

الجمهورية الجزائرية الديمقراطية الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR  
ET DE LA RECHERCHE SCIENTIFIQUE



جامعة محمد بوضياف - المسيلة

UNIVERSITE MOHAMED BOUDIAF - M'SILA

كلية التكنولوجيا

FACULTE DE TECHNOLOGIE

قسم الإلكترونيك

DEPARTEMENT D'ELECTRONIQUE

## **MEMOIRE DE MASTER**

DOMAINE : SCIENCES ET TECHNOLOGIE

FILIERE : GENIE ELECTRIQUE

OPTION : CONTROLE INDUSTRIEL

### **THEME**

METHODES DE REGRESSION POUR LA SURVEILLANCE DES EAUX  
PROPRES

*Réalisé par :*

*Ould Mohamed Cheibeta Mohamed Mostapha*

*Proposé et dirigé par :*

*Mr .DJERIOUI Mohamed*

*N° D'ordre:*

*2016/CI03/89/385*

**PROMOTION: JUIN 2016**

## *Remerciements*

*Je remercie mon Dieu miséricorde de m'avoir donné le courage de réaliser ce projet et de poursuivre mes études supérieures et de les réussir et d'avoir la chance d'atteindre le niveau MASTER II.*

*Je tiens à exprimer mes vifs remerciements :*

*A ma mère pour ses sacrifices en témoignage de tout mon affection au long de mes études.*

*A mon père qui m'a encouragé durant mes études, et son affection qui m'a amené à finir mon projet.*

*A mon encadreur monsieur M. DJREOUI pour sa prise en charge et pour avoir dirigé mon travail durant l'élaboration de ce mémoire.*

*Mes remerciements vont également à tous les enseignants du département d'électronique sans oublier toutes les personnes de ce même département.*

*Mes remerciements vont aussi à messieurs les membres du jury pour avoir accepté de juger mon travail.*

*Enfin, A tous ceux qui ont contribué de près ou de loin à la réalisation de ce projet de fin d'études.*

*Ould Mohamed cheibeta Mohamed Mostapha*

## *Dédicace*

*Je dédie ce modeste travail :*

*A ma mère pour ses sacrifices en témoignage de tout mon affection au long de mes études.*

*A mon père qui m'a encouragé durant mes études, et son affection qui m'a amené à finir mon projet.*

*A mes chers frères.*

*A mes chères sœurs.*

*A toute la famille.*

*A mes collègues qui ont marqué leurs présences par réconfort indispensable.*

*A tous les amis de ma promotion (2016).*

*Ould Mohamed cheibeta Mohamed Mostapha*

# Table des matières

Introduction générale.....	01
----------------------------	----

## CHAPITRE I

### *Généralités sur les eaux propres*

Introduction .....	03
1. Cycle de l'eau .....	03
2. Ressources d'eau .....	06
3. Paramètres physico-chimiques mesurables .....	06
4. Chaîne de traitement de l'eau potable.....	09
4.1. Prétraitement.....	10
4.2. Pré-oxydation .....	10
4.3. Clarification .....	11
4.4. Oxydation-Désinfection.....	12
4.5. Affinage .....	13
5. les capteurs .....	13
5.1. Définitions et caractéristiques générales.....	14
5.2. Capteurs actifs.....	15
5.3. Capteurs passifs.....	16
5.4. Transformation de la grandeur physique.....	16
5.5. Qualité des capteurs .....	17
5.6. Sensibilité, précision, gamme de mesure .....	17
5.7. Fiabilité, environnement et dérive .....	18
5.8. Mesure des paramètres à l'aide des capteurs .....	18
Conclusion .....	20

## CHAPITRE II

*Les réseaux de neurones artificiels*

<b>Introduction</b> .....	<b>21</b>
<b>1. Généralité</b> .....	<b>21</b>
<b>2. Eléments de bases</b> .....	<b>22</b>
<b>2.1 Le modèle neurophysiologique</b> .....	<b>22</b>
<b>2.2. Le modèle mathématique</b> .....	<b>23</b>
<b>2.3. Fonctions d'activations</b> .....	<b>24</b>
<b>3. Architectures des réseaux de neurones</b> .....	<b>26</b>
<b>3.1. Les réseaux de neurones non bouclés</b> .....	<b>26</b>
<b>3.1.1. Les réseaux de neurones complètement connectés</b> .....	<b>27</b>
<b>3.1.2. Les réseaux de neurones à couches</b> .....	<b>27</b>
<b>3.2. Les réseaux de neurones bouclés</b> .....	<b>27</b>
<b>4. Quelques réseaux de neurones usuels</b> .....	<b>29</b>
<b>4.1. Perceptron</b> .....	<b>29</b>
<b>4.2. Perceptron multicouches(PMC)</b> .....	<b>29</b>
<b>4.3. Réseaux RBF (Radial Basis Function)</b> .....	<b>30</b>
<b>4.3.1. Architecture générale d'un réseau RBF</b> .....	<b>30</b>
<b>4.3.2. Avantages et inconvénients des réseaux RBF</b> .....	<b>31</b>
<b>5. Apprentissage</b> .....	<b>31</b>
<b>6. Types d'apprentissage</b> .....	<b>32</b>
<b>6.1. Apprentissage supervisé</b> .....	<b>32</b>
<b>6.2. Apprentissage renforcé</b> .....	<b>33</b>
<b>6.3. Apprentissage non supervisé</b> .....	<b>33</b>
<b>7. Algorithme d'apprentissage</b> .....	<b>33</b>
<b>7.1. Algorithme de rétropropagation</b> .....	<b>34</b>

7.2. Méthodes Quasi-Newton .....	35
7.3. Algorithme de BFGS .....	35
8. Avantages et inconvénients des réseaux de neurones .....	36
8.1. Avantages des réseaux de neurones .....	36
8.2. Inconvénients des réseaux de neurones .....	36
Conclusion .....	37

## CHAPITRE III

### *Les machines à vecteurs de support*

Introduction .....	38
1. Généralité .....	38
2. principe de fonctionnement .....	39
2.1. Notions de base: Hyperplan, marge et support vecteur .....	39
2.2. Maximisation de la marge .....	41
2.3. Linéarité et non-linéarité .....	42
2.4. Cas non linéaire .....	43
3. SVM appliquée à la classification binaire .....	43
3.1. Classifieur linéaire .....	45
3.2. Les fonctions noyau et le cas non-linéaire .....	55
Conclusion .....	59

## CHAPITRE IV

*Simulation et évaluation*

<b>Introduction .....</b>	<b>60</b>
<b>1. problématique.....</b>	<b>60</b>
<b>1.1. Architecture du système de contrôle et de surveillance .....</b>	<b>60</b>
<b>1.2. Approche utilisée dans la surveillance.....</b>	<b>61</b>
<b>2. Description des données d'entrée .....</b>	<b>62</b>
<b>3. Développement du capteur logiciel : Oxygène Dissous .....</b>	<b>63</b>
<b>3.1.Apprentissage et Généralisation .....</b>	<b>64</b>
<b>3.1.1 Modèle PMC.....</b>	<b>65</b>
<b>3.1.2 Modèle neuronal-RBF .....</b>	<b>66</b>
<b>3.1.3. Modèle SVM .....</b>	<b>67</b>
<b>3.2. Evaluation des performances.....</b>	<b>69</b>
<b>3. 3. Discussion des résultats .....</b>	<b>70</b>
<b>Conclusion.....</b>	<b>71</b>
<b>Conclusion générale.....</b>	<b>72</b>

## *Liste des figures*

Figure.1.1. Cycle de l'eau .....	04
Figure.1.2. Représentation schématique du champ de recherche « environnement développement » .....	05
Figure.1.3, a. Problèmes fréquemment rencontrés dans les eaux souterraines .....	06
Figure.1.3, b. Problèmes fréquemment rencontrés dans les eaux superficielles .....	07
Figure.1.4. Chaîne de traitement de l'eau potable.....	09
Figure.1. 5 .Exemple d'évolution d'un mesurande « m » et de la réponse « s » correspondante du capteur.....	14
Figure. 1. 6. Exemples d'application de Capteurs actifs.....	15
Figure.2.1.Structure d'un réseau de neurones biologiques .....	21
Figure.2.2. Un neurone avec son arborisation dendritique .....	22
Figure.2.3, a– Un potentiel d'action.....	23
Figure.2.3, b– Fonctionnement au niveau synaptique .....	23
Figure.2.4.Mise en correspondance du neurone biologique /et du neurone artificiel .....	24
Figure.2.5. Neurone formel .....	24
Figure.2.6.Différentes fonctions d'activation .....	25
Figure.2.7. Réseau de neurones à n entrées, une couche de $N_c$ neurones cachés et $N_0$ neurones de sortie .....	26
Figure.2.8. Réseau de neurone bouclé .....	28
Figure.2.9.Représentation graphique de perceptron .....	29
Figure.2.10. Architecteur générale d'un réseau RBF .....	30
Figure. 3.1. Exemple de classification binaire linéaire et non-linéaire.....	39
Figure .3.2.Exemple d'un hyperplan séparateur .....	40
Figure.3.3.Exemple de vecteurs de support .....	40

<b>Figure.3.4.Exemple de marge maximal (hyperplan valide).....</b>	<b>41</b>
<b>Figure. 3.5.a. Hyperplan avec faible marge .....</b>	<b>41</b>
<b>figure.3.5. b. Meilleur hyperplan séparateur .....</b>	<b>41</b>
<b>Figure. 3.6. Exemple de classification d'un nouvel élément .....</b>	<b>42</b>
<b>Figure. 3.7 a. Cas linéairement séparable .....</b>	<b>42</b>
<b>Figure.3.7.b. Cas non linéairement séparable.....</b>	<b>42</b>
<b>Figure. 3.8. Exemple de changement de l'espace de données.....</b>	<b>43</b>
<b>Figure. 3.9. Hyperplan séparateur linéaire optimal et marge .....</b>	<b>45</b>
<b>Figure. 3.10. Expression de la marge pour l'exemple <math>x_i</math> .....</b>	<b>46</b>
<b>Figure. 3.11. Importance de la marge de l'hyperplan .....</b>	<b>48</b>
<b>Figure. 3.12. Hyperplans canoniques .....</b>	<b>49</b>
<b>Figure. 3.13. Les vecteurs de support .....</b>	<b>52</b>
<b>Figure. 3.14. Distance séparant le point aberrant de l'hyperplan canonique .....</b>	<b>54</b>
<b>Figure. 3.15. Illustration de l'effet de changement d'espace (mapping) par une fonction noyau .....</b>	<b>56</b>
<b>Figure.3.16. Chaîne de traitement générique des méthodes à noyaux .....</b>	<b>58</b>
<b>Figure.4.1. Architecture du système de contrôle et de surveillance .....</b>	<b>61</b>
<b>Figure.4.2. Evolution des dix paramètres descripteurs .....</b>	<b>65</b>
<b>Figure.4.3. Résultats d'apprentissage du modèle PMC .....</b>	<b>65</b>
<b>Figure.4.4. Résultats du test du modèle PMC .....</b>	<b>66</b>
<b>Figure.4.5. Résultats d'apprentissage du modèle neuronal-RBF .....</b>	<b>66</b>
<b>Figure.4.6. Résultats du test du modèle neuronal-RBF .....</b>	<b>67</b>
<b>Figure.4.7. Résultats d'apprentissage du modèle SVM .....</b>	<b>68</b>
<b>Figure.4.8. Résultats du test du modèle SVM .....</b>	<b>68</b>

---

## *Liste des Tableaux*

<b>Tableau. 1. 1. Capteurs actifs : principes physiques de base .....</b>	<b>15</b>
<b>Tableau. 1. 2 .Capteurs passifs : principes physiques et matériaux .....</b>	<b>16</b>
<b>Tableau .4.1.Résumé statistique des paramètres d'une eau brute .....</b>	<b>63</b>
<b>Tableau. 4.2. Résultats d'apprentissage (Modèle PMC) .....</b>	<b>65</b>
<b>Tableau. 4.3. Résultats d'apprentissage (Modèle neuronal-RBF) .....</b>	<b>66</b>
<b>Tableau. 4.4. Résultats d'apprentissage (Modèle SVM) .....</b>	<b>67</b>
<b>Tableau. 4.5. Résultats statistiques des modèles PMC, RBF et SVM.....</b>	<b>69</b>
<b>Tableau. 4.6. Performances des modèles PMC, RBFet SVM.....</b>	<b>70</b>
<b>Tableau. 4.7. Evolution de l'erreur relative des modèles PMC, RBF et SVM.....</b>	<b>70</b>
<b>Tableau. 4.8. Tableau comparatif des caractéristiques des modèles PMC, RBF et SVM.....</b>	<b>71</b>

## **Résumé :**

L'eau est un élément essentiel dans la vie. Il est donc nécessaire de la contrôler, la surveiller pour la rendre potable. Dans ce cadre des efforts ont été effectués durant ces dernières années, pour développer des méthodes de contrôle et de surveillance automatique de l'eau. Ce travail s'inscrit dans ce courant de recherche, en proposant une nouvelle approche pour le contrôle et la surveillance de la qualité des eaux potables. Ce mémoire aborde le problème de la surveillance des eaux potables en fonction des caractéristiques physico-chimiques de l'eau brute à l'aide d'un modèle basé sur les techniques de l'apprentissage statistiques telles que : les PMC, RBF et SVM, appliquées au développement du capteur logiciel d'OD vu comme problème de régression et d'évaluer les performances de chacune après une simulation à l'aide du logiciel MATLAB pour avoir la technique adaptée entre elles.

## **Mots clés :**

Capteur logiciel, Eaux potables, Perceptron Multicouches, Régression, Réseaux de neurones, RBF, Surveillance, SVM.

## ***Introduction générale***

L'eau joue un rôle déterminant dans la vie des hommes, des animaux et des plantes, elle est essentielle pour tous les organismes vivants. L'eau est un bien commun, elle est une ressource indispensable à la vie. La consommation en eau par habitant est désormais considérée comme un indicateur du développement économique d'un pays. Selon une étude des Nations Unies, l'eau pourrait même devenir, d'ici à 50 ans, un bien plus précieux que le pétrole. C'est dire toute l'importance de cette ressource que d'aucuns appellent déjà <l'or bleu> [1].

Donc suivant cette importance de l'eau, il est nécessaire de trouver des solutions pour la contrôler et la surveiller pour avoir une eau de plus grande qualité et avec un coût de production plus faible. Ces considérations motivent les efforts importants dans le développement de nouvelles méthodes et techniques de contrôle et de surveillance.

L'exigence d'une réglementation très stricte des pouvoirs publics est alors bien justifiée. Durant ces dernières années, d'importants efforts ont été déployés dans le développement de méthodes de contrôle et de surveillance automatique dans ce domaine. Les techniques de l'intelligence artificielle qui servent comme outil de base pour l'aide à la décision, présentent une réponse plus élaborée qui peut être obtenue soit à partir de données brutes venant directement des variables de surveillance, soit à partir de données traitées venant des sorties de traitements de bas niveau. Parmi ces techniques, on trouve les réseaux de neurones artificiels (RNAs) et les machines à vecteurs de support (SVM), qui se démarquent des autres outils par leur capacité et rapidité d'apprentissage. Présentant des performances dans plusieurs domaines, ils sont utilisés en reconnaissance de formes comme outil de classification ou de régression [2].

Le travail présenté dans ce mémoire a pour objectif la mise en œuvre de trois méthodes d'apprentissage telles que : Perceptron Multi Couches(PMC), RBF (Radial Basis Function) et SVM, appliquées au domaine de reconnaissance de formes. L'application concerne le contrôle et la surveillance de la potabilité des eaux. Une étude en simulation est effectuée pour valider et évaluer les performances de chacune de ces méthodes dans un but comparatif, permettant un choix décisif de la technique la mieux adaptée.

Le mémoire est structuré en 4 chapitres :

Le premier chapitre présente une généralité sur les eaux propres. Dans cette section, nous parlerons tout spécialement du processus de traitement des eaux potables, en l'introduisant par des généralités sur leur cycle, leur chaîne de traitement, les ressources naturelles, ainsi que leurs paramètres physico-chimiques. Nous terminerons ce chapitre par les capteurs en général et en particulier les capteurs des paramètres descripteurs de l'eau.

Le deuxième chapitre est dédié à une présentation des réseaux de neurones artificiels. On commencera ce chapitre par une présentation de la structure d'un neurone biologique et d'un neurone formel, puis les architectures et l'apprentissage des réseaux PMC et RBF. Nous finaliserons ce chapitre par les avantages et les inconvénients des RNAs.

Le troisième chapitre présente une généralité sur les SVM, leur principe de fonctionnement, ainsi que des notions de bases essentielles concernant les SVM. Nous terminerons ce chapitre par une description détaillée sur les SVM appliquées à la classification binaire.

Enfin, dans le quatrième et le dernier chapitre on va présenter la base de données et les différents paramètres caractéristiques de l'eau. On va mettre en œuvre les trois techniques (PMC, RBF et SVM) appliquées au contrôle de la potabilité de l'eau. Une discussion des résultats de simulation obtenus, conclue cette étude pour le choix de la technique la mieux adaptée à l'application.

Une conclusion générale en fin de ce travail, retrace les différentes étapes réalisées et souligne les perspectives envisagées.

# *Chapitre I*

## *Généralités sur les eaux propres*

### **Introduction**

Le principal objectif d'une station de traitement d'eau potable est de fournir un produit qui satisfait à un ensemble de normes de qualité à un prix raisonnable pour le consommateur. A cet effet, le traitement de l'eau brute est un passage obligé dans le processus de production de l'eau potable. Le but principal dans les plupart des usines de production, est de connaître les différents paramètres de l'eau brute pour décider de son état, et cherche par la suite les méthodes pour la rendre potable. La surveillance permanente de la qualité de cette eau à travers des mesure qualifiables et quantifiables, ainsi que du fonctionnement des installations de traitement, est donc exigée afin d'éviter de ralentir la production et assurer un niveau élevé et uniforme de qualité.

Ce premier chapitre est consacré au processus de contrôle de traitement de l'eau brute pour la rendre potable. Il s'agit de donner quelques généralités à propos de cette ressource naturelle qu'est l'eau, ainsi que les outils et moyens mis en œuvre pour son traitement. La chaîne de traitement d'eau potable la plus courante est de ce fait présentée. Les différents paramètres descripteurs de l'eau sont décrits. Enfin du chapitre nous allons donner une description détaillée des capteurs des paramètres descriptifs de l'eau.

### **1. Cycle de l'eau**

L'eau est très présente sur notre planète, les océans recouvrant près des trois quarts de la surface terrestre (70%). La totalité de l'eau sur terre représente un volume de 1,4 milliard de km<sup>3</sup>, sous forme liquide, solide ou gazeuse [3]. Entre les grands réservoirs d'eau de l'hydrosphère qui sont les mers et océans, les eaux continentales (superficielles et souterraines), l'atmosphère et la biosphère, l'échange d'eau est permanent et forme ce que l'on appelle le cycle externe de l'eau. Le moteur de ce cycle est l'énergie solaire qu'il rayonne, il active et maintient constamment les masses d'eau en mouvement (condensation, évaporation,...). Chauffée par le soleil, l'eau des océans, des mers, des rivières et des lacs s'évapore. A ce volume s'ajoute la vapeur d'eau issue de l'évapotranspiration des plantes. Toute cette vapeur d'eau monte dans l'atmosphère et se refroidit. Elle se condense alors et forme des gouttes ou des cristaux. Les nuages ainsi constitués se déplacent avec le vent. Ils laissent échapper leur contenu sous forme de précipitations (pluie, neige, grêle) ou de

brouillard. La plus grande partie de l'eau tombe directement dans les océans. Le reste atteint les continents, où 60% s'évapore à nouveau, seul 25% s'infiltrate dans la terre et alimente les nappes souterraines, enfin 15% ruisselle et rejoint les cours d'eau, pour retrouver ensuite l'océan, et le cycle recommence. L'évaporation, l'action du vent, la condensation, les précipitations, le ruissellement et l'infiltration, sont les 6 phases du cycle de l'eau (figure 1.1). Ce dernier se divise en deux parties intimement liées :

- Une partie atmosphérique qui concerne la circulation de l'eau dans l'atmosphère, sous forme de vapeur d'eau essentiellement.
- Une partie terrestre qui concerne l'écoulement de l'eau sur les continents, qu'il soit superficiel ou souterrain [4].

Cependant, la majeure partie de l'eau (97,23 %) est contenue dans les océans, et est salée, ce qui la rend relativement inutilisable par l'homme [5, 6]. L'eau douce n'en représente que 2,6 %. La moitié de cette eau douce est disponible pour l'usage humain, soit seulement 0,3 % du volume d'eau de la planète, soit 4 millions de km<sup>3</sup>.

Les eaux utilisées pour l'alimentation humaine qu'elles soient d'origine superficielle ou souterraine 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. Il faut insister sur le fait que le problème de l'eau n'est pas un problème de quantité, mais un problème de flux [7]. Lorsque la quantité d'eau est insuffisante, l'enjeu de la gestion des eaux est d'augmenter le débit disponible dans la zone du cycle utile aux activités humaines. Il s'agira dans tous les cas de prélever de l'eau dans l'une des portions du cycle et de ramener celle-ci à l'entrée de la zone utile, c'est-à-dire en entrée d'une usine de production (figure 1.2).

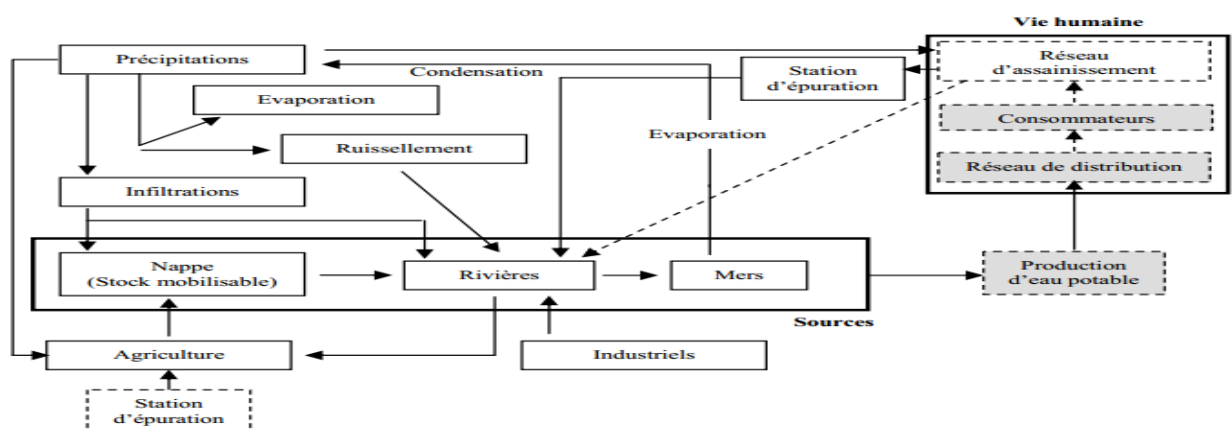
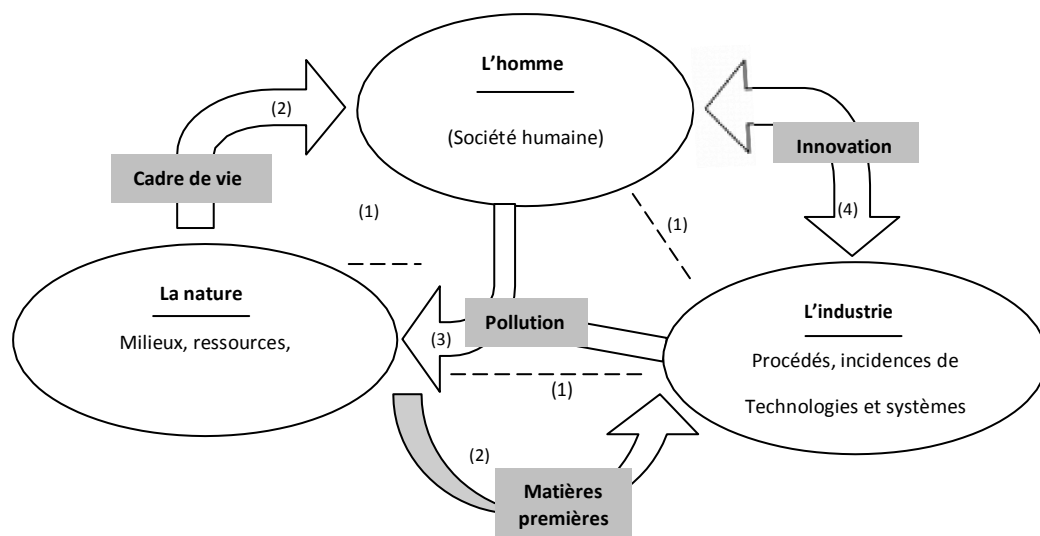


Fig.1.1. Cycle de l'eau.

Pour la gestion qualitative de l'eau, les processus de décision qui intégreront au mieux les trois « sphères » – l'Homme, la Nature, l'Industrie (figure 1.2), constitueront une avancée dans la prise en compte d'une durabilité à la fois écologique, économique et sociale [8]. L'eau du système, représentant l'unicité de la ressource de tout bassin fluvial, est répartie dans divers compartiments de l'hydro système qui sont naturellement reliés par le cycle de l'eau (précipitations, superficielles, souterraines), forment un système cohérent, qui intègre des éléments naturels et des éléments issus d'activités humaines. La persistance d'une pollution sera liée à la nature et au degré d'intensité de toutes les interrelations au sein de l'hydro-système. Toute dégradation de l'un des compartiments aura des répercussions sur les autres.

On distingue deux fonctions essentielles dans la plupart des unités d'exploitation :

- La production d'eau potable : qui consiste à mobiliser les ressources, à les traiter le cas échéant, puis à les transporter sur le lieu où elles seront distribuées.
- La distribution d'eau potable : qui débute à l'aval du château d'eau et dont le réseau s'étend jusqu' au consommateur.



(1) **interfaces** instrumentées pour analyses, mesures, modélisations,...

(2) **apports** (cadre de vie, matières premières,..)

(3) **pollution** (effets de comportements, d'activités,..)

(4) **innovation** et création de richesse (garantie de la prévention active)

Fig.1.2. Représentation schématique du champ de recherche « environnement développement » [8].

## 2. Ressources d'eau

L'eau potable est produite à partir d'eaux brutes : superficielles ou souterraines. Les eaux souterraines sont généralement de meilleure qualité que les eaux superficielles. Le choix de la ressource s'effectue en fonction des trois critères qui sont : la quantité, la qualité et la sécurité [9]. On prend en compte :

- La disponibilité des ressources : (y a-t-il une nappe ou un cours d'eau capable de fournir des débits nécessaires à la satisfaction des besoins ?).
- La qualité des ressources : il est évident qu'on utilise en premier lieu les ressources les moins polluées pour la production d'eau potable dans la mesure où tout incident dans ce domaine touche potentiellement une population importante.
- La sécurité de l'approvisionnement : il est nécessaire de prévoir une "substitution" en cas d'indisponibilité d'un point d'eau par maillage.

Les problèmes les plus fréquemment rencontrés dans les eaux brutes souterraines et superficielles sont montrés dans la figure 1.3 (a, b).

Pour les eaux souterraines, la présence de :	
Fer, Manganèse	Issus de la dissolution des roches traversées par les eaux d'infiltration
Nitrates	Issus de la nitrification naturelle des sols ainsi que des apports agricoles, déjections animales, engrais.
Produits phytosanitaires	Provenant des traitements agricoles, voies ferrées, zones urbaines, ...
Pollutions bactériennes	Qui peuvent provenir des eaux usées domestiques, des rejets hôpitaux, des élevages ou de certaines industries agro-alimentaires ; elles sont entraînées par les eaux de ruissellement et d'infiltration.

Fig.1.3, a. Problèmes fréquemment rencontrés dans les eaux souterraines

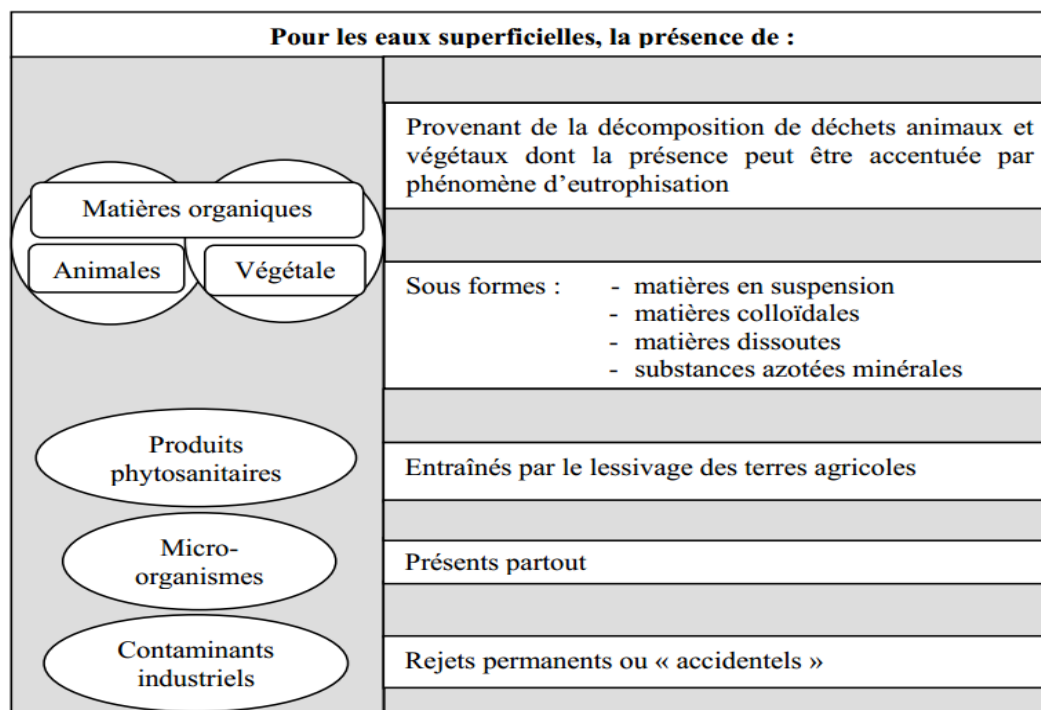


Fig.1.3, b. Problèmes fréquemment rencontrés dans les eaux superficielles.

### 3. Paramètres physico-chimiques mesurables

Ils sont essentiellement représentés par des indicateurs plus globaux comme la température, la conductivité, qui permet d'avoir une idée de la salinité de l'eau, et le titre alcalimétrique, qui permet d'apprécier la concentration de tous les carbonates et bicarbonates dans l'eau ou les sels minéraux (calcium, sodium, potassium, magnésium, sulfates...). Ce sont eux qui font l'identité de la base de l'eau.

▪ **Température** : une température élevée réduit la solubilité des sels et surtout des gaz dans l'eau et en particulier les teneurs en oxygène. Si la température de l'eau varie de 13 à 20°C, la concentration en oxygène chute de 13 %. Or, le rôle de l'oxygène est fondamental pour l'oxydation des déchets. Pour les températures basses, il y a risque de croissance accélérée de microorganismes, d'algues, entraînant des goûts et des odeurs désagréables ainsi qu'une augmentation de l'intensité de la couleur et de la turbidité [10]. La température est donc un paramètre important dans l'étude et la surveillance des eaux qu'elles soient souterraines ou superficielles ; les eaux souterraines gardent généralement une fraîcheur constante, mais la température des eaux de surface varie selon plusieurs facteurs, saisonniers et autres.

▪ **Dureté (ou titre hydrométrique)** : la dureté d'une eau correspond à la somme des concentrations en cations métalliques, excepté celles des métaux alcalins ( $\text{Na}^+$ ,  $\text{K}^+$ ) et  $\text{H}^+$  [11]. Elle est souvent due aux ions  $\text{Ca}^{2+}$  et  $\text{Mg}^{2+}$  (dureté temporaire), en grande partie combinés avec du bicarbonate, carbonate e, sulfates, chlorures, sulfates et anions d'acides

minéraux (dureté permanente). De plus, une eau dure est responsable du dépôt de calcaire dans les canalisations et contient beaucoup de minéraux dissous. Ainsi une eau qui contient peu de minéraux dissous est considérée comme douce. La présence de ces deux cations dans l'eau tend souvent à réduire la toxicité des métaux. La dureté se mesure en mg de  $\text{CaCO}_3$  par litre [12].

▪ **PH** : le potentiel hydrogène (ou pH) mesure l'activité chimique des ions hydrogènes ( $\text{H}^+$ ) en solution. Ce paramètre a une importance principale pour la qualité d'une eau donnée. Il indique si l'eau présente une acidité (ou alcalinité) ou une basicité trop élevée risquant de créer des conditions peu favorables à la vie au moyen d'une échelle universellement graduée de 0 à 14. Un pH inférieur à 7 correspond à une eau acide et un pH supérieur à 7 indique que l'eau est alcaline (ou basique). Un pH de 7 est donc dit neutre. Des pH faibles augmentent notamment le risque de présence de métaux sous une forme ionique toxique. Des pH élevés augmentent les concentrations d'ammoniac.

▪ **Alcalinité** : elle caractérise la possibilité qu'a une eau à maintenir son pH constant. Ainsi un ajout d'une petite quantité d'acide faible dans une eau pure provoque automatiquement une baisse sensible du pH. Une eau alcaline sera capable de neutraliser cet acide et donc de maintenir son pH plus ou moins constant. L'Alcalinité mesure la quantité de composés alcalins dans l'eau, tels que les carbonates ( $\text{CO}_3^{2-}$ ), bicarbonates et hydroxydes. Elle se détermine par acidimétrie et se caractérise par deux paramètres : le titre alcalimétrique (TA) qui correspond à la première neutralisation des ions carbonates et le titre alcalimétrique complet (TAC) qui correspond à la neutralisation de la totalité des espèces carboniques dissociées. L'alcalinité est habituellement donnée dans l'unité mEq/l (milliéquivalents par litre) [13].

▪ **Conductivité** : la conductivité est une mesure de la qualité de l'eau pouvant passer un courant électrique [5]. Il s'agit d'une mesure indirecte de la présence de matières inorganiques solides dissoutes tels que : le chlorure, le nitrate, le sulfate, le phosphate, le sodium, le magnésium, le calcium, le fer et l'aluminium, à l'aide d'un conductivimètre. La plupart de ces sels minéraux en solution sont de bons conducteurs. On obtient une valeur en micro-siemens par centimètre ( $\mu\text{s}/\text{cm}$ ) [5]. La conductivité ne permet pas de mesurer la minéralisation, mais plutôt d'évaluer approximativement la quantité des sels minéraux contenue dans l'eau. Cette quantité est exprimée en PPM : 1 PPM représente un milligramme de sels minéraux pour un litre d'eau. Ainsi, si la conductivité est à  $100 \mu\text{s}/\text{cm}$ , l'eau contient 50 PPM, soit 50 mg/l de sels minéraux. La présence de ces substances augmente la conductivité d'une masse d'eau. Les substances organiques comme l'huile,

l'alcool et le sucre ne conduit pas l'électricité très bien, et par conséquent induisent une faible conductivité de l'eau. Ainsi, plus la conductivité est basse, plus l'eau est « douce », c'est-à-dire pauvre en sels minéraux. Une forte concentration de solides dissous, cependant, peut réduire les concentrations d'oxygène dissous.

- **Oxygène dissous (OD) et % de saturation en oxygène** : les concentrations en oxygène dissous constituent, avec les valeurs de pH, l'un des plus importants paramètres de la qualité de l'eau [11]. La concentration en OD varie de manière journalière et saisonnière car elle dépend de nombreux facteurs tels que la pression partielle en oxygène de l'atmosphère, la température de l'eau, la salinité, la pénétration de la lumière et l'agitation de l'eau. Globalement, plus la concentration en OD n'est proche de la saturation, plus l'aptitude de la rivière à absorber la pollution est grande. Cette concentration peut être exprimée en mg d'O<sub>2</sub> par litre ou en % de saturation en oxygène [14]. Une faible teneur en OD provoque une augmentation de la solubilité des éléments toxiques qui se libèrent des sédiments. Des teneurs supérieures indiquant une eutrophisation du milieu, se traduisent par une activité photosynthétique intense.

- **Éléments en solution (éléments caractéristiques)** : il s'agit du sodium (Na<sup>+</sup>), du potassium (K<sup>+</sup>), de magnésium (Mg<sup>2+</sup>), des chlorures (Cl<sup>-</sup>), des nitrates (NO<sub>3</sub><sup>-</sup>) et des sulfates (SO<sub>4</sub><sup>2-</sup>). Ce sont les éléments majeurs, dont les concentrations dépendent de leurs solubilités dans l'eau. Une forte teneur en chlorure peut indiquer une pollution. Les eaux naturelles contiennent pratiquement des chlorures et des sulfates. Les eaux sont traitées au sulfate d'aluminium ajouté lors du traitement de coagulation, bien qu'une légère fraction d'ions SO<sub>4</sub><sup>2-</sup> soit entraînée par absorption avec le floc [15]. Les éléments en solution sont exprimés en mg/l.

#### 4. Chaîne de traitement de l'eau potable

La chaîne de traitement des eaux propres dans les usines se déroule comme ce qui illustré dans la figure 1.4

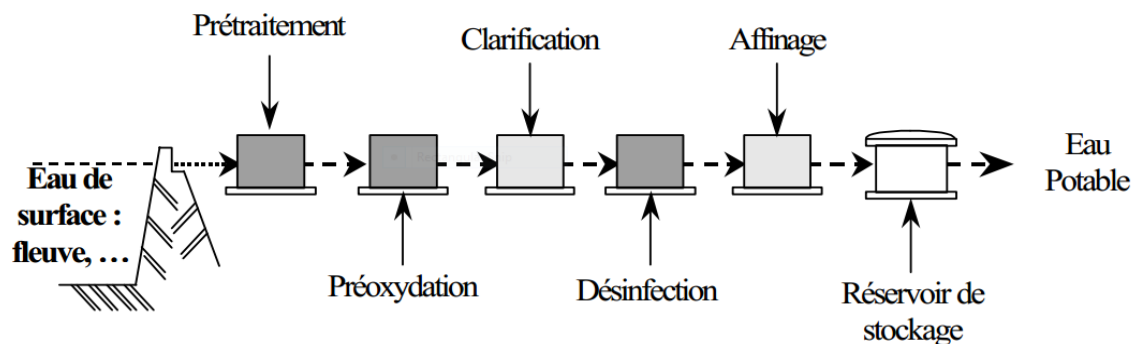


Fig.1.4. Chaîne de traitement

#### 4.1. 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 [1]. Dans le cas d'une eau potable, les prétraitements sont principalement de deux types :

- le dégrillage,
- le tamisage.

Le dégrillage, premier poste de traitement, permet de protéger les ouvrages aval 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, quant à lui, permet d'éliminer des objets plus fins que ceux éliminés par le dégrillage. Il s'agit de feuilles ou de morceaux de plastique par exemple [16].

#### 4.2. Pré-oxydation

A l'issue du prétraitement, on a une eau relativement propre mais qui contient encore des particules colloïdales en suspension et des matières organiques en solution. Celles-ci n'ont en elles-mêmes rien de dangereux. Il nous arrive souvent de consommer de l'eau en contenant : le thé, le café, le vin ou le lait. Ces produits sont des eaux chargées en matières organiques, mais on sait qu'elles s'oxydent spontanément en présence d'air. On va donc les détruire dans la mesure du possible par une pré-oxydation. Celle-ci peut être faite de trois façons différentes:

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

La pré-chloration, effectuée avant le procédé de clarification, s'est surtout développée dans les années 60. Moins en vogue dans les 80, elle tend à disparaître actuellement. Le chlore est le réactif le plus économique, mais il a comme inconvénient de former avec certains micropolluants des composés organochlorés du type chloroforme ou des composés complexes avec les phénols du type chlorophénol dont le goût et l'odeur sont désagréables.

On préfère donc parfois utiliser le dioxyde de chlore qui coûte plus cher mais n'a pas les inconvénients de l'oxydation par le chlore cités ci-dessus. Ce type de traitement est cependant

réservé à des cas spécifiques. En effet, l'utilisation du dioxyde de chlore présente, lui aussi, des inconvénients non négligeables. Premièrement, il est sans effet sur l'ammonium. Deuxièmement, le dioxyde de chlore dans l'eau se décompose à la lumière, ce qui entraîne une augmentation importante du taux de traitement à appliquer en période d'ensoleillement. En conclusion, le dioxyde de chlore est un oxydant plus puissant que le chlore mais il ne s'agit pas d'une solution économique. Il reste très peu utilisé en pré-oxydation mais représente une alternative intéressante à l'utilisation du chlore lorsque celui-ci entraîne des problèmes de qualité d'eau.

Enfin, depuis quinze à vingt ans, on utilise comme pré-oxydant l'ozone, qui non seulement a l'avantage de détruire les matières organiques en cassant les chaînes moléculaires existantes, mais également a une propriété virulicide très intéressante, propriété que n'a pas le chlore.

Généralement utilisée en désinfection finale, cette technique peut être mise en œuvre en pré-oxydation [16].

Elle peut aussi être employée pour l'amélioration de la clarification. L'un des avantages d'une pré-ozonation est l'oxydation des matières organiques, et une élimination de la couleur plus importante. Un autre avantage est la diminution du taux de traitement (taux de coagulation) dans le procédé de clarification. En conclusion, la pré-ozonation est une solution de substitution à la pré-chloration. On évite ainsi les problèmes liés aux sous-produits de la chloration. Néanmoins, ce procédé ne résout pas tous les problèmes car certaines algues résistent à l'ozone. De plus, son coût reste plus élevé que celui du chlore [16].

### **4.3. Clarification**

La clarification est l'ensemble des opérations permettant d'éliminer les matières en suspension MES (minérales et organiques) d'une eau brute ainsi que des matières organiques dissoutes. Suivant les concentrations de l'un et de l'autre des différents polluants, on peut être amené à pratiquer des opérations de plus en plus complexes qui vont de la simple filtration avec ou sans réactif jusqu'à la coagulation – floculation – décantation ou flottation – filtration. La clarification comprend les opérations suivantes :

- Coagulation
- Floculation
- Filtration

La coagulation est l'une des étapes les plus importantes dans le traitement des eaux de surface. 90% des usines de production d'eau potable sont concernées. La difficulté principale est de déterminer la quantité optimale de réactif à injecter en fonction des caractéristiques de l'eau brute.

Un mauvais contrôle de ce procédé peut entraîner une augmentation importante des coûts de fonctionnement et le non-respect des objectifs de qualité en sortie. Cette opération a également une grande influence sur les opérations de décantation et de filtration ultérieures. En revanche, un contrôle efficace peut réduire les coûts de main d'œuvre et de réactifs et améliorer la conformité de la qualité de l'eau traitée.

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

#### 4.4. Oxydation-Désinfection

La désinfection est l'étape ultime du traitement de l'eau de consommation avant distribution. Elle permet d'éliminer tous les micro-organismes pathogènes présents dans l'eau. Il peut cependant subsister dans l'eau quelques germes banals, car la désinfection n'est pas une stérilisation.

Le principe de la désinfection est de mettre en contact un désinfectant à une certaine concentration pendant un certain temps avec une eau supposée contaminée.

Cette définition fait apparaître trois notions importantes : les désinfectants, le temps de contact et la concentration résiduelle en désinfectant. Une bonne désinfection via les réactifs oxydants demande la combinaison d'une concentration  $C$  avec un temps de contact  $T$  ; c'est le facteur  $C \cdot T$  (mg.min/L). Cette valeur varie avec les microorganismes concernés, le type de désinfectant et la température.

Les quatre principaux désinfectants utilisés en production d'eau potable sont les suivants :

- Le chlore
- Le dioxyde de chlore
- L'ozone
- Le rayonnement UV

La concentration en oxydant est pratiquement le seul paramètre sur lequel l'opérateur peut intervenir. Il faut retenir que l'efficacité de la désinfection dépend, en partie, du suivi de la concentration en oxydant. L'évolution de la concentration en oxydant est liée à la demande en oxydant de l'eau. Cette demande dépend de la qualité de l'eau, du pH, des températures (différentes entre été et hiver), des matières organiques, et de la concentration en ammoniacque [17].

#### 4.5. Affinage

Le traitement final traite de la mise à l'équilibre calco-carbonique. L'eau suit un cycle naturel dans lequel les éléments chimiques qu'elle contient évoluent.

L'eau de pluie contient naturellement du dioxyde de carbone ( $\text{CO}_2$ ). Quand celle-ci traverse les couches d'humus, riches en acides, elle peut s'enrichir fortement en  $\text{CO}_2$ . Lors de sa pénétration dans un sol calcaire, c'est-à-dire riche en carbonate de calcium ( $\text{CaCO}_3$ ), elle se charge en calcium  $\text{Ca}^{2+}$  et en ions  $\text{HCO}_3^-$ . En fait, le calcium est dissous par l'eau chargée en  $\text{CO}_2$ . On dit qu'elle est entartrante ou incrustante. En revanche, quand l'eau de pluie traverse une roche pauvre en calcium (région granitique), elle reste très chargée en  $\text{CO}_2$  dissous.

Cette eau est, en générale, acide. On dit qu'elle est agressive.

Il y a typiquement deux problèmes distincts : corriger une eau agressive et corriger une eau incrustante. La correction d'une eau agressive peut s'effectuer de plusieurs façons. Premièrement, on peut éliminer le  $\text{CO}_2$  par aération. Du fait de l'élimination du  $\text{CO}_2$ , le pH augmente et se rapproche du pH d'équilibre. Deuxièmement, on peut ajouter une base à l'eau. L'ajout de base permet d'augmenter le pH et d'atteindre le pH d'équilibre. La correction d'une eau incrustante peut se faire soit par traitement direct soit en réduisant le potentiel d'entartrage par décarbonatation. Le traitement direct correspond à un ajout d'acide [17].

#### 5. les capteurs

La connaissance scientifique s'est développée par un double effort : D'une part, la réflexion sur la nature d'interactions entre différentes grandeurs décrivant les phénomènes physiques. D'autre part, l'expérimentation qui repose sur la mesure des grandeurs physiques et qui, en leur associant une valeur numérique, permet de définir quantitativement les propriétés des objets, de vérifier numériquement les lois physiques ou d'en établir empiriquement la forme. Aussi, la science cherche toujours à exprimer mathématiquement, dans des théories cohérentes, les lois fondamentales régissant les grandeurs physiques. De plus, la connaissance des lois de la nature a permis de rationaliser la technique utilisée par la science et devenir une science de la réalisation. La mesure y joue alors un rôle capital. L'électronique offre à cet égard des moyens divers et puissants. Pour assurer une meilleure détection et exploitation des grandeurs physiques il est alors indispensable de les transformer et les traiter sous forme électrique. C'est donc, le rôle du capteur d'assurer cette duplication de l'information en la transférant, au point même où se fait la mesure, de la grandeur physique (non électrique) qui lui est propre, sur une grandeur électrique.

### 5.1. Définitions et caractéristiques générales

Une grandeur physique (déplacement, température, pression, etc.) est désignée comme étant le '**mesurande**'. l'ensemble des opérations expérimentales servant à la connaissance de ce mesurande (généralement électrique) constitue son-mesurage. Un élément permettant de faire cette dernière est le capteur. Le capteur (premier élément d'une chaîne de mesure) est un dispositif qui transforme la grandeur physique indispensable aux objectifs des mesures, en grandeur électrique facilement mesurable et enregistrable tel que :  $s = F(m)$  (1)

Avec  $m$  : mesurande,  $s$  : sortie du capteur

La relation ci-dessus résulte dans sa forme théorique des lois physiques qui régissent le fonctionnement du capteur .

Pour des raisons de facilité d'exploitation on s'efforce de réaliser le capteur en sorte qu'il établisse une relation linéaire entre les variations  $\Delta s$  de la grandeur de sortie et celles  $\Delta m$  de la grandeur d'entrée .

$$\Delta S = S. \Delta m \quad (2)$$

$S$  : représente la sensibilité du capteur.

Un des problèmes importants dans la conception et l'utilisation d'un capteur est la constance de sa sensibilité «  $S$  » qui doit dépendre aussi peu que possible :

- De la valeur de  $m$  (linéarité) et de sa fréquence de variation (bande passante).
- Du temps (vieillessement).
- De l'action des grandeurs physiques de son environnement qui ne sont pas l'objet de la mesure et que l'on désigne comme grandeurs d'influence.

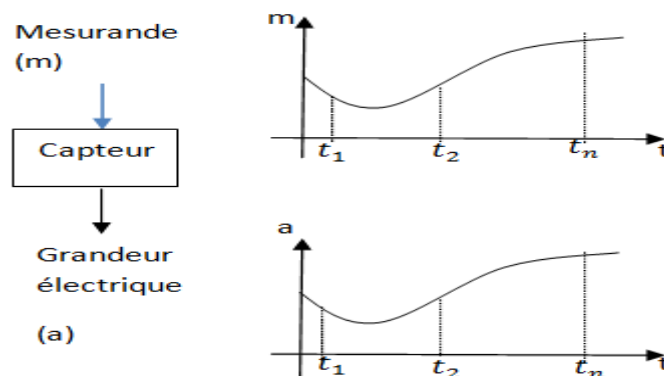


Fig.1. 5 Exemple d'évolution d'un mesurande «  $m$  » et de la réponse «  $s$  » correspondante du capteur.

Selon la grandeur de sortie, un capteur est soit actif c-à-d « S » étant une charge, une tension ou un courant, soit capteur passif c-à-d « S » étant alors une résistance, une inductance ou une capacité .

### 5.2. Capteurs actifs

Dans son principe un capteur actif est généralement fondé sur un effet physique qui assure la conversion en énergie électrique de la forme d'énergie propre au mesurande: énergie thermique, mécanique ou de rayonnement. Les plus importants parmi ces effets sont regroupés dans le Tableau 1.1 et dans la figure1.6[18].

Tab. 1. 1 Capteurs actifs : principes physiques de base

Mésurande	Effet utilisé	Grandeur de sortie
Température	Thermoélectricité	Tension
Flux de rayonnement optique	Pyroélectricité ; Photoémission Effet photovoltaïque, Effet photo électromagnétique	Charge, Courant, Tension Tension
Vitesse	Induction électromagnétique	Tension
Position (aimant)	Effet Hall	Tension

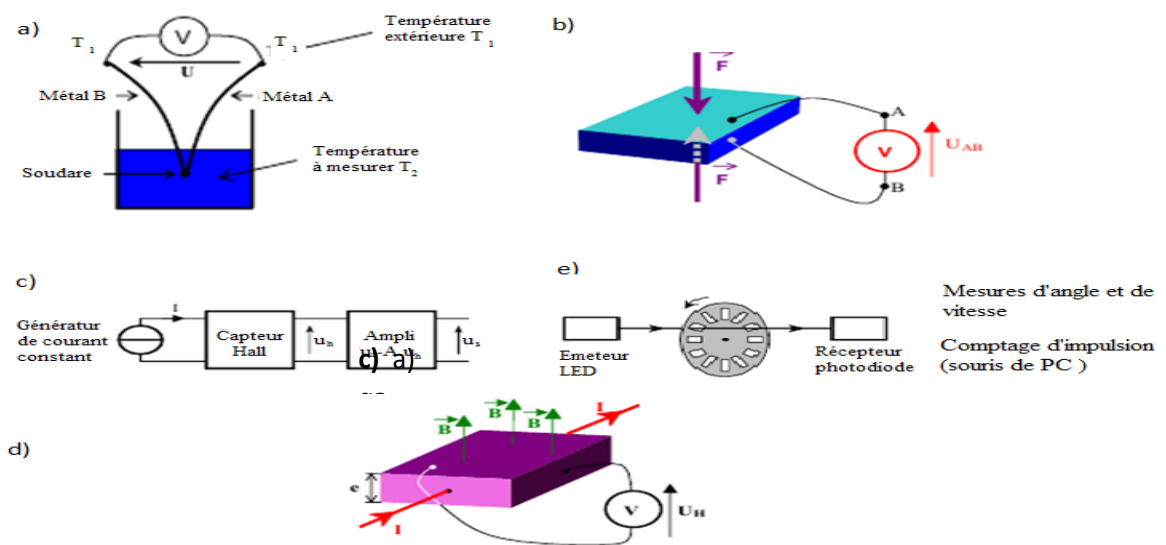


Fig. 1. 6. Exemples d'application de Capteurs actifs :

- (a) : thermoélectricité,
- (b) : pyroélectricité,
- (c) : induction électromagnétique,
- (d) : photoélectricité,
- (e) : Effet Hall [19].

### 5.3. Capteurs passifs

Il s'agit d'impédances dont l'un des paramètres déterminants est sensible au mesurande. Dans l'expression littérale d'une impédance sont présents des termes liés d'une part à sa géométrie et à ses dimensions, d'autre part aux propriétés électriques des matériaux : résistivité  $\rho$ , perméabilité magnétique  $\mu$ . Constante diélectrique  $\epsilon$ .

Le Tableau 1.2 donne un aperçu des divers mesurandes susceptibles de modifier les propriétés électriques de matériaux employés pour la réalisation de capteurs.

Tab. 1. 2 .Capteurs passifs : principes physiques et matériaux

Mesurande	Caractéristique électrique Sensible	Types de matériaux Utilisés
Température	Résistivité	Métaux : platine, nickel, cuivre.
Très basse température	Constante diélectrique	Semi-conducteurs. Verres.
Flux de rayonnement optique	Résistivité	Semi-conducteurs.
Déformation	Résistivité Perméabilité Magnétique	Alliages de nickel silicium dopé. Alliages ferromagnétiques.
Position (aimant)	Résistivité	Matériaux magnéto-résistants: bismuth, antimoniure d'indium
Humidité	Résistivité Constante diélectrique	Chlorure de lithium. Alumine ; polymères.
Niveau	Constante diélectrique	Liquides isolants.

### 5.4. Transformation de la grandeur physique

Une Grandeur physique subira deux types de transformations :

- **Transformation directe**

Les variables physiques sont en général traduisibles directement en variables électriques.

Exemple : le couple thermoélectrique ou certaines sondes de mesure de pH.

- **Transformation indirecte**

Les variables mécaniques et certaines variables physiques nécessitent une transformation préalable permettant la traduction en grandeur électrique.

Exemple : Jauge de Contrainte[20].

## 5.5 Qualité des capteurs

Nous allons maintenant aborder les problèmes liés à la qualité et à l'utilisation en continu de ces différents capteurs. En effet, pour que le fonctionnement de l'ensemble d'une boucle de mesure soit correct, il est essentiel de s'assurer de la compatibilité de chacun des instruments de la boucle et en particulier du capteur. L'information ainsi délivrée, surtout si elle est utilisée dans une régulation automatique ou comme entrée d'un modèle, doit être la plus représentative possible de la valeur vraie du paramètre mesuré et être très fiable [17].

## 5.6. Sensibilité, précision, gamme de mesure

De nombreux facteurs conditionnent l'écart entre la valeur du paramètre mesuré et l'information délivrée. Le premier facteur est la précision du capteur. Celle-ci, exprimée en pourcentage, est le quotient de l'incertitude de la valeur obtenue par l'étendue de mesure pour des conditions de mesure données. La précision du capteur est fonction du processus de mesure mais aussi des corrections annexes qui y sont apportées. Une bonne précision finale dépend d'une bonne corrélation entre une caractéristique ou un phénomène annexe, pouvant être mesuré avec une bonne précision, et le phénomène à mesurer. Un autre facteur peut être l'existence d'erreurs systématiques dues à un étalonnage incorrect ou trop peu fréquent du capteur. Les erreurs accidentelles 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. Celle-ci est la valeur minimum du paramètre à mesurer en dessous duquel l'appareil ne réagit pas. La sensibilité en fonctionnement est la plus petite variation du paramètre mesuré décelable par la mesure. Elle n'est pas nécessairement constante dans toute la gamme de mesure. Il faut enfin tenir compte de la gamme de mesure du capteur, qui correspond aux valeurs de seuils au delà desquels la précision et la sensibilité du capteur se dégradent [17].

### 5.7. Fiabilité, environnement et dérive

La fiabilité est définie comme la capacité du capteur à fonctionner correctement, c'est-à-dire à fournir des données avec la précision annoncée. Elle dépend naturellement de la qualité de conception du matériel qui doit être robuste et avoir été entretenu conformément aux spécifications du constructeur. Mais elle dépend également de son adaptation à l'environnement dans lequel il va se trouver. Les contraintes des capteurs concernant la gestion de l'eau sont principalement l'humidité et la nature de l'eau. L'humidité peut provoquer de la condensation dans les boîtiers du matériel. Ceux-ci doivent être étanches, des submersions étant toujours possibles, et doivent comporter des dispositifs éliminant la condensation. Cette atmosphère humide peut également provoquer des courts circuits au niveau des câbles de jonction ou d'alimentation. La nature de l'eau, notamment celle des rivières, peut perturber les capteurs immergés avec des dépôts en modifiant les réactions. C'est en particulier le cas de nombreuses sondes dont le nettoyage doit être effectué très régulièrement car ces dépôts provoquent une dérive du capteur. C'est le principal défaut de ce type de capteur dont la surveillance doit être constante, les dispositifs de nettoyage automatique sous forme de brosses ou de rétro-lavage de la sonde n'étant pas toujours efficaces.

C'est l'une des raisons pour lesquelles certains de ces capteurs opèrent à partir de bacs extérieurs auxquels on amène l'eau à analyser, notamment dans les stations d'analyse ou d'alerte en rivière. Le problème est alors celui de la prise d'eau, de l'entretien de la crépine d'aspiration et de la bonne représentativité de l'eau dans le bac.

En conclusion, pour tirer pleinement parti des avantages des capteurs de mesure et de l'instrumentation associée, il est indispensable d'accepter certaines contraintes telles que le nettoyage des sondes de mesures, l'étalonnage régulier, etc. Malgré ces précautions, certains facteurs peuvent encore perturber l'information délivrée par les capteurs. Il sera donc indispensable d'avoir une phase de validation de ces informations avant toute utilisation dans une boucle de régulation ou en entrée d'un modèle [17].

### 5.8. Mesure des paramètres à l'aide des capteurs physiques

Le turbidimètre mesure 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. Un faisceau lumineux vient toucher la surface sous une incidence telle que ni lui-même ni le faisceau réfléchi ne peuvent impressionner une cellule photorésistante placée sensiblement perpendiculairement au faisceau incident. Par contre, la

lumière diffusée par les particules en suspension vient modifier d'autant plus l'éclairement de la cellule que leur nombre est élevé, ce qui permet d'obtenir la mesure de la turbidité de cette eau.

Selon l'appareil de mesure, les gammes de mesure s'étendent de 0 à 5 000 unités néphéométriques de la turbidité (NTU).

Le principe mis en œuvre pour la mesure de la conductivité et de la résistivité, est simple puisqu'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, pour éviter les phénomènes de polarisation. La résistivité d'une eau étant fonction du degré de dissociation des molécules dissoutes, la plupart des appareils comportent une compensation automatique de température pour ramener la valeur de la mesure à une température de référence donnée. Industriellement, la mesure du pH se fait toujours par potentiométrie à l'aide de deux électrodes : une électrode de référence et une électrode de mesure. L'électrode de référence est plongée dans une solution de concentration constante en ions hydrogène. Une cloison, laissant passer le courant électrique, sépare la solution de référence de celle dont on veut mesurer le pH et dans laquelle est plongée l'électrode de mesure. Une tension, fonction linéaire de la concentration en ions hydrogène de la solution, apparaît alors aux bornes des électrodes. 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 [17].

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. 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. On mesure ainsi la concentration en chlore, ozone, 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 [17].

**Conclusion**

Ce premier chapitre a servi d'introduction au domaine de contrôle et de surveillance des eaux potables. Les différentes étapes d'une chaîne de traitement sont présentées. Les paramètres physico-chimiques ainsi que les capteurs utilisés comme source d'information ayant trait à la qualité de l'eau ont été particulièrement décrits.

Nous allons parler dans le chapitre suivant des techniques d'intelligence industrielle artificielle utilisées dans le domaine du traitement des eaux propres et précisément les réseaux des neurones artificiels(RNA) et nous allons décrire les deux réseaux de neurones usuels (PMC et RBF).

# Chapitre II

## Les réseaux de neurones artificiels

### Introduction

Le cerveau humain est capable de s'adapter, d'apprendre et de décider, et c'est sur ce fait que des chercheurs se sont intéressés à comprendre son principe de fonctionnement et de pouvoir l'appliquer au domaine de l'informatique. C'est ainsi que dans les années cinquante on formalisa le neurone en un modèle mathématique à partir du modèle biologique.

Nous allons présenter dans ce chapitre une étude générale sur les réseaux de neurones artificiels, le passage du modèle biologique au modèle artificiel, le protocole d'apprentissage, les différentes architectures de ces réseaux, ainsi que les réseaux de neurones bouclés et non bouclés. Nous allons présenter deux exemples usuels des RNAs tels que : les PMC et RBF. Nous allons finaliser ce chapitre par les principales avantages et inconvénients des RNAs.

### 1. Généralité

L'origine des réseaux de neurones vient de l'essai de modélisation mathématique du cerveau humain. Les premiers travaux datent de 1943 et sont l'œuvre de MM. Mac Culloch et Pitts. Ils supposent que l'impulsion nerveuse est le résultat d'un calcul simple effectué par chaque neurone et que la pensée née grâce à l'effet collectif d'un réseau de neurones interconnectés (figure 2.1). Ils ont connu des débuts prometteurs vers la fin des années 50, mais le manque d'approfondissement de la théorie a gelé ces travaux jusqu'aux années 80.

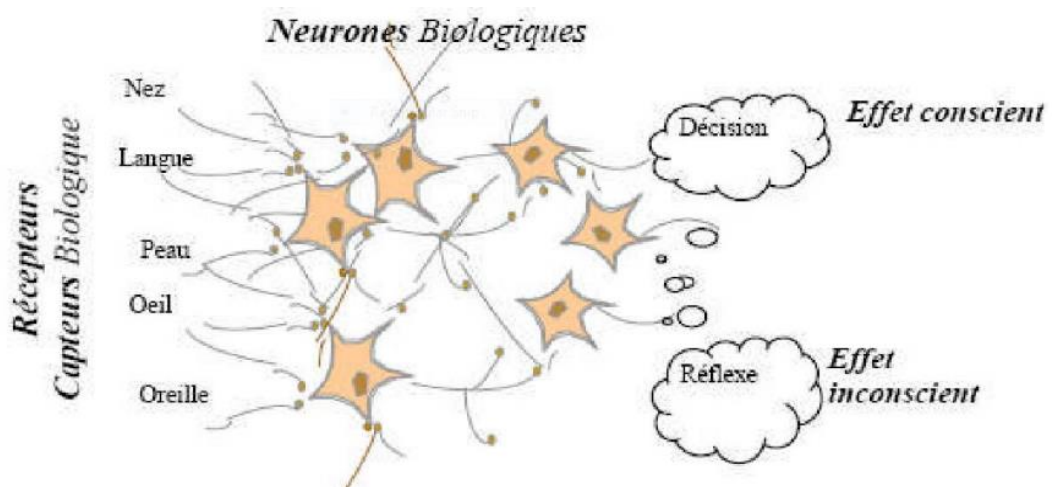


Fig.2.1. Structure d'un réseau de neurones biologiques

Les réseaux de neurones forment une famille de fonctions non linéaires, permettant de construire, par apprentissage, une très large classe de modèles et de contrôleurs. Un réseau de neurones est un système d'opérateurs non linéaires interconnectés, recevant des signaux de l'extérieur par ses entrées, et délivrant des signaux de sortie, qui sont en fait les activités de certains neurones [21].

## 2. Eléments de bases

### 2.1. Le modèle neurophysiologique

L'élément de base du système nerveux central est le neurone. Le cerveau se compose d'environ mille milliards de neurones, avec 1000 à 10000 synapses (connexions) par neurone.

Le neurone est une cellule composée d'un corps cellulaire et d'un noyau (Figure 2.2). Le corps cellulaire se ramifie pour former ce que l'on nomme les dendrites. Celles-ci sont parfois si nombreuses que l'on parle alors de chevelure dendritique ou d'arborisation dendritique. C'est par les dendrites que l'information est acheminée de l'extérieur vers le soma (corps du neurone). L'information est traitée alors par le corps cellulaire. Si le potentiel d'action dépasse un certain seuil, le corps cellulaire répond par un stimulus (Figure 2.3, a). Le signal transmis par le neurone chemine ensuite le long de l'axone (unique) pour être transmis aux autres neurones. La transmission entre deux neurones n'est pas directe. En fait, il existe un espace intercellulaire de quelques dizaines d'Angströms entre l'axone du neurone afférent et les dendrites du neurone efférent. La jonction entre deux neurones est appelée synapse (Figure 2.3, b) [22].

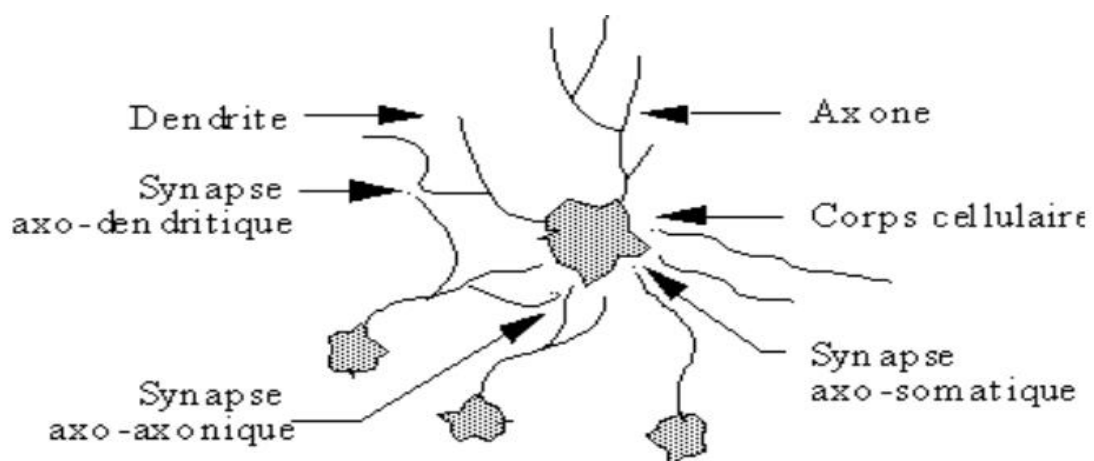


Fig.2.2. Un neurone avec son arborisation dendritique

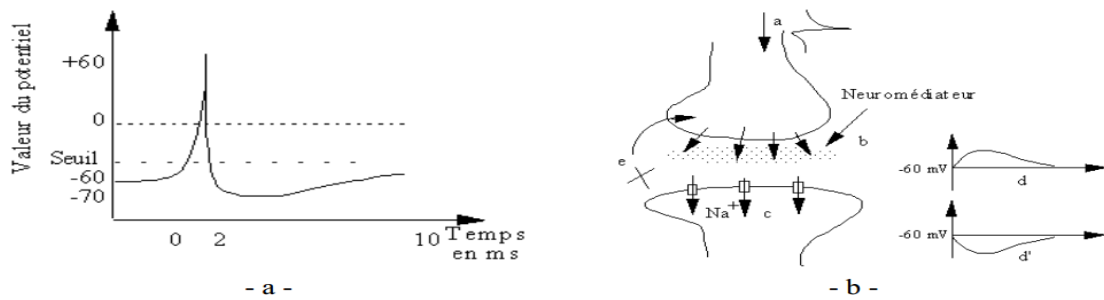


Fig.2.3, a– Un potentiel d'action.

Fig.2.3,b– Fonctionnement au niveau synaptique

a) Arrivée d'un potentiel d'action.

b) Libération du neurotransmetteur dans l'espace synaptique.

c) Ouvertures des canaux ioniques dues au neurotransmetteur.

d) Génération d'un potentiel évoqué excitateur.

d') Génération d'un potentiel évoqué inhibiteur. Les synapses inhibitrices empêchent la génération de potentiel d'action.

e) Fermeture des canaux, élimination du neurotransmetteur [22].

## 2.2. Le modèle mathématique

Les réseaux de neurones biologiques réalisent facilement un certain nombre d'applications telles que la reconnaissance des formes, le traitement du signal, l'apprentissage par l'exemple, la mémorisation et la généralisation. C'est à partir de l'hypothèse que le comportement intelligent émerge de la structure et du comportement des éléments de base du cerveau, que les réseaux de neurones artificiels se sont développés. La figure 2.4 montre la structure d'un neurone artificiel. Chaque neurone artificiel est un processeur élémentaire. Il reçoit un nombre variable d'entrées en provenance de neurones amont. A chacune de ces entrées est associé un poids  $w$ , abréviation de weight (poids en anglais), représentatif de la force de la connexion [22].

Chaque processeur élémentaire est doté d'une sortie unique, qui se ramifie ensuite pour alimenter un nombre variable de neurones aval. A chaque connexion est associé un poids.

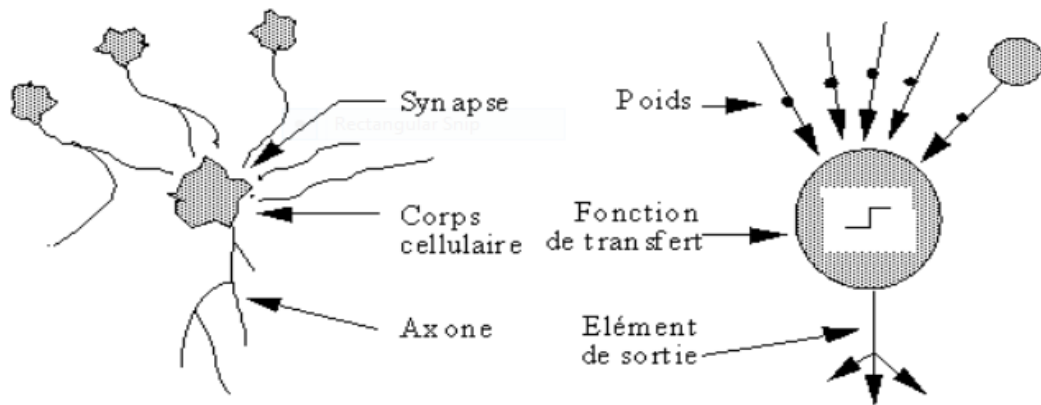


Fig.2.4. Mise en correspondance du neurone biologique /et du neurone artificiel

Par analogie avec le neurone biologique, le comportement du neurone artificiel se compose de deux phases (Figure 2.5).

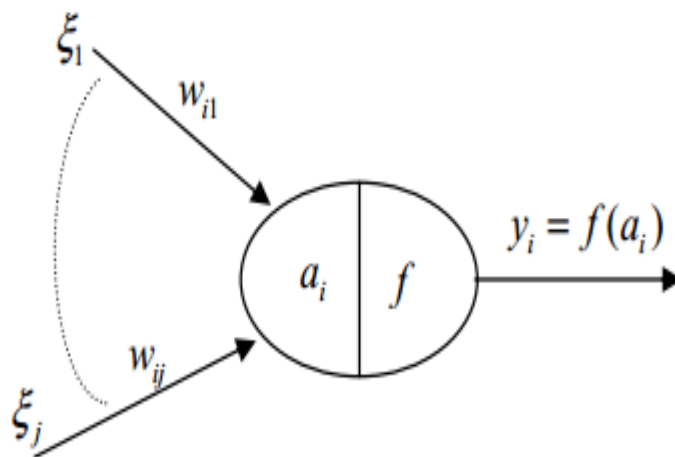


Fig.2.5. Neurone formel

La première phase, appelée activation, représente le calcul de la somme pondérée des entrées selon l'expression suivante :

$$\alpha_i = \sum_j w_{ij} \xi_j \quad (2.1)$$

A partir de cette valeur, une fonction de transfert calcule la valeur de l'état du neurone. C'est cette valeur qui sera transmise aux neurones aval. Il existe de nombreuses formes possibles pour la fonction d'activation.

### 2.3. Fonctions d'activations

Les fonctions d'activations représentent généralement certaines formes de non linéarité. L'une des formes de non linéarité la plus simple, et qui est appropriée aux réseaux discret, est la fonction signe (figure 2.6,a). Une autre variante de ce type des

non linéarité est la fonction de Heaviside (figure 2.6,b). Pour la majorité des algorithmes d'apprentissage il est nécessaire d'utiliser des fonctions sigmoïdes différentiables, telles que la fonction sigmoïde unipolaire (figure 2.6,c) et la fonction sigmoïde bipolaire (figure 2.6,d). La classe, la plus utilisée des fonctions d'activation, dans le domaine de la modélisation et de la commande des systèmes non linéaires est la fonction sigmoïde bipolaire.

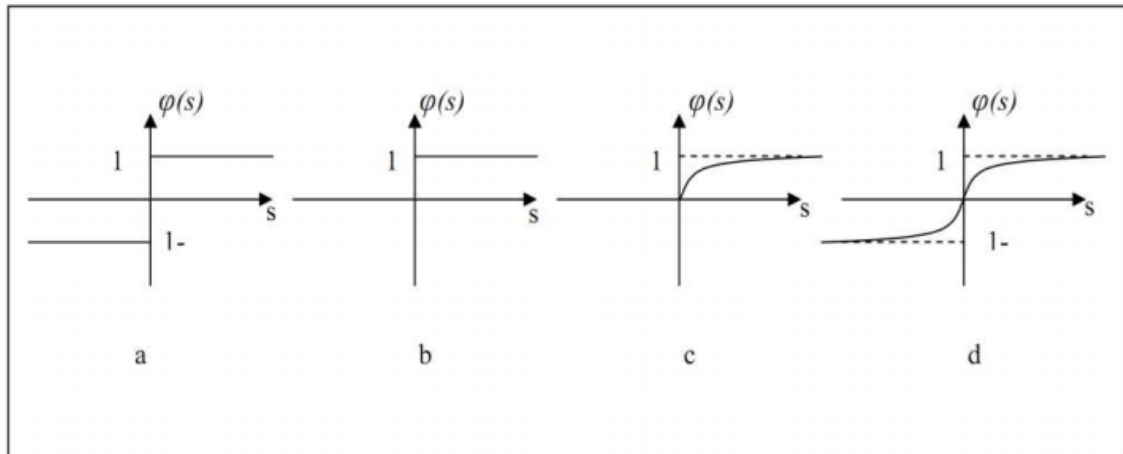


Fig.2.6.Différentes fonctions d'activation

Donc un réseau de neurone est un ensemble d'éléments de traitement de l'information, avec une topologie spécifique d'interconnexions (architecteur du réseau) entre ces éléments et une loi d'apprentissage pour adapter les poids de connexions (poids synaptiques), il est caractérisé par un parallélisme à gain très fin et à forte connectivité. Nous entendons par là que dans un réseau de neurones donné, l'information est traitée par grand nombre de processeurs élémentaires très simples, chacun étant relié à d'autres processus. Ce processus très simple est un neurone formel désigné ainsi son fonctionnement s'inspire d'une modélisation des cellules biologiques. Ils sont dotés de deux propriétés importantes qui sont à l'origine de leur intérêt pratique des domaines très divers :

- Capacité d'adaptation ou d'apprentissage qui permet au réseau de tenir compte des nouvelles contraintes ou de nouvelles données du monde extérieur,
- Capacité de généralisation qui est son aptitude de donner une réponse satisfaisante à une entrée qui ne fait pas partie des exemples à partir desquels il apprend [21].

### 3. Architectures des réseaux de neurones

On distingue deux structures de réseau, en fonction du graphe de leurs connexions, c'est-à-dire du graphe dont les nœuds sont les neurones et les arêtes les «connexions» entre ceux-ci :

Les réseaux de neurones statiques (ou acycliques, ou non bouclés).

Les réseaux de neurones dynamiques (ou récurrents, ou bouclés) [23].

#### 3.1. Les réseaux de neurones non bouclés

Un réseau de neurones non bouclé réalise une (ou plusieurs) fonction algébrique de ses entrées par composition des fonctions réalisées par chacun de ses neurones. Dans un tel réseau (figure 2.7), le flux d'information circule des entrées vers les sorties sans retour en arrière. Si l'on représente le réseau comme un graphe dont les nœuds sont les neurones et les arêtes les « connexions » entre ceux-ci, le graphe d'un réseau non bouclé est acyclique.

Tout neurone dont la sortie est une sortie du réseau est appelé « neurone de sortie».

Les autres, qui effectuent des calculs intermédiaires, sont des « neurones cachés».

Il existe deux types de réseaux de neurones : les réseaux complètement connectés et les réseaux à couche. Le réseau de neurones à une couche cachée et une sortie linéaire est un cas particulier de ce dernier type.

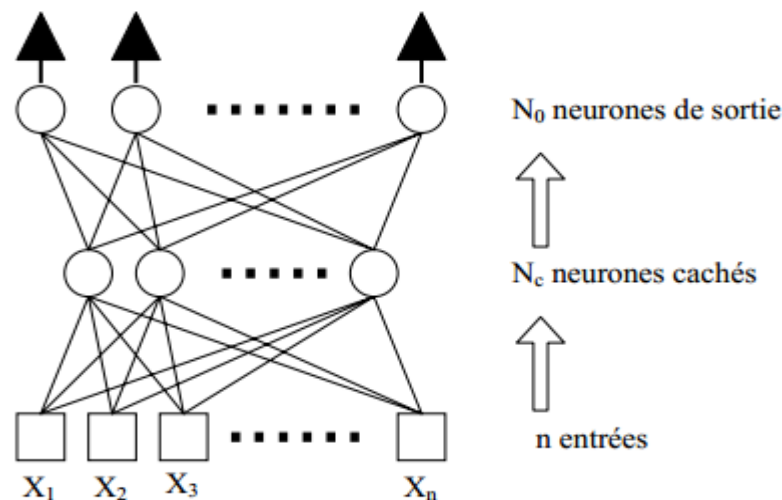


Fig.2.7. Réseau de neurones à n entrées, une couche de  $N_c$  neurones cachés et  $N_0$  neurones de sortie [23].

### 3.1.1. Les réseaux de neurones complètement connectés

Dans un réseau complètement connecté, les entrées puis les neurones (cachés et de sortie) sont numérotés, et, pour chaque neurone :

Ses entrées sont toutes les entrées du réseau ainsi que les sorties des neurones de numéro inférieur.

Sortie est connectée aux entrées de tous les neurones de numéro supérieur [23].

### 3.1.2. Les réseaux de neurones à couches

Dans une architecture de réseaux à couches, les neurones cachés sont organisés en couches, les neurones d'une même couche n'étant pas connectés entre eux. De plus les connexions entre deux couches de neurones non consécutives sont éliminées.

Une telle architecture est historiquement très utilisée, surtout en raison de sa pertinence en classification [23].

#### Remarque :

Dans un réseau de neurones non bouclé, le temps ne joue aucun rôle fonctionnel : si les entrées sont constantes, les sorties le sont également. Le temps nécessaire pour le calcul de la fonction réalisée par chaque neurone est négligeable et on peut considérer ce calcul comme instantané.

Pour cette raison, les réseaux non bouclés sont souvent appelés « réseaux statiques », par opposition aux réseaux bouclés ou « dynamiques ». Ils sont utilisés en classification, reconnaissance des formes (caractères, parole, ...), en prédiction [23].

### 3.2. Les réseaux de neurones bouclés

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

Ainsi, à chaque connexion d'un réseau de neurones bouclé (ou à chaque arête de son graphe) est attaché, outre un poids comme pour les réseaux non bouclés, un retard, multiple entier (éventuellement nul) de l'unité de temps choisie. Une grandeur, à un instant donné, ne

pouvant pas être fonction de sa propre valeur au même instant, tout cycle du graphe du réseau doit avoir un retard non nul.

Les connexions récurrentes ramènent l'information en arrière par rapport au sens de propagation défini dans un réseau multicouche. Ces connexions sont le plus souvent locales.

Pour éliminer le problème de la détermination de l'état du réseau par bouclage, on introduit sur chaque connexion « en retour » un retard qui permet de conserver le mode de fonctionnement séquentiel du réseau (figure 2.8)

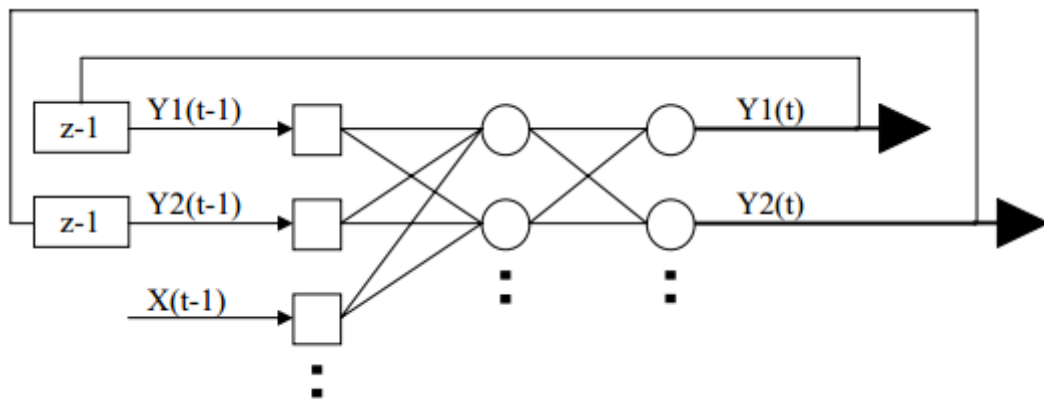


Fig.2.8. Réseau de neurone bouclé

Le graphe des connexions de réseaux récurrents est cyclique. Ces réseaux sont décrits par un système d'équations aux différences [23].

### Forme canonique des réseaux récurrents :

Il a été montré que tout réseau bouclé peut être mis sous une forme particulière, appelée forme canonique, qui est la représentation d'état minimale de la fonction réalisée par ce réseau. Cette forme canonique est constituée d'un graphe acyclique, et de connexions à retard unité reliant certaines sorties de ce graphe à ses entrées. La fonction réalisée par un réseau de neurones ayant cette structure particulière est décrite par les équations aux différences suivantes :

$$x(k+1) = \Phi(x(k), u(k+1)) \quad (2.2)$$

$$y(k+1) = \Psi(x(k+1), u(k+1)) \quad (2.3)$$

Où  $x(k)$  est le vecteur d'état à l'instant  $k$ ,  $u(k)$  est le vecteur des variables de commande exogènes,  $y(k)$  le vecteur des sorties,  $\Psi$  et  $\Phi$  sont deux fonctions qui dépendent de la structure de la partie acyclique du réseau [23].

## 4. Quelques réseaux de neurones usuels

Dans ce qui suit on présente quelques exemples typiques des réseaux de neurones les plus utilisés. [24]

### 4.1. Perceptron

Le perceptron est le premier des réseaux de neurones, conçu en 1958 par Rosenblatt, basé sur le modèle de McCulloch et Pitts (1943), dont le but est de modéliser la vision humaine [25].

Le perceptron généralement comprend trois couches principales, La première couche (d'entrée) représente la rétine. Les neurones de la couche suivante sont les cellules d'association, et la couche finale les cellules de décision. (La figure 2.9) illustre cette architecture.

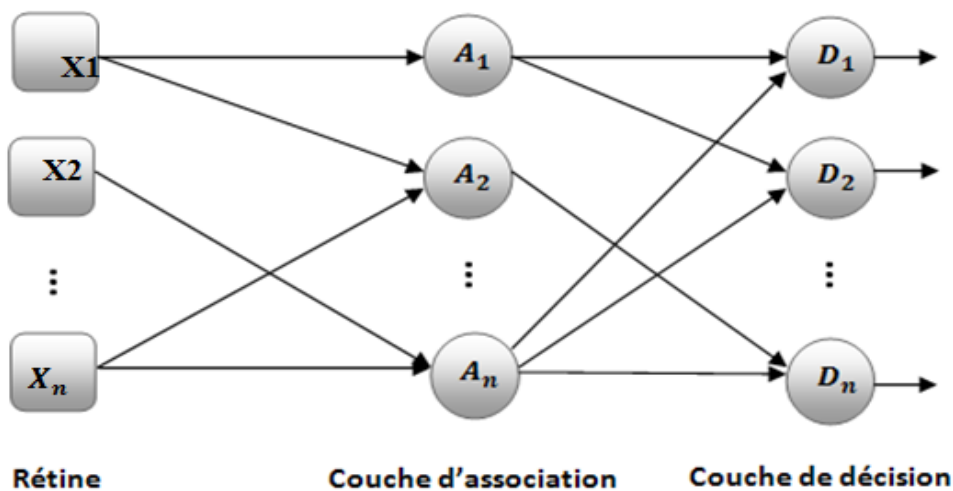


Fig.2.9 Représentation graphique de perceptron.

### 4.2. Perceptron Multicouches(PMC)

Les PMC sont des réseaux de neurones à couches, ils comprennent :

- Une couche d'entrée qui reçoit l'ensemble des formes à classer.
- Une ou plusieurs couches cachées intermédiaires.
- Une couche de sortie qui doit restituer l'ensemble des sorties désirées après apprentissage.

Aujourd'hui les réseaux de neurones multicouches sont les plus utilisés en diagnostic industriel en raison de leur capacité à résoudre les problèmes de classification non linéaire par l'utilisation de l'algorithme de retro-propagation du gradient, qui est une généralisation de la règle de Widrow-Hoff pour le calcul des poids de connexions [26].

### 4.3. Réseaux RBF (Radial Basis Function)

Les Réseaux à fonction de base radiale (RBF) proposés et utilisés par plusieurs auteurs (Moody et Darken en 1989; Renals et Rohwer en 1989;...) . Leur architecture en couches ressemble fortement à celle du PMC, mais avec quelques différences. L'architecture générale d'un réseau RBF comporte trois couches: couche d'entrée, couche cachée et une couche de sortie. La première couche en N entrées, elle est complètement connectée à la deuxième couche. Les neurones cachés ont une fonction RBF comme une fonction d'activation. La figure (Figure 2.10) représente une architecture générale d'un réseau RBF [25].

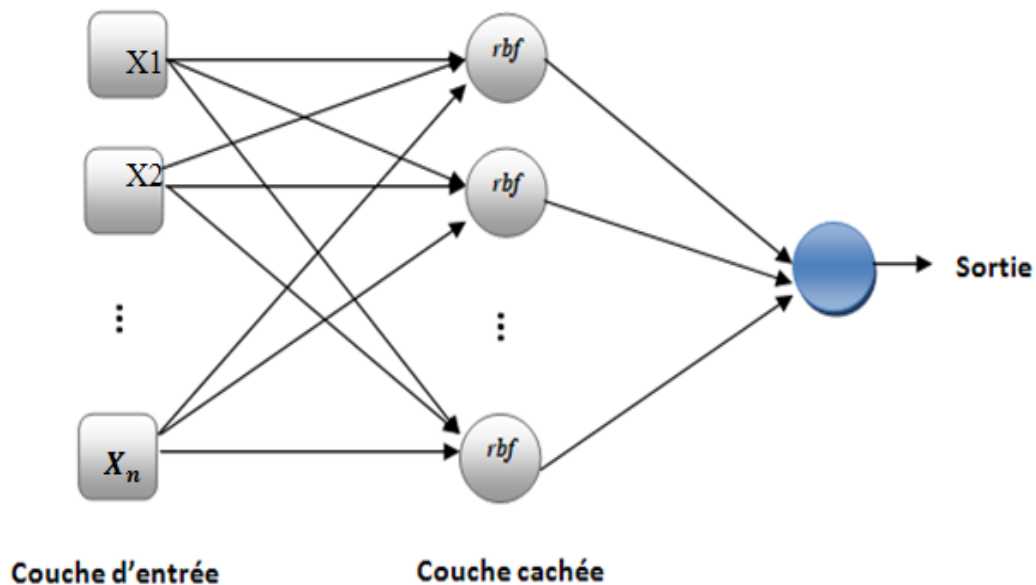


Fig.2.10 Architecture générale d'un réseau RBF.

#### 4.3.1. Architecture générale d'un réseau RBF

Pour des raisons de simplicité, on a décidé de faire une petite dualité entre le réseau RBF et le PMC, en précisant les ressemblances et les différences entre les deux types de réseaux. Ce choix est justifié par la popularité des PMCs et leur vaste utilisation dans les applications industrielles. Un réseau de neurone de type RBF est un PMC spéciale, son architecture est identique à celle d'un PMC à une seule couche cachée donc on peut dire qu'il prend toutes les caractéristiques d'un PMC simple sauf qu'il diffère en quelques points nous citons quelques-uns :

- **Le nombre des couches cachées**

Un réseau RBF ne peut contenir qu'une seule couche cachée, son architecture est fixée pour tous les problèmes à étudier.

- **La fonction d'activation**

Le réseau RBF utilise toujours une fonction dite à base radiale centrée d'un point et munie d'un rayon.

- **Les poids synaptiques**

Les poids entre la couche d'entrée et la couche cachée dans les modèles neuronaux de type RBF sont toujours d'une valeur d'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 les ressemblances entre un réseau RBF et un PMC, on peut mentionner quelques points :

- **La fonction de sortie**

Généralement une simple fonction linéaire qui renvoie une sommation pondérée des valeurs calculées par les neurones de la couche cachée. Bien sûr, ce n'est pas toujours le cas, parfois l'utilisation d'autres fonctions pourrait être plus adéquate dans un problème donné.

- **Le sens des connexions**

Les connexions entre les couches suivent le même sens, on peut dire qu'elles ne sont pas récurrentes, et chaque neurone est entièrement connecté vers les neurones de la couche suivante. [27]

#### 4.3.2. Avantages et inconvénients des réseaux RBF

Alors que l'on pouvait redouter d'avoir inventé un modèle aux belles propriétés théoriques, mais inutilisable en raison de durées d'apprentissage prohibitives, c'est tout le contraire qui se produit : la construction d'un réseau RBF est rapide et facile, et c'est le principal avantage de la technique [28-29].

Mais cet avantage se paye par des performances qui ne peuvent être aussi bonnes que celles de techniques plus sophistiquées (comme le Perceptron Multicouches). En particulier, les réseaux RBF sont peu performants :

- Sur les données dans des espaces de grande dimension (beaucoup de variables d'entrée). Cette faiblesse est propre à toutes les techniques locales.
- Sur des données très bruitées. La reconstruction locale de la fonction empêche le réseau de "moyenner" le bruit sur tout l'espace (comparez avec la Régression Linéaire, dont l'objectif est justement de moyenner le bruit sur les données).

### 5. Apprentissage

L'apprentissage est vraisemblablement la propriété la plus intéressante des réseaux. Elle ne concerne cependant pas tous les modèles, mais les plus utilisés [30].

#### Définition :

L'apprentissage est une phase du développement d'un réseau de neurones durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré.

L'apprentissage neuronal fait appel à des exemples de comportement.

Dans le cas des réseaux de neurones artificiels, on ajoute souvent à la description du modèle l'algorithme d'apprentissage. Le modèle sans apprentissage présente en effet peu d'intérêt. Dans la majorité des algorithmes actuels, les variables modifiées pendant l'apprentissage sont les poids des connexions. L'apprentissage est la modification des poids du réseau dans l'optique d'accorder la réponse du réseau aux exemples et à l'expérience. Il est souvent impossible de décider à priori des valeurs des poids des connexions d'un réseau pour une application donnée. A l'issue de l'apprentissage, les poids sont fixés : c'est alors la phase d'utilisation. Certains modèles de réseaux sont improprement dénommés à apprentissage permanent. Dans ce cas il est vrai que l'apprentissage ne s'arrête jamais, cependant on peut toujours distinguer une phase d'apprentissage (en fait de remise à jour du comportement) et une phase d'utilisation. Cette technique permet de conserver au réseau un comportement adapté malgré les fluctuations dans les données d'entrées.

Au niveau des algorithmes d'apprentissage, il a été défini deux grandes classes selon que l'apprentissage est dit supervisé ou non supervisé. Cette distinction repose sur la forme des exemples d'apprentissage. Dans le cas de l'apprentissage supervisé, les exemples sont des couples (Entrée, Sortie associée) alors que l'on ne dispose que des valeurs (Entrée) pour l'apprentissage non supervisé. Remarquons cependant que les modèles à apprentissage non supervisé nécessitent avant la phase d'utilisation une étape de labellisation effectuée l'opérateur, qui n'est pas autre chose qu'une part de supervision [30].

## 6. Types d'apprentissage

Il existe de nombreux types de règles d'apprentissage qui peuvent être regroupées en trois catégories : les règles d'apprentissage supervisé, non supervisé, et renforcé. Mais l'objectif fondamental de l'apprentissage reste le même : soit la classification, l'approximation de fonction ou encore la prévision. L'apprentissage consiste à extraire des régularités (à partir des exemples) qui peuvent être transférées à de nouveaux exemples [23].

### 6.1. Apprentissage supervisé

Un apprentissage est dit supervisé lorsque l'on force le réseau à converger vers un état final précis, en même temps qu'on lui présente un motif. Ce genre d'apprentissage est réalisé à l'aide d'une base d'apprentissage, constituée de plusieurs exemples de type entrées-sorties (les entrées du réseau et les sorties désirées ou encore les solutions souhaitées pour l'ensemble des sorties du réseau).

La procédure usuelle dans le cadre de la prévision est l'apprentissage supervisé (ou à partir d'exemples) qui consiste à associer une réponse spécifique désirée à chaque signal d'entrée. La modification des poids s'effectue progressivement jusqu'à ce que l'erreur (ou l'écart) entre les sorties du réseau (ou résultats calculés) et les résultats désirés soit minimisée.

Cet apprentissage n'est possible que si un large jeu de données est disponible et si les solutions sont connues pour les exemples de la base d'apprentissage [23].

## 6.2. Apprentissage renforcé

L'apprentissage renforcé est une technique similaire à l'apprentissage supervisé à la différence qu'au lieu de fournir des résultats désirés au réseau, on lui accorde plutôt un grade (ou score) qui est une mesure du degré de performance du réseau après quelques itérations.

Les algorithmes utilisant la procédure d'apprentissage renforcé sont surtout utilisés dans le domaine des systèmes de contrôle [23].

## 6.3. Apprentissage non supervisé

L'apprentissage non supervisé consiste à ajuster les poids à partir d'un seul ensemble d'apprentissage formé uniquement de données. Aucun résultat désiré n'est fourni au réseau.

Qu'est-ce que le réseau apprend exactement dans ce cas ? L'apprentissage consiste à détecter les similarités et les différences dans l'ensemble d'apprentissage. Les poids et les sorties du réseau convergent, en théorie, vers les représentations qui capturent les régularités statistiques des données. Ce type d'apprentissage est également dit compétitif et(ou) coopératif. L'avantage de ce type d'apprentissage réside dans sa grande capacité d'adaptation reconnue comme une auto-organisation, « self-organizing ». L'apprentissage non supervisé est surtout utilisé pour le traitement du signal et l'analyse factorielle [23].

## 7. Algorithme d'apprentissage

L'apprentissage des réseaux de neurones, consiste à adapter les poids synaptiques de telle manière que l'erreur entre la sortie du réseau et la sortie désirée soit aussi petite que possible. La plupart des algorithmes d'apprentissage des réseaux de neurones est basée sur les méthodes du gradient: ils cherchent à minimiser un critère de la forme suivante:

$$J = \frac{1}{2} \sum_{q=1}^m \left( y_{rq}(X_p) - y_q^d(X_p) \right)^2 \quad (2.4)$$

Où  $y_{rq}$  et  $y_q^d$  sont la sortie du réseau et la sortie désirée pour le vecteur d'entrée  $X_p$ . Cette optimisation se fait d'une manière itérative, en modifiant les poids en fonction du gradient de la fonction de coût selon la loi d'adaptation suivante:

$$w_{ji}^l(k+1) = w_{ji}^l(k) - \frac{\eta \partial J}{\partial w_{ji}^l} \quad (2.5)$$

Où  $w_{ji}^l$  est le poids de la connexion entre le  $j^{\text{eme}}$  neurone de la couche l et le  $i^{\text{eme}}$  neurone de la couche l-1.

$i=1, \dots, n+1$  la  $i^{\text{eme}}$  composante du vecteur d'entrée.

$j=1, \dots, m+1$  la  $j^{\text{eme}}$  composante du vecteur de sortie.

$l=1, \dots, L$  l'ordre d'une couche dans le réseau de neurone.

$\eta$  est une constante positive ( $\eta > 0$ ) appelée taux d'apprentissage.

Le gradient est calculé par une méthode spécifique aux réseaux de neurones, dites méthode de rétro propagation. Dans cette méthode il est possible de calculer le gradient de la fonction coût en retropropageant l'erreur commise en sortie vers les couches cachées[21].

### 7.1. Algorithme de rétropropagation

L'algorithme de rétropropagation (ARP) ou de propagation arrière « backpropagation » est l'exemple d'apprentissage supervisé le plus utilisé à cause de l'écho médiatique de certaines applications spectaculaires telles que la démonstration de Sejnowski et Rosenberg (1987) dans laquelle l'ARP est utilisé dans un système qui apprend à lire un texte. Un autre succès fut la prédiction des cours du marché boursier et plus récemment la détection de la fraude dans les opérations par cartes de crédit .

La technique de rétropropagation du gradient (Backpropagation en anglais) est une méthode qui permet de calculer le gradient de l'erreur pour chaque neurone du réseau, de la dernière couche vers la première. L'historique des publications montre que l'ARP a été découvert indépendamment par différents auteurs mais sous différentes appellations . Le principe de la rétropropagation peut être décrit en trois étapes fondamentales : acheminement de l'information à travers le réseau; rétropropagation des sensibilités et calcul du gradient; ajustement des paramètres par la règle du gradient approximé. Il est important de noter que l'ARP souffre des limitations inhérentes à la technique du gradient à cause du risque d'être piégé dans un minimum local. Il suffit que les gradients ou leurs dérivées soient nuls pour que

le réseau se retrouve bloqué dans un minimum local. Ajoutons à cela la lenteur de convergence surtout lorsqu'on traite des réseaux de grande taille (c'est à dire pour lesquels le nombre de poids de connexion à déterminer est important). Pour rendre l'optimisation plus performante, on peut utiliser des méthodes de second ordre telles que les méthodes dites de Quasi-Newton ou de Newton modifiée[31].

## 7.2. Méthodes Quasi-Newton

L'algorithme Quasi-Newton nécessite la connaissance des valeurs de la fonction objectif à minimiser et du vecteur des dérivées premières au point solution courant. Le vecteur des dérivées est évalué analytiquement[31].

Avec les méthodes Quasi-Newton, on veut généraliser la formule itérative de Newton.

$$X^{k+1} = X^k - \alpha^k [\nabla^2 f(X^k)]^{-1} \cdot \nabla f(X^k) \quad (2.6)$$

Le principe des méthodes de résolution de type Quasi-Newton est de générer une séquence de matrices symétriques définies positives qui soient des approximations, toujours améliorées, de la matrice Hessienne réelle ou de son inverse. Désignons par  $H_k$  la suite d'approximation de l'inverse de la matrice Hessienne  $[\nabla^2 f(X^k)]^{-1}$  et par  $B_k$  la suite des estimations de la matrice Hessienne  $\nabla^2 f(X^k)$ .

On recherche une méthode telle que, dans le cas d'un problème quadratique, la matrice  $B_k$  converge vers la valeur exacte des dérivées secondes (constantes dans ce cas), de sorte qu'en fin de convergence, on retrouve une convergence de type Newton. Si l'on applique la méthode à une fonction quelconque,  $B_k$  peut être considéré, à chaque instant, comme une approximation (définie positive) du Hessien[31].

## 7.3. Algorithme de BFGS

L'algorithme de BFGS (du nom de ses inventeurs : Broyden, Fletcher, Goldfarb et Shanno) prend implicitement en compte les dérivées secondes et s'avère donc nettement plus performante que la méthode de rétropropagation. Le nombre d'itérations est nettement plus faible et les temps de calcul réduits d'autant.

L'algorithme de BFGS est une règle d'ajustement des paramètres qui a l'expression suivante :

$$\theta^k = \theta^{k-1} \pm \mu_k M_k \nabla J(\theta^{k-1}) \quad (2.7)$$

Où  $M_k$  est une approximation, calculée itérativement, de l'inverse de la matrice Hessienne.

L'approximation de l'inverse du Hessien est modifiée à chaque itération suivant la règle suivante :

$$M_k = M_{k+1} + \left[ 1 + \left( \frac{\gamma_{k-1}^T M_{k-1} \gamma_{k-1}}{\delta_{k-1}^T \gamma_{k-1}} \right) \right] \frac{\delta_{k-1}^T \delta_{k-1}}{\delta_{k-1}^T \gamma_{k-1}} \pm \frac{\delta_{k-1} \gamma_{k-1}^T M_{k-1} + M_{k-1} \gamma_{k-1} \delta_{k-1}^T}{\delta_{k-1}^T \gamma_{k-1}} \quad (2-8)$$

Avec  $\gamma_{k-1} = \nabla J(\theta^k) \pm \nabla J(\theta^{k-1})$  et  $\delta_{k-1} = \theta^k \pm \theta^{k-1}$ . Nous prenons pour valeur initiale de  $M$  la matrice identité. Si, à une itération, la matrice calculée n'est pas définie positive, elle est réinitialisée à la matrice identité.

Une méthode "quasi-newtonienne", n'est efficace que si elle est appliquée au voisinage d'un minimum. D'autre part, la règle du gradient simple est efficace lorsqu'on est loin du minimum et sa convergence ralentit considérablement lorsque la norme du gradient diminue (c'est à dire lorsqu'on s'approche du minimum). Ces deux techniques sont donc complémentaires. De ce fait, l'optimisation s'effectue en deux étapes : utilisation de la règle du gradient simple pour approcher un minimum, et de l'algorithme de BFGS pour l'atteindre[31].

## 8. Avantages et inconvénients des réseaux de neurones

### 8.1. Avantages des réseaux de neurones

- Capacité de représenter n'importe quelle fonction, linéaire ou pas, simple ou complexe
- Faculté d'apprentissage à partir d'exemples représentatifs, par "rétropropagation des erreurs". L'apprentissage (ou construction du modèle) est automatique .
- Résistance au bruit ou au manque de fiabilité des données .
- Simple à manier, beaucoup moins de travail personnel à fournir que dans l'analyse statistique classique. Aucune compétence en mathématiques, informatique statistique requise .
- Comportement moins mauvais en cas de faible quantité de données .
- Pour l'utilisateur novice, l'idée d'apprentissage est plus simple à comprendre que les complexités des statistiques multivariées[21].

### 8.2. Inconvénients des réseaux de neurones

- L'absence de méthode systématique permettant de définir la meilleure topologie du réseau et le nombre de neurones à placer dans la (ou les) couche(s) cachée(s) .

- Le choix des valeurs initiales des poids du réseau et le réglage du pas d'apprentissage, qui jouent un rôle important dans la vitesse de convergence .
- Le problème du sur-apprentissage (apprentissage au détriment de la généralisation) .
- La connaissance acquise par un réseau de neurone est codée par les valeurs des poids synaptiques, les réseaux de neurones sont donc des boites noires où les connaissances sont inintelligibles pour l'utilisateur[21].

### **Conclusion**

Dans ce chapitre nous avons essayé de donner en bref une description sur les réseaux de neurones artificiels,et aussi les différents types d'architectures et de modele qui existent.Nous avons parleré d'une façon détaillée des deux structures usueles des RNAs telles que : les PMC et les RBF.

Nous avons aussi présenté une définition de l'apprentissage des réseaux de neurones,puis une description générale de l'algorithme de retropropagation,mise en œuvre l'algorithme RNAs et aussi le domaine d'utilisation des réseaux de neurones.Nous avons terminé ce chapitre par la les avantages et les inconvenients des RNAs ,pour aller directement aux SVM qui constituent notre suivant chapitre.

# Chapitre III

## Les machines à vecteurs de support

### Introduction

Parmi les méthodes à noyaux, inspirées de la théorie statistique de l'apprentissage de Vladimir Vapnik, les SVM constituent la forme la plus connue. Cette méthode est donc une alternative récente pour la classification. Cette méthode repose sur l'existence d'un classificateur linéaire dans un espace approprié. Puisque c'est un problème de classification à deux classes, cette méthode fait appel à un jeu de données d'apprentissage pour apprendre les paramètres du modèle [32].

Au cours de ce chapitre nous allons présenter une généralité sur les SVM, leur principe de fonctionnement, ainsi que des notions de bases essentielles concernant les SVM. Nous terminerons notre chapitre par une description détaillée sur les SVM appliquées à la classification binaire.

### 1. Généralité

Les machines à vecteurs de support ou séparateurs à vastes marges (en anglais Support Vector Machines, SVM) sont une classe d'algorithmes d'apprentissage statistique initialement définis pour la discrimination. Destinées à résoudre des problèmes de discrimination et de régression, les SVMs sont considérées comme étant une généralisation des classifieurs linéaires. Cette méthode fût introduite par V. Vapnik à partir de 1995 sur le développement d'une théorie statistique de l'apprentissage appelée théorie de Vapnik-Chervonenkis [33, 34]. Son succès est justifié par les solides bases théoriques qui la soutiennent. Elle permet en fait d'aborder un grand nombre de tâches de classification supervisée telles que : la catégorisation automatique de textes, la reconnaissance de visages, les diagnostics médicaux, la reconnaissance des chiffres manuscrits etc. La méthode SVM est une technique de classification discriminative très populaire, particulièrement bien adaptée au traitement des données de très haute dimension, tels que les textes et les images [35]. Depuis son introduction dans le domaine de la reconnaissance de formes, plusieurs travaux ont pu montrer son efficacité. Fortement basée sur la théorie, il existe en fait un lien direct entre la théorie de l'apprentissage statistique et son algorithme d'apprentissage. Ce dernier suscite un vif intérêt dans la communauté « Machine Learning (ML) » pour ces bonnes performances, et le fait qu'il trouve une solution unique. Toutefois, la véritable force de la méthode reste au

niveau du mécanisme de projection qui lui permet de changer d'espace pour réaliser l'apprentissage. Dans ce qui suit, nous allons rappeler la théorie SVM du point de vue mathématique pour illustrer son mécanisme.

## 2. principe de fonctionnement

Le principe des machines à vecteurs de support (SVM) consiste à déterminer une séparation linéaire entre deux classes en maximisant la marge définie comme la distance minimale entre la surface de séparation (un hyperplan) et le point le plus proche de l'ensemble d'apprentissage (vecteur de support). On parlera de méthode des séparateurs à vaste marge [36].

Lorsque les données ne sont pas linéairement séparables, ou contiennent du bruit (outliers: données mal étiquetées ou étranges), on projette les points de l'ensemble d'apprentissage dans un espace de dimension supérieure où une séparation linéaire peut être trouvée. On parle alors de SVM à noyau [36].

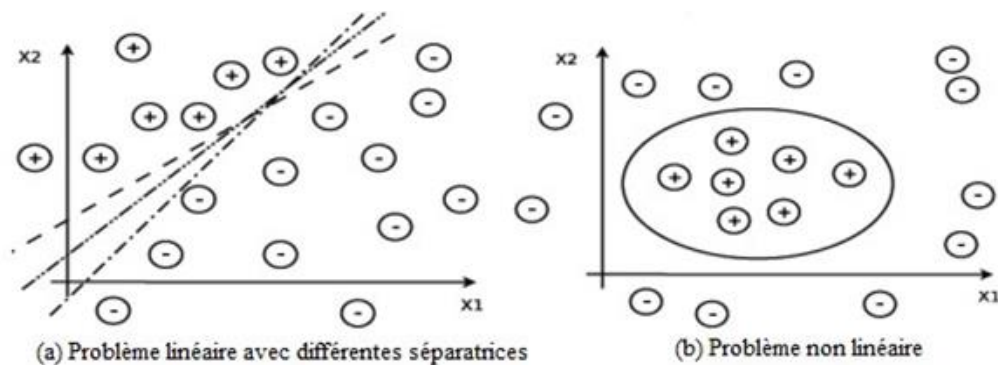


Fig. 3.1.Exemple de classification binaire linéaire et non-linéaire

### 2.1. Notions de base: Hyperplan, marge et support vecteur

Pour deux classes d'exemples donnés, le but de SVM est de trouver un classificateur qui va séparer les données et maximiser la distance entre ces deux classes. Avec SVM, ce classificateur est un classificateur linéaire appelé hyperplan.

Dans le schéma qui suit, on détermine un hyperplan qui sépare les deux ensembles de points [37].

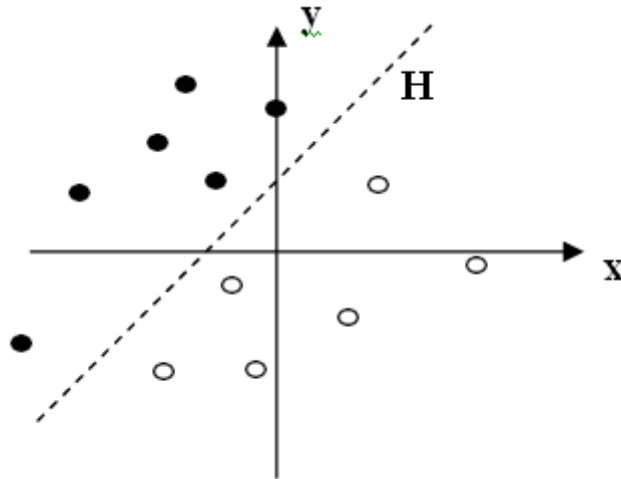


Fig. 3.2 Exemple d'un hyperplan séparateur [37].

Les points les plus proches, qui sont les seuls utilisés pour la détermination de l'hyperplan, sont appelés vecteurs de support.

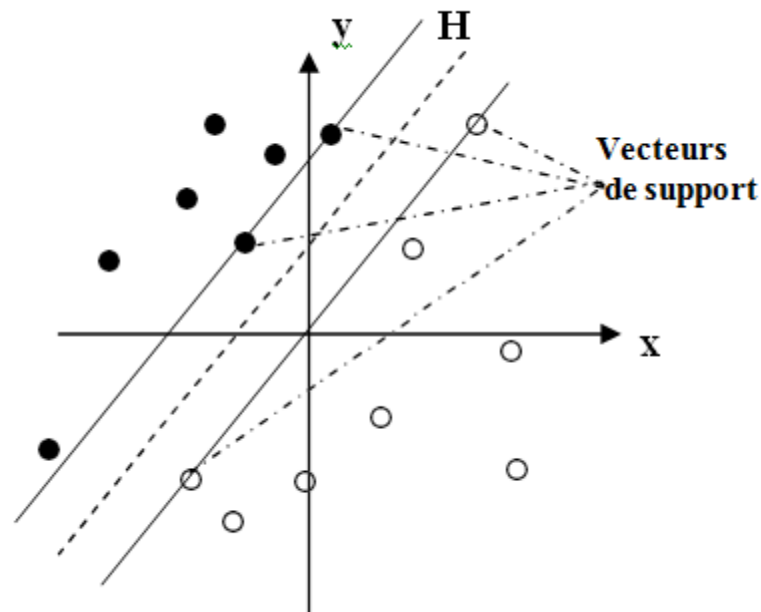


Fig.3.3 Exemple de vecteurs de support [38].

Il est évident qu'il existe une multitude d'hyperplan valide mais la propriété remarquable des SVM est que cet hyperplan doit être optimal. Nous allons donc en plus chercher parmi les hyperplans valides, celui qui passe « au milieu » des points des deux classes d'exemples. Intuitivement, cela revient à chercher l'hyperplan le « plus sûr ». En effet, supposons qu'un exemple n'ait pas été décrit parfaitement, une petite variation ne modifiera pas sa

classification si sa distance à l'hyperplan est grande. Formellement, cela revient à chercher un hyperplan dont la distance minimale aux exemples d'apprentissage est maximale [39]. On appelle cette distance « marge » entre l'hyperplan séparateur optimal et les exemples. L'hyperplan séparateur optimal est celui qui maximise la marge. Comme on cherche à maximiser cette marge, on parlera de séparateurs à vaste marge.

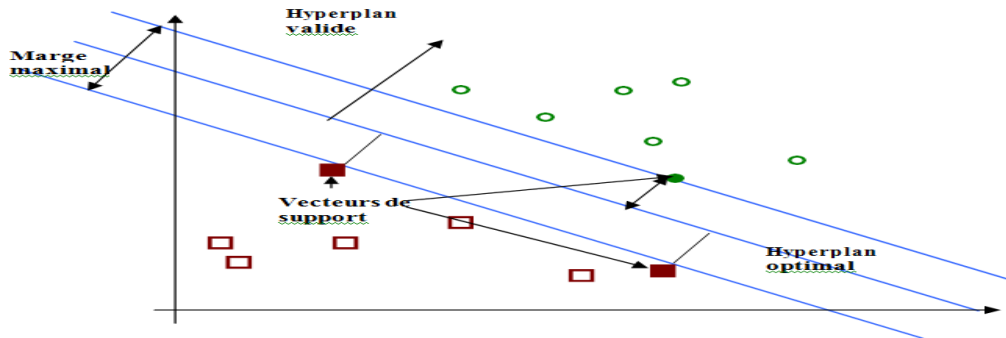


Fig.3.4. Exemple de marge maximale (hyperplan valide) [39].

## 2.2. Maximisation de la marge

Intuitivement, le fait d'avoir une marge plus large procure plus de sécurité lorsque l'on classe un nouvel exemple. De plus, si l'on trouve le classificateur qui se comporte le mieux vis-à-vis des données d'apprentissage, il est clair qu'il sera aussi celui qui permettra au mieux de classer les nouveaux exemples. Dans le schéma qui suit, la partie droite nous montre qu'avec un hyperplan optimal, un nouvel exemple reste bien classé alors qu'il tombe dans la marge. On constate sur la partie gauche qu'avec une plus petite marge, l'exemple se voit mal classé [37].

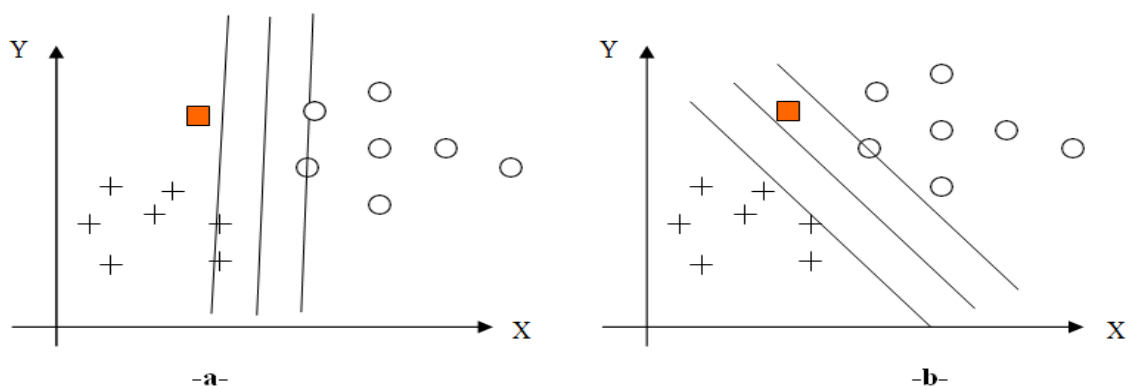


Fig. 3.5 a) Hyperplan avec faible marge, fig.3.5. b) Meilleur hyperplan séparateur [37].

En général, la classification d'un nouvel exemple inconnu est donnée par sa position par rapport à l'hyperplan optimal. Dans le schéma suivant, le nouvel élément sera classé dans la catégorie des « + ».

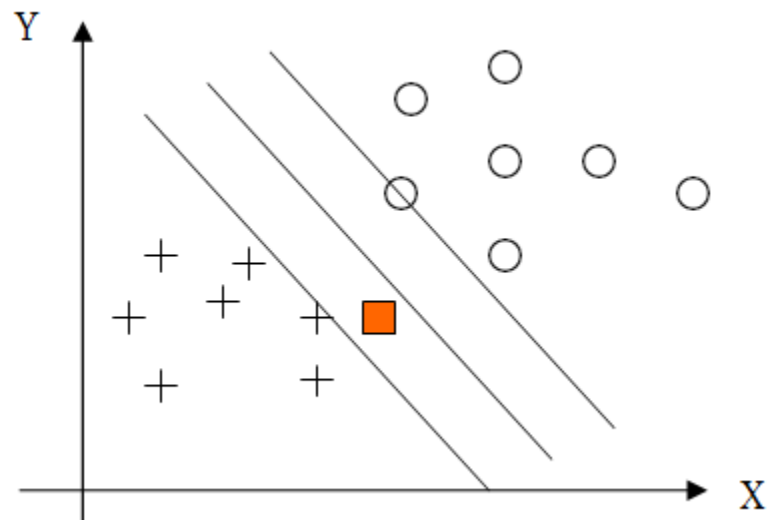


Fig. 3.6. Exemple de classification d'un nouvel élément [37].

### 2.3. Linéarité et non-linéarité

Parmi les modèles des SVM, on constate les cas linéairement séparable et les cas non linéairement séparable. Les premiers sont les plus simples de SVM car ils permettent de trouver facilement le classificateur linéaire. Dans la plupart des problèmes réels il n'y a pas de séparation linéaire possible entre les données, le classificateur de marge maximale ne peut pas être utilisé car il fonctionne seulement si les classes de données d'apprentissage sont linéairement séparables [37].

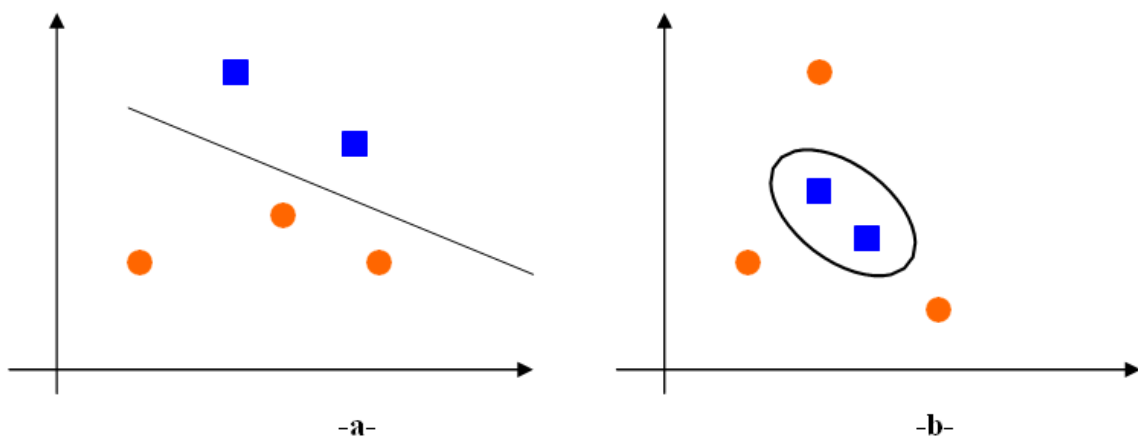


Fig. 3.7 a) Cas linéairement séparable,

fig.3.7.b) Cas non linéairement séparable [37].

## 2.4. Cas non linéaire

Pour surmonter les inconvénients des cas non linéairement séparable, l'idée des SVM est de changer l'espace des données. La transformation non linéaire des données peut permettre une séparation linéaire des exemples dans un nouvel espace. On va donc avoir un changement de dimension. Cette nouvelle dimension est appelé « espace de ré-description ». En effet, intuitivement, plus la dimension de l'espace de ré-description est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les exemples est élevée. Ceci est illustré par le schéma suivant.

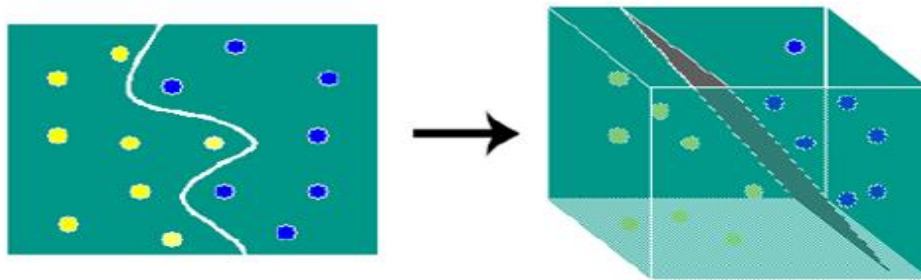


Fig. 3.8. Exemple de changement de l'espace de données [37].

On a donc une transformation d'un problème de séparation non linéaire dans l'espace de représentation en un problème de séparation linéaire dans un espace de ré-description de plus grande dimension. Cette transformation non linéaire est réalisée via une fonction noyau. En pratique, quelques familles de fonctions noyau paramétrables sont connues et il revient à l'utilisateur de SVM d'effectuer des tests pour déterminer celle qui convient le mieux pour son application. On peut citer les exemples de noyaux suivants : polynomiale, gaussien, sigmoïde et laplacien [37].

## 3. SVM appliquée à la classification binaire

En accord avec la théorie de l'apprentissage statistique, une fonction qui décrit correctement un ensemble d'apprentissage  $X$  et qui appartient à un ensemble de fonctions avec une dimension VC réduite, aura un bon pouvoir de généralisation, indépendamment de la dimension de l'espace de l'entrée. Basées sur ce principe, les Machines à Vecteurs de Support ou SVM ont une approche systématique pour trouver une fonction linéaire appartenant à un ensemble de fonctions avec une dimension VC basse [40]. Les systèmes d'apprentissage appelés SVM sont donc les algorithmes basés sur les trois principes mathématiques suivants [41] :

- **Le principe de Fermat (1638) :** les points qui minimisent ou maximisent une fonction dérivable annulent sa dérivée. Ils sont appelés points stationnaires.
- **Le principe de Lagrange (1788) :** pour résoudre un problème d'optimisation sous contraintes, il suffit de rechercher un point stationnaire  $x_0$  du Lagrangien  $L$  de la fonction  $f$  à optimiser, les  $f_i$  expriment les contraintes :

$$L(x, \alpha) = f(x) + \sum_{i=1}^k \alpha_i f(x)_i \quad (3.1)$$

Où les  $a_i$  sont des constantes appelées coefficients (ou multiplicateurs) de Lagrange.

- **Le principe de Karush-Kuhn-Tucker (KKT), (1951) :** les relations de *KKT* peuvent s'appliquer au cas qui nous intéresse. Avec des fonctions  $f$  et  $f_i$  convexes, il est même toujours possible de trouver un *point-selle*  $(x_0, \alpha^*)$  qui vérifie :

$$\min_x L(x, \alpha^*) = L(x_0, \alpha^*) = \max_{\alpha \geq 0} L(x_0, \alpha) \quad (3.2)$$

Le principe de base des SVM consiste donc de ramener le problème de la discrimination à celui, linéaire, de la recherche d'un hyperplan séparateur optimal, dans le cadre de la classification appliquant les principes cités auparavant. Deux idées permettent d'atteindre cet objectif :

- La première consiste à définir l'hyperplan comme solution d'un problème d'optimisation sous contraintes dont la fonction objectif ne s'exprime qu'à l'aide de produits scalaires entre vecteurs et dans lequel le nombre de contraintes "actives" ou vecteurs supports contrôle la complexité du modèle.
- Le passage à la recherche de surfaces séparatrices non linéaires est obtenu par l'introduction d'une fonction noyau (kernel) dans le produit scalaire induisant implicitement une transformation non linéaire des données vers un espace intermédiaire (*feature space*) de plus grande dimension de l'espace initial et dans lequel est résolu le problème linéaire.

En effet, le principal objectif des SVM appliquées à la reconnaissance de formes est de construire un hyperplan séparateur optimal entre deux classes, c'est à dire, avec la plus grande marge (figure 3.9). Lorsqu'une solution linéaire n'est pas possible, la méthode réalise une projection de l'espace d'entrée  $X$  d'apprentissage dans un espace de caractéristiques  $Z$  de dimension plus importante, à travers une fonction noyau. Grâce à la liberté d'utiliser

différents types de noyaux, l'hyperplan séparateur optimal correspond à des estimateurs non linéaires différents dans l'espace original.

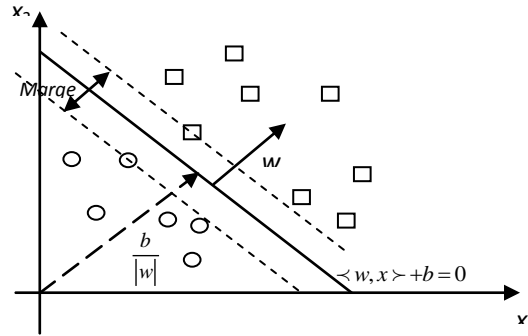


Fig. 3.9. Hyperplan séparateur linéaire optimal et marge.

### 3.1. Classifieur linéaire

Un problème de classification est dit linéairement séparable lorsqu'il existe une fonction de décision linéaire (appelé aussi séparateur linéaire), classant correctement toutes les observations de l'ensemble d'apprentissage.

▪ **Définition 3.1 :** Considérons l'ensemble d'apprentissage  $\{x_1, y_1\}, \dots, \{x_l, y_l\}$ , avec  $x \in X$  et  $y \in \{-1, 1\}$ , où  $l$  est le nombre d'observations et  $X$  est une distribution dans l'espace  $\mathbb{R}^n$ . L'ensemble est linéairement séparable [42], si :

$$\exists w \in \mathbb{R}^n, b \in \mathbb{R} : y_i (\langle w, x_i \rangle + b) \geq 0 \quad \forall i = 1, \dots, n \quad (3.3)$$

▪ **Définition 3.2 :** Un classifieur est dit linéaire lorsqu'il est possible d'exprimer sa fonction de décision par une fonction linéaire en  $x$ . On peut, en toute généralité, exprimer une telle fonction comme ceci [43] :

$$f(x) = \langle w, x \rangle + b = \sum_i w_i x_i + b \quad (3.4)$$

Où  $w \in \mathbb{R}^m$  et  $b \in \mathbb{R}$  sont des paramètres, et  $x \in \mathbb{R}$  est une variable (nous supposons que les exemples nous sont fournis dans un format vectoriel).

Pour ce faire, on doit trouver le vecteur  $w$  et la constante  $b$  (éq. 3.4) qui minimisent la norme  $|w|^2 = w^T w$  (puisque'elle est inversement proportionnelle à la marge), sous les contraintes [41, 44] :

$$\begin{aligned} w^T x_i + b &\geq 1, & \text{si } y_i &= 1 \\ w^T x_i + b &\leq -1, & \text{si } y_i &= -1 \end{aligned} \quad (3.5)$$

Pour décider à quelle catégorie un exemple  $x_i$  appartient, il suffit de prendre le signe de la fonction de décision  $y = \text{sign}(f(x_i))$ . Géométriquement, cela revient à considérer un hyperplan qui est le lieu des points  $x$  satisfaisants [45] :

$$wx + b = 0 \quad (3.6)$$

En orientant l'hyperplan, la règle de décision correspond à observer de quel côté de l'hyperplan se trouve l'exemple  $x_i$  selon (éq. 3.5), on peut combiner les deux inéquations en une seule [46] :

$$y_i(w \cdot x_i + b) - 1 \geq 0 \quad \forall i \quad (3.7)$$

Dans ce cas là, l'algorithme des SVM cherche simplement à trouver l'hyperplan qui minimise la distance (marge) entre les  $x_i$  et l'hyperplan (figure 3.10). Ce qui revient à minimiser  $\|w\|$ , où encore de manière équivalente  $\frac{1}{2}\|w\|^2$ . On voit que le vecteur  $w$  définit la pente de l'hyperplan, donc il est perpendiculaire à l'hyperplan. Le terme  $b$  quant à lui, permet de translater l'hyperplan parallèlement à lui-même. La définition consiste à dire, qu'il doit exister un hyperplan laissant d'un côté toutes les données positives et de l'autre toutes les données négatives.

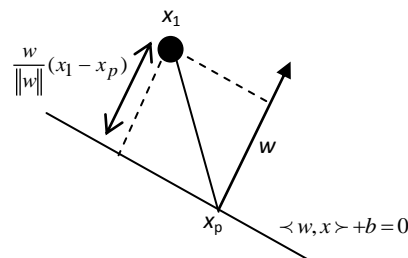


Fig. 3.10. Expression de la marge pour l'exemple  $x_i$  [46].

#### \* Marge d'un classifieur

La notion de marge peut être relative à un exemple particulier ou à l'ensemble d'apprentissage. Pour un échantillon donné, il est possible d'associer deux marges à un même classifieur linéaire : une marge numérique ou fonctionnelle et une marge géométrique [46]. La marge fonctionnelle d'un exemple  $x_i$  par rapport à l'hyperplan séparateur (classifieur) caractérisé par  $w$  et  $b$ , est la quantité :  $y_i(w \cdot x_i + b)$ . La marge fonctionnelle est donnée par la plus petite valeur de la fonction de décision atteinte sur un point de l'échantillon. Elle dépend

de la pente de l'hyperplan de décision. En prenant un point quelconque  $x_p$  se trouvant sur l'hyperplan, la marge fonctionnelle peut s'exprimer par :

$$\frac{w}{\|w\|} \cdot (x_i - x_p) \quad (3.8)$$

La figure 3.10 illustre la situation. En utilisant la marge fonctionnelle et le fait que  $x_p$  est sur l'hyperplan, on peut calculer la valeur de la marge fonctionnelle :

$$\begin{aligned} & \langle w, x_i \rangle + b = f(x_i) \\ - & \langle w, x_p \rangle + b = 0 \\ \rightarrow & w \cdot (x_i - x_p) = f(x_i) \\ \Rightarrow & \frac{w}{\|w\|} \cdot (x_i - x_p) = \frac{f(x_i)}{\|w\|} \end{aligned} \quad (3.9)$$

La marge géométrique d'un exemple  $x_i$ , par rapport à l'hyperplan caractérisé par  $w$  et  $b$  est la quantité [46] :

$$\Psi_{w,b}(x_i, y_i) = y_i \left( \frac{w}{\|w\|} \cdot x_i + \frac{b}{\|w\|} \right) \quad (3.10)$$

Remarquons que la marge d'un exemple mal classé, la marge correspond à la distance négative qui le sépare de l'hyperplan.

**Remarque :** la marge de l'ensemble d'apprentissage (training set) par rapport à l'hyperplan caractérisé par  $w$  et  $b$  est définie comme étant [46] :

$$\Psi_{w,b} = \min_{i=1..m} \Psi_{w,b}(x_i, y_i) \quad (3.11)$$

En général, la marge est la distance minimale entre les échantillons de l'ensemble d'apprentissage et la frontière de décision. Il a été montré que pour la classe des hyperplans, la dimension de VC peut être bornée en fonction de la marge. La marge peut à son tour être mesurée grâce au vecteur poids. Puisque nous supposons que les échantillons sont séparables, on peut redéfinir  $w$  et  $b$  de sorte que les échantillons  $x$  les plus proches de l'hyperplan satisfassent  $|wx + b| = 1$ .

Considérons maintenant deux échantillons  $x_1$  et  $x_2$  de classes différentes tels qu'on ait :  $wx_1 + b = 1$  et  $wx_2 + b = -1$ . La marge  $d$  correspond alors à la distance entre  $x_1$  et  $x_2$  mesurée perpendiculairement à l'hyperplan [44, 47] :

$$d = \left\langle \frac{w}{\|w\|}, x_1 - x_2 \right\rangle = \frac{2}{\|w\|} \quad (3.12)$$

Intuitivement, le fait d'avoir une marge plus large procure plus de « sécurité » lorsque l'on classe un exemple inconnu. La partie gauche de la figure 3.11 nous montre qu'avec l'hyperplan optimal, un nouvel exemple reste bien classé alors qu'il tombe dans la marge. On constate sur la partie droite qu'avec une plus petite marge, l'exemple se voit mal classé.

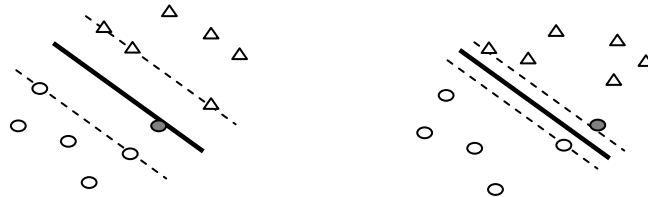


Fig. 3.11. Importance de la marge de l'hyperplan.

Cette intuition est plus précisément exprimée dans un théorème introduit par Vapnik faisant intervenir la dimension VC ( $h$ ).

▪ **Théorème 3.1** (Dimension VC d'un classifieur linéaire) : Soit l'ensemble des hyperplans  $w \cdot x = 0$ , où  $w$  est redimensionné de sorte que les exemples les plus proches se trouvent sur les hyperplans canoniques. La famille des fonctions de décision  $f_w(x) = w \cdot x$  tel que  $\|w\| = A$ , possède une dimension VC bornée par [41, 48] :

$$h \leq \min[(A^2 r^2), d] + 1 \quad \text{et} \quad \|w\| \leq A \quad (3.13)$$

Où :  $r$  est le rayon de la plus petite boule englobant les données.

Ce résultat liant la dimension de VC de la classe des hyperplans de séparation, à la marge et à la longueur du vecteur poids  $w$ . Ce théorème nous dit qu'en diminuant la borne  $A$  sur  $\|w\|$ , c.-à-d en augmentant (indirectement) la marge, la dimension VC du classifieur diminue. On peut donc contrôler la dimension VC en agissant sur la marge.

Les classifieurs ayant pour critère de maximiser la marge de l'ensemble d'apprentissage sont appelés classifieurs à marge maximale SVM.

- **Hyperplans canoniques**

On suppose que l'ensemble d'apprentissage est linéairement séparable. On peut définir deux plans se trouvant de part et d'autre de l'hyperplan et parallèles à celui-ci, sur lesquels reposent les exemples les plus proches. La figure 3.12 illustre cette situation. Il est possible que différentes équations correspondent au même plan géométrique d'après la définition de l'hyperplan :

$$a < w, x > + \frac{b}{a} = 0 \quad \forall a \in \mathbb{R} \quad (3.14)$$

Il est donc possible de redimensionner  $w$  et  $b$  de telle sorte que les deux plans parallèles aient respectivement pour équations montrées dans la figure :

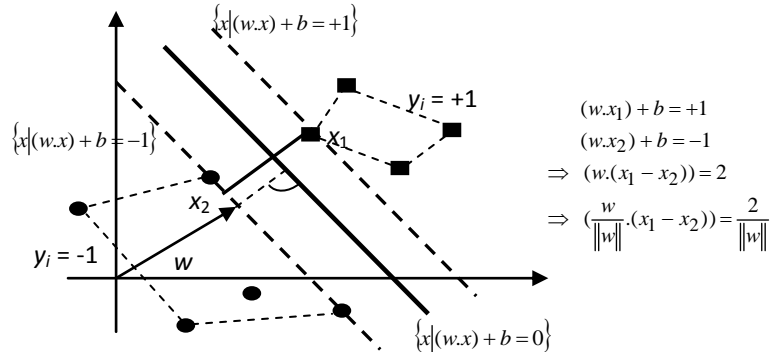


Fig. 3.12. Hyperplans canoniques.

Ces deux hyperplans sont appelés hyperplans canoniques. Notons que la marge des hyperplans canoniques est  $\frac{1}{\|w\|}$ . Le vecteur  $w$  possède à présent une signification géométrique très claire.

- **Maximisation de la marge d'un classifieur**

Lorsque des observations sont linéairement séparables, il existe dans le cas général une infinité de frontières de décision linéaires séparant cet échantillon. La notion de marge et d'hyperplans canoniques offre un critère de choix parmi toutes ces solutions en admettant que, maximiser la marge c'est aussi maximiser la confiance et donc minimiser la probabilité d'erreurs associées au classifieur. Le problème devient un problème d'optimisation mathématique de telle sorte que sa solution nous fournisse l'hyperplan optimal. Cela revient à chercher cet hyperplan dont la distance minimale aux exemples d'apprentissage est maximale. L'hyperplan optimal est défini par le vecteur de poids  $w$  vérifiant l'équation [49] :

$$\text{Arg max min } \left\{ \|x - x_i\| : x_{w,b} \in \mathfrak{R}^d, (w^T x + b) = 0, i = 1, \dots, m \right\} \quad (3.15)$$

Pour cet hyperplan, la marge vaut  $1/\|w\|$  et sa recherche revient à minimiser  $\|w\|$ , soit à résoudre le problème suivant qui porte sur les paramètres  $w$  et  $b$  [50, 51] :

$$QPI \quad \begin{cases} \text{Minimiser}_w & \frac{1}{2} \|w\| \\ \text{Sous les contraintes} & y_i (w^T x_i + b) \geq 1, \quad i = 1, \dots, m \end{cases} \quad (3.16)$$

Il s'agit d'un problème quadratique ( $QP$ ) dont le but est de minimiser une fonction objective, qui est le carré de l'inverse de la double marge. L'unique contrainte stipule que les exemples doivent être bien classés et qu'ils ne dépassent pas les hyperplans canoniques. Cette écriture ( $QPI$ ) du problème est appelée formulation primale. Dans cette formulation, les variables à fixer sont les composantes  $w_i$  et  $b$ . Le vecteur  $w$  possède un nombre de composantes égale à la dimension de l'espace d'entrée. En gardant cette formulation telle quelle, nous allons nous confronter aux mêmes problèmes que les méthodes classiques. Pour éviter cela, il est nécessaire d'introduire une formulation dite duale du problème fournissant la même solution que la primale mais dont la formulation est différente.

Le problème dual est obtenu en introduisant les multiplicateurs de Lagrange  $\alpha_i$ . La solution est fournie par un point-selle  $(w^*, b^*, \alpha^*)$  du Lagrangien. Il s'agit de faire rentrer les contraintes dans la fonction objective et de pondérer chacune d'entre elles par une variable duale [52, 53]. Dans le cas des SVM, il s'écrit [50, 54] :

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum \alpha_i [y_i (< w_i, x_i > + b) - 1] \quad (3.17)$$

Où les  $\alpha_i > 0$  sont les multiplicateurs de Lagrange associés aux contraintes.

Le point-selle vérifie en particulier les conditions :

$$\alpha_i^* [y_i (< w^*, x_i > + b^*) - 1] = 0 \quad \forall i \in \{1, \dots, n\} \quad (3.18)$$

Cette condition est appelée : condition d'optimalité de KKT de programmation quadratique associée aux SVM permettant de caractériser la solution du problème primal  $(w^*, b^*)$ . Ce théorème fondamental en optimisation mathématique, nous fournit une condition suffisante et nécessaire pour l'optimalité d'une solution dans le cadre de problèmes différentiables convexes. Les multiplicateurs de Lagrange  $\alpha^*$  sont associés au système d'équations suivant :

$$\begin{array}{ll} \text{stationarité} & w^* - \sum_{i=1}^n \alpha_i^* y_i x_i = 0 \quad \sum_{i=1}^n \alpha_i^* y_i = 0 \\ \text{complementarité} & \alpha_i^* (y_i (w^{*T} x_i + b^*) - 1) = 0 \quad i = 1, \dots, n \\ \text{admissibilité primale} & y_i (w^{*T} x_i + b^*) \geq 1 \quad i = 1, \dots, n \\ \text{admissibilité duale} & \alpha_i^* \geq 0 \quad i = 1, \dots, n \end{array} \quad (3.19)$$

Les conditions d'annulation des dérivées partielles du Lagrangien permettent de décrire les relations qui vérifient le plan optimal avec les multiplicateurs de Lagrange  $\alpha_i^* > 0$ , non nuls et strictement positifs, seulement par les points support [46, 50] :

$$w^* = \sum_{i=1}^n \alpha_i^* y_i x_i \quad \text{et} \quad \sum_{i=1}^n \alpha_i^* y_i = 0 \quad (3.20)$$

Remarquons que d'après cette formulation, on peut calculer  $w$  en fixant seulement  $n$  paramètres. L'idée va donc être de formuler un problème dual dans lequel  $w$  est remplacé par sa formulation [55, 51] :

$$\underset{w, b, \alpha_i}{\text{Maximiser}} \quad \frac{1}{2} \|w^2\| - \sum_{i,j=1}^n \alpha_i (y_i (w^T x_i + b) - 1) \quad (3.21)$$

$$\text{avec:} \quad w - \sum_{i=1}^n \alpha_i y_i x_i = 0 \quad \sum_{i=1}^n \alpha_i y_i = 0 \quad \text{et} \quad \alpha_i \geq 0 \quad \forall i = 1, \dots, n$$

De cette façon, le nombre de paramètres à fixer est relatif au nombre d'exemples de l'ensemble d'apprentissage et non plus à la dimension de l'espace d'entrée. Ces contraintes d'égalité après l'élimination de la variable primale  $w$ , permettent d'exprimer la formulation du Lagrangien [46] :

$$L(w, b, \alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle \quad (3.22)$$

Donc, nous pouvons formuler le problème dual des SVM séparable par *QP2* [50, 56] :

$$\underset{\alpha_i}{\text{QP2 Maximiser}} \quad W(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle \quad (3.23)$$

$$\text{avec:} \quad \sum_{i=1}^n \alpha_i y_i = 0 \quad \text{et} \quad \alpha_i \geq 0 \quad \forall i = 1, \dots, n$$

La résolution de ce problème dual permet donc de calculer le vecteur  $w$  à moindre coût. Cependant cette formulation ne fait à aucun moment apparaître le terme  $b$ . Pour calculer ce dernier, nous devons utiliser les variables primales [42, 46] :

$$b = - \frac{\max_{y_i=-1} (\langle wx_i \rangle) + \max_{y_i=1} (\langle wx_i \rangle)}{2} \quad (3.24)$$

L'équation de l'hyperplan optimal devient dans ce cas :

$$\sum_{i=1}^n \alpha_i^* y_i \langle x, x_i \rangle + b^* \quad (3.25)$$

Nous avons à présent tous les éléments nécessaires pour exprimer la fonction de décision de notre classifieur linéaire [56] :

$$f(x) = \sum_{i=1}^n \alpha_i y_i \langle x, x_i \rangle + b \quad (3.26)$$

Cette formulation correspond à l'hyperplan solution.

Notons qu'un grand nombre de termes de cette somme est nulle. En effet, seuls les  $\alpha_i$  correspondant aux exemples se trouvant sur la contrainte sont non nuls. Ces exemples sont appelés : « les vecteurs de support (VS), ou (Support Vectors en anglais) », terme que l'on pourrait traduire aussi par points-selles de support [56] qui sont les vecteurs  $x_i$  pour lesquels l'égalité :  $y_i ((w^*.x_i) + b^*) = 1$  est vérifiée et la contrainte est active. Concrètement, ce sont les points les plus proches de l'hyperplan optimal. Pour tous les autres exemples, on a donc  $\alpha_i^* = 0$ . La fonction de décision est calculée à partir d'exemples qui se trouvent sur la marge (figure 3.13).

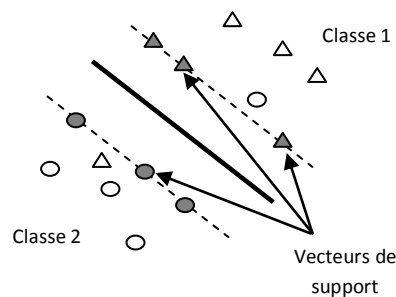


Fig. 3.13. Les vecteurs de support.

À ce point, remarquons deux choses. D'abord, que l'hyperplan solution ne requiert que le calcul des produits scalaires entre des vecteurs de l'espace d'entrée  $X$ . La solution ne dépend plus de la dimension  $d$  de l'espace d'entrée, mais de la taille  $m$  de l'échantillon de données et même du nombre  $m_c$  d'exemples critiques qui est généralement bien inférieur à  $m$ . En effet, plus la dimension de l'espace de description est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les exemples et les contre-exemples est élevée. En transformant l'espace d'entrée en un espace de redescription de très grande dimension, éventuellement infinie, il devient donc possible d'envisager d'utiliser la méthode SVM. En effet, il existe des fonctions bilinéaires symétriques positives  $K(x, y)$ , appelées : « fonctions noyau », faciles à calculer et dont on peut montrer qu'elles correspondent à un produit scalaire dans un espace de grande dimension.

- **Cas des données non séparables**

Pour la plupart du temps, les données d'apprentissage comportent du bruit (erreur d'acquisition, erreur sur la catégorie etc.). Par conséquent, même s'il existe une relation linéaire entre les données et leur catégorie, un classifieur linéaire pourrait commettre des

erreurs. Cependant, cela reviendrait à apprendre le bruit des exemples et donc à perdre une grande partie du pouvoir de généralisation (overfitting). Au lieu de cela, il paraît plus raisonnable d'admettre que certains exemples (supposés bruités) soient mal classés par notre classifieur ; on les appelle souvent des points aberrants (ou outliers) [42] (figure 3.14). Il est possible de suivre la même démarche par l'introduction de termes d'erreurs (ou écart)  $\xi_i$ . L'idée est de modéliser ces erreurs associées à chacune des observations  $(x_i, y_i)$ . Dans le cas où le point vérifie la contrainte de marge  $y_i(w^T x_i + b) \geq 1$ , la variable d'écart est nulle. On a donc les deux cas suivants :

$$\begin{array}{ll} \text{Pas d'erreurs} & y_i(w^T x_i + b) \geq 1 \Rightarrow \xi_i = 0 \\ \text{Erreur} & y_i(w^T x_i + b) < 1 \Rightarrow \xi_i = 1 - y_i(w^T x_i + b) > 0 \end{array} \quad (3.27)$$

Du point de vue de notre problème primal, cela revient à relaxer la contrainte imposant que tous les exemples soient bien classés comme suit [57] :

$$\begin{array}{ll} y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i & \forall i = 1, \dots, n \\ \xi_i \geq 0 & \forall i = 1, \dots, n \end{array} \quad (3.28)$$

On constate que si toutes les variables d'écart sont nulles, on retrouve le problème des SVM linéairement séparable. La résolution du problème précédent peut s'effectuer grâce à l'introduction d'un terme d'équilibrage  $C > 0$  fixé pour assurer que le nombre de points aberrants reste raisonnable, ce qui permet de poser le problème primal sous la forme suivante [56, 58] :

$$\begin{array}{ll} \underset{w, b, \xi}{\text{Minimiser}} & \frac{1}{2} \|w\|^2 + \frac{C}{d} \sum_{i=1}^n \xi_i^d \\ \text{avec:} & y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i \quad \forall i = 1, \dots, n \\ & \xi_i \geq 0 \quad \forall i = 1, \dots, n \end{array} \quad (3.29)$$

Où :  $C$ , est une constante et  $d$ , une puissance.

Géométriquement, la variable d'écart  $\xi_i$  divisée par  $\|w\|$ , correspond à la distance euclidienne prise perpendiculairement entre l'hyperplan canonique du côté de la catégorie de l'exemple et cet exemple. Notons que pour les exemples correctement classés,  $\xi_i$  est nul.

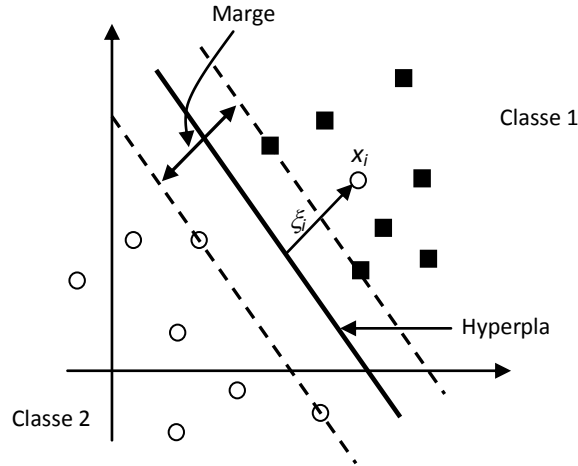


Fig. 3.14. Distance séparant le point aberrant de l'hyperplan canonique.

La constante  $C$  est souvent appelée constante « Trade-off », parce qu'elle permet d'indiquer l'importance que l'on accorde aux erreurs commises sur l'ensemble d'apprentissage par rapport au fait de maximiser la marge. Si on sait que les données d'apprentissage sont très bruitées, on accordera davantage d'importance à la marge en utilisant un  $C$  petit. Par contre, si l'intérêt se porte plutôt sur les résultats obtenus sur l'ensemble d'apprentissage, on utilisera un  $C$  de grande valeur.

La formulation que nous avons présentée du problème est souvent reprise sous la dénomination de « soft margin » dans la littérature [58]. Pour les mêmes raisons qu'auparavant, il est intéressant de dualiser le problème. On recherche un point-selle du Lagrangien :

$$L(w, b, \alpha, \beta) = \frac{1}{2} \|w\|^2 + \frac{C}{d} \sum_{i=1}^n \xi_i^d - \sum_{i=1}^n \alpha_i (y_i (w^T x_i + b) - 1 + \xi_i) - \sum_{i=1}^n \beta_i \xi_i \quad (3.30)$$

Avec des multiplicateurs de Lagrange  $\alpha_i \geq 0$  et  $\beta_i \geq 0$ .

Le calcul des gradients est identique par rapport à  $w$  et  $b$ . Dans le cas où  $d = 1$ , la dérivée partielle du Lagrangien par rapport aux variables d'écart s'écrit :

$$\partial_{\xi_i} L(w, b, \alpha) = C - \alpha_i - \beta_i. \quad (3.31)$$

Cette condition supplémentaire de stationnarité permet d'éliminer les  $\beta$  car :

$$\beta_i \geq 0 \text{ et } C - \alpha_i - \beta_i = 0 \Rightarrow \alpha_i \leq C \quad (3.32)$$

L'ensemble de ces conditions nous permet d'obtenir la formulation duale qui est le programme quadratique à résoudre pour obtenir la solution des SVM. Dans le cas où  $d=2$ , la

contrainte de positivité sur les variables d'écart est inutile. Le problème dual des SVM avec une variable d'écart appelée aussi  $L2$  SVM s'écrit [59] :

$$QP3 \quad \begin{array}{ll} \text{Maximiser} & W(\alpha) = \frac{1}{2} \alpha^T G \alpha - e^T \alpha \\ \text{avec} & \begin{cases} y^T \alpha = 0 \\ 0 \leq \alpha_i \leq C \quad \forall i = 1, \dots, n \end{cases} \end{array} \quad (3.33)$$

Avec :  $G$  la matrice symétrique  $n \times n$  de terme général  $G_{ij} = y_i y_j x_j^T x_i$ .

Ce problème reste un programme quadratique avec contraintes. Les  $\alpha_i$  représentant l'influence de chacun des exemples, sont maintenant bornés supérieurement par  $C$  qui représente l'influence maximale permise pour un exemple.

### 3.2. Les fonctions noyau et le cas non-linéaire

Dans le cas général la frontière optimale est non-linéaire. L'intérêt principal d'un classifieur à marge maximale de ce type réside dans le fait que l'on contrôle facilement la capacité et donc le pouvoir de généralisation. Naturellement, un grand nombre de jeux de données sont non-linéairement séparables. Pour classer ce genre de données on pourrait utiliser une fonction de décision non-linéaire. Géométriquement, cela reviendrait à avoir une (hyper) courbe qui marquerait la frontière entre les exemples positifs et négatifs. Ceci est fait par l'introduction de fonctions de décision non linéaires dites « *noyau* » (kernel en anglais).

$$\begin{array}{ll} \Phi : \mathcal{R}^N \rightarrow F \\ x \rightarrow \Phi(x) \quad \text{où } N \ll \dim(F) \end{array} \quad (3.34)$$

Ce qui permet d'obtenir un nouvel ensemble d'apprentissage :

$$(\Phi(x_1), y_1), \dots, (\Phi(x_l), y_l) \in F \times \{\pm 1\} \quad (3.35)$$

Une fois que les données sont modifiées, nous les utilisons pour faire l'apprentissage. Dans la figure (3.15), nous pouvons voir un exemple simple de cette transformation :

$$\begin{array}{l} k(x, x') = \langle x, x' \rangle^2 = (x_1 x'_1 + x_2 x'_2) = \langle \Phi(x), \Phi(x') \rangle \quad \text{avec } x \text{ et } x' \in \mathcal{R}^2 \\ \text{où } \Phi : \mathcal{R}^2 \rightarrow \mathcal{R}^3 \\ x = (x_1, x_2) \rightarrow \Phi(x) = (z_1, z_2, z_3) = (x_1^2, \sqrt{2}x_1 x_2, x_2^2) \end{array} \quad (3.36)$$

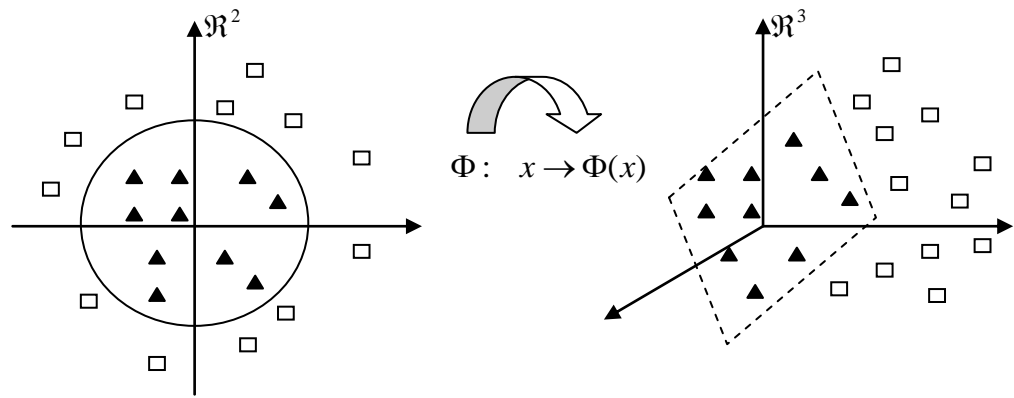


Fig. 3.15. Illustration de l'effet de changement d'espace (mapping) par une fonction noyau.

La dimension de l'espace de caractéristiques (feature space) est généralement très élevée. Cela ne pose pas de problème pour notre classifieur à marge maximale vu que sa formulation duale fixe le nombre de variables à déterminer en fonction de la taille de l'ensemble d'apprentissage (training set). Les nouveaux axes contiennent une surgénérations d'informations par rapport aux précédents. Ce qui permet idéalement d'effectuer une discrimination linéaire là où auparavant ce n'était pas possible.

#### ▪ Mesure de la similarité

De manière générale, il peut être utile de savoir à quel point un exemple est similaire à un autre. Pour ce faire, on utilise souvent en mathématique le produit scalaire qui moyennant une normalisation, correspond au cosinus de l'angle entre deux vecteurs. En utilisant le mapping  $\Phi$  (Changement de dimension), on peut définir une mesure de similarité dans le feature space. Dans ce cas, on montre qu'il existe une espace  $H$  et une fonction de mapping  $\Phi$  tels que :

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle \quad (3.37)$$

La fonction  $k(x, y)$  est appelée *noyau*.

Pour calculer l'hyperplan optimal dans le feature space, il suffit de remplacer toutes les occurrences du produit scalaire par le noyau. Plus généralement, tout algorithme d'apprentissage accédant exclusivement aux exemples au travers du produit scalaire (ou d'une grandeur qui en dérive) est dit kernelisable. Le produit scalaire lui-même peut être vu comme un noyau dont la transformation  $\Phi$  est l'identité. Le produit scalaire peut être très coûteux en temps de calcul étant donné que sa complexité est linéaire en la dimension de  $F$  et que cette dernière peut être très élevée. Cet inconvénient peut rendre le calcul de l'hyperplan optimal fastidieux. Remarquons que, si l'on peut déterminer une autre forme plus économique pour la

fonction noyau on peut se passer d'utiliser explicitement  $\Phi$ . En effet, le résultat du produit scalaire étant un réel, une autre fonction à image dans  $\mathfrak{R}$  peut être utilisée.

- **Condition de Mercer**

Une fonction  $k(.,.)$  symétrique est un noyau, si pour tous les  $x_i$  possibles, la matrice de terme général  $k(x_i, x_j)$  est une matrice définie positive c.-à-d. quelle définit une matrice de produit scalaire.

La matrice contenant les similarités entre tous les exemples du training set est :

$$G = \begin{pmatrix} k_{11} & k_{21} & \dots & k_{1n} \\ k_{21} & k_{22} & \dots & k_{2n} \\ \dots & \dots & \dots & \dots \\ k_{n1} & k_{n2} & \dots & k_{nn} \end{pmatrix} \quad (3.38)$$

Est appelée matrice de Gram du noyau.

- **Définition 3.3 (Matrice de Gram)** [59] : la matrice de Gram du noyau  $k(.,.)$  pour les observations  $(x_1, \dots, x_i, \dots, x_n)$  à  $n$  fini, est la matrice carrée  $G$  de taille  $n$  et de terme général  $G_{ij} = k(x_i, x_j)$ .

- **Définition 3.4 (Matrice définie positive)** [60] : un noyau  $k$  est définie positif si, pour tout entier  $n$  fini et pour toutes les suites de  $n$  observations possibles  $\{x_i, i = 1, \dots, n\}$ , la matrice de Gram de dimension  $n \times n$ , associée est une matrice dont les éléments sont des réels. Elle est symétrique définie positive SSI:

$$\forall v \in \mathbb{R}^n \quad v^T G v \geq 0 \quad (3.39)$$

Ce qui revient à exiger que toutes les valeurs propres de  $K$  soient positives.

L'intérêt des noyaux positifs c'est qu'il est possible de leur associer un produit scalaire. Les noyaux dits de Mercer sont les noyaux positifs définis sur un ensemble compact. Le théorème de Mercer fournissant une condition suffisante et nécessaire pour qu'une fonction soit un noyau.

- **Théorème 3.2 (Condition de Mercer)** [42]: la fonction :  $k(x, z) : X \times X \rightarrow \mathfrak{R}$  est un noyau si :

$$G = (k(x_i, x_j))_{i,j=1}^n \quad (3.40)$$

est définie positive et possède les trois propriétés fondamentales du produit scalaire :

- ✓ Positivité :  $k(x_i, x_j) \geq 0$ .
- ✓ Symétrie :  $k(x_i, x_j) = k(x_j, x_i)$ .
- ✓ Inégalité de Cauchy-Schwartz :  $|k(x_i, x_j)| \leq \|x_i\| \cdot \|x_j\|$ .

La condition de Mercer nous indique que si une fonction est un noyau mais nous n'avons aucun renseignement sur le mapping  $\Phi$  (et donc sur le feature space) induit par ce noyau, cette matrice contient toute l'information utilisée par les méthodes à noyaux sur les données pour l'apprentissage (figure 3.16).

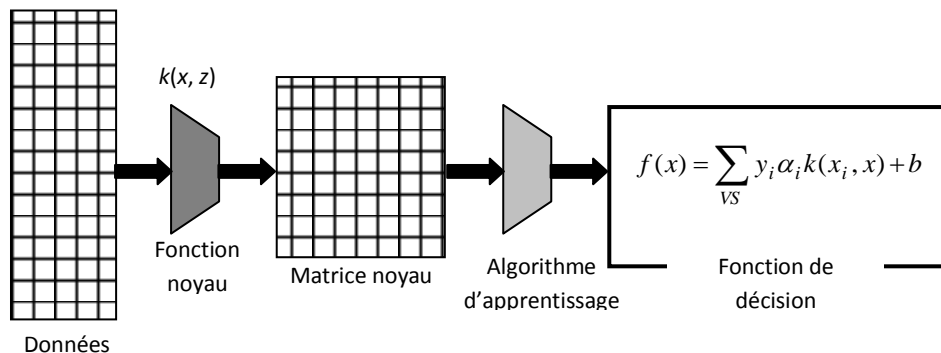


Fig.3.16. Chaîne de traitement générique des méthodes à noyaux.

La fonction de décision de la technique SVM dans ce cas, est [55, 61] :

$$f(x, \alpha) = \text{sign} \left[ \sum_{\text{Vecteurs de Support}} y_i \alpha_i k(x_i, x) + b \right] \quad (3.41)$$

▪ **Exemples de noyaux** [54, 58, 61] :

- **Linéaire** :  $k(x, y) = x \cdot y + c$
- **Polynomiale** :  $k(x, y) = (x \cdot y + c)^d$ , la puissance  $d$  est déterminée à priori par l'utilisateur.
- **Fonctions sigmoïdes** :  $k(x, y) = \tanh((a \cdot x \cdot y) - b)$ . Le théorème de Mercer n'est vérifié que pour certaines valeurs de  $a$  et  $b$ .
- **Fonctions gaussiennes RBF** :  $k(x, y) = \exp\left(-\|x - y\|^2 / 2\sigma^2\right)$ . L'écart type  $\sigma$ , commun à tous les noyaux, est spécifié à priori par l'utilisateur.

Dans le cas du noyau polynomial de degré  $d = 2$ , la transformation du vecteur  $x$  est :

$$(x_1, x_2) \xrightarrow{\Phi} (x_1^2, x_2^2, x_1x_2, x_2x_1) \quad (3.42)$$

Dans l'espace des caractéristiques, le produit simple entre deux vecteurs  $x$  et  $y$  devient :

$$\begin{aligned} (\Phi(x), \Phi(y)) &= (x_1^2, x_2^2, x_1x_2, x_2x_1)(y_1^2, y_2^2, y_1y_2, y_2y_1)^T \\ &= x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2 \\ &= (xy)^2 = k(x, y) \end{aligned} \quad (3.43)$$

La fonction noyau désirée est tout simplement le carré du produit scalaire dans l'espace d'entrée.

- **Composition de noyau**

Il est possible de composer de nouveaux noyaux en utilisant des noyaux existants, en prenant  $k_1$  et  $k_2$  des fonctions satisfaisant à la condition de Mercer. Soient  $a \in \mathfrak{R}^+$  et  $B$  est une matrice définie positive, alors les fonctions suivantes sont des noyaux :

$$\begin{aligned} k(x, z) &= k_1(x, z) + k_2(x, z) \\ k(x, z) &= ak_1(x, z) \\ k(x, z) &= k_1(x, z)k_2(x, z) \\ k(x, z) &= x^T B_z \end{aligned} \quad (3.45)$$

Les fonctions noyaux se substituent au produit vectoriel dans l'espace des caractéristiques et agissent comme une mesure de similitude non linéaire. Tous les algorithmes qui dépendent uniquement d'un produit vectoriel peuvent être dans une certaine mesure rendus non linéaire par l'utilisation de la fonction noyau. Tout ce qui a été dit pour les cas linéaires est applicable pour des cas non linéaires en appliquant une fonction noyau appropriée à la place du produit vectoriel euclidien.

## Conclusion

Ce chapitre a servi de donner en bref une description des SVM. Nous avons présenté une généralité sur les SVM, leur principe de fonctionnement, ainsi que des notions de bases essentielles concernant les SVM. Nous avons terminé notre chapitre par une description détaillée sur les SVM appliquées à la classification binaire.

L'étude comparative des performances des techniques PMC, RBF et SVM, constitue notre principal objectif. Une étude en simulation ayant pour but d'évaluation des performances de ces techniques appliquées à la regression et classification fera l'objet du chapitre suivant. L'évaluation des résultats reflétant les performances obtenues, nous conduira à la validation de notre choix de la méthode adaptée à notre application.

## **CHAPITRE IV**

### ***Simulation et évaluation***

#### **Introduction**

Ce dernier chapitre est consacré à la simulation et vise l'application des techniques étudiées comme étant une solution dans la conception de capteurs logiciels et de surveillance de la qualité de l'eau par reconnaissance de formes. L'objectif est de valider et d'évaluer les performances de chacune des méthodes présentées. Les exigences principales d'efficacité sont formulées sur deux points essentiels à savoir, les tests de spécification qui vérifient que le programme réalise bien la tâche pour laquelle il a été conçu, et les tests de performances qui vont servir à mesurer l'efficacité avec laquelle cette tâche est remplie. Afin de mener une étude comparative permettant un choix décisif de la méthode la mieux adaptée à l'application indiquée, on évaluera pour les méthodes exposées les paramètres liés au RMSE, EAM, temps d'apprentissage et coefficient de corrélation. Une discussion des résultats conclura cette étude de simulation pour choisir la technique la mieux adaptée. Un exemple d'application est prévu dans ce cadre. Le développement d'un capteur logiciel d'Oxygène Dissous (OD) vu comme un problème de régression. En fin de chapitre, la technique sélectionnée pour ses performances, est retenue pour une extension multi-classe où trois classes de la qualité de l'eau sont envisagées.

#### **1. problématique**

##### **1.1. Architecture du système de contrôle et de surveillance**

Il s'agit dans cette partie de travail d'évaluer les performances des trois techniques étudiées précédemment qui sont issues, rappelons-le, du domaine de l'intelligence artificielle à savoir, les PMC, RBF et les SVM. 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 un contrôle permanent de la qualité de l'eau. L'architecture de ce système imaginé est basée sur une approche multi-sensorielle et présentée dans la figure

4.1. Le processus de contrôle est vu comme un problème de reconnaissance de formes, où les classes correspondent aux différents états de l'eau, et les formes représentent l'ensemble des observations ou mesures des paramètres liés à ses caractéristiques.

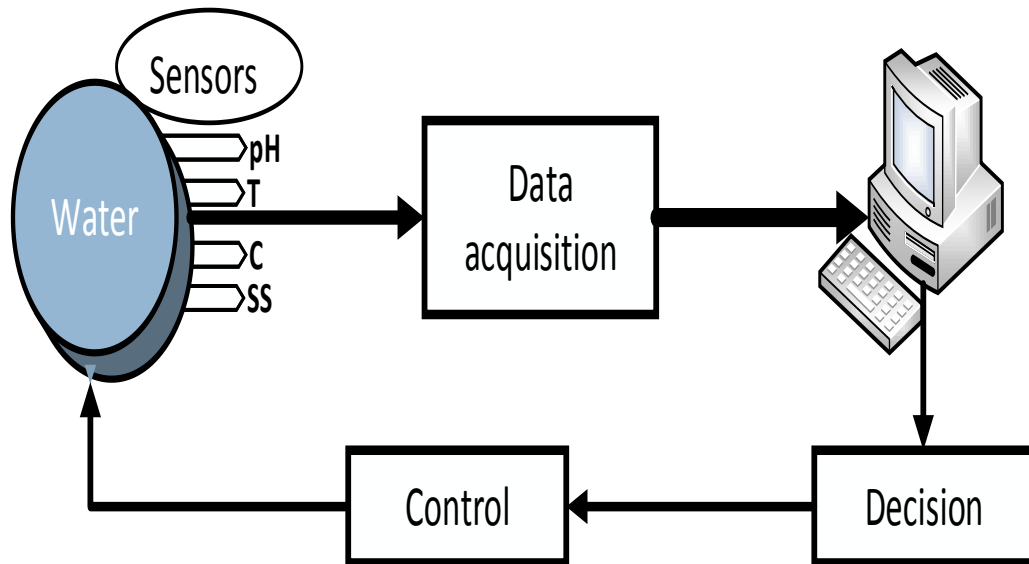


Fig.4.1. Architecture du système de contrôle et de surveillance.

Au niveau du système, on peut supposer que les différents paramètres physico-chimiques utilisées, tels que le pH, la Température ( $T^\circ$ ), la Conductivité (C), la Turbidité (TU), l'Oxygène Dissous (OD), etc. sont transformés en signaux électriques à partir des capteurs, et transmis vers une station de contrôle qui assure l'acquisition, le traitement et l'analyse. La technique de surveillance utilisée effectue après chaque acquisition, la classification et la séparation des données en plusieurs classes bien différentes. Une suite d'acquisitions pourrait être envisagée plusieurs fois par jour, sous des conditions prédéfinies. Un module d'apprentissage supervisé par un expert, permet de collecter de manière continue les paramètres relatifs aux différents états de l'eau pour la mise en œuvre d'une base de connaissance complète.

## 1.2. Approche utilisée dans la surveillance

La solution devant être adoptée par les techniques citées ci-dessus au problème de reconnaissance de formes posé, ne s'applique en fait que si on se trouve dans le cas d'un apprentissage supervisé. Nous procédons donc lors d'une étape préliminaire d'apprentissage, à paramétrer le classificateur pour la reconnaissance. L'étape de test ou de reconnaissance proprement dite, s'effectue une fois le modèle statistique établi. Il y a ici tout l'intérêt pour

dire que cette approche se caractérise par sa souplesse et sa généralité. A souligner toutefois que les méthodes de reconnaissance de formes à base d'apprentissage statistique sont les plus utilisées dans les systèmes de classification à fusion multisensorielle. En général l'apprentissage est une étape assez longue, et nécessite plus de temps de calcul. Les techniques partagent ce point commun mais diffèrent sur un certain nombre d'autres points. L'étude comparative effectuée dans les paragraphes suivants en fera la différence. Ce critère (temps d'apprentissage) aussi important dans le choix du modèle de reconnaissance, évoque un traitement hors ligne devant être effectué par le système de surveillance. Le déroulement de cette opération en permanence contribue sans doute à enrichir une base de connaissance qu'on veut qu'elle soit la plus complète possible pour le modèle de surveillance implanté. Le système de contrôle doit donc pouvoir marier à la fois une surveillance directe de l'eau et un apprentissage en arrière plan (en différé). Un opérateur (ou système) expert supervisant cet apprentissage permet de collecter de manière continue les paramètres relatifs aux différents états de l'eau.

## **2. Description des données d'entrée**

Nous cherchons à 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. Il y a dix paramètres physico-chimiques qui sont souvent utilisés dans plusieurs travaux [62, 63], et qui renseignent sur les dangers majeurs qu'il faut surveiller. Ces paramètres sont résumés comme suit : le pH, la Température ( $T^\circ$ ), la Conductivité (C), Chlorures ( $Cl^-$ ), Calcium ( $Ca^{2+}$ ), Magnésium ( $Mg^{2+}$ ), Bicarbonate ( $HCO_3^-$ ), Ms, OBD en plus de l'Oxygène Dissous (OD).

L'objectif qui se trouve derrière la collecte des données relatives à ces paramètres est de trouver un modèle de classification permettant de distinguer deux états bien distincts de l'eau (état potable et état non potable) dans un premier temps, et trois états différents 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. A noter que la turbidité, le pH, et l'OD sont fortement dépendants des phénomènes saisonniers. Il y a donc intérêt de disposer d'au moins une année pour archiver des données afin de déterminer une base de connaissance assez complète capable de

fonctionner normalement. D'où la nécessité d'une base de connaissance riche en informations exigeant d'abord une collecte des données sur une longue période, et la présence d'un expert.

Dans un but de validation, deux bases de données sont utilisées à savoir, une base de données simulée et une autre réelle. Pour la première, les différents paramètres descripteurs de l'état de l'eau sont générés par simulation selon les normes recommandées par les pouvoirs publics [64, 65]. Deux cas d'étude sont alors à envisager. Le premier consiste en une classification binaire comportant deux classes d'appartenance (eau potable, eau non potable), en respectant les normes de potabilité imposées. Le second est une extension multi-classe représentant un cas de classification plus élaboré, comportant trois classes de potabilité différentes et appliqué sur des données réelles. Le tableau 4.1 contient un résumé statique des paramètres d'eau brute.

Tableau 4.1. Résumé statistique des paramètres d'une eau brute.

Propriété	Température (°C)	Conductivité à 25 °C (ms/cm)	pH	DO	OBD	Sm	Calcium $Ca_2^+$ (mg/l)	Chlorures $Cl^-$ (mg/l)	Magnésium $Mg_2^+$ (mg/l)	Bicarbonate $HCO_3^-$ (mg/l)
Maximum	27,2	329	9,1	15	626	7,8	72	40	10,8	146
Minimum	0,1	11	6,8	7	0,5	0,3	2	0	0,7	8
Moyenne	11,89	146,68	7,65	10,53	22,27	3,27	17,32	11,23	5,10	69,05

Dans ce qui suit nous nous intéressons au développement d'un capteur logiciel OD pris comme exemple parmi plusieurs capteurs pouvant être imaginés et intégrés dans une chaîne de surveillance d'eau potable. L'intérêt n'est pas à discuter vu les grands avantages que peut apporter une telle conception. L'information issue de ce capteur devrait être utilisée au niveau de la chaîne de surveillance comme étant une variable d'entrée manquante (nouveau paramètre descripteur) dans la procédure d'analyse et de classification. Les résultats de prédiction ainsi obtenus pour ce paramètre, sont illustrés à travers une évaluation des performances pour les deux phases à savoir, l'apprentissage et la qualité de prédiction.

### 3. Développement du capteur logiciel : Oxygène Dissous

Dans cette partie de travail on s'intéresse à la conception d'un capteur logiciel d'OD permettant de fournir une information sur l' $O_2$  dissous à partir d'autres paramètres physico-chimiques issus de capteurs physiques. L'intérêt aussi bien économique que technique, permet d'assurer un fonctionnement permanent et stable de l'analyse et la surveillance de l'état de

l'eau. Le problème posé est vu comme un problème de régression, où la sortie du capteur logiciel correspond à la quantité d'oxygène dissous estimée.

Le but recherché quant au choix d'une méthode adaptée, se limite à une simulation appliquée sur des données réelles. Une évaluation des performances de ce capteur en matière de temps d'apprentissage, erreurs relatives, EAM et coefficient de corrélation, est effectuée. Les différents types de modèles de comportement possibles (PMC, RBF et SVM) possèdent l'avantage de pouvoir intrinsèquement décrire des relations non linéaires entre les variables d'entrée et celles de sortie d'un processus donné. Il reste alors à identifier les paramètres du modèle, une étape connue sous le terme d'apprentissage. En effet, les données de l'historique sont d'abord séparées en deux groupes : un groupe de données constituant la base d'apprentissage, et un groupe de données de généralisation (test) non utilisées lors de l'apprentissage, mais servant à « tester » le modèle lors de la phase de reconnaissance une fois celui-ci établi. Dans le but de montrer l'intérêt et l'efficacité des méthodes d'apprentissage statistique dans la conception de ce type de capteurs, trois modèles sont évalués et comparés dans ce qui suit, à savoir, PMC, RBF et SVM.

### 3.1. Apprentissage et Généralisation

Afin de procéder aux tests, on peut d'emblée supposer qu'en matière de reconnaissance, un taux de plus de 75 % est en général jugé satisfaisant [4]. On présente d'abord un ensemble de données extrait d'une base réelle et constitué de 774 échantillons ou vecteurs, correspondant aux neuf paramètres physico-chimiques (pH, T°, et C,..) avec un 10<sup>ème</sup> paramètre qui est l'OD [66]. L'ensemble des échantillons est séparé en deux, à savoir, 443 échantillons utilisés pour l'apprentissage, le reste (331 échantillons) est utilisé pour la généralisation. Notons toutefois que ces paramètres choisis (pH, T°, et C..) sont corrélés avec la sortie désirée qui représente le taux d'oxygène dissous. La corrélation entre paramètres E/S joue donc un rôle essentiel dans la conception de capteurs logiciels. La figure 4.2 montre l'évolution de ces dix paramètres descripteurs utilisés .

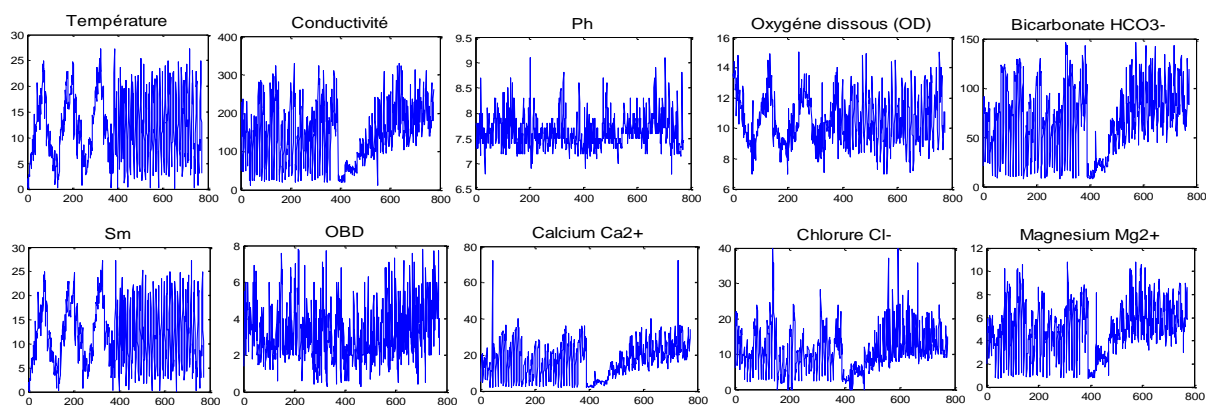


Fig.4.2. Evolution des dix paramètres descripteurs.

### 3.1.1. Modèle PMC

Comme les autres modèles, on a déterminé les paramètres tels que l'erreur d'entraînement (EQMA) ainsi que le temps d'apprentissage ( $t_{appr}$ ) correspondant aux PMC ; comme ce qui est montré dans le tableau 4.2.

Tableau. 4.2. Résultats d'apprentissage (Modèle PMC).

Base de données	$t_{appr}$ (s)	EQMA
443 Vecteurs	128.68	$3.4 \times 10^{-21}$

La figure 4.3 montre les résultats de prédiction ponctuelle obtenue sur l'ensemble d'apprentissage utilisé.

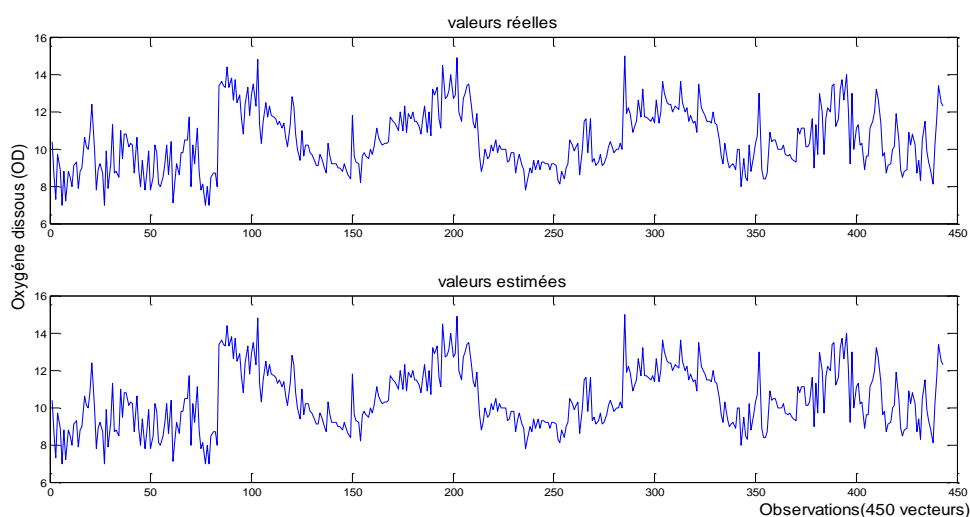


Fig.4.3. Résultats d'apprentissage du modèle PMC

Pour procéder maintenant à la généralisation, la même base de test (331 vecteurs) utilisée précédemment est appliquée. Les résultats de généralisation appliquée sur l'ensemble de test,

sont illustrés dans la figure 4.4. On montre clairement que la sortie calculée par le capteur logiciel et celle de la quantité d'oxygène réel sont nettement plus proches ( $R^2=0.919$ ).

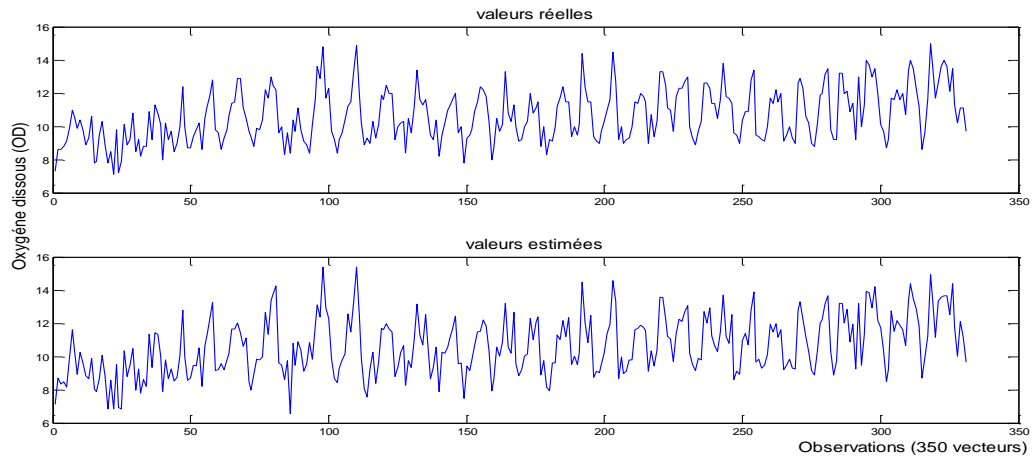


Fig.4.4.Résultats du test du modèle PMC

### 3.1.2. Modèle Neuronal-RBF

Pour ce modèle on a déterminé le temps d'apprentissage, l'erreur d'entraînement(EQMA) comme ce qui est illustré dans le tableau 4.3.on a aussi tracé les courbes des résultats d'apprentissage du modèle neuronal-RBF pour les deux phases (apprentissage et test). La figure 4.5 montre les résultats de prédiction ponctuelle obtenue sur l'ensemble des 443 vecteurs de la base d'apprentissage.

Tableau. 4.3 Résultats d'apprentissage (Modèle neuronal-RBF).

Base de données	t_appr (s)	EQMA
443 Vecteurs	110.01	$4.97 \times 10^{-28}$

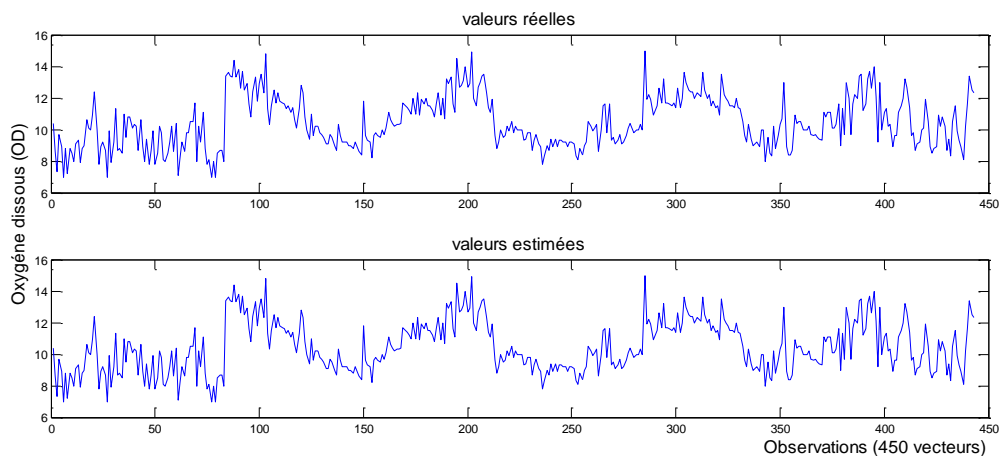


Fig.4.5. Résultats d'apprentissage du modèle neuronal-RBF.

On procède maintenant à la généralisation. Une base de test de 331 vecteurs est appliquée dans ce cas. La comparaison entre la sortie calculée par le réseau (valeurs estimées) et la quantité OD réellement obtenue (valeurs réelles) montre un net rapprochement (figure 4.6).

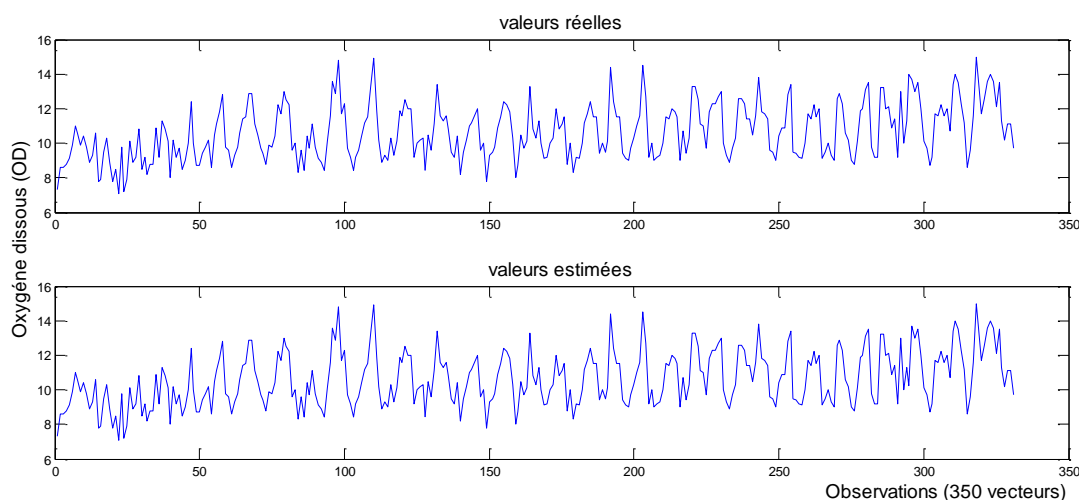


Fig.4.6. Résultats du test du modèle neuronal-RBF.

### 3.1.3. Modèle SVM

Les résultats d'apprentissage obtenus avec le modèle SVM utilisant la même base de 443 vecteurs que précédemment, sont présentés dans le tableau 4.4. Les paramètres tels que l'erreur d'entraînement (EQMA) ainsi que le  $t_{appr}$ , sont tous indiqués pour la valeur du facteur de pénalisation  $C = 150$ , avec fonction noyau RBF ayant les paramètres ( $\varepsilon = 0.01, \sigma = 0.7$ ).

Tableau. 4.4. Résultats d'apprentissage (Modèle SVM).

Base de données	$t_{appr}$ (s)	EQMA
443 Vecteurs	62.34	$1.26 \times 10^{-8}$

. La figure 4.7 montre les résultats de prédiction ponctuelle obtenue sur l'ensemble d'apprentissage utilisé.

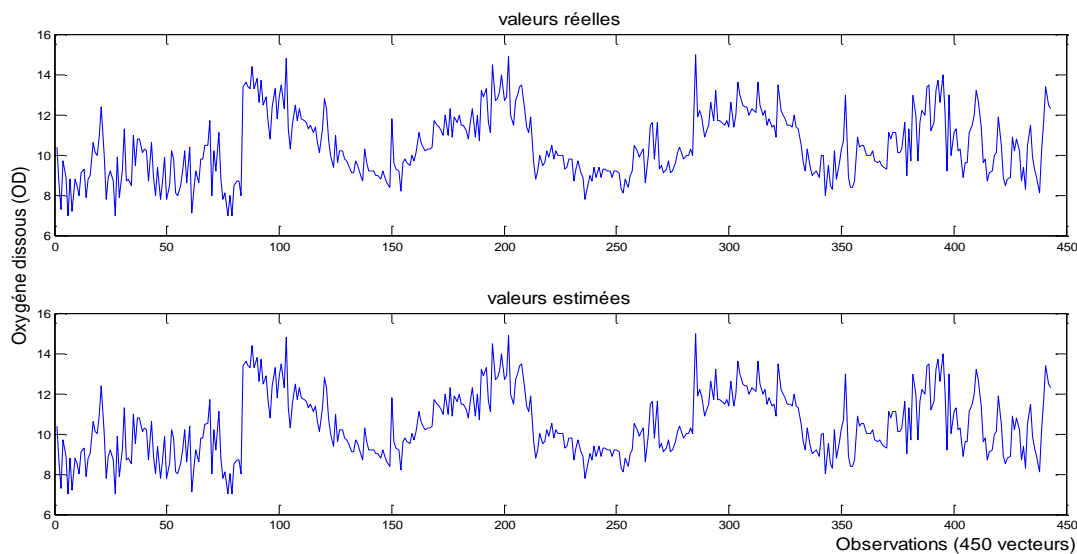


Fig.4.7. Résultats d'apprentissage du modèle SVM.

Pour procéder à la généralisation, la même base de test (331 vecteurs) utilisée précédemment est appliquée. Les résultats de généralisation appliquée sur l'ensemble de test, sont illustrés dans la figure 4.8. On montre clairement que la sortie calculée par le capteur logiciel et celle de la quantité d'oxygène réel sont nettement plus proches ( $R^2=0.98$ ).

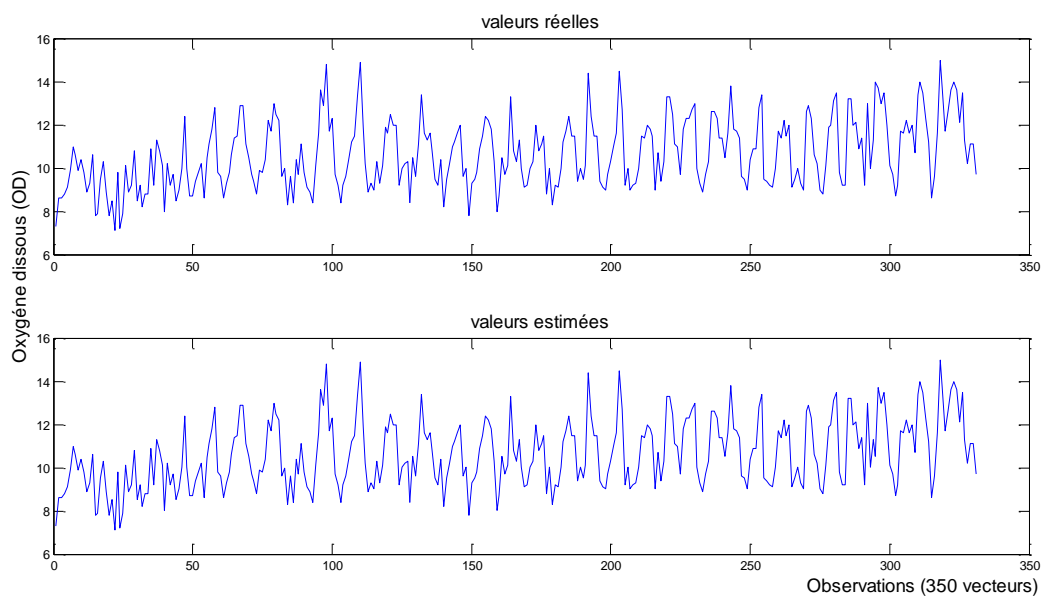


Fig.4.8. Résultats du test du modèle SVM.

### 3.2. Evaluation des performances

Pour une évaluation des performances des trois modèles précédemment testés, définissons quelques mesures statistiques.

Soient :

- Racine de l'erreur quadratique moyenne (RMSE) [67] :

$$RMSE = \left( \frac{1}{n} \sum_{i=1}^n e^2(i) \right)^{\frac{1}{2}} \quad (4.1)$$

- Erreur absolue moyenne (EAM) :

$$EAM = |e(i)| \quad (4.2)$$

- Erreur quadratique moyenne de généralisation (EQMG) :

$$EQMG = \frac{1}{n} \sum_{i=1}^n e^2(i) \quad (4.3)$$

Où :  $e(i) = y_r(i) - y_{e/c}(i)$ , tel que :  $y_r$ : sortie réelle,  $y_{e/c}$ : sortie calculée (apprentissage) ou estimée (test).

Les résultats statistiques obtenus pour les trois modèles testés sont récapitulés dans les tableaux 4.5, 4.6 et 4.7. On peut constater que la valeur moyenne des valeurs estimées de l'oxygène dissous est égale à celle calculée sur les données réelles. La bonne corrélation entre l'oxygène dissous estimé et réel pour les trois modèles est exprimée par les fortes valeurs des coefficients de corrélation.

Tableau. 4.5. Résultats statistiques des modèles PMC ,RBF et SVM.

	Exemples d'apprentissage					Exemples de test				
	Valeur moyenne (mg/l)	Ecart type (mg/l)	Minimum (mg/l)	Maximum (mg/l)	R <sup>2</sup>	Valeur moyenne (mg/l)	Ecart type (mg/l)	Minimum (mg/l)	Maximum (mg/l)	R <sup>2</sup>
<b>Réel</b>	10.446		7	15		10.647		7.100	15	
<b>PMC</b>	10.446	1.604	7	15	1	10.641	1.671	6.576	15.373	0.919
<b>RBF</b>	10.446	1.605	7	15	1	10.635	1.585	7.100	15	0.982
<b>SVM</b>	10.446	1.603	7	14.99	1	10.636	1.585	7.001	14.999	0.985

Tableau. 4.6. Performances des modèles PMC ,RBF et SVM.

	Exemples d'apprentissage			Exemples de test		
	RMSE	EQMA	EAM	RMSE	EQMG	EAM
<b>PMC</b>	$5.86 \times 10^{-11}$	$3.44 \times 10^{-21}$	$4.01 \times 10^{-11}$	0.062	0.003	0.043
<b>RBF</b>	$2.2 \times 10^{-14}$	$4.97 \times 10^{-28}$	$1.01 \times 10^{-14}$	0.207	0.043	0.011
<b>SVM</b>	$1.12 \times 10^{-4}$	$1.26 \times 10^{-8}$	$1.11 \times 10^{-4}$	0.191	0.036	0.010

Tableau. 4.7. Evolution de l'erreur relative des modèles PMC, RBF et SVM.

Erreur relative	Phase de test			
	0.05	0.1	0.2	0.25
<b>MLP</b>	78.54	94.56	99.39	99.69
<b>RBF</b>	99.69	99.69	99.69	99.69
<b>SVM</b>	99.69	99.69	99.69	100

### 3. 3. Discussion des résultats

D'après ces résultats, on trouve que le coefficient de corrélation entre les valeurs estimées et réelles de l'OD est supérieur à 0.91 ( $R^2 > 0.91$ ) ce qui est tout à fait satisfaisant. La précision de l'estimation est confirmée par l'absence de valeurs aberrantes. On peut constater que la moyenne et l'écart type des valeurs estimées et réelles montrées dans le tableau 4.5, sont proches les uns des autres, aussi bien que les valeurs minimales et maximales. L'évolution de l'erreur relative en phases d'apprentissage et de test pour les trois modèles, est montrée dans le tableau 4.7. L'erreur relative moyenne de la phase de génération calculée pour la technique PMC est 3.33%, pour le réseau Neuronal-RBF est de 0.08%, alors que celle calculée pour le modèle SVM est de 0.07%.

Les temps d'apprentissage pour les trois techniques (PMC, RBF et SVM) sont respectivement 128.68s ,110.01s et 62.34s. L'erreur d'entraînement reste relativement (EQAM) faible. On a aussi obtenu des valeurs faibles de RMSE et EAM sur l'apprentissage pour les trois techniques (PMC, RBF et SVM).

On constate que dans la phase d'apprentissage les trois techniques (PMC, RBF et SVM) présentent des coefficients de corrélations égales à 1 ( $R^2 = 1$ ) alors que pour la phase de génération la technique SVM donne un coefficient de corrélations égal à 0.985, ce qui est plus grand que ceux de deux autres techniques (PMC et RBF).

Tableau. 4.8. Tableau comparatif des caractéristiques des modèles PMC , RBF et SVM).

Caractéristiques	PMC	RBF	SVM
Base d'entraînement	443 vecteurs	443 vecteurs	443 vecteurs
Temps d'entraînement	128.68s	110.01s	62.34s
ERM	0.014	0.00034	0.00033
RMSE d'exemples de test	0.062	0.207	0.191
R <sup>2</sup> d'exemples d'apprentissage	1	1	1
R <sup>2</sup> d'exemples de test	0.919	0.982	0.985

## Conclusion

Ce chapitre a fait l'objet d'une étude en simulation concernant la mise en œuvre de trois techniques d'apprentissage statistique appliquées dans le domaine de surveillance des eaux potables avec la conception et la mise en œuvre d'un capteur logiciel en fonction des caractéristiques physico-chimiques de l'eau brute.

Afin de voir que cette méthode est adaptable à l'application indiquée ; une étude comparative dans le but d'un choix décisif de la méthode la mieux adaptée à l'application a été effectuée. La discussion des résultats obtenus des paramètres liés au temps d'apprentissage, à l'erreur d'entraînement, à l'erreur relative moyenne et au coefficient de corrélation des trois techniques, a permis d'opter pour la technique SVM retenue pour ses qualités et avantages adaptés au problème posé. Cette technique a fourni de très bons résultats de simulation au développement du capteur logiciel d'OD qui présente une solution satisfaisante pour notre problème de régression.

## *Conclusion générale*

Le travail présenté dans ce mémoire a été consacré à la mise en œuvre de trois techniques d'apprentissage statistique (PMC,RBF et SVM) appliquées à la reconnaissance de formes dans le domaine de contrôle et de surveillance des eaux propres. Cette étude découle des progrès technologiques importants qui ont été enregistrés ces dernières années, dans le but et l'intérêt d'une surveillance moderne et plus efficace de la qualité des eaux propres. A cet effet, notre modeste travail peut être considéré comme une contribution aux solutions proposées, pour résoudre des problèmes d'intérêt stratégique à préoccupation nationale, utilisant des outils modernes à base de techniques avancées.

Les divers dispositifs et outils de surveillance dans le domaine de l'eau existants actuellement par le monde, sont réalisés dans le but d'assurer une surveillance permanente et efficace de ces ressources. C'est dans l'esprit et l'intérêt considérable que présente la surveillance de la qualité de l'eau potable dans les usines de production et de distribution, que nous avons tenté dans ce travail d'exposer notre application. On veut bien que le système de surveillance proposé soit apte à contrôler de manière efficace et permanente cette ressource précieuse.

Cette étude a été structurée autour de quatre chapitres essentiels. Le premier a été consacré à une généralité sur les eaux propres, il a permis de présenter le cycle de l'eau ,leur chaîne de traitement ainsi que ses différents capteurs des paramètres descripteurs.Le deuxième chapitre concerne les réseaux de neurones artificiels(RNA).Dans ce chapitre nous avons essayé de donner en bref une description des réseaux de neurones artificiels et leurs différents types d'architectures et de modèle qui existent.Nous avons parlé d'une façon détaillée des deux structures usuelles des RNAs telles que : PMC et RBF.

Dans le troisième chapitre, nous avons donné et décrit un aperçu sur les mécanismes des machines à vecteurs de support(SVM) ,leurs architectures ,ainsi que les SVM appliquées à la classification binaire.

La simulation concernant la mise en œuvre de trois techniques (PMC,RBF et SVM) fondée sur ce type d'apprentissage et appliquée dans le domaine de contrôle et de surveillance des eaux potables a fait l'objet du quatrième et dernier chapitre.

Une étude comparative dans le but d'un choix décisif de la méthode la mieux adaptée à l'application a été effectuée. La discussion des résultats obtenus des paramètres liés au temps d'apprentissage, à l'erreur d'entraînement, à l'erreur relative moyenne et au coefficient de corrélation des trois techniques, a permis d'opter pour la technique SVM retenue pour ses qualités et avantages adaptés au problème posé. Cette technique a fourni de très bons résultats de simulation au développement du capteur logiciel d'OD qui présente une solution satisfaisante pour notre problème de régression.

Le domaine de surveillance des eaux propres avec les techniques d'intelligence artificielles est toujours dans la dynamique d'évolution pour améliorer les techniques d'acquisition et d'adaptation au bien tout les techniques de traitement de données.

Ceci met évidemment en jeu le nombre et le type d'exemples à utiliser dans la base d'apprentissage. Le temps correspondant à la phase d'entraînement reste relativement important, ce qui laisse envisager d'autres outils de calcul plus spécifiques DSP (Digital Signal Processor) un processeur spécialisé dans le traitement du signal afin d'améliorer ses performances et aboutir à des systèmes temps réel plus intelligents. Les horizons de l'application de ces techniques dans ce même domaine restent prometteurs. Dans un but préventif, son utilisation est envisageable dans le suivi permanent du degré de potabilité de l'eau. La précision de la décision peut aussi être améliorée davantage en ajoutant de nouveaux capteurs en entrée, des capteurs logiciels entre autres.

## Références

- [1] Les propriétés de développement social, L'eau potable, Document : ACDI, Agence canadienne de développement international.
- [2] A. ABD ERRAZEK, Surveillance des eaux propres à l'aide des techniques d'intelligence artificielle, thème de master, Université de M'sila.2013.
- [3] "Le cycle de l'eau, Observation régionale de l'environnement", Les Agences de l'eau Loire- Bretagne et Adour-Garonne, Poitou-Charentes, France, 2012.
- [4] LADJAL. Mohamed, Contribution au développement de systèmes de surveillance innovants dédiés au contrôle de la qualité des eaux potables, Thèse de Doctorat, Laboratoire LASS, Université de M'sila.2013.
- [5] Zahid Amjad and Peter G. Koutsoukos, "The science and technology of industrial water treatment: Mineral Scales and Deposits: An Overview", CRC Press, Taylor & Francis Group, Lubrizol Advanced Materials Inc., Brecksville, Ohio, USA, 2010.
- [6] Marie Cuq, "Eau : droit de l'homme ou marchandise ? Réflexion sur l'effectivité du droit de l'homme à l'eau", Human Rights Abuses in the Contemporary World: Legal approaches, Tbilissi : Géorgie, 2011.
- [7] F. Valiron, "Gestion des Eaux : principes - moyens – structures", Presses de l'école nationale des ponts et chaussées, Paris, France, 1990.
- [8] L. Zilliox, "Pollution et épuration des eaux", Conférence N° 287 de l'université de tous les savoirs, 13 octobre 2000, Paris, France.
- [9] W. Schon, K. Odeh, T. Denoeux et F. Fotoohi, "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, Montpellier, 2000, France, pp. 695-701
- [10] Jean-luc Celerier et Jean-Antoine Faby, "La dégradation de la qualité de l'eau potable dans les réseaux", Fonds National pour le Développement des Adductions d'Eau, Document technique N°12.

- [11] S. Merabet, "Evaluation de la qualité physico-chimiques des eaux brutes et distribuées du barrage réservoir de Beni haroun", Mémoire de Magister, Université de Constantine, Algérie, 2010.
- [12] "Système d'évaluation de la qualité des eaux souterraines", Rapport de présentation, Agences de l'eau et le Ministère de l'Ecologie et du Développement Durable, France, 2003.
- [13] 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.
- [14] E. Derwich, L. Benaabidate, A. Zian, O. Sadki, D. Belghity, "Caractérisation physico-chimique des eaux de la nappe alluviale de haut Sebou en aval de sa confluence avec Oued Fes", Larhyss Journal, Juin 2010, N° 08, pp. 101-112.
- [15] A. Lounnas, "Amélioration des procédés de clarification des eaux de la station Hamadi-Kroma de Skikda", Mémoire de Magister, Université de Skikda, Algérie, 2009.
- [16] R. Zemouri, Contribution à la surveillance des systèmes de production à l'aide des réseaux de neurones dynamiques, Application à l'e-maintenance, Thèse de doctorat, université de France, 2002.
- [17] N. VALENTIN Construction d'un capteur logiciel pour le contrôle automatique du procédé de coagulation en traitement d'eau potable, Thèse de doctorat, UTC/Lyonnaise des Eaux/CNRS, 2000.
- [6] H. Hernández, Supervision et diagnostic des procédés de production d'eau potable, Thèse de doctorat, L'INSA de Toulouse 2006.
- [18] A. Georges, les capteurs instrumentation industriel, Dunod. BORDAS, paris, 1987.
- [19] « <http://cbissprof.free.fr> ».
- [20] parmantier yves et kratz Frédéric, « capteurs –définitions, principes de detection », n° r400, 2009.
- [21] S. ZEGHLACHE, Commande Intelligente, cours, université Mohamed boudiaf-m'sila.
- [22] R. Zemouri, Contribution à la surveillance des systèmes de production à l'aide des réseaux de neurones dynamiques, Application à l'e-maintenance, Thèse de doctorat, université de France, 2002.

- [23] M. Y. Ammar, mise en œuvre de réseaux de neurones pour la modélisation de cinétiques réactionnels en vue de la transposition batch/continu, thèse de doctorat, Toulouse, France, 2007.
- [24] M. Bouzenita, Conception et implémentation d'un système expert hybride pour le diagnostic d'un système industriel, 2002.
- [25] K. Kasabov. Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering. London, England : The MIT Press, 1998.
- [26] Gilles Zwinglestein. Diagnostic des défaillances, Paris : Hermes, 1995.
- [27] M. Bourezgue, Combinaison de l'Analyse en Composante principale et les réseaux de neurones artificiels pour la prédiction des paramètres descripteurs de la qualité des eaux brutes, Mémoire de Master, Université de M'sila. juin 2012.
- [28] Marc Parizeau. Réseaux de neurones gif-21140 et gif-6432 université Laval 2004
- [29] A. Borgi , Apprentissage supervisé par génération de règles: le système SUCRAGE, Thèse de l'université de Paris 6 ,1999.
- [30] C. TOUZET, les réseaux de neurones artificiels, introduction au connexionnisme, cours, exercices et travaux pratiques, Juillet 1992.
- [31] R. Rakotomalala, Pratique de la régression linéaire multiple ,diagnostic et sélection de variables, cours, Université Lumière Lyon 2.
- [32] B. Abderezek, reconnaissance vocale basée sur les SVM, mémoire de master, Université Mohamed Khider – Biskra, 2011.
- [33] V. Vapnik, "The nature of statistical learning theory", Springer-Verlag, New York, USA, 1995.
- [34] Anis Ben Ishak, "Sélection de variables par les machines à vecteurs de support pour la discrimination binaire et multi-classe en grande dimension", Thèse de Doctorat, Université de Tunis, Tunisie, 2007.
- [35] R. Reyna Rojas, "Conception et intégration VLSI d'un système de vision générique, Application à la détection et la localisation d'objets à l'aide de "support vector machines"", Thèse de Doctorat, Institut National des Sciences Appliquées, Laboratoire LAAS/CNRS, N°02226, Toulouse, France, 2002.

- [36] Melle Z. Nadia, Application d'un apprentissage SVM pour le codage d'images hyper-spectrales, mémoire de magister, université mouloud Mammeri, Tizi-Ouzou, Algérie
- [37] Mohamad ally Hasan, Fomani Boris, « SVM machine à vecteurs de support ou séparateur a vaste marge ». Article. BD Web, ISTY3, Versailles St Quentin, France, janvier 2006.
- [38] Jeremy Mary, « Méthodes d'apprentissage avancées ». Présentation ppt. Centre National de la Recherche Scientifique. Université de Lille. janvier2006.
- [39] Antoine Cornuéjols, « Une nouvelle méthode d'apprentissage : les SVM. Séparateur à vaste marge ». Article. Université de Paris-Sud, Orsay. Juin 2002.
- [40] Miguel Gonzalez Mendoza, "Surveillance temps-réel des systèmes homme-machine : Application à l'assistance à la conduite automobile", Thèse de Doctorat, Institut national des sciences appliquées de Toulouse, France, 2004.
- [41] Y. Guermeur and H. Paugam-Moisy, "Théorie de l'apprentissage de Vapnik et SVM, support vector machines", In M. Sebban and G. Venturini, editors, Apprentissage Automatique, Edition Hermès, 1999, pp. 109-138.
- [42] J. Callut, "Implémentation efficace des machines à supports vecteurs pour la classification", Mémoire de grade de Maître en informatique, Université libre de Bruxelles, Belgique, 2003.
- [43] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, B. Scholkopf, "Support vector machines", IEEE Intelligent Systems and Their Applications, 1998, Vol. 13, pp. 18-28.
- [44] C. Junli, J. Licheng, "Classification mechanism of support vector machines", Signal Processing Proceedings WCCC-ICSP, 5th IEEE International Conference, 21-25 Aug 2000, Vol. 3, pp. 1556-
- [45] P. Mahé, "Noyaux pour graphes et Support Vector Machines pour le criblage virtuel de molécules", Rapport de stage, DEA en Mathématiques, Ecole Normale Supérieure de Cachan, France, 2003.
- [46] H. Majdoulayne, "Extraction de caractéristiques de texture pour la classification d'image satellites", Thèse de Doctorat, Université de Paul Sabatier, Toulouse III, France, 2009.

- 
- [47] J. Kharroubi, G. Chollet, "Nouveau système hybride GMM-SVM pour la vérification du locuteur", 24èmes Journées d'Étude sur la Parole JEP'2002, Nancy, 24-27 juin 2002, France, pp. 101- 104.
- [48] S. R. Gunn, "Support vector machines for classification and regression", Technical Report ISIS-1-98, Department of Electronics and Computer Science, University of Southampton, UK, 1998. 1559,
- [49] P. H. Chen, C. J. Lin, B. Schölkopf, "A tutorial on-Support Vector Machines", Applied Stochastic Models on Business and Industry, 2005, Vol. 21, pp. 111–136.
- [50] B. Schölkopf, A. J. Smola, "Learning with Kernels Support Vector Machines, Regularization, Optimization and Beyond", MIT Press, Cambridge, MA, USA, 2002.
- [51] Ludovic Mercier, "Les machines à vecteur support pour la classification en imagerie hyper spectrale ; implémentation et mise en œuvre", Travail d'étude et de synthèse technique en informatique, Grenoble, France, 2010.
- [52] A. Charef, A. Ghauch, P. Baussand, M. Martin-Bouyer Yin, "Water quality monitoring using a smart sensing system", Measurement, 2000, Vol. 28, pp. 219–224.
- [53] B. Efron and R.J. Tibshirani, "An introduction to the Bootstrap", New York : Chapman & Hall, USA, 1993.
- [54] Mathieu Ramona, "Classification automatique de flux radiophoniques par machines à vecteurs de support", Thèse de Doctorat, l'école télécom Paris Tech, France, 2010.
- [55] A. Cornuéjols, "Une nouvelle méthode d'apprentissage : Les SVM, Séparateurs à vaste marge", Bulletin de l'AFIA, N° 51, Université de Paris-Sud, Orsay, France, 2002.
- [56] C. J.C. Burges, "A toturial on support vector machines for pattern recognition. In Data Mining and Knowledge Discovery", Vol. 2, N° 2, Kluwer Academic Publishers Hingham, 1998, USA, pp. 121 167.
- [57] A. J. Smola, P. L. Bartlett, B. Schölkopf, and D. Schuurmans, "Advances in Large Margin Classifiers", MIT Press, Cambridge, MA, USA, 2000, pp. 147-169.
- [58] Jérôme Louradour, "Noyau de séquences pour la vérification du locuteur par machines à vecteur de support", Thèse de Doctorat, Université de Paul Sabatier, Toulouse III, France, 2007.

- [59] Philippe Besse, "Apprentissage statistique", Institut de Mathématiques de Toulouse, France, 2011.
- [60] Philippe Besse et Béatrice Laurent, "Apprentissage Statistique : Prévion et data mining", Institut de Mathématiques de Toulouse, France, 2012.
- [61] S. Canu, "Machines à noyaux pour l'apprentissage statistique", Techniques d'ingénieur, TE 5 255,2007.
- [62] 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.
- [63] H. R. Hernández 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.
- [64] <http://www.cg58.fr/>
- [65] <http://www.lenntech.fr/applications/potable/normes/normes-oms-eau-potable.htm>
- [66] Qualité des eaux superficielles, l'Agence de Bassin Loire-Bretagne, Orléans, France.
- [67] B. Yeganeh, M. Shafie Pour Motlagh, Y. Rashidi, H. Kamalan, "Prediction of CO concentrations based on a hybrid Partial Least Square and Support Vector Machine model", Atmospheric Environment, 2012, Vol. 55, pp. 357-365.

**Résumé :** L'eau est un élément essentiel dans la vie. Il est donc nécessaire de la contrôler, la surveiller pour la rendre potable. Dans ce cadre des efforts ont été effectués durant ces dernières années, pour développer des méthodes de contrôle et de surveillance automatique de l'eau. Ce travail s'inscrit dans ce courant de recherche, en proposant une nouvelle approche pour le contrôle et la surveillance de la qualité des eaux potables. Ce mémoire aborde le problème de la surveillance des eaux potables en fonction des caractéristiques physico-chimiques de l'eau brute à l'aide d'un modèle basé sur les techniques de l'apprentissage statistiques telles que : les PMC, RBF et SVM, appliquées au développement du capteur logiciel d'OD vu comme problème de régression et d'évaluer les performances de chacune après une simulation à l'aide du logiciel MATLAB pour avoir la technique adaptée entre elles.

**ملخص:** يعتبر الماء عنصرا أساسيا في الحياة ومن الضروري أن يراقب ويعالج حتى يصبح صالحا للشرب. وفي هذا المجال بذلت الجهود في السنوات الأخيرة بغية تطوير طرق المعالجة و المراقبة الآلية للمياه. الهدف من هذه الرسالة هو المساهمة في دراسة وتطوير أنظمة مبتكرة للتحكم والرصد لنوعية المياه الصالحة للشرب وذلك باستخدام تقنيات الذكاء الاصطناعي (Intelligence Artificielle) والتي يمكن أن نذكر منها: SVM وRNA. يأتي هذا العمل عن طريق المحاكاة باستخدام تقنيات التدريب الإحصائي كحل لتصميم الملتقطات المبرمجة والهدف من هذا هو التحقق من صحة وتقييم أداء التقنيات المعروضة.