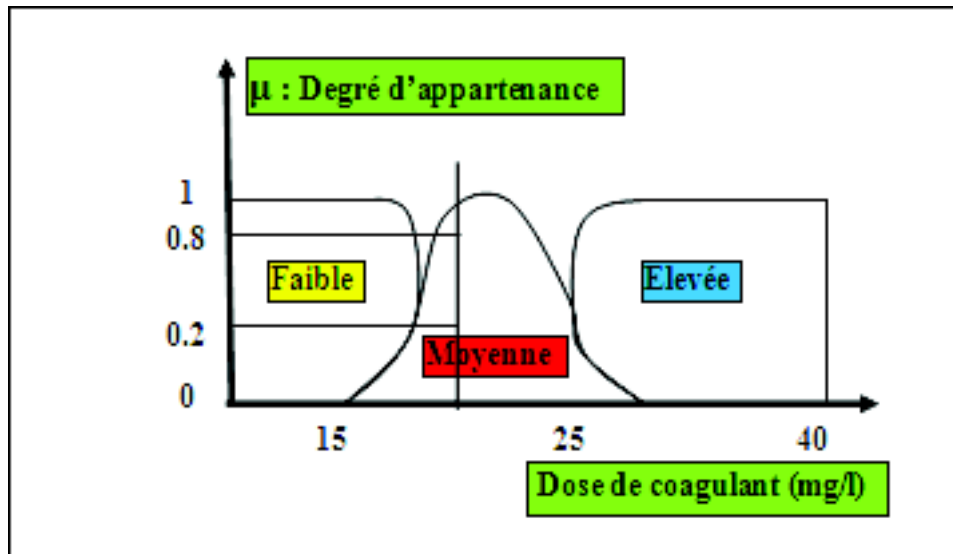


Figure V.5 Exemple d'ensembles considérés en logique floue.

On peut alors remarquer qu'un sous-ensemble classique est un sous-ensemble flou dont la fonction d'appartenance a une forme particulière, puisqu'elle prend ses valeurs uniquement dans le doublet  $\{0; 1\}$ .

Pour la variable floue  $x$ , on définit un ensemble flou  $A$  sur un univers de discours  $X$  par une fonction degré d'appartenance :

$$\begin{aligned} \mu_A : X &\rightarrow [0; 1] \\ x &\mapsto \mu_A(x) \end{aligned} \tag{V.3}$$

L'univers de discours est l'ensemble des valeurs réelles que peut prendre la variable floue  $x$  et  $\mu_A(x)$  est le degré d'appartenance de l'élément  $x$  à l'ensemble flou  $A$  (figure V.6).

Plus généralement, le domaine de définition de  $\mu_A(x)$  peut être réduit à un sous-ensemble de  $X$  (Zadeh, 1965), on peut ainsi avoir plusieurs fonctions d'appartenance, chacune caractérisant un sous-ensemble flou, c'est par l'association de tous les sous-ensembles flous de l'univers de discours, que l'on obtient l'ensemble flou de la variable floue  $x$  (Mamdani, 1975), par abus de langage, les sous-ensembles flous sont fort souvent confondus avec l'ensemble flou (baghli, 1999).

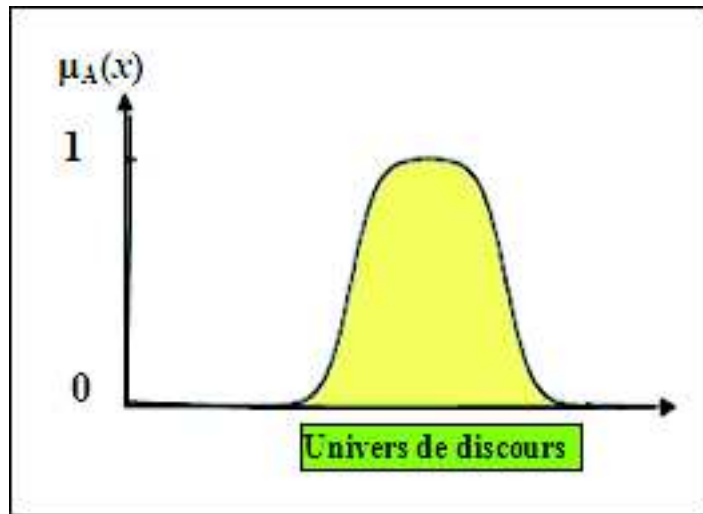


Figure V.6 Représentation d'un ensemble flou par sa fonction d'appartenance

Dans notre exemple, la variable floue est la dose de coagulant, l'univers de discours est l'ensemble des réels de l'intervalle  $[0, 40]$ , on attribue à cette variable trois sous-ensembles flous : faible, moyenne et élevée (figure V.7) , chacun est caractérisé par sa fonction degré d'appartenance:

$$\mu_{\text{faible}}(\text{dose}), \mu_{\text{moyenne}}(\text{dose}), \mu_{\text{élevée}}(\text{dose}).$$

Figure V.7 Cas de l'ensemble flou "moyenne" de la variable dose de coagulant

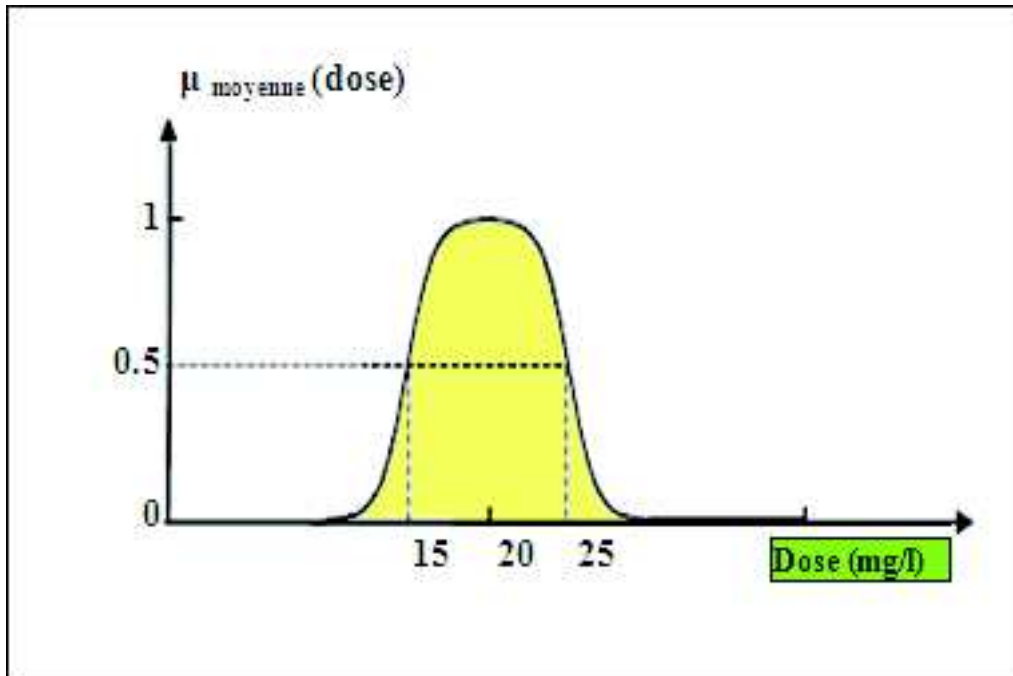


Figure V.7 Cas de l'ensemble flou "moyenne" de la variable dose de coagulant

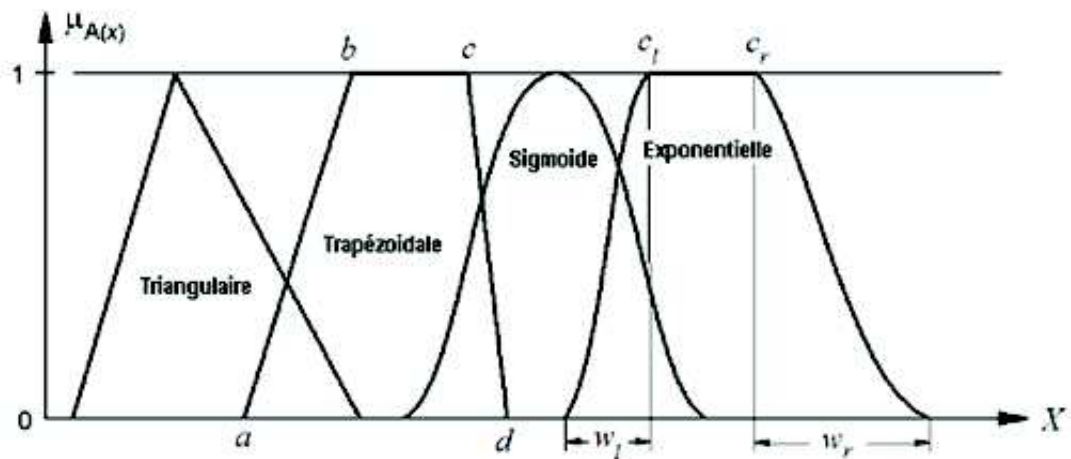


Figure V.8 Exemple de fonctions d'appartenance

Dans les univers de discours continus, les ensembles flous sont définis analytiquement par des fonctions d'appartenance, afin de faciliter les calculs, elles sont souvent paramétrées (Babuska, 1998).

Fonction d'appartenance trapézoïdale :

$$\mu(x; a, b, c, d) = \max\left(0, \min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right)\right) \quad (V.4)$$

Où  $a, b, c, d$  sont les coordonnées des sommets du trapèze (figure V.8). Il faut remarquer que si  $b = c$  on obtiendra une fonction d'appartenance triangulaire.

Fonction d'appartenance exponentielle par morceaux :

$$\mu(x; c_l, c_r, w_l, w_r) = \begin{cases} \exp\left(-\left(\frac{x-c_l}{2w_l}\right)^2\right) & \text{si } x < c_l \\ \exp\left(-\left(\frac{x-c_r}{2w_r}\right)^2\right) & \text{si } x > c_r \\ 1 & \text{les autres cas} \end{cases} \quad (V.5)$$

Où les paramètres  $\{w_l, w_r, c_l \text{ et } c_r\}$  correspondent aux points définis au niveau de la figure V.8, pour les conditions  $c_l = c_r$  et  $w_l = w_r$ , on obtiendra la fonction d'appartenance Gaussienne.

#### V.4.1.3. éléments caractéristiques des sous-ensembles flous

Bien que la fonction d'appartenance  $\mu_A$  d'un sous-ensemble flou A permette de le définir entièrement plusieurs éléments caractéristiques sont couramment employés pour les manipuler (ragot, 2003), nous les énumérons ici en considérant E comme étant l'univers de référence (figure V.9).

1. Le noyau : Le noyau  $Noy(A)$  d'un sous-ensemble flou  $A$  est l'ensemble des éléments de  $E$  dont le degré d'appartenance à  $A$  vaut 1 :

$$Noy(A) = \{e \in E \mid \mu_A(e) = 1\} \quad (V.6)$$

1. Le support : Le support  $Supp(A)$  d'un sous-ensemble flou  $A$  est l'ensemble des éléments de  $E$  dont le degré d'appartenance à  $A$  est non nul :

$$Supp(A) = \{e \in E \mid \mu_A(e) > 0\} \quad (V.7)$$

1. La hauteur : La hauteur  $h(A)$  d'un sous-ensemble flou  $A$  est la valeur maximale de la fonction d'appartenance  $\mu_A$ :

$$h(A) = \sup_{e \in E} \mu_A(e) \quad (V.8)$$

1. La cardinalité: La cardinalité  $|A|$  d'un sous-ensemble flou  $A$  est la quantité (floue) d'éléments de  $E$  qui appartiennent à  $A$  :

$$|A| = \sum_{e \in E} \mu_A(e) \quad (V.9)$$

1. L' $\alpha$ -coupe: L'  $\alpha$ -coupe  $A_\alpha$  d'un sous-ensemble flou  $A$  est l'ensemble des éléments de  $E$  qui appartiennent à  $A$  avec un degré au moins égale à  $\alpha$  :

$$A_\alpha = \{e \in E \mid \mu_A(e) \geq \alpha\} \quad (V.10)$$

Figure V.9 Principaux éléments caractéristiques d'un sous-ensemble flou  $A$

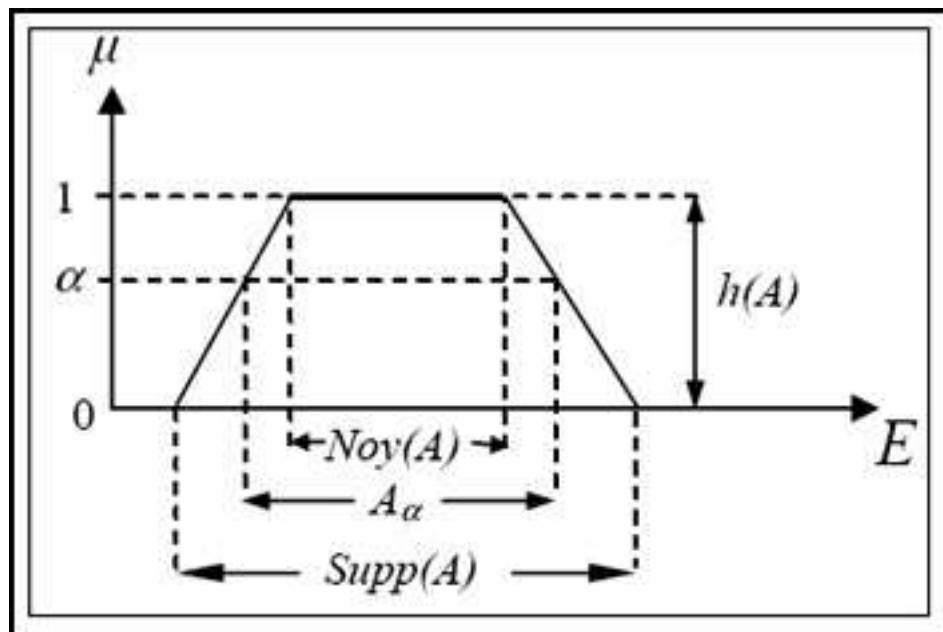


Figure V.9 Principaux éléments caractéristiques d'un sous-ensemble flou  $A$

#### V.4.1.4. opérations sur les sous-ensembles flous

La plupart des opérations que l'on peut effectuer sur les sous-ensembles classiques sont généralisables aux sous-ensembles flous, tout comme les sous-ensembles classiques peuvent être vus comme des cas particuliers des sous-ensembles flous, l'algèbre booléenne et ses trois éléments de base (NON, ET, OU) peut alors être considérée comme une restriction de l'algèbre floue.

Les opérateurs disponibles dans le cadre de la théorie des sous-ensembles flous sont nombreux et ont une propriété importante du point de vue de la fusion d'informations, ils donnent un résultat qui a la même interprétation en termes d'imprécision que leurs entrées, ce qui signifie que la décision peut être reportée à la fin du processus de combinaison, évitant ainsi des décisions partielles qui pourraient conduire à des imperfections (ciofola, 2005).

Nous présentons ci-après les opérations les plus courantes.

1- Égalité : Deux sous-ensembles flous (A) et (B) de S sont dits égaux s'ils ont des fonctions d'appartenance égales en tout point de S :

$$A = B \Leftrightarrow \forall x \in S, \mu_A(x) = \mu_B(x) \quad (V.11)$$

2- Inclusion : Le sous-ensemble flou (A) de S est inclus dans le sous-ensemble B de S si tout élément de S appartient moins à (A) qu'à (B):

$$A \subseteq B \Leftrightarrow \forall x \in S, \mu_A(x) \leq \mu_B(x) \quad (V.12)$$

3- Union : L'union de deux sous-ensembles flous A et B de S est le sous-ensemble flou constitué des éléments de S affectés du plus grand des degrés avec lesquels ils appartiennent à A et B :

$$\forall x \in S, \mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)) \quad (V.13)$$

La figure V.10 donne un exemple de graphe de la fonction d'appartenance correspondant à l'union de deux sous-ensembles flous A et B.

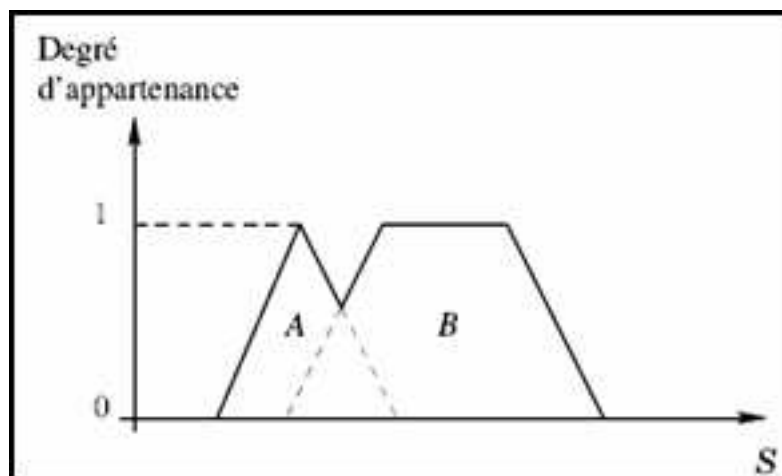


Figure V.10 Fonction d'appartenance de l'union A ∪ B

4- Intersection: L'intersection de deux sous-ensembles flous A et B de S est le sous-ensemble flou constitué des éléments de S affectés du plus petit des degrés avec lesquels ils appartiennent à A et B :

$$\forall x \in S, \mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)) \quad (V.14)$$

La figure V.11 donne un exemple de graphe de la fonction d'appartenance correspondant à l'intersection de deux sous-ensembles flous A et B.

On note que l'union et l'intersection peuvent être définies au moyen d'autres opérateurs, respectivement appelés conormes et normes triangulaires.

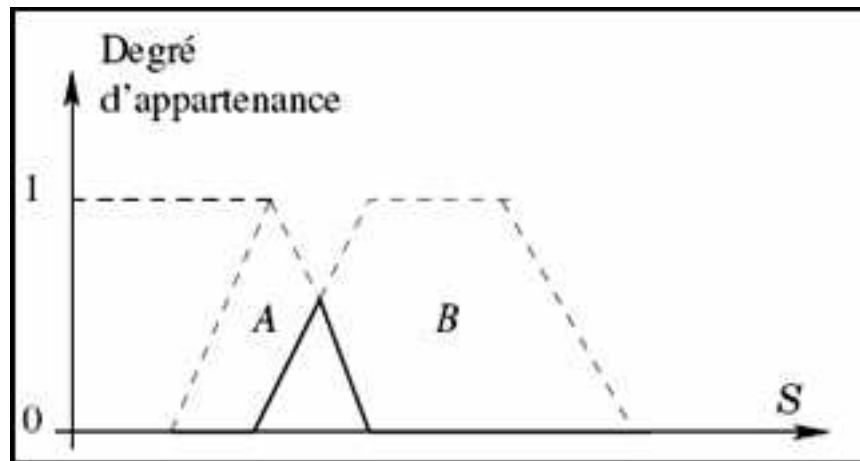


Figure V.11 Fonction d'appartenance de l'intersection  $A \cap B$

#### V.4.1.5. relations floues et compositions

Une relation floue R permet de représenter un lien entre deux univers  $E_1$  et  $E_2$  par un sous-ensemble flou défini dans  $E_1 \times E_2$  par sa fonction d'appartenance  $\mu_R : E_1 \times E_2 \rightarrow [0, 1]$ . La composition  $B : A \circ R$  est alors l'opération qui permet, à partir de la relation R et d'un sous-ensemble flou A de  $E_1$ , de déduire le sous-ensemble flou B de  $E_2$  défini par :

$$\forall e_2 \in E_2, \mu_B(e_2) = \sup_{e_1} \min(\mu_A(e_1), \mu_R(e_1, e_2)) \quad (V.15)$$

### V.4.2. Raisonnement à partir des sous-ensembles flous

Le raisonnement flou est une procédure d'inférence qui permet de déduire certaines conclusions à partir de relations fonctionnelles imprécises ou floues. Ce mécanisme constitue la base de la modélisation par les systèmes flous (Zadeh, 1979). Lors de leur définition les sous-ensembles flous peuvent être associés à un concept à décrire. Il existe alors un lien sémantique fort entre le sous-ensemble flou décrit par sa fonction d'appartenance et l'univers de référence sur lequel il est défini, dans ce contexte des mécanismes de raisonnement peuvent être mis en œuvre pour étendre les principes de la logique classique à ceux de la logique floue.

#### V.4.2.1. Formalisme linguistique

Dans le formalisme linguistique défini par Zadeh (Zadeh, 1965), les propriétés définies par les sous-ensembles flous se nomment des termes flous ou termes linguistiques (exemple Faible, moyenne, forte) chaque terme représente un sous-ensemble de valeurs numériques et caractérise ainsi la variable *floue/linguistique*, le domaine sur lequel ces termes et ces variables sont définies, l'ensemble de référence constitue *l'univers de discours*

(Les réels positifs pour l'objet dose de coagulant), le découpage de cet univers de discours par les termes flous est appelé *une partition floue*, lorsque l'univers de discours est totalement recouvert par les termes flous, et que pour toutes valeurs, la somme des degrés d'appartenance est égale à 1, on parle alors de partition floue forte (Jouffe, 1997).

#### V.4.2.2. variables linguistiques

Une variable linguistique est une variable dont les valeurs sont des mots ou des phrases exprimées dans une langue naturelle ou un langage artificiel, une variable linguistique est définie par le quadruplet suivant (Driankov, 1993 ; Nakoula, 1997):

$$\langle x\_nom, L(X), X, M_X \rangle \quad (V.16)$$

Où :

- $x\_nom$  est le nom de la variable linguistique, comme température, coagulant ...etc.,
- $L(X) = \{L1, \dots, Ln\}$  est l'ensemble des valeurs linguistiques que peut prendre la variable  $x\_nom$ . Par exemple  $L(X) = \{faible, moyenne, élevée\}$  peut être défini pour la variable température.  $L(X)$  est aussi appelé ensemble de termes (term-set), ensemble de symboles ou ensemble de référence (reference-set) associé à la variable  $x\_nom$ .
- $X$  est le domaine physique associé à la variable  $x\_nom$ , il est aussi appelé univers de discours. C'est en fait l'ensemble de toutes les valeurs numériques que peut prendre la variable numérique associée à la variable linguistique  $x\_nom$ , (par exemple,  $X$  peut être défini comme étant l'intervalle  $[0.. 50]$  pour la variable dose de coagulant).
- $M_X$  est une fonction sémantique qui associe à tout symbole de  $L(X)$  une signification floue.

La définition d'une variable linguistique regroupe donc à la fois des informations de nature numérique et linguistique, le lien entre ces deux types d'information nécessite la spécification d'interfaces. Si la signification floue  $M_X$  introduite dans la définition d'une variable linguistique représente l'interface "linguistique-numérique", la notion duale d'interface "numérique-linguistique" n'apparaît pas explicitement dans le concept de variable linguistique.

#### V.4.2.3. calcul des propositions floues

Une variable linguistique permet donc d'une part de synthétiser l'information manipulée grâce aux sous-ensembles flous, et d'autre part de représenter des concepts imprécis tels que l'homme en manipule quotidiennement, la détermination de la forme et de la position de ces sous-ensembles flous est un élément essentiel pour pouvoir effectuer des raisonnements valides robustes et compréhensibles. C'est pour cette raison que dans beaucoup de systèmes comme ceux destinés à l'aide à la décision les sous-ensembles flous sont définis a priori par des experts du domaine afin qu'ils représentent exactement leurs connaissances.



Le calcul d'une proposition floue consiste à déterminer sa fonction caractéristique « $\mu$ ». Lorsque la proposition n'est constituée que d'un prédicat  $P = \langle x \text{ est } X \rangle$ , sa fonction d'appartenance est simplement celle de  $X$ , « $\mu_P \text{ est } \mu_X$ », quant elle est composée de conjonctions ou de disjonctions de plusieurs prédicat, il est nécessaire de définir les opérateurs correspondants (jouffe, 1997), on note trois implémentations principales pour ces opérateurs logiques : la conjonction (  $\wedge$  ), la disjonction (  $\vee$  ) et la négation .

Soient  $X_1^i$  et  $X_2^i$  deux termes flous opérant sur les univers de discours  $X_1$  et  $X_2$  et composant la proposition  $P$  :

1. La conjonction :  $P = \langle x_1 \text{ est } X_1^i \text{ et } x_2 \text{ est } X_2^i \rangle$  est définie par  $\forall x_1 \in X_1, \forall x_2 \in X_2$  :

$$\mu_P = ET(X_1^i, X_2^i) = \begin{cases} \min(\mu_{X_1^i}(x_1), \mu_{X_2^i}(x_2)), & \text{Mamdani} \\ \mu_{X_1^i}(x_1) \times \mu_{X_2^i}(x_2), & \text{Larsen} \\ \max(0, \mu_{X_1^i}(x_1) + \mu_{X_2^i}(x_2) - 1), & \text{Lukasiewicz} \end{cases}$$

2. La disjonction :  $P = \langle x_1 \text{ est } X_1^i \text{ ou } x_2 \text{ est } X_2^i \rangle$  est définie par  $\forall x_1 \in X_1, \forall x_2 \in X_2$  :

$$\mu_P = Ou(X_1^i, X_2^i) = \begin{cases} \text{Mcc}( \mu_{X_1^i}(x_1), \mu_{X_2^i}(x_2) ), & \text{Mamdani} \\ \mu_{X_1^i}(x_1) + \mu_{X_2^i}(x_2) - (\mu_{X_1^i}(x_1) \mu_{X_2^i}(x_2)), & \text{Larsen} \\ \min(1, \mu_{X_1^i}(x_1) + \mu_{X_2^i}(x_2)), & \text{Lukasiewicz} \end{cases}$$

3. La négation : La négation de la proposition « $P = \langle x_1 \text{ est } X_1^i \rangle$ » s'exprime par « $\langle x_1 \text{ n'est pas } X_1^i \rangle$ » ou « $\langle x_1 \text{ est non } X_1^i \rangle$ » ou encore « $\langle x_1 \text{ est } \overline{X_1^i} \rangle$ ». Elle est définie par la fonction :  $\mu_{\overline{x}} : S_v \rightarrow [0,1]$  qui représente son degré de vérité :  $\mu_{\overline{x}} = n(\mu_x)$

Où  $n$  est une fonction de négation de  $[0,1]$  dans  $[0,1]$ , comme, par exemple, pour tout  $x$  dans  $[0,1]$ ,  $n(x) = 1 - x$ .

Après le calcul de toutes les propositions floues l'étape suivante consiste à calculer la contribution de chaque règle afin de pouvoir dans un troisième temps composer la proposition correspondant aux faits observés avec chaque règle.

#### V.4.2.4. Implication floue et règles floues

La notion d'implication entre deux propositions floues est assez délicate à généraliser par rapport à la logique classique. En effet, l'opérateur d'implication devra être choisi en fonction des propriétés souhaitées pour l'application visée (ciofola, 1997).

Une règle floue  $R$  : «Si  $A$  est  $A_i$  Alors  $B$  est  $B_j$ » est une relation entre deux propositions floues ayant chacune un rôle particulier. La première ( $A$  est  $A_i$ ) est appelée prémisse de la règle, alors que la seconde ( $B$  est  $B_j$ ) est la conclusion, dans le cas de propositions floues élémentaires, la prémisse et la conclusion sont définies à partir de deux variables linguistiques ( $A, E_A, F_A$ ) et ( $B, E_B, F_B$ ) qui décrivent les connaissances relatives aux univers de références  $E_A$  et  $E_B$  de manière à prendre en compte l'imprécision relative aux modalités de  $A$  et  $B$ .

Une proposition floue élémentaire est souvent insuffisante pour représenter l'ensemble des informations à manipuler. Plusieurs propositions floues peuvent alors être combinées pour enrichir et détailler la représentation (ragot, 2003), ainsi, la prémisse correspondant à «la dose de coagulant est élevée » peut correspondre à la jonction de deux autres propositions : «la turbidité est élevée et la conductivité est élevée ». Les opérations de conjonction, disjonction et de négations énoncées dans la partie précédente sont souvent utilisés à cet effet.

### V.4.3. Système d'inférence flou (SIF)

---

En général une simple règle de décision ne permet pas à elle seule de condenser toute l'information nécessaire à la représentation d'un problème complexe (Ragot, 2003). Des bases de règles doivent alors être élaborées pour aboutir à une description suffisamment complète qui puisse être utilisable. Ces bases de règles floues associées aux mécanismes de décision permettent de déduire de nouvelles connaissances, elles constituent des systèmes d'inférence floue (SIF).

Un modèle linguistique ou *système d'inférence flou* SIF (Renaud, 1999), est un système décrit à l'aide des règles du type *SI-ALORS* utilisant des propositions floues. C'est un système à base de connaissance qui contient des informations vagues, provenant de l'observation de phénomènes réels. Les systèmes flous ont été appliqués à des domaines très variés, comme le contrôle flou, la classification, la robotique, la reconnaissance des formes ou les séries temporelles, sous des appellations différentes : les mémoires associatives floues (kosko, 1991), les modèles flous (takagiet sugeno, 1985), les systèmes experts flous (Kandel, 1992), ou les systèmes à base de règles floues.

Nous considérerons pour l'instant que les vecteurs d'entrée du système sont multidimensionnels et la sortie monodimensionnelle (système MISO), sans perte de généralité, puisqu'un système à plusieurs sorties peut toujours se décomposer en un ensemble de systèmes à sortie unique. Soient  $x = \{x_1, \dots, x_n\}$  les variables d'entrée du système appartenant aux ensembles de référence  $x = x_1^X \dots x_r^X$  et  $y$ , la variable de sortie appartenant à l'espace  $Y$ .

Prenons une règle linguistique  $R_i$  de la forme générale suivante :

$$R_i : \text{Si } x \text{ est } A_i \text{ Alors } y \text{ est } B_i, \quad i = 1, \dots, r \quad (\text{V.17})$$

Où :  $r$  dénote le nombre de règles du modèle,  $x$  est la variable de l'*antécédent* (*prémisse*), qui représente l'entrée du système et  $y$  est la variable du *conséquent* (*conclusion*), qui représente la sortie du système flou.  $A_i$  et  $B_i$  sont des termes linguistiques qui correspondent aux valeurs qualitatives associées aux variables de base  $x$  et  $y$ . Ces termes linguistiques sont décrits par des ensembles flous définis par des fonctions d'appartenance  $\mu$ , qui établissent une correspondance de chacun des univers de discours vers l'intervalle  $[0,1]$ , ainsi :

$$\mu_{A_i}(x): X \rightarrow [0,1], \mu_{B_i}(y): Y \rightarrow [0,1] \quad (V.18)$$

En revenant sur la règle linguistique générale décrite par l'expression (V.17), les systèmes flous basés sur des règles du type « Si...Alors » ont des antécédents (prémises) et des conséquents (conclusions) qui sont spécifiés de manière symbolique, dans le cadre de la modélisation de systèmes, l'exploitation d'une telle connaissance nécessite en général la mise en place d'interfaces numérique/symbolique (N/S) et symbolique/numérique (S/N), ces dernières sont en effet des passerelles indispensables à l'établissement d'un lien entre l'ensemble de règles (*base de règles*) qui interfacent le système flou et le procédé, sur lequel seules des mesures et des actions numériques sont envisageables (Grisales palacio, 2007).

Un système d'inférence flou est essentiellement composé de deux blocs fonctionnels différents (Renaud, 1999) :

- 1) Une base de connaissance constituée d'une base de règles floues et d'une base de données définissant les fonctions d'appartenance des ensembles flous ;
- 2) Un mécanisme d'inférence flou qui détermine la sortie du système sous la forme d'un ensemble flou.

On peut y ajouter deux blocs (figure V.12) :

- 1) Une interface de fuzzification, qui transforme des entrées ponctuelles en ensembles flous, le cas échéant ;
- 2) Une interface de défuzzification, qui transforme la sortie floue en sortie ponctuelle.

Les deux premiers blocs définissent des relations entre les entrées et sorties floues du système. Les entrées du système  $\{x_1, \dots, x_r\}$  peuvent être fixes ou floues.

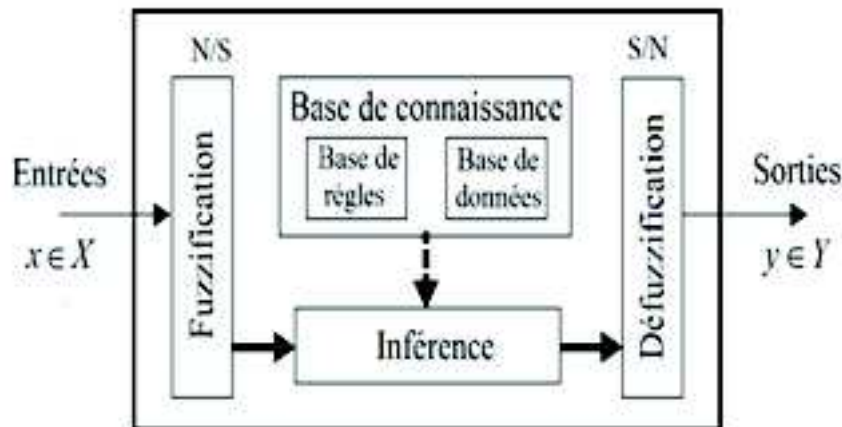


Figure V.12 Architecture d'un système flou

### V.4.3.1. la fuzzification

La fuzzification est la première étape du traitement flou, elle consiste à caractériser au moyen de sous-ensembles flous les valeurs des variables d'entrées représentant l'état du système. La fuzzification n'est qu'une étape technique permettant d'introduire des entrées observées fixes dans le mécanisme de raisonnement proprement dit en convertissant artificiellement ces entrées en ensembles flous (*fuzzification symbolique*), si ces données sont déjà définies par des ensembles flous, cette étape est donc inutile.

Dans le cas contraire (*fuzzification numérique*), on distingue la fuzzification dite «singleton» qui consiste simplement à transformer un élément fixe  $x_k$  de  $X_k$  en un sous-ensemble classique  $\{x_k\}$  de  $X_k$ , notée.  $X_k$  Ce type de fuzzification ne fait pas intervenir l'imprécision éventuelle des données, la méthode du singleton demeure la plus simple, et la plus fréquemment employée. Une autre approche, la fuzzification «non singleton» intègre l'imprécision ou le bruit dû à l'observation, en représentant l'entrée  $x_k$  sous la forme d'un ensemble flou quelconque  $X_k$ , comme par exemple un nombre flou gaussien centré en  $x_k$  (Renaud, 1999). Dans la littérature des SIF, la fuzzification singleton est de loin la plus utilisée, quelques rares cas illustrent cependant une fuzzification non-singleton (Palm, 1994 ; Galichet, 1994).

### V.4.3.2. base de règles et mécanisme d'inférence

L'évaluation des règles floues, autrement appelée *inférence floue*, consiste à mettre en œuvre les règles floues, établies au préalable, afin de calculer les sorties floues à partir des entrées issues de la fuzzification.

En logique floue, les sous-ensembles flous en entrée d'une règle floue ne correspondent pas exactement aux sous-ensembles flous des prémisses de cette règle, initialement définies. Pour obtenir une conclusion à partir d'une règle dont la prémisse est différente de la prémisse réelle, il est nécessaire d'appliquer la règle du *modus ponens généralisé* (MPG), introduite par ZADEH (Zadeh, 1965).

Soit une règle floue de la forme «Si X est A, alors Y est B». En notant « $p = X \text{ est } A$ » et

« $q = Y \text{ est } B$ », il peut exister une proposition « $p' = X \text{ est } A'$ », voisine de  $p$ , avec « $p \neq p'$ », qui implique une conclusion à la règle « $q' = Y \text{ est } B'$ », avec « $q' \neq q$ », les fonctions d'appartenance  $\mu_A$  et  $\mu_{A'}$  des deux sous-ensembles flous A et A' sont alors sensiblement différentes.

Le modus ponens généralisé (MPG) permet alors, à partir d'une prémisse  $p'$ , de déterminer une conclusion  $q'$  à la règle floue établie préalablement,  $p \rightarrow q$ . Il s'écrit de la manière suivante :

Si X est A, alors Y est B, et X est A', alors Y est B'

Le sous-ensemble flou B' induit est déterminé par une fonction d'appartenance,  $\mu_{B'}$  telle que :

$$\mu_{B'}(x) = \sup_{y \in X} (\mu_{A \rightarrow B}(x, y) * \mu_A(x)) \quad (V.19)$$

Où :  $\mu_{A \rightarrow B}$  est la fonction d'appartenance caractérisant la relation  $A \rightarrow B$ .

Où :  $\mu_{A \rightarrow B}$  est la fonction d'appartenance caractérisant la relation  $A \rightarrow B$ .

\* est un opérateur du modus ponens généralisé, les deux principaux opérateurs les plus fréquemment rencontrés sont l'opérateur minimum de ZADEH, et l'opérateur produit.

En conclusion, on peut noter que pour un vecteur d'entrée multidimensionnels du système « $x_1, \dots, x_n$ » et un vecteur de sortie multidimensionnels « $y_1, \dots, y_m$ », la base de règles floues se compose comme suit (castalleno, 2000) :

$$R_k = IF \underbrace{(x_1 \text{ est } A_1^k) \text{ et } \dots \text{ et } (x_n \text{ est } A_n^k)}_{\text{antécédent}} ALORS \underbrace{(y_1 \text{ est } B_1^k) \text{ et } \dots \text{ et } (y_m \text{ est } B_m^k)}_{\text{conséquent}} \quad (V.20)$$

Où  $A_i^k, i=1, \dots, n$  et  $B_j^k, j=1, \dots, m$  correspondents aux variables linguistiques, représentant respectivement les vecteurs d'entrée et de sortie du système, la partie antécédente (prémisse) de la règle, peut être un sous ensemble flou multidimensionnel noté  $H^k$ , obtenu par intersection (voir section V.4.1.4), ou bien des sous ensembles flous multivariés  $A_i^k, i=1, \dots, n$ . Quand l'opérateur d'intersection est utilisé le sous-ensemble flou multidimensionnel peut être représenté par le biais des fonctions d'appartenance des sous-ensembles flous multi variés :

$$\mu_{H^k}(X) = \prod_{i=1}^n \mu_{A_i^k}(x_i) \quad (V.21)$$

La deuxième partie est appelée conséquent, ou conclusion de la règle.

### V.4.3.3. la défuzzification

La défuzzification transforme la partie floue issue de l'inférence en une valeur numérique  $y \in Y$ , le résultat de l'inférence peut être soit une partie floue numérique, soit une partie floue symbolique, afin de les transformer en une valeur numérique, on aura besoin de deux types de défuzzification: une défuzzification numérique  $\delta_1$ , et une défuzzification symbolique  $\delta_2$  (nakoula, 1997), le but de la défuzzification est de trouver une sortie précise  $y^* = y_0$  qui corresponde au mieux à la caractérisation floue.

Il existe plusieurs techniques de défuzzification (Brown et Harris, 1994 ; Jang et Sun 1995 ; Lee, 1990 ; Wang, 1994), cependant la méthode la plus la plus courante et la plus répandait est celle du centre de gravité.

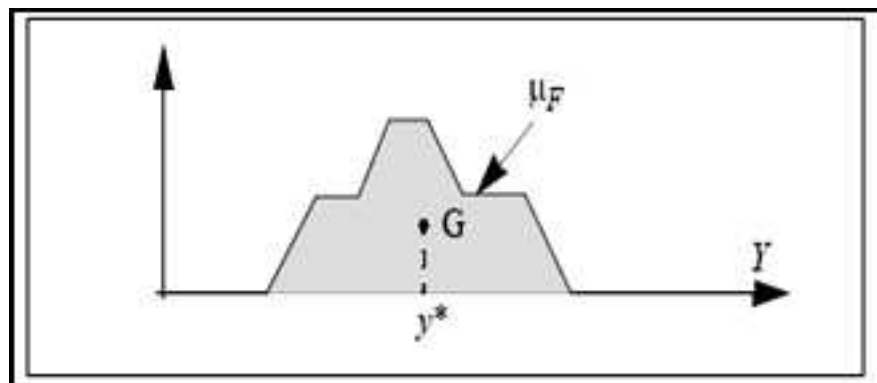


Figure V.13 Défuzzification par le centre de gravité

Soit  $F \in F(Y)$ , la partie floue numérique issue de l'inférence. La valeur numérique  $y^*$  produite par l'application de la méthode du centre de gravité est donnée par:

$$y^* = \mathcal{D}_1(F) = \frac{\int_{y^*} \mu_F(y) \cdot y \cdot dy}{\int_{y^*} \mu_F(y) \cdot dy} \quad \text{si } Y \text{ est continu} \quad (V.22)$$

$$y^* = \mathcal{D}_1(F) = \frac{\sum \mu_F(y) \cdot y}{\sum \mu_F(y)} \quad \text{si } Y \text{ est non (discret)} \quad (V.23)$$

La valeur  $y^*$  produite par la défuzzification d1 est définie comme étant la coordonnée  $y$  du centre de gravité (G) de la surface située sous  $\mu_F$  (Figure V.13).

#### V.4.4.Types de Système d'inférence flou

Si d'un point de vue mathématique les systèmes sont classés selon la nature des équations qui les caractérisent (linéaires, non linéaires, ...etc.), les systèmes flous sont quant à eux répertoriés selon leur nature structurelle (boukezzoula, 2000).

On distingue classiquement deux grandes familles de systèmes flous: systèmes flous à conclusions symboliques (systèmes de Mamdani) (Mamdani, 1974 ; Mamdani, 1975)et systèmes flous à conclusion fonctionnelle (systèmes flous de Takagi-Sugeno-Kang *tsk*) (takagiet sugeno, 1983 ; takagiet sugeno, 1985 ; Sugeno et Kang, 1986).

##### V.4.4.1.Sif de type Mamdani

Les SIF de type Mamdani (*Modèle flou linguistique*)(zede, 1973 ; Mamdani, 1977)se caractérisent par le choix de l'opérateur min que cela soit pour la conjonction dans le calcul des propositions ou pour l'implication T-norme (Jouffe, 1997).

Soit le SIF réduit aux deux règles linguistiques suivantes :

**R<sub>1</sub>** : **Si** turbidité est élevée **et** conductivité est élevée **alors** dose de coagulant est fort.

**R<sub>2</sub>** : **Si** turbidité est moyenne **et** conductivité est moyenne **alors** dose de coagulant est moyenne.

##### V.4.4.2.Sif de type Takagi\_Sugeno

Les SIF de Sugeno sont, comme ceux de Mamdani, construits à partir d'une base de règles «Si ... alors ...». Si les prémisses sont toujours exprimées linguistiquement et donc similaires à celles utilisées dans le modèle de mamdani, les conclusions sont par contre de nature numérique. Elles peuvent prendre différentes formes : équations linéaires ou polynomiales pour les SIF d'ordre un (TS1), ou directement des conclusions constantes pour les SIF d'ordre zéro (TS0). Dans ce cas la défuzzification fait partie intégrante du système d'inférence.



Un système flou de type Takagi\_Sugeno (TS) (Takagiet Sugeno, 1985; Sugenoet Kang, 1988), utilise des règles suivantes :

$$\begin{cases} R^{(i)} = SI\ x_1\ est\ A_1^{(i)}\ ET\ x_2\ est\ A_2^{(i)}\ ET\ \dots\ x_n\ est\ A_n^{(i)}\ ALORS\ y^i = f^i(x_1, x_2, \dots, x_n) = h^{(i)}(x) \\ R^{(i)} = SI\ x_1\ est\ A_1^{(i)}\ ET\ x_2\ est\ A_2^{(i)}\ ET\ \dots\ x_n\ est\ A_n^{(i)}\ ALORS\ y^i = f^i(x_1, x_2, \dots, x_n) = h^{(i)}(x) \end{cases} \quad (V.24)$$

Les fonctions de sortie  $h^{(i)}(x)$ , peuvent être en principe des fonctions arbitraires des entrées, mais d'une manière générale elles sont choisies telles qu'elles soient une combinaison linéaire des entrées.

Ce type de système est appelé un modèle Takagi\_Sugeno linéaire, ou modèle Takagi\_Sugeno homogène, ou encore si un terme constant est ajouté, on parle de modèle Takagi\_Sugeno affine (Morere, 2001).

Le paramétrage d'un SIF de type TS se limite généralement à la définition des éléments suivants :

- 1) Le nombre de fonctions d'appartenance sur chaque domaine d'entrée ;
- 2) Le placement de ces fonctions d'appartenance ;
- 3) La valeur numérique des conclusions.

Deux sources d'information sont habituellement utilisées (Guillaume, 2001) :

- 1) La connaissance a priori, exploitée de façon essentiellement qualitative ;
- 2) Des données numériques, qui vont permettre l'optimisation des règles floues.

#### V.4.5. Caractéristiques d'un Sif

Un Sifest en tout premier lieu caractérisée par son type (Mamdani, Takagi\_Sugeno) et par les partitions floues qu'il met en œuvre. Cependant, la définition totale d'un Sif passe par la spécification d'un ensemble de caractéristiques dites structurelles et paramétriques (jouffe, 1997).

##### V.4.5.1. Caractéristiques structurelles

Ces caractéristiques spécifient tous les éléments du SIF qui influent sur sa structure, ces éléments sont constituées par :

- 1) Le type de fonction d'appartenance utilisée (triangle, trapèze, sigmoïde...) pour chaque terme linguistique ;
- 2) Le nombre de termes linguistiques pour chaque variable ;
- 3) Le nombre de règles ;
- 4) Les variables participant à ces règles.

#### V.4.6. Méthodes d'optimisation des SIF

Les problèmes de synthèse des SIF résident dans la détermination des paramètres des fonctions d'appartenance ainsi que dans l'élaboration des règles floues, ces problèmes



peuvent être scindé en deux groupes interdépendants : l'identification de la structure du SIF et l'ajustement (ou réglage) des paramètres (Glorennec, 1999 ; Lin et Lee, 1991 ; Mitraet Hayashi, 2000).

L'identification de la structure du SIF englobe le problème de partitionnement de l'espace d'entrée, la sélection des variables de la prémisse et celles de la conclusion, ainsi que la définition du nombre de règles floues, tandis que les procédures de réglage des paramètres ont pour objet de déterminer les paramètres des fonctions d'appartenance et des conclusions.

D'une manière générale, l'optimisation d'un SIF se décompose en deux sous problèmes (Glorennec, 1999) :

1) L'optimisation de la structure et des paramètres, qui a pour objectif de déterminer les entrées significatives, le nombre de fonctions d'appartenance par entrée, les paramètres des fonctions d'appartenance et les conclusions des règles ;

2) L'optimisation des paramètres pour une structure donnée, qui a pour but de déterminer les paramètres des fonctions d'appartenance, et les conclusions des règles.

#### **V.4.7. Théorème d'approximation universelle**

---

Dans (Wang et Mendel, 1992a ; Wang et Mendel, 1992b), Wang montre que les systèmes flous sont des approximateurs universels s'ils respectent certains critères, il voit son résultat comme un théorème d'existence d'un système flou optimal pour une large variété d'applications, d'un point de vue mathématique, la construction d'un système flou consiste à trouver une relation de correspondance entre l'espace d'entrée et l'espace de sortie pour approximer, avec un degré de précision donné, la fonction ou la dynamique désirée. Buckley de son côté (Buckley, 1992 ; Buckley et Hayashi, 1993) démontre qu'une classe de systèmes flous qu'il appelle modèles de Takagi\_Sugeno sont des approximateurs universels et des contrôleurs universels. Le mot "approximateurs universel" dans les travaux de Buckley a la même signification que dans les travaux de Wang. Buckley écrit "pour un procédé donné, si on peut le commander par un contrôleur classique, on peut également le commander par un contrôleur de Takagi Sugeno"(boukezzoula, 2000).

### **V.5. systèmes hybrides neuroflous**

Les systèmes neuro-flous permettent de combiner les avantages de deux techniques complémentaires. Les systèmes flous fournissent une bonne représentation des connaissances. L'intégration de réseaux de neurones au sein de ces systèmes amélioré leurs performances grâce à la capacité d'apprentissage des réseaux de neurones. Inversement, l'injection de règles floues dans les réseaux de neurones, souvent critiquées pour leur manque de lisibilité, clarifie la signification des paramètres du réseau et facilite leur initialisation, ce qui représente un gain de temps de calcul considérable pour leur identification (Renaud, 1999).

Cette idée à former par la suite la base des systèmes neuroflous (neurofuzzy modeling), de nombreux auteurs ont donc tout naturellement cherché à combiner ces deux paradigmes depuis le début des années 90, et ceci de plusieurs manières (Brownet Harris, 1994; Janget Sun, 1993; Wang et Mendel, 1992a, Magdalena, 1995), dans ce sens (Hayashiet

buckley, 1994) ont montré qu'un réseau de neurone de type feedforward est capable d'assurer l'approximation de toutes règles à base de système d'inférence flou, réciproquement (Li et Chen, 2000) ont montré qu'un réseau de neurone de type feedforward peut être rapproché par un système à base de règles floues (Abraham, 2005).

Dans les sections suivantes, nous discuterons brièvement les modèles neuroflous hybrides qui font usage de complémentarités entre les réseaux de neurones (RNA) et les systèmes d'inférence flous (SIF) type Mamdani ou Takagi\_Sugeno.

### V.5.1. Système neuroflou type Mamdani.

Le système neuroflou hybride de Mamdani utilise la technique d'apprentissage supervisé (retro propagation) pour l'apprentissage des paramètres des fonctions d'appartenance. L'architecture du système neuroflou hybride de Mamdani est illustrée la Figure (V.14), les détails de chaque couche sont ainsi présentés (Abraham, 2005) :

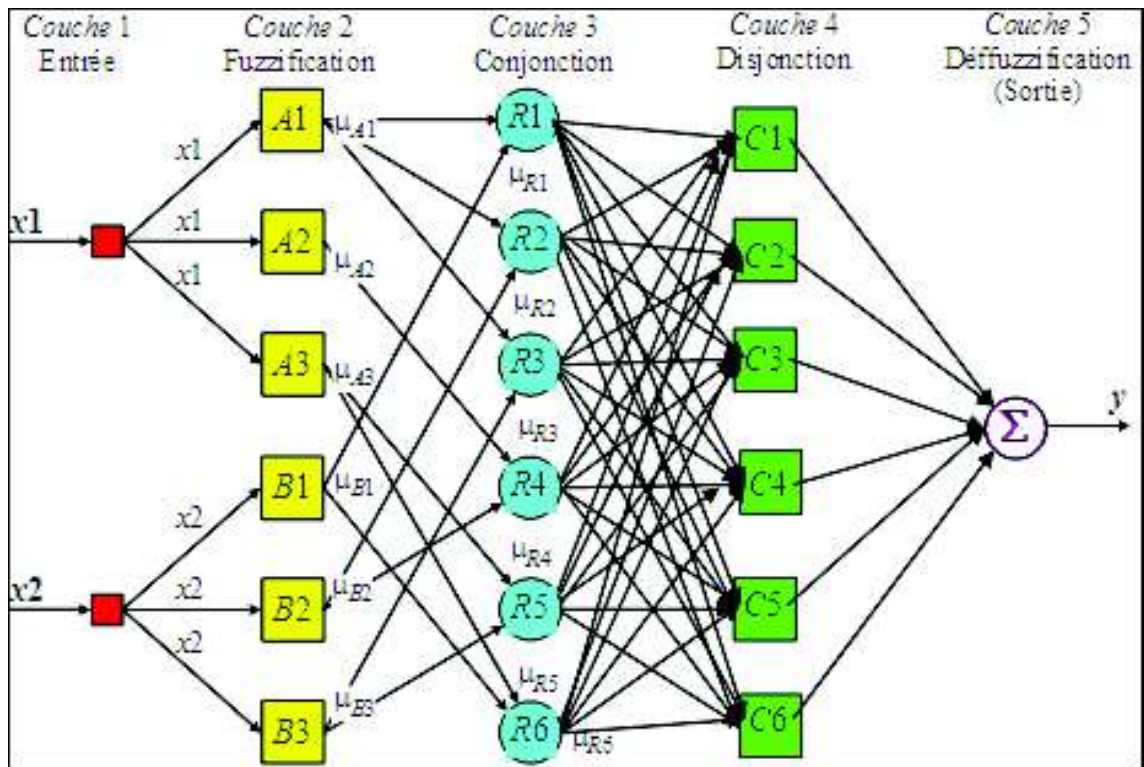


Figure V.14 Système Neuroflou type Mamdani

*Couche 1* : Agit comme un tampon qui transmet simplement les données d'entrée à la couche de Fuzzification. Chaque nœud dans cette couche (la couche d'entrée) correspond à une variable d'entrée, transmet seulement et directement des valeurs d'entrée à la prochaine couche. Le poids dans la couche 1 correspond à l'unité (1), on a :

$$y_i^{(1)} = x_i^{(0)} \tag{V.25}$$

*Couche 2* : Les neurones de cette couche (la couche de Fuzzification) réalisent chacun un ensemble flou ( $A_1, A_2, A_3, B_1, B_2, B_3$ ) qui servira dans les antécédents

(prémises) des règles. Chaque neurone reçoit une valeur d'entrée précise et génère son degré d'appartenance à l'ensemble flou du neurone.

- *Couche 3* : Elle réalise les conjonctions des antécédents (prémises) de règles floues. Chaque neurone dans cette couche (la couche de permise des règles) reçoit comme entrée les degrés d'appartenance ( $\mu_{A1}, \dots$  etc.) définis dans la couche 2, l'intersection floue est réalisée avec l'opérateur produit :

$$y_i^{(3)} = x_{1i}^{(3)} \times x_{2i}^{(3)} \times \dots \times x_{mi}^{(3)} \quad (V.26)$$

$$y_{R1}^{(3)} = \mu_{A1} \times \mu_{B1} = \mu_{R1} \quad (V.27)$$

- *Couche 4* : Elle réalise les disjonctions des antécédents (prémises) des règles floues. Chaque neurone dans cette couche (la couche de la partie conclusion des règles) reçoit comme entrée les degrés d'appartenance définis dans la couche 3, l'union floue est réalisée à l'aide de l'opérateur ou probabiliste :

$$y_i^{(4)} = x_{1i}^{(4)} \oplus x_{2i}^{(4)} \oplus \dots \oplus x_{ni}^{(4)} \quad (V.28)$$

$$y_{C1}^{(4)} = \mu_{R3} \oplus \mu_{R6} = \mu_{C1} \quad (V.29)$$

Le nombre de nœuds dans cette couche sera égale au nombre de règles.  $m \subset 1$  représente la force combinée des conjonctions implémentées par les neurones  $R3$  et  $R6$ . Les nœuds de cette couche réalisent donc deux tâches fondamentale : combiner les Antécédents (prémises) des règles, et déterminer le degré d'appartenance à la variable Linguistique de sorties.

- *Couche 5* : Chaque neurone dans cette couche (la couche de combinaison et de défuzzification) prend les ensembles flous précédents, et les combine en un seul ensemble flou, ce dernier est alors défuzzifié par une méthode standard.

## V.5.2. Système Neuroflou type Takagi Sugeno.

---

Le système neuroflou de type Takagi\_Sugeno (Jang et Sun, 1993 ; Jang et al., 1997), peut se voir à la fois comme des réseaux de neurones et comme des systèmes flous conventionnels. Il s'agit de systèmes flous auxquels on incorpore des concepts issus des réseaux de neurones : apprentissage des paramètres, ou des règles (Renaud, 1999).

Les systèmes neuroflous de type Takagi\_Sugeno utilisent l'algorithme de rétro propagation du gradient pour l'apprentissage des fonctions d'appartenance et la méthode de moindre carré pour déterminer les coefficients des équations linéaires pour les conclusions des règles, parmi ce type de modèles on peut cité la modèle ANFIS, utilisé dans le cadre de notre travail.

## V.6. le modèle utilisé: Adaptive neuro-Fuzzy Inference System (ANFIS)

Les premiers travaux sur l'optimisation des SIF étaient explicitement inspirés de l'algorithme de rétro propagation des réseaux de neurones (Glennec, 1999). Cette idée a conduit à la construction de la base du développement du modèle ANFIS (Jang, 1991a ; 1991b) (Système d'inférence flou organisé en réseau adaptatif), proposé par Jang (Jang, 1993). Ce dernier modèle est considéré comme un approximateur universel, est capable d'approximer n'importe quelle fonction avec un degré de précision très élevée (Jang et al., 1997).

Le modèle ANFIS est un modèle de génération automatique de règles floues basé sur le modèle d'inférence de type Takagi Sugeno (équation V.30), il utilise un algorithme d'apprentissage hybride (Jang et al., 1997) :

$$SI \ x_1 \text{ est } A_1^{(i)} \ ET \ x_2 \text{ est } A_2^{(i)} \ ET \dots \ ET \ x_m \text{ est } A_m^{(i)} \ ALORS \ y^i = f^i(x_1, x_2, \dots, x_m) \quad (V.30)$$

Où  $x_1, x_2, \dots, x_m$  sont des variables d'entrée et  $A_1, A_2, \dots, A_m$  sont des sous ensembles

Où  $x_1, x_2, \dots, x_m$  sont des variables d'entrée et  $A_1, A_2, \dots, A_m$  sont des sous ensembles flous.

- Lorsque  $y = \text{constante}$ , on obtient un modèle de Sugeno d'ordre zéro. La conclusion d'une règle est un singleton.
- Lorsque  $y$  est une combinaison linéaire des entrées :  $y = k_0 + k_1 x_1 + k_2 x_2 + \dots + k_m x_m$  on obtient un modèle de Sugeno de premier ordre.

Pour expliquer le modèle, nous considérons un système à deux entrées  $x_1$  et  $x_2$ , une sortie  $y$  et un modèle flou de type TSK de ce système, composé des deux règles suivantes:

$$SI \ x_1 \text{ est } A_1 \ ET \ x_2 \text{ est } B_1 \ ALORS \ y_1 = f_1(x_1, x_2) = a_1 x_1 + b_1 x_2 + c_1 \quad (V.31)$$

$$SI \ x_1 \text{ est } A_2 \ ET \ x_2 \text{ est } B_2 \ ALORS \ y_2 = f_2(x_1, x_2) = a_2 x_1 + b_2 x_2 + c_2 \quad (V.32)$$

Jang a proposé de représenter cette base de règles par le réseau adaptatif présenté par la Figure (V.15 et V.16). Le réseau adaptatif ANFIS est un réseau multicouche dont les connexions ne sont pas pondérées, ou ont toutes un poids de 1. Les nœuds sont de deux types différents selon leur fonctionnalité: les nœuds carrés (adaptatifs) contiennent des paramètres, et les nœuds circulaires (fixes) n'ont pas de paramètres. Toutefois, chaque nœud (carré ou circulaire) applique une fonction sur ses signaux d'entrées. La sortie  $O_i^k$  du nœud  $i$  de la couche  $k$  (appelé nœud  $(i, k)$ ) dépend des signaux provenant de la couche  $k-1$  et des paramètres du nœud  $(i, k)$ , c'est-à-dire (Nakoula, 1997) :

$$O_i^k = f(O_i^{k-1}, \dots, O_{n_{k-1}}^{k-1}, a_i, b_i, c_i, \dots) \quad (V.33)$$

Figure V.15 le modèle ANFIS

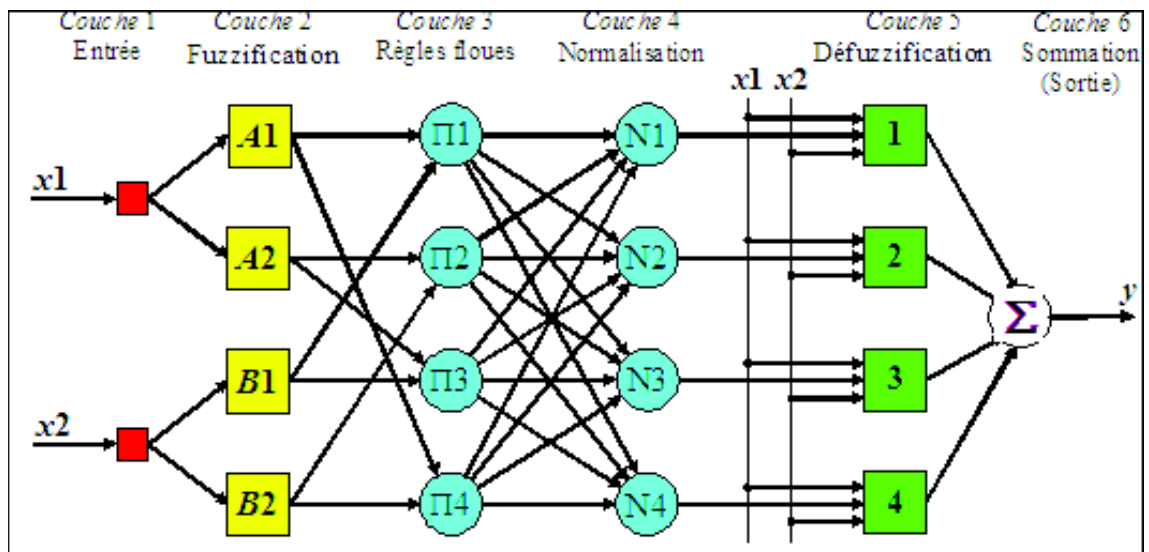


Figure V.15 le modèle ANFIS

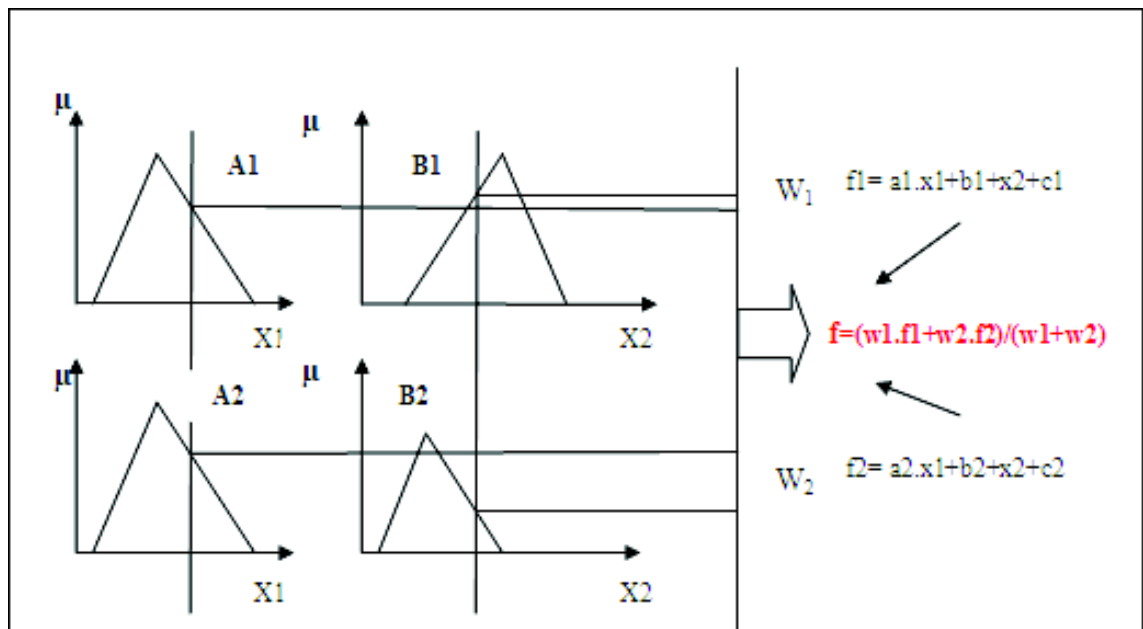


Figure V.16 le modèle flou de Seguro avec deux règles et deux entrées

Où  $n_{k-1}$  est le nombre de nœuds dans la couche  $k-1$ , et  $a, b, c \dots$  sont les paramètres du nœud  $(i, k)$ . Pour un nœud circulaire ces paramètres n'existent pas.

La description détaillée de l'architecture du modèle ANFIS, comme elle a été présentée la première fois dans (Jang, 1993), et présentée dans les travaux de plusieurs chercheurs (Nakoula, 1997) et (Abraham, 2005), est la suivante :

- Couche 1: Transmet simplement les données d'entrée à la couche de fuzzification.
- Couche 2 : Chaque nœud de cette couche est un nœud carré avec une fonction:

$$O_i^2 = \mu_{A_i}(x) \quad (V.34)$$

Où  $x$  est l'entrée du nœud  $i$ , et  $A_i$  le terme linguistique associé à sa fonction. En d'autres termes,  $O_i^2$  est le degré d'appartenance de  $x$  à  $A_i$ . Les paramètres d'un nœud de cette couche sont ceux de la fonction d'appartenance correspondante, dans le modèle de Jang, les fonctions d'appartenance sont des gaussiennes.

- Couche 3 :Chaque nœud  $i$  de cette couche est un nœud circulaire appelé P qui engendre en sortie le produit de ses entrées, ce produit représente le degré d'activation d'une règle :

$$w_i = \mu_{A_i}(x_1) \cdot \mu_{B_i}(x_2), i=1..2 \quad (V.35)$$

En d'autres termes, chaque neurone dans cette couche correspond à une règle floue Sugeno. Il reçoit les sorties des neurones de fuzzification et calcule son activation, la conjonction des antécédents est réalisée avec l'opérateur produit.

- Couche 4 :Chaque nœud de cette couche est un nœud circulaire appelé  $\Pi$  La sortie du nœud  $i$  est le degré d'activation normalisé de la règle  $i$ :

$$\bar{w}_i = \frac{w_i}{w_1 + w_2} \quad (V.36)$$

- Couche 5 : Chaque nœud de cette couche est un nœud carré avec une fonction réalisant le calcul:

$$O_i^5 = \bar{w}_i \cdot f_i = \bar{w}_i \cdot (a_i x_1 + b_i x_2 + c_i), i=1..2 \quad (V.37)$$

Où  $\bar{w}_i$  est la sortie de couche 4, et  $\{a_i, b_i, c_i\}$  est l'ensemble des paramètres de sortie

de la règle  $i$ . À signaler que chaque neurone  $i$  de cette couche est relié à un neurone de normalisation correspondant et aux entrées initiales du réseau

- Couche 6 :Le seul nœud de cette couche est un nœud circulaire qui effectue la somme des signaux provenant de la couche 5, c'est-à-dire :

$$O_i^6 = y = \sum_i \bar{w}_i \cdot f_i \quad (V.38)$$

On remarque que la sortie globale du réseau est équivalente à la sortie du modèle TSK.

### V.6.1. Apprentissage du modèle ANFIS

L'apprentissage à partir d'un ensemble de données concerne l'identification des paramètres des prémisses et des conséquences, la structure du réseau étant fixée.(Nakoula, 1997). L'algorithme d'apprentissage du modèle ANFIS se déroule en deux temps, où on estime



d'abord les paramètres des conséquents par une technique de moindres carrés et ensuite les poids du réseau par une descente de gradient. Chaque époque d'entraînement comprend une propagation en avant et une rétro-propagation. Durant la propagation en avant, les variables d'entrée servent à déterminer les sorties des neurones couche par couche, permettant de déterminer les valeurs des paramètres des conséquents. Durant la phase de rétro-propagation, l'algorithme de rétropropagation d'erreur est appliqué pour optimiser les poids des différentes couches.

### V.6.2. Mise en évidence

Reprenant en détaille le modèle ANFIS en démontrant mathématiquement son algorithme d'apprentissage (figure V.17).

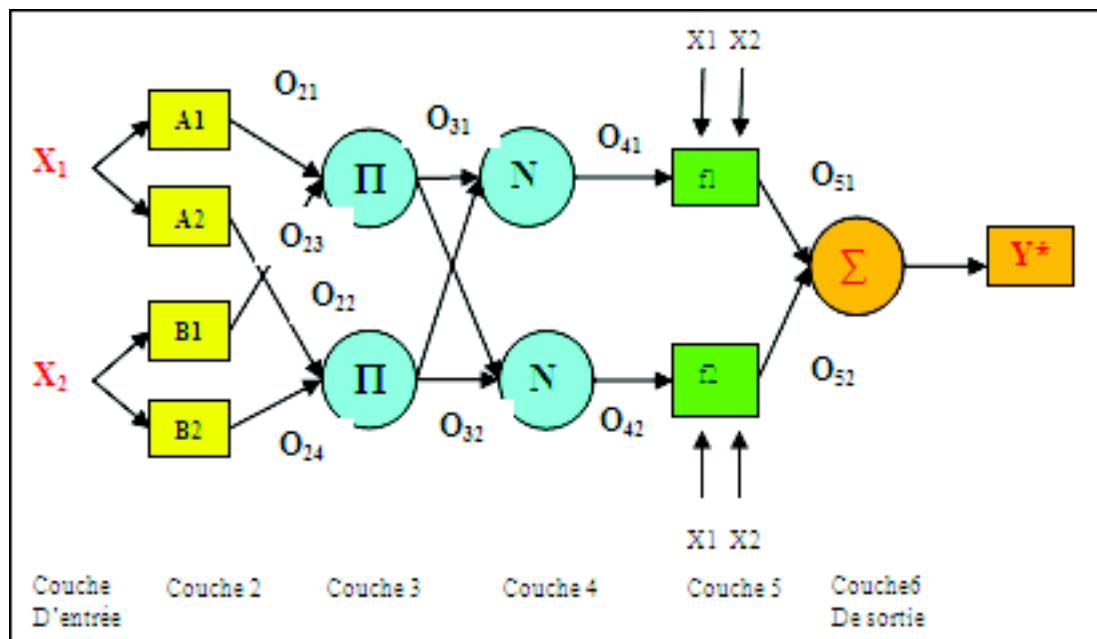


Figure V.17 le modèle ANFIS à deux règles et deux entrées

Supposant un problème quelconque décrit par deux règles floues,  $R_1$  et  $R_2$  pour deux entrées  $X_1$  et  $X_2$  et une sortie  $Y$  :

$$\begin{aligned}
 R_1: & \text{SI } x_1 \text{ est } A_1 \text{ ET } x_2 \text{ est } B_1 \text{ ALORS } y_1 = f_1(x_1, x_2) = Z_{11}x_1 + Z_{12}x_2 + Z_{13} \\
 R_2: & \text{SI } x_1 \text{ est } A_2 \text{ ET } x_2 \text{ est } B_2 \text{ ALORS } y_2 = f_2(x_1, x_2) = Z_{21}x_1 + Z_{22}x_2 + Z_{23}
 \end{aligned}
 \tag{V.39}$$

Quand on présente la variable numérique  $X(x_1, x_2)$ , le mécanisme d'inférence produit ce qui suit :

$$y^* = \frac{A_1(x_1)B_1(x_2)f_1(X) + A_2(x_1)B_2(x_2)f_2(X)}{A_1(x_1)B_1(x_2) + A_2(x_1)B_2(x_2)}
 \tag{V.40}$$



La variable observée  $X(x_1, x_2)$  est présentée à la couche 2 par les sorties de la couche 1. La sortie de la couche 2 est :

$$(O_{21}, O_{22}, O_{23}, O_{24}) = (A_1(x_1), A_2(x_1), B_1(x_2), A_2(x_2)) \quad (V.41)$$

Les sous ensembles flous  $A_i, B_i, i=1, 2$ , sont définis par plusieurs paramètres décrivant les fonctions d'appartenances les plus connus (triangulaires, rectangulaires, Etc.).

La couche3 représente un neurone flou, définie par le résultat d'agrégation de règles floues, en utilisant l'opérateur t-norme :

$$(O_{31}, O_{32}) = (A_1(x_1) \cdot B_1(x_2), A_2(x_1) \cdot B_2(x_2)) \quad (V.42)$$

La couche 4 représente la couche de normalisation, elle présente des sorties qui correspond à :

$$\begin{aligned} (O_{41}, O_{42}) &= \left( \frac{O_{31}}{O_{31} + O_{32}}, \frac{O_{32}}{O_{31} + O_{32}} \right) \\ &= \left( \frac{A_1(x_1) \cdot B_1(x_2)}{A_1(x_1) \cdot B_1(x_2) + A_2(x_1) \cdot B_2(x_2)}, \frac{A_2(x_1) \cdot B_2(x_2)}{A_1(x_1) \cdot B_1(x_2) + A_2(x_1) \cdot B_2(x_2)} \right) \end{aligned} \quad (V.43)$$

La sortie du neurone flou de la couche 5 est :

$$\begin{aligned} (O_{51}, O_{52}) &= (O_{41} \cdot f_1, O_{42} \cdot f_2) \\ &= \left( \frac{A_1(x_1) \cdot B_1(x_2) \cdot (Z_{11} \cdot x_1 + Z_{12} \cdot x_2 + Z_{13})}{A_1(x_1) \cdot B_1(x_2) + A_2(x_1) \cdot B_2(x_2)}, \frac{A_2(x_1) \cdot B_2(x_2) \cdot (Z_{21} \cdot x_1 + Z_{22} \cdot x_2 + Z_{23})}{A_1(x_1) \cdot B_1(x_2) + A_2(x_1) \cdot B_2(x_2)} \right) \end{aligned} \quad (V.44)$$

Finalement la sortie finale  $y^*$  est :

$$\begin{aligned} y^* &= O_{51} + O_{52} \\ &= \left( \frac{A_1(x_1) \cdot B_1(x_2) \cdot (Z_{11} \cdot x_1 + Z_{12} \cdot x_2 + Z_{13}) + A_2(x_1) \cdot B_2(x_2) \cdot (Z_{21} \cdot x_1 + Z_{22} \cdot x_2 + Z_{23})}{A_1(x_1) \cdot B_1(x_2) + A_2(x_1) \cdot B_2(x_2)} \right) \end{aligned} \quad (V.45)$$

Dans la structure du modèle ANFIS, les paramètres des prémisses et des conséquents jouent le rôle des poids. L'algorithme d'apprentissage du modèle ANFIS se résume donc, dans l'ajustement de ses paramètres, et cela pour une base de données représentative du problème à étudier, à savoir :

$$\left\{ (x_1^k, x_2^k), y^k \right\}, k=1, \dots, N \Leftrightarrow \left\{ (x^1, y^1), \dots, (x^N, y^N) \right\} \quad (V.46)$$

Où :  $x^k = (x_1^k, \dots, x_n^k) \in R^n$  et  $y^k \in R$ . Le modèle comme il a été expliqué utilise un nombre  $R_i, i=1, \dots, m$ , de règles floues de types {SI...ALORS} :

$$R_i: \text{SI } x_1^k \text{ est } A_i^1 \text{ et...et } x_n^k \text{ est } A_i^n \text{ donc } y = \sum_{j=1}^n Z_j^i x_j^k + Z_i \quad (V.47)$$

$$O^k = \frac{\sum_{i=1}^m \alpha_i^k (\sum_{j=1}^n Z_j^i x_j^k + Z_i^0)}{\sum_{i=1}^m \alpha_i^k} \quad (V.48)$$

$$O^k = \frac{\sum_{i=1}^m \left( \prod_{j=1}^n A_j^i(x_j^k) \right) (\sum_{j=1}^n Z_j^i x_j^k + Z_i^0)}{\sum_{i=1}^m \prod_{j=1}^n A_j^i(x_j^k)} \quad (V.49)$$

$$E^k = \frac{1}{2} [O^k - y^k]^2 \quad (V.50)$$

Avec respectivement  $O^k$  la sortie du système flou correspondant au vecteur d'entrée  $x^k$  et  $y^k$  la sortie désirée.

## V.7. Résultats de la modélisation par le modèle ANFIS

### V.7.1. Méthodologie

Dans cette partie, nous allons montrer les capacités de la méthode de modélisation par les systèmes neuroflous type ANFIS à partir de données entrée-sortie. Deux catégories de méthodes sont présentées ici. La première concerne la construction de la base de règle par la technique dite "GIRD PARTITION", cette méthode inclus huit type de fonctions d'appartenance (triangulaire, trapézoïdale, gaussienne, ..., etc.), pour la construction de la base de règle. En fait, nous allons montrer que cette catégorie présente un grand handicap dans la

construction des modèles neuroflous, dans le cas où le nombre d'entrée du système dépasse six (06) et lorsque le nombre des variables linguistiques pour chaque variable d'entrée dépasse lui aussi trois(03). Ceci va nous permettre de comparer les performances de cette méthode avec celles obtenues par la deuxième technique qui est nommé, le Groupage soustractif "SUBTRACTIVE CLUSTERING", et de montrer les particularités et les points forts et faibles de chaque méthode (Heddam et al., 2011b).

À cette fin, une base d'apprentissage est nécessaire (60 % de la base de données disponible). L'apprentissage consiste en la correction des paramètres (prémises et conséquents) du réseau afin de généraliser une fonction de transfert (inconnue *a priori*) entre les entrées et la sortie du réseau. Celle-ci est constituée d'un ensemble de couples "entrées /sortie "connus (relevé de données). Le déploiement des algorithmes d'apprentissage sur cette base de données permet de construire une fonction d'approximation de la sortie d (sorties désirées de la dose de coagulant) à partir des nouveaux vecteurs d'entrée (Heddam et al., 2011a).

Une des plus importantes étapes pour la génération de la structure des réseaux neuroflous ANFIS est l'établissement des règles d'inférence. En utilisant un certain mécanisme d'inférence, les règles sont définies comme combinaisons des fonctions d'appartenance pour chaque variable d'entrée. Les variables d'entrée sont divisées en un nombre limité de fonctions d'appartenance (et leurs combinaisons mènent à beaucoup de règles d'inférence différentes).

(Rui *et al.*, 1995), ont montré qu'il n'y a pas de structure du réseau idéal *a priori*. Celle-ci est à définir au cas par cas selon le problème. Concernant le réseau neuro-flou ANFIS, définir une architecture consiste à choisir le nombre de fonctions d'appartenance, leur type (gaussienne, sigmoïde...) et le nombre de règles floues utilisées. Aussi, un compromis simplicité/généralisation doit être trouvé : il n'est pas concevable de construire un système ANFIS trop complexe. Dans ce contexte deux stratégies sont largement utilisées pour la construction de la base de règle d'un modèle ANFIS. La première appelée "*GIRD PARTITION*" ; cette méthode a besoin seulement d'un nombre restreint de fonctions d'appartenance (MF) pour chaque entrée. Cependant, elle rencontre des problèmes quant nous avons un nombre assez grand d'entrées. Par exemple, un modèle ANFIS avec 6 entrées (cas de notre étude) et quatre MF's sur chaque entrée aurait en conséquence  $M = 4^6 = 4096$  règles de type Si-Alors. Ce problème, habituellement désigné sous le nom du problème de la dimensionnalité, peut être allégé par la deuxième stratégie de partage de l'espace des entrées. Cette dernière appelée algorithme flou d'auto organisation ou "*SUBTRACTIVE CLUSTERING*" ou bien "fuzzy selforganization algorithme" (Chiu, 1994).

Afin de tester l'influence des deux stratégies sur les performances du modèle ANFIS, appliqué pour la modélisation de la dose de coagulant, nous avons réalisé une étude comparative entre ces deux approches.

## V.7.2. Application de la méthode *GIRD PATITION*

### V.7.2.1. Formulation du système flou

Comme nous l'avons déjà évoqué, l'application d'un système d'inférence floue (ou neuroflou) à la modélisation et l'estimation de la dose de coagulant suit une démarche assez intuitive, la solution la plus simple consiste à utiliser les valeurs retournées par les différentes observations relatives aux variables descriptives de l'eau brute comme variables d'entrée. Il est possible de définir pour ces variables des sous-ensembles flous correspondant à des distances, les prémises des règles utilisées ont dans ce cas la forme suivante :

**Si** antécédent **Alors** conséquent

**Si** la turbidité est élevée et le pH est faible **Alors** la dose de coagulant est élevée

**Si** la turbidité est élevée et le pH est faible **Alors** dose = 20 \*A +30 \*B+ C

La première règle est celle d'un modèle de type Mamdani utilisant les variables linguistiques associées aux concentrations de la turbidité, et le pH tandis que la seconde est celle d'un modèle de type Takagi\_Sugeno où la variable *dose* du conséquent (dose de coagulant) est numérique sous la forme d'une fonction des variables associées à l'antécédent. Ici *A* et *B* représentent respectivement les valeurs numériques des concentrations de la turbidité et du pH. En général 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 ci-dessus.

Nous avons testés deux types de répartition : une première basée sur la répartition des variables d'entrée en deux sous ensembles flous (deux variables linguistiques pour chaque entrée), et une deuxième méthode basée sur la répartition des variables d'entrée en trois sous ensembles flous (trois variables linguistiques pour chaque entrée).

### V.7.2.2. Représentation des règles floues

Les règles floues établissent des relations logiques entre les variables du système en associant valeurs *qualitatives* d'une variable (la turbidité est élevée) avec celles d'une autre variable (la dose de coagulant est élevée). Les valeurs qualitatives ont typiquement une interprétation linguistique, elles sont nommées *termes linguistiques* (étiquettes). La signification des termes linguistiques par rapport aux variables d'entrée/sortie numériques est définie par des ensembles flous appropriés. Dans ce sens, les ensembles flous, ou plus précisément leurs fonctions d'appartenance, fournissent une interface entre les variables numériques d'entrée/sortie et les valeurs linguistiques qualitatives dans les règles. Dans le cas des modèles neuroflous de types de *flou Takagi\_Sugeno* (TSK), qui est utilisé dans notre

étude, le nombre total de règles (Tableau V.1) à optimiser sera déterminé par la règle suivante (Heddam et al., 2011b):

$$NRF = NSF^{NVE} \quad (V.51)$$

*NRF* : Nombre de règle floue ;

*NSF* : Nombre de sous ensemble floues pour chaque variable d'entrée ;

*NVE* : Nombre de variables d'entrées.

Tableau V.1. Nombre de règles floues pour différents modèles utilisés

Tableau V.1. Nombre de règles floues pour différents modèles utilisés

Modèle	Nombre d'entrées	Nombre de sous ensemble flous	Nombre total de règles
M02-ANFIS	02	02	4
		03	9
M03-ANFIS	03	02	8
		03	27
M04-ANFIS	04	02	16
		03	81
M05-ANFIS	05	02	32
		03	243
M06-ANFIS	06	02	64
		03	729

### V.7.2.3. Fonctions d'appartenance du système

Les variables d'entrée peuvent être fuzzifiées de différentes façons, cela implique l'utilisation des fonctions d'appartenances de différents types. Dans le cadre de notre étude nous avons testé quatre types de fonctions d'appartenances : triangulaires (Trimf), Gaussienne (Gaussmf), trapézoïdales (Trapmf) et la fonction dite Generalized bell-shaped (Gbellmf). Les paramètres à optimiser pour chaque type de fonction sont présentés dans le Tableau (V.2).

Fonction d'appartenance	Expression	Paramètres à optimiser
Gbellmf	$f(x,a,b,c) = \frac{1}{1 + \left  \frac{x-c}{a} \right ^{2s}}$	a, b et c
Gaussmf	$f(x,\sigma,c) = e^{-\frac{(x-c)^2}{2\sigma^2}}$	$\sigma$ et c
Trimf	$f(x,a,b,c) = \begin{cases} 0, & \dots x \leq a \\ \frac{x-a}{b-a}, & \dots a \leq x \leq b \\ \frac{c-x}{c-b}, & \dots b \leq x \leq c \\ 0, & \dots c \leq x \end{cases}$	a, b et c
Trapmf	$f(x,a,b,c,d) = \begin{cases} 0, & \dots x \leq a \\ \frac{x-a}{b-a}, & \dots a \leq x \leq b \\ 1, & \dots b \leq x \leq c \\ \frac{d-x}{d-c}, & \dots c \leq x \leq d \\ 0, & \dots d \leq x \end{cases}$	a, b, c et d

Tableau V.2. Paramètres à optimiser pour chaque type de fonction d'appartenance

### V.7.2.4. Partitionnement avec des fonctions triangulaires

Une fonction d'appartenance triangulaire dans un univers de discours [a, b] peut être définie par, trois paramètres : a, b et c (Tableau V.2) qui prennent leurs valeurs dans l'intervalle [a, b]. Nous avons procédé à une comparaison entre les résultats obtenus par les cinq variantes (à savoir le modèle à deux, trois, quatre, cinq entrées ainsi que la variante à

six entrées qui incluent les six variables descriptives caractérisant l'eau brute). Les tests ont été réalisés avec, premièrement deux sous ensembles flous pour chaque variable d'entrée, puis avec trois sous ensembles flous. Le Tableau (V.3) représente l'ensemble des paramètres des fonctions d'appartenance optimisées pendant l'apprentissage des modèles avec deux ensembles flous.

**Tableau V.3. Paramètres optimisés pour la fonction d'appartenance triangulaire avec deux sous ensembles flous**

Modèle	Variables d'entrée	Fonction d'appartenance	Sous ensemble flou	Paramètres		
				a	b	c
ANFIS	input1	Triangulaire	in1mf1	-6.390	-1.812	2.766
		Triangulaire	in1mf2	-1.812	2.766	7.344
	input2	Triangulaire	in2mf1	-7.485	-2.103	3.278
		Triangulaire	in2mf2	-2.103	3.278	8.660
	Input3	Triangulaire	in3mf1	-9.042	-2.794	3.454
		Triangulaire	in3mf2	-2.794	3.454	9.702
	Input4	Triangulaire	in4mf1	-8.601	-1.569	5.462
		Triangulaire	in4mf2	-1.569	5.462	12.49
	Input5	Triangulaire	in5mf1	-5.911	-1.536	2.838
		Triangulaire	in5mf2	-1.536	2.838	7.213
	Input6	Triangulaire	in6mf1	-20.102	-1.717	16.667
		Triangulaire	in6mf2	-1.717	16.667	35.05

Les résultats obtenus par les modèles neuroflous avec deux sous ensembles flous sont présentés dans le Tableau (V.4). On constate que le modèle donne de médiocres résultats, que se soit pour la phase d'apprentissage ou pour les phases de validation et de test, pour les deux variantes M02 et M03. Le coefficient de corrélation ne dépasse pas les 0.62, tandis que le RMSE reste toujours élevé et avoisine les 5.86. On constate aussi une nette amélioration des résultats pour la phase d'apprentissage à partir de la variante M04, et les meilleurs résultats sont obtenus pour la variante M06, incluant les six variables, à savoir la température, le pH, la conductivité, la turbidité, les UV<sub>254</sub> et l'oxygène dissous, cependant pour les deux phases de validation et de test les performances restent toujours médiocres et aucune amélioration n'est enregistrée, avec une RMSE de 6.21, une erreur moyenne absolue de 3.88 et un coefficient de corrélation de 0.68, pour la phase de test. La Figure (V.18a-d) représente les courbes Scatterplots, pour le modèle à six entrées (M06-ANFIS), pour les phases d'apprentissage, validation, test ainsi que pour l'ensemble des données, Les Figures (V.19) et (V.20), présentent respectivement la forme initiale et finale des fonctions d'appartenance triangulaire utilisée, la comparaison entre les valeurs observées et calculées est présentée dans la figure (V.21a-c).

**Tableau V.4. Performances des modèles ANFIS avec deux sous ensembles flous, pour des fonctions d'appartenances triangulaires**



Modèle	Apprentissage			Validation			Test		
	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE	CC
M02-ANFIS	6.30	5.00	0.52	6.04	4.67	0.64	5.85	4.53	0.58
M03-ANFIS	5.86	4.62	0.60	5.86	4.43	0.66	5.62	4.22	0.62
M04-ANFIS	4.27	3.30	0.81	5.49	4.25	0.72	4.67	3.65	0.76
M05-ANFIS	3.32	2.41	0.89	5.80	4.00	0.71	4.61	3.30	0.78
M06-ANFIS	2.49	1.77	0.94	6.70	4.80	0.69	6.21	3.88	0.68

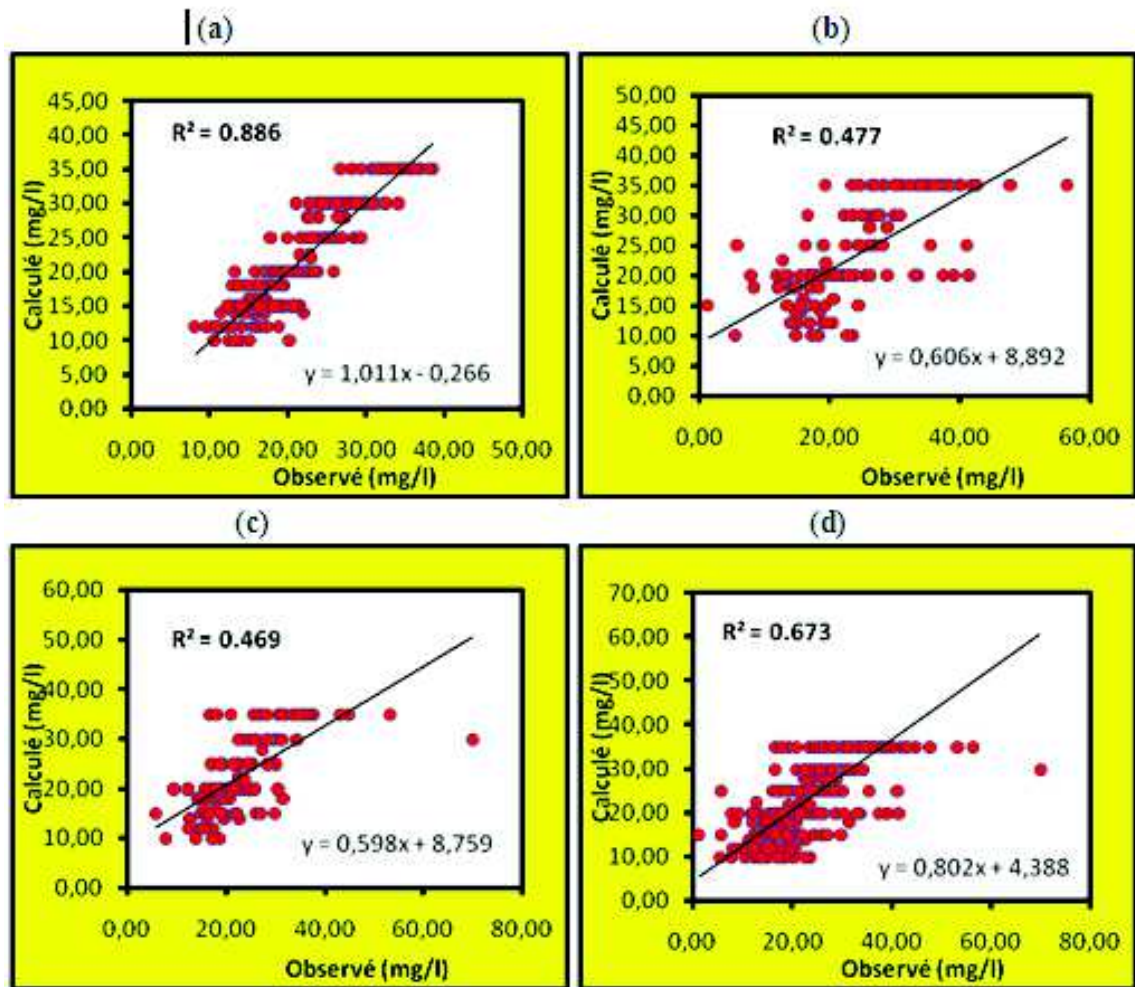


Figure V.18 Comparaison des résultats du modèle à base de réseau neuroflou ANFIS, avec deux sous ensembles flous, pour des fonctions d'appartenances triangulaires : (a) apprentissage, (b) validation, (c) test et (d) l'ensemble des données.



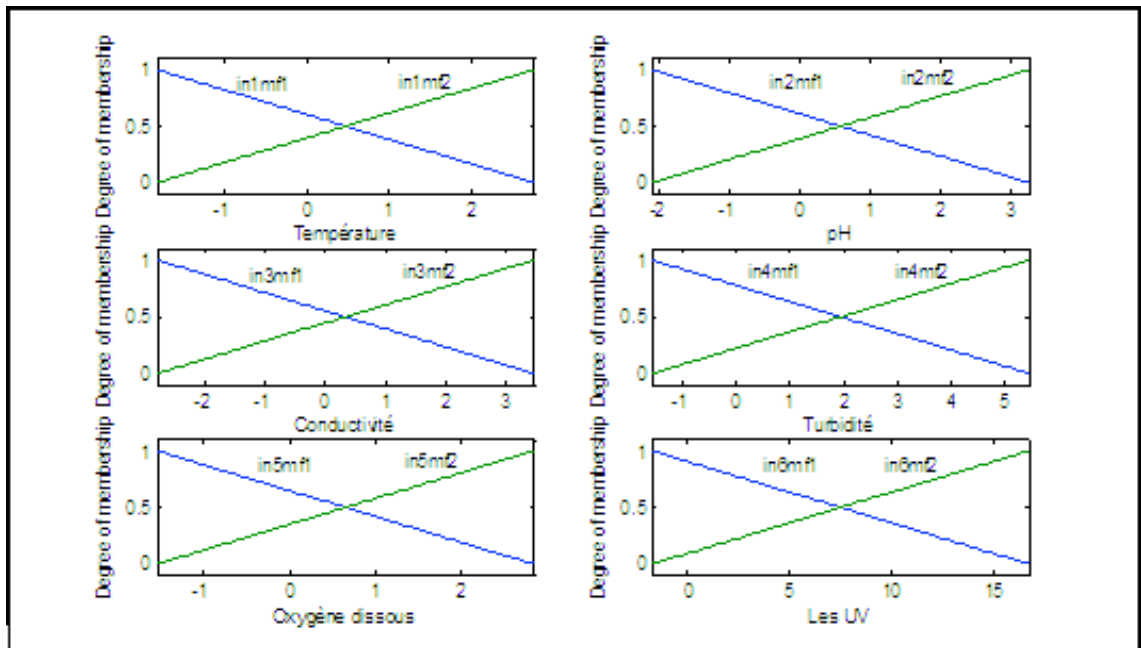


Figure V.19 Forme initiale des fonctions d'appartenances triangulaires utilisées pour le modèle ANFIS avec deux sous ensembles flous.

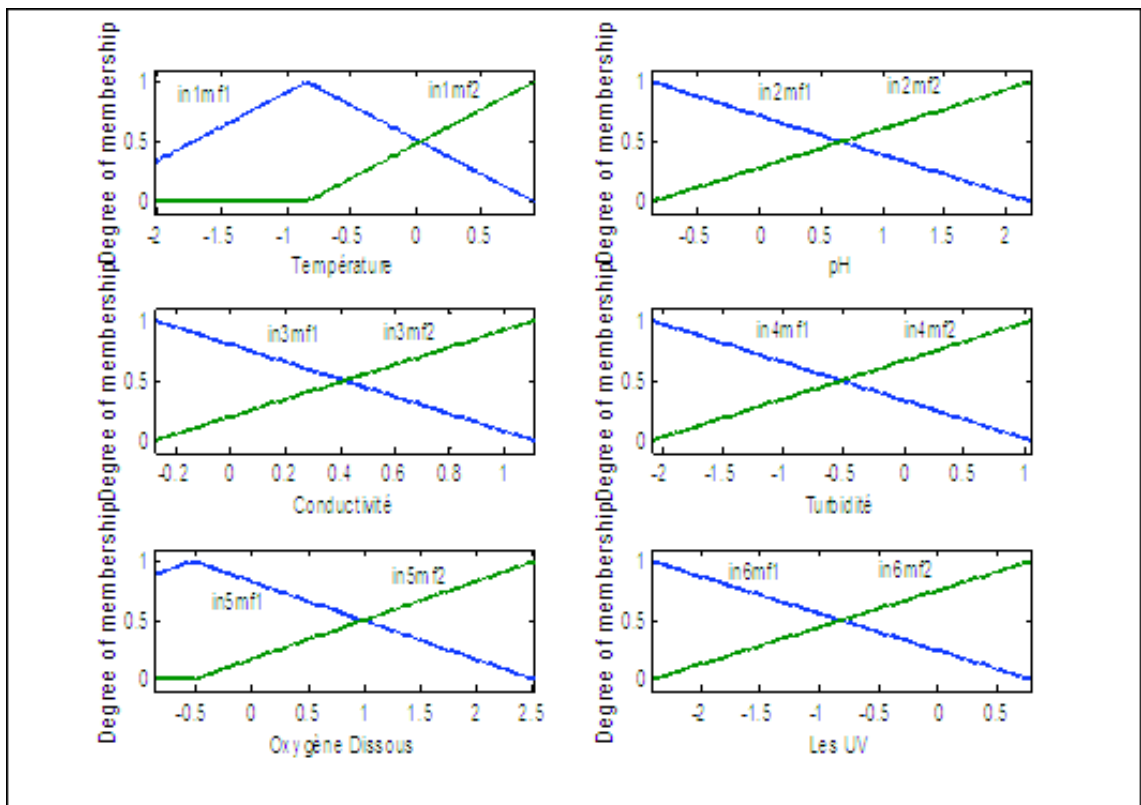


Figure V.20 Forme finale des fonctions d'appartenances triangulaires utilisées pour le modèle ANFIS avec deux sous ensembles flous.

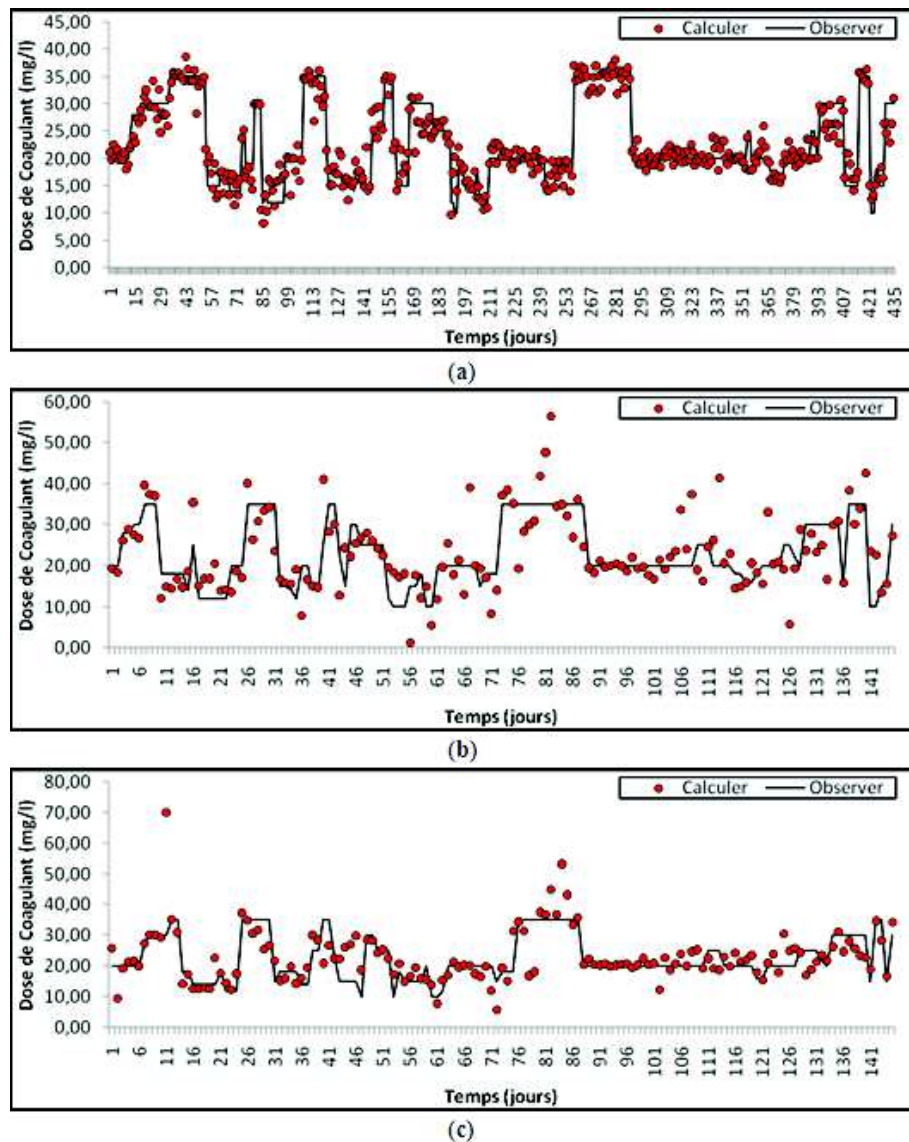


Figure V.21 Modélisation de la dose de coagulant par le modèle à base de réseau neuroflou ANFIS, avec deux sous ensembles flous, pour des fonctions d'appartenances triangulaires: (a) partie apprentissage, (b) partie validation et (c) partie test.

Pour le modèle utilisant trois sous ensembles flous pour chaque variable avec des fonctions d'appartenance triangulaire, le Tableau (V.5) représente l'ensemble des paramètres des fonctions d'appartenance optimisées pendant l'apprentissage des modèles. Le Tableau (V.6) présente une comparaison entre les performances obtenues par les cinq variantes testées. On constate que le modèle donne de très bons résultats en phase d'apprentissage, cependant en phases de validation et de test, les performances sont loin d'être acceptées. La Figure (V.22a-d) représente les courbes Scatterplots, pour le modèle retenu (M06-ANFIS), les deux Figures (V.23 et V.24) présentent respectivement la forme initiale et finale des fonctions d'appartenance triangulaire utilisées, la comparaison entre les valeurs observées et calculées est présentée dans la figure (V.25a-c).

On remarque que les performances du modèle ANFIS s'améliorent en augmentant le nombre des sous ensembles flous de deux à trois en phase d'apprentissage, cependant en phases de validation et de test les résultats restent toujours inacceptables, ce qui nous a

conduits à tester d'autres fonctions d'appartenance, dans les soucis d'améliorer les qualités prédictive du modèle. Cependant il est important de rappeler à ce stade que le nombre total de règles à optimiser pour le modèle à base de deux sous ensembles flous est de l'ordre de 64 règles ( $2^6$ ), cependant pour le modèle à base de trois sous ensembles flous il passe à 729 règles ( $3^6$ ), ce qui rend la tâche d'apprentissage colossale, et l'optimisation du réseau avec un nombre important d'itérations nécessite un temps relativement long.

**Tableau V.5. Paramètres optimisés pour la fonction d'appartenance triangulaire avec trois sous ensembles flous**

Modèle	Variable d'entrée	Fonction d'appartenance	Sous ensemble floue	Paramètres		
				a	b	c
M06-ANFIS	Input1	Triangulaire	in1mf1	-3.930	-1.698	0.534
		Triangulaire	in1mf2	-1.698	0.534	2.766
		Triangulaire	in1mf3	0.534	2.766	4.998
	Input2	Triangulaire	in2mf1	-4.559	-1.946	0.666
		Triangulaire	in2mf2	-1.946	0.666	3.279
		Triangulaire	in2mf3	0.666	3.279	5.891
	Input3	Triangulaire	in3mf1	-5.919	-2.794	0.329
		Triangulaire	in3mf2	-2.794	0.329	3.454
		Triangulaire	in3mf3	0.329	3.454	6.578
	Input4	Triangulaire	in4mf1	-4.492	-1.571	1.325
		Triangulaire	in4mf2	-1.570	1.352	4.274
		Triangulaire	in4mf3	1.352	4.274	7.196
	Input5	Triangulaire	in5mf1	-3.490	-1.537	0.416
		Triangulaire	in5mf2	-1.537	0.416	2.369
		Triangulaire	in5mf3	0.416	2.369	4.322
	Input6	Triangulaire	in6mf1	-11.18	-1.970	7.384
		Triangulaire	in6mf2	-1.970	7.384	16.67
		Triangulaire	in6mf3	7.384	16.67	25.95

**Tableau V.6. Performances des modèles ANFIS avec trois sous ensembles flous, pour des fonctions d'appartenances triangulaires**

Modèle	Apprentissage			Validation			Test		
	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE	CC
M02-ANFIS	6.01	4.67	0.60	6.25	5.00	0.54	5.99	4.80	0.56
M03-ANFIS	4.88	3.43	0.76	11.73	5.24	0.31	6.76	4.45	0.50
M04-ANFIS	2.86	1.90	0.92	10.66	5.37	0.53	49.98	9.81	0.068
M05-ANFIS	1.01	0.52	0.99	8.44	5.33	0.59	11.59	5.51	0.49
M06-ANFIS	0.35	0.17	0.99	9.30	5.87	0.57	12.75	6.99	0.36

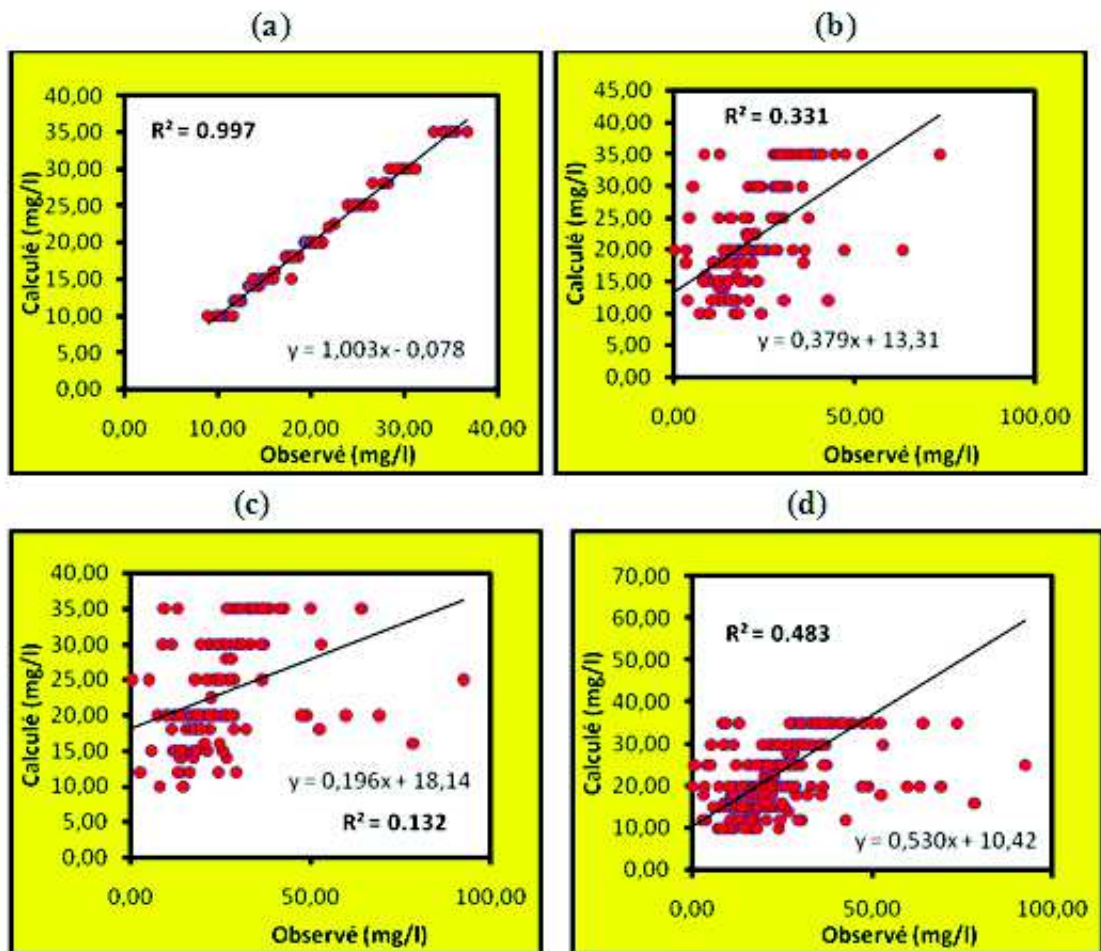


Figure V.22 Comparaison des résultats du modèle à base de réseau neuroflou ANFIS, avec trois sous ensembles flous, pour des fonctions d'appartenance triangulaires: (a) apprentissage, (b) validation, (c) test et (d) l'ensemble des données.

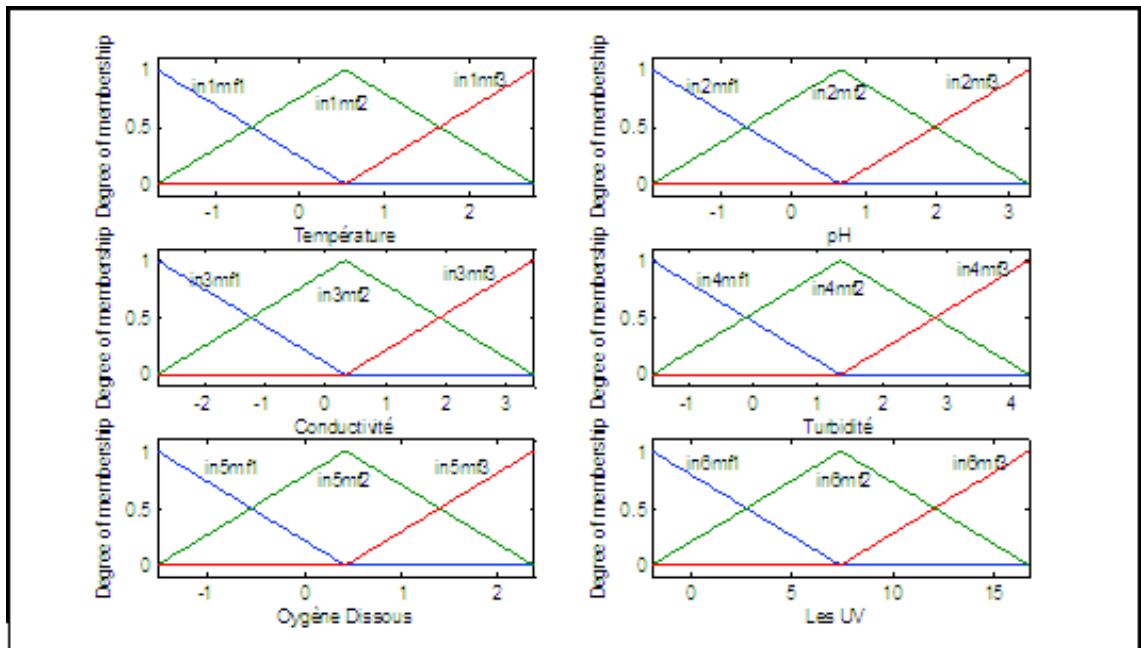


Figure V.23 Forme initiale des fonctions d'appartenances triangulaires utilisées pour le modèle ANFIS avec trois sous ensembles flous

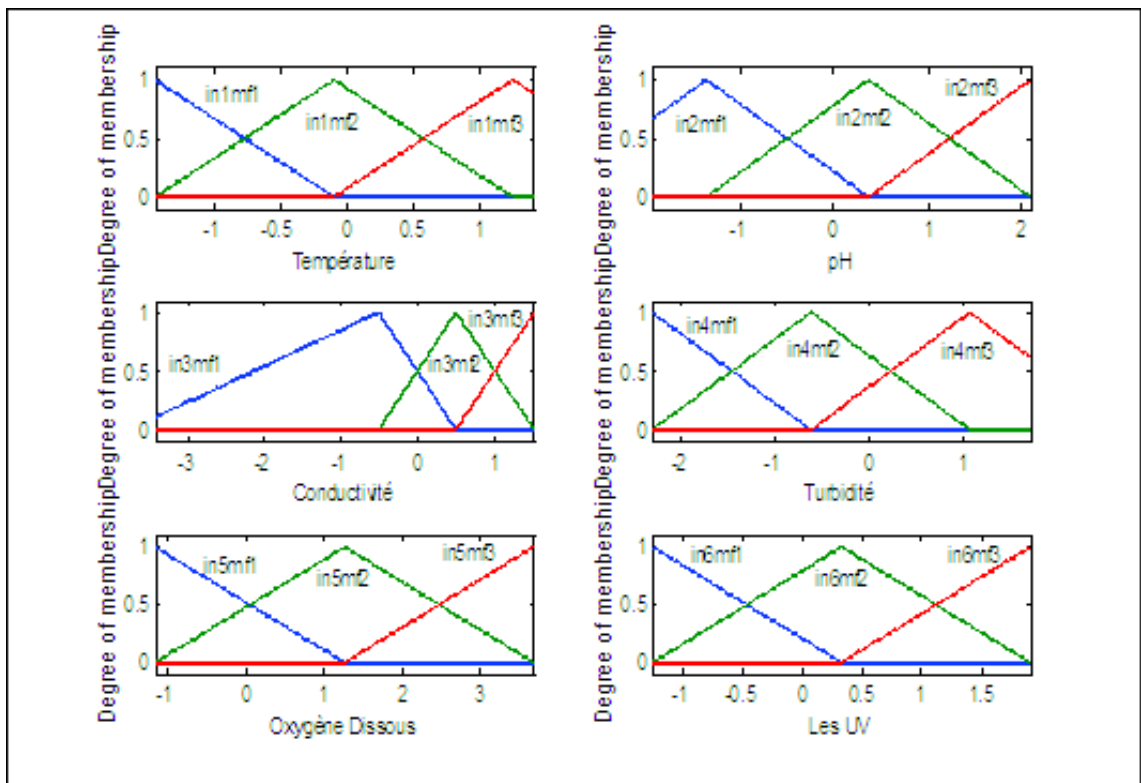
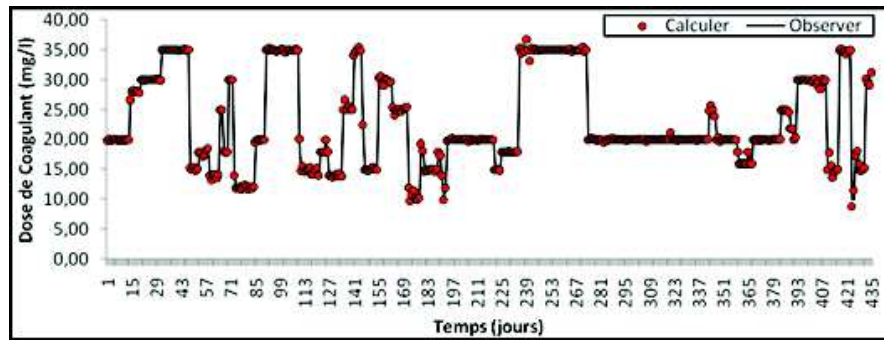
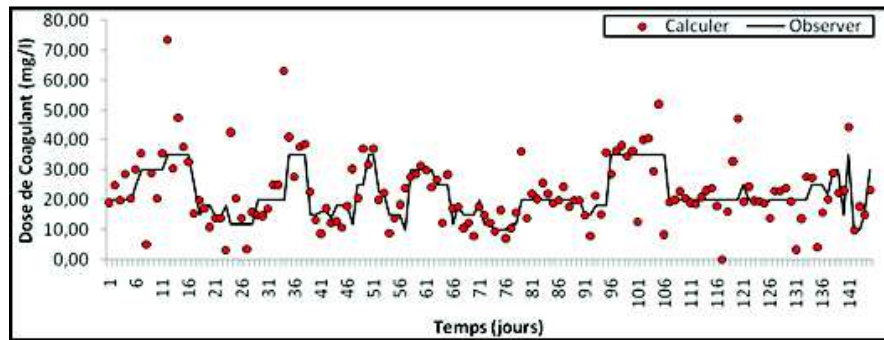


Figure V.24 Forme finale des fonctions d'appartenances triangulaires utilisées pour le modèle ANFIS avec trois sous ensembles flous





(a)



(b)

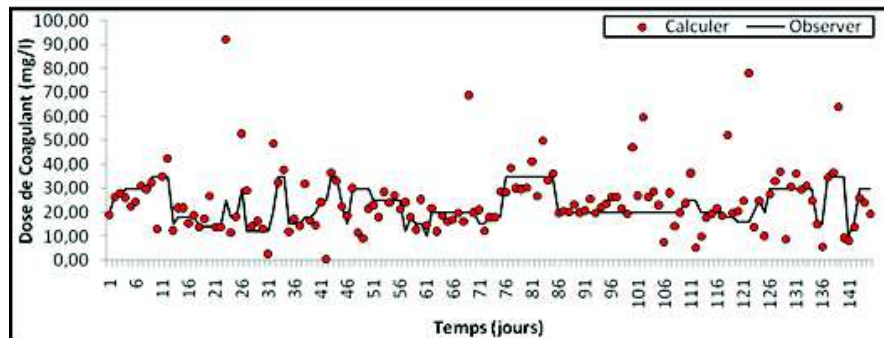


Figure V.25 Modélisation de la dose de coagulant par le modèle à base de réseau neuroflou ANFIS, avec trois sous ensembles flous, pour des fonctions d'appartenance triangulaires: (a) partie apprentissage, (b) partie validation et (c) partie test.

### V.7.2.5. Partitionnement avec des fonctions Gaussienne

Une fonction d'appartenance Gaussienne dans un univers du discours  $[\sigma, c]$  peut être définie par, deux paramètres :  $\sigma$  et  $c$  (Tableau V.2) qui prennent leurs valeurs dans l'intervalle  $[a, b]$ . Nous avons procédé à une comparaison entre les résultats obtenus par les cinq variantes (à savoir le modèle à deux, trois, quatre, et cinq entrées ainsi que la variante à six entrées qui incluent les six variables descriptives caractérisant l'eau brute). Nous avons testés les cinq variantes avec deux sous ensembles flou pour chaque variable d'entrée, puis avec trois sous ensembles flous. Le Tableau(V.7) représente l'ensemble des paramètres des fonctions d'appartenance optimisées pendant l'apprentissage des modèles avec deux ensembles flous.

**Tableau V.7. Paramètres optimisés pour la fonction d'appartenance Gaussienne avec deux sous ensembles flous**

Modèle	Variable d'entrée	Fonction d'appartenance	Sous ensemble flou	Paramètres	
				$\sigma$	c
M06-ANFIS	input1	Gaussienne	in1mf1	1.944	-1.812
		Gaussienne	in1mf2	1.944	2.766
	input2	Gaussienne	in2mf1	2.286	-2.103
		Gaussienne	in2mf2	2.286	3.279
	Input3	Gaussienne	in3mf1	2.653	-2.794
		Gaussienne	in3mf2	2.653	3.454
	Input4	Gaussienne	in4mf1	2.986	-1.570
		Gaussienne	in4mf2	2.986	5.462
	Input5	Gaussienne	in5mf1	1.858	-1.537
		Gaussienne	in5mf2	1.858	2.838
	Input6	Gaussienne	in6mf1	7.807	-1.717
		Gaussienne	in6mf2	7.807	16.67

Les résultats obtenus par les modèles neuroflous avec deux sous ensembles flous sont présentés dans le Tableau (V.8), on constate que le modèle n'a pas donné de résultats exemptés, que se soit pour la phase d'apprentissage ou pour les phases de validation et de test pour les deux variantes M02 et M03. Le coefficient de corrélation ne dépasse pas les 0.64, tandis que le RMSE reste toujours élevé et avoisine les 5.64. On constate aussi une nette amélioration des résultats pour la phase d'apprentissage à partir de la variante M04, et les meilleurs résultats sont obtenus pour la variante M06, incluant les six variables, à savoir la température, le pH, la conductivité, la turbidité, les UV<sub>254</sub> et l'oxygène dissous, cependant pour les deux phases de validation et de test les performances restent toujours médiocre et aucune amélioration n'est enregistrée, avec une RMSE de 6.18, une erreur moyenne absolue de 3.88 et un coefficient de corrélation de 0.67, pour la phase de test. La Figure (V.26 a-d) représente les courbes Scatterplots, pour le modèle à six entrées (M06-ANFIS), pour les phases d'apprentissage, validation, test ainsi que pour l'ensemble de données, Les deux Figures (V.27 et V.28), présentent respectivement la forme initiale et finale des fonctions d'appartenance gaussienne utilisée. La comparaison entre les valeurs observée et calculée est présentée dans la figure (V.29 a-c).

**Tableau V.8. Performances des modèles ANFIS avec des fonctions d'appartenances gaussiennes pour deux sous ensembles flous**

Modèle	Apprentissage			Validation			Test		
	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE	CC
M02-ANFIS	6.27	4.99	0.52	6.00	4.60	0.65	5.78	4.45	0.60
M03-ANFIS	5.64	4.41	0.64	5.73	4.42	0.68	5.67	4.31	0.62
M04-ANFIS	3.92	3.00	0.84	5.41	4.10	0.73	4.76	3.61	0.75
M05-ANFIS	2.66	1.87	0.93	5.18	3.68	0.77	4.71	3.39	0.77
M06-ANFIS	2.14	1.48	0.95	5.72	4.18	0.75	6.18	3.88	0.67



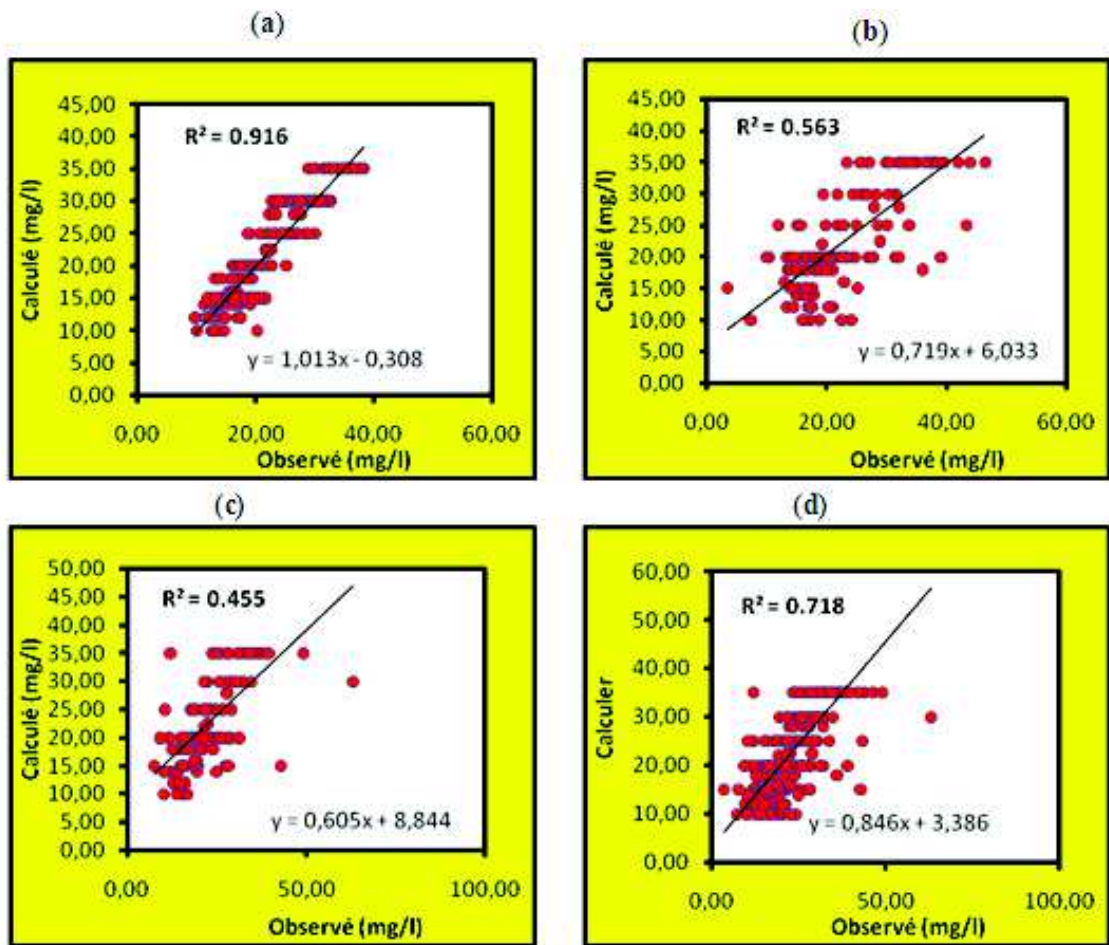


Figure V.26 Comparaison des résultats du modèle à base de réseau neuroflou ANFIS, avec deux sous ensembles flous, pour des fonctions d'appartenances Gaussienne: (a) apprentissage, (b) validation, (c) test et (d) l'ensemble des données.

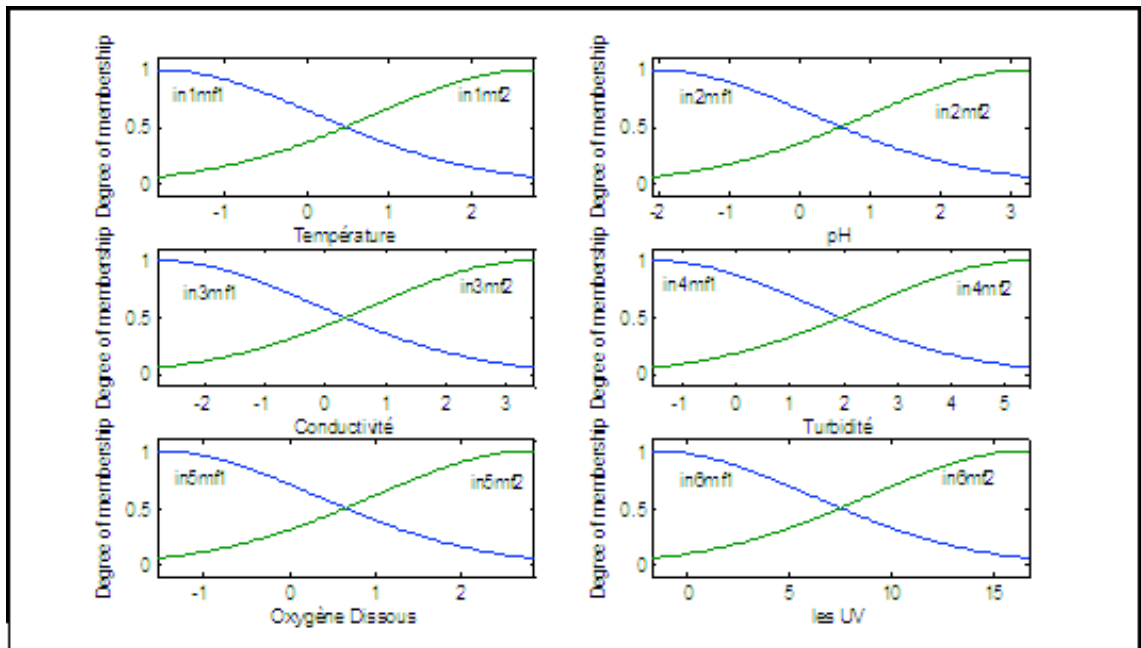


Figure V.27 Forme initiale des fonctions d'appartenances Gaussiennes utilisées pour le modèle ANFIS avec deux sous ensembles flous.

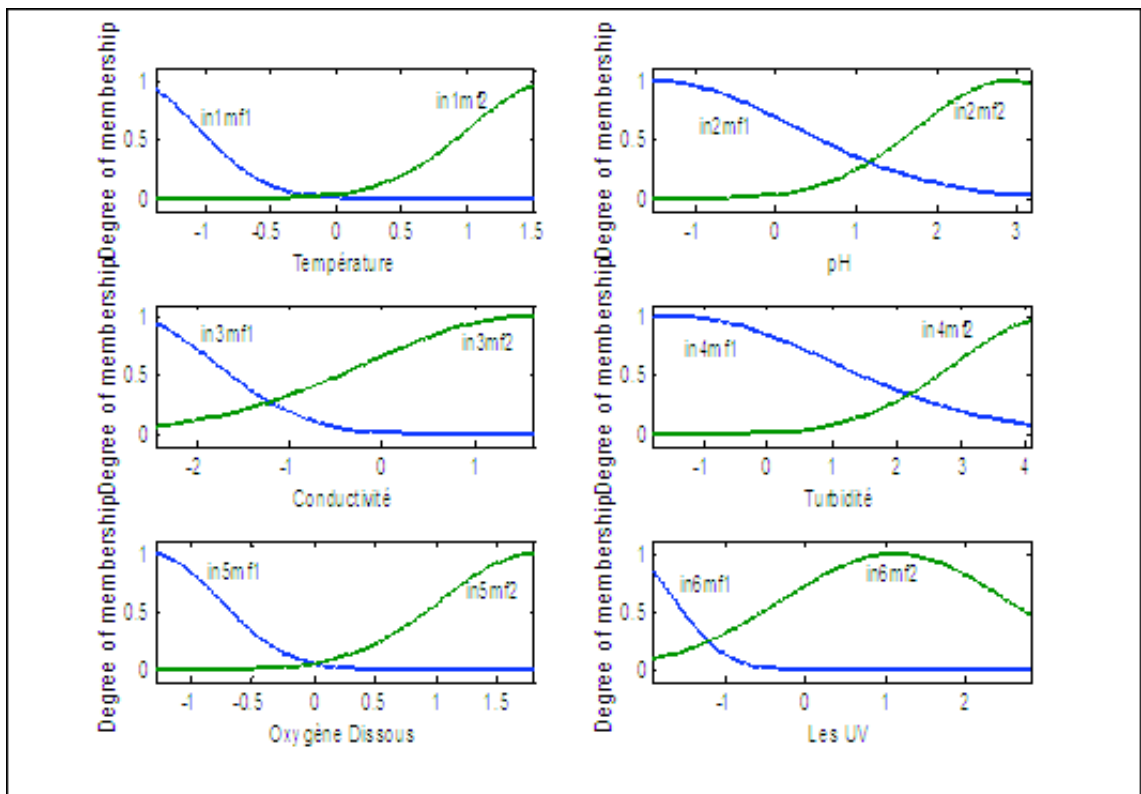
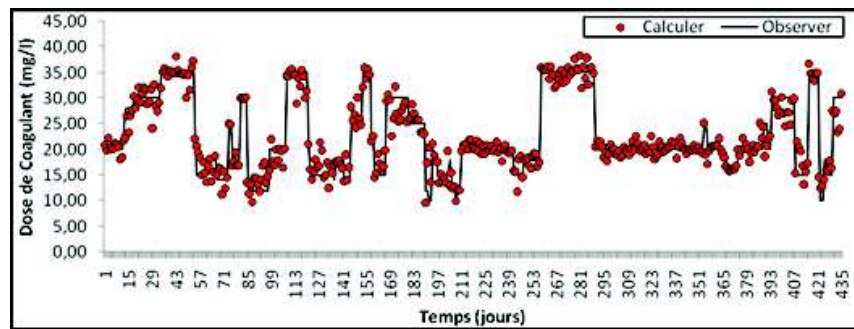
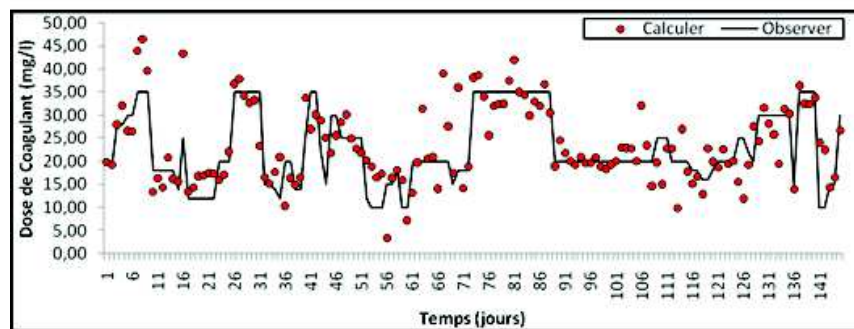


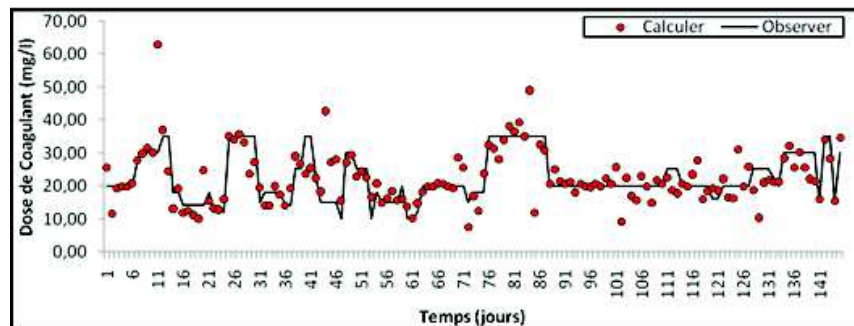
Figure V.28 Forme finale des fonctions d'appartenances Gaussiennes utilisées pour le modèle ANFIS avec deux sous ensembles flous.



(a)



(b)



(c)

Figure V.29 Modélisation de la dose de coagulant par le modèle à base de réseau neuroflou ANFIS, avec deux sous ensembles flous, pour des fonctions d'appartenance Gaussiennes : (a) partie apprentissage, (b) partie validation et (c) partie test.

Pour le modèle utilisant trois sous ensembles flous pour chaque variable avec des fonctions d'appartenance Gaussiennes, le Tableau V.9 représente l'ensemble des paramètres des fonctions d'appartenance, optimisé pendant l'apprentissage des modèles. Le Tableau (V.10) présente une comparaison entre les performances obtenues par les cinq variantes testées. On constate que le modèle donne de très bons résultats en phase d'apprentissage, cependant en phases de validation et de test, les performances sont loin d'être acceptées. La Figure (V.30 a-d) représente les courbes Scatterplots, pour le modèle retenu (M06-ANFIS), Les deux Figures (V.31 et V.32), présentent respectivement la forme initiale et finale des fonctions d'appartenance triangulaire utilisée. La comparaison entre les valeurs observées et calculées est présentée dans la figure (V.33a-c).

L'augmentation du nombre de sous ensembles flous n'a pas amélioré les résultats, cela est dû surtout au nombre important de règles floues mise en jeu (pour 4 sous ensembles

floes à chaque variable et pour six variables d'entrée on obtient  $4^6 = 4096$  règles) à optimiser, auxquelles il faut ajouter le nombre de paramètres pour chaque règle. Donc on peut conclure qu'il est impossible d'élaborer un tel modèle avec un nombre important de règles floues. Cette situation nous a conduits, à tester la deuxième stratégie d'optimisation des nombres de règles floues par le modèle neurone flou ANFIS, il s'agit de la technique à base des méthodes de classification dite de «*Groupage Soustractive*».

Tableau V.9. Paramètres optimisés pour la fonction d'appartenance Gaussienne avec trois sous ensembles floues

Modèle	Variable d'entrée	Fonction d'appartenance	Sous ensemble floue	Paramètres initiaux	
				$\sigma$	c
M06-ANFIS	Input1	Gaussienne	in1mf1	0.953	-1.812
		Gaussienne	in1mf2	0.953	0.434
		Gaussienne	in1mf3	0.953	0.680
	Input2	Gaussienne	in2mf1	1.143	-2.103
		Gaussienne	in2mf2	1.143	0.587
		Gaussienne	in2mf3	1.143	3.279
	Input3	Gaussienne	in3mf1	1.273	-2.729
		Gaussienne	in3mf2	1.273	0.268
		Gaussienne	in3mf3	1.273	3.266
	Input4	Gaussienne	in4mf1	1.489	-1.552
		Gaussienne	in4mf2	1.489	1.955
		Gaussienne	in4mf3	1.489	5.462
	Input5	Gaussienne	in5mf1	0.811	-1.451
		Gaussienne	in5mf2	0.811	0.459
		Gaussienne	in5mf3	0.811	2.369
	Input6	Gaussienne	in6mf1	2.963	-1.900
		Gaussienne	in6mf2	2.963	5.079
		Gaussienne	in6mf3	2.963	12.06

Tableau V.10. Performances des modèles ANFIS pour trois sous ensembles floues, avec des fonctions d'appartenance Gaussiennes

Modèle	Apprentissage			Validation			Test		
	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE	CC
M02-ANFIS	5.62	4.37	0.63	7.72	5.44	0.45	6.33	5.22	0.55
M03-ANFIS	4.25	2.96	0.81	7.67	4.88	0.54	6.12	4.22	0.62
M04-ANFIS	2.46	1.51	0.94	5.35	3.64	0.77	8.15	4.47	0.61
M05-ANFIS	0.72	0.38	0.99	36.42	10.74	0.28	20.47	9.34	0.36
M06-ANFIS	0.34	0.15	0.99	12.11	6.49	0.48	15.40	7.73	0.53

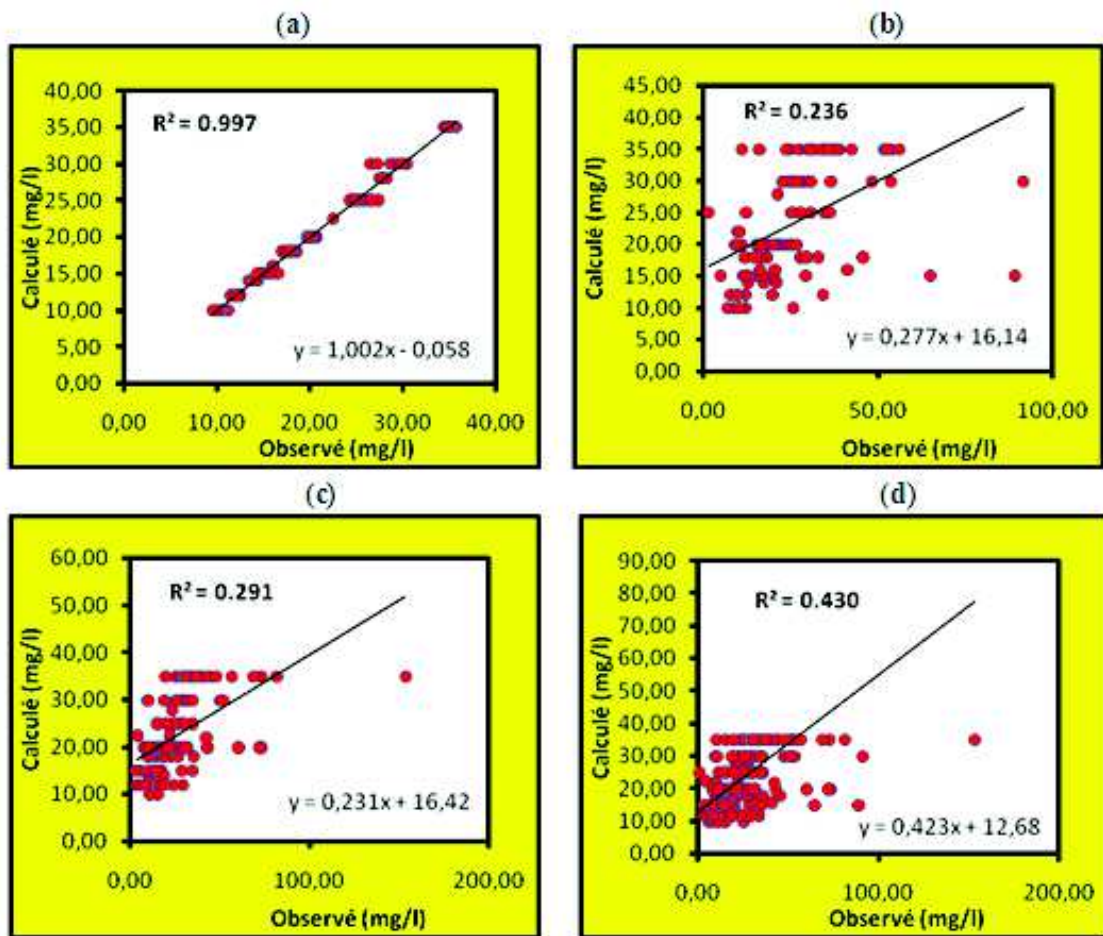


Figure V.30 Comparaison des résultats du modèle à base de réseau neuroflou ANFIS, avec trois sous ensembles flous, pour des fonctions d'appartenances Gaussienne: (a) apprentissage, (b) validation, (c) test et (d) l'ensemble des données.



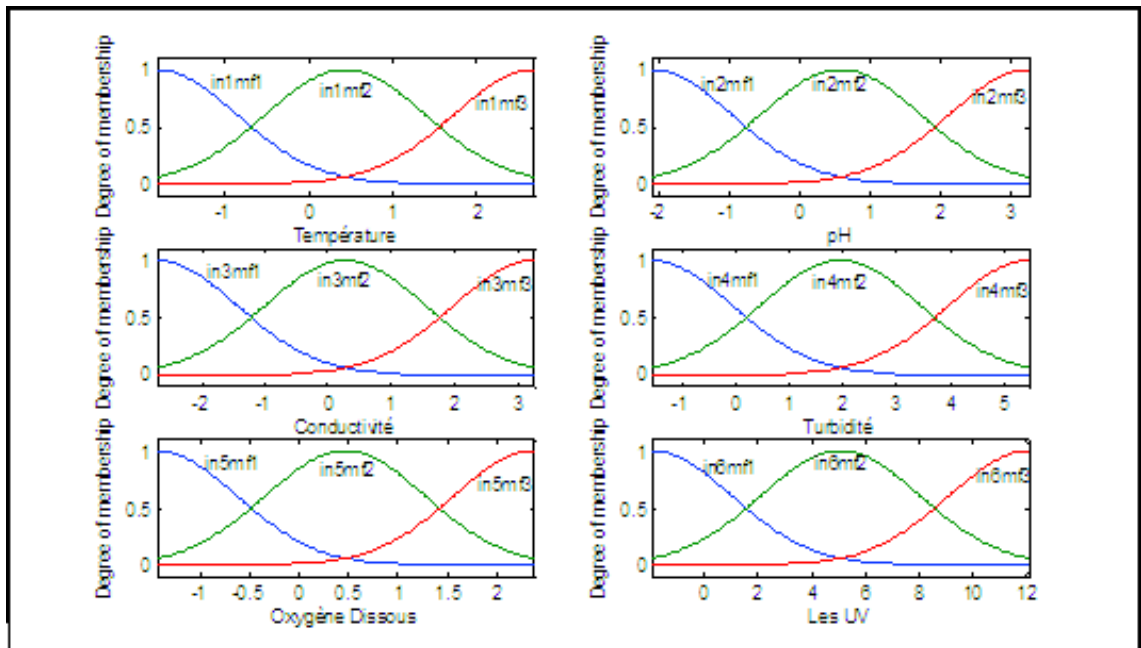


Figure V.31 Forme initiale des fonctions d'appartenances Gaussiennes utilisées pour le modèle ANFIS avec trois sous ensembles flous

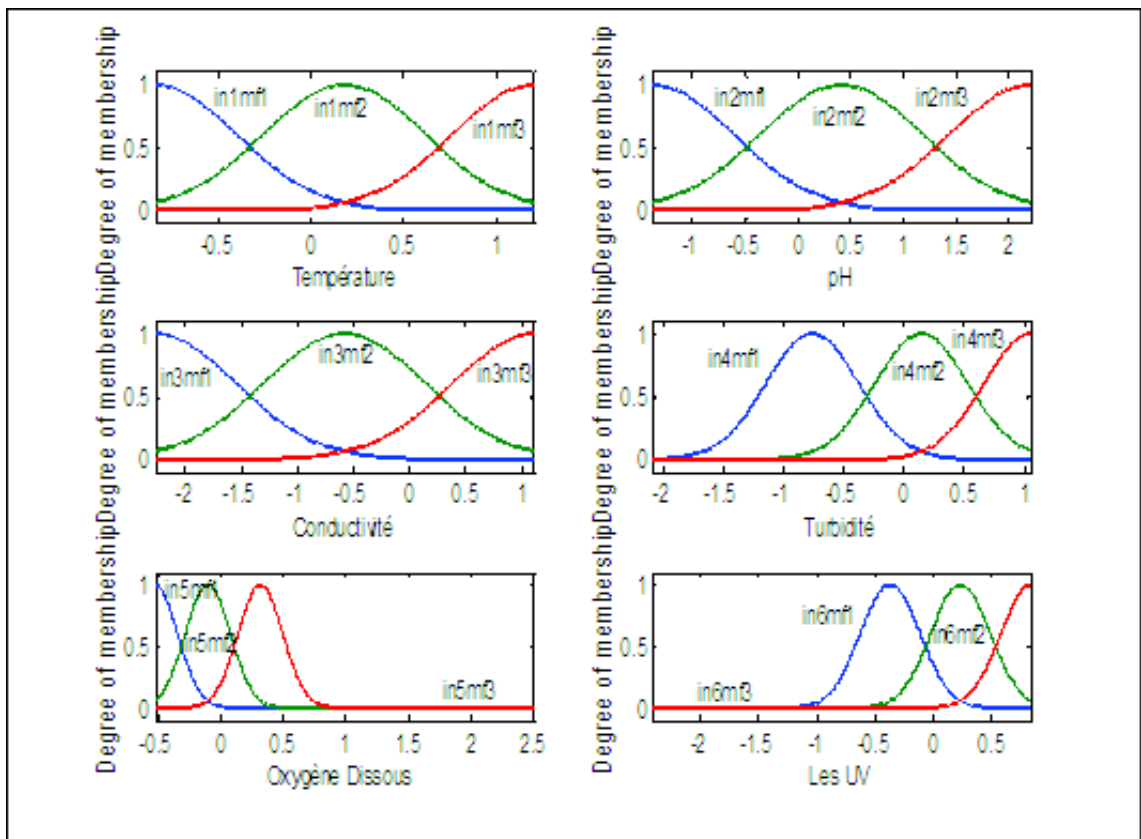


Figure V.32 Forme finale des fonctions d'appartenances Gaussiennes utilisées pour le modèle ANFIS avec trois sous ensembles flous



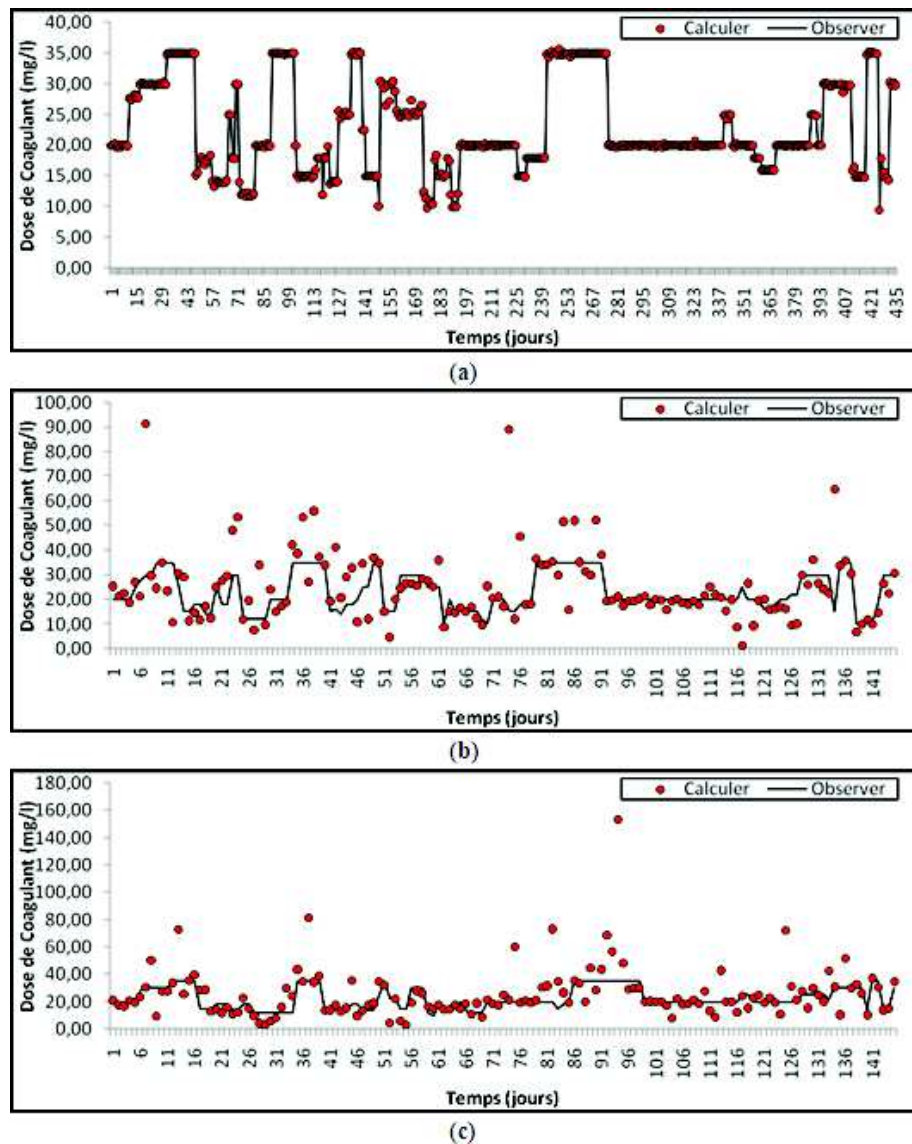


Figure V.33 Modélisation de la dose de coagulant par le modèle à base de réseau neuroflou ANFIS avec trois sous ensembles flous, pour des fonctions d'appartenances Gaussiennes : (a) partie apprentissage, (b) partie validation et (c) partie test

### V.7.3. Application de la méthode du Groupage Soustractive

(Chiu, 1994) a proposé un algorithme soustractif inspiré de la méthode de YAGER (Yager et Filev, 1994). Chaque exemple est considéré comme un centre de groupe éventuel et est affecté d'un potentiel d'attraction qui est lui même fonction d'un voisinage, le voisinage est paramétré par un rayon,  $r_a$ . Pour l'exemple  $i$ , le potentiel d'attraction s'exprime par :

$$P_i = \sum_{j=1}^n e^{-\frac{\|x_i - x_j\|^2}{r_a^2}} \quad \text{avec } r_a > 0 \quad (V.52)$$

$P_i$  : potentiel d'attraction de l'exemple  $x$  ;

$r_a$  : le rayon de voisinage.

L'exemple ayant le plus fort potentiel,  $P_1^*$  noté  $X_1^*$  devient le centre du premier groupe. Puis, le potentiel de chacun des exemples est diminué d'une valeur qui dépend de sa distance à  $X_1^*$ , la nouvelle valeur du potentiel pour l'exemple  $i$  est donnée par :

$$P_i = p_i - p_1^* \sum_{j=1}^n e^{-\frac{\|x_i - x_j\|}{r_a}} \quad \text{avec } r_b \approx 1.5 r_a \quad (V.53)$$

Le processus est réitéré. Afin d'utiliser l'introduction du bruit contenue dans les données, deux seuils sont définis  $s^+$  et  $s^-$ , typiquement 0.5 et 0.15, et un centre candidat,  $x_k^*$  à l'étape  $k$  et son potentiel associé  $P_k^*$ , est géré de la façon suivante :

1. Si  $P_k^* > P_1^* * s^+$ ,  $x_k^*$  est accepté comme centre d'un nouveau groupe ;
2. Si  $P_k^* < P_1^* * s^-$ ,  $x_k^*$  est rejeté comme centre d'un nouveau groupe et l'algorithme s'arrête ;
3. Sinon

et le centre du groupe le plus proche  $x_k^*$ . Soit  $d_{\min}$  la distance entre

Si

$$\frac{d_{\min}}{r_a} + \frac{P_k^*}{P_1^*} \geq 1 \quad x_k^* \text{ est accepté comme centre d'un nouveau groupe}$$

(Il est suffisamment loin du groupe le plus proche)

est rejeté, son potentiel est mis à 0, et l'algorithme continue.  $x_k^*$  Sinon

Cet algorithme nous semble être sensible aux différents paramètres, rayons de voisinage et seuils, malheureusement, le choix de ces valeurs demeure empirique.

Dans le cas de la méthode du «Groupage Soustractive», comme pour la méthode «*Gird Partition*», la construction du modèle neuroflou passe par :

- Le choix de la fonction d'appartenance pour chaque variable d'entrée.
- Le choix du centre du groupage soustractive, pour chaque variable d'entrée.
- Une fois le centre choisi (généralement entre 0.15 – 0.8), la répartition des variables d'entrée en fonction d'appartenance se fait d'une façon automatique et sans aucun choix préalable (l'algorithme décrit plus haut procède à la répartition des variables d'entrée en différents sous ensemble appelés *Cluster*).
- La construction de la base de règles, se fait elle même d'une façon automatique, et on constate que, contrairement à la méthode «*Gird Partition*», chaque *Cluster* de chaque variable d'entrée sera utilisé une seule fois dans une seule règle, ce qui limite considérablement le nombre de règle à optimiser et par conséquent accélère l'apprentissage du modèle, et nous avons constatés que le nombre de règle varie généralement entre 2 et 20 selon le centre du groupage soustractive, choisi.

### V.7.3.1. Résultats et interprétation

Dans ce qui suit nous allons présentés les résultats détaillées obtenus par la méthode du «Groupage Soustractive» ainsi qu'une comparaison avec la méthode «*Gird Partition*». La fonction d'appartenance utilisée est la fonction Gaussienne ; qui a donnée de meilleurs résultats par rapport aux autres fonctions. Nous avons procéder à la variation du paramètre clé qui est le centre du groupage soustractive, entre 0.8 et 0.35. Le tableau (V.11) présente la répartition des variables d'entrées en sous ensembles flous pour la variante finale retenue (M06-ANFIS), Les résultats obtenus par des fonctions d'appartenance Gaussienne sont présentés dans le Tableau (V.12), la base derègles établis pour est présentée dans le Tableau (V.13).

**Tableau V.11. Répartition des variables en sous ensembles floues, à base de la méthode du groupage soustractive pour le modèle ANFIS**

Variable	Désignation	Sous ensembles floues	Nombre de règles
Turbidité	input1	in1Clus1 à in1Clus15	15
Température	input2	in1Clus1 à in1Clus15	
pH	input3	in1Clus1 à in1Clus15	
conductivité	input4	in1Clus1 à in1Clus15	
Oxygène dissous	input5	in1Clus1 à in1Clus15	
Les UV <sub>254</sub>	input6	in1Clus1 à in1Clus15	
Coagulant	output	out1Clus1 à out1Clus15	

**Tableau V.12. Résultats obtenus par la méthode du Groupage Soustractive**

Modèle	Apprentissage			Validation			Test		
	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE	CC
Centre = 0.8									
M02-ANFIS	6.16	4.89	0.56	6.17	4.76	0.59	5.87	4.55	0.55
M03-ANFIS	5.98	4.79	0.60	6.01	4.49	0.62	5.75	4.49	0.58
M04-ANFIS	5.009	4.01	0.74	4.70	3.68	0.79	5.14	4.03	0.68
M05-ANFIS	3.82	2.98	0.86	8.86	6.35	0.21	5.06	3.64	0.72
M06-ANFIS	3.56	2.77	0.87	4.12	3.11	0.84	4.55	3.38	0.78
Centre = 0.6									
M02-ANFIS	5.88	4.58	0.61	5.94	4.80	0.57	6.28	4.97	0.58
M03-ANFIS	5.35	4.10	0.69	5.68	4.47	0.62	5.96	4.63	0.63
M04-ANFIS	4.29	3.22	0.81	5.06	3.87	0.73	4.71	3.63	0.79
M05-ANFIS	2.70	1.93	0.93	8.32	5.84	0.37	4.10	2.83	0.86
M06-ANFIS	2.62	1.88	0.93	4.20	3.12	0.83	3.79	2.72	0.87
Centre = 0.5									
M02-ANFIS	5.13	3.89	0.73	5.85	4.38	0.63	5.81	4.41	0.58
M03-ANFIS	4.76	3.54	0.77	5.35	3.99	0.70	5.01	3.66	0.70
M04-ANFIS	32.3	2.39	0.90	3.97	2.98	0.85	3.92	2.77	0.83
M05-ANFIS	2.20	1.65	0.95	10.19	7.79	0.01	3.86	2.37	0.85
M06-ANFIS	2.56	1.88	0.94	4.17	2.91	0.86	3.50	2.59	0.85
Centre = 0.35									
M02-ANFIS	5.41	4.14	0.69	5.22	4.08	0.71	5.39	4.02	0.67
M03-ANFIS	4.08	2.81	0.83	4.81	3.30	0.76	3.86	2.69	0.84
M04-ANFIS	2.65	1.93	0.93	3.82	2.50	0.86	3.53	2.55	0.87
M05-ANFIS	1.94	1.41	0.96	10.76	8.06	0.07	2.90	2.01	0.91
M06-ANFIS	1.59	1.15	0.97	3.19	2.08	0.90	2.74	1.95	0.92

On constate d'après les deux tableaux (V.13) et (V.14), premièrement le nombre de règles à optimiser est moins important que le nombre établis par la méthode Gird partition, d'ailleurs pour le meilleure modèle utilisant six entrées le nombre est de six uniquement, ce qui accélère l'apprentissage du réseau et par conséquent améliore les résultats obtenus. On remarque aussi que pour la méthode du Groupage Soustractive, plus le critère (*centre*) est petit, plus on constate une amélioration des performances du modèle, les meilleurs résultats sont obtenus avec un centre de 0.35 ; le coefficient de corrélation atteint la barre de 0.97 en phase d'apprentissage tandis qu'on phase de test il est de l'ordre de 0.92, ce qui correspond aux meilleures résultats obtenus à ce stade de la modélisation. Le tableau (V.13) présente la base de règle établis pour le modèle final retenu par la méthode du Groupage Soustractive, Les deux Figures (V.34) et (V.35), présentent respectivement la forme initiale et finale des fonctions d'appartenance gaussiennes utilisées. Le tableau (V.14) présente la formulation mathématique du modèle final retenu, les paramètres finaux correspondant à l'ensemble de la base de règles établis, sont présentés dans le tableau (V.15).

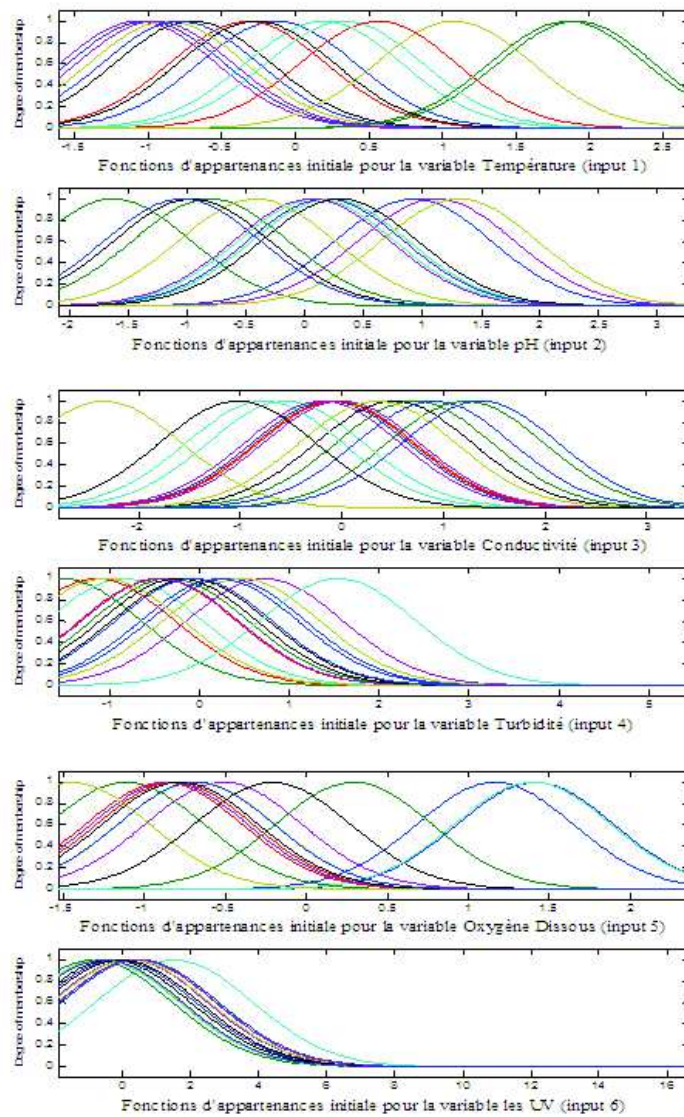


Figure V.34 Forme initiale des fonctions d'appartenances Gaussiennes utilisées pour le modèle final retenu ANFIS, à base de la méthode du groupage soustractive.



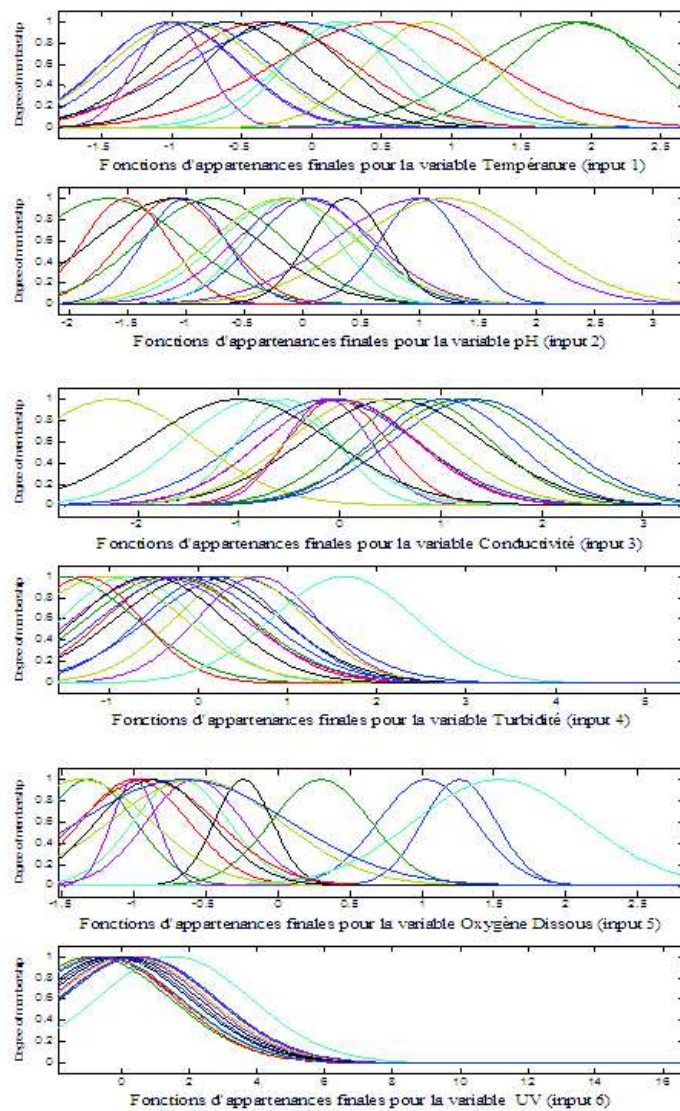


Figure V.35 Forme finale des fonctions d'appartenance Gaussiennes utilisées pour le modèle final retenu ANFIS, à base de la méthode du groupage soustractive

Tableau V.13. Base de règles établie par la méthode du groupage soustractive pour le modèle ANFIS



Construction de la base de règles		
Règle N°	Antécédent	Conséquent
01	Si (input1 est in1Clus1) et (input2 est in2Clus1) et (input3 est in3Clus1) et (input4 est in4Clus1) et (input5 est in5Clus1) et (input6 est in6Clus1)	Alors output estout1Clus1
02	Si (input1 est in1Clus2) et (input2 est in2Clus2) et (input3 est in3Clus2) et (input4 est in4Clus2) et (input5 est in5Clus2) et (input6 est in6Clus2)	Alors output estout1Clus2
03	Si (input1 est in1Clus3) et (input2 est in2Clus3) et (input3 est in3Clus3) et (input4 est in4Clus3) et (input5 est in5Clus3) et (input6 est in6Clus3)	Alors output estout1Clus3
04	Si (input1 est in1Clus4) et (input2 est in2Clus4) et (input3 est in3Clus4) et (input4 est in4Clus4) et (input5 est in5Clus4) et (input6 est in6Clus4)	Alors output estout1Clus4
05	Si (input1 est in1Clus5) et (input2 est in2Clus5) et (input3 est in3Clus5) et (input4 est in4Clus5) et (input5 est in5Clus5) et (input6 est in6Clus5)	Alors output estout1Clus5
06	Si (input1 est in1Clus6) et (input2 est in2Clus6) et (input3 est in3Clus6) et (input4 est in4Clus6) et (input5 est in5Clus6) et (input6 est in6Clus6)	Alors output estout1Clus6
07	Si (input1 est in1Clus7) et (input2 est in2Clus7) et (input3 est in3Clus7) et (input4 est in4Clus7) et (input5 est in5Clus7) et (input6 est in6Clus7)	Alors output estout1Clus7
08	Si (input1 est in1Clus8) et (input2 est in2Clus8) et (input3 est in3Clus8) et (input4 est in4Clus8) et (input5 est in5Clus8) et (input6 est in6Clus8)	Alors output estout1Clus8
09	Si (input1 est in1Clus9) et (input2 est in2Clus9) et (input3 est in3Clus9) et (input4 est in4Clus9) et (input5 est in5Clus9) et (input6 est in6Clus9)	Alors output estout1Clus9
10	Si (input1 est in1Clus10) et (input2 est in2Clus10) et (input3 est in3Clus10) et (input4 est in4Clus10) et (input5 est in5Clus10) et (input6 est in6Clus10)	Alors output estout1Clus10
11	Si (input1 est in1Clus11) et (input2 est in2Clus11) et (input3 est in3Clus11) et (input4 est in4Clus11) et (input5 est in5Clus11) et (input6 est in6Clus11)	Alors output estout1Clus11
12	Si (input1 est in1Clus12) et (input2 est in2Clus11) et (input3 est in3Clus12) et (input4 est in4Clus12) et (input5 est in5Clus12) et (input6 est in6Clus11)	Alors output estout1Clus12
13	Si (input1 est in1Clus13) et (input2 est in2Clus13) et (input3 est in3Clus13) et (input4 est in4Clus13) et (input5 est in5Clus13) et (input6 est in6Clus13)	Alors output estout1Clus13
14	Si (input1 est in1Clus14) et (input2 est in2Clus14) et (input3 est in3Clus14) et (input4 est in4Clus14) et (input5 est in5Clus14) et (input6 est in6Clus14)	Alors output estout1Clus14
15	Si (input1 est in1Clus15) et (input2 est in2Clus15) et (input3 est in3Clus15) et (input4 est in4Clus15) et (input5 est in5Clus15) et (input6 est in6Clus15)	Alors output estout1Clus15

Tableau V.14. Formulation mathématique de la base de règles établie par la méthode du groupage soustractive pour le modèle ANFIS

Règle N°	Antécédent	Conséquent
01	$A_1 * pH + A_2 * CON + A_3 * TE + A_4 * TU + A_5 * OD + A_6 * UV + A_7$	CO
02	$A_1 * pH + A_2 * CON + A_3 * TE + A_4 * TU + A_5 * OD + A_6 * UV + A_7$	CO
03	$A_1 * pH + A_2 * CON + A_3 * TE + A_4 * TU + A_5 * OD + A_6 * UV + A_7$	CO
04	$A_1 * pH + A_2 * CON + A_3 * TE + A_4 * TU + A_5 * OD + A_6 * UV + A_7$	CO
05	$A_1 * pH + A_2 * CON + A_3 * TE + A_4 * TU + A_5 * OD + A_6 * UV + A_7$	CO
06	$A_1 * pH + A_2 * CON + A_3 * TE + A_4 * TU + A_5 * OD + A_6 * UV + A_7$	CO
07	$A_1 * pH + A_2 * CON + A_3 * TE + A_4 * TU + A_5 * OD + A_6 * UV + A_7$	CO
08	$A_1 * pH + A_2 * CON + A_3 * TE + A_4 * TU + A_5 * OD + A_6 * UV + A_7$	CO
09	$A_1 * pH + A_2 * CON + A_3 * TE + A_4 * TU + A_5 * OD + A_6 * UV + A_7$	CO
10	$A_1 * pH + A_2 * CON + A_3 * TE + A_4 * TU + A_5 * OD + A_6 * UV + A_7$	CO
11	$A_1 * pH + A_2 * CON + A_3 * TE + A_4 * TU + A_5 * OD + A_6 * UV + A_7$	CO
12	$A_1 * pH + A_2 * CON + A_3 * TE + A_4 * TU + A_5 * OD + A_6 * UV + A_7$	CO
13	$A_1 * pH + A_2 * CON + A_3 * TE + A_4 * TU + A_5 * OD + A_6 * UV + A_7$	CO
14	$A_1 * pH + A_2 * CON + A_3 * TE + A_4 * TU + A_5 * OD + A_6 * UV + A_7$	CO
15	$A_1 * pH + A_2 * CON + A_3 * TE + A_4 * TU + A_5 * OD + A_6 * UV + A_7$	CO

Tableau V.15. Paramètres finaux retenus pour le modèle ANFIS établis par la méthode du groupage soustractive

N° de règle	Paramètres (Antécédent)						
	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	A <sub>7</sub>
01	-0.401	0.3727	0.3155	-0.4638	0.5645	0.04778	-1.488
02	-0.4791	0.644	2.324	-1.044	-1.039	0.09815	-1.716
03	-10.63	2.523	5.651	2.146	3.373	-0.6377	8.786
04	3.686	0.0191	2.147	-2.38	4.608	0.2693	-2.921
05	-0.3333	-0.5261	1.366	0.03238	-0.3481	0.179	1.282
06	-0.2198	-0.7254	-0.3288	0.1808	0.8257	-0.01511	0.1304
07	1.318	-0.3778	1.951	1.156	0.713	0.7575	-2.973
08	-0.01226	0.04768	0.3306	-0.08397	0.3134	0.09156	-1.601
09	-0.2282	-0.135	-0.1642	-0.2788	-0.3867	0.01601	2.154
10	-2.254	-5.947	-2.046	4.647	-0.997	-1.058	0.389
11	0.8004	0.1078	0.9364	-0.3125	-0.0232	-0.4366	3.173
12	2.895	1.129	-1.731	1.764	3.937	-1.892	4.741
13	-1.015	2.485	-1.716	-0.4327	2.451	0.1881	5.523
14	5.02	0.5454	-0.5888	-1.192	-1.281	0.2373	1.322
15	-8.15	0.2052	3.942	0.4119	2.29	-0.2123	-12.42

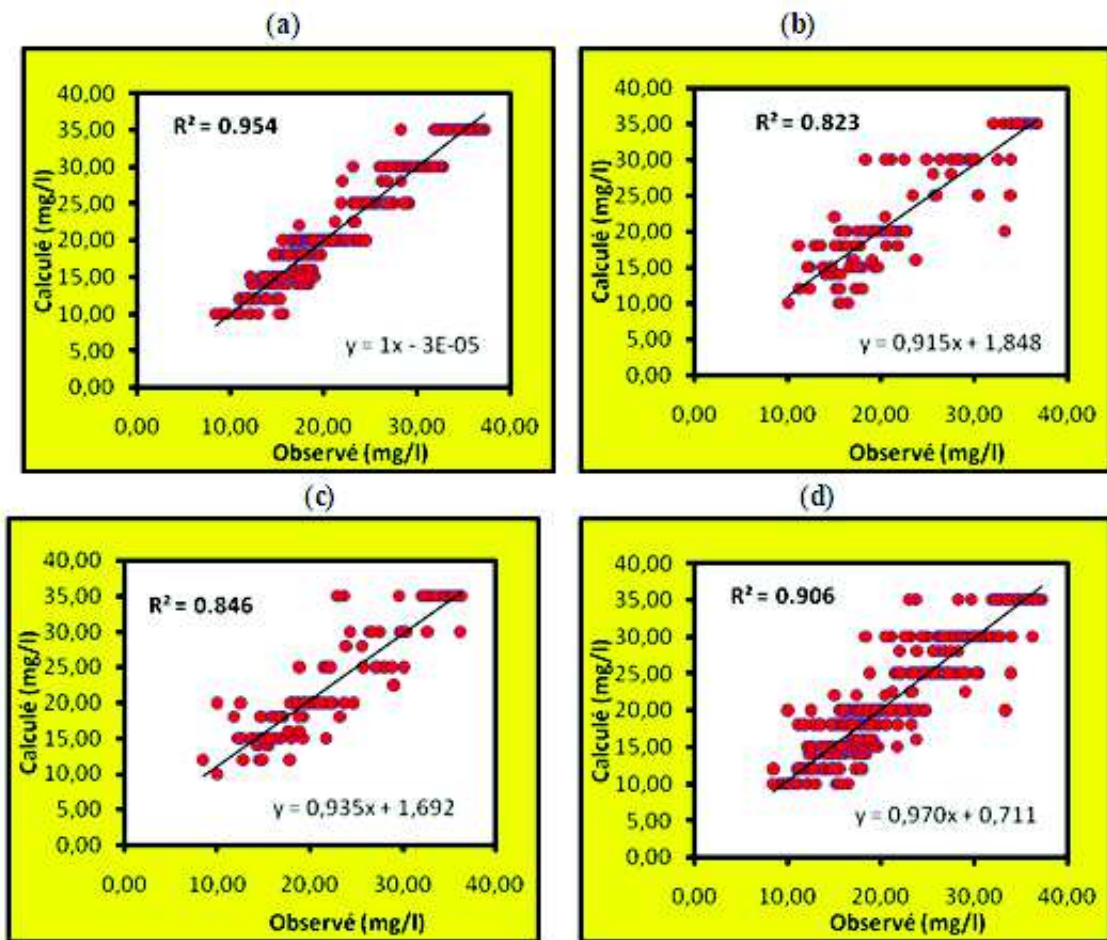


Figure V.36 Comparaison des résultats du modèle neuroflou ANFIS à base de la méthode du groupage soustractive: (a) apprentissage, (b) validation, (c) test et (d) l'ensemble des données.

Tableau V.16. Nombre de règles établies (à base du centre 0.35)

Modèle	M02-ANFIS	M03-ANFIS	M04-ANFIS	M05-ANFIS	M06-ANFIS
Nombre de règles	7	10	13	14	15

La Figure (V.36 a-d) représente les courbes Scatterplots pour le modèle retenu (M06-ANFIS) à base de la méthode du *Groupage Soustractive*, la comparaison entre les valeurs observée et calculée est présentée dans la figure (V.37a-c), le tableau IV.16 présente le nombre de règles établies par chaque variante testée.

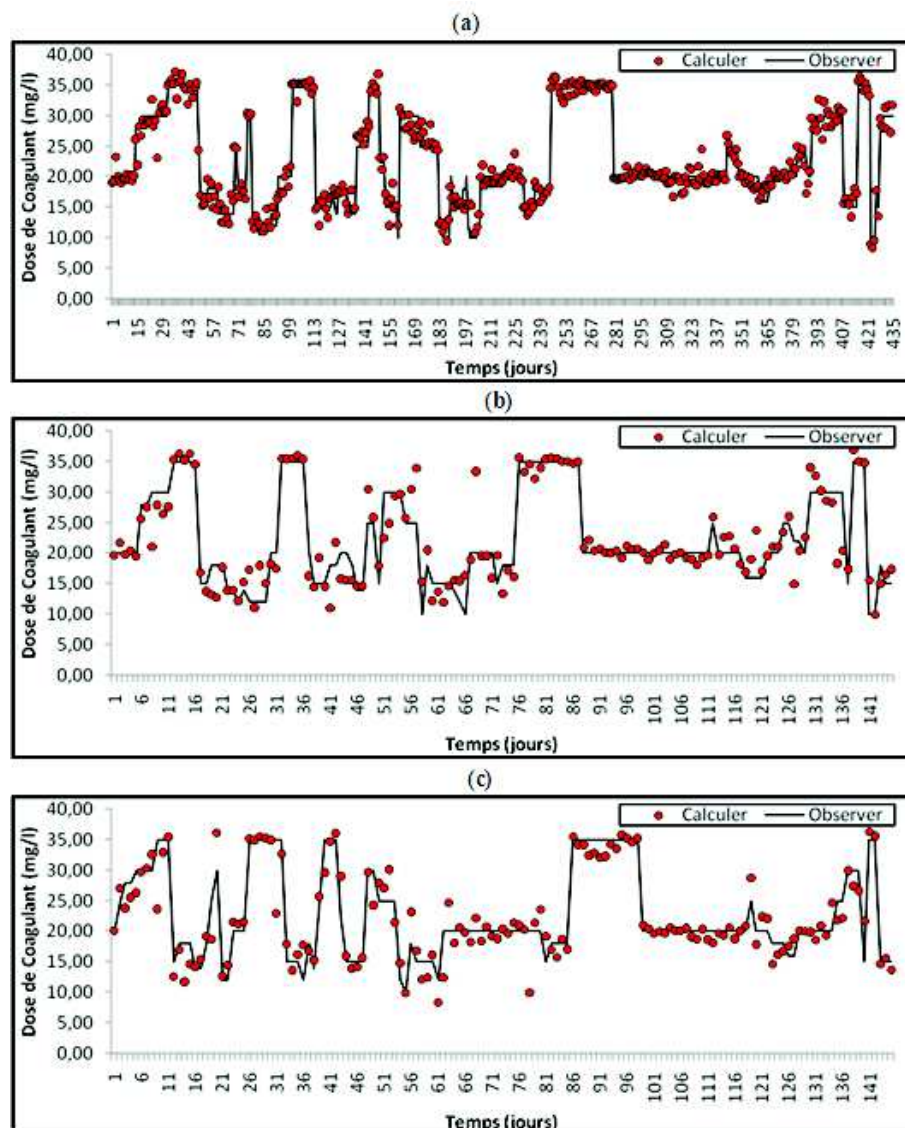


Figure V.37 Modélisation de la dose de coagulant par le modèle neuroflou ANFIS, à base de la méthode du groupage soustractive :  
 (a) partie apprentissage, (b) partie validation et (c) partie test

#### V.7.4. Comparaison des résultats obtenus par le modèle ANFIS

Nous avons obtenu des résultats meilleurs en se basant sur la technique du *Groupage Soustractive* par rapport à la stratégie *Gird Partition*, nous pouvons expliquer cette différence d'une part par la variation des paramètres utilisés par chaque démarche. D'autre part, lorsque le nombre de règles augmente l'apprentissage par la méthode *Gird Partition*, devient difficile et complexe. C'est essentiellement pour cet raison que les résultats obtenus par la stratégie du *Groupage Soustractive* sont généralement supérieurs. Le tableau (V.17) présente une comparaison entre les deux stratégies appliquées, le tableau (V.18) présente une comparaison générale entre le modèle ANFIS et tous les autres modèles utilisés dans le cadre de ce travail de recherche.

Tableau V.17. Comparaison des performances du modèle ANFIS pour les deux méthodes étudiées

Modèle	Apprentissage			Validation			Test		
	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE	CC
Modèle à base de la méthode du <i>groupage soustractive</i>									
M02-ANFIS	5.41	4.14	0.69	5.22	4.08	0.71	5.39	4.02	0.67
M03-ANFIS	4.08	2.81	0.83	4.81	3.30	0.76	3.86	2.69	0.84
M04-ANFIS	2.65	1.93	0.93	3.82	2.50	0.86	3.53	2.55	0.87
M05-ANFIS	1.94	1.41	0.96	10.76	8.06	0.07	2.90	2.01	0.91
M06-ANFIS	1.59	1.15	0.97	3.19	2.08	0.90	2.74	1.95	0.92
Modèle à base de la méthode <i>Gird Partition</i> (Triangulaires)									
Modèle	Apprentissage			Validation			Test		
	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE	CC
M02-ANFIS	6.01	4.67	0.60	6.25	5.00	0.54	5.99	4.80	0.56
M03-ANFIS	4.88	3.43	0.76	11.73	5.24	0.31	6.76	4.45	0.50
M04-ANFIS	2.86	1.90	0.92	10.66	5.37	0.53	49.98	9.81	0.06
M05-ANFIS	1.01	0.52	0.99	8.44	5.33	0.59	11.59	5.51	0.49
M06-ANFIS	0.35	0.17	0.99	9.30	5.87	0.57	12.75	6.99	0.36
Modèle à base de la méthode <i>Gird Partition</i> (Gaussiennes)									
Modèle	Apprentissage			Validation			Test		
	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE	CC
M02-ANFIS	5.62	4.37	0.63	7.72	5.44	0.45	6.33	5.22	0.55
M03-ANFIS	4.25	2.96	0.81	7.67	4.88	0.54	6.12	4.22	0.62
M04-ANFIS	2.46	1.51	0.94	5.35	3.64	0.77	8.15	4.47	0.61
M05-ANFIS	0.72	0.38	0.99	36.42	10.74	0.28	20.47	9.34	0.36
M06-ANFIS	0.34	0.15	0.99	12.11	6.49	0.48	15.40	7.73	0.53

Tableau V.18. Comparaison des performances des modèles étudiées

Modèle	Apprentissage			Validation			Test		
	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE	CC
M06-RLM	5.99	4.97	0.57	6.39	5.42	0.54	6.25	5.24	0.56
M06-MLPNN	3.20	2.43	0.90	4.00	2.93	0.84	4.40	3.32	0.79
M06-RBFNN	2.35	1.74	0.94	3.87	2.71	0.86	3.35	2.46	0.87
M06-GRNN	0.13	0.01	0.99	2.92	1.14	0.92	3.11	1.29	0.90
M06-ANFIS	1.59	1.15	0.97	3.19	2.08	0.90	2.74	1.95	0.92

## V.8. Conclusion

Le développement de modèles mathématiques des systèmes industriels complexes est un sujet central dans plusieurs disciplines des sciences et de l'ingénierie. Néanmoins, le besoin d'une forte compréhension des éléments physiques de base (comme c'est le cas dans l'approche de type logique floue) constitue une grande restriction au niveau pratique quand on est confronté à des systèmes complexes, il est à remarquer que le procédé de coagulation fait partie de ces systèmes. D'un autre côté, dans l'approche de type floue, il est supposé que le procédé étudié peut être représenté en utilisant une structure générale « *Entrée-Sortie* » qui approxime le comportement du système. Dans ce cas, le problème



de modélisation consiste alors à proposer la structure appropriée pour l'approximateur et d'estimer les paramètres du modèle, en utilisant habituellement des données entrée-sortie représentatives du comportement, afin de capturer correctement la non linéarité du système.

Malgré les bons résultats obtenus par les réseaux de neurones, on s'est posé la question sur la possibilité d'améliorer encore les performances de la modélisation de la dose de coagulant. Pour cela la technique neuroflou a été discuté à travers ce cinquième chapitre, l'utilisation conjointe des réseaux de neurones et de la logique floue «systèmes neuroflous», permet de tirer les avantages des deux méthodes : les capacités d'apprentissage de la première, et la lisibilité et la souplesse de la seconde.

Nous avons commencés le présent chapitre par la description des différentes composantes de la logique floue, l'étude bibliographique réalisée a mis en évidence l'existence d'une classification des systèmes d'inférences flous, il s'agit des deux approches «Mamdani» et «Takagi\_Sugeno».

Par la suite on s'est penchée sur la manière de construire un modèle neuroflou de type «Takagi\_Sugeno». Dans ce contexte, les réseaux ANFIS (système d'inférence floue paramétré par apprentissage neuronal) ont été considérés appropriés parce que ces systèmes combinent les capacités d'apprentissage et de généralisation des réseaux de neurones, avec la formalisation de règles imprécises ou entachées d'incertitudes correspondant à une information approximative ou à une information partielle ou partiellement erronée.

Le travail a ensuite porté sur la construction de la base de règles d'un système à base de logique floue, et phase cruciale pour la construction de ce type de modèle. Dans ce contexte deux techniques on été appliquées à savoir la technique *GIRD PARTITION* «GP» et la technique du *GROUPE SOUSTRUCTIVE* «GP». La première technique est basée sur le principe d'une répartition des variables d'entrée en plusieurs sous ensembles flous, une combinaison des sous ensembles permis la construction de la base de règles. Nous avons démontrés que malgré la présence d'une large gamme de fonctions d'appartenance (triangulaire, trapézoïdale, gaussienne,..., etc.), cette catégorie de technique présente un grand handicap dans la construction du modèle ANFIS, dans le cas où le nombre d'entrées du système dépasse six (06) et lorsque le nombre des variables linguistiques pour chaque variable d'entrée dépasse lui aussi trois(03). Nous avons constaté qu'au delà de trois sous ensembles flous pour chaque variable d'entrée, le nombre de règles floues deviennent important à un point où l'apprentissage du modèle devient colossal et nécessite un temps très important.

Pour remédier à cette situation, nous avons proposés une nouvelle technique pour l'optimisation du nombre de règles floues à utiliser par le modèle ANFIS dont les performances de point de vue modélisation de la dose de coagulant ont été mis en évidence, il s'agit de la technique dite «*Groupage Soustractive*». L'idée fondamentale est que chaque exemple est considéré comme un centre de groupe éventuel et est affecté d'un potentiel d'attraction qui est lui même fonction d'un voisinage. La construction de la base de règle, se fait elle même d'une façon automatique, et on constate que, contrairement à la méthode «*Gird Partition*», chaque variable d'entrée se résume un centre ou *Cluster*, ce derniersera utilisé une seule fois dans une seule règle, ce qui limité considérablement le nombre de règle à optimiser et par conséquence accélère l'apprentissage du modèle, et nous avons constatés que le nombre de règle varie généralement entre 2 et 20 selon le centre du groupage soustractive choisi.



Les résultats obtenus sont d'une grande importance, la technique du «*Groupage Soustractive*» améliore nettement les résultats obtenus et elle dépasse la technique «*Gird Partition*», et en comparant les résultats obtenus par les réseaux de neurones, on peut conclure que le modèle ANFIS doit être considéré comme la meilleure technique pour la modélisation de la dose de coagulant.

# Chapitre VI Application Des Technique De Classification

## VI.1. Introduction

Nous consacrons ce dernier chapitre à la partie classification, phase finale de notre travail de recherche. Nous nous intéresserons plus particulièrement à la classification en mode non supervisé, c'est à dire ne nécessitant aucune connaissance a priori sur la classe d'appartenance d'un certain nombre d'observations dont on dispose pour la mise au point du classifieur.

A partir de la base de données dont on connaît les sorties souhaitées, une partition de l'espace des observations en autant de régions qu'il y a de classes doit être élaborée : c'est la phase d'apprentissage. Le classifieur ainsi mis au point permettra ensuite d'affecter à une nouvelle observation inconnue sa classe d'appartenance : c'est la phase de généralisation.

On s'intéresse plus particulièrement à la seconde grande famille de réseaux de neurones, utilisés pour traiter des problèmes de classification: les cartes topologiques auto organisatrices. Ces dernières font partie de la famille des modèles dits à « apprentissage non supervisé » ; par opposition avec les perceptrons multicouches et les autres réseaux de neurones à apprentissage supervisé qui ont été présentés précédemment. Les données à analyser constituées d'observations dont on cherche à comprendre la structure, il n'y a pas de but précis à atteindre, ni de réponse souhaitée.

Les méthodes, dites d'apprentissage « non supervisé », utilisées par les modèles de cartes topologiques auto-organisatrices, proviennent des techniques initialement mises au point pour l'apprentissage compétitif, parmi les premiers travaux dans le domaine, on peut citer ceux de Didday (Didday, 1970).

Comme toutes les méthodes de l'analyse de données, la classification automatique a pour but d'obtenir une représentation simplifiée des données initiales. Il s'agit donc d'une méthode de réduction des données. La classification est la définition de classes alors que le classement (discrimination) est le rangement dans des classes déjà existantes. Cette démarche est utilisée :

- En statistique, elle permet d'identifier plusieurs populations dans une population initiale hétérogène, et ainsi de faciliter une étude statistique ultérieure ;
- En sciences naturelles, la classification du règne animal et du règne végétal est l'une des classifications les plus connues ;
- Et, de manière plus générale, le fait de nommer des objets est une forme de classification.

La classification automatique, encore appelée clustering ou taxonomie numérique, recouvre l'ensemble des méthodes permettant la construction automatique de telles classifications. Une définition formelle de la classification, qui puisse servir de base à un processus automatisé, amène à se poser les questions suivantes :

- Comment les objets à classer sont-ils définis ?

- Comment définir la notion de ressemblance entre objets ?
- Qu'est-ce qu'une classe ?
- Comment sont structurées les classes ?
- Comment juger une classification par rapport à une autre ?

Pour effectuer cette classification, deux démarches sont généralement utilisées :

- On regroupe en classes les objets qui partagent certaines caractéristiques, ce genre de démarche aboutit à une classification mono théétique base de l'approche aristotélicienne (Sutcliffe, 1994), tous les objets d'une même classe partagent alors un certain nombre de caractéristiques ;
- On peut aussi regrouper en classe les objets qui posséderont des caractéristiques

« proches ». Cette démarche est dite polythétique. Généralement, on utilise pour cela, la notion de mesure de proximité (distance, dissimilarité), c'est cette approche qui sera abordée dans ce chapitre.

## VI.2. Formalisation mathématique d'un problème de classification

Le principe de base des différentes méthodes de classification est de former, à partir d'individus, des groupes ou classes qui soient les plus homogènes possibles. L'homogénéité, définie à partir d'un critère de classification et d'une notion de distance (plus la distance entre deux éléments est faible, plus ils se ressemblent), implique que les éléments d'une même classe sont très similaires (notion de compacité et de distance intra-classe), tandis que ceux appartenant à des groupes distincts sont très dissemblables (notion de séparabilité et de distance interclasses).

Traduit en termes mathématiques, un problème de classification comporte les ingrédients suivants :

- Une population de  $N$  individus  $I^i$ , ( $i$  variant de 1 à  $N$ ) ;
- $P$  variables descriptives  $X^i d$ , qui permettent de décrire les individus ; elles sont aussi appelées plus simplement descripteurs ( $d$  varie de 1 à  $P$ ).
- $C$  classes  $C_k$  dans lesquelles on cherche à ranger les individus ( $k$  varie de 1 à  $C$ ) ;

Résoudre un problème de classification, c'est trouver une application de l'ensemble des objets à classer, décrits par les variables descriptives choisies, dans l'ensemble des classes. L'algorithme ou la procédure qui réalise cette application est appelé classifieur.

## VI.3. Les cartes auto organisatrices de KOHONEN

### VI.3.1. Historique

---

L'algorithme des cartes auto-organisatrices de Kohonen est l'un des modèles neuronaux les plus anciens, il représente le prototype de l'apprentissage non supervisé dans le domaine. Un grand nombre de résultats théoriques sont maintenant disponibles, qui explicitent les liens entre ce modèle et les méthodes statistiques de reconnaissance des formes: classification automatique, analyse en composantes principales non linéaires, approximation de fonctions densité (Kohonen, 1995).

Ce n'est que très récemment qu'il a été montré que le cerveau forme des représentations neuronales sur des aires corticales spécialisées (traitement de la parole, vision, audition, ...) qui sont des images des différentes expériences sensorielles vécues. Entre ces aires, sont présentes des aires associatives sur lesquelles des signaux de différentes modalités convergent : ils s'agissent de véritables cartes respectant la topologie des informations qu'elles traitent. Bien que les neurones semblent placés sur des cartes (aires) spécifiques ils peuvent participer à différentes activités grâce à une grande plasticité neuronale qui illustre parfaitement les processus d'auto-organisation simple. La formation de cartes topologiquement correctes de distribution des signaux, en dimension 1 ou 2, à l'aide d'unités de traitement de configuration initiale arbitraire, inspira Kohonen (Kohonen, 1981) pour la formalisation, par un modèle neuronal artificiel, des représentations internes du cerveau. Ce modèle est une généralisation des travaux de Von der Malsburg (Malsburg, 1973), qui a décrit le cortex visuel et ses capacités d'auto-organisation et de ceux de Willshaw (Willshaw et Von der Malsburg, 1976) et Amari (Amari, 1981) portant sur la formation de projections topographiques directes entre deux structures laminaires.

Le premier article de T. Kohonen sur ce sujet a été publié en 1982 (Kohonen, 1982). Son modèle, que nous décrivons plus précisément par la suite porte le nom de *Cartes topologiques auto organisatrices* ou *Self Organising Feature Map*, ou *Self Organizing Map* (SOM). Il publie en 1984 la première version de son ouvrage : *Self Organization and associative memory* (Kohonen, 1984), qui est régulièrement remis à jour.

Dès 1988 de nombreuses études ont commencé à voir le jour pour comprendre le fonctionnement de ce modèle appartenant au domaine de l'apprentissage non-supervisé. De nombreuses améliorations relatives à l'architecture du réseau, au réglage des paramètres, à l'étude théorique du fonctionnement et à la représentation des sorties ont été apportées dans les années 90. Les cartes topologiques n'ont commencé à apparaître en milieu industriel qu'au début des années 90 (Valentin, 2000).

Les cartes auto-organisatrices de Kohonen ont fait l'objet d'intéressantes applications dans des domaines de recherche très variés, allant de l'ingénierie à la chimie (Blasco *et al.* 2000), de la télédétection (Foody, 1999 ; Foody et Cutler 2003 ; Aurelle *et al.* 1999) à l'économie et au suivi de la qualité de l'eau (Aguilera *et al.* 2001 ; Walley *et al.* 2000).

### VI.3.2. Définition

---

On entend par "auto-organisation" un processus par lequel un système trouve, sans apprentissage supervisé, une solution optimale à un problème. D'un point de vue pratique dans les domaines d'application, ces modèles d'auto-organisation ont pour objectif de représenter des données complexes, souvent bruitées (caractérisées par de nombreuses mesures, c'est-à-dire appartenant généralement à un espace de grande dimension) dans un espace discret dont la topologie est limitée à 1, 2 voire 3 dimensions. Globalement, il s'agit donc de modèles de Quantification Vectorielle (QV) qui seront dotés de propriétés topologiques particulières.

### VI.3.3. Principe de base

Teuvo Kohonen a proposé dès 1982 (Kohonen, 1982a, 1982b) un algorithme qui produit une carte d'organisation topologique (Kohonen, 1989 ; Kohonen, 1995). Le processus ne dépend que des entrées et ne nécessite pas l'intervention d'un superviseur, on parle dans ce cas d'auto organisation. Sa fonction principale est de faire correspondre les éléments de l'espace d'entrée avec des unités ordonnées sur une carte – qui est une représentation graphique où chaque unité est entourée de ses voisines, les voisinages ayant été définis a priori. Le résultat est une fonction de l'espace des entrées vers l'ensemble des unités, telle que les images de deux éléments voisins au sens d'une certaine distance dans l'espace des entrées sont la même unité ou des unités voisines sur la carte.

Les propriétés de l'algorithme et de ses cartes de sortie permettent conjointement d'analyser des données qui ont une structure non linéaire, et de réaliser des représentations graphiques qui mettent en évidence certaines de leurs caractéristiques. L'application la plus courante est la classification de l'espace d'entrée, où l'on définit une notion de voisinage entre les classes qui n'est pas prise en compte par les méthodes de classification classiques.

### VI.3.4. Architecture du Réseau

Une carte de Kohonen est un réseau de neurones constitué de deux couches totalement connectées. La couche d'entrée prend la valeur du vecteur d'entrée  $X$ , de dimension  $m$ , la seconde est un treillis régulier à deux ou trois dimensions dont chacun des nœuds est un neurone linéaire. Ce treillis représente un espace discret, de faible dimension  $d \ll m$ , muni d'une topologie engendrée par sa structure. En effet, la géométrie du treillis (Figure VI.1) définit le type de voisinage de chaque neurone.

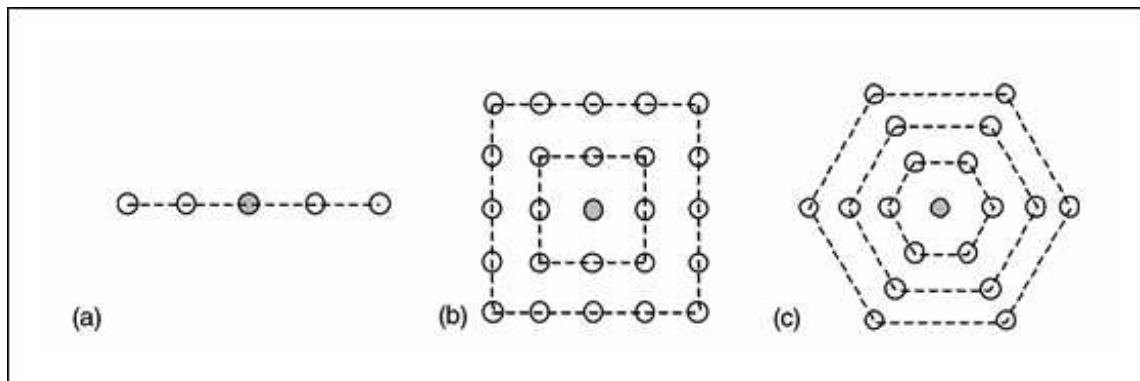


Figure VI.1 Définitions de voisinage employés dans les cartes topologiques de Kohonen, (a) voisinage linéaire, (b) voisinage carré et (c) voisinage hexagonal.

D'une manière générale, les cartes topologiques vont projeter les données initiales sur un espace discret et régulier de faible dimension (en général 1 ou 2). Les espaces utilisés sont des treillis réguliers dont chacun des nœuds est occupé par un neurone. La notion de voisinage entre neurones découle alors directement de la structure et définit une topologie de la carte. Grâce au procédé d'auto-organisation la topologie qui lie les données initiales est conservée au niveau des réponses proposées par le réseau, la localisation des neurones actifs reproduit les liens existants au niveau des données initiales.

La figure (VI.2) illustre un tel réseau dans sa configuration la plus courante, le cas bidimensionnel. L'entrée du réseau, unique et commune à tous les neurones, est le vecteur de caractéristiques  $X$  qui représente un objet. A chaque neurone sont associés

des coordonnées  $(x, y)$  indiquant sa position sur la carte, ainsi qu'un vecteur de *poids synaptiques*, appelé ainsi par analogie avec les synapses rencontrées dans le cerveau.

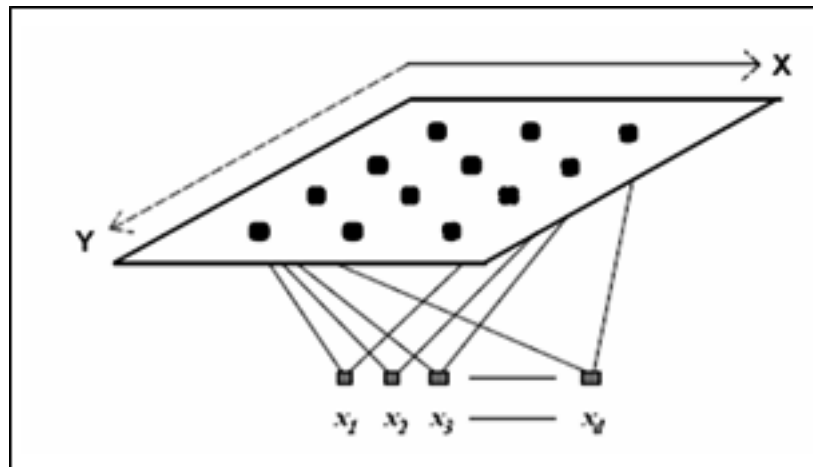


Figure VI.2 Carte auto organisatrice à deux dimensions.

En général, ces neurones sont disposés en un arrangement en 1 ou 2 dimensions. Chacune des  $N$  unités d'entrée est reliée à chacun des neurones de sortie par  $p$  connexions modifiables. Le réseau contient donc  $Nxp$  connexions au total. Enfin, chaque neurone possède des connexions latérales récurrentes, qui viendront modifier le patron d'excitation de façon à créer la forme d'un "chapeau mexicain". Ce chapeau mexicain est caractérisé par 3 zones distinctes (figure VI.3):

- Une région restreinte d'excitation ;
- Une région de pénombre ou d'action inhibitoire ;
- Une région d'excitation plus faible; cette zone possède une action négligeable et est habituellement ignorée.

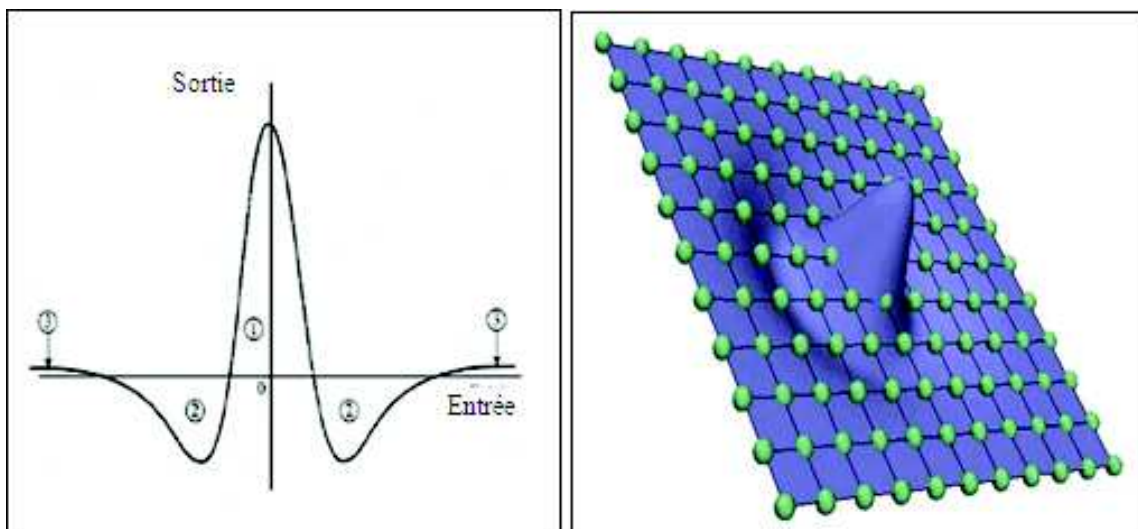


Figure VI.3 Représentation du neurone le plus excité en activité sous forme de chapeau mexicain.

### VI.3.5. Règle de transmission



Le réseau de Kohonen est un réseau de neurones artificiels, qui apprend à modéliser un espace de données, aussi appelé ensemble de stimuli, par un ensemble de prototypes, les « neurones ». Stimuli et neurones sont des vecteurs de l'espace d'entrée. Si le réseau est constitué d'un nombre  $N$  de neurones, on fait donc une partition en  $N$  classes de l'espace d'entrée constitué d'un nombre de stimuli  $Q \gg N$ . Chaque neurone  $i$  étant un prototype, on lui associe un vecteur de poids  $w_i$  appartenant à l'espace d'entrée. Le réseau apprend donc, pour un ensemble de stimuli donné, à positionner dans cet espace d'entrée  $N$  centres  $\mathbf{W} = \{w_1, w_2, \dots, w_N\}$ . Le réseau apprend d'autant mieux que la partition, ou classification réalisée, est meilleure. Par exemple, dans le cas trivial où  $N = 1$ , la meilleure partition possible est évidemment le cas où  $\mathbf{W}$  est le centre de masse des données, qui minimise la distance à l'ensemble des données.

On voit ici que la qualité de modélisation dépend d'une métrique de distance. Dans un espace vectoriel, on utilise la norme  $\|x - y\|$  pour calculer la distance entre deux vecteurs  $x$  et  $y$ . La norme la plus utilisée dans un espace euclidien  $R_n$  est la norme euclidienne (ou  $l_2$ ), que l'on utilisera également par convention, mais qui pourrait être remplacée par toute norme correctement définie. On utilise donc :

$$\|X\| = \sqrt{\langle X, X \rangle} = \sqrt{X^T X} = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \quad (VI.1)$$

Où  $\langle \cdot, \cdot \rangle$  désigne le produit scalaire.

À présent, nous pouvons définir une fonction d'erreur absolue qui soit un critère quantitatif de la qualité de modélisation de nos données. Cette fonction, que nous appellerons  $E^{ext}$ , prend tout simplement en compte la distance qui sépare chaque stimulus du neurone, ou prototype, le plus proche. Elle est définie comme suit :

$$E^{ext} = \sum_{q=1}^Q \left\{ \min_{i \in \{1, \dots, N\}} \|P_q - W_i\| \right\} \quad (VI.2)$$

Où  $Q$  est le nombre de stimuli,  $N$  le nombre de neurones,  $w_i$  le neurone d'indice  $i$  et  $P_q$  le stimulus  $q$ . On peut se servir de cette fonction à la fin d'une simulation d'apprentissage, pour voir si la modélisation est bonne. Cependant, *cette fonction ne peut pas être utilisée par le réseau pendant l'apprentissage*. En effet, puisque l'apprentissage est non supervisé, le réseau ne peut pas être guidé par une fonction de coût ou d'erreur *externe* (d'où la notation  $E^{ext}$ ) pour améliorer ses performances. Il ne peut suivre les conseils d'un « professeur » qui connaît les bonnes et les mauvaises réponses parce qu'il a une connaissance globale du problème. Si, à un moment donné de l'apprentissage, notre exemple simpliste de réseau constitué d'un seul neurone semble diverger du centre de masse des stimuli, on ne peut pas l'influencer et corriger son comportement. Il doit rester autonome. Ainsi, le réseau de Kohonen se nomme alternativement *carte auto-organisée* ou *Self-Organizing Map* (SOM).

Outre la modélisation des entrées avec des prototypes, une carte auto-organisée permet de construire un graphe (généralement une chaîne unidimensionnelle ou grille bidimensionnelle de neurones) permettant de structurer cet espace (*quantification vectorielle*), donc de visualiser, en une ou deux dimensions, la topologie des liens entre les classes. Il convient de rappeler ici que le réseau de Kohonen n'est pas un simple algorithme de classification, mais bien un réseau de neurones artificiels, inspiré des travaux

physiologiques de Hubel et Wiesel (Hubel et Wiesel, 1962 ; 1963 ; 1968). La grille de voisinage est une simplification formelle de la connectivité des neurones, liés entre eux dans la réalité par les synapses.

### VI.3.6. Formalisme mathématique

---

Voyons plus en détail le processus d'apprentissage non supervisé de la topologie des données d'entrée, et les équations qui le régissent.

Les stimuli sont tour à tour tirés aléatoirement, chaque stimulus étant présenté à tous les neurones du réseau (dite couche de Kohonen), parmi lesquels un « vainqueur » est sélectionné. On parle d'*apprentissage compétitif*. Le neurone vainqueur n'est autre que le prototype le plus proche, au sens de la norme définie (Équation. VI.1), du stimulus présenté. Le neurone gagnant  $g$  apprend alors ce stimulus en rapprochant son vecteur de poids de celui-ci. Il est déterminé par l'équation suivante :

$$g(P) = \arg \min_i \|P - w_i\|, \dots, i \in \{1, \dots, N\} \quad (\text{VI.3})$$

Où  $N$  est le nombre total de neurones du réseau.

Les neurones qui lui sont liés, i.e. ses voisins dans la grille de voisinage, apprennent de façon identique. On note  $V_g$  ce voisinage, et l'on obtient l'*équation de mise à jour des poids*, qui suit :

$$\Delta w_i(t) = \begin{cases} \eta(t) [P(t) - w_i(t)] & \text{si } i \in V_g(t) \\ 0 & \text{sinon} \end{cases} \quad (\text{VI.4})$$

Où  $p(t)$  est le stimulus présenté au temps  $t$ ,  $g$  est le neurone gagnant,  $V_g$  l'ensemble de ses voisins,  $w_i$  les poids des neurones mis à jour, et  $\eta$  une quantité appelée le taux d'apprentissage ( $0 \leq \eta \leq 1$ ). Lorsque  $\eta < 1$ , le neurone n'apprend pas totalement le stimulus. L'apprentissage est donc *local*, les neurones se spécialisent individuellement sur des sous-ensembles de stimuli.

Dans l'équation (VI.4), le taux d'apprentissage  $\eta$  et le voisinage  $V_g$  dépendent du temps. L'idée est d'employer au début un grand taux d'apprentissage ainsi qu'un grand voisinage, pour les diminuer ensuite progressivement. De la sorte, les neurones se déplacent dans tout l'espace au début de l'apprentissage, et de façon de plus en plus précise à mesure que l'apprentissage progresse pour que l'algorithme converge. La carte des neurones se déploie rapidement dans la première phase dite d'organisation, puis affine la position de ses centres dans la deuxième phase, dite de convergence.

Nous utilisons une décroissance linéaire pour le taux d'apprentissage et une décroissance exponentielle (par morceaux) pour l'ordre de voisinage, combinaison qui permet une bonne performance de l'algorithme.

$$\eta(t) = \begin{cases} \eta_0 - \left( \frac{\eta_0 - \eta_T}{\tau} \right) t & \text{si } t < \tau \\ \eta_T & \text{sin on} \end{cases} \quad (\text{VI.5})$$

$$V(t) = \begin{cases} E \left[ V_0 \left( \frac{e^{-t/\tau} - e^{-1}}{1 - e^{-1}} \right) \right] + 1 & \text{si } t < \tau \\ 0 & \text{sin on} \end{cases} \quad (\text{VI.6})$$

Où  $\tau$  est le paramètre qui détermine la fin de la phase d'organisation et le début de la phase de convergence.  $\eta_0$  et  $V_0$  sont respectivement le taux d'apprentissage et l'ordre de voisinage initial (à  $t = 0$ ),  $\eta_T$  est le taux d'apprentissage en phase de convergence.  $E[\ ]$  est la fonction partie entière, qui discrétise à des valeurs entières la décroissance exponentielle. Le voisinage d'un neurone  $i$  est défini sur la base de l'ordre de voisinage, selon une métrique de distance propre aux graphes.

L'apprentissage est donc basé sur deux constatations biologiques essentielles, à savoir que, dans le cerveau:

- 1) Chaque cellule nerveuse correspond à un stimulus spécifique;
- 2) Il existe une région d'intense activité autour de la cellule la plus stimulée.

Le modèle d'entraînement de la carte consistera donc à :

- 1) Sélectionner le neurone correspondant le mieux à un signal d'entrée donné;
- 2) Induire dans un voisinage de l'élue une région d'intense activité.

Dans un premier temps, l'apprentissage est *non-supervisé* et consiste à répéter les étapes suivantes:

- 1) Initialiser les poids  $w_{ij}(0)$  avec des valeurs aléatoires ;
- 2) Fixer  $N$ ,  $\eta_0$ ,  $\eta_T$ ,  $V_0$ ,  $\tau$  et  $t_{max}$  ;
- 3)  $t=1$  ;
- 4) Répéter tant que  $t < t_{max}$  :
  - A) Choisir aléatoirement un stimulus  $\mathbf{P}(t)$  parmi l'ensemble des données ;
  - B) Déterminer le neurone gagnant  $g(\mathbf{p})$  (l'équation VI.3) ;
  - C) Mettre à jour les poids (l'équation VI.4) ;
  - D) Mettre à jour  $\eta(t)$  et  $V(t)$  (équations VI.5 et VI.6);
  - E)  $t = t+1$ .

## VI.4. Analyse des données à l'aide de la carte de Kohonen

Dans un premier temps nous avons testés deux types de géométrie du treillis qui définit le type de voisinage de chaque neurone, la géométrie carrée et la géométrie hexagonal. Pour chaque type de géométrie, on commence par une carte de grande dimension (10 x 10), ce qui donne une carte de 100 neurones, et on réduit progressivement la taille de la carte jusqu'à l'obtention de la carte la plus adaptée à notre base de données.

La base de donnée est scindée en trois parties (60% apprentissage, 20 % validation et 20% test), la partie apprentissage pour la construction de la carte topologique et l'identification des différentes classes, tandis que les deux autres parties sont utilisées pour juger la qualité de la classification.

Le réseau est composé de deux couches (Figure VI.4). La couche d'entrée contient un nœud pour chaque variable (ex. la température ou la turbidité de l'échantillon) qui est connecté de façon non-linéaire aux cellules hexagonales ou carrée de la couche de sortie (la carte de Kohonen). Chaque cellule de la carte est connectée aux autres cellules avec une pondération (poids) qui décroît en fonction de la distance relative entre celles-ci. L'ensemble de ces cellules forme la carte de Kohonen (couche de sortie). Les rangées et les colonnes de la carte de Kohonen peuvent être interprétées comme des axes, et la position des différents échantillons comme des coordonnées.

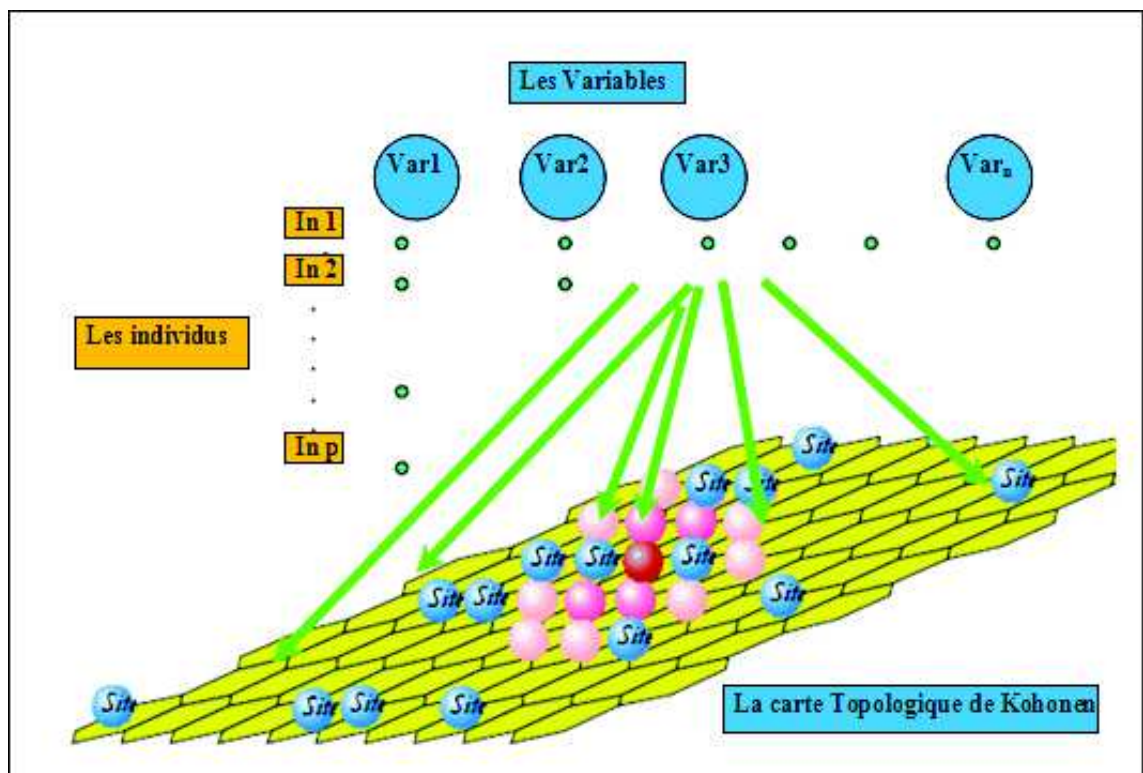


Figure VI.4 Représentation schématique de la carte auto-organisatrice de Kohonen.

La base de données est présentée sous forme de matrice possédant p lignes et n colonnes. Les lignes représentent l'échantillon (les individus), dans notre cas les enregistrements journaliers concernant la qualité de l'eau brute et les colonnes représentent

les variables (température, pH, turbidité, conductivité, ....etc.). Chaque individu peut être considéré comme un vecteur dans un espace à plusieurs dimensions.

Chaque case (hexagone ou carrée) de la carte contient, au départ, des valeurs aléatoires VA (les poids  $w_{ik}$ ) pour chaque variable. La modification des valeurs aléatoires se fait par l'entremise d'un réseau neuronal possédant une capacité d'apprentissage. Les valeurs aléatoires sont modifiées par des ajustements itératifs en fonction des valeurs des échantillons présentés dans la couche d'entrée. Contrairement à un apprentissage supervisé, pour chaque échantillon, la sortie (output) désirée est inconnue. Le but du processus d'apprentissage est que les valeurs aléatoires ressemblent aux valeurs des échantillons. Une fois la phase d'apprentissage terminée, les valeurs aléatoires demeurent inchangées. Chaque échantillon est comparé avec chaque valeur aléatoire. Cette comparaison s'effectue généralement à l'aide de la distance euclidienne.

Un échantillon sera associé à la valeur aléatoire ayant une distance minimale, c'est à dire dont la somme des différences entre les valeurs de chaque variable et celle de la valeur aléatoire est minimale. Une fois la valeur aléatoire sélectionnée, les valeurs de celle-ci sont modifiées de façon à ressembler davantage aux valeurs de l'échantillon. Les valeurs des unités voisines sont également modifiées à l'aide des connexions latérales entre les cellules hexagonales. Cette modification est effectuée à l'intérieur d'un voisinage défini par l'utilisateur, dont la taille est réduite durant la phase d'apprentissage. L'ampleur des modifications est fonction de la distance par rapport à l'unité d'origine. Les valeurs aléatoires des hexagones situés près de la valeur aléatoire sélectionnée sont sujettes à une plus grande modification de leur abondance relative que celles des hexagones distants (Foody, 1999), Ce procédé est répété pour chaque échantillon.

### VI.4.1. Visualisation de la topologie obtenue et phase de classification

---

A la fin de la phase d'apprentissage la carte de Kohonen ainsi former se compose d'un certains nombre de neurones, où chaque neurone correspond à une classe, cette dernière se spécialise à une catégorie de signaux d'entrés (les individus), la somme de tous les signaux correspond à la taille du fichier d'apprentissage.

A la fin de l'apprentissage, le réseau s'est déployé dans l'espace des données pour s'ajuster au mieux à leur distribution. La carte est obtenue par projection dans un espace à deux dimensions des référents associés à chaque neurone. Les individus (435) sont réparties entre les neurones du réseau, et deux neurones voisins sur la carte (distance discrète) représentent deux ensembles d'observations voisines dans l'espace à 7 dimensions (distance euclidienne) que l'on étudie.

Les différents paramètres de l'apprentissage sont fixés préalablement:

- Taille de la carte, en précisant le nombre de neurones et la structure de la carte ;
- Voisinage, en indiquant le type et le seuil (rayon maximum du voisinage) ;
- Le nombre maximum d'itérations

En sortie du modèle, nous avons récupère les fichiers dont l'étude nous a permis de vérifier la qualité de l'apprentissage :

- Fichier des cardinalités des neurones : chaque neurone a capté un certain nombre d'individus contenus dans la base d'apprentissage ;

- Fichier des référents ( $w$ ) des neurones après apprentissage : chaque référent est un vecteur à 7 dimensions qui résume l'ensemble des variables captés par le neurone auquel il est affecté ;
- Chaque neurone est caractérisé par son référent est définie donc, par un taux (pourcentage) de classification (nombre d'individus classés par rapport au nombre d'individus totale).

Une dernière vérification consiste à présenter au réseau de neurones ainsi construit (la carte de Kohonen) des données dites de « validation » qui appartenaient à *la base de données totale* mais qui ne faisaient pas partie de sa base d'apprentissage. Pour ce faire nous avons préparé deux autres bases de données que nous avons appelées données de « validation » et de « test ».

La meilleure carte topologique est celle qui correspond à des neurones capables de classer les données, avec le même taux (pourcentage), pour les trois bases de données (apprentissage, validation et test), pour cela on calcule le coefficient de corrélation entre la partie apprentissage et validation et entre l'apprentissage et test.

On commence par une topologie de (10x10), et on calcule les performances de la classification et on change progressivement le nombre de neurones jusqu'à l'obtention de la meilleure topologie.

Les figures (VI.5) et (VI.6) présentent la cardinalité de chacun des neurones du réseau de la carte carrée et hexagonal. On remarque d'après les deux figures que certains neurones n'ont captés aucune observation ce qui explique que les cartes sont surdimensionnées, le coefficient de corrélation ne dépasse pas 0.62 en validation tandis qu'il est de l'ordre de 0.65 en phase test.

Dans le but d'obtenir des résultats plus performants nous avons testés plusieurs topologies, les résultats obtenus sont présentés dans le tableau (VI.1).



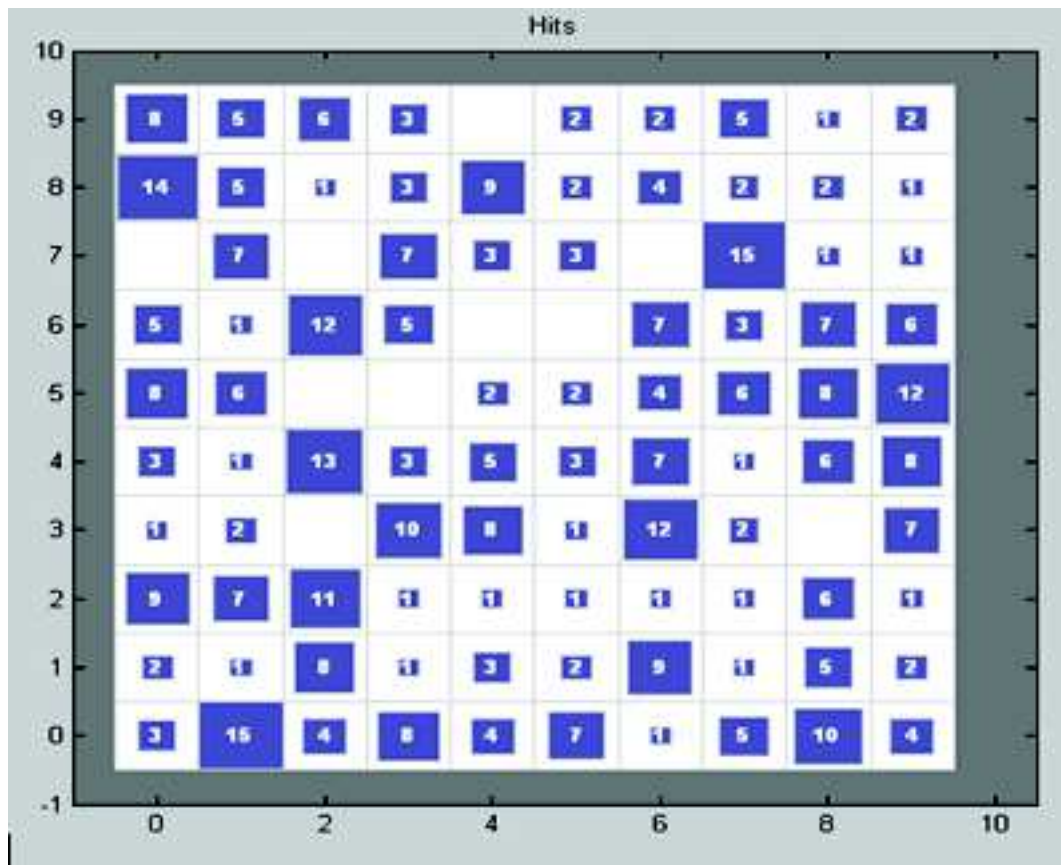


Figure VI.5 Carte topologique carré.

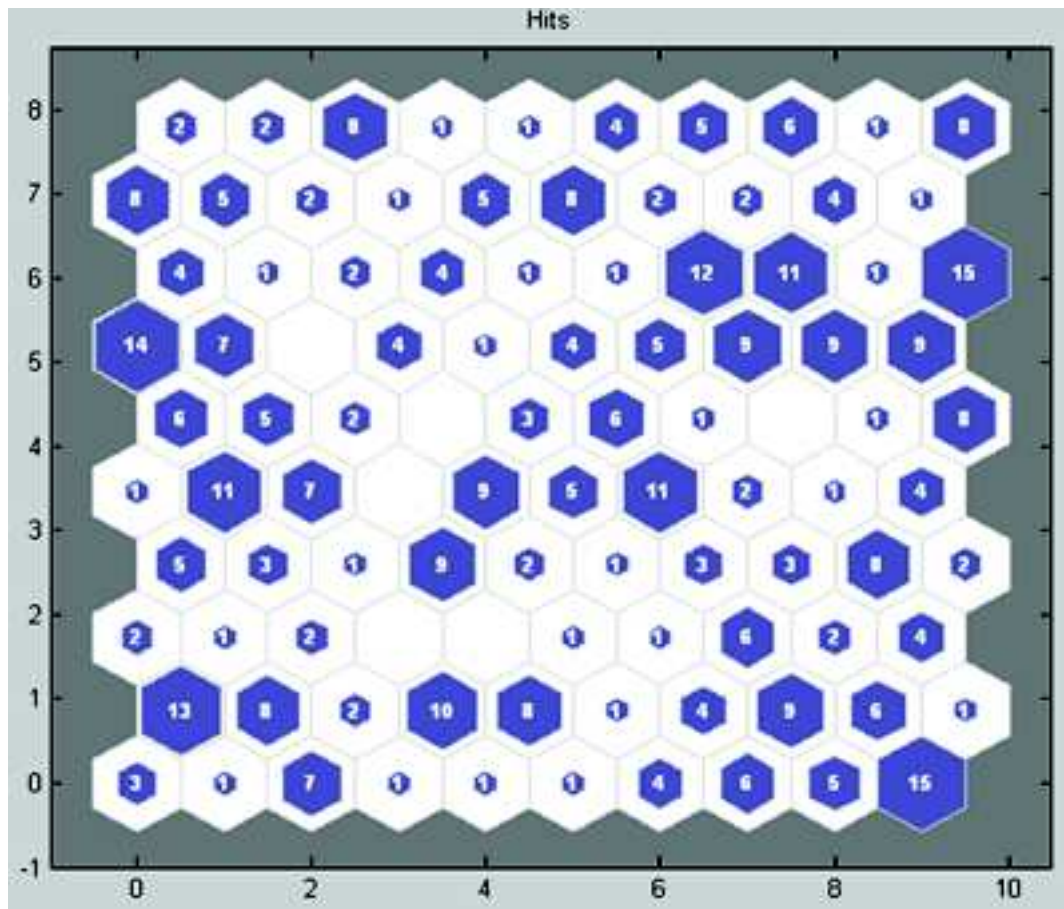


Figure VI.6 Carte topologique hexagonal.

Tableau VI.1 Résultats de la classification pour différentes topologies.

Dimension	Coefficient de corrélation R			
	Topologie carrée		Topologie hexagonale	
	Validation	Test	Validation	Test
10 x10	0.52	0.57	0.62	0.65
9x9	0.57	0.57	0.58	0.66
8x8	0.66	0.63	0.66	0.67
7x7	0.61	0.64	0.66	0.70
6x6	0.64	0.68	0.66	0.46
5x5	0.60	0.53	0.76	0.46
4x4	0.72	0.66	0.70	0.72
3x3	0.93	0.90	0.87	0.83
2x3	0.97	0.92	0.96	0.96

D'après le tableau (VI.1), on constate que plus la taille de la carte diminue, plus les performances de la classification augmentent, et les meilleures performances sont obtenues pour une carte hexagonale de six neurones (3x2), ce qui correspond à six classes. Cela peut être expliqué par le fait que l'amplitude de variation de la qualité de l'eau brute au niveau de la station de Boudouaou est faible et peut être limité à une variation saisonnière.

La figure (VI.7) présente la cardinalité de chacun des neurones du réseau de la carte hexagonale finale retenue. On remarque que le neurone le plus important capte 134 individus tandis que le neurone le moins important capte 34 neurones, le coefficient de corrélation dépasse 0.96 en validation et aussi en phase de test.

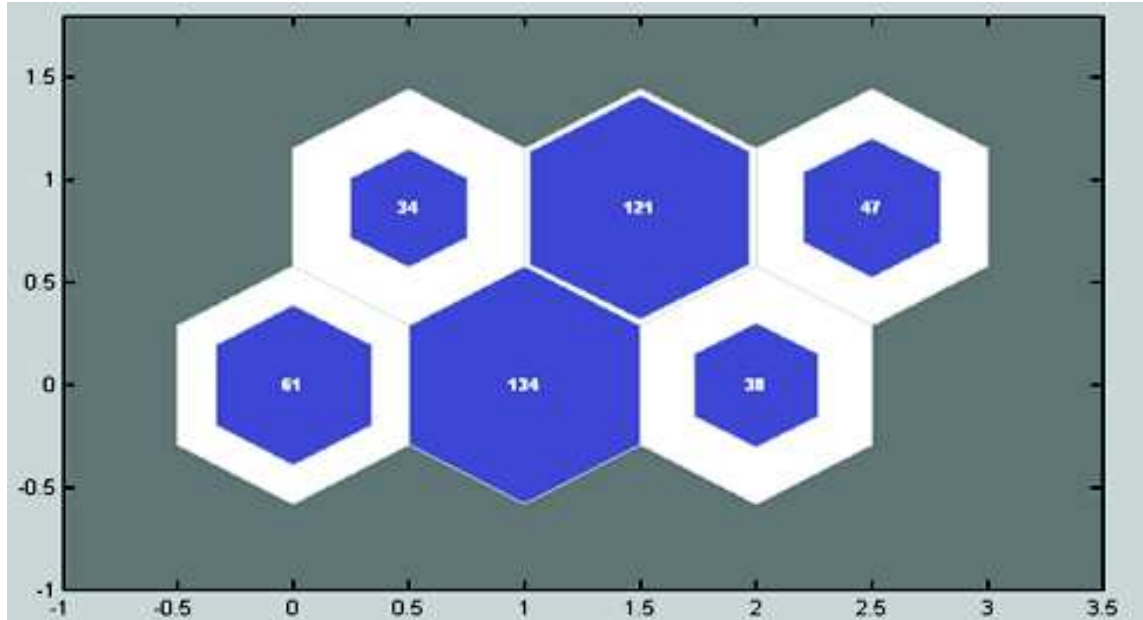


Figure VI.7 Carte topologique hexagonale retenue.

Le tableau (VI.2) présente le détail de la classification obtenu, la cardinalité de chacun des neurones du réseau de la carte hexagonale ainsi que le taux de classification par rapport au nombre total des individus successivement.

Tableau VI.2. Répartition des individus par neurones de la carte.

Apprentissage		Validation		Test	
Taux de classification	Nbr d'individus	Taux de classification	Nbr d'individus	Taux de classification	Nbr d'individus
14.02	61.00	11.03	16.00	9.65	14.00
30.80	134.00	27.58	40.00	32.41	47.00
8.73	38.00	12.41	18.00	11.03	16.00
7.81	34.00	9.65	14.00	8.27	12.00
27.81	121.00	27.58	40.00	31.72	46.00
10.80	47.00	11.72	17.00	6.89	10.00
<b>100.00</b>	<b>435.00</b>	<b>100.00</b>	<b>145.00</b>	<b>100.00</b>	<b>145.00</b>

Le tableau (VI.3) présente le Fichier des référents ( $w$ ) des neurones après apprentissage chaque référent est un vecteur à 7 dimensions qui résume l'ensemble des variables captées par le neurone auquel il est affecté.

Tableau VI.3. Tableau des référents (poids).

N° neurone	Poids $w$						
01	1.45	1.94	-1.07	1.21	-1.10	-0.54	-0.39
02	-0.54	0.01	-0.69	-0.04	-0.64	-0.75	-0.44
03	-0.22	-0.82	0.59	-1.87	0.16	-0.58	-0.02
04	1.47	0.61	-0.17	-0.11	0.96	1.05	1.49
05	-0.56	-0.49	0.53	0.42	0.35	1.17	0.07
06	0.39	-0.84	0.99	-0.51	1.26	-0.65	0.43

## VI.5. Conclusion

Le but de ce chapitre était de classer les données relatives à la station de traitement des eaux de Boudouaou. l'objectif poursuivi était de comprendre la répartition des dites données dans un espace à deux dimensions. La technique d'analyse du réseau de neurones de Kohonen, permet de regrouper un grand nombre de vecteurs d'entrée en familles proches (classes). sans poser aucune connaissance *a priori* sur la composition de ces familles. Cette analyse non supervisée laisse une grande souplesse à l'expérimentateur dans son interprétation des résultats.

---

# Conclusion Générale

Les stations de traitement des eaux se heurtent très souvent à la difficulté de disposer d'un modèle fiable et capable d'exprimer la relation entre toutes les variables descriptives caractérisant l'eau brute et la dose de coagulant injecter durant la phase de clarification. Dans la littérature, des approches de modélisation par des modèles de représentation linéaire et non linéaire ont été proposées. L'hypothèse de base est généralement qu'il s'agit d'une relation entre certaines variables d'entrées (variables explicatives), dont l'étendu varie considérablement d'une variable à une autre, et variable de sortie (variable à expliquer) déterminée dans la plupart des cas par des essais de laboratoire dits Jar-Test. On assiste actuellement au développement d'une technique basée sur l'utilisation de l'intelligence artificiel pour le développement des modèles de procédés complexes.

Le travail présenté dans cette thèse a montré le bien fondé d'une approche basée sur les concepts de la théorie des réseaux de neurones et de la logique floue pour aborder le problème complexe que constitue la détermination en ligne de la dose de coagulant au niveau des stations de traitement des eaux potables. En effet, le traitement des eaux potable est réalisé dans la plupart des cas par un système physico-chimique classique, utilisant la coagulation-floculation comme procédé de base. La gestion du procédé est la plus simple, et les coûts d'installation et d'opération sont les moins chers. Cependant, la qualité d'eau produite et le contrôle d'un système stable sont très sensibles aux variations de la qualité de l'eau brute à traiter, ce dernier étant caractérisé par différentes variables à savoir la température, le pH, la turbidité, etc.

La combinaison (réseau de neurones /logique floue), ou tout simplement système *neuroflou* est une combinaison de la logique floue et des réseaux de neurones qui tire profit des deux approches. Plusieurs architectures neuroflous ont été proposées dans la littérature suivant le type de règles floues qu'elles intègrent (Mamdani ou Sugeno). La puissance de ces structures réside dans la possibilité d'incorporer une base de connaissances, de traiter les données imprécises et vagues par logique floue et en même temps d'introduire l'apprentissage via le réseau de neurones.

Après une revue bibliographique sur les principes et les différentes méthodes de traitement des eaux destinées à la consommation humains, nous avons tout d'abord étudié le phénomène de coagulation-floculation qui représente l'objet de notre travail de recherche. Le but était de montrer les avantages de cette technique qui est parfaitement adaptée à la majorité des stations de traitement des eaux et de préciser les problèmes rencontrés pour la détermination de la dose de coagulant à injecter. En particulier, nous avons expliqué le principe de l'essai Jar Test, méthode utilisée dans la totalité des stations de traitement des eaux.

Dans le troisième chapitre nous avons abordé l'aspect modélisation linéaire de la dose de coagulant, nous avons élaboré une base de données provenant de la station de traitement des eaux de Boudouaou alimentant la ville d'Alger en eau potable, pour cela six variables descriptives ont été retenues, à savoir: la température (TE), le pH, la turbidité (TU), la

conductivité électrique de l'eau (CON), l'oxygène dissous (OX) et les ultraviolets (UV<sub>254</sub>), parallèlement la dose de coagulant est déterminée par les essais Jar-Test effectués au laboratoire. Une fois la base de données élaborée, on s'est posé la question suivante : Quelles sont les variables clés parmi les variables descriptives étudiées qui sont déterminantes pour expliquer la variation de la dose de coagulant?

Pour répondre à cette question, l'Analyse en Composantes Principales (ACP) a été appliquée, elle consiste à chercher un sous-espace de dimension inférieure à celle de l'espace initial où l'on pourra projeter les données en perdant un minimum d'informations. L'ACP effectue donc une *réduction de dimension*. Pour atteindre ce résultat, l'ACP cherche les axes autour desquels l'inertie de l'ensemble des vecteurs est la plus faible, *i.e.* les axes le long desquels la variance est maximale. Les résultats obtenus ont montrés l'absence totale de toutes variables prépondérantes, et on remarque qu'il est indispensable de tenir compte des cinq premières composantes principales, pour avoir 90 % de la variance totale; ce qui nous oblige d'intégrer toutes les variables descriptives et aucune variable ne peut être exclu.

Une fois les variables descriptives choisies, nous avons procédé à la construction du premier type de modèle utilisé dans le cadre de notre travail de recherche : le modèle à base de régression linéaire multiple (RLM), cette dernière est la méthode la plus simple de modélisation. Elle consiste à rechercher une équation linéaire par rapport à ses paramètres, reliant la variable à modéliser (*la dose de coagulant*) au vecteur d'entrées (*les variables descriptives retenues*). Ces entrées peuvent être soit des variables primaires (*brutes*) sans transformation, soit des variables normalisées, et c'est ce que nous avons fait dans le cadre de ce travail, en effet, les variables d'entrées ont été centrées et réduites. Les résultats obtenus ont confirmés la forte complexité et la non linéarité du procédé de coagulation, étant donnée que les critères de validation utilisés ont été très médiocres, le coefficient de corrélation n'a pas dépassé les 0.57, tandis que le RMSE avoisine le 6.25 en phase de test. Malgré que le modèle à six entrées présente des performances meilleures que les autres modèles, les résultats obtenus par la régression linéaire multiple sont loin d'être accepté, et il a été exclu la modélisation de la dose de coagulant par des techniques de régression multiple.

Depuis plusieurs années, les réseaux de neurones artificiels, se sont montrés très efficaces dans le domaine de la modélisation des processus complexes et non linéaires. Ces systèmes sont composés d'un ensemble structuré d'unités de traitement, appelées *neurones*, qui fonctionnent en parallèle et qui sont fortement interconnectées. Ces neurones réalisent chacun une fonction non linéaire de leurs entrées, qui est déterminée par un ensemble de paramètres dont les valeurs sont établies à la suite d'un apprentissage par l'exemple. Nous nous sommes intéressés dans le chapitre IV à cette famille de modèles mathématiques, qui ont aujourd'hui atteint un niveau de maturité qui en fait un outil de choix pour aborder des problèmes de modélisation non linéaire délicats : les fondements théoriques sont solides (propriété d'approximation universelle et de parcimonie), et leur mise en œuvre ne soulève pas de difficultés particulières (existence d'algorithmes pour l'apprentissage de modèles). Les approches basées sur l'utilisation des réseaux de neurones artificiels sont en effet particulièrement bien adaptées aux problèmes qui s'y posent, grâce à leur robustesse, leur capacité à traiter des données de différentes natures et surtout leurs capacités d'adaptation et d'apprentissage. Après avoir posé la problématique, on a présenté les trois familles de réseaux de neurones les plus classiques, les perceptrons multicouches (MLPNN) et les réseaux à fonction radiale de base (RBFNN) ainsi que les réseaux dites à fonction de régression générale (GRNN). Les MLPNN sont des réseaux basés sur des produits scalaires, qui réalisent généralement des approximations plutôt



---

globales de fonctions, c'est à dire sur de grands intervalles, tandis que les RBFNN et les GRNN sont basés sur les distances, et produisent des approximations plutôt locales.

Les résultats obtenus par ce type de modèles ont été très encourageants, les modèles neuronaux font généralement moins d'erreurs que la régression linéaire; peuvent être marginalement pour les ensembles d'apprentissage mais la différence est plus prononcée dans les ensembles de validation. Ceci donne une certaine indication de l'utilité des modèles neuronaux par rapport aux techniques linéaires simplistes. Les ensembles d'apprentissage présentent comparativement un comportement moins démarqué. Il semble que la plupart des architectures apprennent avec une performance comparable pour la base d'apprentissage. Cependant il est important de signaler que le réseau GRNN présente de meilleurs résultats que les deux autres architectures à savoir les MLPNN et les RBFNN.

Dans le souci d'améliorer encore les performances des résultats obtenus, nous avons consacré la totalité du chapitre V à un autre type de modèle : les modèles à base de systèmes d'inférence floue (SIF). Les outils fournis par la logique floue permettent en effet, une modélisation des phénomènes pouvant en un certain sens s'approcher du raisonnement humain. Nous avons présenté dans ce chapitre les concepts de base des SIF, le concept de variable linguistique. Nous avons constaté que les SIF peuvent en effet raisonner avec l'information imprécise. Après une description détaillée des notions de la logique floue, nous nous sommes intéressés à un type particulier de modèle qui est le modèle neuroflou appelé ANFIS (*Adaptive-Neuro-based Fuzzy Inference System*). Les systèmes neuro-flous sont nés de l'association des réseaux de neurones et de la logique floue, de manière à tirer profits des avantages de chacune de ces deux techniques. Cependant, la logique floue permet une spécification rapide des tâches à accomplir à partir de la connaissance symbolique disponible. Le réglage précis du système obtenu et l'optimisation de ses différents paramètres reste néanmoins beaucoup plus difficile dans de nombreux cas. Par contre les modèles les plus courants de RNA, n'autorisent pas l'incorporation de connaissance à priori mais permettent de régler par apprentissage le comportement précis du système.

ANFIS représente un système à inférence floue mis en application dans le cadre des réseaux adaptatifs. Il utilise la procédure d'apprentissage hybride. Cette architecture affine les règles floues obtenues par des experts humains pour décrire le comportement d'entrée-sortie d'un système complexe. Ce modèle donne de très bons résultats en approximation non linéaire. Un système ANFIS réalise une approximation linéaire de la variable de sortie en décomposant l'espace des entrées en différents espaces flous. Exception faite de la couche d'entrée, un système ANFIS est composé de 5 couches, et chacune peut comprendre différents nœuds.

Nous avons démontré, que partant d'un certain nombre de sous ensembles flous et avec des fonctions d'appartenance de type gaussienne, triangulaire ou trapézoïdales, le réseau ANFIS comporte plusieurs paramètres devant être optimisés (certains inhérents aux fonctions gaussiennes et d'autres à la linéarisation des sorties des règles floues). À cette fin, une base d'apprentissage est nécessaire. L'apprentissage consiste en la correction des paramètres (prémises et conséquents) du réseau afin de généraliser une fonction de transfert (inconnue *à priori*) entre les entrées et la sortie du réseau. Celle-ci est constituée d'un ensemble de couples "entrées /sortie" connu (relevé de données). Le déploiement des algorithmes d'apprentissage sur cette base de données permet de construire une fonction d'approximation (de prédiction dans notre cas) de la sortie désirée à partir des nouveaux vecteurs d'entrées. En général, des algorithmes hybrides sont utilisés. Plus particulièrement, une combinaison de l'algorithme de descente du gradient et de l'algorithme LSE (Least

Square Estimate- estimation des moindres carrés) permet de réduire la complexité de l'apprentissage et d'en améliorer l'efficacité. Les points forts de l'ANFIS consistent dans le mécanisme d'inférence distributif et l'algorithme d'apprentissage adaptif. Par contre les règles sont non interprétables.

A travers la présente étude nous avons démontré que l'optimisation du nombre de règles floues à utiliser par la modèle ANFIS est l'étape cruciale dans l'application de ce type de modèle, en testant deux techniques différentes à savoir, la technique dite *GIRD PARTITION* basée sur le principe de la répartition des variables d'entrées en plusieurs sous ensembles flous fixés à priori, dans ce cas le nombre de règle floue dépend directement de cette répartition, et la deuxième technique dite du groupage *SOUSTRUCTIVE*, basée sur le principe d'une optimisation du nombre de règles floues sans aucune condition imposée au préalable. Les résultats obtenus par la méthode du *SOUSTRUCTIVE* sont meilleurs que ceux obtenus par la méthode *GIRD PARTITION*. Nous avons obtenus des résultats très satisfaisants qui témoignent de la faisabilité et de la validité de notre démarche.

Enfin, en comparant les résultats obtenus par le modèle ANFIS avec ceux obtenus par les réseaux de neurones et la régression linéaire multiple, on peut conclure à ce stade de recherche qu'il exclu d'aborder le problème de la détermination de la dose de coagulant par des modèles linéaires.

Ainsi, au terme de ce travail, la faisabilité de la modélisation de la dose de coagulant par modèle neuroflou a été démontrée. Ce modèle a été élaboré uniquement à partir de données obtenues lors d'expériences Jar Test. On peut étendre cette modélisation par exemple au cas d'une automatisation des mesures des différentes variables par l'installation des capteurs en lignes.

---

# Références Bibliographiques

- Abraham A., (2005). Adaptation of Fuzzy Inference System Using Neural Learning, *Stud Fuzzy* 181, pp 53–83 (2005)
- Adgar A., Cox C.S., Daniel P.R., Billington A.J., Lowdon A., (1995). Experiences in the application of the artificial neural networks to water treatment plant management, in *Proceedings of the International COMDEM'95*, Vol. 1, pp 33-38, Canada.
- Adgar A., Cox C.S., Böhme T.J. (2000). Performance Improvements at surface water treatment works using ANN-based automation schemes. *Transactions of the Institution of Chemical Engineers*, Vol. 78, Part A, October 2000, pp 1026-1039.
- Agteberg F.P., (1974). *Geomathematics*, Elsevier, Amsterdam.
- Aguilera, P.A., Frenich, A.G., Torres, J.A., Castro, H., Vidal, J.L.M. et Canton, M. (2001). Application of the Kohonen neural network in coastal management: methodological development for the assessment and prediction of water quality. *Water Research*, 35: pp 4053-4062.
- Amari, S. A., (1980). Topographic organization of nerve fields. *Bulletin of mathematical biology*, Vol. 42, pp 339-364.
- Amirtharajah A., Mills K.J., (1982). Rapid-mix design for mechanism of alum coagulation. *Journal American Water Works Association*, 74, pp 210–216.
- Amirtharajah A., et C. R. O'melia, (1990). Coagulation processes: Destabilization, mixing, and flocculation. *Water quality and treatment*, F. W. Pontius, ed., McGraw-Hill, New York, pp 269– 365.
- Aurette, D., Lek, S., Giraudel, J.L., Berrebi, P. (1999). Microsatellites and artificial neural networks: tools for the discrimination between natural and hatchery brown trout (*Salmo trutta*, L.) in Atlantic populations. *Ecological Modelling*, 120: pp 313–324.
- Babuska, R., (1998). *Fuzzy Modeling for Control*. Kluwer Academic Publishers, Mass., USA.
- baghli L., (1999). Contribution à la commande de la machine asynchrone, utilisation de la logique floue, des réseaux de neurones et des algorithmes génétiques. Thèse de doctorat Université Henri Poincaré, Nancy-I. 215p.
- Barron A., (1993). Universal approximation bounds for superposition of a sigmoidal function, *IEEE Transactions on Information Theory*, 39, pp 930-945.
- Baxter C.W., Stanley S.J., Zhang Q., Smith D.W., (2002). Developing artificial neural network process models of water treatment process: a guide for utilities. *Can. J. Civ. Eng.* 1(3): pp 201-211.
- Baxter, C.W., Stanley, S.J., Zhang, Q., (1999). Development of a full-scale artificial neural network model for the removal of natural organic matter by enhanced coagulation. *J Water SRT-Aqua* 48 (4), pp 129–136.

- Baxter C.W., Zhang Q., Stanley S.J., Shariff R., Tupas R.R.T., L. Stark, (2001a). Drinking water quality and treatment: the use of artificial neural networks. *Can. J. Civ. Eng.* 28(Suppl. S1): pp 26–35.
- Baxter C.W., Tupas R.R.T., Zhang Q., Shariff R., Stanley S.J., Coffey B.M., K.G. Graff, (2001b). Artificial intelligence systems for water treatment plant optimization. American Water works Association Research Foundation and American Water works Association, Denver, Colo. 141 p.
- Baxter C.W., Shariff R., Stanley S.J., Smith D.W., Zhang Q., E.D. Saumer, (2002). Model-based advanced process control of coagulation. *Water Sci. Technol.*, 45(4–5): pp 9–17.
- Bazer-Bachi A., Puech-Coste E., Ben aim R., Probst J.L. (1990) Mathematical modelling of optimum coagulant dose in water treatment plant. *Rev. Sci. Eau*, 1990: 3, pp 377-397.
- Bernazeau F., Hubele C., Brodard E., Dalga N., (1990) Utilisation du streaming current detector pour l'asservissement de la clarification d'une eau de surface, *Water Supply*, Vol. 9, pp 39-47, Amsterdam.
- Bishop C., (1995). *Neural networks for pattern recognition*, Oxford University Press.
- Blasco, J.A., Fueyo, N., Dopazo, C., Chen, J.Y. (2000). A self-organizing-map approach to chemistry representation in combustion applications. *Combustion Theory Modelling*, 4: pp 61–76.
- Brown M., Harris C.J., (1994). *Neurofuzzy Adaptive Modelling and Control*. Prentice Hall, Hemel.
- Brodard E., Bordet J., Bernazeau F., Mallevalle J., Fiessinger F., (1989). Modélisation stochastique d'une usine de traitement de l'eau potable. 2ème Rencontres Internationales Eau et Technologies Avancées, Montpellier.
- Broomhead D.S., Lowe D., (1988). Multi variable functional interpolation and adaptive networks, *Complex Systems*, Vol. 2, pp 321-355.
- Bohme T.J., Cox C., S (1999) Performance assessment of a neuro self-tuning PI controller to be used at a Water Treatment Plant. *Proceedings of the American Control Conference*, 2-4 June. pp 3216-3220. San Diego.
- Boillereaux L., Flaus J.M., (2003). *Comande et supervision. Les procédés agroalimentaires 2*. Lavoisier ISBN 2-7462-0755-9.
- Bombaugh K. J., Dark W. A., Costello L. A., (1967). Application of the Streaming Current Detector to Control Problems, in *Proceedings of 13th National ISA Analysis Instrument Symposium*, Houston, USA.
- Boukezzoula R., (2000). *Commande floue d'une classe de systèmes non linéaires: application au problème de suivi de trajectoire*. Thèse de doctorat université de Savoie. 245p.
- Boucheron S., (1992). *Théorie de l'apprentissage*. Hermès, Paris.
- Buckley J., (1992). "Universal Fuzzy Controllers", *Automatica*, Vol. 28, No. 6, pp.1245-1248.
- Buckley J.J., Hayashi Y., (1993). Hybrid neural nets can be fuzzy controllers and fuzzy expert systems, *Fuzzy Sets and Systems*, 60: pp 135–142.

- 
- Buniet L., (1997) Traitement automatique de la parole en milieu bruité : étude de modèles connexionnistes statiques et dynamiques. Thèse de doctorat .Université Henri Poincaré - Nancy 1.France. 350P.
- Cardot C., (1999). Les traitements de l'eau. Procédés physico-chimiques et biologiques. Ellipses Edition Marketing S.A.
- castalleno G., (2000). A Neurofuzzy Methodology for Predictive Modeling. PhD Thesis in Computer Science University of Bari Faculty of Science Department of Computer Science. 224p
- Chiu S.L., (1994). Fuzzy model identification based on cluster estimation. Journal of Intelligent and Fuzzy Systems, 2, pp 267-278.
- Chen, C.L., and Hou, P.L. (2006). Fuzzy model identification and control system design for coagulation chemical dosing of potable water. Water Science & Technology: Water Supply Vol 6 No 3, pp 97–104. IWA Publishing
- Ciofalo C., (2005).Segmentation de formes guidée par des modèles en neuro-imagerie Intégration de la commande floue dans une méthode de segmentation par ensembles de niveau. Thèse de doctorat Université de Rennes 1. 210p.
- C.I.D.F. (1995).Centre International De Formation, Principes généraux de traitement des eaux, Lyonnaise des Eaux.
- Cox C.S, Rietveld L., Adgar A., Van der Helm A., (2003). A new modelling and control system simulation environment for the rapid design of potable water systems. ADCHEM, Hong Kong.
- Cox C.S., Daniel P., Adgar A., Lowdon A., (1996) Chemical dosing philosophies for a water treatment plant: results of some pilot plant experimentation, in *Proceedings of the IEE Conference Publication*, Vol. 427 (2), pp 1052-1057.
- Cox C., Graham J., (1994). Steps toward automatic clarification control. Computing and Control Division Coloquium on Adv in Control the Process Ind.IEE Coloquium Digest, 81, 6/1-6/4.
- Coulibaly P., Anctil F., Bobee B. (1999). Prévion hydrologique par réseaux de neurones artificiels : Etat de l'art. Journal of civil engineering, 26; pp 293-301.
- Critchley R.F., Smith E.O., P. Pettit, (1990). Automatic coagulation control at water treatment plants in the North-West region of England. Water and Environment Journal 4(6), pp 535-543.
- Cybenko G., (1989) Approximation by Superposition's of a Sigmoidal Function, Mathematics of control, signals and systems, 2, pp 303-314.
- Degrement. (2005). Mémento technique de l'eau : Tome I, Lavoisier – Technique et documentation, Paris.
- Demotier S., Denoeux T., Schön W., Odeh K., (2003). A new approach to assess risk in water treatment using the belief function framework. IEEE 0-7803-7952-7/03, pp 1792-1797.
- Dempsey B.A., Sheu H., Tanzeer Ahmed T.M., Mentink J., (1985).Polyaluminum chloride and alum coagulation of clay-fulvic acid suspensions. Journal AWWA. 77 (3), pp 74– 80.
-

- Dentel S.K, Thomas A.V, Kingery K.M., (1989) Evaluation of the streaming current detector- 1. Use in Jar tests. *Water Research* **23**(4), pp 413-421
- Dentel S.K., (1995) Use of the streaming current detector in coagulation monitoring and control. *Journal of Water Science Research and Technology - Aqua* **44**(2), pp 70-79.
- Desjardins (1997). *Le traitement des eaux* 2ème édition.
- Driankov D., Hellendoorn H., Reinfrank M., (1993). *An introduction to Fuzzy Control*, Springer-Verlag.
- Dreyfus G., Samuelides M., Martinez J., Gordon M., Badran F., Thiria S., Herault L. (2004). *Réseaux de neurones - Méthodologies et applications*. Editions Eyrolles 200 p.
- Didday E., Simon J. C. (1970). Clustering Analysis, *in Digital Pattern Recognition*, K. S. Fu, Springer Verlag.
- Dubois D., Prade H., (1980). *Fuzzy Sets and Systems – Theory and Applications*. Academic Press, New York.
- Dubois D., Prade H., (1988). *Théorie des possibilités : application à la représentation des connaissances en informatique*, Masson, Paris.
- Dubois D., Prade H., (1991). Fuzzy sets in approximate reasoning, part 1: Inference with possibility distributions. *Fuzzy sets and Systems*, vol. 40, pp 143-202.
- Duby C., Robin S., (2006). *Analyse en composante principales*. Institut National Agronomique Paris – Grignon. 54 P
- Evans J and Enoch C (1998). Investigation of advanced control methods to coagulant dosing at a WTW.UK Water Industry Research Limited/Sensors for Water Interest Group (UKWIR/SWIG) Workshop –Knowledge based systems, National Motorcycle Museum, Birmingham, UK.
- Edwards G.A., et Amirtharajah A., (1985). Removing color caused by humic acids. *Journal American Water Works Association*, **77**(3), pp 50–57.
- Edzwald J.K., et J.E. Tobiason, (1999). Enhanced Coagulation: US requirements and a broad review. *Water Science and Technology*, **40**, pp 67–70.
- Ellis G., Collins A.G, Ge Xi., Ford C.R, (1991). Chemical dosing of small water utilities using regression analysis. *Journal of Environmental Engineering*: **117**(3), pp 308-318.
- faucher G., (2006). *Les plans d'expériences pour le réglage de commandes à base de logique floue*. Thèse de doctorat de l'institut national polytechnique de Toulouse. 198p.
- Footy, G.M., 1999. Application of self-organizing feature map neural network in community data analysis. *Ecological Modelling*, **120**: pp 97-107.
- Footy, G.M. et Cutler, M.E.J. 2003. Tree biodiversity in protected and logged Bornean tropical rain forests and its measurement by satellite remote sensing. *Journal of Biogeography*, **30**: pp 1046-1053.
- Fukushima K. (1975). Cognitron:a self-organizing multi-layered neural network, *Biological Cybernetics* vol. 20, pp 121-136.
- Funahashi K. I., (1989). On the approximate realization of continuous mappings by neural networks, *Neural Networks* **2**, pp 183.



- Gagnon C., Grandjean B.P.A., Thibault J., (1997). Modelling of coagulant dosage in a water treatment plant, *Artificial Intelligence in Engineering*, Vol. 11, pp 401- 404.
- Galichet S., Foulloy L., (1994). Fuzzy control with Non Precise Inputs, *Proc. of the 5th Int. Conf. on Information Processing and Management and Uncertainty in Knowledge Based Systems (IPMU'94)*, Paris, France, July 94, pp 84-89.
- Gaudier F., (1999). Modélisation par réseaux de neurones, application a la gestion du combustible dans un réacteur. Thèse de doctorat. École normale supérieure de Cachan. 130 p.
- Gavin J. B., Dandy, G.C., Maier, H.R., (2005a). Input determination for neural network models in water resources applications. Part 1. Background and methodology. *Journal of Hydrology* 301 (2005), pp 75–92.
- Gavin J. B., Dandy, G.C., Maier, H.R., (2005b). Input determination for neural network models in water resources applications. Part 2. Case study: forecasting salinity in a river. *Journal of Hydrology* 301 (2005), pp 93–107.
- Girou A., Franceschi M., Puech-Costes E., Humbert L., (1992) Modélisation des phénomènes de coagulation et étude de la morphologie des floes: optimisation du taux de coagulant. *Récents Prog. Génie Procédés*:6(20), pp 373-385.
- Glorennec, Y. (1999). Algorithmes d'apprentissage pour systèmes d'inférence floue, Hermès Science Publications, France.
- Godart H.,(2008). Traitements des eaux de distribution. *Traité Construction*, volume C 5 200. Techniques de l'Ingénieur.
- Grisales palacio V.H, (2007). Modélisation et commande floues de type Takagi\_Sugeno appliquées à un bioprocédé de traitement des eaux usées. Thèse de Doctorat de l'Université Paul Sabatier - Toulouse III et l'Université de los Andes, Colombie. 230p.
- Grossberg S., (1976). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detector *Biological Cybernetics*, vol. 23, pp 121-134.
- Guillaume s., (2001). Induction de règles floues interprétables. Thèse de doctorat de l'institut national des sciences appliquées de Toulouse. 198p.
- Hardy R.L., (1971). Multiquadratic equations of topography and other irregular surfaces, *J.Geophys. Res*, 76: pp 1905-1915.
- Hartman E. J., Kowalski J. M., (1990) "*Layered Neural Networks with Gaussian Hidden Units as Universal Approximations.*" *Neural Computation*, Vol. 2, pp.210
- Hayashi Y., Buckley J.J., (1994). Approximations between Fuzzy Expert Systems and Neural Networks, *International Journal of Approximate Reasoning*, Vol 10, pp 63–73.
- Haykin S., (1994). *Neural networks: a comprehensive foundation* Macmillan, New York.
- Heddam S., et Dechemi N., (2008). Modeling of optimal coagulant dose using artificial neural network, application to water treatment plant of Boudouaou (Algeria). *EJSR* Vol 20, issue 3 pp 652–663.
- Heddam S., Bermad A., Dechemi N., (2010). Modélisation de la dose de coagulant par les Réseaux de Neurones Artificiels (ANN) application à la station de traitement

- des eaux potable de Boudouaou (Algérie). 1<sup>er</sup> Colloque international de l'eau CIEAU 2010. Université Badji Mokhtar.
- Heddami S., Bermad A., Dechemi N., (2011a). Modelling of Coagulant Dosage by Neuro Fuzzy Technique Application to Boudouaou Drinking Water Treatment Plant (Algeria). 4<sup>ème</sup> colloque international « Ressources en eau et Développement Durable CIRED » février 2011/Alger.
- Heddami, S., Bermad A., Dechemi N. (2011b). ANFIS-based modelling for coagulant dosage in drinking water treatment plant: a case study. *Environmental Monitoring and Assessment (EMAS)*. Doi: 10.1007/s10661-011-2091-x.
- Heddami, S., Bermad A., Dechemi N. (2011c). Applications of Radial Basis Function and Generalized Regression Neural Networks for Modelling of Coagulant Dosage in a Drinking Water Treatment: A Comparative Study. *ASCE Journal of Environmental Engineering*. Doi: 10.1061/(ASCE)EE.1943-7870.0000435
- Heddami, S., Bermad A., Dechemi N. (2011d). Modélisation de la dose de coagulant par les systèmes à base d'inférence floue (ANFIS) application à la station de traitement des eaux de Boudouaou (Algérie). (*RSE*)Revue des sciences de l'eau. ACCEPTED MANUSCRIPT. Accepter pour publication Aout 2011.
- Hernandez De León H.R., Supervision et diagnostic des procédés de production d'eau potable Thèse de doctorat, UTC/Lyonnaise des Eaux/CNRS, 163P.
- Herault J., Jutten C.H., (1994). Réseaux de neurones et traitement de signal. Hermès.
- Hopfield J., (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8): pp 2554–2558.
- Hornik K., Stinchcombe M., White H., (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, *Neural Networks*, 3, pp 551-560.
- Hornik K., Stinchcombe M., White H., (1989). Multilayer feedforward networks are universal approximators, *Neural Networks*, 2, pp 359-366.
- Hornik K., (1991). Approximation capabilities of multilayer feedforward networks, *Neural Networks*, 4, pp 251-257.
- Hornik K., Stinchcombe M., White H., Auer P., (1994). Degree of Approximation Results for Feedforward Networks Approximating Unknown Mappings and Their Derivatives, *Neural Computation*, 6 (6) (1994), pp 1262-1275.
- Hotelling, H., (1933) Analysis of a complex statistical variable into principal component", *J. Edu. Psy.* , Vol 24, pp 417-441 and pp 498-520.
- Hubel D. H. et Wiesel T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *J. Physiol. (Lond.)*, 160, pp 106-154.
- Hubel D. H. et Wiesel T. N. (1963). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens, *J. Neurophysiology*. 26, pp 994-1002.
- Hubel D. H. et Wiesel T. N. (1968). Receptive fields and functional of monkey striate cortex, *J. Physiol. (Lond.)*, 195, pp 215-243.

- Jackson J.E., (1991). A user's guide to principal components. New York Wiley, 351 p
- jouffe L., (1997). Apprentissage de systèmes d'inférence floue par des méthodes de renforcement application a la régulation d'ambiance dans un bâtiment d'élevage porcin. Thèse de doctorat Université de Rennes 1. 192p
- Joo, D.-S., Choi, D. J., Park, H., (2000). The effects of data preprocessing in the determination of coagulant dosing rate. *Water Research* 34 (13), pp 3295–3302.
- Jang J.S. R., (1991a). Fuzzy modeling using generalized neural networks and Kalman filter algorithm. In Proc. of the Ninth National Conference on Artificial Intelligence (AAAI-91), pp 762–767.
- Jang J.S. R., (1991b). Rule extraction using generalized neural networks. In Proc. of the 4th IFSA World Congress, pp 82–86 (in the Volume for Artificial Intelligence).
- Jang, J.S.R., (1993). ANFIS: Adaptive-network-based fuzzy inference systems, *IEEE Transactions on Systems, Man and Cybernetics*, vol.23, n3, pp 665-685.
- Jang J. S.R., Sun C. T., (1993). Functional equivalence between radial basis Function Networks and fuzzy inference systems. *IEEE Transactions on Neural Networks*, 4: pp 156-159.
- Jang J.S., Sun C.T., (1995). Neuro-Fuzzy Modelling and Control", *Proceeding of the IEEE*, vol. 83, No. 3, pp 378-406.
- Jang J.S.R, Sun C.T., Mizutani E., (1997). *Neuro-Fuzzy and Soft Computing*. Prentice-Hall, Upper Saddle River, N.J.
- Kosko B., (1991) *Neural Networks and Fuzzy Systems*, Prentice Hall, Englewood Cliffs, NJ.
- Kandel A., (1992). *Fuzzy expert systems*. CRC Press Inc., Boca Raton, Florida.
- Kohonen, T., (1981). Automatic formation of topological maps of patterns in a self organizing system. In *Proceedings of 2nd Scandinavian Conference on Image Analysis*, pp 214-220.
- Kohonen, T., (1982a). Self organized formation of topological correct feature maps. *Biological Cybernetics*, 43, pp 59-69.
- Kohonen T., (1982b). Analysis of a simple self-organizing process *Biol. Cybernetic*, 44, pp. 135-140.
- Kohonen, T., (1984). *Self organization and associative memory, 2nd Ed*, Springer Verlag.
- Kohonen T., (1989). *Self-organization and Associative Memory* 3rd ed., Springer, New York.
- Kohonen, T., (1995). *Self Organizing Maps*, volume 30 of Springer series of information sciences. Springer Verlag.
- Lamrini B., Benhammou A., Le lann M.V., Karama A., (2005a). A neural software sensor for on-line prediction of coagulant dosage in a drinking water treatment plant. *Transactions of the Institute of Measurement and Control*.

- Lamrini B., Le lann M.V., Benhammou A., Lakhal K., (2005b) Detection of functional states by the "LAMDA" classification technique: application to a coagulation process in drinking water treatment. Elsevier, C.R. Physique 6, pp1161-1168.
- LeCun Y., (1985). Une procédure d'apprentissage pour réseau à seuil asymétrique *Cognitiva*, Vol. 85, pp 599-604.
- LeCun Y., (1987). Modèle connexionniste de l'apprentissage, Thèse de doctorat de l'Université de Paris VI.
- LeCun Y., Boser B., Denker J., Henderson D., Howard R., Hubbard W., Jackel L.,(1990). Handwritten digit recognition with a back-propagation network. In D. Touretzky, editor, *Advances in Neural Information Processing Systems*2, pp 396–404. Denver, CO. Morgan Kaufmann, San Mateo.
- LeCun Y., Bottou L., Bengio Y., Haffner P., (1998a) Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): pp 2278–2324, November.
- LeCun Y., Bottou L., Orr G., Müller K.R., (1998b). Efficient backprop. In G. Orr and K.-R. Müller, editors, *Neural Networks: Tricks of the Trade*, pp 9–50.
- Lee C.C., (1990). Fuzzy logic in control systems: Fuzzy logic controller – part I and II. *IEEE Trans. on Systems, Man and Cybernetics*, 20(2): pp 404-435.
- Lebart L., Piron M. Morineau (2006). Statistique exploratoire multidimensionnelle Visualisation et inférence en fouilles de données. 4e édition Paris: Dunod.
- Lefebvre E., Legube B., (1993). Coagulation floculation par le chlorure ferrique de quelques acides et phénols en solution aqueuse, *Wat. Res.*, 27,3, pp 433-447.
- Lemaire V., (1999). Une nouvelle fonction de coût régularisante dans les réseaux de neurones artificiels : Application à l'estimation des temps de blocage dans un nœud ATM. Thèse de doctorat. Paris VI 165 p.
- Levenberg, K. (1944). A method for the solution of certain nonlinear problem in least squares. *Q Appl Math*; 2: pp 164–168.
- Li X.H., Chen C.L.P., (2000). The Equivalence Between Fuzzy Logic Systems and Feedforward Neural Networks, *IEEE Transactions on Neural Networks*, Vol 11, No. 2, pp 356–365.
- Lin C.T., Lee C.S.G., (1991). Neural-Network-Based Fuzzy Logic Control and Decision Systems. *IEEE Trans. on Computers*, 40(12): pp 1320-1336.
- Lind C., (1994a). Coagulation Control and Optimization: Part one, *Public Works*, pp 56-57.
- Lind C., (1994b). Coagulation Control and Optimization: Part two, *Public Works*, pp 32-33.
- Lind C., (1995). A coagulant road map, *Public Works*, pp 36-38.
- Lippmann R.P., (1987). An introduction to computing with neural nets. *IEEE Acoustics, Speech and Signal Processing Magazine*, 4: pp 4–22.
- Livingstone D.J., Manallack D.T., Tetko I.V., (1997). Data modelling with neural networks: Advantages and limitations. *Journal of Computer-Aided Molecular Design*, vol. 11: pp 135-142.

- 
- Magdalena L., (1995). Analysis of hybrid models: fuzzy logic neural nets. Technical report, Universidad Politécnica de Madrid. MIX Esprit Project 9119.
- Maier H.R., Morgan N., Christopher W.K., (2004). Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. *Environmental Modelling & Software* 19, pp 485–494.
- Maier H.R., Dandy G.C., (1996a). The Use of Artificial Neural Networks for the Prediction of Water Quality Parameters, *Water Research*, 32(4), pp 1013-1022.
- Maier, H.R., Dandy, G.C., (1996b). Neural network models for forecasting univariate time series. *Neural Network World* 6 (5), pp 747–771.
- Makhoul J., El-Jaroudi A., Schwartz R. (1989). Formation of disconnected decision regions with a single hidden layer. In *Proceedings of the International Joint Conference on Neural Networks*, Vol. 1, pp 455-460.
- Malsburg, C. von der, Willshaw, D.J. (1973). How to label nerve cells so that they can interconnect in an ordered fashion. *Proceedings of the National Academy of Science USA*, no. 74, pp 5176
- Matlab 2008Rb. (2008). *Neural Networks Toolbox, User'guide*, Inc.
- Mamdani E.H., (1974). Application of fuzzy algorithms for control of simple dynamic plant, *Proc. of the IEE. Control & Science*, vol. 121, No. 12, Decembre 1974, pp 1585
- Mamdani E.H., (1975). Advances in the linguistic synthesis of fuzzy controllers, *Int. Journal of Man-Machines Studies*, 8, pp 669-678.
- Mamdani, E.H., (1977). Application of fuzzy logic to approximate reasoning using linguistic systems, *Fuzzy Sets and Systems*, vol.26, pp1182-1191.
- Marquardt, D.W., (1963). An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math*; 11: pp 431–441.
- Masschelein W.J., (1999). *Processus unitaires du traitement de l'eau potable*. Editeur: Cebedoc.
- McCulloch W., Pitts W., (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5; pp 115-113.
- Mirsepasi A., Cathers B., Dharmappa H.B., (1995) Application of artificial neural networks to the real time operation of water treatment plants. *IEEE International Conference on Neural Networks Proceedings*, 1, pp 516-521. Perth, Western Australia, 27 November-1 December.
- Mirsepasi A., Dharmappa H.B., Cathers B., (1996) Economic benefits of applying an Artificial Neural Network to the operation of a water treatment plant. *Proceedings, Watertech*, The Centre, Darling Harbour.
- Mirsepasi A., Cathers B., Dharmappa H., (1997). Predicted of chemical dosage in water treatment plants using ANN and Box-Jenkins models. *Preprints of 6th IAWQ Asia-Pacific Regional Conference*. Korea, 16, pp 1561-1568.
- Minsky M., Papert S., (1969). *Perceptrons*, Expanded Edition. MIT Press.
- Mitra S., Hayashi Y., (2000). Neuro-Fuzzy Rule Generation: Survey in Soft Computing Framework', *IEEE Transactions on Neural Networks*, Vol II. N<sup>o</sup>. 3. pp 748–768.
-



- Mohtadi M.F., Rao P.N., (1973). Effect of temperature on flocculation of aqueous dispersions, *water research*, 7, pp 747-767.
- Moody J., Darken J., (1989). Fast Learning in networks of locally tuned processing units, *Neural Computation*, vol. 1, p. 281
- Morere Y., (2001). Mise en œuvre de lois de commande pour les modèles flous de type Takagi\_Sugeno. Thèse de doctorat Université de Valenciennes et du Hainaut Cambrésis.149p
- Mouchet P., (2008).Techniques de l'ingénieur Traitement des eaux avant utilisation filières et applications.
- Murata N., Yoshizawa S., Amari S.I., (1992). Network information criterion: determining the number of hidden units for an artificial neural network model. Rapport technique, 17, Université de Tokyo, Tokyo (Japon).
- Nahm E., Lee S., Woo K., Lee B., Shin S. (1996). Development of an optimum control software package for coagulant dosing process in water purification system, in *Proceedings of the Society of Instrument and Control Engineers Annual Conference*, Vol. 35, pp 1157-1161, Tottori, Japan.
- Nakoula Y., (1997).Apprentissage des modèles linguistiques flous par jeu de règles pondérées. Thèse de doctorat université de Savoie.155P.
- O'melia C.R., (1972). Coagulation and flocculation. Physicochemical processes for water treatment control, W. J. Weber, ed., Wiley, New York, pp 61–109.
- OMM (1994).Office Mondiale de la Météorologie, Guide des pratiques hydrologiques, 829 p.
- Oukhellou, L., (1997): Para métrisation et classification de signaux en contrôle non destructif. Application à la Reconnaissance des Défauts de Rails par Courant de Foucault. Thèse de doctorat université Paris sud, pp211.
- Pagotto C., (1999). Étude sur l'émission et le transfert dans les eaux et les sols des éléments traces métalliques et des hydrocarbures en domaine routier. Thèse de Doctorat Université de Poitiers Ecole des Mines de Nantes Laboratoire Central des Ponts et Chaussées.
- Palm R., (1994). Processing of signal distributions in control loops, Proc. EUFIT'95, Sep 1994, Aachen, Germany, pp 310-321.
- Kandel A., (1992). Fuzzy expert systems. CRC Press Inc., Boca Raton, Florida.
- Palm R., (1998).L'analyse en composantes principales principes et applications. Notes de statistique et d'informatique. Faculté universitaire des Sciences agronomiques Centre de Recherches agronomiques GEMBLOUX (Belgique). 98/233p.
- Park, J., Sandberg I.W., (1991). Universal approximation using radial basis function 212 networks. *Neural Computer*; 3(2): pp 246–57.
- Park, S., Bae H., and Kim, C. (2008). "Decision model for coagulant dosage using genetic programming and multivariate statistical analysis for coagulation/flocculation at water treatment process." *Korean J. Chem. Eng.* 25 (6), pp 1372–1376.
- Parker D. B., (1985). Learning Logic (Tech. Rep. Nos Tr-47). Cambridge, MA: Massachusetts Institute of Technology, Centre for Computational Research in Economics and Management Science.



- 
- Parzen E., (1962). *Annals of Mathematical Statistics*, 33, pp 1065-1076, 1962, On estimation of a probability density function and mode
- Pearson, K., (1901) On lines and planes of closest fit to systems of points in space", *Phil. Mag.*, 2, 11, pp 559-572.
- Press (1972). *Applied multivariate analysis*. New York, Holt. Rinehart and Winston, 521 p.
- Poggio T., Girosi F., (1989). *A Theory of Networks for Approximation and Learning*, AI Memo N°1140, Massachusetts Institute of Technology, Artificial Intelligence Laboratory and Center for Biological Information Processing Whitaker College.
- Powell M.J.D., (1987). *Radial Basis Functions for multivariable interpolation: a review* J.C. Masson and M.G. Gox editors, algorithms for approximation. Clarendon Press, Oxford.
- Plaut, D. C., Nowlan, S. J., Hinton, G. E., (1986). Experiments on learning by back-propagation. Technical report CMU-CS-86-126, Computer Science Département, Carnegie- Mellon Université.
- ragot N., (2003). MELIDIS Reconnaissance de formes par modélisation mixte intrinsèque discriminante à base de systèmes d'inférence floue hiérarchisés. Thèse de doctorat Université de Rennes 1. 222p.
- Ratnaweera H., Blom H., (1995) Optimisation of coagulant dosing control using real-time models selective to instrument errors. *Water Supply* 13(13/14), pp 285-289.
- Renaud S.P., (1999). Application de la théorie des croyances et des systèmes flous à l'estimation fonctionnelle en présence d'informations incertaines ou imprécises. Thèse de doctorat Université de Technologie de Compiègne Laboratoire HeuDiaSyc-UMR CNRS 6599. 204p.
- Roberson, A., Abd-Shukor, S.R., Aziz, N. (2009). Development of Process Inverse Neural Network Model to Determine the Required Alum Dosage at Segama Water Treatment Plant Sabah, Malaysia. *Computer Aided Chemical Engineering* Volume 27, pp 525-530.
- Rumelhart D.E., Hinton G.E., Williams J., (1986). Learning internal representation by error propagation, in *Parallel Distributed Processing*, MIT Press. Cambridge, 1; pp 318-362.
- Rosenblatt F., (1958). The Perceptron: a probabilistic model for information storage and organization in the brain, *Psychological Review* (65), pp 386-408.
- Rosenblatt, F., (1962). *Principles of Neurodynamics*. Spartan New-York.
- Romain B., (2003). *Modèles Impulsionnels de Réseaux de Neurones Biologiques*. Thèse de doctorat. Paris 172 p.
- Riedmiller M., Braun H., (1993). A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In *Proc. of the Int. Conf. on Neural Networks*, San Francisco (CA).
- Rui, Y., A.A. El-Keib (1995). A Review of ANN-based Short-Term Load Forecasting Models, Department of Electrical Engineering, University of Alabama, Tuscaloosa, AL 35487- 92, 93.
-

- Saporta G., (1990) Probabilités, Analyse des Données et Statistique. Éditions Technip, 1990.
- SEAAL (2008). Société eau et assainissement d'Alger. Rapport interne.
- Shafer G., (1976). A mathematical theory of evidence. Princeton University Press,
- Sperring D.A., Chow C.W., Mulcahy D.E., Davey D.E., Haskard M.R., (1992) A neural network applied to sensory signal processing determination of copper in water, *Journal of Intelligent Materials Systems and Structures* 1992: 3(3), 418-431.
- Spearman C., (1904). General intelligence objectively determined and measured. *American Journal of Psychology*, vol. 15, pp. 201-293.
- Specht, D.F., (1991). A general regression neural network. *IEEE Trans Neural Networks*; 2(6):568-76.
- Stum W., Morgan J.J., (1962). Chemical aspect of coagulation. *Journal American Water Works Association*, 54(8):971- 992.
- Stein R., (1993). Selecting data for neural networks. *Ai Expert*, vol. 8 (2): 42-47 pp.
- Sugeno M., Kang G. T., (1986). "Fuzzy modelling and control of multilayer incinerator", *Fuzzy Sets and Systems*, vol. 18, pp. 329.
- Sugeno M., Yasukawa T., (1993). A Fuzzy-Logic-Based Approach to Qualitative Modeling. *IEEE Trans. on Fuzzy Systems*, 1, pp7-31.
- Sugeno M., Kang G. T., (1988). "Structure Identification of Fuzzy Models", *Fuzzy Sets and Systems*, vol. 28, pp. 15-33.
- Sutcliffe J. (1994). On the logical necessity and priority of a monothetic conception of class, and on the consequent inadequacy of polythetic accounts of category and categorization, Dans *New approaches in Classification and data analysis*, E. Diday (coordinateur), pp 53-63, Berlin, Springer-Verlag.
- Sutton R.S., (1992). Special issue on reinforcement learning. *Machine Learning*.
- Takagi T., Sugeno M., (1983). "Derivation of fuzzy control rules from human operator's control actions", *Proc of the IFAC on Fuzzy information*, pp 55-60, Marseille, France.
- Takagi T., Sugeno M., (1985). "Fuzzy Identification of Systems and Its Applications to Modeling and Control", *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 15, No. 1, Feb.
- Thurstone, L., L., (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Rosenblatt F., (1958). The Perceptron: a probabilistic model for information storage and organization in the brain, *Psychological Review* (65), pp 386-408.
- Tran V.B.N., (1999). Contribution à l'étude des diffusiomètres NSCAT et ERS-2 par modélisation neuronale Influence de la hauteur des vagues sur le signal diffusiométrique', Thèse de Doctorat, Université. Paris 6, France.
- Valiron F., (1989). *Gestion des Eaux : alimentation en eau - assainissement*, Presses de l'école nationale des ponts et chaussées, Paris.
- Valentin N., Denoeux T., Fotoohi F., (1999a). A hybrid neural network based system for optimisation of coagulation doing in a water treatment plant. *Proceedings of IJCNN99*, Washington D.C. July 1999, IEEE.

- Valentin N., Denoeux T., and Fotoohi F., (1999b). Modelling of coagulant dosage in a water treatment plant. *Proceedings of EANN '99*, Warsaw, September. pp 165-170.
- Valentin N., (2000). Construction d'un capteur logiciel pour le contrôle automatique du procédé de coagulation en traitement d'eau potable. Thèse de doctorat, UTC/Lyonnaise des Eaux/CNRS, 153P.
- Van Leeuwen J., Chow C.W., Bursill D., Drikas M., (1999) Empirical mathematical models and artificial neural networks for determination of alum doses of southern Australian surface waters *Journal of Water Science Research and Technology - Aqua*48(3), pp115-127.
- Vapnik, V. N., Chervonenkis, A. Y., (1971): On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its applications*, 16: pp264-271.
- Vapnik, V., (1982). *Estimation of Dependences Based on Empirical Data*. Springer-Verlag.
- Vapnik, V., (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag New York Heidelberg Berlin.
- Villa J.L., Duque M., Gauthier A., Rakoto-ravalontsalama N., (2003). Modeling and control of a water treatment plant. *IEEE 0-7803-7952-7/03*, pp171-180.
- Walley, W.J., Martin, R.W. et O'Connor, M.A. (2000). Self-organizing maps for classification of river quality from biological and environmental data. In: Denzer, R., Swayne, D.A., Purvis, M., Schimak, G. (eds) *Environmental software systems: environmental information and decision support*, IFIP Conference Series, Kluwer Academic Publishers, Amsterdam, pp 27-41.
- Wang L.X., (1994). *Adaptive Fuzzy Systems and Control, Design and Stability Analysis*. Prentice Hall, New Jersey.
- Wang L.X., Mendel J.M., (1992a). Fuzzy Basis Functions, Universal Approximation and Orthogonal Least-Squares Learning, *IEEE Trans. on Neural Networks*, Vol. 3, No. 5, pp 807-814.
- Wang L.X., Mendel J.M., (1992b). "Fuzzy Systems are Universal Approximators", *Proc. of the IEEE Conf. on Fuzzy Systems (Fuzz'IEEE 92)*, SanDiego, pp 1163-1170.
- Werbos P.J., (1974). *Beyond regression: New tools for prediction and analysis in the behavioral science*, Thèse de doctorat, Harvard University, 1974.
- Widrow B., Hoff M., (1960). Adaptive switching circuits, *IRE WESCON Convention Record*, New York: IRE, pp 96-104.
- White D., Sofge D., (1992). *Handbook of intelligent control*. Van Nostrand Reinhold, New York.
- Wilson, D. R., Martinez, T. R., (2003). The general inefficiency of batch training for gradient descent learning. *Neural Networks*, 16(10): pp 1429–1451.
- Willshaw, D. J. et Von der Malsburg, C., (1976). How patterned neural connections can be set up by self organization. In *Proceedings of the Royal Society of London*, Vol. B194, pp 431-445.

- Weishaar J.L., Aiken G.R., Bergamaschi B. A., Fram M.S., Fujii R., Mopper K., (2003). Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of dissolved organic carbon. *Environmental Science and Technology*, 37: pp 4702-4708.
- Wu, G.D., and Lo, S.L. (2008). Predicting Real-time Coagulant Dosage in Water Treatment by Artificial Neural Networks and Adaptive Network Based Fuzzy Inference System, *Engineering Applications of Artificial Intelligence*, 21 (8): pp1189-1195.
- Wu, G.D., and Lo, S.L. (2010). Effects of Data Normalization and Inherent-factor on Decision of Optimal Coagulant Dosage in Water Treatment by Artificial Neural Network, *Expert Systems with Applications*, 37: pp 4974 - 4983.
- Yager R.R., Filev D.P., (1994). Generation of fuzzy rules by mountain clustering. *Journal of Intelligent and Fuzzy Systems*, 2, pp 209-219.
- Yu, R.-F., Kang, S.-F., Liaw, S.-L., Chen, M.-C., (2000). Application of artificial neural network to control the coagulant dosing in water treatment plant. *Water Science and Technology* 42 (3-4), pp 403–408.
- Zadeh L.A., (1965). Fuzzy Sets, *Information and control*, vol. 8, pp 338-353.
- Zadeh L.A., (1971). Quantitative fuzzy semantics, *Information Sciences*, vol. 3, pp 159-176.
- Zadeh, L.A., (1973). Outline of a New Approach to Analysis of Complex Systems and Decision Process, *IEEE Trans. on Systems, Man, and Cybernetics*, 3(1), pp 28-44.
- Zadeh L.A., (1978). «PRUF- A Meaning Representation Language for Natural Languages». *Int. Journal. Man-Mach Studies*, Vol. 10, pp 395-460.
- Zadeh L.A., (1979). A Theory of Approximate Reasoning. *Machine Intelligence*, vol. 9, pp 149-194.
- Zhang, Q., Stanley, S.J., (1999). Real-time water treatment process control with artificial neural networks. *Journal of Environmental Engineering*, ASCE 125 (2), pp 153–160.
- Zemouri, R. (2003). Contribution à la surveillance des systèmes de production à l'aide des réseaux de neurones dynamiques : Application à l'e-maintenance. Thèse de doctorat L'UFR des Sciences et Techniques de l'Université de Franche-Comté. 277 p.
- Zemouri R., Racoceanu D., Zerhouni N., (2003). Recurrent Radial Basis Function network for Time-Series Prediction, *Engineering Applications of Artificial Intelligence* (Elsevier Science), Volume 16, Issue 5-6, pp. 453-463.