Université Mohamed Boudiaf - M'sila

FACULTE DE TECHNOLOGIE

DEPARTEMENT D'ELECTRONIQUE

Numéro de série : ……

Numéro d'inscription : …….

Thèse

Présentée pour l'obtention du diplôme de

DOCTORAT SCIENCES

Spécialité : Electronique
Option : Contrôle industriel

THEME

Contribution au Développement de Systèmes Multicapteurs Intelligents Dédiés à la Surveillance et au Contrôle de la Qualité des Eaux Propres.

Présenté par

DJERIOUI Mohamed

Soutenue le : 02/05/2019

Devant le jury composé de :

Nom & Prénom	**Grade**	**Etablissement**	**Qualité**
HOCINI Abdesselam	Professeur	Université Med Boudiaf, M'sila	Président
BOUAMAR Mohamed	Professeur	Université Med Boudiaf, M'sila	Rapporteur
BOUDJEMA Fares	Professeur	E N P, Alger	Examinateur
KHEMLICHE Mabrouk	Professeur	Université Ferhat Abbas, Sétif 1	Examinateur
ZERGUINE Azzedine	Professeur	Université King Fahd, Arabie Saoudite	Invité

Année Universitaire : **2018/2019**

DEDICACE

A la mémoire de mon très cher père "Abdelaziz" et à ma très chère mère "Zohra"
A ma petite famille ; ma femme et mes enfants "Khalil" et "Yasmine" ;
A mes frères et sœurs ;

A tous ceux que j'aime.

بسم الله .. والحمد والشكر لله ..

REMERCIEMENTS

Au-delà de la formalité d'usage, c'est avec un grand plaisir que je tiens à exprimer tout d'abord ma gratitude et mes sincères remerciements à mon directeur de thèse, le Professeur BOUAMAR Mohamed, Chef de l'équipe de recherche "Microsystèmes" du LASS, Laboratoire d'Analyse des Signaux et Systèmes, Université Mohamed BOUDIAF de M'sila, pour m'avoir accepté sous son aile et m'avoir fait découvrir avec générosité, patience et gentillesse quelques uns des secrets du monde de la recherche scientifique. Je le remercie également pour ses précieux conseils, son investissement dans tous les aspects de mon travail, ainsi que son temps consacré à suivre toutes les étapes de ma thèse et ce malgré ses nombreux engagements.

Mes remerciements vont également aux membres du jury : le Professeur HOCINI Abdesselam, pour avoir accepté très aimablement de présider mon jury de soutenance, aux : Professeur BOUDJEMA Fares et Professeur KHEMLICHE Mabrouk, pour l'honneur qu'ils m'ont fait en acceptant d'examiner et de juger objectivement ce travail de recherche.

Je remercie chaleureusement le Professeur ZERGUINE Azzedine, Département de Génie Electrique de l'Université King Fahd du Pétrole et des Minéraux en Arabie Saoudite, pour son aide appréciable et bénéfique ainsi que pour l'intérêt qu'il a manifesté en participant en qualité de membre invité à mon jury de soutenance.

Je tiens à remercier également le Docteur LADJAL Mohamed, chef du Département d'Electronique de l'Université Mohamed BOUDIAF de M'sila, le Docteur CHOUDAR Aissa et le Docteur KHENNOUF Salah qui ont toujours répondu présents pour m'apporter leur précieuse aide et leur soutien moral.

Une place dans ces remerciements est réservée au staff administratif ainsi qu'à tous les enseignants du Département d'Electronique, Faculté de Technologie, Université Mohamed BOUDIAF de M'sila.

Table des matières

DEDICACES i

REMERCIEMENTS ii

TABLE DES MATIERES iii

RESUME ix

ABREVIATIONS ET NOTATIONS x

LISTE DES FIGURES xiv

LISTE DES TABLEAUX xvi

INTRODUCTION GENERALE 01

CHAPITRE I

CONTROLE ET PRODUCTION DES EAUX PROPRES

INTRODUCTION 06

1. DEFINITION DE L'EAU PROPRE 06

1.1. Types des eaux propres 07

1.1.1. L'eau minérale naturelle 07

1.1.2. L'eau de source 07

1.1.3. L'eau rendue potable par traitements 07

2. CARACTERESTIQUES DES EAUX 07

2.1. Les paramètres organoleptiques 07

2.2. Les paramètres physico-chimiques 08

2.2.1. La température 08

2.2.2. La dureté 09

2.2.3. Le potentiel d'Hydrogène 09

2.2.4 L'alcalinité 10

2.2.5. Les Matières en Suspension 10

2.2.6. La conductivité 10

2.2.7. L'Oxygène Dissous 11

2.2.8. Résidu sec 11
2.3. Les paramètres indésirables 12
2.3.1. Les matières organiques 12
2.3.2. L'azote 14
2.4. Les paramètres toxiques : éléments métalliques et non métalliques 14
2.5. Caractéristiques biologiques 15
3. LES NORMES DE POTABILITE 16

4. TRAITEMENT DES EAUX 18
4.1. Importance de traitement 18
4.2. Chaîne de traitement 20
4.2.1. Prétraitement 20
4.2.2. Pré-oxydation 21
4.2.3. Clarification 22
4.2.3.1. Coagulation – Floculation 22
4.2.3.2. La filtration sur sable 23
4.2.4. Oxydation – Désinfection 23
4.2.5. Affinage 25
4.3 Les nouveaux procédés 25

5. LA SURVEILLANCE DE LA QUALITE DES EAUX 26
5.1. La supervision des procédés 27
5.2. Méthodes de surveillance 27
5.2.1 Méthodes sans modèle 29
5.2.2 Méthodes à base de modèle 30

CONCLUSION 31

CHAPITRE II

CAPTEURS DE SURVEILLANCE DES EAUX PROPRES

INTRODUCTION 32
1. GENERALITES 32

1.1. Définition 32
1.2. Structure d'un capteur 32
1.2.1. Le corps d'épreuve 33
1.2.2. Transducteur 33
1.2.3. Le boîtier 33
1.2.4. L'électronique de conditionnement 33
1.3. Types de capteurs 33
1.3.1. Capteurs actifs 33
1.3.2. Capteurs passifs 33
1.4. Caractéristiques métrologiques d'un capteur 34
1.4.1. Etalonnage 34
1.4.2. Sensibilité 34
1.4.3. Temps de réponse 34
1.4.4. Précision, Sensibilité, Gamme et Etendue de mesure 35
1.4.5. Fiabilité en environnement 35

2. CAPTEURS PHYSICO-CHIMIQUES DE SURVEILLANCE DES EAUX 35
2.1. Mesure de la turbidité 35
2.2. Mesure de la conductivité 36
2.3. Mesure du pH 37
2.4. Mesure d'oxygène dissous 38
2.5. Mesure de la température 38
2.6. Les sondes multi-paramètres 39

3. CAPTEURS INTELLIGENTS 40
3.1. Conjonction des besoins et Technologie 41
3.2. Constitution des capteurs intelligents 42
3.3. Conception des capteurs intelligents 45
3.3.1. Approche fonctionnelle 45
3.3.2. Approche informationnelle 45
3.3.3. Approche comportementale 45
3.4. Fonctions primordiales d'un capteur intelligent 46
3.5. Aspect économique 47
3.6. Domaines d'applications des capteurs intelligents 47

4. SYSTEMES A BASE DE CAPTEURS INTELLIGENTS 47
4.1. Capteurs intelligents autonomes 48
4.2. Réseau de capteurs intelligents 48
4.3. Capteurs intelligents en contrôle de la qualité des eaux 49
4.4. Exemple d'application 50

5. CAPTEURS LOGICIELS 52

CONCLUSION 53

CHAPITRE III

TECHNIQUES DE DEVELOPPEMENT DES CAPTEURS LOGICIELS

INTRODUCTION 54
1. LE CAPTEUR LOGICIEL 55
1.1. Capteurs logiciels appliqués en traitement des eaux : état de l'art 56
1.2. Types de capteurs logiciels 57
1.2.1. Capteurs logiciels pilotés par les données 57
1.2.2. Capteurs logiciels pilotés par le modèle 57
1.3. Développement du capteur logiciel 58
1.3.1. Acquisition des données 59
1.3.2. Sélection des variables d'entrées 59
1.3.3. Choix du modèle et apprentissage 60
1.3.4. Validation du modèle 61
1.3.5. Maintenance du modèle 61
1.4. Applications des capteurs logiciels 62
1.4.1. Prédiction en ligne 62
1.4.2. Surveillance de processus et détection de défauts de processus 62
1.4.3. Détection et reconstruction de défauts de capteurs 63

2. TECHNIQUES DE DEVELOPPEMENT 63
2.1. Etat de l'art 63
2.2. Méthode de réduction des dimensions 64
2.2.1. Analyse en Composantes Principales 65
2.2.2. Concepts de base 66
2.3. Méthodes d'apprentissage 70
2.3.1. Les réseaux de neurones artificiels 70
2.3.1.1. Les perceptrons multicouches 71
2.3.1.2. Les réseaux de neurones RBF 74
2.3.2. Les Machines à Vecteurs de Support pour la régression 76
2.3.3. Les Machines d'apprentissage extrême 78
2.3.3.1 Théorie de la technique ELM 80
2.3.3.2 L'algorithme ELM 81

CONCLUSION 82

CHAPITRE IV

SIMULATION ET EVALUATION DES PERFORMANCES

INTRODUCTION 83
1. PROBLEMATIQUE 83
2. ARCHITECTURE DU SYSTEME DE SURVEILLANCE 84
3. DEVELOPPEMENT DU CAPTEUR LOGICIEL 86
3.1. Description des données d'entrée 86
3.2. Prétraitement des données 88
3.2.1. Normalisation des données 88
3.2.2. Sélection des variables d'entrées 89
3.3. Choix du modèle, apprentissage et test 91
3.3.1. Modèle Neuronal-MLP 91
3.3.2. Modèle Neuronal-RBF 93

3.3.3. Modèle SVM 94
3.3.4. Modèle ELM 96
3.3.5. Evaluation des performances 98
4. APPLICATION 102

CONCLUSION 106

CONCLUSION GENERALE 107

REFERENCES 110

RESUME

Actuellement, dans le domaine de traitement des eaux propres, l'utilisation de procédés automatiques devient impérative pour atteindre deux objectifs principaux : la maîtrise de la qualité de l'eau, et la diminution des contraintes de coût de fonctionnement. Les variables de qualité jouent un rôle important dans le contrôle et la surveillance de l'eau. Bien que certains paramètres puissent être mesurés en continu à l'aide de capteurs physiques à faible coût, il existe d'autres paramètres nécessitant des analyses de laboratoire spécifiques et coûteuses en raison de l'absence de capteurs dédiés. En outre, il existe un grand nombre de capteurs hétérogènes qui peuvent prendre beaucoup de temps lors des étapes de mesure et de traitement. Néanmoins, le système à base des capteurs logiciels peut fournir un moyen efficace et économique pour résoudre ces problèmes.

Notre travail peut être considéré comme une contribution aux solutions proposées, pour résoudre des problèmes d'intérêt stratégique à préoccupation nationale et internationale, utilisant des outils modernes à base de techniques avancées. On présente dans ce cadre, une contribution à l'étude et au développement de capteurs logiciels pouvant être intégrés au niveau de capteurs intelligents et utilisés dans la surveillance et le contrôle de la qualité des eaux propres. Ces capteurs algorithmiques à base de techniques de l'intelligence artificielle et représentant un moyen très attrayant pour faire face au manque de capteurs spécifiques, sont devenus alors très convoités. Dans un but de choix de la technique la mieux adaptée à la conception de ces capteurs logiciels, une étude comparative d'évaluation de plusieurs méthodes, est à cet effet présentée. L'objectif, est de mettre en œuvre une architecture de système de surveillance fondée sur l'emploi de ce type de capteurs, pouvant être intégrés au sein de plateformes plus élaborées, construites autour de capteurs intelligents.

Mots clés : Eaux propres, Contrôle de la qualité, Intelligence artificielle, Capteurs logiciels, Régression, Classification, Evaluation, Simulation.

ABREVIATIONS ET NOTATIONS

ANRH : Agence Nationale des Ressources Hydrauliques

CEE : Communauté Economique Européenne

DX : Nom de l'international workshop sur les principes du diagnostic

OMS : Organisation Mondiale de la Santé

UNICEF : Fonds des Nations unies pour l'enfance

ACP : Analyse en Composantes Principales

CAO : cartes auto-organisées

ELM : Extreme Learning Machine

IA : Intelligence Artificielle

MCP : Moindres carrés partiels

MLP : Multi-Layer-Perceptron

PMC : Perceptron Multi Couches

RBF : Radial Basis Function

RNs : Réseaux de neurones

RNA : Réseaux de Neurones Artificiels

RNSCC : Réseaux de Neurones à Simple Couche Cachée

SLFN: Single Layer FeedForward Neural Network

SVM : Support Vector Machines

SVR : Support Vector Regression

Ag^+ : Argent

BF_4^- : Fluoroborate

Br^- : Bromure

C : Conductivité

Ca : Calcium

$CaCl_2$: Chlorure de calcium

Cl : Chlore

CO_3^{-2} : Carbonate

COD : Carbone Organique Dissous

COP : Carbone Organique Particulaire

COT : Carbone Organique Total

Cu : Cuivre

Cu^{2+} : Cuivrique

DBO : Demande Biochimique en Oxygène

DCO : Demande Chimique en Oxygène

F^- : Fluorure

Fe : Fer

HCO_3^- : Hydrogénocarbonates

I^- : Iodure

Li^+ : Lithium

MES : Matières En Suspension

Mg : Magnesium

Mn : Manganèse

MO : Matières Organiques

NH_4^+ : Ammoniaque

NH4-N : Azote ammoniacal

NO_2^- : Nitrites

NO_3^- : Nitrates

OD : Oxygène Dissous

OH^- : Hydroxyde

ORP : Potentiel d'oxydoréduction

pH : Potentiel d'Hydrogène

$PKMnO_4$: Permanganate de potassium

T : Température

TA : Titre alcalimétrique

TAC : Titre Alcalimétrique Complet

TP : Phosphore Total

TU : Turbidité

Zn : Zinc

AMDEC : Analyse des Modes de défaillance, de leurs Effets et de leur Criticité

AdD : Arbre De Défaillance

AN : Analogique-Numérique

CAN : Convertisseur Analogique/Numérique

CDMA : Accès multiple par division de code

dsPIC : Digital Signal Peripheral Interface Controller

FDI : Failure Detection and Isolation

GPRS : Service général de radiocommunication par paquets

GSM : Système global de communication mobile

LoRaWAN: Long Range Wide Area Network

OTAP: Programmation en liaison radio

PROM: Programmable Read Only Memory

RAM: Random Access Memory

RCSF : Réseau de Capteurs Sans Fil (WSN)

RF : Radio Frequency

UV : Rayonnements ultraviolets

WiFi : Wireless Fidelity

dbar : decibar

mg/l : milligrammes par litre

mV : milliVolt

NTU : Unité néphélométrique de turbidité

ppm : partie par million

µS/cm : Micro-Siemens par centimètre

b **:** biais

C : Le minimum du critère quadratique

COV : La matrice de covariance

cr :centrées et réduites

EAM : L'erreur absolue moyenne

EQM : L'erreur quadratique moyenne

$H^{\dagger}$: La solution inverse généralisée de Moore-Penrose

KL : Karhunen-Loève

m : mesurande

$max(X)$ **:** La valeur maximale de X

$min(X)$ **:** La valeur minimale de X

N **:** Le nombre d'exemples de la base d'apprentissage.

NCC : Le nombre de couches cachées

NNCC : Le nombre de neurones dans la couche cachée

Q **:** La fonction objective

$\mathbf{R^2}$ **:** Le coefficient de corrélation

REQM : La racine de l'erreur quadratique moyenne

s : La sensibilité

t_{appr} **:** Le temps d'apprentissage

$\mathbf{t_{ex}}$ **:** Le temps d'exécution

$\mathbf{t_{\alpha}}$: Le temps de réponse

V **:** La matrice de changement de base

$\mathbf{w}$ **:** Poids

$X = [x_1, x_2, \dots, x_i]$ **:** Le vecteur à normalisé

$\bar{x}$ **:** La valeur moyenne

x_{i_N} **:** La valeur normalisée

σ_x **:** Ecart type

λ **:** La valeur propre

ϕ **:** Fonction d'activation

μ **:** Les centres

σ **:** Les largeurs

δ **:** La marge

α_i **:** Les multiplicateurs non négatifs de Lagrange

$\boldsymbol{\varepsilon}$ **:** La marge de tolérance

σ **:** L'écart type

LISTE DES FIGURES

Fig.1.1. Chaîne de traitement des eaux de surface. 20

Fig.1.2. Classification globale des méthodes de surveillance. 29

Fig.2.1. Structure d'un capteur. 33

Fig.2.2. Courbe d'étalonnage d'un capteur. 34

Fig.2.3. Exemple de capteur de turbidité. 36

Fig.2.4. Exemple de Conductimètre. 37

Fig.2.5. Exemple de pH-mètre. 37

Fig.2.6. Exemple de sonde pour oxygène dissous. 38

Fig.2.7. Exemple de capteur de température de l'eau. 39

Fig.2.8. Exemple de sonde de qualité de l'eau (Borehole). 39

Fig.2.9. Capteur « smart » et capteur intelligent. 43

Fig. 2.10. Architecture matérielle d'un capteur intelligent. 44

Fig.2.11. Topologie en étoile de capteurs intelligents autonomes. 48

Fig.2.12. Communication multi-sauts ou coopérative dans un RCSF. 49

Fig.2.13. Plateforme de capteurs intelligents « Waspmote smart water ». 51

Fig.2.14. Capteur logiciel décrit comme un modèle d'entrées-sortie. 53

Fig.3.1. Principales étapes de développement d'un capteur logiciel. 58

Fig.3.2. Principales familles de méthodes employées dans la conception de capteurs logiciels. 64

Fig.3.3. Éléments constituant un neurone formel. 70

Fig.3.4. Les réseaux de neurones artificiels multicouches. 72

Fig.3.5. Hyper-tube modélisant la fonction de régression. 77

Fig.3.6. Utilisation des noyaux pour la résolution de la régression non linéaire. 78

Fig.3.7. Structure du réseau de neurone SLFN. 80

Fig.4.1. Architecture du système pour la surveillance et le contrôle de la qualité des eaux propres. 85

Fig.4.2. Décision à base de capteurs logiciels. 85

Fig.4.3. Etapes de développement d'un capteur logiciel. 86

Fig. 4.4. Evolution temporelle des paramètres descripteurs. 87

Fig. 4.5. Histogramme des valeurs propres des composantes. 90

Fig. 4.6. Organigramme d'élaboration du modèle ELM pour le capteur logiciel. 97

Fig4.7. Temps d'entraînement des différents modèles élaborant le capteur logiciel de Cl. 101

Fig.4.8. Données réelles et estimées de la concentration Cl pour les modèles MLP, RBF, SVM et ELM. 102

Fig.4.9. Temps d'entraînement correspondant aux modèles MLP, RBF, SVM et ELM. 104

Fig. 4.10. Classifieur et capteur logiciel Cl à base du modèle ELM. 104

Fig. 4.11. Taux de classifications obtenus pour les 3 scénarios (cas). 105

Fig. 4.12. Temps d'exécution obtenus pour les 3 scénarios (cas). 105

LISTE DES TABLEAUX

Tableau. 1.1. Valeurs guides de l'ANRH et concentrations limites des directives CEE utilisées pour la production d'eau destinée à la consommation humaine. 17

Tableau 2.1. Caractéristiques de la sonde multi-paramètre *Borehole*. 40

Tableau 2.2. Intérêt fonctionnel des capteurs intelligents. 42

Tableau 3.1. Problèmes de capteurs physiques. 55

Tableau.4.1. Statistiques descriptives des paramètres descripteurs utilisés. 87

Tableau.4.2. Statistiques des composantes principales. 89

Tableau.4.3. Valeurs propres, proportions de variance et corrélations entre variables des quatre premières composantes principales. 91

Tableau.4.4. Les bases de données utilisées. 91

Tableau.4.5. Résultats d'entraînement et de test pour différentes architectures du modèle neuronal-MLP. 92

Tableau.4.6. Résultats d'entraînement et de test pour différentes architectures du modèle neuronal-RBF. 93

Tableau.4.7. Analyse de sensibilité de SVM pour les trois bases de données. 94

Tableau.4.8. Résultats d'entraînement et de test pour différentes architectures du modèle ELM. 97

Tableau.4.9. Paramètres statistiques des différents modèles élaborant le capteur logiciel Cl. 99

Tableau.4.10. Paramètres de performance des différents modèles élaborant le capteur logiciel Cl. 100

Tableau.4.11. Performances des modèles MLP, RBF, SVM et ELM. 103

INTRODUCTION GENERALE

Aujourd'hui, l'eau propre en tant que source de vie, représente l'un des produits alimentaires les plus surveillés dans le monde. En effet et depuis quelques années, l'importance des technologies modernes en vu de son traitement attire de plus en plus l'attention des pays en voie de développement [1-3]. Selon l'Organisation Mondiale de la Santé (OMS) et le Fonds des Nations Unies pour l'enfance (UNICEF), en 2015, 844 millions de personnes manquaient encore de services d'eau potable [4,5]. Ajouter à cela, les activités de traitements de moindre qualité qui augmentent davantage les risques de maladies par l'utilisation d'eaux de surface polluées [6]. A des fins d'amélioration de la qualité de cette ressource vitale, l'usage, le développement et la diffusion de technologies de traitement de plus en plus modernes deviennent alors prioritaires quant aux choix stratégiques des pays.

En effet, le domaine de la surveillance de l'eau se trouve le plus souvent sous une pression croissante pour produire une eau potable à la fois d'une plus grande qualité et à moindre coût. Ceci représente une économie en termes de coût, mais aussi en termes de respect de l'environnement. Des caractéristiques bien spécifiques qu'il est indispensable dans ce cas d'en tenir compte dans la construction d'une démarche globale de prévention des risques. La maîtrise de la qualité sanitaire de l'eau pallie sans aucun doute les conséquences graves qui se concrétisent au niveau des risques encourus pour la santé publique. L'exigence d'une réglementation très stricte des pouvoirs publics dans ce domaine est donc bel et bien justifiée. Il faut souligner à ce stade qu'une usine moderne de production d'eau potable assure en fait deux principales fonctions : la satisfaction de la demande en eau, et l'assurance d'un niveau de qualité élevé et uniforme [7]. Des systèmes de surveillance permanents doivent alors assurer le contrôle des divers procédés et installations de traitement, particulièrement les paramètres relatifs à la qualité de l'eau en sortie d'une station de production [8].

Quand on parle de surveillance de l'eau, il s'agit en fait de connaître son état en continu à partir de différents paramètres, dits descripteurs, ayant trait à sa qualité. Traditionnellement, on procède généralement à un certain nombre de mesures utiles des paramètres descripteurs de l'eau brute pour le test de sa qualité tels que : le contrôle bactériologique, le contrôle de désinfection et le contrôle physico-chimique, pour décider enfin sur son état propre après

avoir appliqué les techniques et méthodes adéquates pour la rendre potable. L'inconvénient de cette procédure est qu'elle nécessite de façon non stop des interventions et des déplacements sur site de l'opérateur. Elle ne permet pas en fait de suivre finement l'évolution de la qualité de l'eau brute et a le désavantage d'avoir un temps de réponse relativement long. Ceci représente tout l'intérêt de disposer d'un contrôle automatique pour une meilleure efficacité de traitement et une réduction des coûts d'exploitation. Le contrôle de l'eau au niveau des usines de production doit donc se faire de façon immédiate en se basant sur une surveillance continue des paramètres descripteurs de la qualité à travers toute une chaîne de traitement qui assure à la fois surveillance et contrôle. L'automatisation de la surveillance permet en fait de s'affranchir de certaines faiblesses humaines, d'accroître la fiabilité et d'améliorer les conditions d'exploitation et des performances de l'installation par la suppression d'un maximum de tâches répétitives et pénibles pour l'opérateur. Automatiser les procédés de traitement peut donc avoir des objectifs divers, les plus fréquents sont d'augmenter les performances du système de production, de garantir la qualité du produit, de la surveillance des installations et des équipements de traitement, de diminuer les coûts de production et d'améliorer la sécurité de l'installation industrielle et de son environnement. De plus, l'automatisation couplée à un stockage d'informations, permet d'envisager des études statistiques de données recueillies ouvrant la voie à des études de modélisation et donc d'optimisation des procédés de traitement. Dans ce cadre, l'automatisation et la supervision par ordinateur ont conduit à une augmentation du confort en permettant une meilleure maîtrise même à distance d'un grand nombre d'informations, ce qui permet de simplifier les tâches d'exploitation, de surveillance, de maintenance et de gestion. L'utilisation des techniques issues du domaine de l'intelligence artificielle apparaît alors comme la principale alternative pour aborder ces problèmes lorsqu'il est nécessaire de prendre en compte l'intervention des experts du domaine ou de traiter l'information de nature qualitative.

Dans la surveillance de la qualité de l'eau, comme dans d'autres domaines similaires, un contrôle efficace dépend principalement de la précision et de la fiabilité des capteurs de mesure utilisés [9]. D'autre part, l'eau est considérée comme une entrée de processus et a plusieurs paramètres descripteurs à contrôler et à surveiller. De plus, l'hétérogénéité de ces paramètres mesurés par les capteurs, est considérée comme un réel défi en matière de qualité de l'eau [10]. Ces paramètres de la qualité jouent en réalité un rôle fondamental pour garantir une eau propre de qualité. Bien que certains de ces paramètres puissent être mesurés en continu à l'aide de capteurs à faible coût, il existe d'autres paramètres nécessitant des analyses

de laboratoire spécifiques et coûteuses en raison de l'absence de capteurs dédiés [11]. En fait, et pour plusieurs raisons, telles que le coût élevé de certains capteurs, leur nombre, le temps consacré à leur contrôle, leur nettoyage, les routines de calibrage et de remplacement, rendent souvent leur bon fonctionnement difficile.

En vu des problèmes susmentionnés, les capteurs logiciels (Soft Sensors) conçus à base d'algorithmes inspirés des techniques de l'intelligence artificielle, sont devenus un moyen efficace et très attractif pour pallier au manque de capteurs physiques et/ou spécifiques [12-18]. En effet, le déploiement des dispositifs microélectroniques a conduit à une réduction drastique des coûts informatiques, ce qui a permis l'émergence et le développement de ces capteurs logiciels ainsi que leur application généralisée, telle que dans les systèmes biochimiques [19], la surveillance des bioprocédés [20], le traitement biologique des eaux usées [12] et la prédiction de processus [13,18]. Dans ce cadre, plusieurs recherches ont porté sur le développement de ce type de capteurs, en particulier dans le secteur de l'eau. Dans [21], les auteurs ont développé un capteur logiciel en utilisant un algorithme génétique basé sur le système neuro-flou dans le traitement des eaux usées pour estimer la dynamique des nutriments de l'opération anoxique / oxique en utilisant des mesures en ligne plus fiables de perméat-concentration d'hydrogène où le potentiel de réduction et l'oxygène dissous (OD) sont utilisés comme données d'entrée. Lamrini et al [22] ont présenté l'élaboration et la validation d'un capteur logiciel en utilisant un Perceptron Multi- Couche (PMC) pour la prédiction en ligne du dosage optimal de coagulant à partir de mesures de qualité d'eau brute dans une usine de traitement d'eau potable. D'autre part, un modèle de capteur logiciel basé sur la fonction de base radiale (RBF) a été étudié pour estimer les principaux paramètres de performance, à savoir la concentration du potentiel d'Hydrogène (pH), le flux de perméat et la concentration d'hydrogène résiduel [23]. La technique des « Machines à Vecteurs de Support » ou Support Vector Machine (SVM) [24-26], conçue pour les problèmes de classification au départ, a été par la suite appliquée à la régression [27,28]. Récemment, la technique « Extrême Learning Machine » (ELM) [29], qui est un algorithme d'apprentissage rapide pour les Réseaux de Neurones à Simple Couche Cachée (RNSCC), est aussi bien utilisée pour des applications de classification que de régression. Elle permet de générer aléatoirement les paramètres de nœuds cachés, puis détermine analytiquement les poids de sortie au lieu d'un réglage itératif. En plus de la facilité de mise en œuvre, le temps de traitement se voit être considérablement réduit [30,31]. Dans [32], la technique ELM a été comparée aux réseaux neuronaux à Rétro propagation et à la régression des vecteurs de

support (SVR) afin de mesurer les concentrations d'OD dans un processus d'aquaculture. On a trouvé que la précision de prédiction de l'algorithme ELM était relativement élevée et que celui-ci pouvait être choisi dans des opérations de contrôle et d'optimisation des processus d'aquaculture. Comme étudié dans [33], l'algorithme ELM présente plusieurs avantages par rapport aux autres algorithmes d'entraînement ou d'apprentissage statistique tels que SVM, least square SVM et proximal SVM en termes de complexité informatique, de précision de résolution et de rapidité d'apprentissage pour la régression, la classification binaire et la classification multi-classe. Fei Kang et al [34] ont présenté la performance du modèle de suivi de santé basé sur l'ELM pour la prédiction de déplacement des barrages gravitaires et l'ont comparée avec celle des réseaux de neurones à Rétro propagation, des régressions linéaires multiples et des modèles de régression pas à pas en termes de prédiction et de vitesse d'entraînement.

L'ELM, en tant que nouvel algorithme d'apprentissage et pouvant être utilisé en régression comme en classification, a montré des avantages inégalés en matière de rapidité pour résoudre les problèmes rencontrés dans la mesure continue en ligne. A souligner dans ce cadre, que les données manipulées au niveau des systèmes de surveillance et de contrôle dans les usines de traitement des eaux, sont d'un très grand nombre d'échantillons, soit d'importantes bases de données à élaborer et à explorer.

Dans ce travail thèse, une étude d'évaluation de plusieurs techniques de régression issues de l'intelligence artificielle et appliquées au domaine de traitement des eaux propres, telles que : MLP, RBF, SVM et ELM, est présentée. L'objectif est de mettre en œuvre une architecture de système de surveillance fondée sur l'emploi de capteurs logiciels à base de ces techniques et pouvant être intégrés au sein de plateformes de capteurs intelligents. D'un usage plus économique, ces capteurs hybrides permettent de prendre une décision adaptée dans le contrôle et le suivi des processus pour une meilleure qualité de l'eau.

Le travail réalisé dans ce cadre, est structuré autour de quatre chapitres :

Le premier chapitre présente une introduction au domaine de l'eau propre et à son traitement. Des généralités, des définitions de caractéristiques ainsi que des normes de potabilité des eaux propres sont présentées. Les méthodes de contrôle et de surveillance employées dans le domaine sont décrites à cet effet.

Le deuxième chapitre introduit la notion de capteur en tant qu'élément primordial dans une chaîne de mesure. Les différents capteurs physico-chimiques utilisés dans le domaine de traitement des eaux propres sont particulièrement présentés. Les capteurs intelligent et logiciel en tant qu'éléments innovants dans une chaîne moderne de mesure sont introduits à cet effet.

Le troisième chapitre est particulièrement dédié aux capteurs logiciels. Considérés comme étant un moyen intéressant pouvant remplacer éventuellement des capteurs physico-chimiques, ceux-ci sont conçus à base de techniques de l'intelligence artificielle. Quatre modèles (Neuronal, Neuronal-RBF, SVM et ELM) à base de la conception de ces capteurs sont donc étudiés et évalués dans ce cadre. Les aspects théoriques et fondements de l'apprentissage statistique de ces techniques sont ainsi décrits. Les formulations des techniques SVM et ELM appliquées à la régression sont à cet effet détaillées et présentées.

Le quatrième et dernier chapitre présente plutôt une application réelle justifiant la conception et la mise en œuvre d'un capteur logiciel dédié à la mesure du Chlore (Cl) et utilisé dans le contrôle de potabilité de l'eau. La technique la mieux adaptée pour la conception de ce capteur repose sur une évaluation préliminaire des performances de chacune des techniques présentées dans cette étude. Plusieurs paramètres sont évalués à cet effet, tels que le temps d'apprentissage, l'erreur d'entraînement etc.

Une conclusion générale en fin de cette thèse est prévue, elle retrace les différentes étapes réalisées et souligne les perspectives envisagées.

CHAPITRE I

CONTROLE ET PRODUCTION DES EAUX PROPRES

INTRODUCTION

L'usage de l'eau à des fins alimentaires nécessite une excellente qualité physico-chimique et microbiologique. L'eau distribuée par réseau constitue un des produits alimentaires les plus contrôlés et ses analyses sont réalisées depuis son origine jusqu'au robinet. L'eau de distribution doit nécessairement répondre aux exigences de la qualité. Ainsi, elle ne doit contenir aucun micro-organisme, aucun parasite ni aucune substance constituant un danger potentiel pour la santé des personnes ; elle doit également être conforme vis-à-vis d'un ensemble de normes de potabilité [35]. Le but principal dans les usines de production, est de connaître les différents paramètres de l'eau brute pour décider de son état, et chercher par la suite les méthodes pour la rendre une eau potable. La surveillance permanente de la qualité de l'eau à travers des mesures qualifiables et quantifiables avec le bon fonctionnement des installations de traitement, est donc exigée dans le souci de ne pas ralentir la production et assurer en même temps un niveau de qualité élevé et uniforme [36,37].

Ce premier chapitre se veut introductif, il a pour objectif de donner une idée générale du domaine des eaux propres. La première partie est consacrée à la définition de l'eau propre ainsi que de ses différentes propriétés. La chaîne de production et de traitement d'eau potable la plus courante est particulièrement présentée. Sont cités dans ce cadre, les différentes méthodes de surveillance industrielle, les installations ainsi que les équipements de traitement.

1. DEFINITION DE L'EAU PROPRE

Est une eau que l'on peut boire sans risque pour la santé. Afin de définir précisément une eau potable, des normes ont été établies qui fixent notamment les teneurs limites à ne pas dépasser pour un certain nombre de substances nocives et susceptibles d'être présentes dans l'eau. Le fait qu'une eau soit conforme aux normes, c'est-à-dire potable, ne signifie donc pas qu'elle soit exempte de matières polluantes, mais que leur concentration a été jugée suffisamment faible pour ne pas mettre en danger la santé du consommateur [8].

1.1. Types des eaux propres

1.1.1. L'eau minérale naturelle

C'est une eau d'origine souterraine, microbiologiquement saine, qui doit être tenue à l'abri de tout risque de pollution. Elle répond à des exigences de qualité microbiologique et physico-chimique strictes. Elle se distingue des autres eaux par la présence de minéraux, oligoéléments ou autres constituants, et témoigne d'une stabilité de ses caractéristiques essentielles. L'eau minérale naturelle ne peut faire l'objet que de quelques traitements autorisés par la réglementation (séparation des constituants naturellement présents, la désinfection de l'eau est interdite).

1.1.2. L'eau de source

C'est une eau d'origine souterraine, microbiologiquement saine et qui doit être protégée contre les risques de pollution. L'eau de source, à l'émergence et conditionnée, répond aux mêmes exigences de qualité microbiologique que l'eau minérale naturelle ; et l'eau de source conditionnée répond aux mêmes exigences de qualité physico-chimique et radiologique que l'eau du robinet. Au même titre que l'eau minérale naturelle, l'eau de source ne peut faire l'objet que de quelques traitements autorisés par la réglementation (séparation des constituants naturellement présents, la désinfection de l'eau est interdite).

1.1.3. L'eau rendue potable par traitements

C'est une eau d'origine souterraine ou superficielle. L'eau rendue potable par traitements, à l'émergence et conditionnée, répond à des exigences de qualité microbiologique strictes. L'eau rendue potable par traitements conditionnée, répond aux mêmes exigences de qualité physico-chimique et radiologique que l'eau du robinet.

2. CARACTERESTIQUES DES EAUX

2.1. Les paramètres organoleptiques

Les paramètres organoleptiques sont liés à la couleur, à la saveur et à l'odeur de l'eau. C'est d'ailleurs principalement pour ces aspects que le consommateur apprécie la qualité d'une eau. Néanmoins, il faut tout de même noter que ce sont des paramètres de confort. En

effet, ces critères n'ont pas de valeur sanitaire directe. Une eau peut être trouble, colorée ou avoir une odeur particulière et néanmoins être consommable.

2.2. Les paramètres physico-chimiques

La qualité physico-chimique de l'eau informe sur la localisation et l'évaluation d'un niveau de pollution, en fonction d'un ensemble de paramètres. Basée sur des valeurs de référence, elle s'apprécie à l'aide de plusieurs paramètres [38] :

Qualité physique : matière en suspension (MES), turbidité (TU), transparence, température (T°), conductivité (C) et salinité ;

Qualité chimique : pH, sels minéraux, OD, nutriments (nitrites, nitrates, ammonium, phosphate, silice), pesticides, etc.

Ces paramètres permettent d'acquérir des connaissances de base, de développer une surveillance pour détecter des perturbations et de mettre en place un suivi pour rétroagir sur la gestion.

2.2.1. La température

Elle représente une grandeur physique liée à la notion immédiate de chaud et de froid. Elle se mesure au moyen d'un thermomètre. C'est un paramètre physique qui influence considérablement la multiplication microbienne ainsi que le métabolisme. Selon la température optimale de développement, les micro-organismes sont classés en trois catégories:

- les germes mésophiles qui supportent une température moyenne comprise entre 20 et 40°C ;
- les germes psychrophiles dont la température optimale de croissance est située entre 0 et 15°C ;
- les germes thermophiles qui se multiplient préférentiellement entre 45 et 85°C.

Il faut noter que la majorité des bactéries pathogènes sont des mésophiles.

2.2.2. La dureté

La dureté ou titre hydrotimétrique correspond essentiellement à la présence de sels de calcium (Ca) et de magnésium (Mg). Elle est directement liée à la nature des terrains traversés (roches calcaires ou dolomitiques). Dans une eau naturelle, on peut distinguer différents types de dureté :

Dureté totale : somme des concentrations en Ca et Mg ;

Dureté calcique : concentration globale en Ca ;

Dureté magnésienne : concentration globale en Mg.

Les ions Ca et les ions Mg sont positifs, donc dans l'eau, ils peuvent se lier à d'autres ions négatifs ; de ce fait on peut diviser la dureté en deux catégories :

Dureté permanente ou non carbonatée : dureté qui persiste après ébullition de l'eau et qui correspond uniquement aux sels de Ca et de Mg solubles à chaud, sous forme de chlorures et de sulfates, car les hydrogénocarbonates (HCO_3^-) sont décomposés et précipitent sous forme de carbonate de Ca ;

Dureté carbonatée ou temporaire : dureté qui correspond aux sels de Ca sous forme HCO_3, et qui est la différence entre la dureté totale et la dureté permanente [37,39].

La pollution peut aussi, dans une certaine mesure augmenter la dureté de l'eau ; c'est le cas dans certaines régions où le chlorure de calcium ($CaCl_2$) est utilisé massivement pour faire fondre la neige. La dureté calcique et la dureté magnésienne peuvent s'exprimer en mg/l de Ca et en mg/l de Mg.

Ces diverses teneurs ne peuvent évidemment pas être additionnées et pour calculer la dureté totale, on doit exprimer la concentration de chaque ion participant dans une unité commune.

2.2.3. Le potentiel d'Hydrogène

Cette mesure physico-chimique appelée pH, est effectuée à l'aide d'un pH-mètre. Elle permet de savoir si l'échantillon d'eau est acide, basique ou neutre. L'échelle des pH varie de 0 à 14. Le pH de neutralité étant 7. Pour rendre une solution proche de la neutralité, il suffit d'ajouter une base si celle-ci est acide, ou d'ajouter un acide si elle est basique. Notons que

l'on peut avoir une idée des pHs en trempant dans le liquide à tester, un papier spécialement étudié à cet effet et qui change de couleur avec le pH. Les eaux en sortie de station d'épuration doivent avoir un pH aux alentours de 7,5.

2.2.4 L'alcalinité

L'alcalinité d'une eau correspond à sa capacité à réagir avec les ions hydrogène (H) qui est due à la présence des ions HCO_3^-, carbonate (CO_3^{-2}) et hydroxyde (OH^-). Autrement dit, on est amené à considérer deux types d'alcalinité qui correspondent, encore une fois à deux bornes de pH :

Alcalinité au virage du rouge de méthyle : elle correspond à l'alcalinité totale au pH de 4.5, ce qui revient à déterminer les ions HCO_3^-, CO_3^{-2} et OH^-. Cette alcalinité se nomme également titre alcalimétrique complet (TAC).

Alcalinité au point de virage de la phénophtaléine (alcalinité composite) : elle correspond à l'alcalinité entraînée par les ions OH^- et à la moitié des ions CO_3^{-2}. Cette alcalinité composite est nulle pour une eau dont le pH est inférieur ou égal à 8.3. L'alcalinité composite se nomme également titre alcalimétrique (TA) [40,41].

2.2.5. Les Matières en Suspension

Les matières en suspension comprennent toutes les matières minérales ou organiques qui ne se solubilisent pas dans l'eau. Elles incluent les argiles, les sables, les limons, les matières organiques et minérales de faible dimension, le plancton et autres micro-organismes de l'eau. La quantité de MES varie notamment selon les saisons et le régime d'écoulement des eaux. Ces matières affectent la transparence de l'eau et diminuent la pénétration de la lumière et par suite la photosynthèse. Par ailleurs, les MES peuvent accumuler des quantités élevées de matières toxiques (métaux, pesticides, huiles minérales, hydrocarbures aromatiques polycycliques etc.).

2.2.6. La conductivité

La conductivité mesure la capacité de l'eau à conduire le courant entre deux électrodes. La plupart des matières dissoutes dans l'eau se trouvent sous forme d'ions chargés électriquement. La mesure de la conductivité permet donc d'apprécier la quantité de sels dissous dans l'eau. La conductivité est également fonction de la température de l'eau : elle est

plus importante lorsque la température augmente. Les résultats de mesure doivent donc être présentés en termes de conductivité équivalente à 20 ou 25°C. Les appareils de mesure utilisés sur le terrain effectuent en général automatiquement cette conversion. Ce paramètre doit impérativement être mesuré sur le terrain. La procédure est simple et permet d'obtenir une information très utile pour caractériser l'eau. Comme la température, des contrastes de conductivité permettent de mettre en évidence des pollutions, des zones de mélanges ou d'infiltration. La conductivité est également l'un des moyens de valider les analyses physico-chimiques de l'eau : la valeur mesurée sur le terrain doit être comparable à celle mesurée au laboratoire.

2.2.7. L'Oxygène Dissous

L'eau absorbe autant d'oxygène que nécessaire pour que les pressions partielles d'oxygène dans le liquide et dans l'air soient en équilibre. La solubilité de l'oxygène dans l'eau est fonction de la pression atmosphérique (donc de l'altitude), de la température et de la minéralisation de l'eau : la saturation en O_2 diminue lorsque la température et l'altitude augmentent. L'OD est un paramètre essentiel dans le maintien de la vie, et donc dans les phénomènes de dégradation de la matière organique et de la photosynthèse. C'est un paramètre utilisé essentiellement pour les eaux de surface. Une eau très aérée est généralement sursaturée en oxygène (torrent), alors qu'une eau chargée en matières organiques dégradables par des micro-organismes est sous-saturée. En effet, la forte présence de matières organiques, dans un plan d'eau par exemple, permet aux micro-organismes de se développer tout en consommant de l'oxygène. L'OD est donc un paramètre utile dans le diagnostic biologique du milieu d'eau.

2.2.8. Résidu sec

La détermination du résidu sec sur l'eau non filtrée permet d'évaluer la teneur en matières dissoutes et en suspension ; la mesure après filtration correspond aux matières dissoutes. Ces valeurs peuvent être recoupées à partir de la conductivité et permettent d'apprécier la minéralisation de l'eau.

2.3. Les paramètres indésirables

Certaines substances dites indésirables parce qu'elles peuvent créer soit un désagrément pour le consommateur : goût et odeur (matières organiques, phénols, fer etc.), couleur (fer, manganèse etc.), soit causer des effets gênants pour la santé (nitrates, fluor etc.). On surveille donc prioritairement la contamination des eaux par des matières organiques (mesurée par l'oxydabilité au permanganate de potassium), la concentration en ammonium, la présence de nitrites et de nitrates et la concentration en fer.

2.3.1. Les matières organiques

L'oxydabilité au permanganate de potassium ($KMnO_4$) : correspond à l'oxydation des matières organiques (MO) par le $KMnO_4$ de concentration connue en milieu acide selon un protocole normalisé. Le résultat est exprimé en mg/l d'O_2. L'eau de bonne qualité, pauvre en matières organiques, a une oxydabilité inférieure ou égale à 1 mg/l d'O_2. Les eaux de surface présentent en général une oxydabilité comprise entre 3 et 12 mg/l d'O_2. La limite pour les eaux brutes est de 10 mg/l d'O_2. La référence de qualité est de 5 mg/l d'O_2 [41,42].

Le Carbone Organique Total (COT) : correspond à une mesure globale de la matière organique particulaire et dissoute présente dans l'eau. Il est souvent utilisé comme indicateur non spécifique de la qualité de l'eau. La méthode rapide de détermination du COT par méthode colorimétrique constitue une méthode facile pour garantir des résultats satisfaisants. Le COT des eaux superficielles comprend en moyenne de 90-95 % de carbone organique dissous (COD) et 5 à 10 % de carbone organique particulaire (COP) [43]. La mesure de COT fait intervenir une aération de l'eau (pour éliminer le CO_2 dissous), puis une acidification (pour déplacer les formes carboniques) et enfin une minéralisation par oxydation ou par combustion [44]. La matière organique est ainsi complètement minéralisée en CO_2 que l'on peut détecter en spectrophotométrie infrarouge. Les eaux souterraines sont généralement pauvres en COT. Les eaux de surface présentent des valeurs comprises entre 2 et 12 mg/l. La référence de qualité est de 2 mg/l. Le COT est une méthode rapide qui présente plusieurs avantages ; on peut analyser l'eau quelle que soit sa qualité (de forte salinité, acide, basique, ou contient des composés toxiques). Malgré les avantages offerts, cette mesure exige un appareillage coûteux [45,46].

Charge en matières organiques : demande biochimique en oxygène (DBO) et demande chimique en oxygène (DCO) : deux méthodes permettant d'évaluer la quantité en matières organiques présentes dans l'eau sont généralement utilisées : DBO et DCO. Ces deux méthodes se basent sur la différence entre la teneur en OD initiale et la teneur en OD finale, et ce après oxydation de la matière organique présente dans un échantillon d'eau. La DBO représente la quantité d'oxygène utilisée par les bactéries pour décomposer partiellement ou oxyder totalement les matières biochimiques oxydables présentes dans l'eau, qui constituent d'ailleurs leur source de carbone (graisses, hydrates de carbone, tensioactifs, etc.). En ce qui concerne les eaux domestiques, environ 70% des composés organiques sont généralement dégradés après 5 jours et la dégradation est pratiquement complète au bout de 20 jours. L'indicateur utilisé est généralement la DBO_5 qui correspond à la quantité d'oxygène (exprimée en mg/l) nécessaire aux microorganismes décomposeurs pour dégrader et minéraliser en 5 jours à une température de 20°C, la matière organique présente dans un litre d'eau polluée. Plus la DBO_5 est élevée, plus la quantité de matières organiques présentes dans l'échantillon est élevée [47].

La DCO correspond à la quantité d'oxygène nécessaire pour la dégradation par voie chimique, effectuée à l'aide d'un oxydant puissant, des composés organiques présents dans l'eau [48]. Elle permet de mesurer la teneur en matières organiques totales (excepté quelques composés qui ne sont pas dégradés), y compris celles qui ne sont pas dégradables par les bactéries. Il s'agit donc d'un paramètre important permettant de caractériser la pollution globale d'une eau par des composés organiques.

La différence entre la DCO et la DBO est due aux substances qui ne peuvent pas être décomposées biologiquement. Le rapport entre la DBO et la DCO constitue une mesure indicative de la « dégradabilité » biochimique des composés présents dans l'eau. Le rapport DCO/DBO évolue d'environ 2,5 (eau résiduaire récemment déversée) à 10-20 après décomposition totale. Dans ce dernier cas, on parle d'une eau bien minéralisée. Cependant, lorsque des composés toxiques sont présents, l'activité biologique est ralentie et, de ce fait, la quantité d'oxygène consommée après 5 jours est moindre. Ceci se traduit également par un rapport DCO/DBO élevé. La DBO et la DCO se mesurent en mg d'O_2 par litre.

2.3.2. L'azote

L'azote organique est également un paramètre descripteur de la pollution. Il est dosé après minéralisation en milieu sulfurique à chaud et distillation en milieu basique de l'ammoniaque ainsi obtenu. C'est un paramètre à surveiller, car il réagit avec le chlore pour former des composés organochlorés pouvant être responsables du mauvais goût de l'eau. On retrouve l'azote organique dans les eaux superficielles à des concentrations comprise entre 0.5 et 3 mg/l. C'est sous forme d'azote minéral que l'azote est essentiellement présent dans le milieu hydrique [41]. Il existe principalement sous forme ionique, à savoir l'ammoniaque (NH_4^+), les nitrites (NO_2^-) et les nitrates (NO_3^-). La pollution en ions NH_4^+ existe principalement pour les eaux de surface. Leur oxydation conduit à la formation d'ions NO_2^- . Ils sont en équilibre avec la forme gazeuse NH_3. Ces deux espèces sont très toxiques et en conséquence dangereuses pour la santé publique. Elles induisent une prolifération bactérienne sur le réseau ayant pour conséquence, la dégradation des qualités organoleptique de l'eau et la corrosion des canalisations. Mais les nitrates sont la principale source d'inquiétude. Ces ions se transforment en milieu acide faible et à hauteur de 5%, en ions nitrites. Ceux-ci sont toxiques pour l'organisme humain. Ce sont des agents vasodilatateurs puissants qui causent vertiges et hypotension. Ils s'oxydent au niveau des ions ferreux de l'hémoglobine pour redonner des nitrates. On obtient alors de la méthémoglobine contenant des ions ferriques incapables de transporter les molécules d'oxygène [45].

La valeur limite de l'ammonium pour les eaux brutes est de 4 mg/l. La référence de qualité est de 0.2 mg/l et on n'en retrouve pas généralement dans les eaux souterraines. Bien que l'azote ammoniacal n'ait aucun effet direct sur la santé aux concentrations rencontrées dans les eaux de consommation, sa présence contribue au développement des microorganismes dans les réseaux et dans les réservoirs de distribution. Les nitrites sont présents dans les eaux superficielles, leur valeur doit être inférieure à 0.1 mg/l en sortie des installations de traitement. La limite de qualité des nitrates est de 50 mg/l pour les eaux superficielles et de 100 mg/l pour les eaux souterraines. La limite de qualité est de 50 mg/l de nitrates et 0.5 mg/l de nitrites, quelque soit l'origine de la ressource.

2.4. Les paramètres toxiques : éléments métalliques et non métalliques

De nombreux métaux sont mis en évidence dans l'eau, généralement on trouve des éléments non métalliques : arsenic, bromates, chlorites, monomères, solvants chlorés, benzène, pesticides, etc., ou des éléments métalliques comme : fer, manganèse, aluminium,

plomb, etc. Ces éléments parfois sont indésirables ou toxiques s'ils sont en dehors de leurs limites de qualité. Une pollution industrielle du captage ou une dégradation des réseaux de distribution peut entraîner la présence d'éléments toxiques dans l'eau, dangereux pour la santé en cas de consommation régulière. Les matières toxiques sont constituées de micropolluants minéraux (métaux lourds : cadmium, mercure, cuivre, zinc, nickel, chrome, cobalt, etc.) ainsi que de substances telles que les cyanures, ou des molécules organiques présentant une action d'inhibition des mécanismes biologiques**.** Leurs concentrations dans l'eau dépendent essentiellement de leur solubilité. Leurs concentrations maximales acceptables dans l'eau potable sont très faibles dans les eaux brutes et en distribution [49].

Le fer (Fe) et le manganèse (Mn) comme des exemples, ne sont pas des polluants majeurs et leur ingestion ne présente pas de dangers pour l'homme. Le Cuivre (Cu) ou le Zinc (Zn) présentent peu d'inconvénients, du moins lorsque les concentrations sont faibles. Cependant, il est nécessaire de limiter leur concertation dans les eaux potables, car ils influencent la qualité organoleptique de l'eau et peuvent être à l'origine du développement de microorganismes initiant des phénomènes de corrosion [49]. L'étape de coagulation-floculation de la filière de potabilisation est suffisante parfois pour éviter tous problèmes liés à quelques paramètres indésirables ou toxiques.

2.5. Caractéristiques biologiques

L'eau potable ne doit pas contenir de germes pathogènes (bactéries, virus, parasites, etc.) qui provoqueraient des maladies chez les consommateurs. C'est la qualité la plus importante ; car le critère de potabilité de l'eau se fixe à partir de la présence ou de l'absence de ces germes témoins [50,51].

En général, les caractéristiques physiques de l'eau doivent être comme suit :

- pas de saveur ni d'odeur ;
- coloration : inférieure à 20 unités de l'échelle colorimétrique au platine-cobalt ;
- turbidité : inférieure aux valeurs guides ;
- pH de 6,5 à 9.

3. LES NORMES DE POTABILITE

Les normes définissant une eau potable sont variables suivant la législation en vigueur et selon qu'il s'agit d'une eau industrielle ou destinée à la consommation. Il existe ainsi près de 63 critères pour une eau propre à la consommation : ces critères, décidés selon le principe de précaution maximale qui permet de protéger les personnes dont la santé est la plus fragile, portent sur plusieurs paramètres fixés [39]**.**

Voici les principaux paramètres définissant la qualité de l'eau potable :

Les paramètres organoleptiques sont liés à la couleur, à la saveur et à l'odeur de l'eau. En résumé, afin d'obtenir une qualité organoleptique, l'eau doit être agréable à boire, claire, fraîche et sans odeur. C'est d'ailleurs principalement pour ces aspects que le consommateur apprécie la qualité d'une eau. Néanmoins, il faut tout de même noter que ce sont des paramètres de confort. En effet, ces critères n'ont pas de valeur sanitaire directe. Une eau peut être trouble, colorée ou avoir une odeur particulière et néanmoins être consommable.

Les paramètres physico-chimiques correspondent aux caractéristiques de l'eau telles que le pH, la température et la conductivité. Ils concernent donc tout ce qui est relatif à la structure naturelle de l'eau et délimitent des concentrations maximales pour un certain nombre d'éléments, notamment des ions comme les chlorures, le potassium et les sulfates.

Les paramètres microbiologiques permettent de vérifier que l'eau ne contient pas de germes pathogènes (c'est-à-dire bactéries, virus, parasites etc.) qui provoqueraient des maladies chez les consommateurs. C'est le critère le plus important concernant la potabilité de l'eau.

Les paramètres concernant les substances indésirables concernent des substances dont la règlementation tolère la présence en faible quantité. On peut citer par exemple la teneur maîtrisée en fluor, en nitrates, en nitrites, en sels minéraux etc.

Les paramètres concernant des substances toxiques telles que les pesticides, les métaux lourds comme le plomb ou le chrome. Les teneurs tolérées sont extrêmement faibles car ce sont des poisons mortels pour l'homme.

- **Qualité des eaux superficielles en Algérie**

Les eaux superficielles constituent une part importante des ressources hydrauliques en Algérie. Le contrôle et la surveillance de l'eau se font par l'analyse systématique et régulière de certains paramètres physico-chimiques (pH, dureté, température, conductivité, etc.) qui sont analysés périodiquement par des laboratoires spécialisés tels que le laboratoire d'hydrochimie de l'Agence Nationale des Ressources Hydrauliques (ANRH) ou les laboratoires régionaux rattachés aux entreprises de production d'eau potable. D'autre part, des analyses plus spécifiques sont parfois effectuées en cas de pollution accidentelle (métaux lourds, détergents etc.). Le tableau 1.1 présente les normes de potabilité nationales par les valeurs guides ANRH, il apparaît que les eaux algériennes sont souvent minéralisées, avec une dureté assez importante, notamment pour les eaux du sud algérien [45].

Tableau. 1.1. Valeurs guides de l'ANRH et concentrations limites des directives CEE utilisées pour la production d'eau destinée à la consommation humaine [45, 52].

	Valeurs guides de qualité de l'ANRH				**Directives CEE**
Désignation	**Bonne**	**Moyenne (eau potable)**	**Mauvaise (eau polluée)**	**Très mauvaise (eau non potable)**	**Concentrations limites**
pH	6.5 – 8.5	6.5 – 8.5	8.5 – 9	> 9 et < 6.5	-
OD (%)	90 – 120	50 – 90	30 – 50	< 30	> 30%
NH_4^+ (mg/l)	0 – 0.01	0.01 – 0.1	0.1 – 3	> 3	4
NO_2^- (mg/l)	0 – 0.01	0.01 – 0.1	0.1 – 3	> 3	-
NO_3^- (mg/l)	0 – 10	10 – 20	20 – 40	> 40	50
PO_4^{3-} (mg/l)	0 – 0.01	0.01 – 0.1	0.1 – 3	> 3	-
DBO_5 (mg/l)	5	5 -10	10 – 15	> 15	-
DCO (mg/l)	20	20 – 40	40 – 50	> 50	-
MO (mg/l)	5	5 -10	10 – 15	> 15	10
RS (mg/l)	300 - 1000	1000 – 1200	1200 – 1600	> 1600	-
Température (°C)	5 – 15	15 – 30	30	30	-
Nombre de germes	0 – 10	10 - 100	10000	> 10000	-

4. TRAITEMENT DES EAUX

Qu'elles soient d'origine souterraine ou superficielle, les eaux utilisées pour l'alimentation humaine sont rarement consommables telles quelles. Il est souvent nécessaire de leur appliquer un traitement plus ou moins sophistiqué, ne serait-ce qu'une désinfection dans le cas des eaux souterraines.

4.1. Importance de traitement

Une analyse régulière de l'eau est importante pour les raisons suivantes [40] :

- elle permet de définir les problèmes existants ;
- elle garantit une eau qui convient à l'utilisation prévue ;
- elle garantit une eau potable sûre ;
- elle permet de vérifier l'efficacité des procédés dans un système de production.

La qualité d'une réserve d'eau peut changer au fil du temps et même subitement. Si l'apparence, l'odeur et le goût de l'eau restent les mêmes, le changement de qualité risque de passer inaperçu. La seule façon de connaître la salubrité de l'eau potable, est de la faire analyser. Comme les bactéries, les parasites et les virus nuisibles sont invisibles à l'œil nu, une eau au goût et à l'apparence agréables n'est pas forcément potable. Ces microbes, qui vivent parfois dans l'eau souterraine et de surface, risquent de causer rapidement des maladies chez les humains qui consomment l'eau sans la traiter adéquatement. Certains contaminants chimiques que l'on retrouve dans les réserves d'eau peuvent causer des problèmes de santé à long terme, qui n'apparaissent que des années après la consommation. Une analyse fréquente de l'eau permet de déterminer le niveau de salubrité de l'eau et de vérifier si le système de traitement a atteint un degré de purification satisfaisant. Plusieurs analyses disponibles sont utiles pour déterminer la salubrité et la sûreté des réserves d'eau. L'analyse de base de l'eau potable comprend plusieurs aspects d'analyse tels que celui des bactéries coliformes, des nitrates, du pH, du sodium, du chlorure, du fluorure, des sulfates, du fer, du manganèse, des matières totales dissoutes et celui de la dureté [40].

- Les nitrates sont des contaminants que l'on retrouve couramment, surtout dans l'eau souterraine. Une eau à forte teneur en nitrates risque d'être particulièrement dangereuse pour les bébés de moins de six mois, car les nitrates nuisent au transport de l'oxygène dans le sang.

- Les ions comme le sodium, le chlorure, les sulfates, le fer et le manganèse peuvent conférer à l'eau un goût ou une odeur désagréable.

- Une quantité excessive de sulfates risque d'avoir un effet laxatif et de provoquer une irritation gastro-intestinale.

- Le fluorure est un oligo-élément essentiel, mais en trop grandes quantités, il risque de causer des problèmes dentaires.

- Les matières totales dissoutes représentent la quantité de substances inorganiques (le sodium, le chlorure et les sulfates) dissoutes dans l'eau. Une eau à forte teneur en matières totales dissoutes acquiert un goût désagréable [40].

- L'analyse des bactéries coliformes indique la présence de microorganismes potentiellement nocifs pour la santé humaine. L'eau potable ne doit contenir aucun microorganisme pathogène (bactéries, virus et protozoaires) et être exempte de toute bactérie indiquant une pollution par excréments. Une eau qui est porteuse d'un seul Escherichia Coli par 100 ml devient potentiellement dangereuse [45]. Ce point est important, car le critère de potabilité de l'eau se fixe à partir de la présence ou de l'absence de ces germes témoins [50,51]. Différentes maladies sont transmises par l'eau en présence des différents microorganismes (bactéries, virus et protozoaires). Parmi ces maladies on trouve par exemple : la gastro-entérite, la pharyngite, la méningite encéphalique, correspondant à des microorganismes pathogènes : Escherichia coli, Adénovirus, Naegleria grubéria respectivement.

Si on soupçonne la présence d'un contaminant particulier dans l'eau, on peut procéder à d'autres analyses. On analyse parfois l'eau souterraine afin d'y détecter la présence d'arsenic, de sélénium ou d'uranium, par exemple. On peut aussi évaluer la contamination de l'eau de surface ou souterraine par les pesticides. Les réserves d'eau domestique doivent faire l'objet d'une analyse au moins une fois par an. L'eau potable provenant de puits peu profonds ou de réserves de surface, plus sujette à la contamination que l'eau souterraine ; doit être analysée plus souvent (chaque saison).

Il est important d'analyser l'eau potable au robinet et à la source. Ces deux analyses permettent de vérifier l'efficacité du système de traitement et de détecter tout changement dans la qualité de l'eau à la source. Il est important de souligner que l'eau avant qu'elle parvienne au consommateur, subi des traitements plus ou moins poussés, elle est stockée,

acheminée, puis distribuée. L'eau potable est donc une denrée rare et précieuse qui a un coût, qu'il ne faut pas gaspiller. Par ailleurs, il faut garder à l'esprit qu'elle est produite à partir de ressources naturelles qu'il convient de protéger [40].

4.2. Chaîne de traitement

La transformation d'une eau de surface en une eau propre à la consommation nécessite de faire appel à un ensemble de procédés de traitement extrêmement divers qu'il faut assembler dans un ordre déterminé afin de fournir un produit fini conforme aux normes de potabilité.

L'efficacité du traitement adopté dépendra, de plus, de la façon dont sera conduite l'exploitation de l'usine. Pour atteindre l'objectif souhaité, l'exploitant devra d'une part respecter certains principes élémentaires pour assurer le contrôle du processus de traitement et le contrôle de l'eau traitée, et d'autre part disposer d'un certain nombre de moyens techniques et humains. Cela étant, les techniques existantes se développent et se perfectionnent. Cette évolution se traduit par la recherche de procédés nouveaux et par la mise en place de nombreux automatismes. Nous présentons ci-dessous la filière la plus complète et la plus courante du traitement des eaux de surface, sans entrer dans les détails particulièrement fins des procédés. On peut admettre que la chaîne habituelle complète comporte cinq grandes étapes (Figure 1.1) [53].

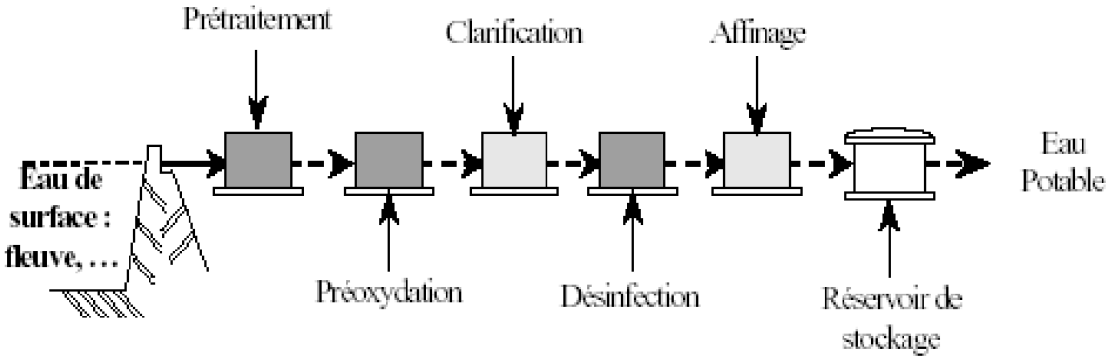

Fig.1.1. Chaîne de traitement des eaux de surface [53].

4.2.1. Prétraitement

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. Dans le cas d'une eau potable, les prétraitements sont principalement de deux types, à savoir, le dégrillage et le tamisage. Le dégrillage, en tant que premier poste de traitement, permet de protéger les ouvrages avals de l'arrivée de gros objets susceptibles de provoquer des bouchages dans les différentes unités de traitement. Ceci permet également de séparer et d'évacuer facilement les matières volumineuses charriées par l'eau brute, qui pourraient nuire à l'efficacité des traitements suivants, ou en compliquer l'exécution. Le dégrillage est avant tout destiné à l'élimination de gros objets : morceaux de bois, etc. Le tamisage permet par contre d'éliminer des objets plus fins que ceux éliminés par le dégrillage. Il s'agit de feuilles ou de morceaux de plastique par exemple [53].

4.2.2. Pré-oxydation

Après l'étape de prétraitement, on peut y adjoindre une pré-oxydation ; procédé chimique essentiel à tout traitement des eaux, elle est aussi incluse en fin de filière de désinfection. On a donc une eau relativement propre mais qui contient encore des particules colloïdales en suspension et des matières organiques en solution non dangereuse [53]. Cette étape peut être faite de trois façons différentes :

- ajout de chlore ou pré-chloration ;
- ajout de dioxyde de chlore ;
- ajout d'ozone ou pré-ozonation.

Le chlore est le réactif le plus économique. Il présente l'inconvénient de former des composés organochlorés avec certains micropolluants qui peuvent avoir des effets nocifs sur la santé humaine (le goût et l'odeur sont désagréables). Il faut donc éviter une pré-chloration dans le cas d'une eau brute fortement chargée en matière organiques [54].

L'oxydation par le dioxyde de chlore n'a pas les inconvénients de l'oxydation par le chlore, il s'applique quelque soit le type d'eau à traiter (pas d'influence du pH), et son action sur la matière organique naturelle conduit à des productions beaucoup plus faibles de composés organo- halogénés indésirables responsables entre autres des goûts et odeurs. En effet, l'utilisation de ce produit, lui aussi, a des inconvénients non négligeables. Cependant, le dioxyde de chlore est sans effet sur l'ammonium. Il se décompose à la lumière, ce qui entraîne une augmentation importante du taux de traitement à appliquer en période d'ensoleillement.

Bien qu'il soit un très bon oxydant pour le Fer et le Manganèse, mais il ne peut pas être une solution économique. Il représente une alternative intéressante par rapport au chlore au point de vue qualité de l'eau.

L'ozone est énormément utilisé pour le traitement de l'eau potable grâce à ses excellentes qualités de désinfection et d'oxydation. L'ozone peut être utilisé pour différents objectifs dans les systèmes de traitement, tels que pour une pré-oxydation, une oxydation intermédiaire, une amélioration de clarification ou une désinfection finale. Généralement, il est recommandé d'utiliser l'ozone pour la pré-oxydation avant la filtration. Cette combinaison a plusieurs avantages : l'élimination de la matière organique et inorganique, l'élimination des micropolluants, tels que les pesticides, l'élimination des goûts et des odeurs et la diminution du taux de traitement (taux de coagulation) dans le procédé de clarification. Néanmoins, certaines algues résistent à l'ozone.

4.2.3. Clarification

La clarification est une étape indispensable pour les eaux de surface et les eaux souterraines provenant de plateaux calcaires. Elle permet d'obtenir une eau limpide par élimination plus poussée des particules en suspension (sable, limons, débris organiques), des matières colloïdales plus petites (argiles fines, bactéries) et d'une partie des matières dissoutes (matières organiques, sels etc.). La clarification combine des procédés de coagulation floculation, de décantation ou flottation et de filtration.

Le contrôle de cette opération est essentiel pour trois principales raisons [53] :

- la maîtrise de la qualité de l'eau traitée en sortie (abattement de la turbidité) ;
- le contrôle du coagulant résiduel en sortie ;
- la diminution des contraintes et des coûts de fonctionnement.

4.2.3.1. Coagulation - Floculation

Pour éliminer les matières colloïdales et les particules en suspension présentes dans les eaux brutes, on utilise des procédés couplés de coagulation et de floculation. Ils consistent à agréger les particules afin de les extraire plus facilement. L'introduction d'un coagulant, généralement un sel de fer ou d'aluminium, permet de réduire les forces de répulsion

électrostatique qui existent entre les particules. Ces particules déstabilisées se rassemblent pour former de petits agrégats.

Après l'étape de coagulation, la floculation a pour objectif d'accroître la cohésion des particules par agitation de l'eau. Elle conduit à la formation d'amas de plus en plus volumineux appelés "flocs". Pour améliorer la floculation, on peut ajouter un floculant : minéral (silice activée), polymère naturel (alginate ou amidon) ou polymère de synthèse (à base d'acrylamide). L'extraction des "flocs" se fait ensuite par décantation ou flottation. Sous l'effet de son poids, le floc se dépose au fond de la cuve de décantation. Il est possible d'accélérer sa décantation en le lestant avec du micro sable.

À l'inverse de la décantation, la flottation consiste à faire remonter les flocs en utilisant de l'air sous pression. Les flocs s'accrochent aux bulles d'air qui les entraînent à la surface. Ils sont ensuite récupérés par un système de raclage.

4.2.3.2. La filtration sur sable

Elle contribue à l'élimination des particules en suspension de certains parasites et des précipités de fer, aluminium et manganèse ainsi qu'à une réduction de la matière organique.

La filtration rapide : C'est la technique la plus répandue. Elle sera d'autant plus efficace que les grains de sable seront fins. On distingue le filtre monocouche (sable) du filtre bicouche (sable anthracite ou charbon actif) qui retient davantage les particules en suspension. La filtration sur lit de sable est efficace et simple. C'est pourquoi elle s'est imposée, en raison des énormes volumes d'eau à filtrer.

La filtration biologique lente : Elle nécessite l'usage d'une couche de sable très fin. L'écoulement est alors suffisamment lent pour permettre le développement d'un bio-film (algues, bactéries) à la surface du sable. L'activité épuratrice de ce bio-film permet la dégradation partielle de la matière organique, des bactéries, ainsi que des virus et parasites.

4.2.4. Oxydation – Désinfection

En fin de traitement, la désinfection permet l'élimination de tous les microorganismes pathogènes (bactéries et virus) présents dans l'eau à la sortie des usines. Elle doit être effectuée de manière à maintenir un résiduel bactériologique sur tout le réseau de distribution afin d'éviter toute dégradation de la qualité de l'eau par prolifération de microorganismes. On

utilise pour cela soit un désinfectant chimique comme le chlore, le dioxyde de chlore ou l'ozone, soit des rayonnements ultraviolets (UV). L'efficacité de ces désinfectants qui sont classés dans l'ordre décroissant : UV, ozone, dioxyde de chlore, puis chlore, dépend de plusieurs facteurs dont [55] :

- les désinfectants ;
- le temps de contact du réactif avec l'eau supposée contaminée ;
- la concentration résiduelle en désinfectant ;
- du pH, de la température, des matières organiques, de la concentration en ammoniaque, et de la turbidité.

Pour chaque type de traitement, il est nécessaire de contrôler les divers paramètres physico-chimiques suivants [53] :

- La température : lorsqu'elle augmente, la prolifération microbiologique s'accélère. Par ailleurs, la consommation en désinfectant est plus importante. Il est donc nécessaire d'être vigilant sur l'évolution de ce paramètre et d'ajuster les consignes du désinfectant en conséquence.

- Le pH : l'efficacité du chlore présent dans l'eau varie avec le pH. Il est donc indispensable de mesurer ce paramètre en même temps que le chlore libre afin d'évaluer au mieux l'efficacité du traitement.

- La turbidité : elle caractérise la présence des particules dans l'eau. La présence de turbidité est le signe d'un traitement incomplet. Les particules non retenues lors du traitement peuvent « véhiculer » des microorganismes qui seront plus difficiles à inactiver par les désinfectants. Enfin, la turbidité révèle la présence de matières en suspension qui pourront former des dépôts dans le réseau, lesquels favorisent, à l'abri de l'action des désinfectants, la prolifération microbienne.

- L'ammonium (NH_4^+) et le NO_2^- : ils sont consommés par le chlore et peuvent être considérés comme des indicateurs de contamination.

Ces contrôles sont le plus souvent effectués en entrée ou en sortie des contacteurs de désinfection, mais aussi en différents points du réseau.

4.2.5. Affinage

L'affinage a pour but de dégrader les matières organiques et d'éliminer certains micropolluants. Il améliore les qualités organoleptiques de l'eau (saveur, odeur, limpidité).

L'affinage par charbons actifs : Les charbons actifs sont obtenus par traitement thermique, jusqu'à 1000°C de charbons naturels (anthracite, tourbe) ou de végétaux (bois, noix de coco). Ce traitement va augmenter considérablement leur surface interne et leur porosité. Ainsi, le charbon actif sera capable de fixer les composés organiques dissous qui ont échappé à la clarification ainsi que la majeure partie des hydrocarbures, pesticides, substances pharmaceutiques, métaux lourds etc. Le charbon actif en poudre est ajouté dans le décanteur après la coagulation, juste avant la floculation, alors que le charbon actif en grains est utilisé sous forme de lit filtrant lors de la filtration.

L'affinage par stripping : Pour extraire les gaz dissous dans l'eau, on applique un contre-courant de gaz d'entraînement. Le stripping retient principalement les composés volatils (benzène, trihalométhane, composés soufrés responsables d'odeurs etc.).

4.3 Les nouveaux procédés

- **La filtration membranaire**

Les procédés membranaires permettent un traitement de l'eau qui évite un certain nombre d'étapes de clarification et limite la désinfection chimique. L'eau circule au travers des membranes. Celles-ci sont constituées de longues fibres creuses et poreuses d'environ 1 mm de diamètre sur 1,30 m de long, assemblées en faisceaux dans une cartouche cylindrique.

Elles peuvent être de nature organique (polymère de synthèse) ou de nature minérale (type céramique). Il existe quatre procédés membranaires classés en fonction de la taille des microorganismes ou des composés chimiques arrêtés par la membrane.

Les membranes de microfiltration : permettent la rétention des particules de taille supérieure à 0,1 µm, c'est-à-dire les bactéries, les parasites, les levures, les matières colloïdales etc.

Les membranes d'ultrafiltration : arrêtent en plus la quasi-totalité des virus de taille plus petite de l'ordre de 0,01 µm.

Les membranes de nanofiltration : retiennent des particules encore plus petites de l'ordre de 0,001 µm, comprenant la plupart des pesticides mais aussi les ions bivalents comme le calcium et le magnésium.

L'osmose inverse : utilise des membranes plus denses qui stoppent aussi des particules de l'ordre de 0,0001 µm. Cette dernière technique est appliquée au dessalement de l'eau de mer et à la production d'eau ultra pure et d'eau de process pour l'industrie et le secteur médical.

5. LA SURVEILLANCE DE LA QUALITE DES EAUX

Quand on parle de surveillance des eaux potables, il s'agit en fait de connaître l'état de l'eau en continu (à chaque instant) à partir des différents paramètres ayant trait à sa qualité. Traditionnellement, on procède généralement à un certain nombre de mesures utiles des différents paramètres descripteurs de l'eau brute pour le test de sa qualité tels que : le contrôle bactériologique, le contrôle de désinfection et le contrôle physico-chimique, pour décider enfin sur son état propre après avoir appliqué les techniques et méthodes pour la rendre potable. Ces méthodes sont déterminées à l'aide d'un essai expérimental appelé Jar-test.

L'inconvénient de cette technique est qu'elle nécessite de façon non stop des interventions et des déplacements sur site de l'opérateur. Elle ne permet pas en fait de suivre finement l'évolution de la qualité de l'eau brute et a le désavantage d'avoir un temps de réponse relativement long. En voici donc tout l'intérêt de disposer d'un contrôle automatique de ce procédé pour une meilleure efficacité de traitement et une réduction des coûts d'exploitation. Le contrôle de l'eau au niveau des usines de production doit donc se faire de façon immédiate en se basant sur une surveillance continue des paramètres de la qualité à travers toute une chaîne de traitement qui assure à la fois la surveillance et le contrôle. L'automatisation de ces procédés pour la surveillance va permettre de s'affranchir de certaines faiblesses humaines, d'accroitre la fiabilité et d'améliorer les conditions d'exploitation et des performances de l'installation par la suppression d'un maximum de tâches répétitives et pénibles pour l'opérateur. Automatiser les procédés de traitement peut donc avoir des objectifs divers, les plus fréquents sont d'augmenter les performances du système de production, de garantir la qualité du produit, de la surveillance des installations et des équipements de traitement, de diminuer les coûts de production et d'améliorer la sécurité de l'installation industrielle et de son environnement. De plus, l'automatisation couplée à un stockage d'informations, permet d'envisager des études statistiques de données recueillies ouvrant la voie à des études de modélisation et donc d'optimisation des procédés de traitement [8].

Plus récemment, l'automatisation et la supervision par ordinateur ont conduit à une augmentation du confort en permettant une meilleure maîtrise même à distance d'un grand nombre d'informations, ce qui permet de simplifier les tâches d'exploitation, de surveillance, de maintenance et de gestion [56]. L'utilisation des techniques issues du domaine de l'intelligence artificielle apparaît alors comme la principale alternative pour aborder ces problèmes lorsqu'il est nécessaire de prendre en compte l'intervention des experts du domaine ou de traiter l'information de nature qualitative.

5.1. La supervision des procédés

L'ensemble : procédé, actionneurs, capteurs et partie de commande, constitue le système automatisé. Cette boucle s'échange des informations avec l'opérateur dont elle reçoit des consignes et à qui elle fournit des informations. Un système de supervision d'une installation consiste à réaliser cette boucle qui passe par l'opérateur. L'acquisition des données est la fonction qui relie la partie commande à la fonction de supervision. Celle-ci consiste à gérer et à surveiller l'exécution d'une opération ou d'un travail accompli par l'homme ou une machine, puis à proposer des actions correctives [56]. La surveillance est une opération de recueil en continu des signaux de commande d'un procédé afin de reconstituer l'état de fonctionnement réel. Ainsi, la surveillance utilise les données à partir des capteurs provenant du système ou procédé pour représenter l'état de fonctionnement puis en détecter les évolutions.

5.2. Méthodes de surveillance

Le développement de nouvelles technologies a introduit une complexité supplémentaire au niveau des systèmes temps réel par le fait qu'ils combinent à la fois des éléments matériels et des éléments logiciels (systèmes embarqués). Le maintien de la sécurité et d'un fonctionnement ininterrompu de ces systèmes est devenu un enjeu important. Le but de la surveillance dans un but de diagnostic d'un processus est d'identifier les causes d'un ensemble de symptômes observés qui indiquent une dégradation ou une panne de certains composants du système conduisant à un comportement anormal du système physique. Les méthodologies de surveillance sont généralement divisées en deux catégories : méthodologies de surveillance avec modèle et sans modèle [57]. De ce fait, il est souvent délicat de déterminer la méthode la plus appropriée pour la résolution du problème de surveillance. Les méthodes de surveillance fondées sur l'utilisation de modèle sont souvent préférables,

notamment dans un contexte dynamique. Ces méthodes et le type de modèle associé se différencient non seulement par la nature de la connaissance dont on dispose sur le système physique et ses défaillances, mais également par la manière dont cette connaissance est exploitée.

Les techniques avec modèle sont donc basées sur l'existence d'un modèle formel du processus et utilisent généralement des techniques de l'Automatique [58].

Les techniques sans modèle sont plutôt divisées en deux parties. La première partie correspond aux outils statistiques de traitement du signal qui sont généralement qualifiés d'outils de traitement de bas niveau, parce qu'ils sont en contact direct avec le signal capteur, et ne servent généralement que pour la génération d'alarmes brutes sans aucune information concernant leur signification. Les outils statistiques établissent des tests sur les signaux d'acquisition, tests qui ne sont capables d'assurer que la fonction de détection des défaillances. La deuxième partie est celle des techniques dites de haut niveau qui sont plutôt orientées vers la communication avec un expert. Celles-ci représentent les techniques de l'intelligence artificielle (IA), et servent comme outil de base d'aide à la décision. Leurs réponses sont donc plus élaborées que celles des techniques de bas niveau et elles sont capables de détecter, interpréter et diagnostiquer les défaillances [59].

Les deux approches Automatique et IA peuvent être combinées pour plus d'avantages. Les méthodes de l'Automatique sont par nature proches du système surveillé, puisqu'elles travaillent directement à partir de données issues des capteurs ; elles sont ainsi principalement utilisées pour la génération d'alarmes. Les méthodes de l'IA sont, elles, plus tournées vers la communication avec l'opérateur et se focalisent sur la transformation d'un ensemble d'informations brutes et non reliées entre elles en une information interprétable directement par l'opérateur chargé de la conduite ; elles sont donc utilisées pour l'interprétation des alarmes et l'aide à la décision. Dans le cas de l'utilisation des outils de l'IA, la fonction surveillance est souvent vue comme une application de reconnaissance des formes où celles-ci représentent le vecteur d'entrée composé des différents paramètres du processus. Toutefois, un problème de diagnostic peut se définir comme un problème de reconnaissance de formes. L'ensemble des états est homologue à un ensemble de classes et le vecteur forme est le vecteur des composantes des paramètres observés sur le système [60].

Une classification globale des méthodes de surveillance est montrée dans la figure 1.2, cependant de nombreuses classifications sont proposées dans la littérature [56,59] [61-63].

Ces classifications sont influencées par les domaines d'application particuliers à chaque communauté et ne sont donc pas homogènes. La partie suivante est consacrée à la classification des méthodes de surveillance, en tenant compte particulièrement des critères associés au diagnostic.

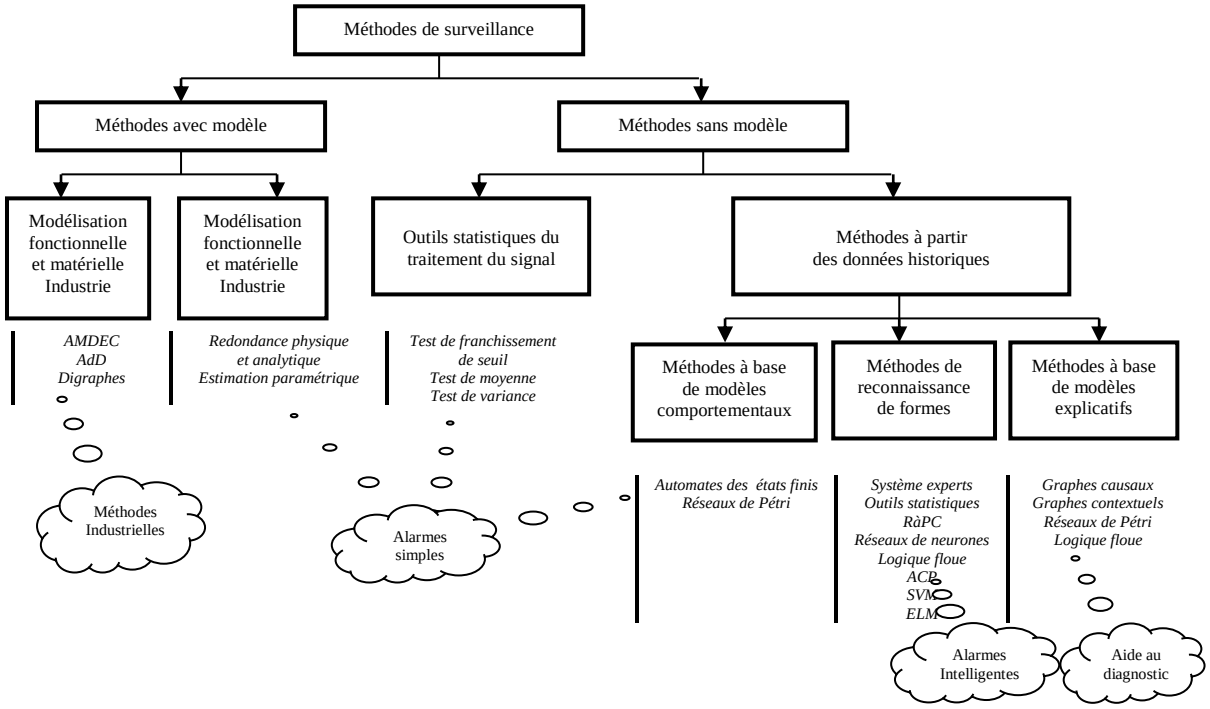

Fig.1.2. Classification globale des méthodes de surveillance [8].

5.2.1 Méthodes sans modèle

Dans certaines applications industrielles, il est difficile, voire impossible, d'obtenir le modèle du système. Cette difficulté est justifiée par la complexité accrue due à de nombreuses reconfigurations intervenantes durant le processus de production. En effet, seules les méthodes de surveillance sans modèles sont opérationnelles pour ce type d'applications. Ces méthodes de diagnostic se basent sur des informations issues d'une expérience préalable, sur des règles heuristiques ou encore sur des exemples de résolution. Parmi ces méthodes, on trouve :

La méthode du seuillage : les signaux fournis par les capteurs sont comparés avec des valeurs limites constantes ou adaptatives (évoluant en fonction du point de fonctionnement) [64,65]. Un premier niveau indique la présence probable d'un défaut alors qu'un second niveau peut en caractériser la gravité. Le franchissement d'un seuil révèle la présence d'une anomalie.

Les méthodes statistiques : les méthodes statistiques supposent que les signaux fournis par les capteurs possèdent certaines propriétés statistiques, sur lesquelles des tests de seuil sont établis [65-67]. En effet, l'étude de l'évolution de la moyenne ou de la variance d'un signal peut favoriser la mise en évidence d'une anomalie.

La reconnaissance des formes : ces méthodes reposent sur l'utilisation des algorithmes de classification des formes et des mesures (continues ou discrètes). Le fonctionnement d'un système de diagnostic par reconnaissance des formes se déroule en trois phases [68-77] :

- une phase d'analyse qui consiste à déterminer et à réduire l'espace de représentation des données et à définir l'espace de décision permettant de spécifier l'ensemble des classes possibles ;

- une phase de choix d'une méthode de décision permettant de définir une règle de décision qui a pour fonction de classer les nouvelles observations dans les différentes classes de l'ensemble d'apprentissage ;

- une phase d'exploitation qui détermine, en appliquant la règle de décision, le mode de fonctionnement du système en fonction de chaque nouvelle observation recueillie sur le processus.

Les systèmes experts : les systèmes experts utilisent une information heuristique pour lier les symptômes aux défauts. Ce sont des systèmes à base de règles qui établissent des associations empiriques entre effets et causes [65]. Ces associations sont généralement fondées sur l'expérience de l'expert plutôt que sur une connaissance de la structure et/ou du comportement du système. Leur fonctionnalité est de trouver la cause de ce qui a été observé en parcourant les règles par un raisonnement inductif par chaînage avant ou arrière.

5.2.2 Méthodes à base de modèle

Ces méthodes reposent sur une comparaison du comportement du système avec le comportement du modèle qualitatif et/ou quantitatif établi où tout écart est alors synonyme d'une défaillance [65]. Selon le type du modèle (qualitatif et/ou quantitatif), on peut distinguer deux branches de méthodes : la méthode FDI (Failure Detection and Identification) issue de la communauté automatique, s'appuyant sur des modèles quantitatifs, et la méthode DX (du Nom de l'international workshop sur les principes du diagnostic) de la communauté de l'IA utilisant des modèles qualitatifs. La dissociation entre les méthodes qualitatives et les

méthodes quantitatives n'implique pas que ces deux aspects sont disjoints. En réalité, ces deux types d'approche peuvent coexister au sein d'une même méthode de diagnostic.

CONCLUSION

Ce chapitre a servi d'introduction au domaine de contrôle et de surveillance de l'eau propre. Les paramètres descripteurs de celle-ci selon les normes de potabilité, ainsi que les procédures de son traitement et analyse, ont été évoqués. Les différentes techniques existantes actuellement dans le domaine de son contrôle et de sa surveillance, ont été présentées. Il est d'ores et déjà apparu qu'un contrôle automatique et permanant basé sur ses paramètres descripteurs ne pouvait être qu'indispensable. Dans ce domaine la mesure joue un rôle primordial, et l'utilisation de capteurs de différents types demeure incontournable. L'objet du chapitre suivant va être orienté vers le vaste domaine des capteurs, notamment ceux utilisés dans la surveillance de la qualité des eaux propres.

CHAPITRE II

CAPTEURS DE SURVEILLANCE DES EAUX PROPRES

INTRODUCTION

En instrumentation industrielle, particulièrement au niveau des chaînes de mesures, toute quantité physique ou chimique à mesurer est d'abord convertie en un signal électrique par un dispositif appelé capteur. Celui-ci joue le rôle de transducteur délivrant à sa sortie un signal comme support d'information pouvant être par la suite exploité dans différentes opérations telles que l'acquisition, l'analyse, l'enregistrement et la transmission.

Dans ce chapitre, la notion de capteur comme étant un élément fondamental d'une chaîne de mesure est présentée. Des généralités sur ces capteurs, leurs différents types, leurs structures, ainsi que leurs caractéristiques métrologiques sont introduites. Les capteurs utilisés particulièrement dans le domaine de surveillance des eaux potables sont à cet effet présentés. Pour de plus amples détails touchant l'évolution technologique de ces dispositifs, les notions de capteurs intelligent et logiciel sont de ce fait introduites.

1. GENERALITES

1.1. Définition

Un capteur est un organe de prélèvement d'information qui élabore à partir d'une grandeur physique, une autre grandeur physique de nature différente (très souvent électrique). Cette grandeur représentative de la grandeur prélevée est utilisable à des fins de mesure ou de commande. [78].

1.2. Structure d'un capteur

Les différentes parties pouvant constituer un capteur en général sont décrites dans la figure 2.1 [79].

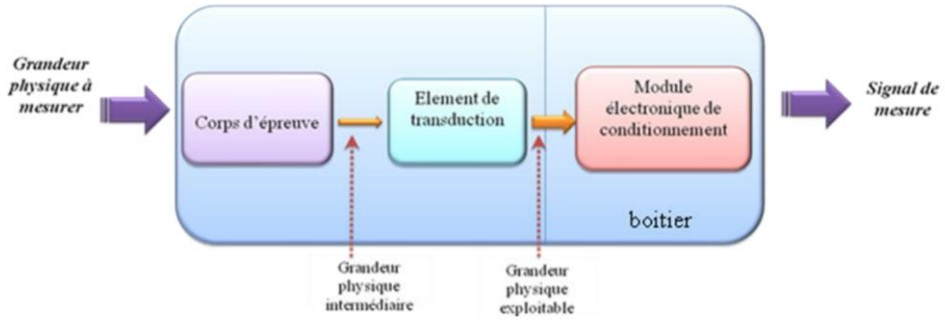

Fig.2.1. Structure d'un capteur [79].

1.2.1. Le corps d'épreuve

Le corps d'épreuve est un élément sensible qui réagit à la grandeur à mesurer. Il a pour rôle de transformer la grandeur à mesurer en une autre grandeur physique intermédiaire dite mesurable.

1.2.2. Transducteur

Le transducteur est un élément sensible lié au corps d'épreuve. Il traduit les réactions du corps d'épreuve en une grandeur électrique constituant le signal de sortie.

1.2.3. Le boîtier

Le boîtier est un élément mécanique de protection, de maintien et de fixation du capteur.

1.2.4. L'électronique de conditionnement

C'est un dispositif qui convertit le signal de sortie du capteur en un signal de mesure standard.

1.3. Types de capteurs

Suivant l'origine du signal électrique de sortie, on peut classer les capteurs en deux types:

1.3.1. Capteurs actifs

Ce type de capteurs fonctionne en générateur, dont une partie de l'énergie physique prélevée sur le mesurande est transformée directement en énergie électrique qui constitue le signal de sortie.

1.3.2. Capteurs passifs

Ils sont constitués d'un matériau spécifique dont l'impédance (résistance, capacité ou inductance) est sensible au mesurande.

1.4. Caractéristiques métrologiques d'un capteur

Les caractéristiques métrologiques d'un capteur constituent les liens effectifs entre le capteur et la grandeur qu'il mesure.

1.4.1. Etalonnage

L'étalonnage permet d'ajuster et de déterminer, sous forme graphique, la relation entre le mesurande et la grandeur électrique de sortie (figure 2.2). Très souvent l'étalonnage n'est valable que pour une seule situation d'utilisation du capteur.

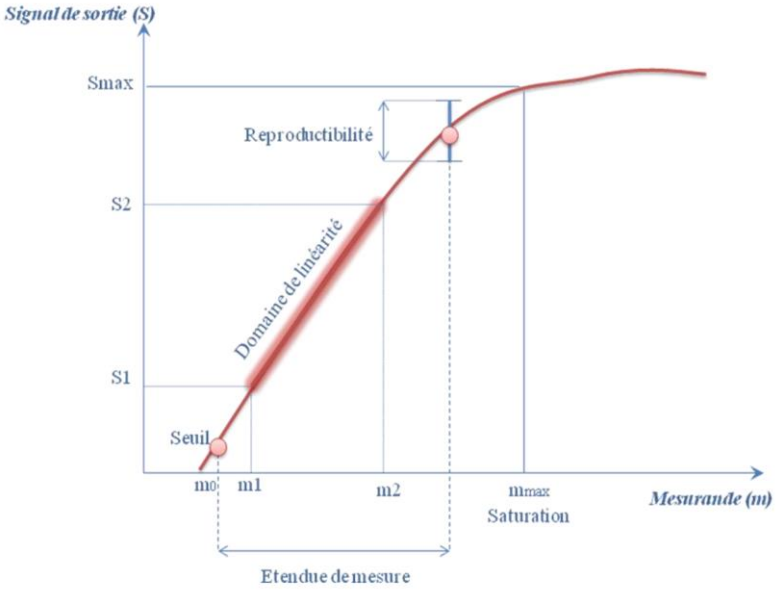

Fig.2.2. Courbe d'étalonnage d'un capteur [78].

1.4.2. Sensibilité

La sensibilité (**s**) est une caractéristique importante pour l'exploitation et l'interprétation des mesures. Elle est définie comme étant la variation du signal de sortie (S) par rapport à la variation du mesurande (m) (pente de la portion linéaire de la courbe d'étalonnage).

1.4.3. Temps de réponse

La rapidité est caractérisée par le temps que met le capteur à réagir à une variation brusque du mesurande. Cependant la valeur finale étant le plus souvent atteinte de manière asymptotique, elle correspond au temps nécessaire pour que le capteur délivre une certaine portion α de la pleine amplitude du signal. Le temps de réponse noté t_α est tel que α vaut généralement 90%. La connaissance du temps de réponse d'un capteur est un élément essentiel lors de la réalisation de mesures.

1.4.4. Précision, Sensibilité, Gamme et Etendue de mesure

La **précision** d'un capteur représente le pourcentage d'erreur de son signal de sortie par rapport à la vraie valeur ; plus celle-ci est grande et plus le coût du capteur serait important. C'est pourquoi il est essentiel de connaître cette caractéristique nécessaire à l'application afin d'optimiser les coûts. La classe d'un appareil de mesure correspond à la valeur du rapport entre la plus grande erreur possible et l'étendue de mesure. L'existence d'erreurs systématiques dues à un étalonnage incorrect ou trop peu fréquent du capteur est un autre facteur. D'autres erreurs de type accidentel peuvent également être causées par des signaux parasites, ou des absences de correction de température, de pression, etc. La **sensibilité** initiale d'un appareil de mesure est un autre facteur à prendre en compte, pour une valeur donnée du mesurande, elle détermine l'évolution de la grandeur de sortie du capteur en fonction de la grandeur d'entrée. Celle-ci est la valeur minimum du paramètre à mesurer en dessous duquel l'appareil ne réagit pas. Il faut enfin tenir compte de la **gamme de mesure** du capteur, qui est l'ensemble des valeurs du mesurande pour lesquelles un instrument de mesure est supposé fournir une mesure correcte. Ces valeurs de seuils au delà desquelles la précision et la sensibilité du capteur se dégradent. L'**étendue de mesure** correspond à la différence entre la valeur maximale et la valeur minimale de la gamme de mesure.

1.4.5. Fiabilité en environnement

Les capteurs sont conçus pour être utilisés dans des environnements généralement hostiles. Ces capteurs doivent alors être de grande robustesse. Un environnement qualifié de difficile, combine bien souvent plusieurs contraintes appliquées en même temps au capteur : chocs, vibrations, accélération et température. Dans le domaine de surveillance de l'eau, les contraintes imposées aux capteurs sont principalement l'humidité et la nature de l'eau. Ces capteurs souvent immergés, nécessitent un nettoyage périodique des sondes de mesure, car les dépôts les affectant provoquent plus de dérives, d'où un étalonnage régulier est nécessaire. La fiabilité est définie comme la capacité du capteur à fonctionner correctement, c'est-à-dire à fournir des données avec une précision adéquate et avec une conception robuste du matériel.

2. CAPTEURS PHYSICO-CHIMIQUES DE SURVEILLANCE DES EAUX

2.1. Mesure de la turbidité

La mesure de la turbidité de l'eau correspond à une mesure optique des particules en suspension dans l'eau qui lui donnent un aspect trouble. L'unité employée est appelée unité

néphélométrique de turbidité (NTU). Les particules sont d'origines diverses : argiles, limons, organismes microscopiques, dépôts dans les canalisations, corrosion, etc. Le turbidimètre mesure alors la quantité de lumière diffusée par un échantillon d'eau brute du fait de la présence de particules dans l'eau. Cette valeur est directement proportionnelle à la turbidité de l'échantillon mesuré. En effet, la lumière diffusée par les particules en suspension modifie d'autant plus l'éclairement de la cellule que le nombre de ces particules est élevé, ce qui permet d'obtenir une mesure de la turbidité de cette eau [53,80]. La figure 2.3 présente le capteur de turbidité utilisé dans une station d'eau potable.

Fig.2.3. Exemple de capteur de turbidité [81].

2.2. Mesure de la conductivité

L'appareil servant à évaluer la conductivité spécifique de l'eau s'appelle « Conductimètre » (fig.2.4). Il consiste à mesurer l'intensité du courant électrique recueilli aux bornes de deux électrodes de géométries connues, plongées dans l'eau et soumises à une différence de potentiel alternatif, dont la fréquence doit être d'autant plus élevée que la concentration en acides, sels ou bases dissous est grande, afin d'éviter les phénomènes de polarisation [80]. Le courant électrique mesuré est proportionnel à la concentration d'ions dans l'eau ; plus il est élevé et plus il y a d'ions dans l'eau. Le résultat se traduit en micro-Siemens par centimètre (µS/cm) à une température normalisée de 25°C.

Fig.2.4. Exemple de Conductimètre [82].

2.3. Mesure du pH

Plusieurs méthodes sont possibles pour mesurer le pH d'une eau. Les appareils électroniques (pH-mètres) permettent une mesure assez précise de la valeur du pH, à condition d'être correctement étalonnés [80]. Le principe de fonctionnement est basé sur le rapport qui existe entre la concentration en ions H_3O^+ et la différence de potentiel électrochimique qui s'établit dans le pH-mètre une fois plongé dans l'eau. Celui-ci est constitué de deux électrodes, l'une standard dont le potentiel est constant et connu (appelée électrode de référence), l'autre à potentiel variable (fonction du pH, appelée électrode de verre). Ces deux électrodes peuvent être combinées ou séparées. Il suffit donc de relier ces bornes à un voltmètre pour connaître la valeur du pH. En pratique, les électrodes sont réunies pour former une sonde. La figure 2.5 présente un exemple d'instrument pH-mètre.

Fig.2.5. Exemple de pH-mètre [83].

2.4. Mesure d'oxygène dissous

La teneur OD des eaux d'un lac est mesurée à l'aide d'une sonde munie d'une électrode sensible à cette molécule (fig.2.6). L'OD est généralement mesuré dans les parties les plus profondes du lac, à chaque mètre en partant de la surface jusqu'au fond. L'ampèremètre est utilisé industriellement en traitement des eaux pour la mesure en continu de la concentration en agents oxydants et met en œuvre une méthode simplifiée d'analyse par ampérométrie [80]. La cellule de mesure, qui est alimentée à débit constant en eau à analyser, comporte une cathode inattaquable, par exemple en platine, et une anode qui peut être en cuivre, en cadmium, en argent, etc. En l'absence d'agent oxydant, la pile ainsi formée est polarisée et n'est traversée que par un courant très faible. Sa dépolarisation, et par conséquent l'intensité du courant qu'elle débite, sont sensiblement proportionnels à la concentration de l'agent oxydant qui vient se réduire à la cathode. On mesure ainsi la concentration en chlore, ozone et oxygène d'une eau. L'inconvénient de ces appareils réside dans le fait qu'ils mesurent la somme des agents oxydants et qu'ils ne peuvent être vraiment utilisés que dans le cas où un seul corps se trouve en solution à concentration variable. L'effet d'un autre corps, éventuellement présent, à concentration constante, peut être annulé par action sur le zéro de l'appareil [53]. La concentration en OD dans l'eau est communément exprimée en milligrammes par litre (mg/l) ou en pourcentage (%) de saturation.

Fig.2.6. Exemple de sonde pour oxygène dissous [84].

2.5. Mesure de la température

La température de l'eau est généralement mesurée à l'aide d'un thermomètre à résistance afin de pouvoir la transmettre à distance. On utilise pour cela la croissance de la résistivité des métaux avec la température. Le courant électrique traversant une telle résistance alimentée

sous tension sera donc inversement proportionnel à la température du milieu environnant. La figure 2.7 présente un exemple de capteur de mesure de la température de l'eau.

Fig.2.7. Exemple de capteur de température de l'eau [85].

2.6. Les sondes multi-paramètres

Il existe sur le marché des dispositifs de contrôle de la qualité de l'eau, appelés sondes de qualité de l'eau. Ces sondes permettent d'effectuer des mesures de quelques paramètres descripteurs de la qualité de l'eau.

Exemple : ***Sonde de qualité de l'eau (Borehole)*** : elle est basée sur la sonde standard de l'industrie Ocean Seven 302TM, fabriquée par Idronaut (fig.2.8). Elle peut être configurée selon diverses options et comprend des fonctions de mesures de température et de pression de fluide et peut être dotée d'un maximum de quatre détecteurs complémentaires choisis dans une liste qui comporte : la conductivité de fluide, l'OD, le pH et le potentiel d'oxydoréduction. Cette sonde permet de dresser rapidement un profil en continu de tous les paramètres sélectionnés, soit pour un usage direct ou comme guide pour des échantillonnages ultérieurs.

Fig.2.8. Exemple de sonde de qualité de l'eau (Borehole) [86].

Principe de mesure : cette sonde est dotée d'un ensemble de six capteurs spécifiques destinés à effectuer des mesures de propriété de fluides dans un puits d'eau. Le recours à la technologie récente des capteurs intelligents, évite sans doute le besoin d'effectuer des calibrages complexes sur site. Les corrections des mesures en fonction des températures et pressions standards sont effectuées automatiquement sur les systèmes de surface. Les différentes caractéristiques de ce capteur multi-paramètre sont montrées dans le tableau 2.1.

Tableau 2.1. Caractéristiques de la sonde multi-paramètre Borehole.

Mesures	**Plage**	**Précision**	**Constante du temps**
Pression	0 à 1500 dbar	0.25 % pleine échelle	50 ms
Température	-1 à 49°C (des plages plus étendues sont disponibles	0.02°C	50 ms
Conductivité	0 à 62 µs/cm	0.02 µs/cm	50ms
Oxygène dissous	0 à 50 ppm	0.1 ppm	3 s
pH	0 à 14	0.01	3 s
Oxydoréduction	-1000 à 1000 mV	10 mV	3 s

3. CAPTEURS INTELLIGENTS

L'intégration des fonctions de traitement du signal au sein même des capteurs physiques est représentative de la notion de « capteur intelligent ». Les incessants progrès dans ce domaine sont un archétype d'innovation technologique. L'incorporation des traitements dédiés au traitement du signal permet d'augmenter la flexibilité de conception des dispositifs de mesure et de réaliser ainsi de nouvelles fonctions complexes. L'autre conséquence est une réduction de la charge sur les unités centrales de traitement et les dispositifs de transmission de l'information obtenue, en distribuant les opérations dans les systèmes de mesure eux-mêmes [87, 88].

Les progrès rapides dans les technologies connexes à la perception et à la commande impactent très largement le domaine. L'accroissement de la puissance de calcul disponible et l'amélioration des capteurs, tant en performances qu'en coût, augmentent le nombre des systèmes automatisés, offrant par ailleurs une qualité de service accrue. Ces systèmes autorisent et soutiennent à leur tour la progression et le développement de capteurs intelligents de plus en plus performants. D'ailleurs, les exemples typiques des domaines nouvellement développés sont la robotique, la mesure environnementale et le secteur biomédical [87].

3.1. Conjonction des besoins et Technologie

Les apports des capteurs intelligents sont relatifs, en premier lieu, à leurs fonctions primitives et relèvent alors d'amélioration de performances (exactitude, temps de réponse etc.). Les capteurs étant utilisés dans de nombreux systèmes de production de biens (processus continus ou manufacturiers) ou de services (tertiaire, systèmes embarqués), et l'apport de l'intelligence est lié aux performances générales et à l'accroissement particulier de la crédibilité de mesure [89].

Cependant, les utilisateurs formulent d'autres besoins d'informations qui concernent :

- **le contrôle-commande**, afin d'être renseigné en temps réel de façon qualitative ou quantitative sur l'état du système et/ou sur les caractéristiques du produit fabriqué ou du service fourni ;

- **la sûreté,** certaines mesures, étant à l'origine d'arcs réflexes qui permettent de protéger les opérateurs, l'équipement ou l'environnement ;

- **la maintenance** pour connaître l'état de dégradation du système et par conséquent son état de fonctionnement ;

- **la gestion de production** pour obtenir des bilans, des caractéristiques des différents flux de produit ou d'énergie ;

- **la gestion technique** pour connaître la disponibilité du processus.

Pour les fabricants, l'intelligence d'un équipement conduit à une diminution des coûts et des délais de son développement, ainsi que les coûts de sa production. Une fois amortis les investissements en matériel et en personnel, les intérêts résident essentiellement dans :

- **l'amélioration des performances** (auto compensation, prise en compte des non-linéarités, auto calibration etc.) ;

- **la traçabilité** de la vie de l'équipement ;

- **les aides à la configuration** et à la mise au point.

En résumé, les principaux gains apportés par les capteurs intelligents, tant en terme métrologique qu'en terme de fonctionnalité sont comme suit [90] :

- accroissement de l'exactitude des mesures par la prise en compte des grandeurs d'influence et la compensation locale des dérives ;

- crédibilité accrue des informations ;
- possibilité de validation, de diagnostic des mesures et de la communication ;
- élaboration des données directement exploitables au niveau supérieur, permet de distribuer une partie non négligeable de la puissance totale de calcul requise.

Pour l'utilisateur qui conçoit des systèmes à base de capteurs intelligents, le concept de capteur intelligent est synonyme de :

- diminution des coûts par simplification des câblages (bus de terrain), réduction des transmetteurs, convertisseurs ou adaptateurs intermédiaires ;
- augmentation des performances par compensation au niveau des capteurs des grandeurs d'influence ou des non linéarités ;
- simplification de la maintenance et de la configuration (auto-calibration, diagnostic etc.).

L'intérêt fonctionnel des capteurs intelligents est résumé dans le tableau 2.2 :

Tableau 2.2. Intérêt fonctionnel des capteurs intelligents.

Métrologie	**Installation – Configuration (mise en service)**	**Exploitation – Maintenance**
Fiabilité Précision accrue Etendues de mesure plus grande Traitement du signal Validation des mesures Corrections des grandeurs d'influence	Identification Configuration à distance Communication numérique bidirectionnelle Dialogue local (console portable)	Identification Autodiagnostics Auto-surveillance Aide à la maintenance Etalonnage Contrôle à distance

3.2. Constitution des capteurs intelligents

Un capteur intelligent est obtenu par l'association de la technologie des capteurs, de l'électronique et de l'informatique. Un instrument intelligent est un équipement qui intègre des fonctionnalités supplémentaires ou évoluées aptes à améliorer ce pourquoi il a été conçu. C'est au travers des nouveaux services inclus qu'un instrument intelligent se distingue d'un composant standard. D'une manière générale, l'évolution des instruments fait apparaître une

gradation allant de l'instrument analogique jusqu'à l'instrument intelligent en passant par à l'instrument numérique.

Un instrument est souvent considéré comme intelligent dès qu'il intègre au moins un traitement numérique (complexe ou non) et ce quel que soit son apport en terme de services. Outre des raisons de « marketing », ceci est dû à une très grande similitude entre l'architecture matérielle de l'instrument numérique, et celle de l'instrument intelligent. La distinction entre « smart sensor » et « intelligent sensor » est illustrée dans la figure 2.9 [89].

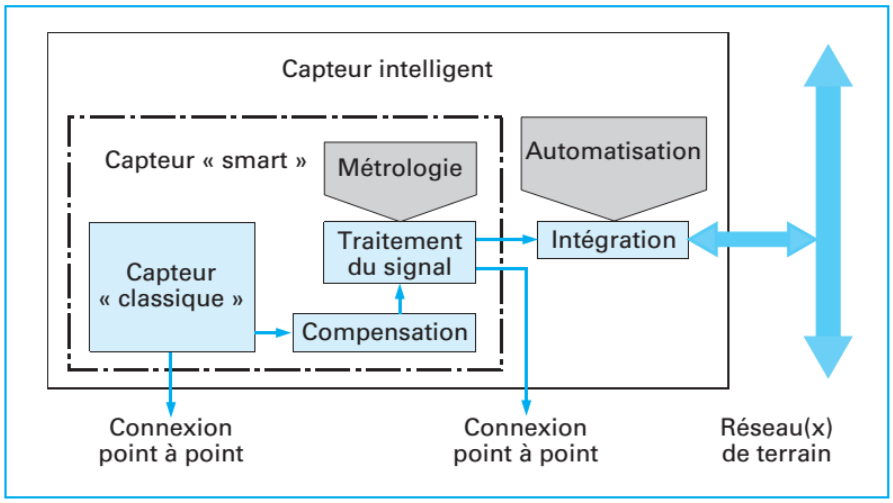

Fig.2.9. Capteur « smart » et capteur intelligent [89].

- **L'instrument smart,** possède des fonctionnalités qui améliorent ses performances métrologiques, par des fonctions embarquées de mémorisation et de traitement de données.

- **L'instrument intelligent,** enrichit cela d'une capacité à crédibiliser sa fonction associée à une implication plus importante dans la réalisation des fonctions du système auquel il appartient.

Le capteur intelligent correspond principalement à l'intégration dans le corps du capteur d'un organe de calcul interne (microprocesseur, microcontrôleur), d'un système de conditionnement du signal (programmable ou contrôlé) et d'une interface de communication.

D'un point de vue matériel, un capteur intelligent se compose alors de quatre unités montrées dans la figure 2.10 [91-99] :

- **Un capteur principal** spécifique au mesurande avec ses dispositifs d'acquisition et de numérisation du signal : transducteur, conditionneur qui adapte le signal électrique en vue de sa transmission, multiplexeur, amplificateur, échantillonneur bloqueur, convertisseur analogique/numérique. Le transducteur permet de détecter toute variation de la grandeur

physique en entrée du capteur. Sa conception est étroitement liée au domaine d'application pour lequel le capteur sera utilisé.

- **Un organe de calcul** numérique (microcontrôleur, microprocesseur ou dsPIC) servant au calcul et à la gestion de l'acquisition, la correction des effets des grandeurs d'influence au moyen de paramètres stockés en mémoire PROM, la linéarisation et le diagnostic des capteurs.

- **Une interface de communication** (radio ou filaire) assurant la liaison du capteur à un calculateur central et permettant un dialogue bidirectionnel des données numériques avec le système d'automatisation. Elle permet également au capteur intelligent de recevoir les informations du système nécessaires à l'élaboration de sa mesure et de sa validation. Elle peut également être utilisée dans les phases de calibration et de mise en service d'un équipement dans son environnement du travail.

- **Une alimentation** assurant une stabilisation des tensions est nécessaire à l'électronique de l'instrument. Une batterie peut être envisagée pour maintenir certaines activités en l'absence de source d'énergie extérieure (horloge, mémoire etc.).

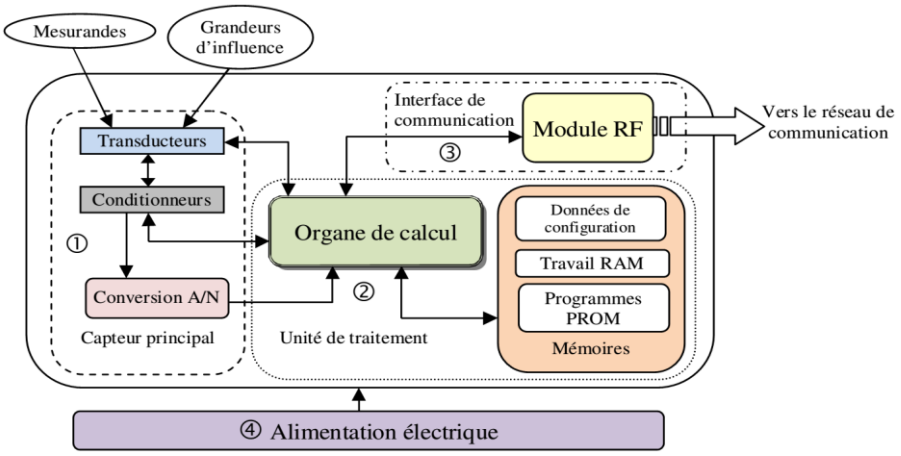

Fig. 2.10. Architecture matérielle d'un capteur intelligent [91].

Un capteur intelligent peut donc être considéré comme un véritable système embarqué autonome, qui devra posséder son propre système d'exploitation lui permettant de coopérer au sein d'une organisation d'autres capteurs intelligents.

3.3. Conception des capteurs intelligents

3.3.1. Approche fonctionnelle

L'approche orientée « fonction » s'articule autour de ce que fait l'équipement sans s'intéresser très finement à comment il le réalise exactement. Ce qui intéresse le concepteur c'est la sortie d'une « boîte noire ». Ce point de vue laisse le choix au concepteur-réalisateur d'adopter la technologie la mieux adaptée en termes de performances et de coût attendus [87].

3.3.2. Approche informationnelle

Cette approche incite à s'intéresser à ce qu'est l'équipement, dans son ensemble et, en particulier, à ses propriétés intrinsèques ; l'équipement est alors décrit minutieusement. Cette particularité de l'approche orientée « donnée » implique qu'on s'intéresse au type d'équipement et à la technologie utilisée. Néanmoins, les informations ainsi modélisées ne traitent pas certains concepts comme le service, en termes de comportement et de structure, de liens généralisation-spécialisation et de composés-composants. Ainsi, les approches orientées « données » ne suffisent pas en elles même à l'entière conception d'un équipement intelligent mais elles restent incontournables pour prendre en charge l'aspect informationnel [87].

3.3.3. Approche comportementale

Les approches orientées « comportement » s'intéressent plus à la dynamique du système et à son évolution dans le temps qu'à son état statique qui est plus représenté par les approches orientées « fonctions (ou traitements) ». Ainsi, il est courant de joindre au modèle obtenu par une approche orientée « données » ou par une approche orientée « fonctions », un modèle de comportement « dynamique ». En effet, les approches orientées « comportement » s'appuient sur le modèle développé par l'une des deux approches précédentes afin de déterminer les changements du système dans le temps, puisque le comportement du système est étroitement lié à ce qu'il fait et/ou sur quoi il le fait. Parmi les approches les plus connues orientées « comportement », on trouve :

- approches par systèmes de transitions ;
- approches par Réseaux de Petri ;
- approches par GRAFCET ;
- langages asynchrones.

Il existe un nombre important de langages permettant d'étudier le comportement dynamique des systèmes sans autant pouvoir exclure une analyse « statique préalable » afin de modéliser correctement le comportement. La multitude d'approches s'explique par le fait qu'il existe plusieurs moyens d'aborder le comportement d'un système : les notions de séquencement de tâches, parallélisme, d'exclusion, de préemption, de synchronisation, etc., ne sont pas toutes prises en charge par les modèles proposés [87].

3.4. Fonctions primordiales d'un capteur intelligent

Parmi les fonctions primordiales d'un capteur intelligent, on peut citer :

- Réaliser l'acquisition d'une mesure primaire de la caractéristique physique à mesurer : c'est le travail du corps d'épreuve, du transducteur et du conditionneur de fournir un signal électrique représentatif de la grandeur mesurée.

- Élaborer une image de la grandeur principale en tenant compte des grandeurs d'influence, et des éventuelles non-linéarités de type hystérésis, par exemple : il s'agit d'intégrer d'autres sources d'informations (capteurs locaux ou données externes) afin d'affiner et de corriger la mesure.

- Valider la mesure : il s'agit de crédibiliser la mesure en s'assurant du bon fonctionnement de l'instrument (mesure de tension d'alimentation, de la température de l'électronique), de la cohérence de la mesure soit de manière locale (détection de mesures aberrantes dans une série temporelle), soit en la confrontant à d'autres sources extérieures.

- Fournir une mesure opérationnelle, qui peut être aussi bien une mise en forme dans une unité particulière que l'élaboration d'une grandeur « virtuelle » c'est-à-dire non mesurée directement, mais élaborée à partir de la mesure de plusieurs points (cas d'une moyenne) ou de plusieurs phénomènes physiques.

- Fournir une mesure repérée dans le temps, pour traitement avec d'autres grandeurs issues d'autres équipements en exploitation et également pour constitution d'historiques.

A souligner toutefois les qualités toujours attendues d'un capteur qui sont la justesse, la fidélité, l'exactitude, la capacité de portée, la sensibilité, la linéarité, la finesse, la rapidité résolution, la traçabilité, la répétabilité et la reproductibilité.

3.5. Aspect économique

Les caractéristiques métrologiques et fonctionnelles probables des capteurs intelligents ont des incidences économiques appréciables, comme notamment [89] :

- une réduction des stocks de capteurs (grâce à la rangeabilité accrue) ;
- une réduction du temps d'étalonnage grâce à la correction rapide des imperfections des transducteurs et de calibration (configuration à distance) ;
- une fiabilité accrue des mesures grâce aux possibilités d'auto-contrôle et d'alarme ;
- une diminution des couts de câblage (grâce au bus de communication externe) ;
- un allègement des tâches routinières du système de calcul central.

3.6. Domaines d'applications des capteurs intelligents

On dénombre quatre grands domaines d'application dans lesquels les capteurs intelligents sont plus particulièrement usités :

- l'industrie manufacturière, caractérisée par des contraintes de temps-réel fortes et des contraintes environnementales variables suivant le type d'application ;
- l'industrie des « processus continus », caractérisée par des contraintes de temps-réel plutôt faibles mais qui doit satisfaire des contraintes environnementales fortes (processus chimiques, thermiques, etc.) ;
- les systèmes embarqués, qui doivent également satisfaire des contraintes temporelles très variables suivant les applications, mais des contraintes relatives à l'environnement très fortes ;
- enfin, le tertiaire, qui est le plus souvent le moins contraignant tant au niveau temps réel qu'environnemental.

4. SYSTEMES A BASE DE CAPTEURS INTELLIGENTS

Les systèmes fonctionnant à base de capteurs intelligents sont classifiés selon leur topologie de communication en deux classes principales : la classe à base de capteurs autonomes et la classe à base de capteurs organisés sous forme de réseau [91].

4.1. Capteurs intelligents autonomes

Dans cette architecture, lorsque la portée du module radiofréquence est suffisamment large, les capteurs intelligents réalisant une mission de détection relative à un procédé, peuvent transmettre directement les mesures d'une manière autonome à la station de base à travers une communication point à point en formant une topologie en étoile comme le montre la figure 2.11.

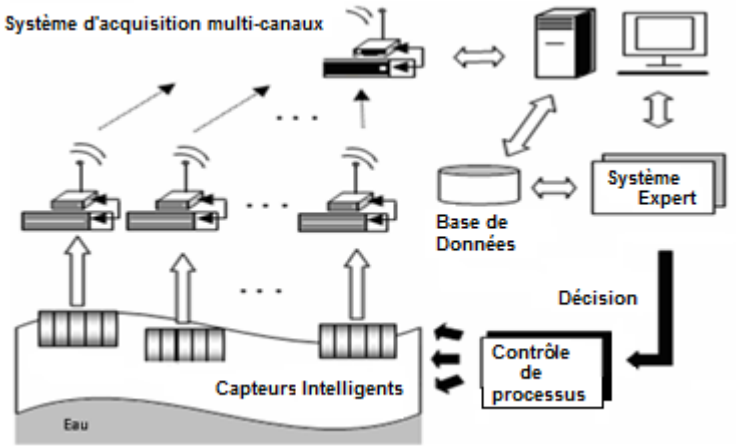

Fig.2.11. Topologie en étoile de capteurs intelligents autonomes [8].

La station de base envoie ensuite les données collectées pour les superviseurs via une passerelle, souvent en utilisant Internet ou tout autre canal de communication. Cependant, ce modèle de liaison directe n'est pas pratique, car la communication entre un capteur et sa station de base n'est pas toujours possible vue la limitation de la portée de transmission et les atténuations causées par des obstacles ou par des bruits dans le canal de communication [97,100].

4.2. Réseau de capteurs intelligents

Dans une zone géographique de déploiement plus vaste ayant une dimension dépassant la portée de la couverture radiofréquence des capteurs, la détection devient une tâche très difficile. Pour pallier à cette difficulté, un réseau ad-hoc de capteurs appelé réseau de capteurs sans fil (RCSF) est construit au fur et à mesure en fonction des besoins et des ressources disponibles et sans infrastructure comme représenté sur la figure 2.12.

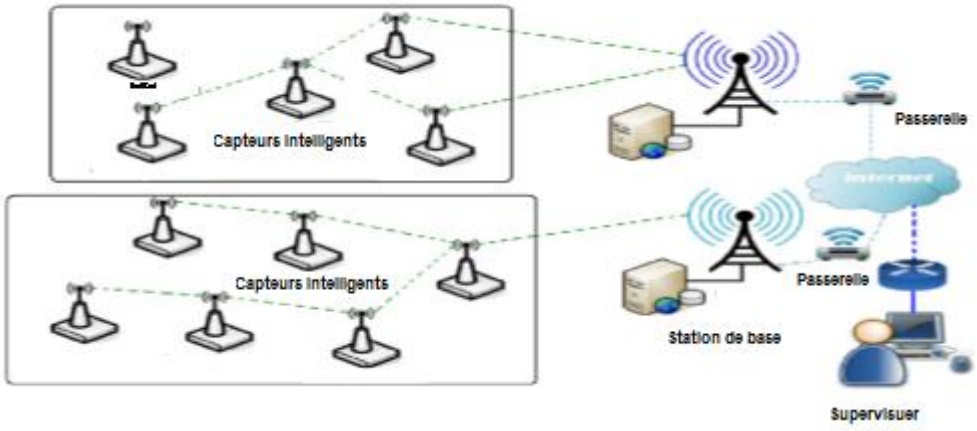

Fig.2.12. Communication multi-sauts ou coopérative dans un RCSF [91].

Les capteurs autonomes distribués dans l'espace doivent coopérer en formant un réseau de capteur sans fil maillé pour accomplir une tâche de détection globale demandée. Chacun de ces nœuds ne doit pas seulement collecter et diffuser ses propres données à la station passerelle, mais aussi servir de relai pour transférer les données de ses voisins qui sont plus proches par une liaison multi-sauts. La station de base transmet ensuite ces données par Internet ou par satellite à l'ordinateur central gestionnaire de tâches pour analyser ces données et prendre des décisions. Les paquets de données se propagent point à point en collaboration depuis le capteur source jusqu'à ce qu'ils arrivent à la station de base. Ce problème de routage, qui est la tâche de trouver un chemin multi-sauts d'un nœud de capteur à la station de base, est l'un des défis les plus importants. Il a reçu une immense attention de la communauté de recherche [97,101,102].

4.3. Capteurs intelligents en contrôle de la qualité des eaux

On trouve dans la littérature diverses études qui ont porté sur la mise en œuvre de systèmes de surveillance de la qualité des eaux utilisant la technologie de réseaux de capteurs sans fil (WSN). Dans [103], par exemple, un système distribué de mesure de la qualité de l'eau est conçu et mis en œuvre. Les capteurs de température, de conductivité, de pH et de turbidité sont connectés à un point de champ, à partir duquel des données sont envoyées via un réseau GSM de télécommunications mobiles, vers une station terrestre. Cette étude porte sur le traitement des données du capteur à l'aide de cartes de Kohenen (réseaux de neurones auto-associatifs). Dans [104] on présente un système d'environnement aquatique basé sur WSN qui détecte et surveille les données vidéo de zones clés et de paramètres de l'eau tels que la température, la turbidité, le pH, l'oxygène dissous et la conductivité. Les données sont envoyées dans ce cas depuis les nœuds de surveillance des données et la station de base vidéo

de données, vers un centre de surveillance distant utilisant la technologie ZigBee et CDMA (accès multiple par division de code). Dans [105], le système de surveillance de l'eau mis en place analyse et traite les paramètres de qualité de l'eau (pH, conductivité, oxygène dissous et température) et sonne l'alarme en cas de contamination de l'eau ou de modification de la qualité de l'eau. Les paramètres sont mesurés avec des capteurs standards et les données sont envoyées vers une station de base via GPRS (service général de radiocommunication par paquets). Dans [106], un système de surveillance et de mesure de la qualité de l'eau basé sur ZigBee WSN est présenté. Le système permet un sondage à distance et une surveillance en temps réel des paramètres descripteurs, ainsi que l'observation de l'état actuel et historique de la qualité de l'eau. Un réseau WSN à l'échelle du bassin hydrographique pour l'agriculture et la surveillance de l'eau, appelé « Soil Weather », est mis en œuvre dans [107]. Le réseau utilise les technologies GSM et GPRS pour la transmission des données des capteurs. Dans [108], le projet DEPLOY est introduit pour surveiller la distribution spatiale et temporelle de la qualité de l'eau ainsi que les paramètres environnementaux d'un bassin hydrographique. L'objectif est de démontrer qu'un réseau autonome de capteurs peut être déployé sur une vaste zone et que le système mesure des paramètres tels que le pH, la température, la profondeur, la conductivité, la turbidité et l'oxygène dissous. Dans [109], un système WSN est utilisé pour mesurer la qualité de l'eau douce et utilise la récupération de la lumière solaire pour une gestion optimisée de l'énergie. Les données collectées à partir des différents nœuds de capteur sont envoyées à un nœud de base secondaire et de là à une station de surveillance utilisant un réseau GSM. Les auteurs Adamo et al. Dans [110] ont plutôt proposé une solution de surveillance continue spatio-temporelle de la qualité des eaux de surface basée sur un réseau de capteurs intelligents. Une sonde d'eau de mer peu coûteuse, capable de mesurer la chlorophylle-a, la conductivité, la turbidité et la température a été utilisée. Les auteurs ont suggéré que la mise en œuvre de la solution entraînerait une amélioration immédiate de l'efficacité du contrôle de la qualité des eaux de surface. Cependant, l'utilisation d'une seule sonde pour détecter tous les paramètres de l'eau n'est pas très fiable, car toutes les données seront perdues en cas de dysfonctionnement de la sonde.

4.4. Exemple d'application

La société « Libelium » a lancé une plateforme de capteurs sans fil intelligents pour simplifier la surveillance à distance de la qualité de l'eau. Doté de multiples capteurs qui mesurent une douzaine de paramètres de qualité de l'eau les plus pertinents, « *Waspmote smart water* » est la première plateforme de surveillance de la qualité de l'eau à proposer des

nœuds autonomes qui se connectent au « Cloud » pour un contrôle de l'eau en temps réel. « *Waspmote Smart Water* » convient donc à la surveillance de l'eau potable, à la détection des fuites chimiques dans les rivières ainsi qu'à la mesure à distance des niveaux de pollution de l'eau de mer. Les paramètres de qualité de l'eau mesurés incluent le pH, l'oxygène dissous, le potentiel d'oxydoréduction, la conductivité (salinité), la turbidité, la température et les ions dissous (Fluorure (F^-), Calcium (Ca^{2+}), Nitrate(NO_3^-), Chlorure (Cl^-), Iodure (I^-), Cuivrique (Cu^{2+}), Bromure (Br^-), Argent (Ag^+), Fluoroborate (BF_4^-), Ammoniac (NH_4), Lithium (Li^+), Magnésium (Mg^{2+}), Nitrite (NO_{2-}), Perchlorate (ClO_4), Potassium (K^+), Sodium (Na^+) [111].

La figure 2.13 présente la plateforme de capteurs intelligents « *Waspmote smart water* ».

Fig.2.13. Plateforme de capteurs intelligents « *Waspmote smart water* » [111].

Les caractéristiques de la plateforme s'énoncent comme suit :

- Ajout et/ou changement de sonde simple et rapide.
- Option de panneau externe alimenté à l'énergie solaire.
- Radios disponibles : 802.15.4, 868 MHz, 900 MHz, WiFi, 4G, Sigfox et LoRaWAN.
- Programmation en liaison radio (OTAP) de plusieurs nœuds à la fois (via des radios WiFi ou 4G).
- Protocoles industriels en option : RS-232, RS-485, Modbus, bus CAN.
- Récepteur GPS.
- Module de batterie externe.
- Connecteur SIM externe pour les modèles 4G.

Vu les fonctions primordiales pouvant caractériser un capteur intelligent, il serait tout à fait possible dans un but de flexibilité et de performances, d'intégrer d'éventuels capteurs logiciels au sein d'une plateforme de capteurs intelligents, ou même, au niveau d'un capteur intelligent.

5. CAPTEURS LOGICIELS

Les processus industriels sont bien équipés avec une variété de capteurs, conçus pour la supervision, la surveillance et le contrôle en ligne, ainsi que pour maintenir une qualité de produit constante. Certaines variables, ne peuvent pas être automatiquement mesurées en ligne, en raison du manque de capteurs ou du coût élevé du capteur, ce qui entraîne un manque d'informations suffisantes sur l'état du système en temps réel. Habituellement, des tests de laboratoire sur des échantillons de produit sont effectués pour mesurer la qualité du produit hors ligne sur une base d'intervalle spécifié. Afin de mesurer les variables de qualité en temps réel, on peut utiliser des méthodologies d'intelligence informatique pour construire des capteurs intelligents/informatiques permettant de déduire la valeur ou les variables cibles de qualité d'autres variables de processus mesuré en ligne. La construction de tels capteurs repose sur le fait que les valeurs des variables cibles, ont une relation fonctionnelle avec d'autres variables de processus pouvant être mesurées en ligne. Ce type de capteurs est généralement appelé « capteurs logiciels ». En termes généraux, les capteurs logiciels peuvent être définis comme des modèles inférentiels utilisant des mesures de capteurs disponibles en ligne (variables faciles à mesurer) pour une estimation en ligne de variables de qualité (variable difficile à mesurer) qui ne peuvent pas être mesurées automatiquement du tout, ou ne peuvent être mesurées qu'à des coûts élevés, de manière sporadique ou avec des retards importants (par exemple, une analyse en laboratoire).

Faut-il souligner qu'au cours de ces deux dernières décennies, les chercheurs ont commencé à utiliser les données mesurées et stockées dans l'industrie pour concevoir des modèles prédictifs basés sur ces données. Ces modèles ont été donc appelés capteurs logiciels. Le terme « capteur logiciel » est une combinaison du mot « logiciel », car ces modèles sont basés sur des programmes informatiques ; et le terme « capteur » parce que ces modèles fournissent des informations similaires à celles des capteurs matériels. Les capteurs logiciels sont des instruments virtuels qui peuvent être utilisés à des fins similaires à leurs équivalents matériels. Ces sont des programmes informatiques et un modèle à la base traitera l'information produite typiquement par des instruments matériels [112].

Un capteur logiciel est classiquement décrit comme un modèle d'entrées-sortie. Les entrées du modèle sont généralement constituées des variables secondaires faciles à mesurer de façon fiable, avec des coûts raisonnables (souvent présentées comme des variables faciles à mesurer). Les sorties du modèle sont constituées d'informations associées à ces variables primaires dont la mesure fiable est difficile ou très chère (souvent indiqué comme variable difficile à mesurer). Les informations que contiennent les variables d'entrée et de sortie sont modélisées empiriquement dans le capteur logiciel (fig.2.14) [112].

Fig.2.14. Capteur logiciel décrit comme un modèle d'entrées-sortie.

CONCLUSION

Dans ce chapitre, la notion de capteur en tant qu'élément primordial dans une chaîne de mesure a été introduite. Les différents capteurs physico-chimiques utilisés dans le domaine de l'eau ont été particulièrement présentés. Le capteur intelligent en tant qu'élément microélectronique de base dans une chaîne moderne de mesure a été aussi introduit ; une solution qui pourrait être profitable aux systèmes de contrôle de la qualité des eaux. Selon les caractéristiques susmentionnées de ce type de capteurs, les capteurs logiciels à base d'approches algorithmiques pourraient éventuellement être intégrés au niveau de ces architectures pour pallier au manque de certains capteurs spécifiques et surtout pour plus de flexibilité et de performances.

Le chapitre suivant étudie ces capteurs logiciels ainsi que les différentes techniques utilisées dans leur développement.

CHAPITRE III

TECHNIQUES DE DEVELOPPEMENT DES CAPTEURS LOGICIELS

INTRODUCTION

Dans un système de surveillance et de contrôle de la qualité des eaux propres, certaines variables, ne peuvent pas être automatiquement mesurées en ligne, en raison du manque de capteurs, pour des raisons de panne, du coût élevé ou de la nature même du capteur (analyse en laboratoire), ce qui entraîne un manque d'informations suffisantes sur l'état du système en temps réel. En termes généraux, les capteurs logiciels peuvent être définis comme des modèles pouvant utiliser des mesures de capteurs disponibles en ligne pour une estimation en ligne de variables difficiles à mesurer qui ne peuvent pas être mesurées automatiquement du tout, ou ne peuvent être mesurées qu'à des coûts élevés, de manière sporadique ou avec des retards importants.

La première partie de ce chapitre est donc dédiée aux capteurs logiciels et décrit brièvement les environnements industriels dans lesquels ces capteurs sont déployés. Les types de capteurs logiciels ainsi que leurs procédures de conception, sont à cet effet décrits. Dans la seconde partie, un état de l'art sur les méthodes utilisées dans le développement de ces capteurs est présenté. La technique de prétraitement des données utilisée en vue d'une réduction de la dimension des variables d'entrée appelée : analyse en composantes principales (ACP), les propriétés générales des réseaux de neurones artificiels (RNA) statiques (perceptrons multicouches ou MLP), ainsi que les réseaux de neurones RBF sont ainsi décrits. Les aspects théoriques et fondements de l'apprentissage statistique sont ainsi rappelés. La formulation générale de la méthode SVM appliquée à la régression et à la classification, est particulièrement décrite. Enfin, les aspects théoriques et fondements d'une autre technique appelée ELM sont aussi rappelés et décrits.

1. LE CAPTEUR LOGICIEL

Un capteur logiciel est un modèle de régression qui utilise des variables faciles à mesurer pour prédire une variable difficile à mesurer. Il fait l'objet de recherches dans de nombreux domaines. À l'origine, les capteurs logiciels ont été étudiés dans le cadre de la chimiométrie, ce qui signifie des méthodes statistiques pour extraire des informations à partir d'ensembles de données qui consistent souvent en de nombreuses variables mesurées [113].

Les capteurs logiciels sont capables de remédier aux insuffisances des capteurs matériels et des techniques de mesure conventionnelles. Selon les résultats d'une enquête menée sur 26 sociétés au Japon, les capteurs logiciels peuvent désormais résoudre au moins les problèmes rapportés dans le tableau 3.1 [114,115].

Tableau 3.1. Problèmes de capteurs physiques.

Pourcentage (%)	Problème reconnu
27	Maintenance longue durée
21	Besoin d'étalonnage
15	Vieillissement
13	Précision insuffisante
10	Long temps d'arrêt, dynamique lente
8	Grand bruit
2	Faible reproductibilité
4	Autres

Outre les problèmes mentionnés ci-dessus, que les capteurs logiciels peuvent résoudre, le principal objectif de l'utilisation de ces capteurs dans l'industrie des procédés consiste à :

- mesurer la stabilisation de la qualité du produit via des prédictions en ligne ;
- réduire le temps et les coûts grâce à un fonctionnement efficace ;
- validation en ligne des analyses via un fonctionnement parallèle ;

- dispositifs de sauvegarde et de détection de panne ;
- tenir compte de certains paramètres nécessitant une analyse de laboratoire spécifique, difficile et chère ;
- coût élevé de certains capteurs, leur nombre, le temps consacré à leur vérification, l'opération de nettoyage, les routines d'étalonnage et leur remplacement.

Pour les avantages mentionnés ci-dessus et étant une technologie clé pour le contrôle des variables rarement mesurées, les capteurs logiciels sont de plus en plus déployés dans les installations industrielles, en particulier dans les procédés chimiques.

1.1.Capteurs logiciels appliqués en traitement des eaux : état de l'art

Différents travaux de recherche ayant pour application les capteurs logiciels dans le domaine du traitement des eaux ont été réalisés. Par exemple, Aguado et al.ont introduit un capteur logiciel à base d'un RNA de type MLP pour surveiller la concentration en phosphore total (TP) dans une usine pilote [116]. Ce capteur utilisait les données disponibles d'autres capteurs en ligne afin de surmonter d'éventuels retards pouvant découler d'analyses de laboratoire complémentaires. Hong et al. conçurent aussi un capteur logiciel, basé sur un RNA, pour estimer les concentrations de nutriments dans le processus d'élimination des nutriments biologiques [117]. Dans le même contexte, Bagheri et al. ont développé un capteur logiciel, basé cette fois-ci sur RBF et MLP, et ce pour surveiller à la fois le phosphore total (TP) et l'azote ammoniacal (NH4-N) [118]. Kim et al. ont plutôt construit des capteurs logiciels pour les concentrations de NOx-N et l'ammoniac [119]. Dans [21], les auteurs ont abordé le traitement des eaux usées à l'aide d'un algorithme génétique basé sur un capteur logiciel pour estimer la dynamique nutritionnelle de l'opération anoxique / oxique en utilisant des mesures en ligne plus fiables de la concentration en pH, du potentiel d'oxydoréduction (ORP) et l'OD en tant que données d'entrée. Lamrini et al. [22] ont aussi présenté l'élaboration et la validation d'un capteur logiciel utilisant un MLP pour la prédiction en ligne du dosage optimal de coagulant à partir de mesures de la qualité de l'eau brute dans une station de traitement d'eau potable. Dans [23], un modèle de capteur logiciel basé sur RBF a été aussi étudié pour estimer les principaux paramètres de performance, à savoir la concentration en pH, le flux de gaz de perméat et la concentration en hydrogène résiduel.

1.2.Types de capteurs logiciels

Les capteurs logiciels peuvent être classés en fonction de différents critères. Sur la base de leur méthodologie de modélisation, les capteurs logiciels sont fondamentalement classés en deux catégories principales [14,115,120,121].

1.2.1. Capteurs logiciels pilotés par les données

Il s'agit de modèles basés sur des données de type boîte noire, décrivant les conditions réelles d'une installation, induits par les données de mesure observées (entrées/sorties). Comme leur nom l'indique, elles sont basées sur des données mesurées, qui peuvent être collectées à partir de l'installation pour laquelle le capteur logiciel est développé, ou à partir d'expériences spécialement conçues (données expérimentales) [14, 115,122, 123].

1.2.2. Capteurs logiciels pilotés par le modèle

Ce sont des modèles de premier principe qui se concentrent sur l'état des processus, décrivant leurs antécédents physiques et chimiques. Contrairement aux modèles basés sur les données, ce type de capteur est principalement basé sur des équations dérivées pouvant décrire les caractéristiques de processus. Ils sont également appelés modèles basés sur les connaissances ou modèles à boîte blanche [124]. Les procédés industriels étant si complexes et nécessitant une grande quantité de connaissances phénoménologiques pour le développement de tels capteurs, ils sont souvent peu pratiques, voire impossibles à développer [122].

Un modèle combinant à la fois des modèles empiriques et physiques pour estimer la variable cible est appelé capteur logiciel hybride ou de type boîte grise. Les capteurs logiciels basés sur les données sont plus susceptibles de renvoyer des prévisions plus précises que leurs deux homologues en raison de la flexibilité intégrée qui est réalisée aux dépens de la quantité et de la qualité considérables des données de formation nécessaires pendant la phase de conception. En outre, ils sont relativement faciles à développer car une expertise approfondie n'est pas requise. Inversement, les modèles basés sur des modèles sont entièrement basés sur des connaissances phénoménologiques, qui ne sont souvent pas réalisables dans un processus complexe. Par conséquent, le travail présenté dans le cadre de cette recherche est axé sur les capteurs logiciels pilotés par les données [121].

1.3.Développement du capteur logiciel

Les capteurs logiciels pilotés par les données sont conçus pour constituer une alternative intéressante, non seulement aux capteurs matériels, mais également aux capteurs logiciels pilotés par les modèles. Ceci est principalement dû au coût relativement élevé pour le premier et au coût et à la complexité de ce dernier. De plus, les techniques de mesure modernes utilisées dans les installations industrielles permettent de collecter, de stocker et d'analyser des données, ce qui suggère fortement l'utilisation de capteurs logiciels basés sur les données en manipulant l'utilisation des mesures disponibles dans les bases de données d'installations [125]. Par conséquent, la qualité d'un capteur logiciel dépend de la qualité des données utilisées pour son développement [126]. De plus, la conception d'un capteur logiciel dépend en grande partie de l'objectif pour lequel il est conçu et des tâches à accomplir. Cela signifie que les procédures de conception de capteurs souples à des fins de sauvegarde diffèrent de celles conçues pour les prévisions en ligne. Cependant, ils partagent tous un cadre pratiquement unifié qui peut être divisé en étapes, comme indiqué dans le diagramme de la figure 3.1 [13,14,127].

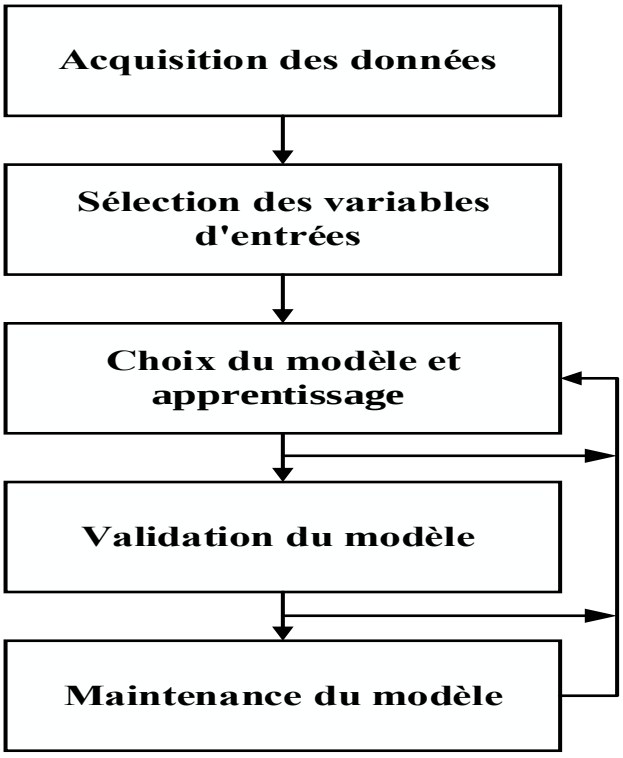

Fig.3.1. Principales étapes de développement d'un capteur logiciel.

La conception d'un capteur logiciel piloté par les données repose principalement sur des données historiques extraites des bases de données d'installations industrielles. Dans ces

usines, la production constitue l'objectif principal du processus, tandis que les mesures variables, l'enregistrement des échantillons, la récupération des données, la réconciliation et la synchronisation sont des objectifs subsidiaires. Cela pose un certain nombre de problèmes aux développeurs de capteurs logiciels, tels que l'évaluation de la qualité des données (par exemple, la fiabilité, la précision, l'exhaustivité et la représentativité), la sélection des variables et des mesures indépendantes pertinentes. En conséquence, les étapes de prétraitement des données qui traitent de ces problèmes sont essentielles à la qualité du capteur logiciel, comme indiqué dans les sections suivantes [126].

1.3.1. Acquisition des données

Les industries sont généralement tenues de stocker leurs données dans les processus. Ceci est la base de l'utilisation ultérieure de ces données pour l'optimisation du système ou d'autres méthodes liées aux données. La collecte des données et l'inspection ultérieure des données sont les premières étapes du développement d'un capteur logiciel. L'inspection des données est une étape préliminaire dans la conception de capteur logiciel, car elle aide à choisir les variables indépendantes les plus influentes, à identifier les profils de données, à classer les profils de processus et à extraire les informations pertinentes contenues dans les données. Elle est généralement effectuée en tant qu'étape initiale, car elle permet au concepteur d'obtenir une vue d'ensemble de la structure de données.

1.3.2. Sélection des variables d'entrées

Les caractéristiques remarquables des données acquises dans les installations de systèmes sont redondantes et peuvent être insignifiantes. Des perturbations qui corrompent les mesures sont également parfois présentes. Souvent, la quantité et la qualité des données ainsi que leur haute dimensionnalité peuvent être un facteur limitant pour le développement du capteur logiciel. Par conséquent, il est nécessaire de préparer les données avant qu'elles ne soient traitées par le modèle du capteur logiciel. La compréhension du processus et la connaissance à priori sont requises dans cette phase [14,112].

Le choix des variables d'entrée est une étape cruciale. La sélection des variables consiste à choisir les variables secondaires qui sont les plus informatives pour le processus en cours de modélisation, ainsi que celles qui fournissent le plus généralisable. Cette étape est fondamentale, car les modèles sont construits à partir d'un nombre fini d'observations et un modèle avec trop d'entrées peut conduire à un ajustement excessif et engendrer une charge de calcul importante. En outre, décrire un processus en fonction de quelques variables

sélectionnées permet de conserver l'interprétabilité. Un nombre réduit de variables présente plusieurs avantages, tels que la réduction du temps de développement du modèle, la possibilité d'agrégation de l'information sur l'interprétation physique du processus ou l'amélioration de la performance du modèle. De plus, une réduction du nombre de variables implique un nombre plus faible de capteurs réels requis, une diminution des coûts et une augmentation de la faisabilité des applications.

1.3.3. Choix du modèle et apprentissage

Le choix du modèle est une étape critique dans le développement du capteur logiciel. En particulier, la structure du modèle définit la tâche d'application spécifique et la sélection des paramètres du modèle détermine la capacité de généralisation du capteur logiciel. Cependant, une approche cohérente de la tâche n'existe pas.

Il existe deux approches de modèle distinctes appliquées pour le développement de capteurs logiciels. La première est basée sur des modèles en boîte blanche, obtenus par une connaissance physique du processus et la seconde est basée sur des modèles basés sur des boîtes noires ou des données, basées exclusivement sur la construction d'un modèle à partir de données empiriques du processus. La modélisation par l'approche de la boîte blanche nécessite de solides connaissances sur le processus et demande un long travail de modélisation pour construire les modèles [112,128]. Pour les systèmes complexes, l'approche de modélisation boîte blanche peut être pratiquement impossible. Les modèles à boîte noire sont basés sur des observations empiriques du processus (les méthodes elles-mêmes sont des méthodes prédictives empiriques). La modélisation en boîte noire est capable de décrire les conditions réelles du processus et nécessite peu de connaissances sur le système à modéliser. Toutefois, cela nécessite un travail intensif sur les données de processus. Certaines difficultés avec ces types d'approches sont liées à la difficulté de choisir le type et la structure du modèle, les fonctions à utiliser et la quantité de termes de fonction nécessaires au développement. Dans la modélisation en boîte noire, le premier aspect à décider est de quel type de modèle s’agira t-il ?

Les modèles choisis pour étude dans le cadre de ce travail de thèse, sont les réseaux de neurones artificiels de types MLP et RBF, les SVMs et l’ELM. Dans la majorité des applications de capteurs logiciels, un seul modèle est réglé en utilisant tous les échantillons d'apprentissage disponibles, sans distinguer les modes de fonctionnement du processus pendant la phase d'apprentissage. Cependant, l'existence de plusieurs modes de

fonctionnement dans un processus est une caractéristique inhérente à la plupart des applications industrielles.

1.3.4. Validation du modèle

La validation du modèle est une étape très importante dans le développement du capteur logiciel, dans laquelle le concepteur estime à quel point le modèle va fonctionner sur les données invisibles. Idéalement, si suffisamment de données étaient disponibles, le développeur du capteur logiciel mettrait de côté un ensemble de validation et l'utiliserait pour évaluer un modèle dont les paramètres de base sont étalonnés sur un ensemble d'apprentissage, pour les différentes valeurs de ses méta-paramètres. Après avoir trouvé l'ensemble optimal de méta-paramètres, le développeur calibrerait alors le modèle pour définir ses paramètres de base, en utilisant toutes les données d'entraînement disponibles. Le modèle résultant est finalement évalué sur un ensemble de données de test indépendant. L'objectif de l'étape de validation du modèle est d'évaluer la capacité du modèle formé à effectuer une généralisation à de nouveaux échantillons. La précision de la généralisation peut également être utilisée comme estimateur pour le classement d'un modèle dans une approche de sélection de variables.

1.3.5. Maintenance du modèle

Lors de la conception d'un capteur logiciel, les données historiques du processus sont utilisées pour apprendre le modèle de capteur logiciel. Cependant, les données historiques contiennent des informations limitées, correspondant à une période de temps limitée, et pouvant également se concentrer sur un ensemble limité de zones d'opérations de l'espace d'état. Lorsqu'il s'agit de nouveaux événements, non décrits dans les données historiques, le capteur logiciel tend à diminuer ses performances. Dans ce contexte, et pour surmonter une telle détérioration des performances, l'objectif de la maintenance du capteur logiciel est de maintenir une bonne réponse du capteur logiciel même en présence de variations de processus, ou de certaines modifications de données. Généralement, cela se fait en mettant à jour le modèle de capteur logiciel en ligne / récursivement, en mode batch ou échantillon, en utilisant les échantillons entrants du processus (dans ce contexte, les capteurs logiciels sont appelés "capteurs logiciels adaptatifs") [113,129].

1.4. Applications des capteurs logiciels

Les capteurs logiciels ont une large gamme d'applications et peuvent remplir une grande variété de tâches. Cela est principalement dû à leur coût relativement faible par rapport à leurs homologues matériels ; de plus, les capteurs matériels impliquent des coûts d'acquisition, d'installation, de fonctionnement et de maintenance. Les capteurs logiciels deviennent des outils courants, non seulement en tant que source d'information pour les opérateurs de salle de contrôle, mais également pour le contrôle inférentiel en boucle fermée [13]. A cet effet, il a été rapporté dans [130] qu'il y a dix ans, 80% des capteurs logiciels étaient déployés dans des applications en boucle ouverte, alors que 20% étaient utilisés dans des applications en boucle fermée. Actuellement, 80% de ces capteurs sont plutôt utilisés dans des applications de contrôle en boucle fermée.

Le domaine d'application dominant des capteurs logiciels est la prédiction en ligne de variables de processus difficiles à mesurer. Cependant, d'autres domaines d'application incluent la surveillance de processus et la détection de défauts [131].

1.4.1. Prédiction en ligne

L'application la plus courante des capteurs logiciels est la prédiction de valeurs ne pouvant pas être mesurées en ligne à l'aide de mesures automatisées [112]. Cela peut être dû à des raisons technologiques (par exemple, aucun équipement n'est disponible pour la mesure requise), ou à des raisons économiques (par exemple, l'équipement nécessaire est trop coûteux), etc. D'ailleurs, la prévision en temps réel des variables cibles de processus est un domaine d'application qui a fait augmenter l'usage des capteurs logiciels comme outils d'optimisation essentiels dans les processus industriels. Contrairement à cette approche, des échantillons sont envoyés aux laboratoires pour une analyse hors ligne, ce qui crée des retards importants, ou des variables sont mesurées en ligne par un analyseur, ce qui crée des retards et des problèmes de fiabilité.

Les capteurs logiciels développés pour les prédictions en ligne sont basés sur des approches d'apprentissage supervisées ; les données d'apprentissage comprennent à la fois des données d'entrée et des mesures de variable cible correspondante.

1.4.2. Surveillance de processus et détection de défauts de processus

La surveillance des processus est un autre domaine d'application des capteurs logiciels. La surveillance du processus peut être une tâche d'apprentissage non supervisée ou de

classification binaire. Les systèmes peuvent être formés pour décrire/analyser l'état de fonctionnement normal ou pour reconnaître d'éventuels défauts de processus [112].

La qualité des produits et la sécurité des processus sont deux critères principaux pris en compte lors de la conception des processus industriels. Il est souvent impossible d'avoir un processus entièrement surveillé, car cela nécessite la construction d'usines fortement instrumentées et à coût élevé. Par conséquent, les capteurs logiciels sont considérés comme une alternative permettant de surveiller la sécurité des processus et d'améliorer la qualité du produit.

1.4.3. Détection et reconstruction de défauts de capteurs

Comme déjà mentionné, les usines chimiques modernes sont équipées d'un grand nombre de capteurs. Parmi ces nombreux paramètres de processus, certains sont des paramètres clés sans lesquels les opérateurs de processus auront du mal à maintenir la qualité et le débit du produit. Étant donné que ces paramètres de processus sont mesurés par des capteurs physiques, il est toujours possible que ces capteurs échouent. Une fois qu'un capteur physique défectueux est détecté et identifié, il peut être reconstruit ou le capteur peut être remplacé par un capteur logiciel, formé pour agir en tant qu'élément de secours du dispositif de mesure câblé. Si le capteur logiciel s'avère efficace en remplacement du capteur physique, il fournira alors à l'opérateur une indication en attente des variables de processus. Dans certains cas, les capteurs souples robustes peuvent remplacer l'appareil de mesure même s'il est en état de fonctionnement. L'outil logiciel peut donc facilement être mis à jour, il ne présente pas de défaillance mécanique, est facile à mettre en œuvre dans le système de contrôle et peut donc fournir un avantage financier substantiel.

2. TECHNIQUES DE DEVELOPPEMENT

2.1. Etat de l'art

Il existe un certain nombre de techniques pour développer des capteurs logiciels. Sur la base des fréquences de leur utilisation rapportée dans la littérature recensée, les principales familles de modélisation suivantes ont été reconnues :

- analyse en composantes principales [14,132];
- moindres carrés partiels (MCP) [16, 17] [133-135] ;
- réseaux de neurones artificiels supervisés [116] [136-140];
- cartes auto-organisées (CAO) [141,142];

- systèmes neuro-flous [143-147];
- machines à vecteurs de support [18,27,148].

Différentes méthodologies multivariées ont été utilisées pour développer les capteurs logiciels dans le domaine de traitement des eaux. La distribution des principales familles de ces méthodes est présentée dans la figure 3.2.

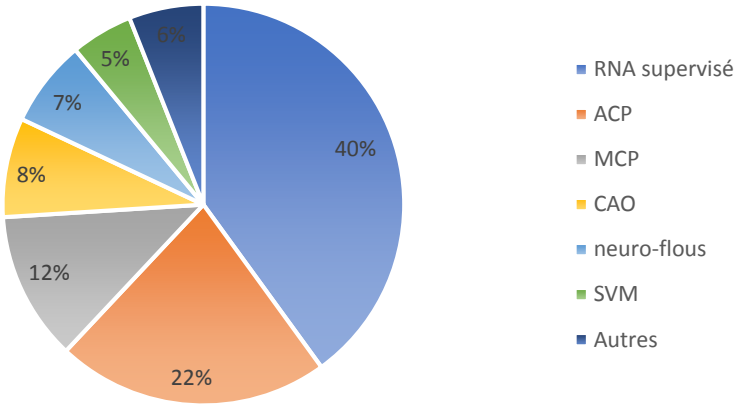

Fig.3.2. Principales familles de méthodes employées dans la conception de capteurs logiciels [14,112,127].

Parmi les principales familles de modélisation, différentes techniques et extensions des méthodologies conventionnelles ont été déployées en fonction de l'application considérée. Les techniques RNA supervisées sont les méthodes les plus courantes pour prévoir les principales variables. Les techniques ACP sont plutôt utilisées dans la majorité des études visant à surveiller les processus et les instruments de traitement. En outre, ces techniques sont les plus utilisées dans le prétraitement appliqué à de nombreux capteurs logiciels, ce qui indique clairement le potentiel de celles-ci dans la compression des informations contenues dans les données de grande dimension.

2.2.Méthode de réduction des dimensions

Tout système de reconnaissance de formes comporte une phase très importante basée sur la réduction des dimensions. La réduction de la dimension se pose comme une étape primordiale dans le processus de prétraitement des données. En effet, pour des données appartenant à un espace de grande dimension, certains attributs n'apportent aucune information voire expriment du bruit, d'autres sont redondants ou corrélés. Ceci rend les

algorithmes de décision complexes, inefficaces, moins généralisables et d'interprétation délicate. En effet, les principaux objectifs de la réduction de dimension sont le fait de :

- faciliter la visualisation et la compréhension des données ;
- réduire l'espace de stockage nécessaire ;
- réduire le temps d'apprentissage et d'utilisation ;
- identifier les facteurs pertinents.

Parmi les méthodes de réduction des données, on trouve l'ACP.

2.2.1. Analyse en Composantes Principales

L'ACP est une technique également connue sous le nom de la décomposition de Karhunen-Loève (KL) ou la décomposition orthogonale. Elle joue un rôle fondamental dans l'analyse statistique. Elle a été introduite par Karl Pearson en 1901 [149] pour décrire et résumer l'information contenue dans un ensemble de données. Par la suite, dans les années 1930, elle a été de nouveau développée par Harold Hotelling [150] comme une méthode d'analyse des relations existantes entre les variables. L'idée de base consiste à remplacer l'ensemble des variables inter-corrélées par un nombre réduit de variables de synthèse qui retiennent l'essentiel de l'information. Cette nouvelle représentation réduite, facilite l'interprétation du contenu de ces variables. En fait, d'un point de vue géométrique, l'ACP peut être vue comme une méthode de rotation des données afin de permettre à l'observateur de mieux comprendre les relations entre les données.

En effet, l'ACP est une transformation algébrique, qui permet de mettre en évidence des composantes, en tenant compte de la variance totale de toutes les variables à étudier. Cette transformation effectue un changement de base qui permet de projeter des variables liées entre elles (inter-corrélées) dans un nouvel espace orthonormé où un autre nombre réduit, de variables décorrélées, est obtenu. Ces nouvelles variables, appelées composantes principales, expliquent au mieux la variabilité des données originales. Les domaines d'application de cette technique sont multiples, allant de la biologie, l'économie et plus récemment le traitement d'images. L'ACP est globalement utilisée pour décrire, visualiser, décorréler, débruiter et réduire la dimension des données.

L'idée est de mettre au point une technique permettant de résumer l'information apportée par « p » variables quantitatives par « n » unités (appelées *individus*) en la détruisant le moins possible. Cette technique utilise des combinaisons linéaires des variables pour la réduction de la dimension d'un ensemble de données en trouvant un nouvel ensemble constitué d'un petit

nombre de facteurs, de dimension plus faible que l'ensemble original des variables. Produire un résumé de cette information, c'est projeter ces points dans un espace de dimension inférieure à « *p* », le nombre de variables initiales. Toutefois, cette réduction doit garder le maximum d'information. Les axes de ce sous-espace sont dits « *axes factoriels* » ou « *facteurs* ». Chaque variable « *p* »porte en elle, une part d'information originale ou part d'inertie et une part d'information originale redondante avec les autres, venant des corrélations entre variables.

C'est cette part d'information redondante qui va être regroupée dans le résumé factoriel. Les facteurs sont hiérarchisés de la manière suivante [151] :

- le 1er axe concentre le maximum de l'information : c'est l'axe de la plus grande dimension du nuage de points et il fournit le meilleur résumé dans un espace à une dimension, mais il laisse des résidus d'information ;
- le 2ème axe concentre le maximum de l'information restante, il est orthogonal au premier et c'est le meilleur résumé dans un espace à deux dimensions. Mais, de même il laisse aussi des résidus ;
- le 3ème axe prend encore une part d'information moindre, il est orthogonal aux deux premiers. Et ainsi de suite, pour les axes suivants tant que l'on pense qu'ils apportent encore de l'information.

Le nombre de composantes en théorie est égal au nombre de variables originelles. Mais, en pratique, les premières directions permettent de couvrir un pourcentage élevé (80-90%) de toutes les données originelles et sont donc utilisées pour restreindre l'espace d'observation [151].

2.2.2. Concepts de base

D'une façon générale, le traitement des données par ACP consiste à :

- calculer la moyenne de chaque vecteur de caractéristiques ;
- centrer les vecteurs pour réduire le temps du calcul ;
- former la matrice de covariance (matrice COV) ;
- calculer les valeurs et les vecteurs propres de la matrice de covariance ;
- utiliser les premiers vecteurs propres correspondant aux plus grandes valeurs propres pour former les axes principaux (nouvel espace propre).

L'analyse des données se fera à partir de la matrice de données de n lignes et p colonnes :

$$X = [X_1, X_2, \dots, X_j, \dots, X_p] = \begin{bmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{bmatrix} \quad (3.1)$$

Chaque variable $x = (x_{11}, \dots, x_{n1})$ a une moyenne $\bar{x}$ et un écart type σ_x.

Une fois la matrice de données définie, on utilise soit les variables centrées soit les variables centrées réduites pour la détermination des axes principaux. La matrice de données centrées est définie par [152-154] :

$$X_c = [X_{c1}, X_{c2}, \dots, X_{cj}, \dots, X_{cp}] = \begin{bmatrix} x_{11} - \overline{x_1} & \dots & x_{1p} - \overline{x_p} \\ \vdots & \ddots & \vdots \\ x_{n1} - \overline{x_1} & \dots & x_{np} - \overline{x_p} \end{bmatrix} \quad (3.2)$$

Avec : $\overline{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$

Elle peut être ainsi réduite :

$$\widetilde{X_c} = \begin{bmatrix} \frac{x_{11}-\overline{x_1}}{\sigma_{x_1}} & \dots & \frac{x_{1p}-\overline{x_p}}{\sigma_{x_p}} \\ \vdots & \ddots & \vdots \\ \frac{x_{n1}-\overline{x_1}}{\sigma_{x_1}} & \dots & \frac{x_{np}-\overline{x_p}}{\sigma_{x_p}} \end{bmatrix} \quad (3.3)$$

La matrice COV de p variables $X_1, \dots, X_j, \dots, X_p$ mesurées sur un ensemble de n unités est la matrice à p lignes et p colonnes contenant sur sa diagonale principale les variances des p variables, et ailleurs les covariances de ces variables qui sont données par [154] :

$$COV = \frac{1}{n} X_c^T X_c = \begin{bmatrix} \Sigma_{11} & \Sigma_{1p} \\ \Sigma_{p1} & \Sigma_{pp} \end{bmatrix} \quad (3.4)$$

Où :

$$\Sigma_{ij_{i=j(diagonal)}} = Var(X_j) = \frac{1}{n}\sum(x_{ij} - \overline{x_j})^2 \quad (3.5)$$

Et :

$$COV(X_k, X_m) = \frac{1}{n}\sum_{i=1}^{n}(x_{ik} - \overline{x_k})(x_{im} - \overline{x_m}) \quad (3.6)$$

Cette matrice de covariance est une matrice carrée, réelle et symétrique. Elle est définie positive si les p variables ne sont pas liées linéairement. Sa trace est égale à la somme des variances empiriques des p variables.

Si on veut travailler avec des variables centrées et réduites, on passe du tableau des valeurs centrées au tableau des valeurs centrées et réduites de la façon suivante :

$$X_{cr} = AX_c \tag{3.7}$$

A est la matrice diagonale qui a sur sa diagonale principale les inverses des écarts-type des variables. Elle est donnée par :

$$A = diag(\Sigma)^{\frac{-1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{\Sigma_{11}}} & \dots & 0 \\ \vdots & \frac{1}{\sqrt{\Sigma_{jj}}} & \vdots \\ 0 & \dots & \frac{1}{\sqrt{\Sigma_{pp}}} \end{bmatrix} \tag{3.8}$$

Si on calcule la matrice de covariance à partir d'un tableau de données centrées et réduites (cr), on obtient la matrice des corrélations empiriques :

$$R_c = \frac{1}{n} X_{cr}^t X_{cr} \tag{3.9}$$

Qui peut s'exprimer de la façon suivante : chaque terme R_{ckl}est égal à :

$$R = \frac{\Sigma_{kl}}{\sqrt{\Sigma_{kk}\Sigma_{ll}}} \; avec \; 1 \leq k \leq p \; et \; 1 \leq l \leq p \tag{3.10}$$

C'est à partir de cette matrice de covariance que l'on va trouver les axes factoriels. On peut de même travailler sur la matrice de corrélation et trouver de nouveaux axes factoriels. Pour cela on calcule les valeurs propres et les vecteurs propres associés. Les valeurs propres trouvées représentent la quantité d'information récupérée dans l'axe.

On remplace les anciens axes (donc les anciennes variables X_k) par de nouveaux axes (variables nouvelles Z_k). Ces nouvelles variables sont appelées composantes principales ; elles s'expriment comme des combinaisons linéaires des anciennes variables $X_1, X_2, .., X_j, \dots, X_p$.

La $k^{ième}$ composante principale de ce tableau de données est fournie par la transformationlinéaire :

$$Z_k = \sum_{j=1}^{p} V_{kj} V_{cj} = X_c V_k \tag{3.11}$$

Avec : $V_k^t = \left[v_{1k}, v_{2k}, \dots, v_{pk}\right] et \; Z_k^t = [z_{1K}, z_{2K}, \dots, z_{nK}]$

Un changement de base sur les variables produit le tableau :

$$Z = X_c V \tag{3.12}$$

où V est la matrice de changement de base :$\boldsymbol{V} = (V_1, .., V_j, .., V_p)$, de dimensions $(p \times p)$. Ce changement de base permet de déterminer les premières nouvelles variables (premières colonnes de Z) qui concentrent l'information (variables significatives) et qui possèdent de bonnes propriétés descriptives : absence de corrélation entre variables.

Le vecteur V_1 est choisi tel que var (Z_1) soit maximale, ceci revient à trouver V_1 qui maximise $V_1^t COV V_1$. Nous devons donc maximiser cette variance des observations projetées. Il s'agit d'un problème d'optimisation sous contrainte. On utilise n'importe quel algorithme de détermination des vecteurs propres pour trouver les valeurs propres de la matrice de covariance des données :

$$COV\ V_i = \lambda V_i \tag{3.13}$$

Où λ est la valeur propre et V_i le vecteur propre correspondant. Chacune de ces nouvelles variables Z_i étant non corrélée entre elles et de variance maximale. Par construction, ces variances sont décroissantes :$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$**.** D'une façon générale, on a :

$$var[Z_k] = V_k^t \Sigma V_k = \lambda_k \tag{3.14}$$

La $k^{ième}$ plus grande valeur propre de COV est la variance de la $k^{ième}$ composante principale.

Comme la matrice variance-covariance est symétrique définie positive, les valeurs propres sont réelles positives et les vecteurs propres peuvent être choisis d'une manière orthonormée. Si la $r^{ième}$ variance var Z_{r+1} est très petite, les variables $Z_{r+1}, Z_{r+2}, \ldots, Z_p$ sont quasiment constantes pour tous les individus.

Il est alors naturel de ne conserver que les composantes principales $Z_1, Z_2, \ldots, Z_r$. Dans la pratique, on estime que Z_{r+1} est petite si on a [155] :

$$\frac{(\lambda_1+\lambda_2+\cdots+\lambda_r)}{(\lambda_1+\lambda_2+\cdots+\lambda_p)} \approx 90\% \tag{3.15}$$

Dans le meilleur des cas, 2 à 3 composantes principales suffisent et, dans ce cas, on est passé de p variables corrélées à 2 ou 3 variables non corrélées qu'on peut facilement représenter à l'aide de graphiques.

Enfin le but de l'ACP est donc d'extraire de nouvelles variables classées par ordre d'importance décroissante. On estime alors que l'usage de plusieurs paramètres descripteurs

(paramètres physico-chimiques) en entrée d'une chaîne de contrôle et de surveillance de l'eau, ne laisse aucun doute à envisager leur restriction (réduction de la dimension), et ce pour des raisons aussi bien technique qu'économique.

2.3.Méthodes d'apprentissage

2.3.1. Les réseaux de neurones artificiels

Les réseaux de neurones artificiels ont été inventés par Mc Culloch et Pitt dans les années 40 qui ont proposé les premières notions de neurone formel (figure 3.3) [156]. Ils ont été rendus populaires dans [157]. Les RNA sont été utilisés par conséquent dans de nombreux domaines d'application tels que : le diagnostic, le traitement du signal, la reconnaissance de formes, la reconnaissance de la parole, la compression de données, le contrôle, la prévision, la régression, l'analyse de données, etc. Un réseau neuronal est essentiellement un outil mathématique qui permet d'identifier des fonctions mathématiques. Les réseaux neuronaux sont des outils qui sont donc bien adaptés pour aider les spécialistes dans le domaine de surveillance et de contrôle. Ils devraient être particulièrement performants pour effectuer des tâches de régression et de classification dans le domaine de traitement des eaux.

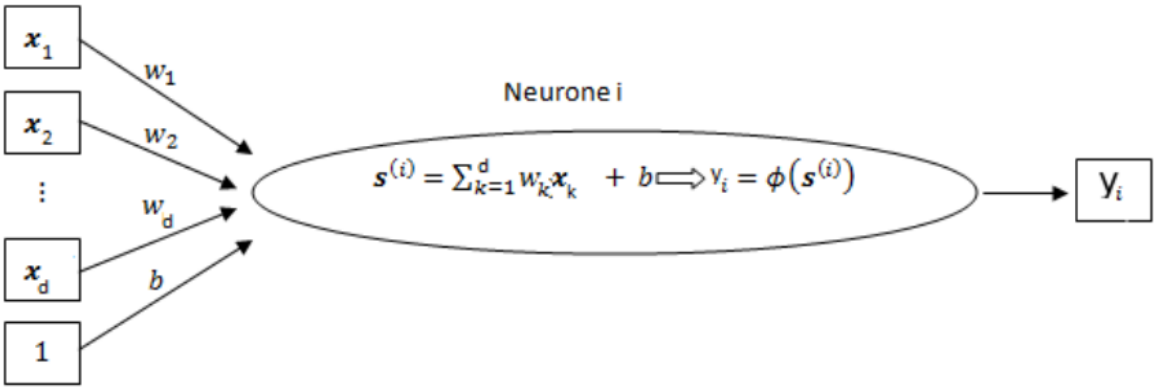

Fig.3.3. Éléments constituant un neurone formel.

Les RNAs offrent l'avantage de construire des approximations simples et dérivables. Un réseau de neurones est constitué d'un ou de plusieurs neurones formels, ceci forme un graphe plus ou moins complexe. Un neurone, comme montré dans la figure 3.3, est constitué :

- des variables d'entrée qui proviennent soit directement des observations (pour $i = 1, \ldots, n, (x_1^{(i)}, \ldots, x_d^{(i)})$), soit d'autres neurones ;

- d'un biais b qui est toujours égal à 1. Il permet d'ajouter de la flexibilité lors de l'apprentissage ;
- des poids w, $w_i(i = 1, \ldots, d)$ qui affectent l'influence de chaque variable d'entrée sur la sortie du neurone ;
- du cœur ou du noyau du neurone qui intègre une fonction d'activation ϕappliquée à :

$$s = \sum_{k=1}^{d} w_k x_k + b \quad (3.16)$$

- de la sortie y qui peut être la sortie finale ou une sortie vers d'autres neurones.

L'évaluation de ϕ, qui est intégrée dans le noyau d'un neurone artificiel, se fait en deuxétapes : tout d'abord, la combinaison linéaire des variables d'entrée s est calculée, ce qui correspond à la sommation des informations reçues, ensuite la fonctionϕ est appliquée sur s pour indiquer l'état du neurone et transférer le résultat soit au neuronesuivant, soit à la sortie finale. Les paramètres $(w_1, \ldots, w_d, b)$ sont estimés au cours de l'élaboration du réseau, ce qui est appelée, la phase d'apprentissage [158].

Les différents types de neurones se différencient suivant la nature de ϕ. Les principaux types de fonctions d'activation existant dans la littérature sont comme suit :

- linéaire : $\emptyset = la\ fonction\ identité$;
- sigmoïde : $\emptyset = \frac{1}{1+e^x}$;
- seuil :$\emptyset = 1_{[0,+\infty[}(x), avec\ 1_{[0,+\infty[}(.)$la fonction indicatrice sur $[0, +\infty[$;
- radiale : $\emptyset = \sqrt{\frac{1}{2\pi}} e^{\frac{-x^2}{2}}$.

Quelle que soit la fonction d'activation, les réseaux de neurones artificiels ne sont pas interpolant.

2.3.1.1.Les perceptrons multicouches

La figure 3.4 montre la composition générique d'un perceptron multicouche, en anglais Multi-Layer Perceptron (MLP). Dans ce type de réseaux, les neurones sont organisés de manière à constituer plusieurs couches. La couche d'entrée contient autant de neurones que la dimension de l'espace d'entrée. La couche de sortie contient un neurone ou plusieurs neurones, selon le type d'application utilisée. Enfin, une ou plusieurs couches cachées contiennent un nombre variable de neurones.

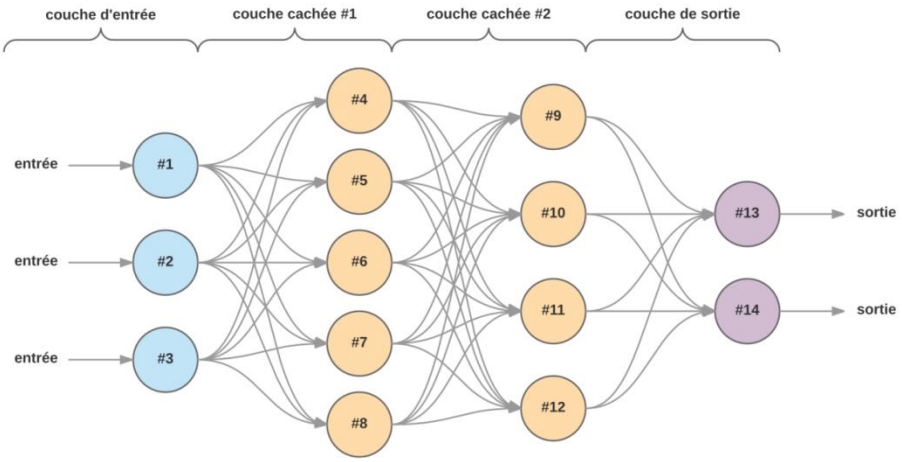

Fig.3.4. Les réseaux de neurones artificiels multicouches.

a- Etape d'apprentissage

Les RNAs sont l'outil puissant pour résoudre différents problèmes tels que la classification ou la prédiction. Les perceptrons multicouches ont été utilisés pour résoudre des problèmes difficiles par apprentissage supervisé à partir d'un algorithme connu sous le nom d'algorithme de rétropropagation du gradient qui se base sur la règle d'apprentissage par correction d'erreur. L'amélioration s'obtient avec le temps à partir de mesures réelles effectuées. Un réseau de neurones apprend de son environnement à partir de processus itératifs en ajustant les poids synaptiques et les seuils. Théoriquement, le réseau s'améliore après chaque itération d'apprentissage [159-162].

Le principe d'apprentissage est l'estimation des paramètres $\alpha_{j=0,p,k=1,q}$ et $\beta_{k=0,q}$ par la minimisation d'une fonction coût (fonction perte) qui représente le but de l'apprentissage. Il faut noter que les deux paramètres α et β sont les poids des entrées à estimer lors de la procédure d'apprentissage [8,163].

$$Q(\alpha,\beta)=\sum_{i=1}^{n}Q_i=\sum_{i=1}^{n}\left[y_i-\Phi(x;\alpha,\beta)\right]^2 \tag{3.17}$$

b- Rétropropagation de l'erreur

L'algorithme de rétropropagation de l'erreur du gradient est une généralisation de la règle de *Windrow-Hoff* pour un réseau multicouche que l'on désigne par « back-propagation » [164].

Il figure aujourd'hui parmi les algorithmes d'apprentissage le plus utilisés. Il a été appliqué avec succès à une grande variété de problèmes tels que la prévision de consommation d'eau [165].

Il s'agit donc d'évaluer la dérivée de la fonction coût en une observation et par rapport aux différents paramètres.

Soit :

$$y = \phi(x;\alpha,\beta) = \beta_0 + \beta^{'} z \tag{3.18}$$

Avec : $z_{ki} = f(\alpha_{k0} + \alpha_{k'} x_i)$ et $z_i = \{z_{1i},...,z_{qi}\}$ (3.19)

Les dérivées partielles de la fonction perte quadratique s'écrivent [163] :

$$\frac{\partial Q_i}{\partial \beta_k} = -2(y_i - \phi(x_i))(\beta^{'} z_i) z_{ki} = \delta_i z_{ki} \tag{3.20}$$

$$\frac{\partial Q_i}{\partial \alpha_{kj}} = -2(y_i - \phi(x_i))(\beta^{'} z_i)\beta_k f^{'}(\alpha_k^{'} x_i) x_{ip} = s_{ki} x_{ip} \tag{3.21}$$

Les termes δ_i et s_{ki} sont respectivement les termes d'erreur du modèle courant à la sortie et sur chaque neurone caché. Ces termes d'erreur vérifient les équations dites de rétropropagation [163] :

$$s_{ki} = f^{'}(\alpha_k^{'} x_i)\beta_i \delta_i \tag{3.22}$$

dont les termes sont évalués en deux passes. Une passe "*avant*", avec les valeurs courantes des poids, l'application des différentes entrées x_i au réseau permet de déterminer les valeurs ajustées $\hat{\phi}(x_i)$. La passe "*retour*" permet ensuite de déterminer les δ_i qui sont "rétro propagés" afin de calculer les s_{ki} et ainsi obtenir les évaluations des gradients.

c- Etape de généralisation

Une fois le réseau de neurones entraîné (phase d'apprentissage), il est nécessaire de le tester sur une base de données différente de celle utilisée pour l'apprentissage. Ce test permet à la fois d'apprécier les performances du système neuronal et de détecter le type de données qui pose problème. Si les performances ne sont pas satisfaisantes, il faudra soit modifier l'architecture du réseau, soit modifier la base d'apprentissage.

2.3.1.2.Les réseaux de neurones RBF

Le modèle d'approximation de fonctions à base radiale, appelé aussi RBF selon la terminologie anglo-saxonne généralement utilisée, est simple à mettre en œuvre [158,166]. Ce métamodèle est souvent utilisé pour les problèmes d'optimisation de grande dimension du fait de sa rapidité et de sa performance [158] [167-171]. Ce métamodèle consiste, dans un cadre très général, à combiner un ensemble de fonctions de base entre elles.

Une fonction de base radiale est une fonction $\emptyset_j$ symétrique autour d'un centre $C^{(j)} i.e. \emptyset_j(x) = \emptyset\left(\left\|x - C^{(j)}\right\|_2\right)$, où$\|\,.\,\|_2$est la norme euclidienne usuelle. De manière à avoir un modèle interpolant, les points d'apprentissage sont souvent choisis comme centres des fonctions de base. L'expression du modèle RBF est donnée par :

$$\hat{y}(x) = \Phi(x)\beta + p(x) = \sum_{i=1}^{n} \beta_i \, \phi_i(x) + p(x) \qquad (3.23)$$

avec :$\beta = [\beta_1, \dots, \beta_n]$, $\Phi(x) = [\phi_1(x), \dots, \phi_n(x)]$ et $p(x)$une tendance linéaire sur les d variables. Il existe plusieurs types de fonctions radiales dans la littérature [166] comme les fonctions (on note $h = \left\|x - C^{(j)}\right\|_2$) :

- cubiques :h^3 ;
- thin plate splines:$h^2 \ln(h)$;
- multi quadratiques: $\sqrt{(1 + \varepsilon h)}$, avec un paramètre $\varepsilon \in \mathbb{R}$;
- gaussiennes:$e^{-(\varepsilon h)^2}$, avec un paramètre $\varepsilon \in \mathbb{R}$.

❖ **Architecture générale d'un réseau RBF**

Le réseau RBF est un réseau à trois couches : une couche d'entrée, une couche cachée composée de fonctions noyaux et une couche de sortie, dont les neurones sont généralement animés par une fonction d'activation linéaire. Le nombre de neurones des couches d'entrée et de sortie est choisi en fonction du problème à traiter. Les poids entre la couche d'entrée et la couche cachée dans les modèles neuronaux de type RBF sont toujours d'une valeur unité, c'est-à-dire que l'information inscrite sur la couche d'entrée sera retransmise sans distorsion vers les neurones de la couche cachée. En ce qui concerne la couche cachée, la technique RBF prévoit une fonction noyau, c'est-à-dire un neurone, par point de donnée apprise. En d'autres termes, le nombre de neurones de la couche cachée est égal au nombre d'exemples présentées au réseau lors de la phase d'apprentissage.

❖ Apprentissage des modèles RBF

L'apprentissage d'un modèle RBF consiste à déterminer son architecture (le nombre N de fonctions radiales) et à fixer les valeurs des paramètres. L'apprentissage d'un réseau RBF est de type *supervisé* : on dispose d'un ensemble d'apprentissage constitué de l couples (vecteur d'entrée, valeur cible) :$(x_1, y_1), \dots, (x_m, y_m), x_i \in \mathbb{R}^d, y_i \in \mathbb{R}$., et du coût associé à chaque exemple :

$$E_i = \frac{1}{2}(y_i - F(x_i))^2 \qquad (3.24)$$

Une caractéristique intéressante des modèles RBF est que l'on peut diviser les paramètres en trois groupes : les centres μ, les largeurs σ et les poids w [172]. En pratique, il existe deux grandes approches distinctes pour entraîner un réseau RBF.

La première approche, qui est également la plus populaire, prône qu'il est raisonnable de séparer l'apprentissage du RBF en deux processus distincts puisque les deux couches de neurones du réseau accomplissent des tâches distinctes. Selon cette approche, les neurones radiaux sont d'abord positionnés dans l'espace d'entrée à l'aide d'un algorithme d'apprentissage non supervisé, puis, par la suite, un algorithme supervisé est utilisé afin d'ajuster les coefficients synaptiques qui relient les neurones de la couche de sortie du réseau. A titre indicatif, J. Moody et C. Darken [172] suggèrent de déterminer la position des neurones selon un algorithme de type k-means. Ils se fondent sur certaines heuristiques afin de pouvoir estimer la valeur d'écart-type qui doit être assignée à chaque neurone radial. Par exemple, une fois les neurones correctement positionnés, les coefficients synaptiques de la couche de sortie sont déterminés selon une méthode basée sur le calcul du gradient de l'erreur, ou encore, si les données sont limitées, en nombre, selon une méthode pseudo-inverse. Durant cette seconde étape, les paramètres des neurones radiaux ne sont plus ajustables.

La seconde approche consiste plutôt en un apprentissage pleinement supervisé, basé sur la méthode du gradient qui considère tous les paramètres du réseau en bloc. Des travaux ont montré expérimentalement que les capacités de généralisation de cette approche sont comparables à celles du réseau MLP [173]. Toutefois, en pratique, ce type d'apprentissage pleinement supervisé a souvent été négligé principalement parce qu'il est beaucoup moins rapide que l'approche hybride non supervisée/supervisée décrite précédemment [174,175].

2.3.2. Les Machines à Vecteurs de Support pour la régression

Les Machines à Vecteurs de Support sont une famille de classifieurs supervisés qui ont été introduits dans les années 1990 par Vapnik et al. [176,177] et qui depuis sont devenus parmi les modèles les plus utilisés dans la classification [178-192].

Les SVMs opèrent sur des données plongées dans un espace vectoriel, et le but est de les catégoriser. L'idée clé est d'augmenter la dimension de la représentation de ces données, et ensuite de calculer la surface de séparation entre les classes dans cet espace de représentation (appelé "hyperplan"), partant du principe que les données sont linéairement séparables dans cet espace de grande dimension (principe appelé : « principe de Cover ») [193].

Dans leur origine, les SVMs ont été développées pour des problèmes de classification. Une formulation des SVMs pour la régression, appelée SVR, existe aussi. Cette méthode a récemment été appliquée avec succès sur un certain nombre de problèmes de régression. La régression est un cas particulier de classification où les classes des exemples ne sont pas dénombrables c-à-d continues. Le problème consiste à trouver, en utilisant $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$ une fonction $\hat{f}: \mathbb{R}^m \to \mathbb{R}$ qui rapproche le plus possible des y_i, en d'autres termes qui minimise la différence entre les $\hat{f}(x_i)$ et les y_i.

Souvent, $\hat{f}$ est considérée comme fonction linéaire :$\hat{f} = \langle w, x \rangle + b$, où $w$est un vecteur et$b$est un scalaire. Le problème revient donc à trouver un hyperplan caractérisé par w^*et b^*qui minimise l'écart global entre$\hat{f}$ et lesy_i (équation 3.25).

$$(w^*, b^*) = argmin_{\omega,b} \sum_{i=1}^{n} (y_i - \langle w, x_i \rangle - b)^2 \quad (3.25)$$

Pour résoudre ce problème, les SVMs utilisent une astuce semblable à celle utilisée en classification [180] [194-196]. On propose de modéliser la fonction de régression par un hyperplan qui se situe au centre d'un hyper-tube de largeur 2ε contenant tous les exemples d'entrainement (figure 3.5).

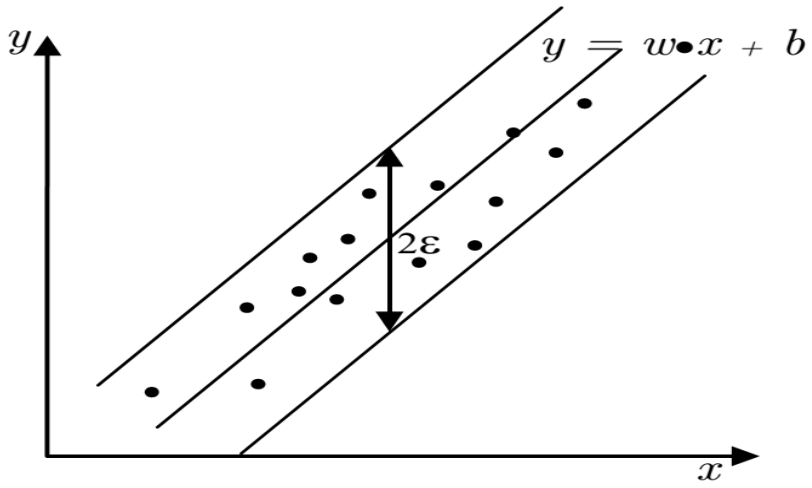

Fig.3.5. Hyper-tube modélisant la fonction de régression.

Plusieurs hyper-tubes, de largeur 2ε contenant tous les exemples d'entrainement, peuvent exister. L'hyper-tube optimal est celui qui minimise la distance entre les exemples d'entrainement et ses frontières, autrement dit, qui maximise la distance des exemples de l'hyperplan du centre. La détermination de l'hyper-tube optimal est semblable à la détermination de l'hyperplan optimal de marge maximale dans le cas de classification. On doit donc rechercher un hyper-tube de marge maximale avec tous les exemples d'entrainement à l'intérieur. Par une analyse similaire à celle du problème de classification, la solution du problème de régression est réduite à la résolution du problème dual d'optimisation quadratique de l'équation (3.26).

$$\begin{cases} Maximiser & \frac{-1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i')(\alpha_j - \alpha_j')\langle x_i, x_j\rangle \\ & -\varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i') + \sum_{i=1}^{n} y(\alpha_i - \alpha_i') \\ sous\ contraintes & \sum_{i=1}^{n}(\alpha_i - \alpha_i') = 0 \\ & 0 \le \alpha_i, \alpha_i' \le C \end{cases} \quad (3.26)$$

Où les $\alpha_i$et les α_i' sont les coefficients des exemples respectivement au-dessus et au-dessous de l'hyperplan et Cest un paramètre de pénalisation. La fonction de sortie $f^(x)$ peut être donnée par l'équation (3.27).

$$\hat{f}(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i')\langle x_i, x\rangle + b \quad (3.27)$$

Où b est calculé à partir d'un exemple dont $0 < \alpha_i < C$ (vecteur support) par l'équation (3.28).

$$b = y_i - \langle w, x_i\rangle - \varepsilon \quad (3.28)$$

- **Utilisation des noyaux**

Parmi les motivations les plus solides du développement des SVMs pour la régression, est leur extension simple aux cas non linéaires, grâce à l'utilisation des noyaux. En effet, d'une manière similaire au cas de classification, on fait une transformation d'espace $\emptyset$ pour se trouver toujours face à une régression linéaire. La transformation d'espace inverse ϕ^{-1}, permet de retourner à l'espace d'origine après la résolution dans le nouvel espace (figure 3.6) [180].

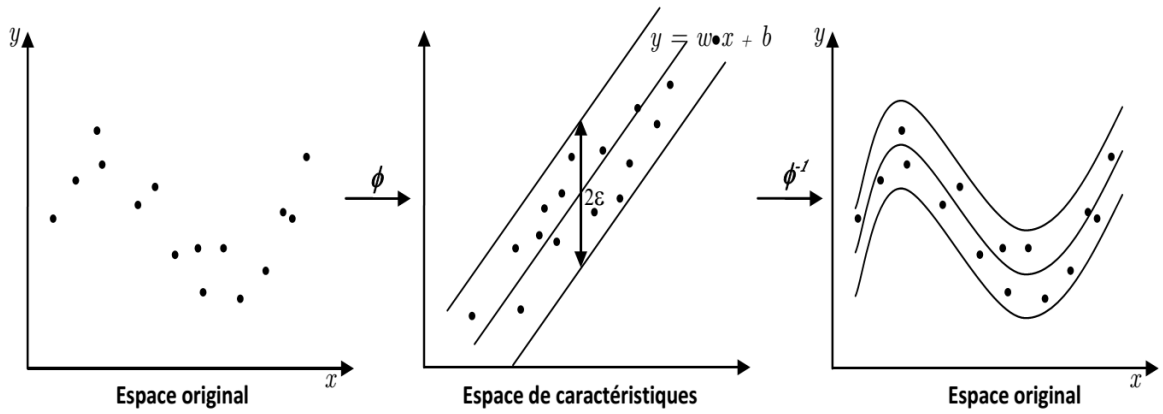

Fig.3.6. Utilisation des noyaux pour la résolution de la régression non linéaire.

Comme dans le cas de classification, la transformation φ et son inverse sont réalisées grâce à une fonction réelle $K(x_i, x_j)$ appelée *Noyau (Kernel)*. Le produit scalaire dans les équations (3.26) et (3.27) est remplacé par la fonction du noyau.

2.3.3. Les Machines d'apprentissage extrême

Les réseaux de neurones artificiels, ont été largement utilisés dans de nombreux domaines en raison de leur capacité à approximer des fonctions non linéaires complexes directement à partir des données et pour fournir des modèles pour une grande classe de phénomènes naturels et artificiels qui sont difficiles à manipuler en utilisant des techniques paramétriques classiques. D'autre part, il y a un manque d'algorithmes d'apprentissage rapides pour les réseaux de neurones (RNs).

Les algorithmes d'apprentissage traditionnels sont généralement beaucoup trop lents que nécessaire. On peut voir que cela peut prendre plusieurs heures, plusieurs jours, et encore plus de temps pour entraîner les RNs en utilisant des méthodes traditionnelles. D'un point de vue

mathématique, la recherche sur les capacités d'approximation des RNs à propagation avant, a mis l'accent sur deux aspects :

- approximation universelle sur ensembles d'entrée compacts ;
- approximation dans un ensemble fini des échantillons de formation.

De nombreux chercheurs ont exploré la capacité d'approximation universelle des RNs à propagation avant. Traditionnellement, tous les paramètres des réseaux doivent être ajustés et donc il existe une dépendance entre les différentes couches de paramètres (poids et biais).

Dans les dernières décennies, différentes méthodes du gradient ont été utilisés dans divers algorithmes d'apprentissage des RNs à propagation avant. Cependant, il est clair que les méthodes d'apprentissage fondées sur la descente du gradient sont généralement très lentes, ils peuvent aussi facilement converger vers des minimas locaux. Beaucoup d'étapes itératives peuvent être exigées par ces algorithmes d'apprentissage afin d'obtenir de meilleures performances d'apprentissage. Il a été démontré, que des réseaux à une seule couche cachée (avec Nœuds cachés) avec des poids d'entrée et des biais choisis aléatoirement peuvent apprendre exactement N observations distinctes. Contrairement à la croyance établie et aux mises en œuvre pratiques, que tous les paramètres des réseaux doivent être ajustés, il n'est pas nécessaire d'ajuster les poids d'entrée et les biais de la couche cachée.

Après que les poids et les biais de la couche cachée soient fixés aléatoirement, les réseaux peuvent être tout simplement considérés comme un système linéaire. Les poids de sortie (reliant la couche cachée à la couche de sortie) peuvent être déterminés analytiquement par l'opération de l'inverse généralisée de la matrice de sortie de la couche cachée. Sur la base de ce concept, il est possible de construire un algorithme d'apprentissage pour les réseaux à une seule couche cachée appelé apprentissage ELM dont la vitesse d'apprentissage peut être des milliers de fois plus rapide que les algorithmes d'apprentissage traditionnels comme l'algorithme de rétropropagation, l'obtention d'une meilleure performance de généralisation est également garantie [29] [197-199].Cet algorithme, non seulement, tend à atteindre la plus petite erreur, mais il atteint aussi la plus petite norme de poids et il peut être :

- facilement mis en œuvre ;
- atteindre l'erreur d'entraînement la plus petite ;
- obtenir la plus petite norme de poids ;
- avoir une bonne performance de généralisation ;
- fonctionner de manière extrêmement rapide en phase d'entraînement.

2.3.3.1.Théorie de la technique ELM

Le principe du réseau neuronal n'est pas modifié, mais le rôle de l'adaptation est reconsidéré, ainsi que le nombre de couches cachées qui se restreint à une seule couche. En effet, plutôt que d'ajuster tous les poids d'un réseau pour émuler une fonction, le réseau est constitué d'un grand nombre de neurones dans la couche interne. Les poids d'entrée sont initialisés aléatoirement une seule fois et restent avec cette valeur. L'adaptation, qui se fait en une seule fois aussi, porte donc uniquement sur les poids de la couche de sortie.

On considère, en premier lieu, un problème d'apprentissage supervisé avec un ensemble d'apprentissage de N signaux, $(X, Y) = (x_i, y_i)_N$ avec $x_i \in \mathbb{R}^d$ et $y_i$étant la classe correspondante de la donnéex_i.

- Cas de la multi-classification : y_i est un vecteur binaire de taille n_0, et prend une valeur égale à 1 en une seule position.
- Cas de la régression : $y_i \in \mathbb{R}^d$.

L'ELM vise à apprendre une règle de décision ou une fonction d'approximation sur la base des données d'entraînement [33,200,201].

En général, l'entraînement se compose de deux étapes. La première étape est de construire la couche cachée en utilisant un nombre fixe de neurones dont les poids sont générés aléatoirement, et de fonctions non linéaires continues par morceaux.

L'algorithme ELM évite la procédure lente d'entraînement itératif et ne nécessite qu'une seule opération pour apprendre le réseau de neurone SLFN (Single Layer FeedForward Neural Network) [29]. Ceci est principalement dû au fait qu'il n'y a pas de biais pour le neurone de la couche de sortie (figure 3.7).

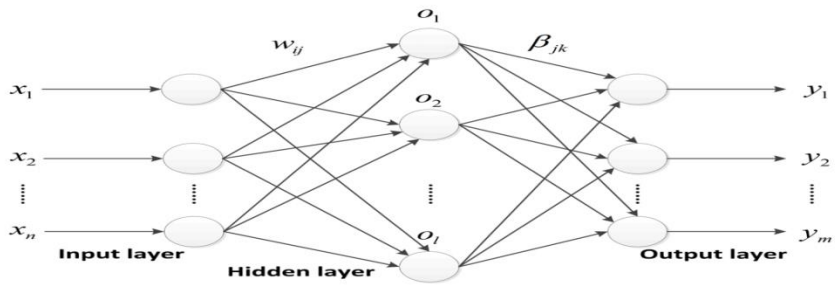

Fig.3.7. Structure du réseau de neurone SLFN.

En bref, pour initier une opération d'apprentissage en une passe, les paramètres du réseau de neurones cachés (poids et biais) sont générés de manière aléatoire sans aucune connaissance préalable ou procédure d'apprentissage. Par conséquent, l'ELM se transforme en

un système d'équations linéaires et les poids inconnus entre les neurones de la couche cachée et de sortie peuvent être obtenus analytiquement en appliquant seulement la procédure inverse généralisée de Moore-Penrose [202].

Notons n et m les nombres d'entrées et de sorties (i,e., les cibles), N le nombre d'échantillons de données d'apprentissage (x_i, t_i), où $i \in [1 \dots N]$, $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbb{R}^n$ and $t_i = [t_{i1}, t_{i2}, , \dots, t_{im}]^T \in \mathbb{R}^m$, et $\widetilde{N}$ le nombre de nœuds cachés, chacun avec unefonction d'activation $f(x)$. Pour chaque échantillon j, la sortie o_j est mathématiquement exprimée comme :

$$\sum_{k=1}^{\widetilde{N}} \beta_k . f\left(w_k . x_j + b_k\right) = t_j, j = 1,2, \dots, N \tag{3.29}$$

qui peut être exprimée sous forme matricielle en tant que,

$$H\beta = T \tag{3.30}$$

$$H(w_1, \dots, w_{\widetilde{N}}, x_1, \dots, x_{\widetilde{N}}, b_1, \dots, b_{\widetilde{N}}) = \begin{bmatrix} f(w_1 . x_1 + b_1) & \dots & f(w_{\widetilde{N}} . x_1 + b_{\widetilde{N}}) \\ \vdots & \dots & \vdots \\ f(w_1 . x_N + b_1) & \dots & f(w_{\widetilde{N}} . x_N + b_{\widetilde{N}}) \end{bmatrix}_{NX\widetilde{N}} \tag{3.31}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\widetilde{N}}^T \end{bmatrix}_{\widetilde{N}Xm} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{NXm} \tag{3.32}$$

La solution des moindres carrés du système linéaire défini dans l'équation (3.30), avec la norme minimale (grandeur) des poids de sortie β est :

$$\hat{\beta} = H^{\dagger}T = (H^T H)^{-1} H^T T \tag{3.33}$$

Où $H^{\dagger}$ représente la solution inverse généralisée de Moore-Penrose pour la matrice de sortie de couche cachée H [203]. Enfin, l'entraînement de l'ELM peut être résumé dans l'algorithme suivant.

2.3.3.2.L'algorithme ELM

Étant donné un ensemble d'observations n entrées, m sorties,$\widetilde{N}$ neurones cachés $\left(k = 1 \dots \widetilde{N}\right)$,la fonction d'activation $f(x)$ et N échantillons de données d'apprentissage.

Étape 01 : Assigner aléatoirement des valeurs aux poids d'entrée $w_k$et aux biais b_k.

Étape 02 : Calculer la matrice H de sortie de couche cachée.

Étape 03 : Calculer les poids de sortie $\beta = H^{\dagger}T$, tel que :$T = [t_1, \ldots, t_N]^T$.

CONCLUSION

Ce chapitre a fait l'objet d'une présentation des techniques de développement des capteurs logiciels vus comme moyen intéressant de remplacement des capteurs physiques. Les techniques utilisées pour développer ces capteurs sont des méthodes paramétriques non linéaires qui relient des variables d'entrée à des variables de sortie, appliquées à la régression et la classification. Des généralités sur les réseaux RNAs et RBF ont été décrits à cet effet. Un exposé non exhaustif des bases théoriques de l'apprentissage statistique et de la méthode SVM en particulier, a été présenté. Une attention particulière à la technique ELM, a été toutefois donnée en rappelant les concepts les plus importants pour sa mise en œuvre en tant qu'outil de régression et de classification.

Une étude et analyse des performances apportées par ces techniques appliquées au domaine de surveillance et de contrôle de la qualité des eaux propres, constitue notre principal objectif dans ce travail. D'ailleurs, une simulation ayant pour but l'évaluation des performances de ces techniques, fera l'objet du chapitre suivant. L'évaluation des résultats obtenus, reflétera les performances attendues de chacune de ces méthodes et conduira au choix de la technique la mieux adaptée à l'application. L'architecture du système de contrôle et de surveillance à base du capteur logiciel à proposer dans ce cadre, sera présentée.

CHAPITRE IV

SIMULATION ET EVALUATION DES PERFORMANCES

INTRODUCTION

Dans un domaine d'application donné, la résolution d'un problème de régression et/ou de classification s'effectue en comparant des modèles afin de choisir le plus apte à résoudre le problème posé. L'évaluation des modèles est donc un préalable inévitable à la sélection. L'état de l'art proposé dans le chapitre précédent montre le nombre important d'approches, tant pour la classification que pour la régression.

Ce dernier chapitre est consacré à une étude en simulation et vise principalement l'évaluation des performances des techniques proposées comme étant une solution dans la conception de capteurs logiciels utilisés dans le domaine de surveillance de la qualité des eaux propres. Ce travail présente donc une contribution à l'étude et au développement de capteurs logiciels utilisés dans ce domaine. Une étude comparative entre les techniques MLP, RBF, SVM et ELM en termes de temps d'apprentissage et d'autres paramètres de régression et de classification est alors présentée. L'objectif est de mettre en place un système multicapteur à base de capteurs logiciels afin de prendre une décision adaptée au contrôle et au suivi des problèmes de la qualité de l'eau. Un exemple d'application est prévu à cet effet ; il s'agit de développer un capteur logiciel de Chlore vu comme un problème de régression, permettant de contribuer au contrôle de la potabilité de l'eau, vu cette fois comme un problème de classification.

1. PROBLEMATIQUE

Une usine moderne de production d'eau potable assure en fait deux principales fonctions : la satisfaction de la demande en eau, et l'assurance d'un niveau de qualité élevé et uniforme [7]. Des systèmes de surveillance permanents doivent alors assurer le contrôle des divers procédés et installations de traitement, particulièrement les paramètres relatifs à la qualité de l'eau en sortie de la station de production. Dans la surveillance de la qualité de l'eau, comme dans d'autres domaines similaires, un contrôle efficace dépend principalement

de la précision et de la fiabilité des capteurs de mesure [9]. D'autre part, l'eau est considérée comme une entrée de processus et a plusieurs paramètres à contrôler et à surveiller. De plus, l'hétérogénéité des variables mesurées par chaque capteur est considérée comme un réel défi en matière de qualité de l'eau [10]. Les variables de qualité jouent donc un rôle important pour assurer le produit de ressource vitale. Bien que certains paramètres puissent être mesurés en continu à l'aide de capteurs à faible coût, il existe d'autres paramètres nécessitant des analyses de laboratoire spécifiques et coûteuses en raison de l'absence de capteurs dédiés [11]. En fait, et ce pour plusieurs raisons, telles que le coût élevé de certains capteurs, leur nombre, le temps de vérification, l'opération de nettoyage, les routines d'étalonnage et de remplacement, fait que leur bon fonctionnement est difficile à assurer pour un suivi rigoureux de la qualité de l'eau.

Selon les problèmes mentionnés ci-dessus, le capteur logiciel offre une aubaine très attractive pour traiter ce manque de capteurs spécifiques [12-15]. La construction de capteurs logiciels peut être vue dans ce cadre comme un problème d'apprentissage de relations entre variables à partir de données observées. Ils sont spécialement programmés avec des algorithmes dérivés de l'IA où le problème à résoudre serait vu comme un problème de régression.

Il s'agit dans cette partie de travail d'évaluer les performances des quatre techniques énoncées précédemment qui sont issues, rappelons-le, du domaine de l'IA à savoir, les RNAs, les RBF, les SVMs et l'ELM. Des techniques servant comme outils de base pour l'aide à la décision et présentant une réponse plus élaborée par rapport aux autres techniques se basant sur des données brutes, venant directement des variables de surveillance, ou à partir de données traitées venant des sorties de traitements de bas niveau. Le choix effectué sur la base des résultats obtenus, conduira à l'intégration de la technique sélectionnée au niveau d'un système de surveillance assurant une surveillance et un contrôle permanent de la qualité de l'eau.

2. ARCHITECTURE DU SYSTEME DE SURVEILLANCE

Le schéma présenté dans la figure 4.1, présente l'architecture du système multicapteur dédié à la surveillance et au contrôle de la qualité des eaux propres. Il se compose d'une entrée des données obtenues à partir de capteurs physico-chimiques, d'un système d'acquisition de données et de logiciels permettant le traitement des données et la prise de décision quant à la qualité de l'eau surveillée.

Les capteurs logiciels sont en fait utilisés pour diverses raisons (comme déjà souligné), à savoir le prix élevé, le grand nombre de capteurs, l'opération de nettoyage, les routines d'étalonnage, le remplacement des capteurs ainsi que le temps consacré à la vérification de ces capteurs, pour n'en citer que quelques-uns.

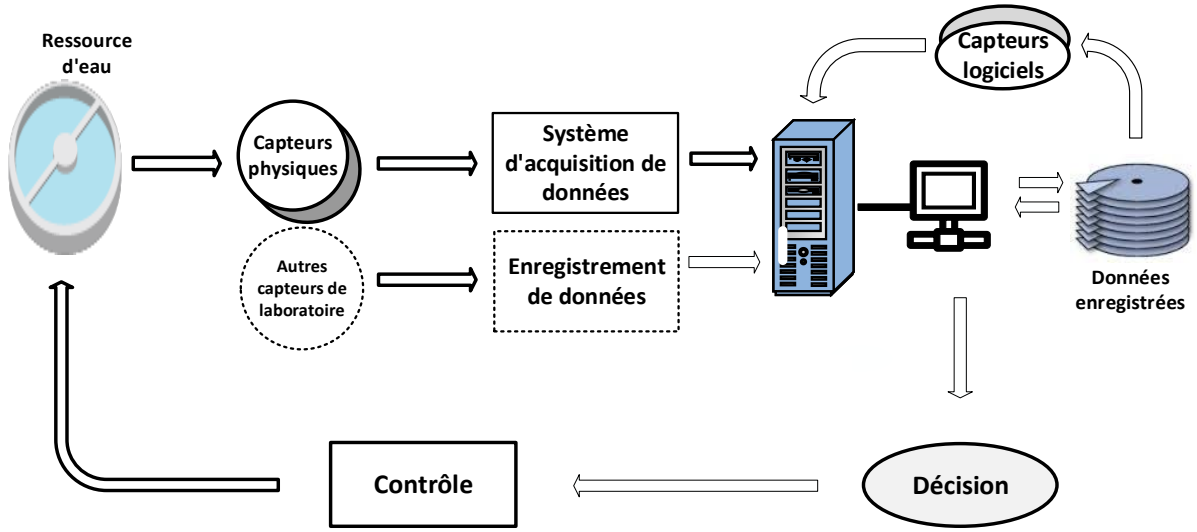

Fig.4.1. Architecture du système pour la surveillance et le contrôle de la qualité des eaux propres.

La décision à base de capteurs logiciels est composée de capteurs physiques et logiciels comme il est illustré dans la figure 4.2 où (n) désigne le nombre total de paramètres physico-chimiques (pH, T°, C, TU, OD, Cl, etc.) qui existent dans la ressource en eau à surveiller. Avec cette approche, il pourrait être possible de reproduire la sortie de défaillance des capteurs ou/et certains capteurs de laboratoire.

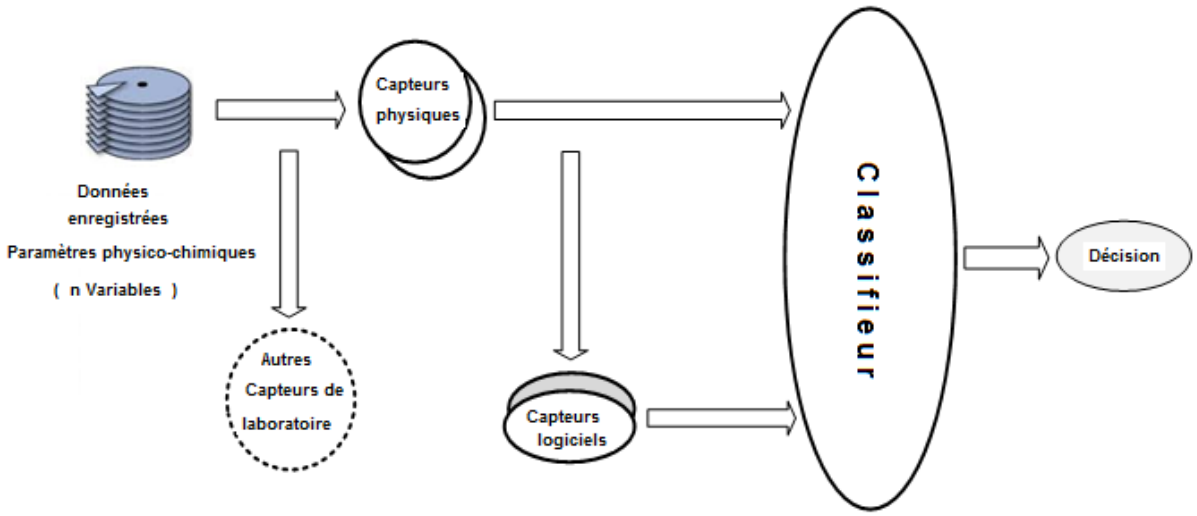

Fig.4.2. Décision à base de capteurs logiciels.

3. DEVELOPPEMENT DU CAPTEUR LOGICIEL

Le développement d'un capteur logiciel consiste à associer des mesures disponibles ou facilement réalisables, représentatives de l'évolution du procédé étudié, et des modèles mathématiques reliant les grandeurs mesurées et les grandeurs à déterminer. Le développement d'un capteur logiciel comprend les étapes principales suivantes (fig. 4.3).

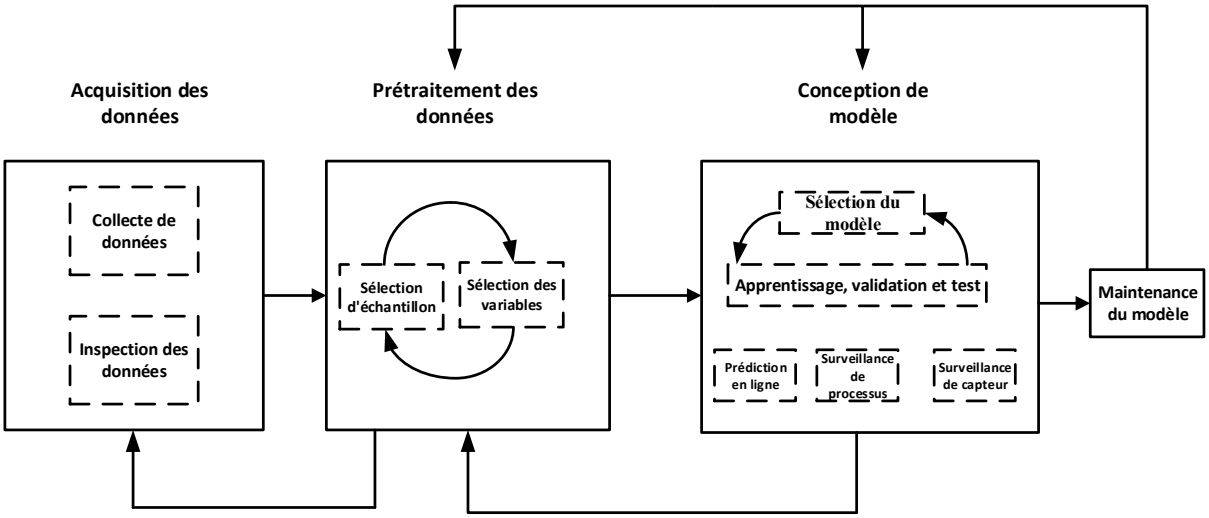

Fig.4.3. Etapes de développement d'un capteur logiciel.

3.1. Description des données d'entrée

Nous cherchons à construire le capteur logiciel et décider sur la qualité de l'eau à travers ses paramètres descripteurs. Nous n'avons en fait aucune connaissance a priori sur un type de modèle représentant parfaitement ce procédé, par contre nous pouvons porter notre jugement sur la qualité de cette eau à partir de quelques données descriptives.

L'objectif qui se trouve derrière la collecte des données relatives à ces paramètres est de trouver un modèle de régression permettant de distinguer le capteur logiciel, et un modèle de classification par la suite après sélection de la technique adoptée. La qualité de cette eau reflétée par sa potabilité repose en fait sur une corrélation qui ne peut être identifiée que statistiquement. Des données descriptives expérimentales recueillies sur une longue période (plusieurs années) pourraient atteindre cet objectif. Dans ce travail, un total de 774 échantillons est obtenu à partir de 10 variables de données sur la qualité de l'eau. Des statistiques descriptives de ces paramètres physico-chimiques sont illustrées dans le tableau

4.1. La figure 4.4 représente graphiquement l'évolution temporelle de ces différents paramètres.

Tableau.4.1. Statistiques descriptives des paramètres descripteurs utilisés.

Entrées (Variables)	Minimum	Maximum	Moyenne	Ecarts-types
T	0,10	27,20	11,91	6,25
C	11,00	329,00	146,59	74,99
pH	6,80	9,10	7,65	0,33
OD	7,00	15,00	10,53	1,60
MES	0,50	626,00	22,28	48,82
DBO	0,30	7,80	3,27	1,54
Ca	2,00	72,00	17,32	9,68
Cl	0,00	40,00	11,22	6,17
Mg	0,70	10,80	5,11	2,21
B	8,00	146,00	69,04	34,65

Les variables d'entrée T, C, pH, OD, MES, DBO, Ca, Cl, Mg et B représentent la température, la conductivité, le logarithme négatif de la concentration d'ions hydrogène dans l'eau, oxygène dissous, matières en suspension, demande biochimique en oxygène, calcium, chlore, magnésium et bicarbonate, respectivement. Ces entrées sont sélectionnées pour l'analyse en raison de leur continuité de mesure dans l'échelle de temps.

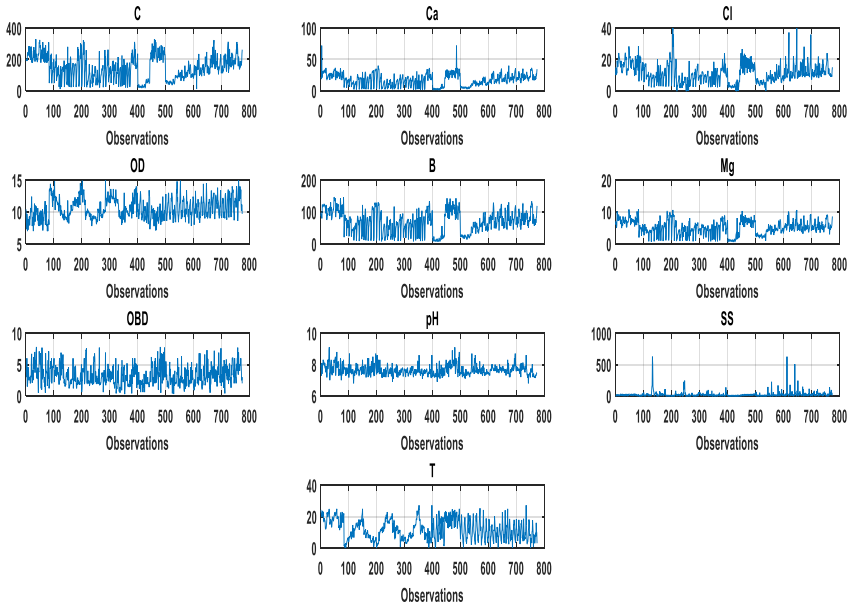

Fig. 4.4. Evolution temporelle des paramètres descripteurs.

3.2. Prétraitement des données

Les caractéristiques remarquables des données acquises dans les installations de traitement des eaux sont redondantes et possiblement insignifiantes, sans parler de la présence d'autres perturbations qui altèrent les mesures. Très souvent, la quantité et la qualité des données ainsi que leur haute dimensionnalité peuvent être un facteur limitant pour le développement des capteurs logiciels. Par conséquent, il est nécessaire de préparer les données avant qu'elles ne soient traitées par le modèle du capteur logiciel. La compréhension du processus et la connaissance a priori sont requises dans cette phase ; les opérateurs de l'installation et les experts peuvent fournir des informations précieuses sur la pertinence des variables et des observations. Une telle connaissance peut être soutenue et complétée par de nombreuses techniques statistiques pour la sélection de variables et la sélection d'échantillons. En ce qui concerne la surveillance de la qualité de l'eau, il existe un grand nombre d'échantillons correspondant à plusieurs capteurs installés, ce qui peut prendre beaucoup de temps à la phase du traitement. Par conséquent, l'utilisation de l'ACP est très utile pour réduire le nombre de variables et donc le nombre de données.

3.2.1. Normalisation des données

La normalisation des données permet de s'affranchir des différences de « normes » des variables. En effet, des variables avec des grandes valeurs peuvent avoir une plus grande influence que des variables avec des petites valeurs, sans pour autant être plus significatives. La normalisation est un outil de prétraitement de données utilisé dans le système d'exploration de données [204]. Un attribut d'un ensemble de données est normalisé en mettant ses valeurs à l'échelle afin qu'elles se situent dans une plage spécifiée de petite taille, telle que 0 à 1. La normalisation est particulièrement utile pour les algorithmes de classification et de régression.

Dans le but d'optimiser les modèles de prédiction, nous introduisons la normalisation des données par la méthode min-max. Cette technique permet de fournir une transformation linéaire sur la plage de données originale. Le rééchelonnement est souvent accompli en utilisant une formule d'interprétation linéaire, tels que :

$$x_{i_N} = \frac{x_i - min(X)}{max(X) - min(X)} \quad (4.1)$$

avec :

x_{i_N} :la valeur normalisée ;

$X = [x_1, x_2, \dots, x_i]$: le vecteur à normalisé ;

$min(X)$:la valeur minimale de X ;

$max(X)$:la valeur maximale de X.

Lorsque la normalisation min-max est appliquée, chaque entité se trouvant dans la nouvelle plage de valeurs restera la même. La normalisation min-max a l'avantage de préserver exactement toutes les relations dans les données.

3.2.2. Sélection des variables d'entrées

La technique ACP est utilisée pour sélectionner les variables qui ont le plus d'influence sur le capteur logiciel en utilisant la matrice de corrélation, ses valeurs propres et leur histogramme. En outre, nous pouvons comprendre qu'il y a un changement de caractéristiques de données à des composants qui ne sont pas corrélés. A souligner toutefois, que l'ensemble des 10 variables d'entrée de cette base de données est retenu en raison de l'importance de ces paramètres pour la qualité de l'eau et la continuité de leurs mesures dans le temps. L'Analyse par ACP appliquée à l'ensemble des données de la base, fourni le tableau 4.2 et l'histogramme de la figure 4.5.

Tableau.4.2. Statistiques des composantes principales.

	Fact. 1	**Fact. 2**	**Fact. 3**	**Fact. 4**	**Fact. 5**	**Fact. 6**	**Fact. 7**	**Fact. 8**	**Fact. 9**	**Fact. 10**
					Valeurs propres					
	5,04	1,85	1,16	0,75	0,670	0,187	0,134	0,103	0,078	0,024
					Pouercentage de la variance totale					
	50,42	18,54	11,55	7,54	6,702	1,867	1,335	1,033	0,779	0,236
					Pourcentage de la variance cumulative totale					
	50,42	68,96	80,51	**88,05**	94,75	96,62	97,95	98,99	99,76	100
					Les valeurs propres obtenues par l'application de l'ACP					
T°	0,21	-0,92	-0,05	0,12	0,18	-0,14	0,17	0,10	-0,06	-0,01
C	**0,95**	0,03	-0,05	-0,13	0,09	-0,15	0,05	-0,08	0,20	-0,04
pH	0,56	-0,23	-0,04	**0,66**	-0,43	0,04	-0,04	-0,04	0,03	-0,00
DO	-0,03	**0,93**	0,13	0,24	-0,05	-0,18	0,14	0,11	-0,03	0,00
MES	0,08	-0,20	**0,86**	-0,27	-0,38	0,02	0,06	0,01	0,01	-0,00
DBO	0,49	0,05	0,61	0,32	0,53	0,05	-0,09	-0,01	-0,00	0,00
Ca	0,94	0,07	-0,01	-0,13	-0,07	-0,18	-0,09	-0,13	-0,17	-0,02
Cl	0,91	0,21	-0,10	-0,03	0,07	0,25	0,22	-0,07	-0,06	-0,01
Mg	0,94	0,07	-0,12	-0,14	-0,08	0,09	-0,12	0,22	-0,01	-0,06
B	0,98	-0,01	-0,07	-0,12	-0,03	-0,02	-0,02	0,05	0,03	0,13

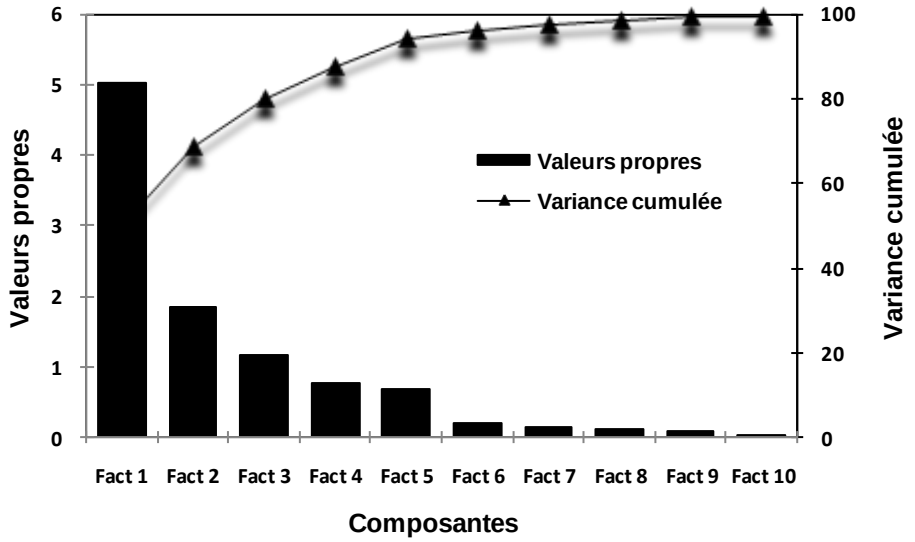

Fig. 4.5. Histogramme des valeurs propres des composantes.

Une diminution rapide des valeurs propres est d'abord constatée sur cette figure 4.5. A noter que les quatre premières composantes du tableau 4.2 représentent 88,05% de la proportion totale de la variance. En fin de compte, l'ensemble des dix variables est simplifié et quatre variables représentant les quatre premiers axes principaux :

- l'axe un, qui représente 50,42% de l'inertie totale, est défini positivement par cinq variables étroitement groupées C, Ca, Cl, Mg et B ;
- l'axe deux (18,54%) représente OD ;
- l'axe trois (11,55%) représente les deux variables MES et DBO ;
- l'axe quatre (7,54%) représente le pH variable.

Ceux-ci sont présentés dans le tableau 4.3. Dans celui-ci, nous rapportons la corrélation entre les variables pour les quatre premiers composants principaux. Nous pouvons observer qu'il existe une forte corrélation entre les variables de deux groupes différents : un groupe de cinq variables (C, Ca, Cl, Mg et B), et un groupe de deux variables (MES et DBO). Cependant, OD et pH ne sont pas corrélés avec d'autres variables.

Enfin, nous pouvons prédire la concentration de Cl en fonction des caractéristiques de l'eau qui sont plus facilement mesurées en continu et sans aucun résultat de laboratoire. Par conséquent, nous ne pouvons conserver que les variables : C, OD, MES et le pH.

Tableau.4.3. Valeurs propres, proportions de variance et corrélations entre variables des quatre premières composantes principales.

Variables	Composantes principales			
	Fact. 1	**Fact. 2**	**Fact. 3**	**Fact. 4**
T°	0,21	-0,92	-0,05	0,12
C	**0,95**	0,03	-0,05	-0,13
pH	0,56	-0,23	-0,04	**0,66**
OD	-0,03	**0,93**	0,13	0,24
MES	0,08	-0,20	**0,86**	-0,27
DBO	0,49	0,05	**0,61**	0,32
Ca	**0,94**	0,07	-0,01	-0,13
Cl	**0,91**	0,21	-0,10	-0,03
Mg	**0,94**	0,07	-0,12	-0,14
B	**0,98**	-0,01	-0,07	-0,2
Valeurs propres	5,042	1,85	1,155	0,75
Proportion de variance (%)	50,42	18,54	11,55	7,54
Proportion de variance cumulative (%)	50,42	68,96	80,51	**88,05**

3.3. Choix du modèle, apprentissage et test

Après la sélection des variables en utilisant la technique PCA, les paramètres C, pH, OD et MES sont considérés comme des variables d'entrée pour le capteur logiciel. Cependant, la sortie de celui-ci est prise comme la concentration de Cl. Afin de comparer l'efficacité des modèles : MLP, RBF, SVM et ELM en termes de temps d'entraînement et d'autres paramètres, l'ensemble de données a été subdivisé en quatre bases, comme illustrées dans le tableau 4.4 (trois bases pour la validation et la quatrième pour l'application).

Tableau.4.4.Les bases de données utilisées.

	Nombre d'échantillons	**Nombre d'échantillons (base d'entraînement)**	**Nombre d'échantillons (base de test)**
Base de données 1	400	300	100
Base de données2	500	375	125
Base de données 3	600	450	150
Base de données4	174	-	174

3.3.1. Modèle Neuronal-MLP

Dans un premier temps on commence par la conception du modèle neuronal de type MLP pour être appliqué dans le développement du capteur logiciel. Plusieurs réseaux sont testés en utilisant l'algorithme d'entraînement le plus connu celui de Levenberg-Marquardt qui est d'ailleurs préféré pour ses biens meilleurs résultats obtenus.

L'apprentissage supervisé du réseau neuronal-MLP, consiste à déterminer les poids de celui-ci qui minimisent sur l'ensemble des données de la base d'entraînement, les écarts entre les valeurs de sortie désirée y_d, et les valeurs de sortie calculée y_c. Ceci consiste à trouver le minimum du critère quadratique suivant :

$$C = \frac{1}{N}\sum_{i=1}^{N}(y_{c_i} - y_{d_i})^2 \qquad (4.2)$$

où Nest le nombre d'exemples de la base d'entraînement.

Les résultats d'entraînement utilisant les trois bases de données sont présentés dans le tableau 4.5. Les paramètres tels que le nombre de couches cachées (NCC), le nombre de neurones dans la couche caché (NNCC), le coefficient de corrélation (R^2) et l'erreur quadratique moyenne (EQM) sont indiqués pour différents types de réseaux testés.

Tableau.4.5. Résultats d'entraînement et de test pour différentes architectures du modèle neuronal-MLP.

Bases de données			**Phase d'entraînement**		**Phase de test**	
	NCC	**NNCC**	**EQM**	R^2	**EQM**	R^2
Base de données 1	01	10	0.1	0.54	0.2	0.50
	02	(12, 20)	0.02	0.65	0.1	0.60
	03	(4, 20, 25)	0.001	0.78	0,01	0.75
Base de données 2	01	10	0.08	0.60	0.1	0.53
	02	(12, 20)	0.015	0.68	0.03	0.63
	03	(4, 20, 25)	0.001	0.78	0.01	0.77
Base de données 3	01	10	0.01	0.66	0.04	0.65
	02	(12, 20)	0.003	0.75	0.009	0.70
	03	(4, 20, 25)	0.0008	0,87	0.001	0.80

On remarque une nette diminution de l'erreur d'entraînement quand il y a association des couches cachées avec augmentation du nombre d'échantillons de la base de données.

3.3.2. Modèle Neuronal-RBF

Différentes architectures sont testées pour déterminer le nombre adéquat de neurones dans la couche cachée du modèle Neuronal-RBF. Le tableau 4.6 montre pour plusieurs architectures de réseaux testés, les résultats correspondants aux différents paramètres d'entraînement. La structure du réseau composé de 400 neurones cachés est la mieux placée pour les différentes bases de données. Cette architecture est caractérisée par une erreur d'entraînement située entre 0.0001 et 0.0008 et un coefficient de corrélation supérieur à 0.97.

Tableau.4.6. Résultats d'entraînement et de test pour différentes architectures du modèle neuronal-RBF.

Bases de données		**La phase d'apprentissage**		**La phase de test**	
	NNCC	**EQM**	R^2	**EQM**	R^2
Base de données 1	10	0.006	0.86	0.007	0.84
	20	0.005	0.87	0.007	0.86
	25	0.005	0.88	0.006	0.86
	50	0.004	0.89	0.006	0.86
	100	0.002	0.93	0.005	0.87
	200	0.001	0.97	0.005	0.87
	400	0.0008	0.98	0.003	0.90
Base de données 2	10	0.0055	0.87	0.0024	0.89
	20	0.0052	0.88	0.0023	0.89
	25	0.005	0.89	0.0023	0.89
	50	0.0043	0.90	0.0021	0.90
	100	0.0031	0.93	0.0018	0.90
	200	0.0017	0.96	0.0014	0.92
	400	0.0001	0.98	0.001	0.94

Base de données 3	10	0.005	0.89	0.007	0.86
	20	0.004	0.89	0.007	0.87
	25	0.004	0.89	0.006	0.87
	50	0.004	0.91	0.006	0.87
	100	0.003	0.93	0.005	0.90
	200	0.001	0.96	0.003	0.93
	400	0.0001	0.98	0.002	0.95

3.3.3. Modèle SVM

Les résultats d'entraînement et de test obtenus pour le modèle SVM et utilisant les trois bases de données que précédemment, sont présentés dans le tableau 4.7. Les paramètres, tels que : R^2 et EQM pour les deux phases sont indiqués pour différentes valeurs de marge de tolérance(ε), avec pour fonction noyau RBF. Une fois de plus, cette fonction noyau est bien convenable pour ce type d'application, en raison de sa meilleure conformité pour la relation non-linéaire E/Set parce qu'il y a peu de paramètres [205].

Tableau.4.7. Analyse de sensibilité de SVM pour les trois bases de données.

Bases de données	**Epsilon** (ε)	**Phase d'entraînement**		**Phase de test**	
		EQM	**R²**	**EQM**	**R²**
Base de données 1	0,0001	$7x10^{-5}$	0,999	4.25	0,519
	0,001	$7x10^{-5}$	0,999	4,45	0,519
	0,01	$2x10^{-4}$	0,999	4,70	0,519
	0,1	0,01	0,999	4,90	0,509
	0,2	0,039	0,999	5,20	0,509
	0,3	0,086	0,998	5,38	0,508
	0,4	0,148	0,997	5,49	0,507

	0,5	0,228	0,995	6,20	0,505
	0,8	0,558	0,98	6,35	0,505
	1,3	1,347	0,97	6,75	0,504
Base de données 2	0,0001	$5x10^{-5}$	0,999	0,703	0,959
	0,001	$5x10^{-5}$	0,999	0,703	0,959
	0,01	$2x10^{-4}$	0,999	0,704	0,959
	0,1	0,009	0,998	0,721	0,958
	0,2	0,038	0,998	0,756	0,956
	0,3	0,085	0,998	0,807	0,954
	0,4	0,149	0,997	0,870	0,951
	0,5	0,226	0,995	0,940	0,947
	0,8	0,546	0,989	1,235	0,932
	1,3	1,345	0,974	1,957	0,904
Base de données 3	0,0001	$4x10^{-5}$	0,999	10^{-7}	0,999
	0,001	$5x10^{-5}$	0,999	$1,3x10^{-6}$	0,999
	0,01	10^{-4}	0,999	10^{-4}	0,999
	0,1	0,0099	0,999	0,01	0,999
	0,2	0,0387	0,999	0,0393	0,999
	0,3	0,085	0,998	0,087	0,998
	0,4	0,147	0,997	0,152	0,997
	0,5	0,224	0,995	0,23	0,995
	0,8	0,54	0,989	0,566	0,988
	1,3	1,324	0,975	1,427	0,972

Comme on peut le remarquer dans le tableau 4.7, la valeur de ε = 0,0001 donne de meilleurs résultats. Dans la suite, cette valeur de ε, sera prise en considération dans l'application pour l'usage de la technique SVM.

3.3.4. Modèle ELM

Vu ses performances, la technique ELM peut être adoptée pour établir le modèle du capteur logiciel dans le domaine de surveillance des eaux propres. La procédure d'élaboration de ce modèle est décrite comme suit.

Etape 1. Recueillir d'abord des données du système de surveillance pour créer des ensembles d'échantillons pouvant être utilisés pour l'entraînement et le test. Les données doivent être normalisées avant l'établissement du modèle.

Etape 2. Définir le NNC et le seuil d'erreur d'apprentissage.

Etape 3. Établir le modèle ELM. Pour former les SLFN sur la base des échantillons d'entraînement, l'inverse généralisé de Moore-Penrose est utilisé pour calculer les poids de sortie β.

Etape 4. Vérifier si la précision d'entraînement obtenue à l'étape 3 correspond au seuil d'erreur d'entraînement. Si la précision de l'entraînement atteint l'objectif, la procédure passe à l'étape suivante, sinon répétez les étapes 2 à 4.

Etape 5. Effectuer l'évaluation des performances du modèle ELM pour le capteur logiciel. Testez les performances de prédiction du modèle ELM établi à l'étape 4 sur la base de tests préparés à l'étape 1.

L'organigramme d'élaboration du modèle ELM peut donc être représenté comme indiqué à la figure 4.6.

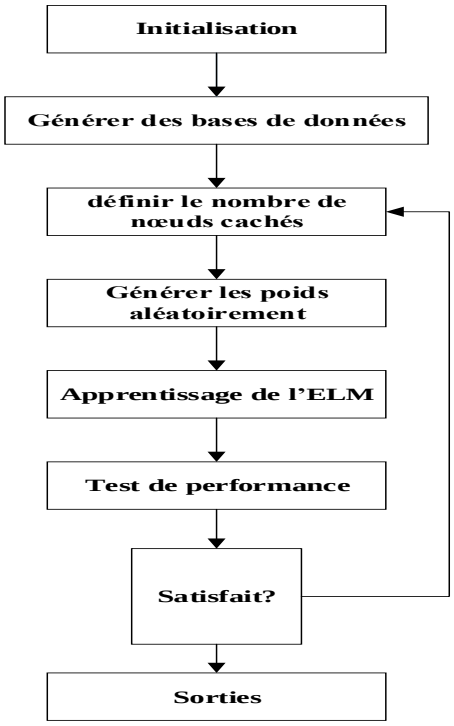

Fig. 4.6. Organigramme d'élaboration du modèle ELM pour le capteur logiciel.

Comme les poids sont initialisés de manière aléatoire, différents résultats peuvent être obtenus par des analyses séparées. Par conséquent, sept (07) architectures ont été proposées pour la formation du réseau ELM.

Tableau.4.8. Résultats d'entraînement et de test pour différentes architectures du modèle ELM.

Bases de données		**Phase d'entraînement**		**Phase de test**	
	NNCC	**EQM**	R^2	**EQM**	R^2
Base de données 1	10	0.0063	0.85	0.006	0.85
	20	0.0059	0.87	0.006	0.80
	25	0.0055	0.87	0.006	0.82
	50	0,005	0.89	0.005	0.82
	100	0,0037	0.92	0.0027	0.83
	200	0,0014	0.97	0.0023	0.85
	400	$4x10^{-8}$	0.99	$7x10^{-4}$	0.86

Base de données 2	10	0.0059	0.85	0.004	0.90
	20	0.0054	0.87	0.004	0.90
	25	0.0052	0.87	0.004	0.90
	50	0.0046	0.89	0.003	0.91
	100	0.0036	0.92	0.0027	0.94
	200	0.0019	0.97	0.0023	0.95
	400	$3x10^{-8}$	0.99	$2x10^{-4}$	0.99
Base de données 3	10	0.0056	0.88	0.0066	0.57
	20	0.0053	0.88	0.006	0.69
	25	0.0051	0.89	0.0055	0.75
	50	0.0048	0.90	0.0052	0.90
	100	0.0037	0.92	0.0044	0.92
	200	0.0021	0.96	0.0025	0.95
	400	10^{-5}	0.99	$4.8x10^{-6}$	0.99

3.3.5. Evaluation des performances

Pour une évaluation des performances des quatre modèles précédemment testés, définissons quelques mesures statistiques telles que la valeur moyenne($\bar{x}$), l'écart type (σ), la valeur minimale (x_{min}), la valeur maximale (x_{max}) et le coefficient de corrélation R^2qui sont calculés et présentés dans le tableau 4.9 dans le but de fournir une comparaison entre les trois bases de données pour le capteur logiciel Cl.

Tableau.4.9. Paramètres statistiques des différents modèles élaborant le capteur logiciel Cl.

Bases de données		**Phase d'entraînement**					**Phase de test**				
		$\bar{x}$	σ	x_{min}	x_{max}	R^2	$\bar{x}$	σ	x_{min}	x_{max}	R^2
Base de données 1	**Données réelles**	10,67	6,26	0	40		12,1	6,08	2,2	28,3	
	MLP	10.34	6.11	0	38.9	0.78	11.2	6.54	0.9	27	0.75
	RBF	10.70	6.15	0	39	0.98	11.34	6.48	0.75	28	0.90
	SVM	10.67	6.26	0	40	0.99	11.92	5.91	2.2	28	0.51
	ELM	10.69	6.26	0	40	0.99	11.42	6.09	0	26	0.86
Base de données 2	**Données réelles**	11,15	6,23	0	40		9,04	5,82	0	40	
	MLP	11.10	6.12	0	38.9	0.79	7.6	5.23	0	34	0.77
	RBF	11.15	6.04	0	39	0.97	6.15	4.08	0	24	0.94
	SVM	11.15	6.23	0	40	0.99	9.24	5.94	0	40	0.96
	ELM	11.15	6.22	0	40	0.99	8.24	5	0	39	0.99
Base de données 3	**Données réelles**	10,23	6,35	0	40		10,76	6,19	0	40	
	MLP	10.15	6.20	0	38.9	0.87	10.33	5.42	0	39	0.80
	RBF	10.24	6.16	0	35.8	0.97	10.44	5.60	1.51	32.59	0.95
	SVM	10.23	6.35	0	40	0.99	10.76	6.19	0	40	0.99
	ELM	10.24	6.34	0	40	0.99	10.76	6.19	0	40	0.99

D'autres paramètres, indicateurs de performance, obtenus des différents modèles pour chaque base de données sont présentés dans le tableau 4.10. Il s'agit de l'erreur absolue moyenne(EAM), l'EQM, la racine de l'erreur quadratique moyenne(REQM), le temps d'entraînement ou d'apprentissage(t_{appr})et le temps d'exécution (t_{ex}).

Tableau.4.10. Paramètres de performance des différents modèles élaborant le capteur logiciel Cl.

Bases de données		**Phase d'entraînement**				**Phase de test**			
		EAM	**EQM**	**REQM**	$\mathbf{t_{appr}(s)}$	**EAM**	**EQM**	**REQM**	$\mathbf{t_{ex}(s)}$
Base de données 1	**MLP**	0.1	0.0015	0.039	15.4	0,2	0.03	0.173	0.12
	RBF	0.018	$8x10^{-4}$	0.028	13.28	0.04	0.005	0.072	0.04
	SVM	$9x10^{-4}$	$7x10^{-5}$	$8x10^{-3}$	8.53	1.19	4.25	2.06	0.031
	ELM	$1.7x10^{-5}$	$4x10^{-8}$	$2x10^{-3}$	1.89	0.011	$7x10^{-4}$	0.027	0.36
Base de données 2	**MLP**	0.09	0.001	0.032	24.78	0.1	0.01	0.1	0.18
	RBF	0.023	$1.4x10^{-3}$	0.037	21.19	0.022	0.001	0.033	0.06
	SVM	$8x10^{-4}$	$5x10^{-5}$	$7x10^{-3}$	12.08	0.179	0.703	0.839	0.047
	ELM	$1.8x10^{-5}$	$3x10^{-8}$	$2x10^{-4}$	2,51	10^{-3}	$2x10^{-4}$	0.01	0.28
Base de données 3	**MLP**	0.001	0.0008	0.028	26.9	0,04	0.001	0.032	0,22
	RBF	0.023	0.001	0.037	25.7	0.03	0.0021	0.046	0.08
	SVM	$7x10^{-4}$	$4x10^{-5}$	0.007	16.45	$2x10^{-4}$	10^{-7}	$3x10^{-4}$	0.06
	ELM	$5x10^{-4}$	10^{-5}	0.003	3.73	$5x10^{-4}$	$4.8x10^{-6}$	$2x10^{-3}$	0.3

À partir des résultats obtenus des tableaux 4.9 et 4.10, on peut noter ce qui suit.

- La valeur calculée R^2 indique un degré de corrélation élevé entre les données réelles et les données prédites en utilisant les techniquesRBF, SVM et ELM. Comme on le voit dans le tableau 4.9, le coefficient R^2 est compris entre 0,95 et 0,99 pour la base de données 2 et la base de données 3.
- Généralement, la valeur de la REQM obtenue avec ELM a la valeur la plus faible par rapport à la valeur obtenue avec les autres techniques.
- ELM assure un temps d'entraînement minimum par rapport aux autres techniques et ce pour les trois bases de données utilisées. En outre, comme indiqué dans la figure 4.7, la différence des temps d'entraînement peut augmenter lors de l'utilisation d'une base de données plus grande.

On sait par ailleurs, que la base de données élaborée à partir de mesures, au niveau d'une station de traitement des eaux, contient un très grand nombre d'échantillons provenant d'un seul capteur ; ce qui nous laisse imaginer la dimension de cette base lorsqu'il y a plusieurs capteurs utilisés au niveau du même système de surveillance.

On peut dire de ce fait que le modèle ELM est bien adapté à l'apprentissage de grandes bases de données.

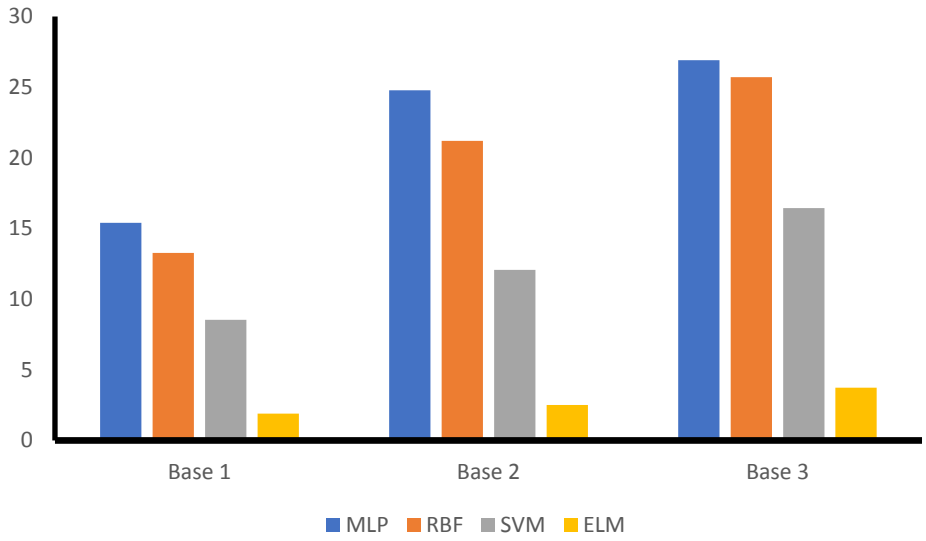

Fig4.7. Temps d'entraînement des différents modèles élaborant le capteur logiciel de Cl.

Dans la figure 4.8, on montre l'évolution temporelle de la concentration Cl pour les trois bases de données, où sont représentées les valeurs mesurée et estimée de cette concentration pour les modèles MLP, RBF, SVM et ELM.

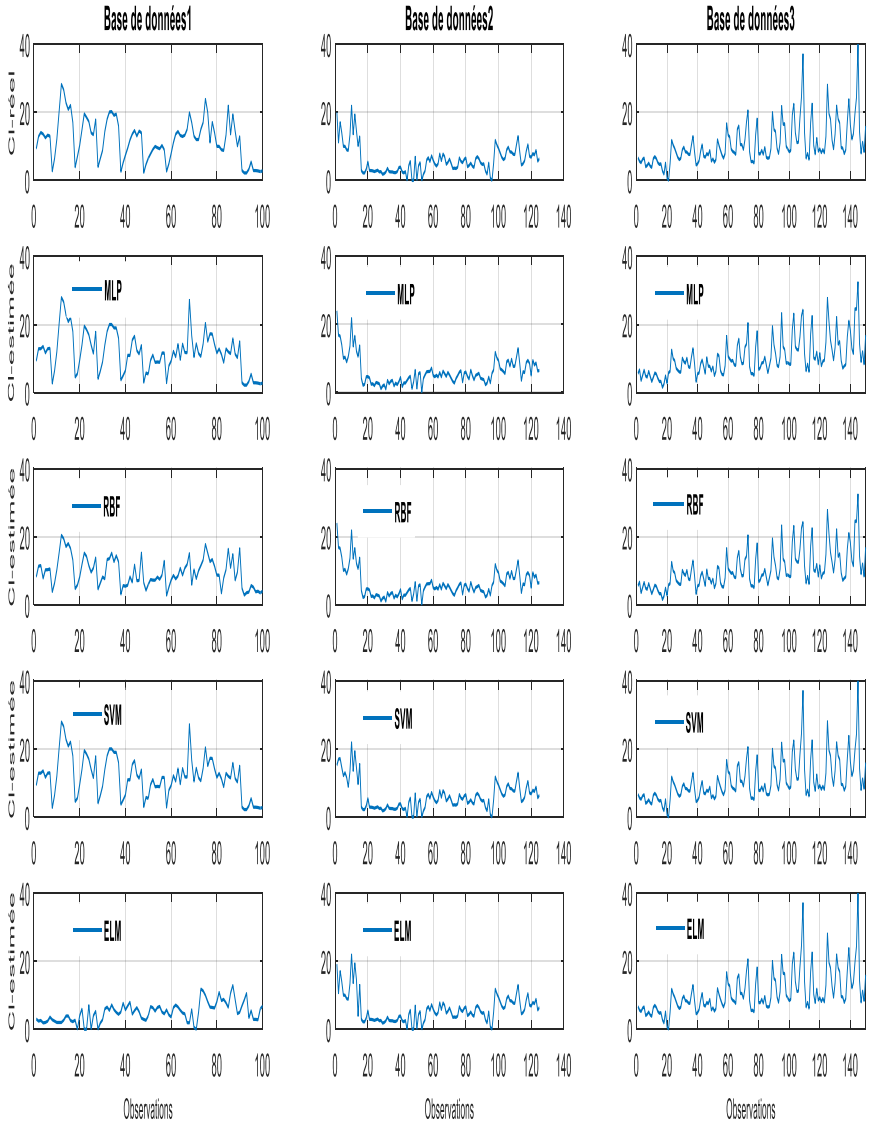

Fig.4.8. Données réelles et estimées de la concentration Cl pour les modèles MLP, RBF, SVM et ELM.

4. APPLICATION

Pour valider les performances de la technique ELM, une application réelle est réalisée pour montrer l'importance et l'intérêt du capteur logiciel élaboré dans contrôle de la qualité de l'eau (eau potable et eau non-potable) [206-208]. Les modèles MLP, RBF, SVM et ELM sont utilisés dans ce cadre avec dix paramètres physico-chimiques comme entrées. Pour cela, différentes bases de données sont testées pour comparer ces quatre modèles. Le tableau 4.11 représente les résultats obtenus des trois bases de données appliquées au taux de classification, au temps d'entraînement et au temps d'exécution.

Tableau.4.11. Performances des modèles MLP, RBF, SVM et ELM.

Bases de données		**Phase d'entraînement**	**Phase de test**	
		Temps d'entraînement (s)	**Taux de classification (%)**	**Temps d'exécution(s)**
Base de données 1	**MLP**	15.43	68	0.1
	RBF	14.83	70	0.03
	SVM	7,5	90	0,1
	ELM	2,75	82	0,1
Base de données 2	**MLP**	19.21	75	0.15
	RBF	17.67	99	0.06
	SVM	13,8	95	0,1
	ELM	3,79	92,8	0,15
Base de données 3	**MLP**	24.32	80	0.2
	RBF	21.92	99.3	0.08
	SVM	18,3	99,3	0,2
	ELM	4,82	99,3	0,25

Les taux de classification sont importants pour les modèles SVM et ELM. Cependant, le temps d'entraînement avec ELM est presque 3 fois inférieur à celui de SVM, et 5 fois à celui de MLP et RBF ; autrement dit le temps d'entraînement avec ELM est beaucoup plus rapide qu'avec les autres modèles. Cependant, les temps d'exécution sont presque similaires pour les quatre modèles. La figure 4.9 montre l'histogramme des temps d'entraînement correspondant aux trois bases de données pour les quatre modèles.

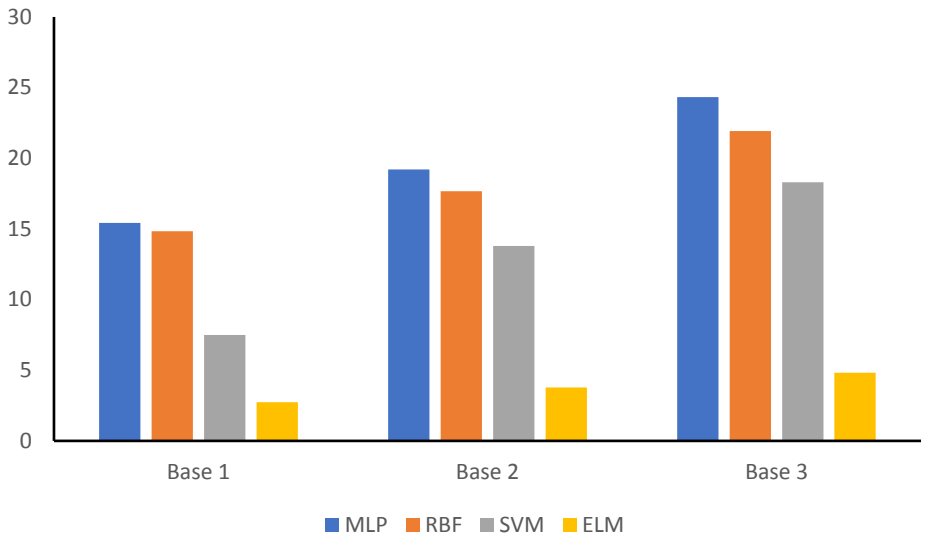

Fig.4.9. Temps d'entraînement correspondant aux modèles MLP, RBF, SVM et ELM.

Afin de montrer l'importance et l'intérêt du capteur logiciel ainsi élaboré, une application de surveillance de la qualité de l'eau basée sur un classifieur ELM est présentée dans la figure 4.10. Dans cet exemple, une simulation à base de ce classifieur et utilisant une quatrième base de données de 174 échantillons, est effectuée. Trois scénarios sont proposés à cet effet : le premier utilise dix capteurs, le second neuf avec un capteur défaillant, et le troisième, neuf capteurs avec un capteur logiciel Cl à base du modèle ELM.

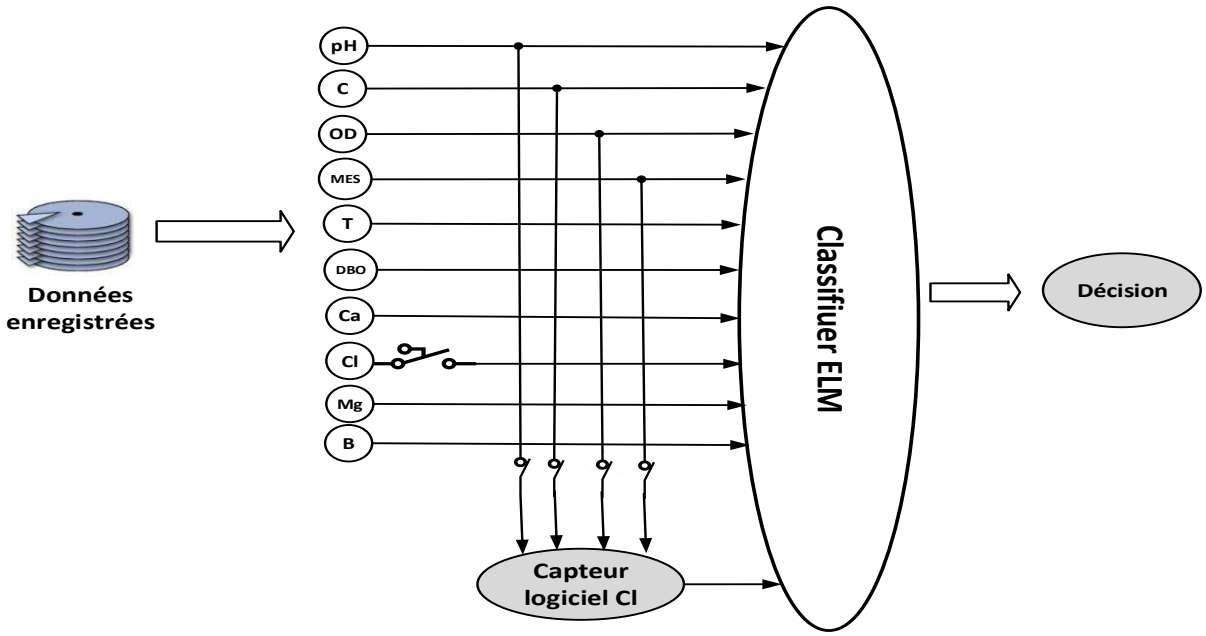

Fig. 4.10. Classifieur et capteur logiciel Cl à base du modèle ELM.

Les résultats obtenus, montrant le taux de classification quant à la décision de potabilité pour ces trois cas, sont montrés dans la figure 4.11 [209]. On peut remarquer à partir de cette figure, une détérioration importante du taux de classification dans le cas du deuxième scénario (cas de capteur en panne), alors que dans le troisième scénario, un taux de classification élevé est obtenu, soit 97,33% (cas de remplacement du capteur défaillant par le capteur logiciel Cl). En comparant les temps d'exécution, on remarque que dans le cas du troisième scénario, le temps est plus élevé que celui des scénarios 1 et 2, car c'est le temps de classification qui est rajouté au temps d'exécution du capteur logiciel ; alors que dans les deux scénarios (1 et 2), le temps d'exécution ne représente que celui de la classification (Fig.4.12).

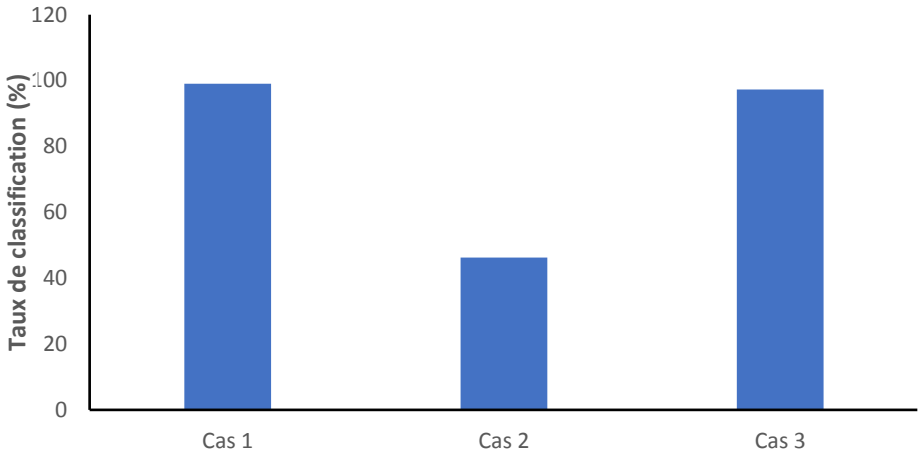

Fig. 4.11. Taux de classifications obtenus pour les 3 scénarios (cas).

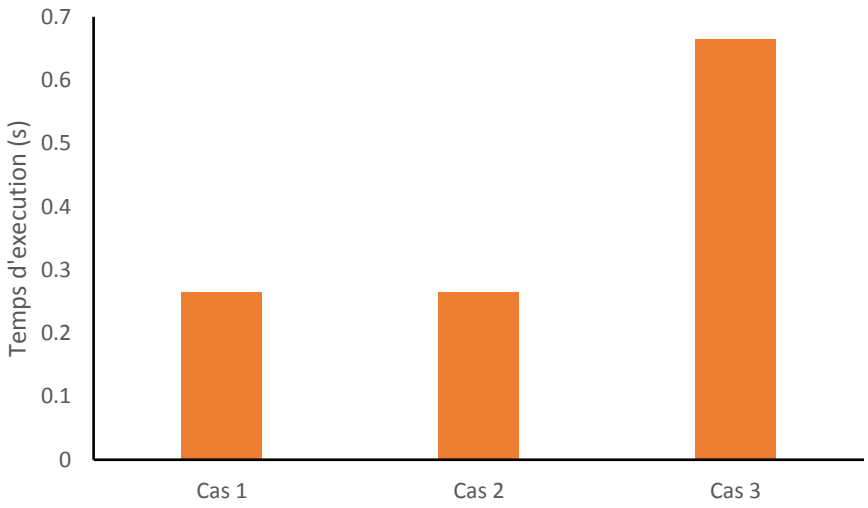

Fig. 4.12. Temps d'exécution obtenus pour les 3 scénarios (cas).

CONCLUSION

Ce dernier chapitre a fait l'objet d'une étude en simulation qui consistait à évaluer les performances de quatre techniques d'apprentissage statistique appliquées au domaine du contrôle et de surveillance des eaux propres. Cette étude a permis la comparaison entre les modèles MLP, RBF, SVM et ELM, en vu de choisir la technique la mieux adaptée pour la régression et la classification. L'objectif a été de mettre en place un système multicapteur à base de capteurs logiciels permettant de prendre une décision adaptée au contrôle et au suivi des problèmes de la qualité de l'eau.

Les résultats obtenus ont montré que ces quatre modèles étaient similaires en terme de temps d'exécution ce qui leur confère l'avantage de s'intégrer dans un système de surveillance dynamique. Cependant, les modèles MLP, RBF et SVM ont montré particulièrement une invalidité majeure liée au temps d'entraînement. Un désavantage qui a été heureusement levé par le modèle ELM, choisi à cet effet pour sa rapidité en cette phase d'apprentissage, et de ses qualités présentées en termes de RMSE et de coefficient de corrélation. Bien adapté à l'apprentissage de grandes bases de données, ce modèle a présenté en fait de nombreux avantages, que ce soit en termes de classification et / ou de régression. L'usage et l'application de celui-ci pourrait avoir un impact direct touchant aussi bien les aspects de conception, d'environnement et d'économie.

CONCLUSION GENERALE

Le travail présenté dans le cadre de cette thèse a fait l'objet d'une contribution à la conception de systèmes de surveillance et de contrôle de la qualité des eaux propres à base de capteurs logiciels. Equipés de capteurs physico-chimiques, tels que les capteurs de température, de pH, d'oxygène dissous, de conductivité etc., ces systèmes assurent de façon continue la surveillance et le contrôle en ligne de cette ressource. Le capteur logiciel a été délibérément introduit car certaines variables liées au contrôle de la qualité ne peuvent être mesurées automatiquement en ligne, soit à cause de pannes pouvant survenir au niveau des capteurs, au coût élevé lié au remplacement, ou encore à la difficulté même de la mesure, nécessitant d'éventuels déplacements sur site.

L'étude présentée découle des progrès technologiques importants qui ont été enregistrés ces dernières années, où l'intérêt d'une surveillance moderne et efficace de la qualité des eaux propres s'est vu imposé. Dans ce domaine, les capteurs logiciels sont devenus un moyen très attrayant pour faire face au manque de capteurs spécifiques dans les chaînes de contrôle et de surveillance. Notre travail peut être considéré comme une contribution aux solutions proposées, pour résoudre des problèmes d'intérêt stratégique à préoccupation nationale et internationale, utilisant des outils modernes à base de techniques avancées.

Le travail présenté a été structuré autour de quatre chapitres essentiels. Le premier a été consacré à des généralités concernant les eaux propres, les normes de potabilité ainsi que les différentes étapes de traitement et méthodes de contrôle employées. Dans le deuxième chapitre, il a été rappelé la notion de capteur en tant qu'élément fondamental dans une chaîne de mesure. Les différents capteurs physico-chimiques utilisés dans le domaine ont été particulièrement décrits. Les notions de capteurs intelligent et logiciel ont été à cet effet introduites. Dans le troisième chapitre, il a été question des capteurs logiciels en particulier, comme étant un moyen intéressant pouvant contourner l'usage de certains capteurs physico-chimiques. Quatre modèles (Neuronal, Neuronal-RBF, SVM et ELM) à base de la conception de ces capteurs ont été étudiés et évalués. Enfin, le quatrième et dernier chapitre a été dédié à une application réelle justifiant la conception et la mise en œuvre d'un capteur logiciel pour la mesure du Chlore utilisé dans le contrôle de potabilité de l'eau. La technique la mieux adaptée

pour la conception de ce capteur a été choisie sur la base d'une étude d'évaluation préliminaire des performances de chacune des méthodes présentées au chapitre trois. Plusieurs paramètres ont été évalués à cet effet, tels que le temps d'apprentissage, l'erreur d'entraînement etc.

Ce travail a donc permis la validation et l'évaluation des performances de chacune des techniques étudiées. L'étude comparative entreprise entre les modèles MLP, RBF, SVM et ELM dans le but d'un choix décisif de la méthode la mieux adaptée à la régression et à la classification a donc été effectuée. Les résultats obtenus ont montré d'abord que ces quatre modèles étaient similaires en termes de temps d'exécution, ce qui leur donne l'avantage de s'intégrer dans un système de surveillance en ligne. Cependant, les modèles MLP, RBF et SVM souffraient particulièrement d'un handicap majeur lié au temps d'entraînement. Ce désavantage a été plutôt levé par le modèle ELM, choisi en raison de sa vitesse en matière d'entraînement qui lui confère un rang privilégié et bien adapté pour un apprentissage rapide sur d'importantes bases de données. Pour valider davantage le modèle le plus approprié en régression comme en classification, une application réelle a été élaborée afin de montrer l'importance du capteur logiciel conçu (capteur Cl) en matière de décision. Pour cela, différentes bases de données ont été testées dans un but de comparaison entre les modèles proposés dans le cadre de cette étude. Les résultats obtenus à cet effet, ont montré que les taux de classification réalisés par les modèles SVM et ELM ont été importants. Cependant, le temps d'entraînement avec l'ELM a été presque 3 fois inférieur à celui de l'SVM et 5 fois inférieur à celui de MLP et RBF. Cela confirme encore une fois la rapidité d'apprentissage qui caractérise la technique ELM relativement aux autres. Néanmoins, les temps d'exécution sont presque similaires pour les quatre modèles étudiés. On peut dire alors que le modèle ELM dispose de nombreux avantages que se soit sur le plan de son adaptation pour les problèmes de classification binaire, des problèmes de régression, de sa mise œuvre du point de vu algorithmique, ou de son adaptation pour résoudre des problèmes de classification. L'usage d'un prétraitement des données par ACP dans un but de réduction de dimensionnalité, a confirmé davantage cet intérêt.

On peut dire en conclusion que les capteurs logiciels dérivés de données utilisant des techniques d'apprentissage statistiques, se sont révélés capables d'extraire des informations facilement compréhensibles et praticables que contiennent des données de grande dimension. L'utilisation de tels capteurs permet une utilisation plus intensive des données mesurées en

temps réel, et par conséquent, des opérations de traitement plus avancées qui satisfont les objectifs de surveillance des installations des eaux de manière rentable.

La possibilité d'utiliser à la fois des mesures physiques et des estimations de capteurs logiciels d'une manière complémentaire dans l'opération de traitement a été montrée au moyen du système de commutation proposé. Les estimations de capteurs logiciels servent généralement d'option de secours pour les instruments existants lorsqu'ils subissent des temps d'arrêt ou sont utilisées lorsque les capteurs physiques ne sont pas du tout utilisés pour mesurer les variables de processus. Cette proposition, en particulier, ouvre de nouvelles perspectives pour la communauté scientifique, non seulement en ce qui concerne le traitement des eaux, mais aussi pour d'autres secteurs industriels. L'application de la technique peut avoir un impact direct sur l'économie, l'environnement, et la conception et le développement de systèmes de surveillance en général.

De nombreuses perspectives restent encore envisageables telles que :

- La base de données réelle qui peut être importante et plus significative, contribuera sans doute à augmenter la précision de reconnaissance (classification et/ou régression).
- Les approches de prétraitement des données devraient être développées pour suivre les changements de mesurandes.
- Les méthodologies pour la sélection dynamique des variables qui influencent la sortie du modèle et leurs délais respectifs, devraient aussi être étudiées.
- Les techniques pour la mise à l'échelle des données en ligne, la détection des valeurs aberrantes en ligne et le dépassement en ligne des données manquantes devraient être proposées.
- L’application des capteurs logiciels peut également être étendue à d'autres domaines tels que la surveillance de la santé, les énergies renouvelables, etc.
- Prévoir l’implémentation pratique de capteurs logiciels au sein de plateformes de capteurs intelligents pour des systèmes multicapteurs étendus, plus souples et plus performants.

REFERENCES

[1] K.J. Howe, et al., "Principles of Water Treatment", John Wiley & Sons: Hoboken, p. 672, NJ, USA, 2012.

[2] United Nations World Water Assessment Programme (WWAP), "The United Nations World Water Development Report 2017, Wastewater: The Untapped Resource", United Nations World Water Assessment Programme, p.198, Paris, France, 2017.

[3] M.A. Shannon, et al., "Science and technology for water purification in the coming decades", Nature, Vol. 452, pp. 301-310, 2008.

[4] World Health Organization (WHO), United Nations Children's Fund (UNICEF), "Progress on Drinking Water, Sanitation and Hygiene: 2017 Update and SDG Baselines", WHO and UNICEF, p. 110, Geneva, Switzerland, 2017.

[5] H. Fujii, S. Managi, "Decomposition Analysis of Water Treatment Technology Patents", Water, Vol.09, p.860, 2017.

[6] United Nations Water (UN-Water), "Towards a Worldwide Assessment of Freshwater Quality", A UN-Water Analytical Brief, UN-Water: Geneva, Switzerland, 2016.

[7] W. Schon, et al., "Maîtrise des risques dans le domaine de l'eau potable", In Actes du 12ème Colloque National de Sûreté de Fonctionnement, Laboratoire SIME Système Intelligents pour la Maîtrise de l'Eau, pp. 695-701, France, 2000.

[8] M. Ladjal, "Contribution au développement de systèmes de surveillance innovants dédiés au contrôle de la qualité des eaux potables", Thèse de doctorat, Université de M'sila, Algérie, 2013.

[9] M. H. Banna, et al., "Miniaturized water quality monitoring pH and conductivity sensors", Sensors and Actuators B, Vol. 193, pp. 434 -441, 2014.

[10] S. Zhuiykov, "Solid-state sensors monitoring parameters of water quality for the next generation of wireless sensor networks", Sensors and Actuators B, Vol. 161, pp. 1-20, 2012.

[11] K. Murphy, B. Heery, T. Sullivan, "A low-cost autonomous optical sensor for water quality monitoring", Talanta, Vol. 132, pp. 520-527, 2015.

[12] J. Qiao, Z. Hu, W. Li, "Soft Measurement Modeling Based on Chaos Theory for Biochemical Oxygen Demand (BOD) ", Water, Vol.08, p. 581, 2016.

[13] L. Fortuna, et al., "Soft sensors for monitoring and control of industrial processes", Springer Verlag, London, U.K, 2007.

[14] P. Kadlec, B. Gabrys, S. Strandt, "Data-driven Soft Sensors in the process industry", Computers and Chemical Engineering, Vol. 33, pp. 795-814, 2009.

[15] A.K. Pani, V.K. Vadlamudi, H.K, Mohanta, "Development and comparison of neural network based soft sensors for online estimation of cement clinker quality", ISA Transactions, Vol. 52, pp. 19-29, 2013.

[16] J. Zheng, Z. Song, "Semi supervised learning for probabilistic partial least squares regression model and soft sensor application", Journal of Process Control, Vol. 64, pp. 123-131, 2018.

[17] D.V. Poerio, S.D. Brown, "Highly-overlapped, recursive partial least squares soft sensor with state partitioning via local variable selection", Chemometrics and Intelligent Laboratory Systems, Vol. 175, pp. 104-115, 2018.

[18] J. Ma, et al., "Novel Field-Support Vector Regression-Based Soft Sensor for Accurate Estimation of Solar Irradiance ", IEEE transactions on circuits and systems–I, Vol. 64, pp. 3183 - 3191, 2017.

[19] S. Sharma, S.S. Tambe, "Soft-sensor development for biochemical systems using genetic programming", Biochemical Engineering Journal, Vol. 85, pp. 89-100, 2014.

[20] P. Sagmeister, P. Wechselberger, M. Jazini, A. Meitz, "Soft sensor assisted dynamic bioprocess control: Efficient tools for bioprocess development", Chemical Engineering Science, Vol. 96, pp. 190-198, 2013.

[21] M. Huang, Y. Mab, J. Wan, "A sensor-software based on a genetic algorithm-based neural fuzzy system for modeling and simulating a wastewater treatment process", Applied Soft Computing, Vol. 27, pp. 1-10, 2015.

[22] B. Lamrini, et al., "A neural software sensor for on-line prediction of coagulant dosage in a drinking water treatment plant", Transactions of the Institute of Measurements and Control, Vol. 27, pp. 195-213, 2005.

[23] L. Wang, C. Shao, H. Wang, H. Wu, "Radial Basis Function Neural Networks-Based Modeling of the Membrane Separation Process: Hydrogen Recovery from Refinery Gases", Journal of Natural Gas Chemistry, Vol. 15, pp. 230-234, 2006.

[24] S. Chen, A.K. Samingan, L. Hanzo, "Support Vector Machine Multiuser Receiver for DS-CDMA Signals in Multipath Channels", IEEE Transactions On Neural Networks, Vol. 12, pp. 604-611, 2001.

[25] F. Melgani, Y. Bazi, "Classification of Electrocardiogram Signals With Support Vector Machines and Particle Swarm Optimization", IEEE Transactions on Information Technology in Biomedicine, Vol. 12, pp. 667-677, 2008.

[26] F. Kang, Q. Xu, J. Li, "Slope reliability analysis using surrogate models via new support vector machines with swarm intelligence", Applied Mathematical Modelling, Vol. 40, pp. 6105-6120, 2016.

[27] C. Shang, X. Gao, F. Yang, D. Huang, "Novel Bayesian Framework for Dynamic Soft Sensor Based on Support Vector Machine With Finite Impulse Response", IEEE Transaction on control systems technology, Vol. 22, pp. 1550-1557, 2014.

[28] J. Su, et al., "A Structurally Simplified Hybrid Model of Genetic Algorithm and Support Vector Machine for Prediction of Chlorophyll a in Reservoirs", Water, Vol.07, pp. 1610-1627, 2015.

[29] G.B. Huang, Q.Y. Zhu, C.K. Siew, "Extreme learning machine: Theory and applications", Neurocomputing, Vol. 70, pp. 489-501, 2006.

[30] W.W. Zong, G.B. Huang, Y.Q. Chen, "Weighted extreme learning machine for imbalance learning", Neurocomputing, Vol. 101, pp. 229-242, 2013.

[31] H.G. Han, L.D. Wang, J.F. Qiao, "Hierarchical extreme learning machine for feedforward neural network", Neurocomputing, Vol. 128, pp. 128-135, 2014.

[32] W. Wang, et al., "Soft Sensing of Dissolved Oxygen in Fishpond via Extreme Learning Machine", Proceeding of the 11th World Congress on Intelligent Control and Automation Shenyang, China, pp. 3393-3395, 2014.

[33] G.B. Huang, et al., "Extreme Learning Machine for Regression and Multiclass Classification", IEEE Transactions on systems, man, and cybernetics-part b: cybernetics, Vol. 42, pp. 513-529, 2012.

[34] F. Kang, et al., "Concrete dam deformation prediction model for health monitoring based on extreme learning machine", John Wiley & Sons, Ltd. Struct Control Health Monit, Vol. 24: e1997, 2017.

[35] M. Kahoul, M. Touhami, "Evaluation de la qualité physico-chimique des eaux de consommation de la ville d'Annaba (Algérie) ", Larhyss/Journal No 19, Septembre 2014.

[36] S. Bandyopadhyay, et al., "Instrumentation for continuous in situ monitoring of water quality", IEEE Transactions on Instrumentation and Measurement, Vol. 38, pp. 820-823, 1989.

[37] W. Yan, et al., "Comprehensive assessment and visualized monitoring of urban drinking water quality", Chemometrics and Intelligent Laboratory Systems, Vol. 155, pp. 26-35, 2016.

[38] J. Rodier, "L'Analyse de l'eau", 9e édition. Dunod, Paris, 2009.

[39] M. Salari, " Quality assessment and artificial neural networks modeling for characterization of chemical and physical parameters of potable water", Food and Chemical Toxicology, Vol. 118, pp.212-219, 2018.

[40] S. Harley, "L'analyse de la qualité de l'eau, Administration du rétablissement agricole", TRE-116-2002-03, Canada, Mars, 2000.

[41] E. Jestin, "La production et le traitement des eaux destinées à l'alimentation et à la préparation de denrées alimentaires", Agence de l'eau de Seine-Normandie, France. www.eau-seine-normandie.fr.

[42] N. Harrat, S. Achour, "Pollution physico-chimique des eaux de barrage de la région d'El-Taref : Impact sur la chloration", Larhyss Journal, No. 8, pp. 47-54, juin, 2010.

[43] "La matière organique dans le traitement de l'eau potable", Memotec N°25, 2006. ww.gls.fr.

[44] "Qualité physico-chimique et chimique des eaux de surface : Cadre général", Institut Bruxellois pour la gestion de l'environnement/ observatoire des données de l'environnement. Fiche 2, Nov. 2005. www.ibgebim.be.

[45] A. Lounnas, "Amélioration des procédés de clarification des eaux de la station Hamadi-Kroma de Skikda", Mémoire de Magister, Université de Skikda, Algérie, 2009.

[46] Degrémont, "Mémento technique de l'eau : Tome 1", Lavoisier - Technique et documentation, Paris, France, 1989.

[47] S. Merabet, "Evaluation de la qualité physico-chimique des eaux brutes et distribuées du barrage réservoir de Beni Haroun", Mémoire de Magister, Université de Constantine, Algérie, 2010.

[48] V. A. Gonzalez, "Estimation et commande robuste non linéaires des procédés biologiques de dépollution des eaux usées : Application à la digestion anaérobie", Thèse de Doctorat, Université de Perpignan, France, 2001.

[49] "Procédés reconnus destinés au traitement de l'eau potable", Office fédéral de la santé publique (OFSP), Suisse, 2010.

[50] A. Kettab, "Traitement des eaux : les eaux potables", O.P.U, Alger, Algérie, 1992.

[51] B. Meghzili, "Suivi de la qualité des eaux de la station de traitement Hamadi-Kroma (Skikda), Essais d'optimisation de la clarification des eaux", Mémoire de Magister, Université de Biskra, 2003.

[52] M. Khiati, "L'eau en Algérie : Une source de vie et un impératif de développement", fondation nationale pour la promotion de la santé et le développement de la recherche, Centre Culturel d'Hussein Dey, Alger, Algérie, 2002.

[53] N. Valentin, "Construction d'un capteur logiciel pour le contrôle automatique du procédé de coagulation en traitement d'eau potable", Thèse de doctorat, UTC, Centre International de Recherche sur l'Eau et l'Environnement, CNRS, France, 2000.

[54] Centre International De Formation, "Principes généraux de traitement des eaux", Lyonnaise des Eaux, France, 2000.

[55] A. Ben Messaoud, "Etude expérimentale et modélisation d'un procédé de désinfection par rayonnement UV", Thèse de Doctorat, Institut National des Sciences Appliquées de Toulouse, France, 2009.

[56] H. R. Hernandez De León, "Supervision et diagnostic des procédés de production d'eau potable", Thèse de Doctorat, Institut National des Sciences Appliquées de Toulouse, CNRS, N° 06584, France, 2006.

[57] S. Dash, V. Venkatasubramanian, "Challenges in the industrial applications of fault diagnostic systems", Proceedings of the Conference on Process Systems Engineering Computing & Chemical Engineering, Keystone, Colorado, USA, pp. 785-791, 2000.

[58] M. Combacau, "Commande et surveillance des systèmes à évènements discrets complexes : Application aux ateliers flexibles", Thèse de doctorat, Université Paul Sabatier, Toulouse, France 1991.

[59] D. Racoceanu, "Contribution à la surveillance des systèmes de production en utilisant les techniques de l'intelligence artificielle", Thèse d'HDR, Université de Franche-Comté de Besançon, France, N° d'ordre : 140, 2006.

[60] B. Dubuisson, et al., "Diagnostic, Intelligence artificielle et reconnaissance de formes", Edition Hermes, France, 2001.

[61] P.Y. Zhou, K.C.C. Chan, "Fuzzy Feature Extraction for Multichannel EEG Classification ", IEEE Transactions on Cognitive and Developmental Systems, Vol. 10, pp. 267-279, 2018.

[62] A. Mokhtari, "Diagnostic des systèmes hybrides ; développement d'une méthode associant la détection par classification et la simulation dynamique", Thèse de Doctorat, Institut national des sciences appliquées de Toulouse, France, 2007.

[63] C.V.I. Narvez, "Diagnostic par technique d'apprentissage floues ; conception d'une méthode de validation et d'optimisation des partitions", Thèse de Doctorat, Université de Toulouse, France, 2007.

[64] J. Montmain, "Interprétation qualitative de simulation pour le diagnostic en ligne de procédé continus", Thèse de doctorat, Institut National Polytechnique de Grenoble, France 1992.

[65] H. Derbel, "Diagnostic á base de modèles des systèmes temporisés et d'une sous-classe de systèmes dynamiques hybrides", Thèse de doctorat, Université Joseph-Fourier - Grenoble I, France 2009.

[66] M. Basseville, "Detecting changes in signals and systems - A survey", Automatica, Vol. 24, pp. 309-326, 1988.

[67] M. Zemouri, "Contribution à la surveillance des systèmes de production à l'aide des réseaux de neurones dynamiques : Application à la e-maintenance", Thèse de doctorat, Université de Franche-Comté, France, 2003.

[68] O. Ondel, "Diagnostic par reconnaissance des formes : Application à un ensemble convertisseur-machine asynchrone", Thèse de doctorat, École Centrale de Lyon, France, 2006.

[69] K. Hameed, et al., "A comprehensive review of fruit and vegetable classification techniques", Image and Vision Computing, Vol. 80, pp. 24-44, 2018.

[70] A. Belaout, et al., "Multiclass adaptive neuro-fuzzy classifier and feature selection techniques for photovoltaic array fault detection and classification", Renewable Energy, Vol. 127, pp. 548-558, 2018.

[71] J. Mushava, M. Murray, "An experimental comparison of classification techniques in debt recoveries scoring: Evidence from South Africa's unsecured lending market", Expert Systems with Applications, Vol. 111, pp. 35-50, 2018.

[72] M. Wesoły, P. Ciosek-Skibińska, " Comparison of various data analysis techniques applied for the classification of pharmaceutical samples by electronic tongue ", Sensors and Actuators B: Chemical, Vol. 267, pp. 570-580, 2018.

[73] M. Elkano, et al., " CHI-BD: A fuzzy rule-based classification system for Big Data classification problems ", Fuzzy Sets and Systems, vol. 348, pp. 75-101, 2018.

[74] K. Rajesh, R. Dhuli, "Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier", Biomedical Signal Processing and Control, Vol. 141, pp. 242-254, 2018.

[75] S. Rajeshbabu, B.V. Manikandan, "Detection and classification of power quality events by expert system using analytic hierarchy method", Cognitive Systems Research, Vol. 52, pp. 729-740, 2018.

[76] A. Goyal, et al., "Recent Named Entity Recognition and Classification techniques: A systematic review", Computer Science Review, Vol. 29, pp. 21-43, 2018.

[77] S. Breker, et al., "Hosting capacity of low-voltage grids for distributed generation: Classification by means of machine learning techniques", Applied Soft Computing, Vol. 70, pp. 195-207, 2018.

[78] G. Asch, "Les capteurs en instrumentation industrielle", Dunod, Edition 4, p. 816, 1993.

[79] H. Imen, "Etudes physico-chimique de capteurs à base de nanomatériaux pour des applications biomédicales", Thèse de doctorat, INSA, Lyon, France, 2009.

[80] L. Parra, et al., "Design and development of low cost smart turbidity sensor for water quality monitoring in fish farms", Aquacultural Engineering, Vol. 81, pp. 10-18, 2018.

[81] www.fr.endress.com.

[82] www.swan.ch.

[83] www.openpr.com.

[84] www.au.endress.com.

[85] www.partech.co.uk.

[86] www.idronaut.it.

[87] M. Zayed, "Véhicules Intelligents : Etude et développement d'un capteur intelligent de vision pour l'attelage virtuel", Thèse de Doctorat, Université des Sciences et Technologie de Lille, France, 2005.

[88] J.M. Giachino, "Smart Sensors", Sensors & Actuators, Vol.10, pp.239, 1986.

[89] M. Bayart et al., "Capteurs et actionneurs intelligents", Techniques de l'Ingénieur, S 7520.

[90] D. Luttenbacher, "Modélisation du concept capteur intelligent par une approche orientée objet : Application à un capteur intelligent de température", Thèse de Doctorat, Université Henri Poincaré - Nancy 1, France, 1997.

[91] S. Chakkor, "E-diagnostic de processus physiques à base des méthodes de haute résolution Application : machines éoliennes", Thèse de doctorat, Université Abdelmalek Essaâdi, Maroc, 2015.

[92] R. Verdone, et al., "Wireless Sensor and Actuator Networks, Technologies, Analysis and Design", Academic Press, Elsevier, London, 2008.

[93] J. Qiu et al., "A Low-Frequency Resonant Electromagnetic Vibration Energy Harvester Employing the Halbach Arrays for Intelligent Wireless Sensor Networks", IEEE Transactions on Magnetics, Vol. 51, 2015.

[94] S.C. Mukhopadhyay, J.A. Jiang, "Wireless Sensor Networks and Ecological Monitoring", Springer-Verlag Berlin Heidelberg, 2013.

[95] W. Dargie, C. Poellabauer, "Fundamentals of Wireless Sensor Networks Theory and Practice", Wiley series on Wireless Communications and Mobile Computing, first edition, John Wiley & Sons Ltd., 2010.

[96] Y. Zhang, et al., "RFID and Sensor Networks Architectures, Protocols, Security and Integrations", CRC Press Taylor & Francis Group, LLC, 2010.

[97] I. Stojmenovic, "Handbook of Sensor Networks Algorithms and Architectures", Wiley Series on Parallel and Distributed Computing, John Wiley & Sons Ltd, 2005.

[98] A.B.G. Hernando et al., "Problem Solving for Wireless Sensor Networks", Computer Communications and Networks series, Springer-Verlag London Limited, 2008.

[99] K. Sohraby, et al., "Wireless Sensor Networks Technology, Protocols, and Applications", John Wiley & Sons Ltd, 2007.

[100] J. Park, et al., "Practical Data Communications for Instrumentation and Control", Newnes, Great Britain,2003.

[101] E. Gaura, et al., "Wireless Sensor Networks Deployments and Design Frameworks", Springer-Verlag Berlin Heidelberg, 2010.

[102] L. Parra, et al., "Design and Deployment of Low-Cost Sensors for Monitoring the Water Quality and Fish Behavior in Aquaculture Tanks during the Feeding Process", Sensors, Vol.18, p. 750, 2018.

[103] M. Zhang, et al., "Design and Development of Water Quality Monitoring System Based on Wireless Sensor Network in Aquaculture", In Proceedings of the International Conference on Computer and Computing Technologies in Agriculture, Nanchang, China, pp. 629-641, 2010.

[104] L. Qi, et al., "Developing WSN-based traceability system for recirculation aquaculture", Math. Comput. Model, Vol.53, pp. 2162-2172, 2011.

[105] J. Huang, et al., "Development and test of aquacultural water quality monitoring system based on wireless sensor network", Trans. Chin. Soc. Agric. Eng, Vol.29, pp. 183-190, 2013.

[106] D.S. Simbeye, S.F. Yang, "Water Quality Monitoring and Control for Aquaculture Based on Wireless Sensor Networks", Trans. Chin. Soc. Agric. Eng, Vol. 9, pp. 840-849, 2014.

[107] X. Zhu, et al., "A remote wireless system for water quality online monitoring in intensive fish culture", Comput. Electron. Agric, Vol. 71, pp. S3-S9, 2010.

[108] B. O'Flynn, et al., "Experiences and receommendations in deploying a real-time, water quality monitoring system", Measurement Science and Technology, Vol. 21, pp. 4004-4014, 2010.

[109] M. Nasirudin, U. Za'bah, O. Sidek, "Fresh water real-time monitoring system based on wireless sensor network and gsm", IEEE Conference on Open Systems (ICOS), pp. 354-357, Malaysia, 2011.

[110] F. Adamo, et al., "A smart sensor network for sea water quality monitoring", IEEE Sensors Journal, Vol. 15, pp. 2514-2522, 2015.

[111] www.libelium.com.

[112] H. Haimi, "Data-derived soft sensors in biological wastewater treatment. With application of multivariate statistical methods", Thèse de doctorat, Université Aelto, Finland, 2016.

[113] F.A. A. de Souza, "Computational Intelligence Methodologies for Soft Sensors Development in Industrial Processes", Thèse de doctorat, Université Coimbra, Portugal, 2014.

[114] N. M. Khalfe, "Soft sensor development and optimization of the commercial petrochemical plant integrating support vector regression and genetic algorithm" Chemical Industry & Chemical Engineering Quarterly, Vol. 15, pp. 175-187, 2009.

[115] A. Abusnina, "Gaussian Process Adaptive Soft Sensors and their Applications in Inferential Control Systems", Thèse de doctorat, University of York, United Kingdom, 2014.

[116] D. Aguado, et al., "A methodology for sequencing batch reactor identification with artificial neural networks: a case study", Comput. Chem. Eng, Vol. 33, pp. 465-472, 2009.

[117] S.H. Hong, et al., "Monitoring of sequencing batch reactor for nitrogen and phosphorus removal using neural networks", Biochemical Engineering Journal, Vol. 35, pp. 365-370, 2007.

[118] M. Bagheri, et al., "Modeling of a sequencing batch reactor treating municipal wastewater using multi-layer perceptron and radial basis function artificial neural networks", Process Safety and Environmental Protection, Vol. 93, pp. 111-123, 2015.

[119] Y. Kim, et al., "Soft Sensor Using PNN Model and Rule Base for Wastewater Treatment Plant". Advances in Neural Networks - ISNN, pp. 1261-1269, 2006.

[120] Y.L. He, et al., "Data driven soft sensor development for complex chemical processes using extreme learning machine Author links open overlay panel", Chemical Engineering Research and Design, Vol. 102, pp. 1-11, 2015.

[121] H. Li, D. Yu, et J. Braun, "A review of virtual sensing technology and application in building systems", Applied Soft Computing, HVAC&R Research, Vol. 17, no. 5, pp. 619-645, 2011.

[122] X. Wang, "Soft sensor modeling method for freezing point of diesel fuel based on PCA and LS-SVM", in Intelligent Control and Automation, WCICA. 7th World Congress on. IEEE, pp. 9157-9161, 2008.

[123] W. Ni, et al., "Moving-window GPR for nonlinear dynamic system modeling with dual updating and dual preprocessing", Industrial & Engineering Chemistry Research, Vol. 51, no. 18, pp. 6416-6428, 2012.

[124] S. Khatibisepehr, B. Huang, S. Khare, "Design of inferential sensors in the process industry: A review of bayesian methods", Journal of Process Control, Vol. 23, No. 10, pp. 1575-1596, 2013.

[125] B. Lin, et al., "A systematic approach for soft sensor development", Computers & chemical engineering, Vol. 31, No. 5, pp. 419-425, 2013.

[126] P. Bangert, "Optimization for Industrial Problems", Springer, 2012.

[127] T. Kulcsar, "Data mining and machine learning algorithms for soft sensor development", Thèse de doctorat, Université Pannonia, Hongrie, 2016.

[128] G. Zahedi, et al., "Hybrid Artificial Neural Network-First Principle Model Formulation for the Unsteady State Simulation and Analysis of a Packed Bed Reactor for CO2 Hydrogenation to Methanol", Chemical Engineering Journal, Vol. 115, No. 1-2, pp. 113-120, 2005.

[129] P. Kadlec, R. Grbic, B. Gabrys, "Review of Adaptation Mechanisms for Data-Driven Soft Sensors", Computers & Chemical Engineering, Vol. 35, No. 01, pp. 1-24, 2011.

[130] C. K. Amos, "Soft sensors win hard jobs", Chemical Processing, Vol. 68, No. 12, pp. 31-33, 2005.

[131] B. O. Teixeira, et al., "Data driven soft sensor of downhole pressure for a gas-lift oil well", Control Engineering Practice, Vol. 22, pp. 34-43, 2014.

[132] C. Rosen, "A Chemometric Approach to Process Monitoring and Control with Applications to Wastewater Treatment Operation", Thèse de Doctorat, Université de Lund, Suède, 2001.

[133] S.H. WOO, et al., "On-line estimation of key process variables based on kernel partial least squares in an industrial cokes wastewater treatment plant ", J. Hazard. Mater, Vol. 161, pp. 538-544, 2009.

[134] J. Liu, "Developing a soft sensor based on sparse partial least squares with variable selection", Journal of Process Control, Vol. 24, pp. 1046-1056, 2014.

[135] L. Bao, X. Yuan, Z. Ge, "Co-training partial least squares model for semi-supervised soft sensor development", Chemometrics and Intelligent Laboratory Systems, Vol. 147, pp. 75-85, 2015.

[136] M.W. Lee, et al., "Real-time remote monitoring of small-scaled biological wastewater treatment plants by a multivariate statistical process control and neural network-based software sensors", Process. Biochem, Vol. 43, pp. 1107-1113, 2008.

[137] J. Qiao, et al., "The chemical oxygen demand modelling based on a dynamic structure neural network", In: Einschlag, F.S.G. (Ed.), Waste Water Evaluation and Management. In Tech, 2011.

[138] A.A. Francisco, et al., "A multilayer-perceptron based method for variable selection in soft sensor design", Journal of Process Control, Vol. 23, pp. 1371-1378, 2013.

[139] X. Wang, H. Liu, "Soft sensor based on stacked auto-encoder deep neural network for air preheater rotor deformation prediction", Advanced Engineering Informatics, Vol. 36, pp. 112-119, 2018.

[140] Q. Cong, W. Yu, "Integrated Soft Sensor with Wavelet Neural Network and Adaptive Weighted Fusion for Water Quality Estimation in Wastewater Treatment Process", Measurement, In Press, 2018.

[141] T. Kohonen, "Self-Organizing Maps", Springer, Vol. 30, 2001.

[142] D. Dürrenmatt, W. Gujer, "Data-driven modeling approaches to support wastewater treatment plant operation. Environ", Modell. Softw, Vol. 30, pp. 47-56, 2012.

[143] J. Wan, et al., "Prediction of effluent quality of a paper mill wastewater treatment using an adaptive network-based fuzzy inference system", Appl. Softw. Comput, Vol. 11, pp. 3238-3246, 2011.

[144] T.Y. Pai, et al., "Predicting effluent from the wastewater treatment plant of industrial park based on fuzzy network and influent quality", Appl. Math. Model, vol. 35, pp. 3674-3684, 2011.

[145] D. Dovžan, V. Logar, I. Skrjanc, "Implementation of an Evolving Fuzzy Model (eFuMo) in a Monitoring System for a Waste-Water Treatment Process", IEEE Transactions on Fuzzy Systems, Vol. 23, pp. 1761-1776, 2015.

[146] J. Mendes, et al., "Genetic fuzzy system for data-driven soft sensors design", Applied Soft Computing, Vol. 12, pp. 3237-3245, 2012.

[147] M. Zhang, X. Liu, "A Soft Sensor based on Adaptive Fuzzy Neural Network and Support Vector Regression for industrial Melt Index prediction", Chemometrics and Intelligent Laboratory Systems, Vol. 126, pp. 83-90, 2013.

[148] H. Kaneko, K. Funatsu, "Adaptive soft sensor based on online support vector regression and Bayesian ensemble learning for various states in chemical plants", Chemometrics and Intelligent Laboratory Systems, Vol. 137, pp. 57-66, 2014.

[149] K. Pearson, "On Lines and Planes of Closest Fit to Systems of Points in Space", Philosophical Magazine, Vol. 2, pp. 559-572, 1901.

[150] H. Hotelling, "Analysis of a complex of statistical variables into principal components", Journal of Educational Psychology, Vol. 24, pp. 498-520, 1933.

[151] H. Hernandez, "Développement d'un capteur logiciel pour la prédiction de la dose de coagulant dans une station de traitement d'eau potable en vue de son diagnostic", Rapport LAAS N° 05175, 6ème Congrès des Doctorats de l'Ecole Doctorat Systèmes (EDSYS), 17-20 Mai, Toulouse, France, 2005.

[152] J. de Lagarde, "Initiation à l'analyse des données", Dunod, Paris, France, 1983.

[153] C. Archaux, "Conception d'un système d'information dédie à l'estimation de la valeur des clients en téléphonie mobile prépayée", Thèse de Doctorat, Université de Nantes, France, N° ED 366-193, 2005.

[154] H. Najmeddine, "Méthode d'identification et de classification de la consommation d'énergie par usage en vue de l'intégration dans un compteur d'énergie électrique", Thèse de Doctorat, Université Blaise Pascal-Clermont II, France, 2009.

[155] S. Tertois, "Réduction des effets des non linéarités dans une modulation à l'aide de réseaux de neurones", Thèse de Doctorat, Université de Rennes 1, France, N° d'ordre : 2924, 2003.

[156] W. S. Mc Culloch, W. Pitts, "A Logical Calculus of the Ideas Imminent in Nervous Activity", Bulletin of math. biophysics, Vol. 05. pp. 115-133, 1943.

[157] J.J. Hopfield, "Neural Networks and physical systems with emergent collective computational abilities", Proceedings of the National Academy of Sciences, Vol. 79. pp. 2554-2558, 1982.

[158] M. Bouamar, M. Ladjal, "Performance evaluation of three pattern classification techniques used for water quality monitoring", International Journal of Computational Intelligence and Applications, Vol. 11, N° 2, pp. 1-14, June 2012.

[159] Y. Es-saady, "Contribution au développement d'approches de reconnaissance automatique de caractères imprimés et manuscrits, de textes et de documents Amazighes", Thèse de doctorat, Université Ibn Zohr, Agadir, Maroc, 2012.

[160] M. Bouamar, M. Ladjal, "Système multicapteur utilisant les réseaux de neurones artificiels pour la surveillance des eaux potables", Proceedings of SETIT'07, 4th International Conference on Sciences of Electronics, Technologies of Information and Telecommunication, 25-29 Mars, Hammamet, Tunisia, pp. 124, 2007.

[161] S. Haykin, "Neural Networks: A Comprehensive Foundation", Prentice-Hall, New Jersey, USA, 1999.

[162] M. Djerioui, M. Ladjal, M. Bouamar, "Capteur logiciel utilisant les réseaux de neurones artificiels pour la surveillance des eaux potables", Première conférence nationale de l'électronique et les nouvelles technologies NCENT, Algérie, 2015.

[163] P. Besse, B. Laurent, "Apprentissage Statistique : Prévision et data mining", Institut de Mathématiques de Toulouse, France, 2012.

[164] E. Davalo, P. Naïm, "Des réseaux de neurones", Edition Eyrolles, 2ème édition, Paris, France, 1993.

[165] X. Ding, S. Canu, T. Denoeux, "Neural network models for forecasting", In J. G. Taylor, editor, Neural Networks and their Applications, John Wiley and Sons, pp. 243-252, Chichester, UK,1996.

[166] M. J. D. Powell, "The Theory of Radial Basis Function Approximation in 1990", pp. 105-210. Oxford University Press, USA, 1992.

[167] M. Bouamar, M. Ladjal, "A comparative study of RBF Neural Network and SVM classification techniques performed on real data for drinking water quality", Proceeding of (IEEE). Systems, Signals and Devices, 5th IEEE International Multi-Conf Systems, Signals & Devices (IEEE SSD'08), pp. 201, July 20-23, Amman, Jordan, 2008.

[168] G. H. Cheng, et al., "Trust region based mode pursuing sampling method for global optimization of high dimensional design problems", Journal of Mechanical Design, Vol. 137, No. 02, 2015.

[169] R.G. Regis, "Constrained optimization by radial basis function interpolation for high-dimensional expensive black-box problems with infeasible initial points", Engineering Optimization, Vol. 46, pp. 218-243, 2014.

[170] R.G. Regis, "Evolutionary programming for high-dimensional constrained expensive black-box optimization using radial basis functions", IEEE Trans. Evolutionary Computation, Vol. 18, pp. 326-347, 2014.

[171] R.G. Regis, C.A. Shoemaker, "A stochastic radial basis function method for the global optimization of expensive functions", INFORMS Journal on Computing, Vol. 19, pp. 479-509, 2007.

[172] J. Moody, C. Darken, "Fast Learning in Networks of Locally-tuned Processing Units", Neural Computation, Vol. 01, pp. 281-294, 1989.

[173] D. Wettschereck, T. Dietterich. "Improving the Performance of Radial Basis Function by Learning Center Locations", In Advances in Neural Information Processing System 4, pp. 1133-1140. Morgan Kaufmann Publishers, San Mateo,CA, 1992.

[174] M. H. Hassoun. "Fundamentals of Articial Neural Networks", MIT Press, 1995.

[175] J. F. Hebert, "Architecture neuronale hybride pour l'apprentissage incrémental des connaissances: application à la reconnaissance d'écriture cursive", Thèse de doctorat, Université Laval, Québec, 1999.

[176] C. Cortes, V. Vapnik, "Support-vector networks", Machine learning, Vol. 20, pp. 273-297, 1995.

[177] V. Vapnik, "Statistical Learning Theory", John Wiley & Sons, 1998.

[178] M. Bouamar, M. Ladjal, "Evaluation of the performances of ANN and SVM techniques used in water quality classification", Proceedings of ICECS'07, 14th IEEE International Conference on Electronics, Circuits and Systems, pp. 1047-1050, Marrakech, Morocco,11-14 December 2007.

[179] M. Baccouche, "Apprentissage neuronal de caractéristiques spatio-temporelles pour la classification automatique de séquences vidéo", Thèse de doctorat, Université Lyon, France, 2013.

[180] A. Djeffal, "Utilisation des méthodes Support Vector Machine (SVM) dans l'analyse des bases de données", Thèse de doctorat, Université Mohamed Khider - Biskra, Algérie, 2012.

[181] F. W. Smith, "Pattern classifier design by linear programming", IEEE Trans. on Computers, Vol. 17, pp. 367-372, 1968.

[182] K. P. Bennett, O. L. Mangasarian, "Robust linear programmating discrimination of two linearly inseparable sets", Optimization Method and Software, Vol. 01, pp. 23-34, 1992.

[183] W. Karush, "Minima of functions of several variables with inequalities as side constraints", Master's thesis, Dept. of Mathematics, Univ. of Chicago, 1939.

[184] H.W. Kuhn, A.W. Tucker, "Nonlinear programming", In University of California Press, editor, 2nd Berkeley Symposium on Mathematical Statistics and Probabilistics, pp.481-492, 1951.

[185] A. Amirou, "Optimisation des SVMs pour la discrimination de signaux", Thèse de doctorat, Université Mouloud Mammeri, Tizi-Ouzou, Algérie, 2015.

[186] J. Mercer, "Functions of positive and negative type, and their connection with the theory of integral equations", Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, Vol. 209, pp. 415-446, 1909.

[187] H. König, "Eigenvalue distribution of compact operators", Birkhäuser Verlag Basel Boston Stuttgart, 1986.

[188] M. Bouamar, M. Ladjal, "La technique SVM appliquée à la surveillance des eaux potables", Proceedings of SETIT'09, 4th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications, pp. 191, Tunisia, 2009.

[189] M. Bouamar, M. Ladjal, "Multisensor system using support vectors machines for water quality classification", Proceeding of (IEEE) ISSPA'07, 9th IEEE International Symposium on Signal Processing and its Applications, Feb 12-15, Sharjah, UAE, 2007.

[190] S. Canu, "Machines à noyaux pour l'apprentissage statistique", Techniques d'ingénieur,TE 5 255, 2007.

[191] M. Ramona, "Classification automatique de flux radiophoniques par machines à vecteurs de support", Thèse de Doctorat, l'école télécom ParisTech, France, 2010.

[192] M. Ladjal, M. Djerioui, M. Bouamar, "Contribution de l'analyse statistique multivariée à l'étude de la classification supervisée de données environnementales à l'aide des techniques d'intelligence artificielle", Première conférence nationale de l'électronique et les nouvelles technologies, NCENT, Algérie, 2015.

[193] T.M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition", IEEE Transactions on Electronic Computers, Vol. 14. pp. 326-334, 1965.

[194] D. Basak, S. Pal, D.C. Patranabis, "Support vector regression ", Neural Information Processing-Letters and Reviews, Vol. 11, pp. 203-224, 2003.

[195] A. Farag, R.M. Mohamed, "Regression using support vector machines: Basic foundations", Technical report, CVIP Laboratory, University of Louisville, 2004.

[196] L. Hamel, "Knowledge discovery with support vector machines", Wiley Edition, 2009.

[197] D. Zerdazi, "L'approche neuronale de l'inférence statistique", Thèse de doctorat, Université Frères Mentouri, Constantine, Algérie, 2017.

[198] G. B. Huang, Q.Y. Zhu, C.K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks", Neural Networks, Proceedings. IEEE International Joint Conference on, pp. 985-990, 2004.

[199] G. Huang, et al. "Trends in extreme learning machines: A review", Neural Networks, Vol. 61, pp. 32-48, 2015.

[200] S. Bazi, "Contribution à la Détection et au Diagnostic des Défauts dans un Système Machine à Induction-Convertisseur ", Thèse de doctorat, Université de Batna 2, Algérie, 2016.

[201] G. Huang, et al., "Semi-supervised and unsupervised extreme learning machines", IEEE Transactions on Cybernetics, Vol. 44, pp. 2405-2417, 2014.

[202] K. Javed, "A robust and reliable data-driven prognostics approach based on Extreme Learning Machine and Fuzzy Clustering", Thèse de doctorat, université de franche comté, France, 2014.

[203] C.R. Rao, S.K. Mitra, "Generalized Inverse of Matrices and its applications", John Wiley and Sons, New York, 1971.

[204] J. Sola, J. Sevilla, "Importance of Input Data Normalization for the Application of Neural Networks to Complex Industrial problems", IEEE Transactions on Nuclear Science, Vol.44, No.03, pp.1464-1468, 1997.

[205] M. Bouamar, M. Ladjal, M. Djerioui, Y. Brik, "Performance evaluation of ANN and SVM multiclass models for intelligent water quality classification using Dempster-Shafer Theory", IEEE International Conference on Electrical and Information Technologies (ICEIT), Morocco, 2016.

[206] M. Ladjal, M. Bouamar, M. Djerioui, " Application of ANN and SVM multiclass models used for multi-sensor monitoring and measurement of surface water quality", The Mediterranean Journal of Measurement and Control, Vol.12, No.03, pp. 627-635, 2016.

[207] M. Ladjal, M. Bouamar, M. Djerioui, " Adaptive Neuro-Fuzzy Inference System for intelligent water quality classification in Tilesdit dam from Algeria ", Th international conference on electrical and electronics engineering (ELECO), Turkey, 2017.

[208] M. Ladjal, M. Bouamar, M. Djerioui, " SVM-PSO Approach Based Modeling and Optimization of Water Quality Classification ", International conference on electronics and new technologies ICENT, Algérie, 2017.

[209] M. Djerioui, M. Bouamar, M. Ladjal, A. Zerguine, "Chlorine Soft Sensor Based on Extreme Learning Machine for Water Quality Monitoring", Arabian Journal for Science and Engineering, Vol 44, No 3, pp 2033–2044, 2019.

ملخص :في الوقت الحالي، في مجال معالجة المياه، صار استخدام العمليات الأوتوماتيكية أمرًا حتميًا لتحقيق هدفين رئيسيين: التحكم في جودة المياه، وتخفيض تكاليف التشغيل. تلعب متغيرات الجودة دورًا مهمًا في مراقبة المياه ورصدها. بعض المتغيرات يمكن قياسها باستمرار باستخدام أجهزة استشعار منخفضة التكلفة ، إلا أن هناك متغيرات أخرى تتطلب تحليلات مخبرية محددة ومكلفة بسبب عدم وجود أجهزة استشعار مخصصة. بالإضافة إلى ذلك ، هناك عدد كبير من أجهزة الاستشعار غير المتجانسة التي يمكن أن تستغرق الكثير من الوقت أثناء خطوات القياس والمعالجة. النظام القائم على أجهزة الاستشعار المبرمجة هو وسيلة فعالة واقتصادية لحل هذه المشاكل.

يمكن اعتبار هذا العمل كمساهمة في الحلول المقترحة ، لحل المشاكل ذات الأهمية الاستراتيجية الوطنية والدولية، باستخدام الأدوات الحديثة القائمة على التقنيات المتقدمة. في هذا السياق ، نقدم مساهمة في دراسة وتطوير المستشعرات المبرمجة التي يمكن دمجها على مستوى أجهزة الاستشعار الذكية واستخدامها في مراقبة جودة المياه الصالحة للشرب. لقد أصبحت هذه المستشعرات المبرمجة المستندة إلى تقنيات الذكاء الاصطناعي تمثل طريقة مهمة جدا للتعامل مع عدم وجود مستشعرات محددة على مستوى مراقبة جودة المياه. من أجل اختيار التقنية الأنسب للتصميم ، يتم تقديم دراسة مقارنة لتقييم العديد من التقنيات لهذا الغرض. الهدف هو توفير نظام مراقبة يعتمد على استخدام المستشعرات المبرمجة التي يمكن دمجها في ارضية أجهزة الاستشعار الذكية.

Abstract: Currently, in the field of water treatment, the use of automatic processes becomes imperative to achieve two main objectives: the control of water quality, and the reduction of operating cost constraints. Quality variables play an important role in the control and monitoring of water. Although some parameters can be measured continuously using low-cost physical sensors, there are other still parameters that require specific and expensive laboratory analyses because of the lack of dedicated sensors. In addition, there is a large number of heterogeneous sensors that can take a long time in the measurement and processing stages. Nevertheless, the system based on soft sensors can provide an efficient and economical way to solve these problems.

Our work can be considered as a contribution to the proposed solutions, to solve problems of strategic interest of national and international concern, using modern tools based on advanced techniques. In this context, we present a contribution to the study and development of soft sensors that can be integrated at the level of intelligent sensors and used in the monitoring and control of the quality of clean water. These algorithmic sensors, based on artificial intelligence techniques and representing a very attractive way to cope with the lack of specific sensors at the level of these control chains, have become very coveted for this purpose. For the purpose of selecting the technique best suited to the design, a comparative study of the evaluation of several techniques is presented. Finally, the objective is to implement a surveillance system architecture based on soft sensors that can be integrated into intelligent sensors platforms.

Résumé : Actuellement, dans le domaine de traitement des eaux propres, l'utilisation de procédés automatiques devient impérative pour atteindre deux objectifs principaux : la maîtrise de la qualité de l'eau, et la diminution des contraintes de coût de fonctionnement. Les variables de qualité jouent un rôle important dans le contrôle et la surveillance de l'eau. Bien que certains paramètres puissent être mesurés en continu à l'aide de capteurs physiques à faible coût, il existe d'autres paramètres nécessitant des analyses de laboratoire spécifiques et coûteuses en raison de l'absence de capteurs dédiés. En outre, il existe un grand nombre de capteurs hétérogènes qui peuvent prendre beaucoup de temps lors des étapes de mesure et de traitement. Néanmoins, le système à base des capteurs logiciels peut fournir un moyen efficace et économique pour résoudre ces problèmes.

Notre travail peut être considéré comme une contribution aux solutions proposées, pour résoudre des problèmes d'intérêt stratégique à préoccupation nationale et internationale, utilisant des outils modernes à base de techniques avancées. On présente dans ce cadre, une contribution à l'étude et au développement de capteurs logiciels pouvant être intégrés au niveau de capteurs intelligents et utilisés dans la surveillance et le contrôle de la qualité des eaux propres. Ces capteurs algorithmiques à base de techniques de l'intelligence artificielle et représentant un moyen très attrayant pour faire face au manque de capteurs spécifiques, sont devenus alors très convoités. Dans un but de choix de la technique la mieux adaptée à la conception de ces capteurs logiciels, une étude comparative d'évaluation de plusieurs méthodes, est à cet effet présentée. L'objectif est de mettre en œuvre une architecture de système de surveillance fondée sur l'emploi de ce type de capteurs, pouvant être intégrés au sein de plateformes plus élaborées, construites autour de capteurs intelligents.

Mots clés : ***Eaux propres, Contrôle de la qualité, Intelligence artificielle, Capteurs logiciels, Régression, Classification, Evaluation, Simulation.***