

Université Mohamed Boudiaf - M'sila

FACULTE DE TECHNOLOGIE

DEPARTEMENT D'ELECTRONIQUE



Numéro de série :

Numéro d'inscription :

Mémoire

Présentée pour l'obtention du diplôme de

Master

Filière : Electronique

Spécialité : Instrumentation

THEME

Contribution à l'étude et l'évaluation des indices de qualité des eaux par l'élaboration d'un outil décisionnel intelligent

Présenté par

BRIK Sadjia & DECHOUCHA Ahlem

Soutenue le : / /

Devant le jury composé de :

<u>Nom & Prénom</u>	<u>Grade</u>	<u>Qualité</u>
DJERIOUI Mohamed	MCA	Président
LADJAL Mohamed	MCA	Encadreur
OUALI Mohammed Assam	MCA	Co- Encadreur
KHENNOUF Salah	MCB	Examinateur

Année universitaire : 2020/2021

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
الْحَمْدُ لِلَّهِ الَّذِي
خَلَقَ السَّمَوَاتِ وَالْأَرْضَ
وَالَّذِي يُضَوِّبُ الْمَوْتَى
إِنَّ رَبَّهُ لَسَدِيدٌ
إِلَىٰ عَرْشِهِ الرَّحِيمُ
الَّذِي يُخْرِجُ الْمَوْتَىٰ
وَيُدْخِلُهُمْ فِي الْأَرْوَاحِ
الْحَمْدُ لِلَّهِ الَّذِي
خَلَقَ السَّمَوَاتِ وَالْأَرْضَ
وَالَّذِي يُضَوِّبُ الْمَوْتَى
إِنَّ رَبَّهُ لَسَدِيدٌ
إِلَىٰ عَرْشِهِ الرَّحِيمُ
الَّذِي يُخْرِجُ الْمَوْتَىٰ
وَيُدْخِلُهُمْ فِي الْأَرْوَاحِ

Remerciements

Nous remercions ALLAH tout puissant de nos avoir accordé la volonté et le courage pour réaliser notre mémoire

Nos sincères remerciements et nos profonde reconnaissance vont a nos directeur de mémoire **DR. LADJAL Mohamed** Sans lui ce travail n'existerait tout simplement pas.

Nous tenons à lui exprimer ici toute nos gratitude pour nos avoir accueilli dans son groupe de recherche, et pour nos avoir permis de mener avec succès ce travail de mémoire. Ses compétences scientifiques, ses qualités humaines, sa disponibilité et son aide précieuse au quotidien en font un encadrant exceptionnel. Nous le remercions aussi pour l'autonomie qu'il nous avons accordé pendant nos travaux de recherche, tout en restant disponible, nos permettant ainsi de réaliser ce travail dans les meilleures conditions possibles. Nous remercions aussi le **C .ENCADREUR DR. MOHAMED**

ASSAM OUALI, Nous le remercions tout spécialement pour nos avoir encourage

Nous voudrions remercier aussi les membres du jury qui nos ont fait l'honneur d'examiner ce travail.

Nos profonds remerciements vont également à chef de département **Dr .E.BAKHTI** et à tous nos enseignants, ainsi que tous les membres du Département d'Electronique.

Nous remerciment vont enfin à toute personne ayant contribué de près ou de loin à la réalisation de ce travail.



Dédicace

Nous dédions ce modeste travail :

A nos très chers parents leurs Affections et leur
amour Que dieu les garde et les protège.

A nos très chers frères et sœurs.

A toute nos familles.

A tous les amis

Et collègues et à tous ceux qui connaissent la famille

BRIK et **DECHOUCHA**

SADJIA Et AHLEM



Table des matières

TABLE DES MATIERES.....	iii
LISTE DES FIGURES	v
LISTE DES TABLEAUX	vi
ABREVIATION ET NOTATION	vii
INTRODUCTION GENERALE	1

CHAPITRE I . CONTROLE ET PRODUCTION DES EAUX PROPRES

Introduction	5
1.1. Définition de L'eau propre	5
1.2. Analyse de l'eau potable.....	7
1.2.1. Les paramètres organoleptiques.....	7
1.3. Le problème d'eaux	8
1.4. Le traitement de l'eau	8
1.4.1. Prétraitement.....	8
1.4.2. Pré-oxydation	9
1.4.3. Clarification.....	9
1.5. LES CAPTEURS	11
1.5.1. Mesure des paramètres physico-chimiques	12
1.5.2. Capteurs physiques	12
1.5.3. Mesure de la conductivité.....	12
1.6. Importance de l'analyse et du traitement.....	12
1.7. La surveillance de la qualité des eaux.....	13
1.7.1. Les techniques avec modèle.....	13
1.7.2. Les techniques sans modèle.....	13
1.8. Les avantages et les inconvénients de la surveillance.....	14
1.9. Normes de la qualité des eaux	14
1.10. L'indice de qualité de l'eau (IQE).....	16
Conclusion	18

CHAPITRE II. LA TECHNIQUE SVM APPLIQUEE A LA REGRESSION

Introduction	19
2.1. LES MACHINES A VECTEURS DE SUPPORT APPLIQUEE A LA CLASSIFICATION	19
2.2. Principe.....	20
2.2.1. Support vectors (vecteurs support)	21
2.2.2. Marge	21
2.3. Principe des machines à vecteurs de support.....	21
2.3.1. Principes fondamentaux	22
2.2. LA TECHNIQUE SVM APPLIQUEE A LA REGRESSION	25
2.2.1. Définition générale de la régression	25
2.2.2. Régression par SVM	25
Conclusion	35

CHAPITRE III. SIMULATION

INTRODUCTION.....	37
1. PROBLEMATIQUE.....	37
1.1. Architecture du système de contrôle et de surveillance	37
1.2. Approche de la surveillance.....	38
2. MATERIELS ET METHODES	40
2.1. Présentation de la zone d'étude (Barrage de Tilesdit).....	40
2.1.1. La station de production d'eau potable TILES DIT	41
2.2. Méthodologie et données utilisées	42
2.2.1. Base de données.....	42
2.2.2. Calcul de l'Indice de qualité de l'eau (IQE)	47
3. PREDICTION D L'INDICE IQE PAR SVM.....	52
3.1. Apprentissage et Généralisation.....	53
CONCLUSION.....	55
CONCLUSION GENERALE	56
REFERENCES BIBLIOGRAPHIQUES.....	59

Liste des figures

CHAPITRE I. CONTROLE ET PRODUCTION DES EAUX PROPRES

Fig. 1.1. Chaîne de traitement des eaux de surface.	8
Fig. 1.2. Bassin de mélange	9
Fig. 1.3. Bassin de coagulation	10
Fig. 1.4. Bassin de floculation.....	10

CHAPITRE II. LA TECHNIQUE SVM APPLIQUEE A LA REGRESSION

Fig. 2.1. Les vecteurs de support.....	21
Fig. 2.2. Représentation de la marge.....	21
Fig. 2.3. Meilleur hyperplan séparateur.	22
Fig. 2.4. Séparateur a vaste marge	23
Fig. 2.5. Exemple de projection dans un espace de redescription.....	24
Fig. 2.6. La fonction d'erreur " <i>-insensible</i>	28
Fig.2.7. Tube d'insensibilité de rayon ϵ	28
Fig. 2.8. Architecture d'une machine à vecteurs de support (d'un nombre N).	32
Fig.2.9. Régression linéaire.....	33
Fig. 2. 10. Régression non linéaire rendue possible par la transformation en un problème de régression linéaire dans l'espace de redescription F.	34
Fig. 2.11. Dans la régression par SVM, au lieu d'imposer une marge entre les points des deux classes.....	35

CHAPITRE III. SIMULATION

Fig.3.1. Architecture du système de contrôle et de surveillance.	38
Fig. 3.2 : Carte géographique situant le barrage « Tilesdit » [Google Maps].....	40
Fig. 3. 3 : « Image représentant le site du barrage « Tilesdit » [Google Earth].	41
Fig. 3.4 : Image représentant le site de la station « Tilesdit » [Google Earth].....	42
Fig. 3.5. Evolution temporelle des paramètres descripteurs.....	42
Fig. 3.6. Diagramme en boîte à moustaches des paramètres physico-chimiques de l'année 2011. ..	44
Fig. 3.7. Modèle de prédiction de l'indice IQE par SVM.....	52
Fig.3.8. Résultats d'apprentissage du modèle SVM.	53
Fig.3.9. Résultats de test du modèle SVM.	54

Liste des tableaux

Tableau 1.1. Normes d'eau potable selon l'OMS (2006).....	15
Tableau 1.2. Valeur maximal standard de certains paramètres qui utiliser dans cette étude.....	16
Tableau 1.3. Classification et usage possible de l'eau selon l'IQE [25].....	18
Tableau. 3.1. Statistiques descriptives des paramètres physico-chimiques relatives aux douze mois de prélèvement de l'année 2011.....	45
Tableau. 3.2. Normes des paramètres descripteurs de l'eau brute.	46
Tableau. 3.3. Classification et usage possible de l'eau selon l'IQE (Brown et al, 1972 ; Chatterji et Raziuddin, 2002 ; Aher et al, 2016).	48
Tableau. 3.4 (1/2). Poids des paramètres physico-chimiques et norme Algérienne de qualité des eaux de surface (Norme Algérienne de qualité des eaux, 2011).....	49
Tableau. 3.4 (2/2). Poids des paramètres physico-chimiques et norme Algérienne de qualité des eaux de surface (Norme Algérienne de qualité des eaux, 2011).....	50
Tableau. 3.5. Valeurs de l'indice IQE et classe de qualité des eaux superficielles du barrage Tilesdit pendant les douze mois de l'année 2011.	51

ABREVIATIONS ET NOTATIONS

MES : d'éliminer les matières en suspension.

NTU : unité céphalométrique de turbidité(NTU).

PH : potentiel hydrogène.

OMS : l'organisation mondiale de la santé.

T° : la température.

C : la conductivité.

TU : la turbidité.

OD : l'oxygène dissous.

NO-2: Nitrite.

Ca²⁺: Calcium.

Cl-: Chlorures.

HCO₃⁻: Bicarbonate.

TH: Dureté totale.

CaCO₃ : Dureté permanente.

CaCO₃ : Titre alcalin complet.

IQE : L'indice de la qualité de l'eau.

WQI : water qualité index.

Wi : le poids de chaque paramètre selon son importance relative dans la qualité des eaux de boisson.

K : constante de proportionnalité.

ABREVIATIONS ET NOTATIONS

N: nombre de paramètres.

Si : valeur maximale de la norme standard algérien des eaux de surface de chaque paramètre.

Qi : échelle d'évaluation de la qualité de chaque paramètre.

Ci : la concentration de chaque paramètre en mg/l.

SVM : Support Vector Machines.

SVR : Support Vector Régression.

RMSE : Racine de l'erreur quadratique moyenne.

MAXPE : Maximum de l'erreur positive.

MAXNE : Maximum de l'erreur négative.

EAM : Erreur absolue moyenne.

EQMG : Erreur quadratique moyenne de généralisation.

RBF: Radial Basis Function.

H+: ions hydrogen's.

W : vecteur de pondération.

W_i: Poids associé à la caractéristique.

$\kappa(\mathbf{z})$: Fonction du noyau.

σ : Largeur du noyau.

$\xi(\mathbf{w})$: La vraie précision de leave-one-out.

λ : Paramètre de régularisation.

α : Longueur l'étape initiale.

η : Petite constante positive.

X_i: Vecteur caractéristique de dimension d.

Y_i: Étiquette de classe.

Introduction générale



INTRODUCTION GENERALE

L'eau est une ressource naturelle principale et indispensable pour la survie de l'humanité et de toute espèce animale ou végétale et pour l'environnement. Aucune substance ne peut remplacer l'eau, c'est pourquoi elle a besoin d'être protégée, traitée et économisée. Une eau est dite potable si elle respecte les valeurs imposées par la loi de potabilité traduite par des normes. La norme pour un paramètre dans l'eau est représentée par un chiffre, qui fixe une limite supérieure à ne pas dépasser et une limite inférieure à respecter [1] L'appréciation de la qualité des eaux de surface se base sur la mesure de paramètres physico-chimiques ainsi que sur la présence ou l'absence d'organismes et de micro-organismes.

Une des problématiques majeures de l'humanité est le risque de contamination et la mauvaise qualité de l'eau de consommation. La détérioration de la qualité de ce ressource vitale représente un grand problème sanitaire, les maladies transmises par l'eau causent la mort de plus d'un million et demi d'enfants chaque année. Pour cela, il faut effectuer un traitement pour protéger les consommateurs des microorganismes pathogènes et d'impuretés désagréables ou dangereuses pour la santé et assurer un approvisionnement de l'eau potable aux citoyens au niveau de leurs domiciles.

L'évaluation de la qualité de l'eau de surface peut être un processus complexe qui implique de multiples paramètres susceptibles de provoquer divers stress sur la qualité globale de l'eau. Pour l'analyse de la qualité de l'eau, différentes approches comme analyses statistiques, multi-facteurs de stress, les indices de qualité de l'eau,...etc. Ont été proposées. Les indices de qualité de l'eau (IQE) sont des outils pour déterminer les conditions de qualité de l'eau et, comme tout autre outil exige la connaissance des principes et des concepts de base de l'eau [2], c'est une méthode bien connue d'exprimer la qualité de l'eau qui offre une unité stable et reproductible de mesure qui répond aux évolutions dans les principales caractéristiques de l'eau [3]. L'importance de l'IQE peut être facilement appréciée du fait que les ressources en eau jouent un rôle crucial dans l'environnement général et cet indice a également été reconnu comme l'un des 25 indicateurs de performance environnementale de la

holistique environnemental Performance Index (PEI) [4]. Les indices de la qualité de l'eau appelés communément WQI (water quality index), ont fait l'objet d'un grand nombre d'étude ces trois derniers décennies, mais le concept en sa forme rudimentaire date de plus de 150 ans, C'est en Allemagne en 1848 que les premiers indicateurs environnementales ont été utilisés. Au départ c'était la présence ou l'absence de certains organismes dans l'eau qui a été utilisée comme indicateur des différentes aptitudes de l'eau [5]. Dans la même période, l'importance de la qualité de l'eau pour la santé publique a été reconnue au Royaume-Uni (1854) [6]. En 1965 Horton a donné la première formulation de l'indice de qualité de l'eau (Horton 1965), où il a tenu compte de la réduction des variables et de leurs fiabilités (pour faciliter la manipulation de l'indice) et de la signification des sites de prélèvement. De nombreux auteurs ont par la suite donnés de nombreuses formulations à l'indice de qualité et qui ont amené des changements à la méthodologie en fonction des objectifs fixés par l'étude. On peut citer ceux de : Prati et al (1971), Deininger et Landwehr (1971), Dinius (1972), Walski et Paker(1974), indice de Stoner(1978), indice de Smith(1990) et indice de qualité de l'eau du conseil canadien des ministres de l'environnement(2001).

Durant ces dernières années, d'importants efforts ont été déployés dans le développement de méthodes de contrôle et de surveillance automatique de procédé et de leurs installations. Ces méthodes peuvent être classées en deux grandes catégories : celles qui se basent sur l'existence d'un modèle formel, et celles qui se basent sur l'analyse des variables, ainsi que sur les connaissances à priori d'un expert humain. L'inconvénient de la première catégorie est l'existence d'incertitudes du modèle physique, qui ne peut prendre en considération tous les paramètres nécessaires de surveillance. La seconde, plutôt divisée en deux classes, correspond aux outils de traitement du signal qui sont généralement qualifiés d'outils de traitement de bas niveau, et celle dite de haut niveau, dont les outils sont plutôt orientés vers la communication avec un opérateur expert [18]. Celle-ci représente les techniques de l'intelligence artificielle (IA) qui servent comme outil de base pour l'aide à la décision. Leur réponse est plus élaborée et peut être obtenue soit à partir de données brutes venant directement des variables de surveillance, ou à partir de données traitées venant des sorties de traitements de bas niveau. Il est judicieux de supposer que le problème de contrôle et de surveillance de la qualité de l'eau peut être vu comme un problème de reconnaissance de formes, où les formes représentent l'ensemble des observations ou mesures liées aux caractéristiques de l'eau, et les sorties correspondent aux différents états de l'eau.

En effet, et pour de nombreuses raisons, il est souvent indispensable de disposer un modèle d'aide à la décision, dont la fonction consiste à fournir à l'utilisateur une information complète, telle qu'un paramètre physico-chimique non mesurable en continu (ou à un coût prohibitif), sous forme d'un indice global de la qualité de l'eau. Ces informations sont ensuite utilisées comme des données d'entrée dans la procédure de diagnostic de la partie amont d'une station de production d'eau potable. La construction d'un tel modèle peut être vue aussi comme un problème d'apprentissage de relations entre variables à partir de données observées. Ils sont spécialement programmés avec des algorithmes dérivés de l'intelligence artificielle. Le problème devient dans ce cas là, comme un problème de régression. Parmi les techniques d'IA utilisées, on trouve les machines à vecteurs de support (SVM). Celle-ci se démarque des autres outils par leur capacité d'apprentissage et de généralisation, notamment dans les applications de grande dimension. Cette technique peuvent être utilisées en raison de leurs robustesses et de leurs capacités à tenir en compte de la nature dynamique et complexe du procédé. Cette technique est de plus en plus acceptée dans l'industrie de production d'eau potable en tant qu'outil de modélisation et de contrôle de procédés.

Le développement d'un système de surveillance de type « machine - environnement » est un problème complexe. Le modèle mathématique d'un tel système est quasiment impossible à construire à cause de ses caractéristiques dynamiques et stochastiques. Comme c'est souligné auparavant, le problème est vu comme un problème de reconnaissance de formes (régression) à partir de différentes formes représentant l'ensemble des observations ou mesures des paramètres liés aux caractéristiques physico-chimiques de l'eau. Au niveau du système de surveillance imaginé pour cette situation, les différents paramètres physico-chimiques utilisés dans l'analyse de l'eau, peuvent être transformés en signaux électriques et transmis vers une station de contrôle qui assure l'acquisition et le traitement des données reçues. La technique devant être utilisée avant la décision du système, doit pouvoir effectuer une décision globale spécifiant la qualité de l'eau.

Le travail réalisé est axé autour de trois chapitres qui sont présentés comme suit :

Le premier chapitre est une description générale non exhaustive du vaste domaine de contrôle et de traitement des eaux brutes. Des généralités sont présentées à propos de cette ressource naturelle, l'eau, ainsi que les outils et moyens mis en œuvre pour son traitement. La chaîne de production et de traitement d'eau potable la plus courante, est de ce fait décrite. Les différentes données et mesures nécessaires pour établir une méthodologie d'analyse basée sur l'interprétation des informations obtenues, sont décrites. Nous détaillons les principaux

paramètres descripteurs de l'eau, en particulier, ceux relatifs à la qualité de l'eau potable et les formulations nécessaires pour la formulation de l'indice de qualité à travers les paramètres descripteurs.

Dans le deuxième chapitre, nous détaillons les fondements théoriques de la méthode SVM. La formulation générale de la méthode SVM appliquée à la régression, est aussi décrite.

Le troisième et dernier chapitre est consacré à la simulation et vise l'application de la technique étudiées précédemment comme étant une solution dans la conception de modèle de prédiction de l'indice de la qualité de l'eau (IQE) par reconnaissance de formes. Dans ce cadre, un modèle de prédiction utilisant les techniques de l'intelligence artificielle comme étant une solution de rechange les calcul traditionnels de l'indice de qualité de l'eau pour des raisons d'usage, du temps et les efforts nécessaires au calcul et de coût, sont présentés. Le barrage de Tilesdit (Bouira – Algérie) est considéré comme une zone d'étude de ce modeste travail.

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

CHAPITRE I

CONTROLE ET PRODUCTION DES EAUX PROPRES

Introduction

L'eau à l'état naturel est souvent impropre à la consommation. Selon la qualité de l'eau prélevée, différentes étapes de traitement peuvent être nécessaires pour rendre l'eau potable et maintenir sa qualité jusqu'au robinet du consommateur. Dans ce chapitre, nous définissons l'eau en mentionnant ses caractéristiques chimiques et physiques et ses différents types, les séries de production et de traitement de l'eau potable sont présentées et à travers cela, nous avons également passé en revue les méthodes les plus importantes de contrôle industriel. L'évaluation de la qualité de l'eau de surface peut être un processus complexe qui implique de multiples paramètres susceptibles de provoquer divers stress sur la qualité globale de l'eau. Les indices de qualité de l'eau sont des outils pour déterminer les conditions de qualité de l'eau et, comme tout autre outil exige la connaissance des principes et des concepts de base de l'eau et les questions connexes. Dans ce sens, une description de cet indice de qualité en mentionnant les normes, règles et lois mathématiques utilisées dans le calcul, est décrite. Nous présentons les différentes étapes de calcul de cet indice comme élément clés dans la décision de la qualité de l'eau dans notre application

1.1. Définition de L'eau propre

L'eau est une substance transparente, incolore et inodore, c'est le principal composant des cours d'eau, des lacs, des mers et des océans, ainsi que des fluides dans tous les organismes vivants, et c'est le composé chimique le plus répandu à la surface de la terre. Une molécule d'eau est constituée d'un atome d'oxygène central, auquel deux atomes d'hydrogène sont liés par une liaison covalente de sorte que sa formule chimique est H_2O . Donc, l'eau est quelque chose de spécial, Chaque vivante sur terre (micro-organismes, plantes, animaux, êtres humains et même notre cerveau) consiste principalement en eau [7].

Le monde contient aussi beaucoup d'eau .plus que 70% de la totalité de l'eau contenue sur terre, seulement une petite partie qui est convenable pour la consommation humaine ou l'usage agricole (approximativement 0.5% de toute l'eau dans le monde) cette petite fraction d'eau douce doit pourtant satisfaire l'ensemble des besoins de l'humanité. Pour cela, la nécessité d'un traitement adéquat des eaux usées par lequel de contribuer à l'amélioration des performances du traitement [9].

1.1.1. Types des eaux propres

a). Eaux souterraines

De point de vue hydrogéologique les couches aquifères se divisent en :

- Nappes phréatiques ou alluviales : Peu profondes et alimentées directement par les précipitations pluvieuses ou les écoulements d'eau en dessus,
- Nappes captives : Plus profondes que les premiers et séparées de la surface par une couche imperméables, l'alimentation de ces nappes est assurée par l'infiltration sur leurs bordures.

Quand une eau souterraine contient une concentration en certains minéraux dépassant les normes de potabilité, mais elle représente des propriétés thérapeutiques on la distribue en bouteilles avec parfois un traitement bien défini, ces eaux sont dites eaux minérales [7].

b). Les eaux de surface

Les eaux de surface sont constituées par les fleuves, les lacs, les rivières, barrages, réservoirs de stockage (citerne de récolte d'eau de pluie), elles sont fréquemment contaminées elles se caractérisent par une forte charge en impuretés et par une pollution biologique surtout chimique cette pollution est surtout dû aux rejets d'eaux usées dans la nature qui détient des polluants toxique [8].

c). Eau de consommation

Ce sont les eaux destinées à la consommation domestique, elles ont connues une énorme croissance suite au développement démographique et à l'amélioration des conditions de vie des populations. La consommation domestique en eau varie de quelques litres par jour dans les pays sans adduction publique et à faible confort ménager jusqu'à plusieurs centaines de litres dans les pays très développés [7].

1.2. Analyse de l'eau potable

1.2.1. Les paramètres organoleptiques

Odeur

Une eau destinée à l'alimentation doit être inodore. En effet, toute odeur est un signe de pollution ou de la présence de matières organiques en décomposition [8].

Couleur

La couleur de l'eau provient de matières organiques, comme par exemple les substances humiques, les tanins mais également les métaux comme le fer et le manganèse ainsi que les résidus industriels fortement colorés [8].

1.2.2. Les paramètres physico-chimiques

Les caractéristiques physico-chimiques sont des paramètres qui distinguent l'eau propre [9] :

Le pH

Le pH est un indicateur de pollution qui mesure la nature de l'eau. Les eaux naturelles ont un pH voisin de 7, le plus souvent compris entre 6 et 8. Plus le pH est bas, plus la

solution est dite acide. Plusieurs espèces de poissons et autres organismes aquatiques ne peuvent pas supporter une eau trop acide [10].

La température

C'est une caractéristique physique importante, elle joue un rôle dans la solubilité des sels et surtout des gaz, dans la détermination du pH pour la connaissance de l'origine de l'eau des mélanges éventuels [11].

La conductivité

La conductivité mesure la capacité de l'eau à conduire le courant électrique. Ce paramètre donne une indication de la concentration totale de l'eau en ions [12].

Les chlorures

Les chlorures sont présents en grande quantité dans l'eau de mer. Leur concentration dans l'eau de pluie est approximativement de 3mg/l. La teneur en chlorures d'une eau dépend de l'origine de l'eau et de la nature du terrain qu'elle traverse. Ils participent à la conductivité électrique des cours d'eau [13].

1.3. Le problème d'eaux

Durant les dernières années, la qualité des eaux de surface s'est fortement dégradée en raison des activités anthropiques telles que l'agriculture, le changement des usages du sol, l'imperméabilisation des surfaces à l'origine d'un ruissellement plus important et plus chargé en polluants, les rejets d'eaux usées non traitées, et pour une part difficile à évaluer, le changement climatique qui influence la température et les caractéristiques chimiques des plans d'eau.

D'autre part, le problème de qualité de l'eau le plus préoccupant est l'eutrophisation, résultat d'une forte concentration en nutriments (principalement phosphore et nitrogène), ce qui réduit significativement l'exploitation utile de l'eau. [53]

1.4. Le traitement de l'eau

Le traitement de l'eau est un terme pour décrire les processus qui ont lieu pour rendre l'eau, après qu'elle a été traitée, adaptée à un usage spécifique. Cela comprend son utilisation comme eau potable, ainsi que dans des usages industriels, médicaux et autres. En général, l'objectif principal du traitement de l'eau est d'éliminer ou de réduire tout plancton ou polluants afin que l'eau soit adaptée à son objectif. [10]

Cette chaîne de traitement a été conçue à la base pour la réduction de la turbidité en éliminant toutes sortes de matières en suspension. La chaîne de traitement conventionnelle des eaux de surfaces destinées à la consommation comprend dans la plupart du temps les étapes suivantes :

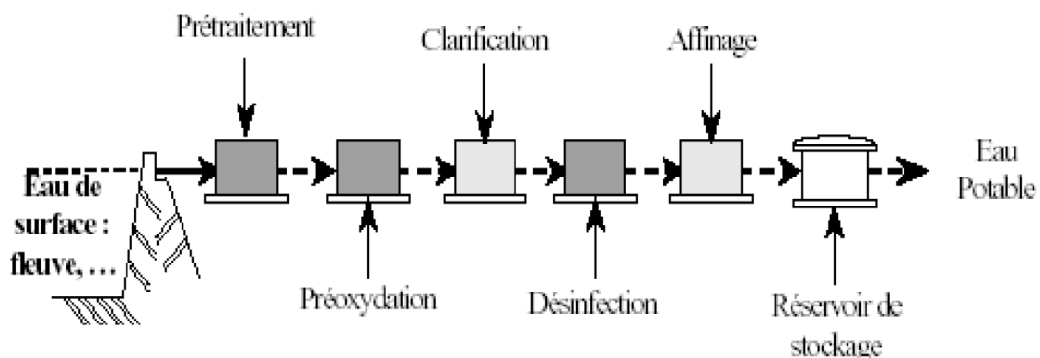


Fig.1.1.Chaîne de traitement des eaux de surface.

1.4.1. 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, les prétraitements sont principalement de deux types [14] :

- le dégrillage,
- le tamisage.

1.4.2. Pré-oxydation

A l'issue du prétraitement, on à une eau relativement propre mais qui contient encore des Particules colloïdales en suspension, La pré-oxydation se fait à l'entrée du bassin de mélange par l'injection d'hypochlorite de sodium d'une concentration de 48°. Cette opération sert à la destruction des germes pathogènes et à l'oxydation des algues et les matières organiques et métaux lourds (fer, manganèse...etc.)[14].



Fig. 1.2. Bassin de mélange

1.4.3. Clarification

La clarification est l'ensemble des opérations permettant d'éliminer les matières en Suspension (MES) d'une eau brute ainsi que la majeure partie des matières organiques. La clarification comprend les opérations de coagulation, de floculation et de filtration.

Coagulation

La coagulation est un processus qui consiste à neutraliser les charges portées par les substances colloïdales ou dissoutes indésirables à l'aide d'un produit chimique de charge opposée, appelé coagulant avec une agitation rapide, afin de faciliter leur agglomération en flocons décantables ou filtrables.



Fig. 1.3. Bassin de coagulation

Floculation

La floculation est une opération complémentaire à la coagulation. Elle vise à favoriser la croissance de flocons par une agitation lente et prolongée de l'eau provenant des bassins de coagulation.



Fig. 1.4. Bassin de floculation.

La filtration

La filtration est un procédé de séparation permettant de séparer les constituants d'un mélange qui possède une phase liquide et une phase solide au travers d'un milieu poreux. La filtration peut aussi désigner le phénomène passif d'épuration naturelle ou de diminution de la turbidité qui se produit quand l'eau pénètre un lit de sable ou de sédiment pour rejoindre la nappe [15].

- La filtration rapide

C'est la technique la plus répandue. Elle sera d'autant plus efficace que les grains de sable seront fins. C'est pourquoi elle s'est imposée, en raison des énormes volumes d'eau à filtrer [16].

- La filtration biologique lente

C'est une méthode d'épuration biologique qui consiste à faire passer l'eau à traiter à travers un lit de matériau filtrant à une vitesse de 0.1 à 0.2 m/h. le matériau filtrant le plus approprié est le sable. Au cours de ce passage, la qualité de l'eau s'améliore considérablement par la diminution du nombre de micro-organismes (bactéries, virus, kystes), par l'élimination de matières en suspension et colloïdales et par des changements dans sa composition chimique [17].

Affinage

Ensemble des procédés de traitement de l'eau par ozonation, filtration sur charbons actifs ou filtration sur membranes, pour oxyder et biodégrader les matières organiques et certains micropolluants par l'élimination absorption .les qualités organoleptiques de l'eau s'en trouvent améliorée [18].

1.5. LES CAPTEURS

Le contrôle permanent et automatique des processus de traitement implique la mesure d'un certain nombre de paramètres à l'aide de capteurs en continu. Ceux-ci peuvent se classer en deux grandes familles: les paramètres usuels et les paramètres spécifiques de l'eau.

La mesure en continu d'un certain nombre de paramètres spécifiques permet de libérer l'opérateur de l'astreinte d'analyse de routine et d'optimiser les traitements en réduisant les temps de réponse. Elle porte principalement sur la turbidité, la conductivité, le pH et la concentration de certains corps dissous. La turbidité est la réduction de la transparence d'un liquide due à la présence de matières en suspension non dissoutes. Elle permet de connaître le

degré de pollution physique des eaux à traiter ainsi que la qualité d'une eau livrée à la consommation humaine. [17]

Les paramètres usuels sont principalement les débits, les niveaux de liquides ou de solides, les pressions, les températures. Dans toute installation de traitement d'eau la connaissance du débit est impérative. De plus, le traitement d'une eau conduit à lui ajouter un certain nombre de réactifs [19].

1.5.1. Mesure des paramètres physico-chimiques

On peut trouver deux types paramètres qui sont :

Mesure des paramètres usuels

On peut citer quelques paramètres usuels qui sont essentiellement les débits, les niveaux de liquide ou de solide, la pression, la température et conductivité.....etc.

Mesure des paramètres spécifiques

Dans les appareils utilisés pour la mesure des paramètres spécifiques de l'eau, les différentes méthodes d'analyse sont mises en œuvre de façon automatique, en particulier : la néphélométrie (mesure de turbidité), la mesure de résistivité ou de conductivité, la potentiomètre (mesure de pH), l'ampèremètre (mesure de concentration en agent oxydant, chlore, ozone), la photo-colorimétrie et la titrimétrie (mesure de la concentration de certaines substances dissoutes dans l'eau[20]. On peut classer ces différents appareils en deux grandes catégories : celle des capteurs physiques et celle des analyseurs chimiques.

1.5.2. Capteurs physiques

Mesure de la turbidité

La mesure de turbidité de l'eau correspond à une mesure optique des particules en suspension dans l'eau qui lui donnent un aspect trouble. L'unité employée est appelée unité céphalométrique de turbidité (NTU) [21].

Mesure de pH

Le pH (potentiel Hydrogène) mesure la concentration en ions H⁺ de l'eau [22].

5.2.3. Mesure de la conductivité

La conductivité mesure la capacité de l'eau à conduire le courant entre deux électrodes. La plupart des matières dissoutes dans l'eau se trouvent sous forme d'ions chargés

électriquement. La mesure de la conductivité permet donc d'apprécier la quantité de sels dissous dans l'eau [22].

1.6. Importance de l'analyse et du traitement

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

- Elle garantit une eau potable sûre.
- Elle permet de définir les problèmes existants.
- Elle garantit une eau qui convient à l'utilisation prévue.
- Elle permet de vérifier l'efficacité du système de traitement [16].

1.7. La surveillance de la qualité des eaux

La surveillance de la qualité des eaux est un processus passif, informationnel qui analyse l'état du système et fournit des indications. La surveillance consiste notamment à détecter et classer les défaillances en observant l'évolution du système puis à les diagnostiquer en localisant les éléments défaillants et en identifiant les causes premières. La surveillance se compose donc de deux fonctions principales qui sont la détection et le diagnostic [23]. Dans la méthodologie de la surveillance, nous trouvons deux formes : méthodologie avec modèle et méthodologie sans modèle, chacune d'elles a son propre concept.

1.7.1. Les techniques avec modèle

À partir de là, les techniques qui ont un modèle dépendent de l'existence d'un certain modèle pour le processus, car elles utilisent apparemment des techniques qui sont automatiques [24].

1.7.2. Les techniques sans modèle

Il se compose de deux parties : la première partie avec les outils de traitement statistique du signal, qui sont de bas niveau et qui est due au contact direct avec le mouvement du capteur, tandis que la deuxième partie est ce que l'on appelle les technologies de haut niveau. Qui sont plus orientés vers l'établissement de la communication avec l'expert [25].

Donc Le domaine de surveillance de l'eau potable présente des caractéristiques bien spécifiques qu'il est indispensable d'en tenir compte dans la construction d'une démarche globale de prévention des risques. La qualité doit être garantie par une surveillance et un contrôle permanents à travers des mesures qualifiables et quantifiables dans le but de ne pas ralentir la production et d'en assurer un niveau de qualité élevé et uniforme. Le but de cette

maîtrise est de pallier sans aucun doute les conséquences graves qui se concrétisent au niveau des risques encourus pour la santé publique. L'exigence d'une réglementation très stricte des pouvoirs publics est alors bien justifiée [26, 27].

1.8. Les avantages et les inconvénients de la surveillance

- Diligence raisonnable;
- Respect des exigences législatives;
- Augmentation de la confiance des clients;
- Création de données de base relatives à la qualité de l'eau;
- Réduction des risques pour la santé publique grâce à la détection et à l'atténuation rapides d'une qualité inadéquate ou qui se détériore;
- Prise de décisions plus facile en ce qui concerne les activités d'exploitation et d'entretien liées à la qualité de l'eau dans le réseau de distribution;
- Maximisation de l'efficacité des produits chimiques ajoutés à la station de traitement;
- Création de données servant à justifier des améliorations importantes qui pourraient être apportées au réseau de distribution;
- Façon proactive de régler les nouveaux problèmes liés à la qualité de l'eau dans le réseau de distribution [28].
- L'inconvénient de cette technique est qu'elle nécessite de façon non-stop des interventions et des déplacements sur site de l'opérateur. Elle ne permet pas en fait de suivre finement l'évolution de la qualité de l'eau brute et a le désavantage d'avoir un temps de réponse relativement long [29].

En général, l'eau souterraine et superficielle se forme à deux types de l'eau dans le monde, ils sont eau potable et eau non potable. Nous disons que:

- L'eau est potable si elle est satisfaite aux normes internationales de l'Organisation mondiale de la santé (OMS) et n'expose pas la santé humaine à un danger.
- L'eau non potable est l'eau dont la qualité ne conforme pas les normes de potabilité établies par l'Organisation mondiale de la santé (OMS), ne pouvant être destinée à la consommation humaine, il est souvent utilisé pour des besoins industriels etc.

1.9. Normes de la qualité des eaux

Une eau de consommation ne doit pas contenir de germes des maladies à transport hydrique, de substances toxiques ni de quantité excessive de matières minérales et organiques. Elle doit par ailleurs, être limpide, incolore et ne posséder aucun goût ou odeur désagréable.

En outre l'eau potable doit contenir sans excès un certain nombre d'éléments minéraux dont la présence lui confère une saveur agréable à l'exclusion de ceux qui seraient l'indice d'une contamination ainsi que toute substance toxique [30]. Ainsi pour notre étude nous nous référons aux normes de l'OMS dans les tableaux qui suivent :

Tableau 1.1. Normes d'eau potable selon l'OMS (2006).

Elément/ substance	Symbole/ formule	Concentration normalement trouvée dans l'eau de surface	Lignes directrices fixées par l'OMS
Aluminium	Al		0,2 mg/l
Ammonium	NH ₄ ⁺	<0.2 mg/l (peut aller jusqu'à 0.3mg/l dans une eau anaérobie)	Pas de contrainte
Antimoine	Sb	< 4 µg/l	0.02 mg/l
Arsenic	As		0,01 mg/l
Baryum	Ba		0,7 mg/l
Béryllium	Be	< 1 µg/l	Pas de valeur guide
B	B	< 1 mg/l	0.5mg/l
Cadmium	Cd	< 1 µg/l	0,003 mg/l
Chlore	Cl		Pas de valeur mais on peut noter un goût à partir de 250 mg/l
Chrome	Cr+3, Cr+6	< 2 µg/l	chrome total : 0,05 mg/l
Cuivre	Cu ₂ ⁺		2 mg/l
oxygène dissous	O ₂		Pas de valeur guide
Fluorure	F ⁻	< 1,5 mg/l (up to 10)	1,5 mg/l
Dureté	CaCO ₃		200 ppm
Fer	Fe	0,5 - 50 mg/l	Pas de valeur guide
Plomb	Pb		0,01 mg/l
Manganèse	Mn		0,4 mg/l
Mercure	Hg	< 0,5 µg/l	inorganique : 0,006 mg/l
Molybdène	Mb	< 0,01 mg/l	0,07 mg/l
Nickel	Ni	< 0,02 mg/l	0,07 mg/l
Nitrate et nitrite	NO ₃ , NO ₂		50 et 3 mg/l (exposition à court terme) 0.2 mg/l (exposition à long terme)
Turbidité			Non mentionnée
pH			Pas de valeur guide mais un optimum entre 6.5 et 9.5
Sélénium	Se	<< 0,01 mg/l	0,01 mg/l
Argent	Ag	5 – 50 µg/l	Pas de valeur guide
Sodium	Na	< 20 mg/l	Pas de valeur guide
Sulfate	SO ₄		500 mg/l
Étain inorganique	Sn		Pas de valeur guide : peu Toxique
Uranium	U		0.015 mg/l
Zinc	Zn		3 mg/l

En L'Algérie, il existe des normes nationales officielles pour la qualité de l'eau potable et sont actuellement mises en œuvre avec les recommandations de l'Organisation mondiale de la santé(OMS) relatives à la qualité de l'eau de boisson.

Tableau 1.2. Valeur maximal standard de certains paramètres qui utiliser dans cette étude

Paramètres	S_i (valeur maximale standard, Algérie)
pH	7
Conductivité (à 25 °C ms/cm)	2700
Température (°C)	25
Turbidité (NTU)	5
Ammonium (mg/l)	0.5
Nitrite (NO ⁻² mg/l)	0.2
Calcium (Ca ²⁺ mg/l)	200
Chlorure (Cl- mg/l)	500
Bicarbonate (HCO ₃ ⁻ mg/l)	50
Dureté totale (mg/l)	200
Dureté permanente (mg/l)	200
Titre alcalin complet (mg/l)	500

Les paramètres

Une eau potable et/ou une eau destinée à fabrication d'eau potable doit respecter 48 paramètres parmi lesquels on compte 6 des 9 paramètres suivis dans l'eau potable. Dans cet étude on utilisés 12 paramètres différents, on classés en deux groupes des paramètres sont :

- **Paramètres physico-chimiques :** Comprennent le PH, conductivité, température, Turbidité, NTU.
- **Paramètres de pollution :** Comprennent l'Ammonium, Nitrite NO-2, Calcium Ca₂⁺, Chlorures Cl⁻, Bicarbonate HCO₃⁻, Dureté Totale TH CaCO₃, Dureté permanente CaCO₃, Titre alcalin complet CaCO₃.

1.10. L'indice de qualité de l'eau (IQE)

L'indice de la qualité de l'eau (IQE) est une méthode mathématique simple utilisée comme Partie de l'analyse de la qualité générale de l'eau à l'aide d'un groupe de paramètres réduisant les grandes quantités d'informations à un seul nombre, généralement sans dimension [31].cet nombre de l'indice de la qualité des eaux (IQE) est représenté l'état de la qualité d'une eau. Il est variant entre 0 et 100 ou 100 représente une eau mauvaise qualité (eau non potable) et 0 une eau excellente qualité (eau potable) [57].

La méthode de l'indice de qualité des eaux souterraines et superficielles reflète l'influence composite des différents paramètres de qualité de l'eau sur sa potabilité. L'indice est calculé en utilisant les équations suivantes :

$$W_i = \frac{K}{S_i} \quad (1.1)$$

$$K = \frac{1}{\sum_{i=1}^n \left(\frac{1}{S_i}\right)} \quad (1.2)$$

W_i : le poids de chaque paramètre selon son importance relative dans la qualité des eaux de boisson,

Ou :

K : constante de proportionnalité et peut également être calculée à l'aide de l'équation suivante :

n : nombre de paramètres,

S_i : valeur maximale de la norme standard algérien des eaux de surface de chaque paramètre en mg/l sauf pour le pH, la T°C et la conductivité électrique.

Ensuite, une échelle d'évaluation de la qualité (Q_i) est calculée pour chaque paramètre en divisant la concentration par la norme dudit paramètre et en multipliant l'ensemble par 100 comme dans la formule suivante :

$$Q_i = \left(\frac{C_i}{S_i}\right) \times 100 \quad (1.3)$$

Q_i : échelle d'évaluation de la qualité de chaque paramètre.

C_i : la concentration de chaque paramètre en mg/l.

Finalement l'indice global de la qualité de l'eau est calculé par l'équation suivante [31] :

$$IQE = \frac{\sum_{i=1}^n Qi \times Wi}{\sum_{i=1}^n Wi} \quad (1.4)$$

Nous pouvons classer la qualité de l'eau à Cinq classes de qualité peuvent être identifiées selon les valeurs de l'indice de qualité de l'eau IQE

Tableau 1.3. Classification et usage possible de l'eau selon l'IQE [31].

Classe d'IQE	Type d'eau	Usage possible
0 - 25	Excellente qualité	Eau potable, irrigation et industrie
>25 - 50	Bonne qualité	Eau potable, Irrigation et industrie
>50 - 75	Mauvaise qualité	Irrigation et industrie
>75 - 100	Très mauvaise qualité	Irrigation
> 100	Eau non potable	Traitement approprié requis avant utilisation

Conclusion

Ce premier chapitre présente une introduction au domaine de la surveillance de l'eau propre et sa maîtrise en tant qu'élément important de la vie humaine, ses paramètres descriptifs selon les normes de qualité de l'eau potable, et son traitement ont été mentionnés. Les différentes technologies les plus importantes qui existent dans le monde dans le domaine de la surveillance ont été décrites. L'utilisation des techniques analytique et intelligente qui aident à obtenir une décision sur la qualité de l'eau, et c'est ce que nous avons étudié, à savoir l'indice de qualité de l'eau, sera l'objet principale de notre étude dans ce travail.

Le chapitre suivant fera l'objet d'une mise en œuvre théorique d'une technique de l'intelligence artificielle basée sur l'apprentissage statistique et appliquées sur des données multi-sensorielles. L'objectif, rappelons-le encore une fois, est l'application de cette technique comme étant une solution dans le développement d'un modèle d'aide à la décision de la qualité de l'eau par reconnaissance de formes pour la prédiction de l'indice de la qualité de l'eau (IQE).

CHAPITRE II

LA TECHNIQUE SVM APPLIQUEE A LA REGRESSION

Introduction

La résolution des problèmes par la construction de machines capables d'apprendre à partir des entrées et des sorties, caractérise l'approche fondamentale de la théorie d'apprentissage (*Machine Learning*). Le problème typique de la théorie de l'apprentissage statistique se résume dans le contexte où des données engendrées par une distribution de probabilité (phénomène physique), se répartissent en plusieurs classes dans le cas d'une classification des données, ou permettent de prédire une sortie désirée dans le cas d'une régression. Ce problème d'apprentissage supervisé des données, est identifié comme une des problématiques majeures dans l'extraction des connaissances à partir des données. Pour traiter un problème dans ce sens, diverses méthodes ont été développées. Parmi celles-ci on trouve particulièrement les machines à vecteurs de support (SVM) [59].

Dans le cadre de notre travail, la solution adoptée rappelons-le une fois de plus, est une surveillance et un contrôle de la qualité de l'eau par les techniques de reconnaissance de formes. Dans ce chapitre, nous allons donc pouvoir passer en revue de la méthode SVM appliquées à la régression. Après une brève introduction, nous présenterons le principe de la technique SVM appliqué à la régression.

2.1. LES MACHINES A VECTEURS DE SUPPORT APPLIQUEE A LA CLASSIFICATION

Les séparateurs à vastes marges (ou SVM : Support Vector Machines) reposent sur deux idées clés : la notion de marge maximale et la notion de fonction noyau. Ces deux notions existaient depuis plusieurs années avant qu'elles ne soient mises en commun pour construire les SVM.

L'idée des hyperplans à marge maximale a été explorée dès 1963 par Vladimir Vapnik et A. Lerner, et en 1973 par Richard Duda et Peter Hart dans leur livre *Pattern Classification*. Les fondations théoriques des SVM ont été explorées par Vapnik et ses

collègues dans les années 70 avec le développement de la Théorie de Vapnik-Chervonenkis, et par Valiant. [52]

L'idée des fonctions noyaux n'est pas non plus nouvelle: le théorème de Mercer date de 1909, et l'utilité des fonctions noyaux dans le contexte de l'apprentissage artificiel a été montrée dès 1964 par Aizermann, Bravermann et Rozenner. Ce n'est toutefois qu'en 1992 que ces idées seront bien comprises et rassemblées par Boser, Guyon et Vapnik dans un article, qui est l'article fondateur des séparateurs à vaste marge.

L'idée des variables ressorts, qui permet de résoudre certaines limitations pratiques importantes, ne sera introduite qu'en 1995. À partir de cette date, qui correspond à la publication du livre de Vapnik, les SVM gagnent en popularité et sont utilisés dans de nombreuses applications. Un brevet américain sur les SVM est déposé en 1997 par les inventeurs originaux [32]

Dans les nombreuses études et applications, ces outils de modélisation se sont avérés très efficaces et parfois supérieurs à plusieurs méthodes différentes, notamment dans le domaine du traitement de l'eau potable, en raison de problèmes de classification ou de régression.

2.2. Principe

SVM (Support Vector Machines) ou Machines à vecteurs supports sont des classifiées faisant partie des techniques d'apprentissage supervisé. Introduits par Vapnik (1995) pour résoudre des problèmes de classification, ils ont connu depuis un grand succès car utilisés massivement dans divers domaines : reconnaissance de formes, OCR, bioinformatique... L'usage s'est également répandu vers la résolution des problèmes de régression, aussi bien gérés que les classifications. [54]

Cette technique fait appel à un jeu de données dit « d'apprentissage » dont les instances contiennent une valeur cible (Target) également appelée « étiquette de classe » ainsi que plusieurs attributs (attribuâtes) représentant l'ensemble des variables observées, et ce pour produire un modèle permettant de prédire les valeurs cibles d'un autre ensemble dit « de test », en ne fournissant à ce modèle-là que les attributs des données de test. En d'autres termes, si on considère un jeu de données divisé en deux groupes : un groupe pour les exemples connus et un autre pour les exemples non connus, le but des SVM est d'apprendre une fonction qui

traduit le comportement des exemples connus pour prédire les cibles des exemples inconnus [33].

2.2.1. Support vectors (vecteurs support)

Ce sont les points les plus proches de l'hyperplan optimal et qui déterminent la marge.

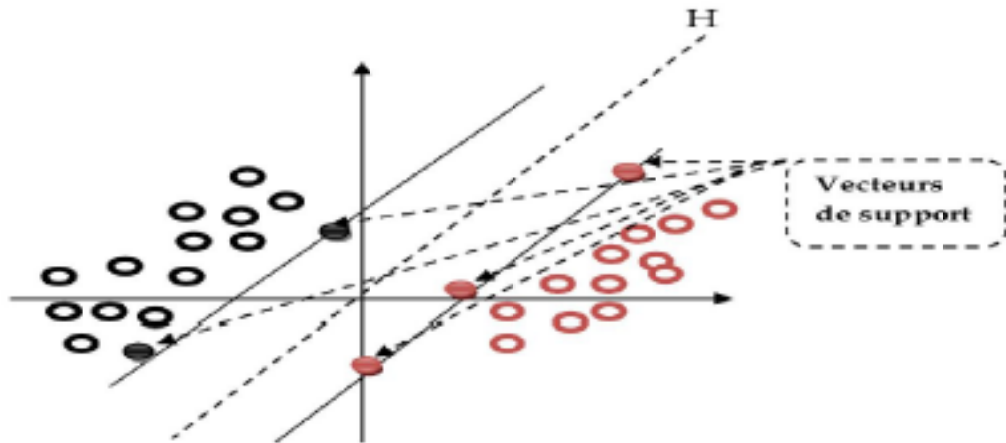


Fig. 2.1. Les vecteurs de support

2.2.2. Marge

Il existe une infinité d'hyperplans capable de séparer parfaitement les deux classes d'exemples. Le principe des SVM est de choisir celui qui va maximiser la distance minimale entre l'hyperplan et les exemples d'apprentissage (i.e. la distance entre l'hyperplan et les vecteurs de support), cette distance est appelée la marge [34].

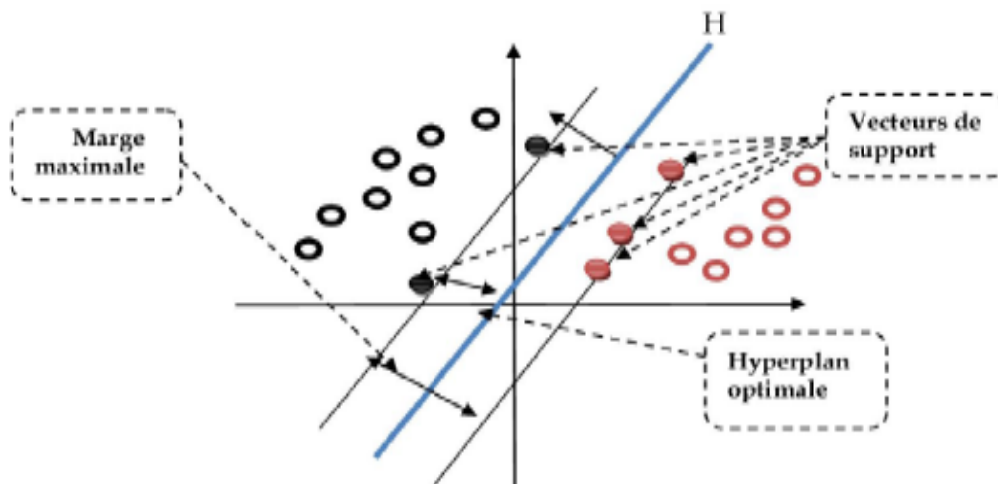


Fig. 2.2. Représentation de la marge.

2.3. Principe des machines à vecteurs de support

Les SVM constituent une classe d'algorithmes basée sur le principe de minimisation de « Risque structurel » décrit par la théorie de l'apprentissage statistique qui utilise la séparation linéaire. Cela consiste à séparer par un hyperplan des individus représentés dans un espace de dimension égale au nombre de caractéristiques, les individus étant alors séparés en deux classes. Cela est possible quand les données à classer sont linéairement séparables. Dans le cas contraire, les données seront projetées sur un espace de plus grande dimension qu'elles deviennent linéairement séparables [35] [58]

2.3.1. Principes fondamentaux

Maximisation de la marge

Intuitivement, le fait d'avoir une marge plus large procure plus de sécurité lorsqu'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.

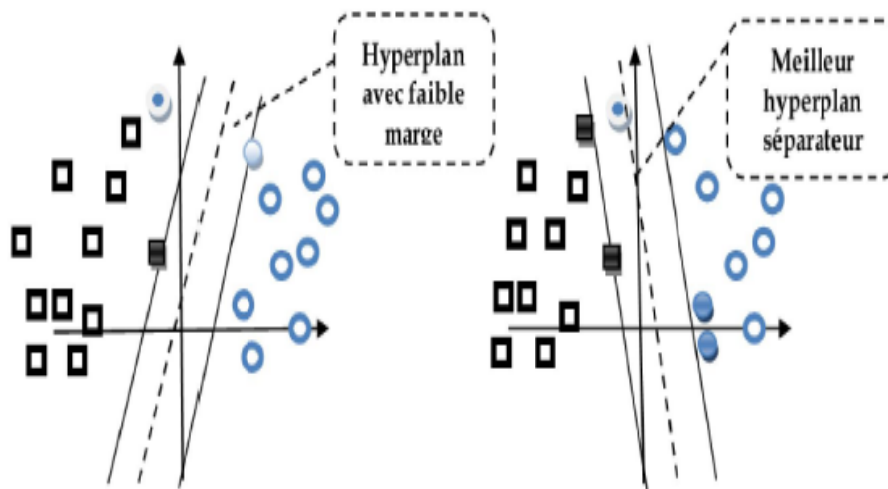


Fig. 2.3. Meilleur hyperplan séparateur.

Cas de données linéairement séparables

Considérons « l » points, $(x_1, y_1, x_2, y_2, \dots, x_l, y_l)$, $x_i \in R^N$ Avec $i=1 \dots L$ et $y_i \in \{\pm 1\}$ Ces points sont classés en utilisant une famille de fonctions linéaires définis par:

$$\langle w \rangle; +b = 0 \quad (2.1)$$

Avec w et $b \in \mathbb{R}$ de telle sorte que la fonction de décision concernant l'appartenance d'un point à l'une des deux classes soit donnée par :

$$f(w, b) = \text{sgn}((w; +b)) \quad (2.2)$$

La fonction (2.1) représente l'équation de l'hyperplan H. La fonction de décision (2) va donc observer de quel côté de H se trouve l'élément de x .

On appelle la marge d'un élément, la distance euclidienne prise perpendiculairement entre H et x . Si on prend un point quelconque t sur H, cette marge peut s'exprimer en :

$$M_x = \frac{w}{\|w\|} (x - t) \quad (2.3)$$

La marge de toutes les données est définie comme étant :

$$M = \min_{x \in E} M_x \quad (2.4)$$

L'approche de classification par SVM tend à maximiser cette marge pour séparer le plus clairement possible deux classes. Intuitivement, avoir une marge la plus large possible sécurise mieux le processus d'affectation d'un nouvel élément à l'une des classes. Un SVM fait donc partie des classificateurs à marge maximale. Un classificateur à marge maximale est un classificateur dont l'hyperplan optimal séparant deux classes est une solution du problème d'optimisation mathématique suivant (forme primale) :

$$\text{minimiser } \frac{1}{2w} \|w\|^2 \quad (w; +b) z \geq x \in E \quad (2.5)$$

La fonction objective de ce problème est le carré de l'inverse de la double marge qu'on veut maximiser. La contrainte unique correspond au fait que les éléments x doivent être bien placés. La résolution de ce problème nécessite de fixer les paramètres w et b qui constituent les variables α_i de la machine d'apprentissage. Les classificateurs à marge maximale donnent de bons résultats lorsque les données sont linéairement séparables.

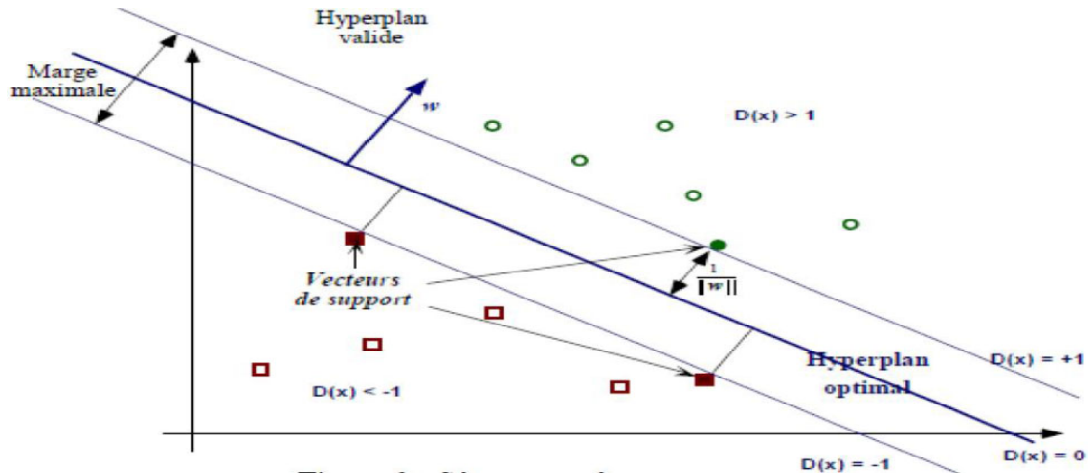


Fig. 2.4. Séparateur a vaste marge

Forme duale

La formulation primale peut être transformée en formulation duale en utilisant les multiplicateurs de Lagrange. L'équation (2.5) s'écrit alors sous la forme suivante [36] :

$$(W.b.\alpha) = \frac{1}{2} \|W\|^2 - \sum_{i=1}^l \alpha_i (y_i (W \cdot x_i) + b) - 1 \quad (2.6)$$

Cas des données non-linéairement séparables

En pratique, il est assez rare d'avoir des données linéairement séparables. Afin de traiter également des données bruitées ou non linéairement séparables, les SVM ont été généralisées grâce à deux outils : la marge souple (soft margin) et les fonctions noyau (kernel functions).

Le principe de la marge souple est d'autoriser des erreurs de classification. Le nouveau problème de séparation optimale est reformulé comme suit :

$$\text{MIN}_{(W,b,\epsilon)} \frac{1}{2} W^T W + c \sum_i^l \epsilon_i ; c \geq 0$$

source contraintes (2.7)

$$y_i((W, X) + b) \pm 1 - \epsilon_i$$

$$\epsilon_i \geq 0 \text{ Pour } i = 0 \dots \dots l$$

Un terme de pénalité est introduit dans la formule (2.7), Le paramètre C est défini par l'utilisateur. Il peut être interprété comme une tolérance au bruit de classificateur [37]

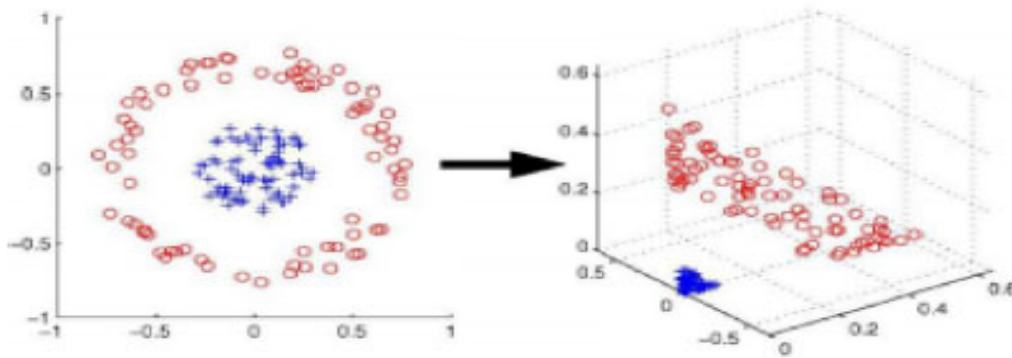


Fig. 2.5.Exemple de projection dans un espace de redescription.

La figure 2.5.Montre un exemple de transformation d'un problème de séparation non linéaire dans l'espace de représentation (2 dimensions) en un problème de séparation linéaire dans un espace de redescription de plus grande dimension (3 dimensions). Cette transformation non linéaire est réalisée via une fonction noyau [37].

2.2. LA TECHNIQUE SVM APPLIQUEE A LA REGRESSION

2.2.1. Définition générale de la régression

La régression est un ensemble de méthodes statistiques très utilisées pour analyser la relation d'une variable par rapport à une ou plusieurs autres. Pendant longtemps, la régression d'une variable aléatoire y sur le vecteur de variables aléatoires x désignait la moyenne conditionnelle de y sachant x . aujourd'hui, le terme de régression désigne tout élément de la distribution conditionnelle de y sachant x , considérée comme une fonction de x . on peut par exemple s'intéresser à la moyenne conditionnelle, à la médiane conditionnelle, au mode conditionnel, à la variance conditionnelle.

Le terme " régression" a été introduit par Francis Galton à la suite d'une étude sur la taille des descendants de personnes de grande taille, qui diminue de générations en générations vers une taille moyenne (donc leur taille régression). [38]

En apprentissage automatique, on distingue les problèmes de régression des problèmes de classification. Ainsi, on considère que les problèmes de prédiction d'une variable quantitative sont des problèmes de classification. Certains méthodes, comme la régression logistique, sont à la fois des méthodes de régression au sens où il s'agit de prédire la probabilité d'appartenir à chacune des classes et des méthodes de classification.

Le modèle de régression le plus connu est le modèle de régression linéaire. Lorsque le modèle n'est pas linéaire, on peut effectuer une régression approchée par des algorithmes itératifs, on

parle de régression non linéaire. Si on s'intéresse au quantile conditionnel de la distribution de la variable aléatoire x , on utilise un modèle de régression quantile. [39] Si la variable expliquée est une variable aléatoire binomiale, il est courant d'utiliser une régression logistique ou un modèle probité. Si la forme fonctionnelle de la régression est inconnue, on peut utiliser un modèle de régression non paramétrique.

2.2.2. Régression par SVM

Lorsque les SVM sont utilisés dans des problèmes de régression pour prédire des valeurs réelles, on parle des SVR (Support Vector Régression). « Les SVM peuvent également être mis en œuvre en situation de régression, c'est-à-dire pour l'approximation de fonctions quand Y est quantitative. Dans le cas non linéaire, le principe consiste à rechercher une estimation de la fonction par sa décomposition sur une base fonctionnelle. La forme générale des fonctions calculées par les SVM se met sous la forme :

$$\varnothing(x, w) = \sum_{i=1}^{\infty} W_i V_i(x) \quad (2.8)$$

Le problème se pose toujours comme la minimisation d'une fonction coût, mais plutôt que d'être basée sur un critère d'erreur quadratique (moindres carrés), celle-ci s'inspire des travaux de Huber sur la recherche de modèles robustes et utilise des écarts absolus. On note ρ_{ϵ} la fonction qui est paire, continue, identiquement nulle sur l'intervalle $[0, \epsilon]$ et qui croît linéairement sur $[\epsilon, +\infty]$. La fonction coût est alors définie par :

$$E(W, \gamma) = \frac{1}{n} \sum_{i=1}^n |Y_i - \varnothing(x_i, w)| + \gamma \|w\|^2 \quad (2.9)$$

Où γ est, comme en régression rigide, un paramètre de régularisation assurant le compromis entre généralisation et ajustement. De même que précédemment, on peut écrire les solutions du Problème d'optimisation. Pour plus de détails, se reporter à Schölkopf et Smola (2002).

Les points de la base d'apprentissage associés à un coefficient non nul sont là encore nommés vecteurs support. Dans cette situation, les noyaux k utilisés sont ceux naturellement associés à la définition de bases de fonctions. Noyaux de splines ou encore noyau de Dériclet associé à un développement en série de Fourier sont des grands classiques. Ils expriment les produits scalaires des fonctions de la base. [40]

En suite, nous avons présentés l'algorithme SVR pour un modèle linéaire, puis pour le cas non-linéaire.

Le cas linéaire

La méthode est d'abord décrite pour une fonction $F(x)$ linéaire de la forme :

$$F(x) = (w \cdot (x)) + b \quad (2.10)$$

Avec $x \in R^p$ le vecteur d'entrée, $w \in R^p$ le vecteur des paramètres (ou poids) et b une constante à déterminer. Pour assurer la platitude de la fonction f , la norme des poids $\|w\|$ est minimisée (contrainte sur la dérivée de la fonction F). Le problème revient donc à minimiser cette norme en garantissant que les erreurs sont inférieures à ϵ et peut s'écrire :

$$\min \frac{1}{2} \|w\|^2$$

$$s. c. |y_i - (w \cdot x_i) - b| \leq \epsilon; \quad i = 1 \dots \dots \dots N \quad (2.11)$$

On remarque que dans cette formulation du problème, l'objectif n'est pas de minimiser l'erreur d'apprentissage comme dans les réseaux de neurones ou la plupart des algorithmes de régression, mais d'assurer la platitude de la fonction. La minimisation de l'erreur n'apparaît que sous forme de contraintes, ici inviolables. Autrement dit, aucune erreur n'est autorisée. Cette description du problème considère donc qu'une fonction linéaire f qui approxime tous les exemples avec une précision ϵ existe. Dans la pratique, ce n'est pas toujours le cas. En présence de bruit trop important ou de valeurs aberrantes, il est aussi plus important d'autoriser certaines erreurs. Dans ce cas, le concept de marge souple (*soft margin*) est utilisé. Il consiste à introduire des variables de relâchement (*slack variables*) $\epsilon_i; \epsilon_i^*$ pour rendre faisables les contraintes du problème d'optimisation qui devient :

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\epsilon_i + \epsilon_i^*) \quad (2.12)$$

Sous les constraints :

$$y_i - (w \cdot x_i) - b \leq (\epsilon_i + \epsilon_i^*) \quad (2.13)$$

$$(\epsilon_i + \epsilon_i^*) \geq 0; \quad i = 1 \dots \dots \dots N$$

Où $(\epsilon_i + \epsilon_i^*)$ représentent respectivement les erreurs positives et négatives. La constante $C > 0$ est un hyper-paramètre permettant de régler le compromis entre la quantité d'erreur autorisée et la platitude de la fonction F . Cette formulation du problème revient à utiliser une fonction d'erreur $\|\epsilon\|$ appelée ϵ -insensible de la forme :

$$|y - f(x)| \varepsilon = \begin{cases} 0 & ; \text{pour } |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon & ; \text{pour } |y - f(x)| > \varepsilon \end{cases} \quad (2.14)$$

Et représentée Figure 2.6. On peut interpréter cette fonction comme créant un tube d'insensibilité de rayon ε au tour de la fonction $f(x)$ (voir la Figure 2.7). Les variables ε_i ou ε_i^* représentent alors la distance selon l'axe y entre le point $(x_i; y_i)$ et le bord ce tube.

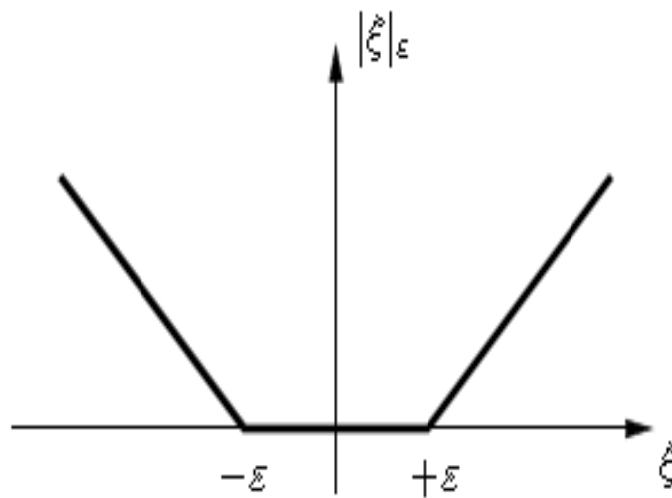


Fig. 2.6. La fonction d'erreur "*insensible*".

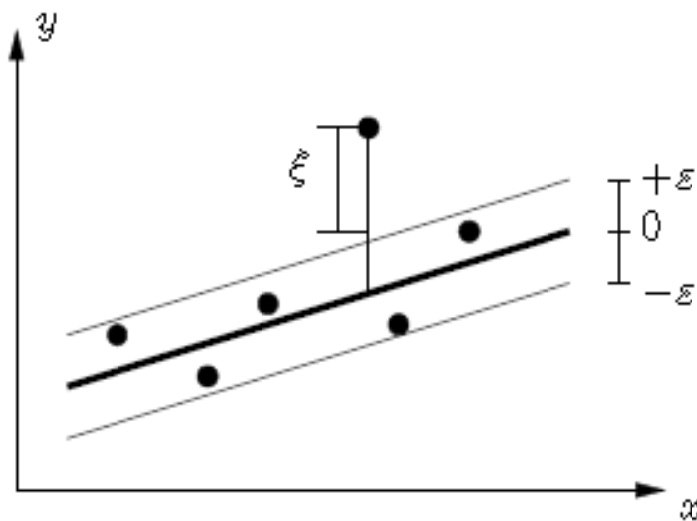


Fig.2.7. Tube d'insensibilité de rayon ε .

Le problème (2.14) se résout en minimisant le Lagrangien L donné par :

$$L = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N 1 = 1(\varepsilon_i + \varepsilon_i^*) - \sum_{i=1}^N (\eta_i \varepsilon_i + \eta_i^* \varepsilon_i^*) \quad (2.15)$$

$$- \sum_{i=1}^N \alpha_i (\varepsilon + \varepsilon_i) - y_i + (w, x_i) + b - \sum_{i=1}^N \alpha_i^* (\varepsilon + \varepsilon_i^*) + y_i - (w, x_i) - b$$

Où les $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ positifs représentent les multiplicateurs de Lagrange.

Le calcul des dérivées de L par rapport aux variables *primales* $(w, b, \varepsilon_i, \varepsilon_i^*)$ donne :

$$\partial L / \partial b = \sum_{i=1}^N 1 (\alpha_i^* - \alpha_i) = 0 \quad (2.16)$$

$$\partial L / \partial W = w - \sum_{i=1}^N 1 (\alpha_i^* - \alpha_i) x_i = 0 \quad (2.17)$$

$$\partial L / \partial \varepsilon_i = C - \alpha_i - \eta_i = 0 \quad (2.18)$$

$$\partial L / (\partial \varepsilon_i^*) = C - \alpha_i^* - \eta_i^* = 0 \quad (2.19)$$

Les équations (2.16) et (2.17) permettent de supprimer les Variables $\eta_i = C - \alpha_i$ et $\eta_i^* = C - \alpha_i^*$

En réinjectant ces résultats dans (2.19), on obtient le Lagrangien dual qui doit être maximisé :

$$LD = -\frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) (x_i, x_j) \quad (2.20)$$

$$- \varepsilon (\alpha_i + \alpha_i^*) + \sum_{i=1}^N (\alpha_i - \alpha_i^*)$$

Sous les contraintes :

$$\sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0$$

$$\text{Et } \alpha_i, \alpha_i^* \in [0, C] \quad (2.21)$$

A partir de (2.21), les poids du modèle sont déterminés par :

$$W = \sum_{i=1}^N (\alpha_i - \alpha_i^*) x_i \quad (2.22)$$

Et le modèle s'écrit donc :

$$F(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) (x_i, x) + b \quad (2.23)$$

Le paramètre de biais b peut se calculer grâce aux conditions de Karush-Kuhn-Tucker (KKT) selon lesquelles les produits entre les variables duales et les contraintes sont nuls à la solution

$$\alpha_i(\varepsilon + \varepsilon_i - y_i + (W \cdot x_i) + b) = 0 \quad (2.24)$$

$$\alpha_i^*(\varepsilon + \varepsilon_i^* + y_i - (W \cdot x_i) - b) = 0 \quad (2.25)$$

$$\eta_i \varepsilon_i = (C - \alpha_i) \varepsilon_i = 0 \quad (2.26)$$

$$\eta_i \varepsilon_i^* = (C - \alpha_i^*) \alpha_i^* = 0 \quad (2.27)$$

b peut par exemple être déterminé par (2.24) sur un point particulier pour lequel x_i ; y_i et α_i sont connus et α_i a été précédemment évalué par l'algorithme de maximisation du Lagrangien dual LD. On peut remarquer que, d'après (2.26) et (2.27), les seuls points pouvant être à l'extérieur du tube d'insensibilité ($\varepsilon_i, \varepsilon_i^* \neq 0$) sont ceux pour lesquels les multiplicateurs de Lagrange correspondants respectent $\alpha_i = C$ ou $\alpha_i^* = C$. D'autre part, pour les exemples à l'intérieur du tube ($|y_i - f(x_i)| < \varepsilon$ et $\varepsilon_i, \varepsilon_i^* = 0$), on a $\alpha_i, \alpha_i^* = 0$ pouvant vérifier les conditions (2.24) et (2.25). Cela entraîne la propriété de parcimonie (*sparsity*) des SVMs : seuls les exemples avec des correspondants non nuls ne sont nécessaires pour implémenter le modèle (2.23). Ces exemples sont appelés Vecteurs de Support ou *Support Vectors* (SVs) et ne constituent dans la pratique

Qu'une petite fraction des données d'apprentissage [41]

Le cas non-linéaire

L'extension des SVMs au cas non-linéaire est assez simple et repose sur la projection des données dans un espace de dimension supérieure dans lequel le problème devient linéaire. L'exemple suivant montre l'implémentation d'une fonction quadratique par un SVM linéaire. Soit la projection Φ définie par :

$$\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3 \quad \Phi(x_1; x_2) = (x_1^2, \sqrt{2}x_1x_2, x_2^2) \quad (2.28)$$

L'algorithme SVM précédemment décrit appliqué aux données projetées dans \mathbb{R}^3 par donnerait une fonction linéaire dans \mathbb{R}^3 correspondant à une fonction quadratique dans \mathbb{R}^2 .

Cette approche semble raisonnable pour cet exemple en dimension 2, mais peut vite devenir coûteuse en temps de calculs pour des non-linéarités polynomiales d'ordre supérieur ou des problèmes à grande dimension d'entrée. [42]

Pour résoudre ces problèmes, une projection implicite par fonctions noyaux est utilisée. Le fait que les données x_i n'apparaissent que sous forme de produit scalaire aussi bien dans la formulation du problème d'optimisation (2.20) que dans la fonction F (2.23) constitue un des principaux avantages des SVMs. Il n'est pas nécessaire de connaître la projection :

$\Phi : \mathbb{R}^2 \rightarrow F$ explicitement, mais uniquement le résultat $k(x_i; x) = (\Phi(x_i), \Phi(x))$ du produit scalaire entre les images des points dans l'espace étendu F , aussi appelé *feature space*. Le problème non-linéaire peut donc être formulé ainsi :

$$\text{Max } LD = -\frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) k(x_i, x_j)$$

$$-\varepsilon (\alpha_i + \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \quad (2.29)$$

Sous les contraintes :

$$\sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0$$

$$\text{Et } \alpha_i, \alpha_i^* \in [0, C] \quad (2.30)$$

De même, les poids sont donnés par :

$$W = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \Phi(x_i) \quad (2.31)$$

Et le modèle par :

$$F(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) k(x_i, x) + b \quad (2.32)$$

La principale différence avec le cas linéaire est que le vecteur des poids w n'est plus exprimé explicitement dans \mathbb{R}^p mais implicitement dans F . De même, l'algorithme cherchera la fonction la plus plate dans l'espace étendu F et non plus dans l'espace d'entrée \mathbb{R}^p . La fonction k est appelée fonction noyau ou noyau (*kernel function* ou *kernel*). Dans la pratique, seule cette fonction est connue, la projection Φ ne l'est pas. Les noyaux les plus couramment utilisés pour les SVMs sont les noyaux polynomiaux, sigmoïdaux et à fonction de base radiale (*Radial Basis Function, RBF*) définis ainsi :

$$- \text{Linéaire: } k(x; x') = (x, x');$$

- *polynomial* : $k(x; x') = (\gamma(x; x') + c)^d$;
- *Sigmoïdal* : $k(x; x_0) = \tanh(\gamma(x; x') + c)$;
- *RBF* : $k(x; x') = \exp(-\|x - x'\|^2 / 2\sigma^2)$;

Pour être admissible, une fonction noyau doit vérifier les conditions de Mercer qui se résument à vérifier que la fonction $k : R^{P*P} \rightarrow R$ corresponde bien à un produit scalaire définissant une norme dans un espace étendu F [42].

Les fonctions noyau

Afin de résoudre le problème de données non linéaires, la fonction noyau joue le rôle central de liaison des vecteurs d'entrées à l'espace de caractéristiques de grande dimension [36].

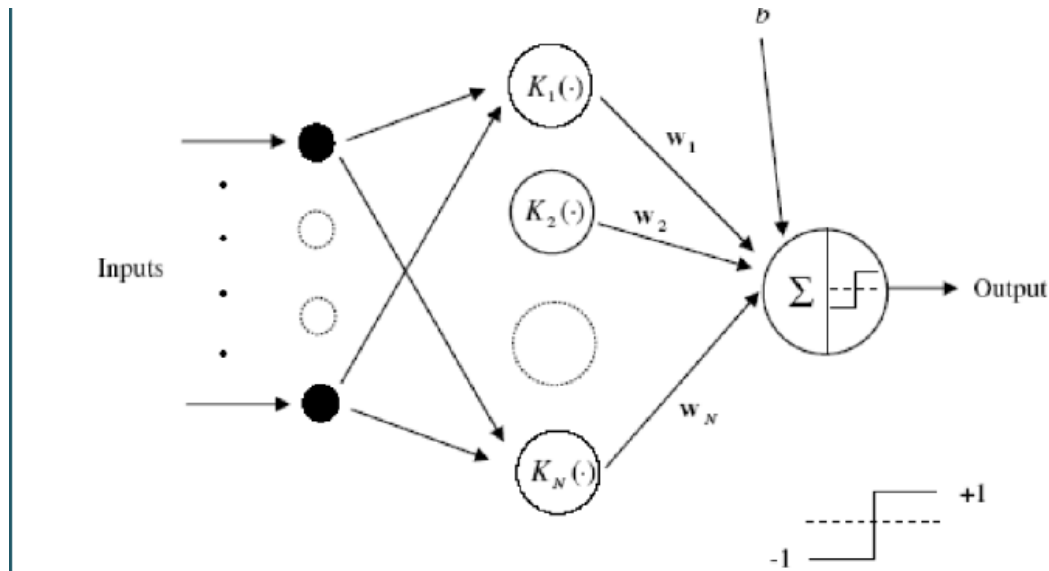


Fig. 2.8. Architecture d'une machine à vecteurs de support (d'un nombre N).

Les choix typiques pour la fonction noyau sont : Noyau gaussien à base radiale (ou RBF : Radial Basis Function) :

$$K(x, x_i) = \exp[-\gamma \|x - x_i\|^2]$$

Noyau polynômiale :

$$K(x_i - x_j) = (x_i^T x_j + 1)^p$$

Avec p : une constante qui spécifie le degré du polynôme.

À titre d'exemple, nous considérons ici le cas de la régression linéaire déjà étudié dans ce chapitre. On cherche à trouver une fonction linéaire h définie dans $X = \mathbb{R}^d$, qui interpole au mieux des données d'apprentissage $S = ((x_1, u_1), \dots, (x_m, u_m))$ de points x_i pris dans X et de valeur associée $u_i \in \mathbb{R}$:

$$h(x) = w_0 + \sum_{i=1}^m w_i x_i$$

Où les x_i sont les coordonnées d'un point x . Nous écrivons aussi :

$h(x) = (w, x)$ En étendant le vecteur d'entrée par l'ajout d'une composante zéro de valeur 1, ou bien $(w \cdot x + w_0)$ quand nous voudrions mettre en exergue le coefficient w_0 . Utilisons le principe inductif de minimisation du risque empirique, en prenant l'écart quadratique comme fonction de perte :

$$REmp(h) = REmp(w) = 1/m \sum_{i=1}^m l(h(x_i), u_i) = 1/m \sum_{i=1}^m \varepsilon_i^2$$

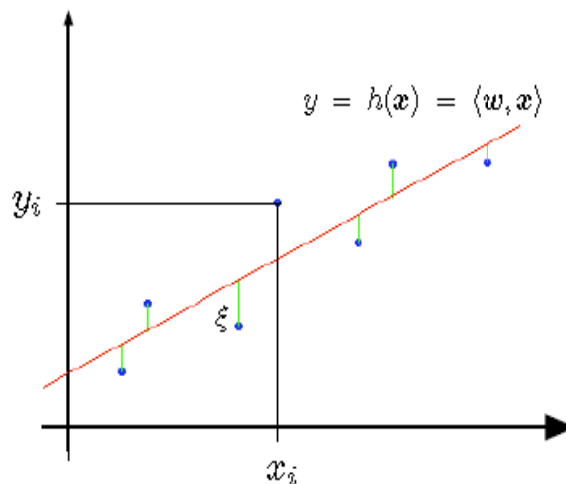


Fig.2.9. Régression linéaire.

On cherche le vecteur de paramètres w minimisant ce risque empirique. Écrivons ε le vecteur des écarts ε_i : $\varepsilon = y - Xw$.

On a alors : $REmp(w) = \|\varepsilon\|_2^2 = (y - Xw)^T (y - Xw)$.

On peut chercher la valeur optimale de w en calculant la dérivée du risque par rapport au vecteur de paramètres w et en la posant égale au vecteur nul :

$$\frac{\partial \text{REmp}(w)}{\partial w} = -2X^T y + 2X^T X w$$

On obtient alors les équations dites « normales » : $X^T X w = X^T y$.

Si l'inverse de XX^T existe, la solution peut s'exprimer sous la forme :

$$w = (X^T X)^{-1} X^T y = X^T X (X^T X)^{-2} X^T y = X^T \alpha$$

Ce qui signifie que w est une combinaison linéaire des points d'apprentissage : $w = \sum_{i=1}^m \alpha_i x_i$

L'hypothèse recherchée a alors la forme, appelé représentation duale:

$$h(x) = \sum_{i=1}^m \alpha_i(x_i, x)$$

Supposons que l'on cherche à résoudre cette fois-ci un problème de régression non linéaire dans X (voir figure 2.10), on peut être tenté d'utiliser pour ce faire une fonction de redescription Φ de X dans un nouvel espace F (« *featurespace*»), de telle manière que le problème devienne un problème de régression linéaire dans ce nouvel espace.

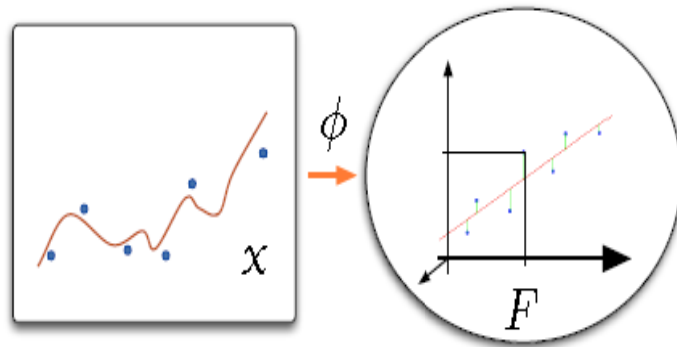


Fig. 2. 10. Régression non linéaire rendue possible par la transformation en un problème de régression linéaire dans l'espace de redescription F .

Dans ce cas, la solution du problème de régression linéaire, dans l'espace de redescription F , prendra la forme primale $h(x) = (w, \Phi(x)) = 0$, que l'on peut ré-exprimer en mettant en évidence la base de fonctions non linéaires $(\phi_1(x), \phi_2(x) \dots \dots)$ éventuellement infinie, de l'espace F :

$$h(x) = \sum_{j=1}^m w_j \phi_j(x) + w_0 = 0 \quad (2.33)$$

En résolvant comme précédemment, cette fois-ci dans l'espace de redescription, on obtient à

Nouveau une expression duale :

$$h(x) = \sum_{i=1}^m \alpha_i (\phi(x_i), \phi(x)) \quad (2.34)$$

Et, en rapprochant les équations (1) et (2), on tire :

$$h(x) = \sum_{j=1}^{\infty} w_j \phi_j(x) + w_0 = w^T \cdot \phi(x)$$

$$= \sum_{i=1}^m \alpha_i \phi(x_i) \phi(x) = 0$$

D'où :

$$w^T = \sum_{i=1}^m \alpha_i \phi(x_i)$$

Un problème est évidemment que la nouvelle formulation implique, d'une part, de trouver une fonction de redescription Φ adéquate, et, d'autre part, de calculer des produits scalaires dans un espace F dont la dimension peut être grande. Cependant, il peut parfois être possible de ne pas avoir à effectuer explicitement ces produits scalaires dans F grâce à l'utilisation de fonctions noyau. C'est ce que l'on appelle l'« astuce des noyaux » (kernel trick) [43].

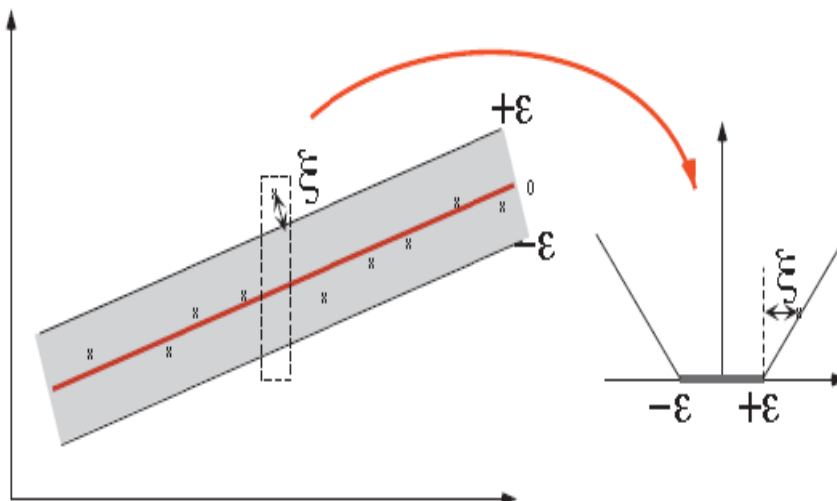


Fig. 2.11. Dans la régression par SVM, au lieu d'imposer une marge entre les points des deux classes, on impose une sorte de « chaussette » autour des points grâce au paramètre ϵ . Le compromis entre la complexité du modèle (la fonction de régression) et la fidélité aux points d'apprentissage est réglée par la variable ressort ϵ . [43]

Conclusion

La caractéristique importante de la prise en charge des bus de machines est qu'ils peuvent capturer des dépendances non linéaires de haut niveau entre des variables explicatives. La prise en charge des vecteurs de régression est un outil puissant de modélisation et de prédiction, et elle a été appliquée dans divers domaines d'application. Ce chapitre a fait l'objet de rappels des fondements de la technique SVM appliqués à la régression. Un exemple de modélisation développé à base de cette technique d'apprentissage sera d'ailleurs proposé dans ce qui suit, et ce afin de répondre aux critères de performances et exigences de la chaîne de contrôle et de surveillance proposée.

Le chapitre suivant consacré totalement à la simulation, est dédié à l'évaluation des performances de la technique étudiée. L'objectif, vise l'étude et la contribution de cette méthode en vue de leur application dans le domaine de contrôle de la qualité de l'eau. A titre d'exemple, nous développerons à cet effet un modèle de prédiction de l'indice de qualité de l'eau (IQE). L'information issue de ce modèle pourrait être utilisée au niveau de la chaîne de production de l'eau comme étant une donnée unique pouvant décider sur la qualité de l'eau.

CHAPITRE III

SIMULATION

INTRODUCTION

Ce dernier chapitre est consacré à la simulation et vise l'application de la technique SVM étudiée comme étant une solution dans la conception d'un outil d'aide à la décision de la qualité de l'eau. L'objectif principal de cette recherche est de réduire considérablement le temps et les efforts nécessaires au calcul de l'IQE dans les zones humides artificielles à surface libre. Les SVM ont été utilisés comme techniques robustes pour déterminer l'IQE. 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. Deux étapes d'application sont prévues dans ce cadre. Le calcul de l'indice de qualité de l'eau (IQE) analytiquement et le développement de cet indice par le modèle SVM et dans ce cas la l'élaboration de ce modèle est vu comme un problème de régression.

1. PROBLEMATIQUE

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 réseaux Neuronal, les réseaux neuronal-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.

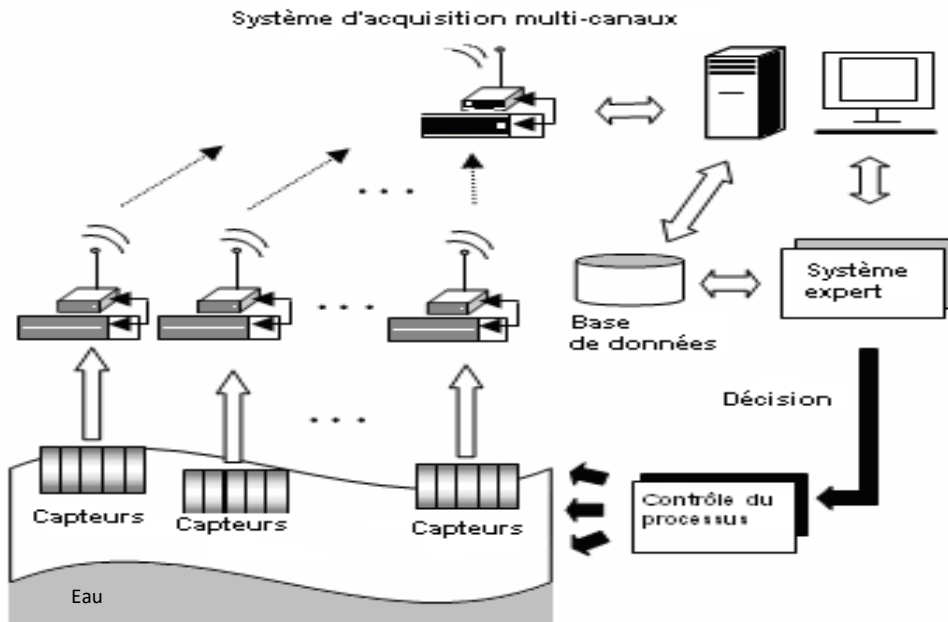


Fig.4.1. Architecture du système de contrôle et de surveillance.

Au niveau du système, on peut supposer que les différents paramètres physico-chimiques utilisées, tels que le pH, la température (T°), 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, le traitement des données et la décision appropriée. 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 de la surveillance

La qualité de l'eau (QE) est donc une préoccupation majeure pour les ressources en eau, les systèmes environnementaux et l'écosystème. C'est une terminologie utilisée pour décrire les caractéristiques chimiques, physiques et biologiques de l'eau en lien avec un ensemble de normes [44]. L'évaluation de la qualité de l'eau peut être utilisée pour évaluer les propriétés de l'eau en référence à la qualité naturelle et aux effets sur la santé humaine [45]. Il peut être évalué en mesurant un large éventail de variables pour représenter le niveau de pollution de l'eau. Par conséquent, une technique mathématique robuste est nécessaire pour combiner la

caractérisation physico-chimique de l'eau en une seule variable qui décrit la qualité de l'eau. Compte tenu de cela, un indice de qualité de l'eau (IQE) a été développé sous la forme d'un nombre unique qui utilise un ensemble de variables physico-chimiques de l'eau pour expliquer la qualité de l'eau à un certain endroit et à un certain moment [46].

L'IQE est un nombre sans unité qui reflète l'état de la qualité de l'eau. Le concept d'IQE est basé sur la comparaison du paramètre de qualité de l'eau avec les normes réglementaires respectives [47]. Il existe plusieurs équations pour l'IQE dans différents pays tels que les États-Unis, le Canada et la Malaisie, qui sont développées sur la base des normes de la US National Sanitation Foundation (Said et al. 2004). Dans notre étude, douze (12) paramètres ont été choisis comme principales variables de qualité de l'eau (WQV) pour développer IQE pour les eaux de surface de la zone d'étude de barrage de Tilesdit de Bouira – Algérie. Ces variables doivent être converties en paramètres non dimensionnels par des fonctions de sous-indice. La méthode conventionnelle recommandée par le DOE nécessite des transformations de longue durée pour calculer les sous-indices. De plus, les sous-indices nécessitent l'inclusion de différentes équations, qui nécessitent beaucoup d'efforts et de temps pour estimer l'IQE final. Par conséquent, l'estimation d'un tel IQE est lourde et peut conduire à des erreurs occasionnelles [48], et des techniques robustes peuvent être utilisées pour résoudre ces problèmes [49]. Les techniques de l'apprentissage statistique comme les machines à vecteurs de support (SVM) peuvent être suggérées comme techniques alternatives pour l'estimation de l'IQE, car cette méthode utilise les données brutes au lieu de sous-indices.

La solution devant être adoptée par la technique citée 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 modèle 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 généricité. A souligner toutefois que les méthodes de reconnaissance de formes à base d'apprentissage statistique sont les plus utilisées dans les systèmes d'aide à la décision à fusion multi-sensorielle. En général l'apprentissage est une étape assez longue, et nécessite plus de temps de calcul. Le critère de 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. MATERIELS ET METHODES

2.1. Présentation de la zone d'étude (Barrage de Tilesdit)

Le barrage « Tilesdit » est situé géographiquement dans la commune de Bechloul à 20 km du Sud-Est de la wilaya de Bouira, Algérie. Ce barrage est situé entre les coordonnées cartographiques et les coordonnées Lambert suivantes (figure 3.1) :

- Latitude : 35° 13' 22'' Nord.

- Longitude : 4° 14' 23'' Est.



Fig. 3.1 : Carte géographique situant le barrage « Tilesdit » [Google Maps].

Ce barrage disposant d'un volume de 167 million de mètres cubes d'eau, est conçu de façon à juguler la tension qui persiste dans la distribution d'eau au niveau de 12 communes (figure 3.2). De part son implantation dans la région de Bouira, le barrage Tilesdit dont la mise en eau à été effectuée vers la fin de l'année 2005, garantira de l'eau potable aux communes relevant de l'Est et du Sud-Est de la wilaya, c.-à-d: El-Asnam, Bechloul, El-Adjiba, Ahl-Ksour, Oud El-Berdi, Mesdour, Taguedit, Ahnif, Bordj O'khris, Ouledrached, Ath Mansour, Bouira et la zone industrielle de Sidi Khaled. Le transfert d'eau dont le lancement des travaux de réalisation à été prévu au début 2011, est destinée à l'alimentation de cinq autres communes de la daïra de Mansourah, dans la wilaya de Bordj Bou Arréridj. Des travaux sont en cours en vue de raccorder les communes de Takdit et Ait Laaziz, ainsi que d'autres communes rurales du Sud-Est de la wilaya de Bouira au réseau d'alimentation de ce barrage, qui devrait s'étendre jusqu'à Hammam k'sana. Parallèlement, la même direction a prévu 12 opérations portant sur la réalisation de réseaux AEP, en plus de 12 réservoirs d'une

capacité globale de 74000 m³ d'eau en vue de l'amélioration de l'approvisionnement en eau des communes alimentées par ce barrage notamment avec un débit de 72650 m³/jour. Il est également important de signaler qu'à l'horizon 2020, ce barrage garantira l'eau en faveur de 307200 habitants, selon les prévisions du secteur de l'hydraulique. Les travaux menés ont porté à 6 le nombre de stations de pompage pour un débit de 41 à 981 l/s. Une station de traitement d'une capacité de 74000 m³/jour, un réservoir d'eau traitée de deux compartiments d'un volume total de 13000 m³ et de 6 réservoirs de capacité allant de 200 à 5000 m³.

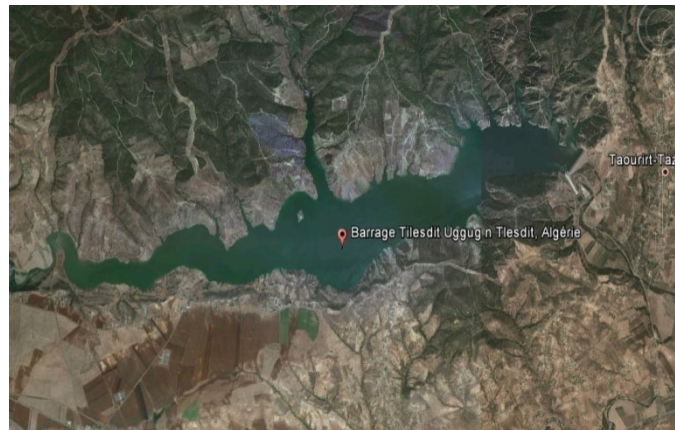


Fig. 3. 2 :«Image représentant le site du barrage « Tiledit » [Google Earth].

2.1.1. La station de production d'eau potable TILES DIT

L'eau prélevée dans le barrage est pompée jusqu'à la station de traitement. Celle-ci se trouvant au même lieu, est mise en service depuis 2009 (figure 3.3). Elle effectue le processus d'épuration à travers les cinq étapes de traitement indiquées au premier chapitre à savoir : le prétraitement, la pré-oxydation, la clarification, la désinfection et l'affinage. L'étape de clarification est assurée par le procédé de coagulation-floculation, décantation et filtration, grâce à un décanteur et un étage de filtration sur sable.

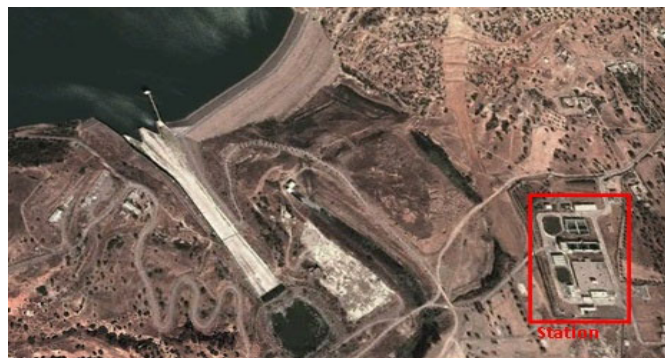


Fig. 3.3 :Image représentant le site de la station « Tiledit » [Google Earth]

2.2. Méthodologie et données utilisées

2.2.1. Base de données

Nous cherchons à décider sur la qualité de l'eau à travers ses paramètres descripteurs et le développement de l'indice de qualité de l'eau. Dans cette étude, il y a douze paramètres physico-chimiques qui sont souvent utilisés. Ces paramètres comprennent les données relatives aux analyses physicochimiques des eaux superficielles de barrage de Tilesdit (Bouira-Algérie). Les échantillons d'eau ont été prélevés au niveau de la station de production sur une période de temps s'étalant sur douze mois. Les paramètres sont résumés comme suit : pH, Conductivité électrique (CE) et Température (T°), Turbidité(TU), Ammonium(NH_4^+), Nitrite (NO_2^-), Calcium(Ca^{2+}), Chlorure(Cl^-), Bicarbonate (HCO_3^-), Dureté totale (TH), Dureté permanente (PEH), Titre alcalin complet (TAC). La figure 3.4 présente graphiquement l'évolution temporelle de ces différents paramètres.

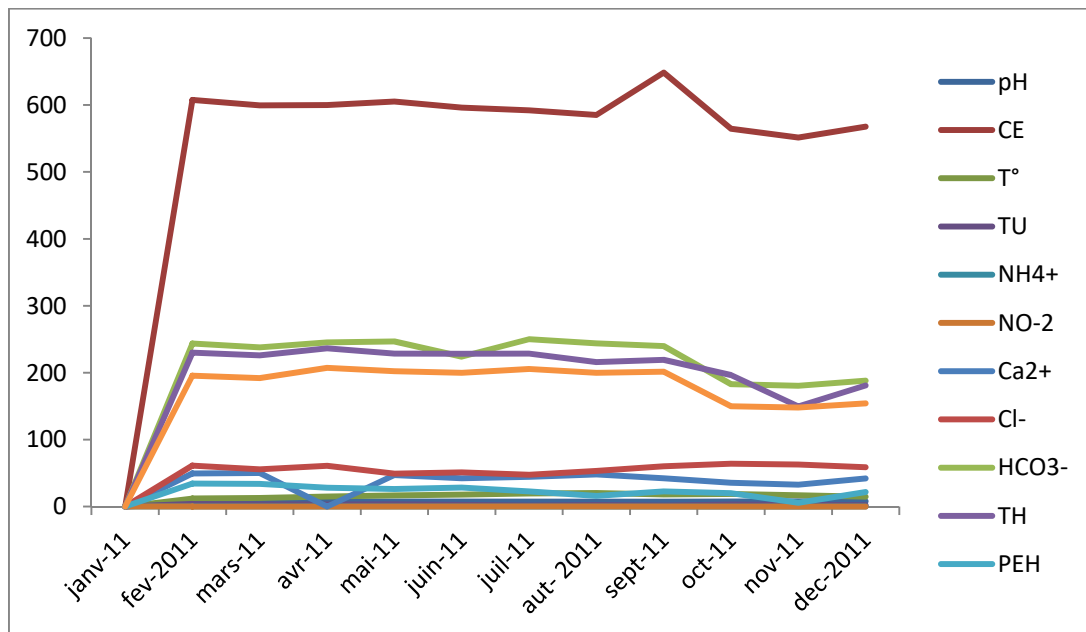


Fig. 3.4. Evolution temporelle des paramètres descripteurs.

Les caractéristiques statistiques descriptives des variables physico-chimiques utilisées dans cette étude concernent les valeurs minimales et maximales, la moyenne (tableau 3.1 et figure 3.5). Les résultats montrent que le pH est légèrement alcalin varie entre 6,9 et 8,73 avec une valeur maximale observée durant le mois octobre. La minéralisation des eaux est relativement élevée avec une conductivité électrique à 25 °C qui varie entre 14,9 et 6129 ms/cm, cette dernière est observée pendant le mois septembre. La température varie entre 9,5 et 24,6°C, cette dernière est observée pendant le mois août. La Turbidité varie entre 0,32 et 8,59 avec une valeur maximale enregistrée pendant le mois d'avril. En ce qui concerne les

concentrations en Ammonium et en Nitrite, les résultats augmentent respectivement de 0 à 0,93 et de 0 à 0,46 mg/l. La concentration de la Calcium Ca^{2+} varie entre 27,2 et 60,8 mg/l, avec une valeur maximale observée durant le mois avril. Le Chlorures Cl^- varie entre deux valeurs minimales et maximale respectivement 36,21 et 79,52 mg/l, cette dernière est observée pendant le mois janvier. La concentration de la Bicarbonate HCO_3^- dans l'eau d'étude varie entre 141,56 et 289,14 mg/l avec une valeur maximale enregistrée pendant avril. Les résultats d'étude de la concentration de la Dureté Totale et de la Dureté permanente augmentent respectivement de 122 à 252 et de 0 à 44 mg/l. Le Titre alcalin complet varie entre les valeurs suivant 130 et 237 mg/l, cette dernière est observée pendant le mois avril.

L'objectif qui se trouve derrière la collecte des données relatives à ces paramètres est de trouver deux modèles à savoir : un modèle analytique par le calcul de l'indice IQE et un autre modèle par apprentissage, permettant de distinguer les états de la qualité de l'eau. Des données descriptives expérimentales recueillies sur une longue période (plusieurs années) pourraient atteindre cet objectif. 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 étapes de simulation sont distinguées à savoir, une étape préliminaire de calcul de l'IQE analytiquement et une autre étape par la suite utilisant une approche par reconnaissance de formes dans un but de prédiction de l'indice IQE. Pour la première, les différents paramètres descripteurs de l'état de l'eau sont calculés par un ensemble des formules analytiques (3.1, 3.2, 3.3, 3.4) en respectant les normes recommandées par les pouvoirs publics. Le deuxième cas d'étude représentant un cas de prédiction plus élaboré, comportant le calcul analytique de l'indice IQE comme sortie de modèle à prédire et appliqué sur les douze paramètres physico-chimiques comme des entrées correspondantes.

Dans ce qui suit nous nous intéressons au développement d'un outil d'aide à la décision de la qualité de l'eau par les deux étapes citées auparavant (Analytique et prédiction). L'information de l'indice de qualité (IQE) issue de ce modèle de prédiction après un calcul analytique devrait être utilisée au niveau de la chaîne de surveillance comme étant une décision sur la qualité de l'eau brute dans la procédure d'analyse et de classification de différents états de la qualité de l'eau. Les résultats de prédiction ainsi obtenus pour ce paramètre (IQE), sont illustrés à travers une évaluation des performances pour les deux phases à savoir, l'apprentissage et la qualité de prédiction.

L'objectif principal de cette recherche est de réduire considérablement le temps et les efforts nécessaires au calcul de l'IQE dans les zones à surface libre. La technique SVM a été utilisée comme technique robuste pour déterminer l'IQE. L'eau de surface de barrage de Tilesdit a été utilisée sur une période de 12 mois de l'année 2011 et un vaste ensemble de données a été collecté pour 12 variables de qualité de l'eau. Une analyse statistique a été utilisée. Enfin, la précision de la méthode SVM a été évaluée en utilisant une comparaison entre les résultats obtenus.

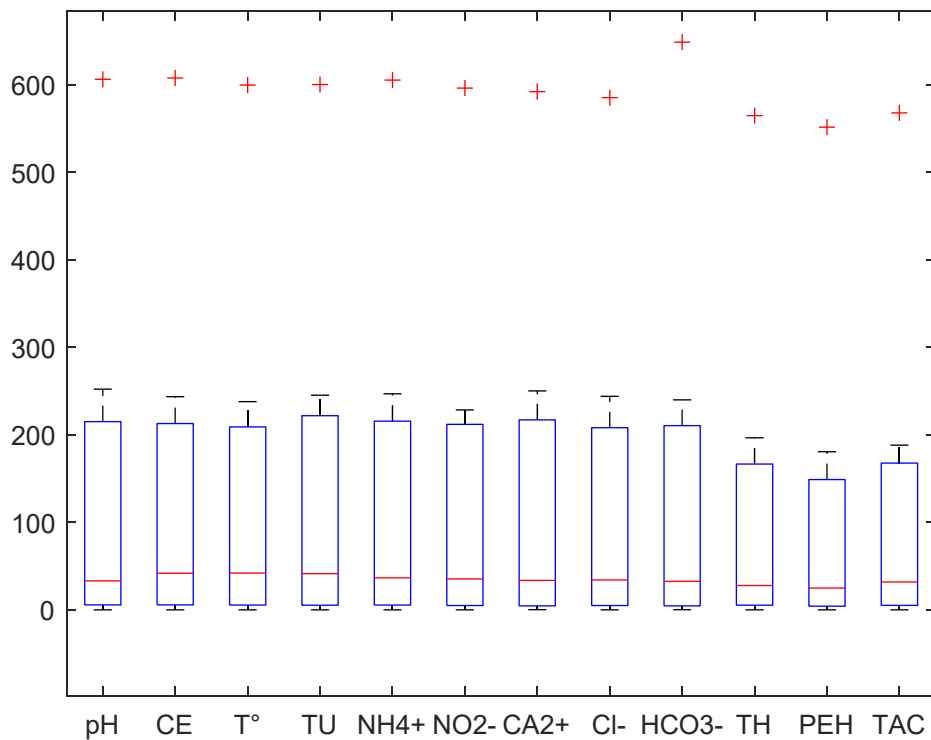


Fig. 3.5. Diagramme en boîte à moustaches des paramètres physico-chimiques de l'année 2011.

Tableau. 3.1. Statistiques descriptives des paramètres physico-chimiques relatives aux douze mois de prélèvement de l'année 2011.

		PARAMETRES PHYSICO-CHIMIQUES				PARAMETRES DE POLLUTION (Mg/l)							
		pH	Conductivité à 25°C (ms/cm)	Température (°C)	Turbidité (NTU)	NH ₄ ⁺ (mg/l)	NO ⁻² (mg/l)	Ca ²⁺ (mg/l)	Cl ⁻ (mg/l)	HCO ₃ ⁻ (mg/l)	TH (CaCO ₃) (mg/l)	PEH(CaCO ₃) (mg/l)	TAC (mg/l)
Jan 2011	Min	7.34	558	9.5	1.89	0	0	44.8	44.02	242.78	218	6	199
	Moy	7.54	606.08	12.30	3.89	0.36	0.07	48.8	61.20	252.05	223.6	17.52	206.6
	Max	7.91	650	14.6	4.99	0.93	0.197	53.6	79.52	261.08	238	39	214
Fev 2011	Min	7.26	578	10.2	2.35	0.12	0.032	44	52.54	225.7	224	21	185
	Moy	7,60	607,52	12,18	3,83	0,23	0,05	49,6	61,30	243,60	230	34,33	195,67
	Max	7.85	624	14.7	5.6	0,31	0,07	55.2	73,13	254,98	236	43	209
Mar 2011	Min	7.16	577	10.8	1.79	0.11	0.03	48	51.12	233.02	222	31	190
	Moy	7,68	599,46	12,96	3,45	0,20	0,07	50,13	55,62	237,9	226	34	192
	Max	7.97	609	15.7	5.91	0,27	0,10	52	58.93	242,78	230	36	195
Avr 2011	Min	6.9	14.7	13.2	1.41	0,08	0,07	42.4	56.8	207.4	218	13	195
	Moy	7,59	600	15,18	3,21	0,15	0,09	54.4	61,07	245,22	236,5	28,25	207,25
	Max	8.16	643	18.2	8.59	0,20	0,13	60.8	68.9	289.14	250	44	237
Mai 2011	Min	7.23	546	12.4	1.54	0	0,03	45.6	36.21	237.9	228	20	195
	Moy	7,58	605,15	16,82	3,38	0,01	0,09	47	49,23	246,85	228,67	26,33	202,33
	Max	8,12	631	19,8	7,01	0,026	0,29	49,6	56,8	253,76	230	35	208
Jui 2011	Min	7.15	575	13.8	1.48	0,01	0,02	33,6	49,7	141,56	220	18	195
	Moy	7,46	596,06	18,01	2,54	0,04	0,28	42,24	51,16	224	228,4	28,4	200
	Max	7,72	615	21,4	3,7	0,08	0,46	51,2	53,96	250.1	238	33	205
Juil 2011	Min	7,31	579	17,8	0,57	0,10	0	30	42,6	245,22	218	9	201
	Moy	7,469	591,99	19,54	1,608	0,252	0,14	44,56	47,57	250,09	228,6	22,8	205,6
	Max	7,72	612	24,4	3,07	0,47	0,43	52,8	51,12	254,98	252	44	209
Aout 2011	Min	7.16	546	18.2	1.1	0,01	0	48	53,25	244	216	16	200
	Moy	7,44	585	20,34	2,29	0.25	0,02	48	53,25	244	216	16	200
	Max	7,66	596	24,6	4,05	0,486	0,0329	48	53,25	244	216	16	200

Sep 2011	Min	7,2	560	14,9	0,32	0,128	0	32,8	58,93	225,7	210	5	200
	Moy	7,46	648,59	18,08	1,73	0,22	0,15	42,4	60,35	239,93	219,33	22,67	201,67
	Max	7,76	6129	22,7	5,43	0,37	0,30	48	61,77	250,1	230	33	205
Oct 2011	Min	7,26	504	12,8	1,61	0	0	29,6	58,93	165,92	156	2,2	136
	Moy	7,71	564,51	18,79	3,09	0,09	0	35,6	64,08	183	196,5	20,05	150
	Max	8,73	614	22,5	5,51	0,28	0	44,8	74,55	225,7	250	35	185
Nov 2011	Min	7,06	533	11,4	1,05	0	0	27,2	58,93	174,46	122	0	143
	Moy	7,88	551,46	16,99	2,71	0,08	0	32,76	62,91	180,63	149,6	5,74	148,06
	Max	8,26	588	20,8	4,83	0,179	0	36,6	64,61	187,02	160	11	153,3
Dec 2011	Min	7,33	554	12,1	1,13	0,13	0	39,2	53,96	158,62	166	8,85	130
	Moy	7,89	567,69	14,71	2,22	0,19	0,02	42	58,93	188,19	181	21,71	154,25
	Max	8,22	594	17,5	4,32	0,35	0,03	44,8	62,48	213,5	204	36	175

Tableau. 3.2. Normes des paramètres descripteurs de l'eau brute.

Propriété	TU (mg/l)	Conductivité (C) (□S)	pH (uph)	Température (T°) (°C)	Oxygène dissous (OD) (mg/l O ₂)
Potabilité	≤2	400 ≤ · ≤ 2500	6.5 ≤ · ≤ 9	12 ≤ · ≤ 25	4 ≤ · ≤ 8

2.2.2. Calcul de l'Indice de qualité de l'eau (IQE)

Comme mentionné précédemment, pour déterminer l'IQE de la surface de l'eau, Douze paramètres importants telles que pH, CE, T°, TU, NH₄⁺, NO₂⁻, Ca²⁺, Cl⁻, HCO₃⁻, TH, PEH et TACont été sélectionnés pour calculer l'Indice de qualité de l'eau IQE. Le développement de l'IQE est une technique de classification de la qualité de l'eau qui repose sur la comparaison des paramètres de qualité de l'eau avec les normes internationales ou nationales dans le cadre de cette étude. Ces variables doivent être converties en variables non dimensionnelles à l'aide de fonctions de sous-indice (SI). Le tableau en annexe montre les différentes concentrations et les valeurs requises.

L'IQE résume de grandes quantités de données sur la qualité de l'eau en termes simples (Excellente, Bonne, Mauvaise, Très mauvaise, etc.). Dans cette étude l'indice IQE est appliqué pour estimer l'influence des facteurs naturels et anthropiques sur la base de plusieurs paramètres clés du chimisme des eaux de surface de barrage Tilesdit. Cet indice est calculé en suivant la méthode de l'indice arithmétique pondéré [50]. Dans cette approche, une valeur numérique appelée poids relatif (W_i), spécifique à chaque paramètre physico-chimique, est calculée (Tableau. 3.3) selon la formule suivante :

$$W_i = \frac{K}{S_i} \quad (1.1)$$

$$K = \frac{1}{\sum_{i=1}^n \left(\frac{1}{S_i}\right)} \quad (1.2)$$

Où :

W_i : le poids de chaque paramètre selon son importance relative dans la qualité des eaux de boisson,

K : constante de proportionnalité et peut également être calculée à l'aide de l'équation suivante :

n : nombre de paramètres,

S_i : valeur maximale de la norme standard algérienne des eaux de surface de chaque paramètre en *mg/l* sauf pour le pH, T° et la conductivité électrique.

Ensuite, une échelle d'évaluation de la qualité (Q_i) est calculée pour chaque paramètre en divisant la concentration par la norme dudit paramètre et en multipliant l'ensemble par 100 comme dans la formule suivante :

$$Q_i = \left(\frac{C_i}{S_i} \right) \times 100 \quad (1.3)$$

Q_i : échelle d'évaluation de la qualité de chaque paramètre.

C_i : la concentration de chaque paramètre en mg/l.

Finalement l'indice global de la qualité de l'eau (IQE) est calculé par l'équation suivante :

$$IQE = \frac{\sum_{i=1}^n Q_i \times W_i}{\sum_{i=1}^n W_i} \quad (1.4)$$

Cinq classes de qualité peuvent être identifiées selon les valeurs de l'indice de qualité de l'eau IQE (tableau. 3.3).

Tableau. 3.3. Classification et usage possible de l'eau selon l'IQE (Brown et al, 1972 ; Chatterji et Raziuddin, 2002 ; Aher et al, 2016).

Classe d'IQE	Type d'eau	Usage possible
0 - 25	Excellente qualité	Eau potable, irrigation et industrie
>25 - 50	Bonne qualité	Eau potable, Irrigation et industrie
>50 - 75	Mauvaise qualité	Irrigation et industrie
>75 - 100	Très mauvaise qualité	Irrigation
> 100	Eau non potable	Traitement approprié requis avant utilisation

Le poids relatif (W_i) de chaque paramètre physico-chimique et la constante de proportionnalité K sont calculés en premier lieu en utilisant les valeurs maximales de la norme standard algérienne des eaux de surface des paramètres physicochimiques étudiés [51]. Le calcul de l'indice de qualité de l'eau (IQE) de ces paramètres physico-chimiques sont illustrées dans le tableau 3.4.

Tableau. 3.4 (1/2). Poids des paramètres physico-chimiques et norme Algérienne de qualité des eaux de surface (Norme Algérienne de qualité des eaux, 2011).

paramètres	S_i valeur maximale standard	$1/S_i$	W_i	Q_i (Janv-11)	Q_i (Fevr-11)	Q_i (Mars-11)	Q_i (Avri- 11)	Q_i (Mai-11)	Q_i (Juin-11)	Q_i (Juil-11)
pH	7	0.142	0.019	107.711	108.581	109.674	108.397	108.352	106.521	106.631
Conductivité à 25 °C Ms/cm	2700	0.000	0.000	22.447	22.500	22.202	22.222	22.412	22.076	21.925
Température °C	25	0.040	0.005	49.187	48.704	51.858	60.712	67.27	72.02	78.151
Turbidité (NTU)	5	0.200	0.026	77.885	76.686	69.027	64.289	67.641	50.762	32.067
Ammonium (mg/l)	0.5	2.000	0.268	72.975	45.85	40.008	30.916	1.776	7.588	49.624
NO ⁻² (mg/l)	0.2	5.000	0.67	33.286	27.2	33.943	45.045	45.755	138.25	69.09
Ca ²⁺ (mg/l)	200	0.005	0.000	24.4	24.8	25.066	27.2	23.5	21.12	22.28
Cl ⁻ (mg/l)	500	0.002	0.000	12.240	12.259	11.123	12.213	9.845	10.231	9.514
HCO ₃ ⁻ (mg/l)	50	0.020	0.002	504.104	487.186	475.8	490.44	493.693	448	500.184
TH (mg/l)	200	0.005	0.000	111.8	115	113	118.25	114.333	114.2	114.3
Dureté permanente (mg/l)	200	0.005	0.000	8.76	17.166	17	14.125	13.166	14.2	11.4
Titre alcalin complet (mg/l)	500	0.002	0.000	41.32	39.133	38.4	41.45	40.466	40	41.12

Tableau. 3.4 (2/2). Poids des paramètres physico-chimiques et norme Algérienne de qualité des eaux de surface (Norme Algérienne de qualité des eaux, 2011).

Paramètres	S_i valeur maximale standard, Algérie	$1/S_i$	W_i	Q_i (Aout-11)	Q_i (Sept-11)	Q_i (Oct-11)	Q_i (Nov-11)	Q_i (Dec-11)
pH	7	0.142	0.019	106.231	106.603	110.141	112.539	112.765
Conductivité à 25 °C Ms/cm	2700	0.000	0.000	21.666	24.021	20.907	20.424	21.025
Température °C	25	0.040	0.005	81.375	72.323	75.158	67.981	58.857
Turbidité (NTU)	5	0.200	0.026	45.718	34.624	61.730	54.244	44.354
Ammonium (mg/l)	0.5	2.000	0.268	49.8	43.34	18.537	15.622	37.93
NO ⁻² (mg/l)	0.2	5.000	0.67	8.225	77.372	0	0	7.862
Ca ²⁺ (mg/l)	200	0.005	0.000	24	21.2	17.8	16.38	21
Cl ⁻ (mg/l)	500	0.002	0.000	10.65	12.07	12.815	12.581	11.786
HCO ₃ ⁻ (mg/l)	50	0.020	0.002	488	479.866	366	361.264	376.38
TH (mg/l)	200	0.005	0.000	108	109.666	98.25	74.8	90.5
Dureté permanente (mg/l)	200	0.005	0.000	8	11.333	10.025	2.87	10.856
Titre alcalin complet (mg/l)	500	0.002	0.000	40	40.333	30	29.612	30.85
	$\sum \frac{1}{S_i}$	7.42						
	$K = \frac{1}{\sum (\frac{1}{S_i})}$	0.13						

Après le calcul de l'indice global de qualité IQE par l'équation (3.4) en utilisant les résultats d'analyses physico-chimiques et les valeurs standards de la norme Algérienne [51], la classification des différents états de qualité des eaux est déterminée. IQE est un nombre sans unité qui varie entre 0 et 100, où une valeur faible d'IQE représente une qualité d'eau élevée (bonne) et une valeur élevée de WQI représente une qualité d'eau faible (mauvaise). Sur la base de cet indice, la qualité de l'eau peut être classée en cinq classes. Le tableau 3.3 montre Classification et usage possible de l'eau selon l'IQE. Le tableau 3.5 présente les résultats de classification de qualité de l'eau vis-à-vis l'indice IQE de l'année 2011.

Tableau. 3.5. Valeurs de l'indice IQE et classe de qualité des eaux superficielles du barrage Tilesdit pendant les douze mois de l'année 2011.

Mois	IQE	Classe de qualité de l'eau
Janv-11	47.66	Bonne qualité
Fevr-11	36.15	Bonne qualité
Mars-11	38.94	Bonne qualité
Avr-11	43.92	Bonne qualité
Mai-11	36.64	Bonne qualité
Juin-11	100.26	Eau non potable
Juil-11	64.48	Mauvaise qualité
Aout-11	23.68	Excellente qualité
Sept-11	68.38	Mauvaise qualité
Oct-11	9.87	Excellente qualité
Nov-11	8.89	Excellente qualité
Dec-11	19.97	Excellente qualité

Quatre classes de qualité dans notre cas peuvent être identifiées selon les valeurs de l'indice de qualité de l'eau IQE obtenus (Excellente, bonne, mauvaise et non potable) sont identifiées lors des mois Juin, Juillet et Septembre. L'indice de qualité de l'eau IQE montre qu'au passage automne -hiver -printemps Octobre - Mai, sont de bonne à excellente qualité ($8.89 < IQE \leq 47.66$). Les mois de Juin montre une qualité de l'eau non potable ($IQE > 100$). Les mois de Juillet et Septembre montrent une qualité mauvaise ($IQE > 50 - 75$).

3. PREDICTION D L'INDICE IQE PAR SVM

Lorsque les SVM sont utilisés dans des problèmes de régression pour prédire des valeurs réelles, on parle des SVR (Support Vector Régression). « Les SVM peuvent également être mis en œuvre en situation de régression, c'est-à-dire pour l'approximation de fonctions quand Y est quantitative. Dans le cas non linéaire, le principe consiste à rechercher une estimation de la fonction par sa décomposition sur une base fonctionnelle. La forme générale des fonctions calculées par les SVM se met sous la forme :

$$\emptyset(x, w) = \sum_{i=1}^{\infty} W_i \cdot V(x) \quad (2.8)$$

Le problème se pose toujours comme la minimisation d'une fonction coût, mais plutôt que d'être basée sur un critère d'erreur quadratique (moindres carrés), celle-ci s'inspire des travaux de Huber sur la recherche de modèles robustes et utilise des écarts absolus [55]. On note $|\cdot|_{\epsilon}$ la fonction qui est paire, continue, identiquement nulle sur l'intervalle $[0, \epsilon]$ et qui croît linéairement sur $[\epsilon, +\infty]$. La fonction coût est alors définie par :

$$E(W, \gamma) = \frac{1}{n} \sum_{i=1}^n |Y - \emptyset(x, w)| + \gamma \|w\|^2 \quad (2.9)$$

Où γ est, comme en régression rigide, un paramètre de régularisation assurant le compromis Entre généralisation et ajustement. De même que précédemment, on peut écrire les solutions du Problème d'optimisation. Pour plus de détails, se reporter à Schölkopf et Smola[56].

Dans cette partie de travail on s'intéresse à la conception d'un modèle de prédiction de l'indice de qualité de l'eau (IQE) (figure 3.6) permettant de fournir une information sur la valeur unique de l'indice IQE à partir d'autres paramètres physico-chimiques. 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 modèle est l'indice de qualité de l'eau estimée.

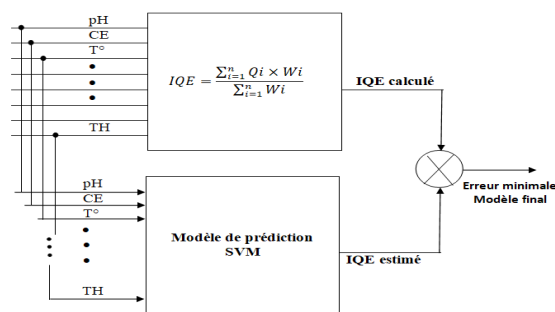


Fig. 3.6. Modèle de prédiction de l'indice IQE par SVM.

Le but recherché quant à l'application de la méthode choisie, se limite à une simulation appliquée sur des données réelles utilisés auparavant. L'indice de qualité de l'eau (IQE) est déduit d'une combinaison non-linéaire des 12 paramètres précédemment annoncés (figure 3.4 et tableau 3.1). Le modèle SVM possède 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.

3.1.Apprentissage et Généralisation

On présente d'abord un ensemble de données extrait d'une base réelle et constitué de 144 échantillons, correspondant aux douze paramètres physico-chimiques avec la valeur de l'IQE. Les résultats d'apprentissage obtenus avec le modèle SVM utilisant la même base précédente, sont présentés dans la figure 3.7 pour la valeur du facteur de pénalisation $C = 10$, avec pour fonction noyau polynomiale ayant les paramètres ($\epsilon = 0.05, d = 2$).

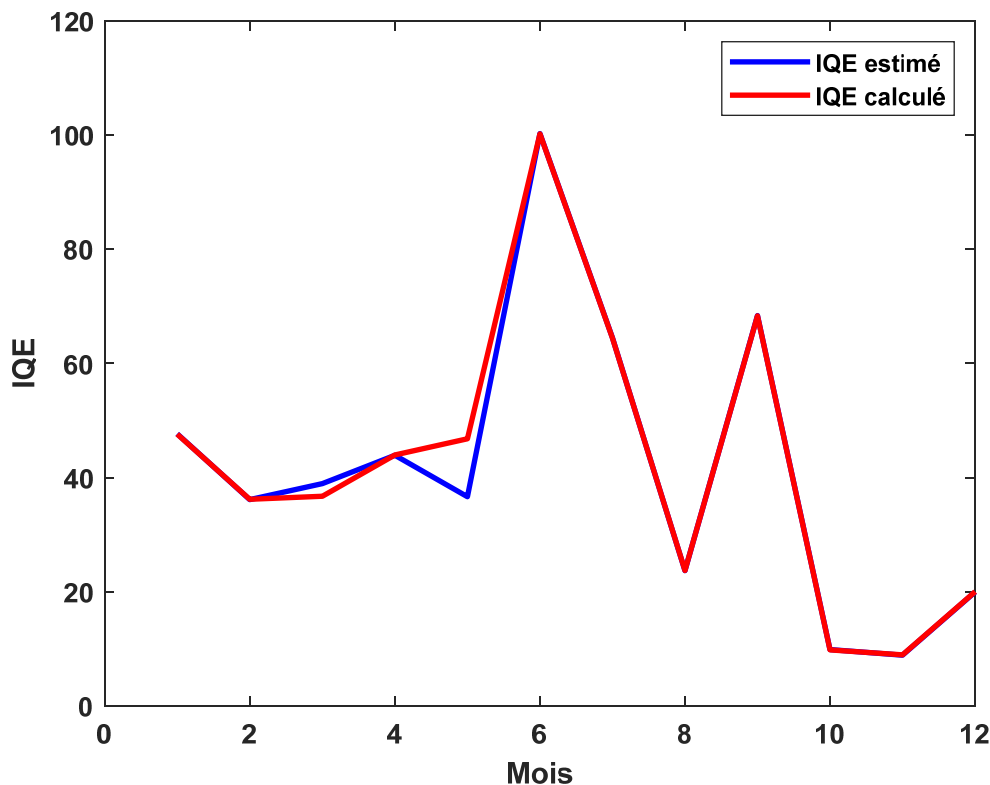


Fig.3.7. Résultats d'apprentissage du modèle SVM.

Une bonne convenance de la fonction noyau polynomiale comparativement aux autres noyaux testés pour ce modèle est constatée. La figure 3.7 montre les résultats de prédiction ponctuelle obtenue sur l'ensemble d'apprentissage utilisé.

Pour procéder à la généralisation, les résultats de généralisation appliquée sur l'ensemble de test, sont illustrés dans la figure 3.8. On montre clairement que la sortie estimée de l'indice IQE par le modèle SVM et celle de la valeur de l'indice IQE calculé sont plus proches ($R^2=0.99$).

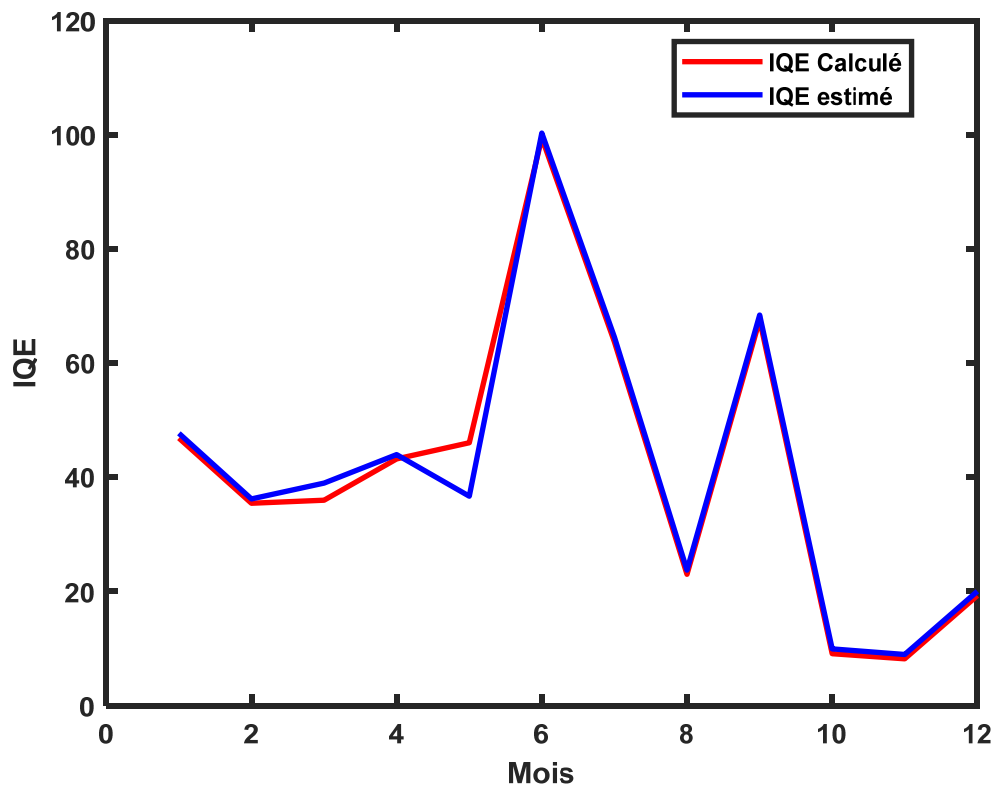


Fig.3.8. Résultats de test du modèle SVM.

D'après ces résultats, on trouve que le coefficient de corrélation ($R^2=0.99$) entre les valeurs estimées et réelles de l'IQE est tout à fait satisfaisant. La précision de l'estimation est confirmée par l'absence de valeurs aberrantes. Une bonne aptitude en phase de généralisation justifiée par de forts taux de reconnaissance, a été obtenue pour le modèle SVM. Il est clairement démontré que les sorties calculées et estimées dans les deux phases sont significativement plus proches. C'est confirmer ce fait par la comparaison visuelle donnée dans ces figures où il y a une superposition presque totale des données de sortie simulées et réelles.

CONCLUSION

Dans ce dernier chapitre, la technique SVM a été utilisée pour développer l'IQE dans la zone d'étude à surface libre. Présentée pour un problème de prédiction, la technique a fourni de très bons résultats de simulation. La base de données réelle provenant de la station Tilesdit de Bouira et appliquée à la technique en question sur une période de 12 mois, et un vaste ensemble de données a été collecté pour 12 variables de qualité de l'eau, a montré l'efficacité de cette approche. L'indice de la qualité globale de l'eau IQE était un outil très utile pour prendre la bonne décision et évaluer de façon comparative la qualité de l'eau dans le temps et dans l'espace. Par la suite, les douze paramètres importants, ont été choisis pour développer la méthode SVM. Une valeur élevée du coefficient de corrélation ($R^2 = 0,99$) est indiqué. Cette méthode SVM était capable de prédire avec succès l'IQE avec une grande précision. Cette recherche met en évidence que les méthodes à base d'apprentissage statistique peuvent être utilisées comme des méthodes puissantes et très fiables pour estimer la qualité de l'eau dans les zones humides et autres plans d'eau. Ces techniques sont fortement recommandées pour être utilisées pour des évaluations précises, rapides et rentables de la qualité de l'eau pour tout système de production dans le monde.

Conclusion générale



CONCLUSION GENERALE

L'évaluation de la qualité de l'eau par des indices avec plusieurs types de fonctions d'agrégation est appliquée depuis de nombreuses années. Cependant, l'utilisation de ces indices doit toujours être liée à la zone dans laquelle la masse d'eau est influencée, le climat de la région, la géologie, les activités développées autour de celle-ci et son utilisation par la population, afin qu'il soit possible d'identifier quelles variables physiques, chimiques et biologiques doivent être mesurées. Le type de masse d'eau, lenticule ou l'otique, influencera également ce choix de variables pour la construction de l'indice et un autre facteur important qui doit être inséré dans celui-ci concerne les impacts du changement climatique. Dans ce contexte, pour atteindre ces objectifs, il est nécessaire de développer des indices moins influencés par les effets d'éclipse et d'ambiguïté, nécessitant l'utilisation de meilleures fonctions d'agrégation et de statistiques, comme le suggèrent plusieurs autres.

En ce sens, les techniques statistiques multi-variées, les régressions, l'apprentissage automatique et l'intelligence artificielle, sont désormais pertinentes et importantes pour la création d'indices qui englobent les variables les plus diverses pour chaque usage spécifique de l'eau. De plus, cette création facilitera la prise de décision et la communication entre les gestionnaires de l'environnement et la société. Il est important de souligner que de telles méthodes permettraient une plus grande flexibilité dans la sélection des variables, rendant les indices de qualité de l'eau plus affirmés, efficaces et plus faciles à comprendre et à manipuler.

Cette étude a été structurée autour de trois chapitres essentiels. Le premier consacré à une introduction au domaine de l'eau, a permis de présenter des généralités ainsi que les différentes méthodes de contrôle et de surveillance employées. Une description détaillée des différents paramètres caractéristiques de l'état l'eau, appelés : paramètres descripteurs qui sont relatifs à la qualité de l'eau potable en particulier, ainsi que leurs capteurs physico-chimiques correspondants et le calcul de l'indice de qualité de l'eau à travers ces paramètres descripteurs, ont été décrits. Le deuxième chapitre a été particulièrement dédié aux mécanismes théoriques des méthodes de régression de données à apprentissage statistique

Références bibliographiques



supervisé. Dans ce chapitre, le modèle SVM fondé sur ce type d'apprentissage ont été exposés.

Enfin le troisième et dernier chapitre, a fait l'objet d'une étude en simulation concernant la mise en œuvre de modèle SVM d'apprentissage statistique appliqués dans le domaine du contrôle et de surveillance des eaux potables après une étude préliminaire où on fait une modélisation théorique et analytique de l'indice de la qualité de l'eau (IQE). Cette étude a permis la validation et l'évaluation des performances de la méthode présentée. La notion de modèle de prédiction, conçus dans le domaine, comme étant un moyen intéressant pouvant remplacer le calcul analytique dans un but de réduire considérablement le temps et les efforts nécessaires au calcul de l'IQE, a été présentée. Présentée pour un problème de régression, la technique étudiée a obtenu des résultats de simulation très impressionnants dans la décision de la qualité de l'eau en un temps très court que le temps pris par les méthodes de calcul traditionnelles a fourni de très bons résultats. La base de données réelle provenant de la station Tilesdit de Bouira et appliquée à la technique en question, a montré l'efficacité de l'approche. On peut dire qu'à l'avenir nous pouvons développer et trouver des techniques plus rapides et plus précises qui nous aident de décider sur la qualité de l'eau afin de contribuer à faciliter la consommation d'eau pour les citoyens.

Les performances ainsi obtenues peuvent être alors améliorées. En effet, une base de données réelle plus importante et plus significative, contribue sans doute à augmenter la précision de reconnaissance. Toutefois, le principal souci pour l'application de tel modèle (SVM) est l'obtention d'une base de données « optimale ». Ceci met évidemment en jeu le nombre et le type d'exemples à utiliser dans la base d'apprentissage. Comme souligné auparavant, la présence d'un expert (ou système expert) serait indispensable dans ce cas là. Le temps correspondant à la phase d'entraînement reste relativement important, ce qui laisse envisager d'autres outils de calcul plus puissants afin d'améliorer les capacités et obtenir plus de performances. Le contrôle de potabilité peut par contre être pris en charge de façon dynamique par le système de surveillance multi-capteur, puisque le temps d'exécution est faible.

Les horizons de l'application de cette technique SVM restent prometteurs. La décision du système peut être améliorée par l'exploitation de nouveaux paramètres d'entrée et d'autres indices de qualité plus générale et plus étudiés. Les capteurs logiciels peuvent dans ce cas jouer un rôle primordial en se substituant davantage à des paramètres descripteurs chimiques ne pouvant être mesurés en continu. Il reste à noter que la sensibilité du domaine à des

CONCLUSION GENERALE

menaces imprévues, exigent de plus grands efforts pour maximiser l'immunité du système et apporter d'autres améliorations afin de minimiser les risques encourus pour la santé publique. Enfin, cette application montre une alternative prometteuse pour notre pays dans l'avenir, pour une surveillance intelligente, automatique et efficace de la qualité des eaux potables.

Références bibliographiques

- [1] Boeglin Jean-Claude. Propriétés des eaux naturelles. Technique de l'ingénieur, traité environnement, G1 110.
- [2] Bremond R., Vuichard R., 1973. Paramètres de la qualité des eaux, OSPEPE, Paris.
- [3] Bharti N. and Katyal D. (2011). Water Quality Indices Used for Surface Water Vulnerability Assessment. *International Journal of Environmental Sciences*2(1), 154-71.
- [4] Cude C. G. (2001). Oregon water quality index a tool for evaluating water quality management effectiveness. *JAWRA Journal of the American Water Resources Association* 37 (1), 125-37.
- [5] Abbasi T. and Abbasi S. A. (2012). *Water quality indices*. Elsevier.
- [6] Lumb A., Sharma T. C. and Bibeault J.-F. (2011). A Review of Genesis and Evolution of Water Quality Index (WQI) and Some Future Directions. *Water Quality, Exposure and Health*3(1), 11-24.
- [7] Première année GPEE CHIMIE DES EAUX) (<http://www.ensa-agadir.ac.ma>).
- [8] Bouziane, M., (2000) - L'eau de la pénurie aux maladies. Ed., Ibn khaldoun, pp (11), Alger. (2006)
- [9] GAUJOU, 1995 La pollution des milieux aquatiques : aide-mémoire (2ème éd.) Traitement des eaux de refroidissement.
- [10] M. D. D. E. P, 2006 ; Ministère du Développement durable, de l'Environnement et des Parcs du Québec.
- [11] HAMED. M et al, 2012 Etude des propriétés physicochimiques et bactériologiques de l'eau du barrage DJORF- TORBA Bechar. Mémoire d'Ingénieur d'état en Biologie. Université de Bechar. pp 7, 9.
- [12] BRADAI. M. N, 1994, Mémoire sur la floraison et la fructification de deux phanérogames marines sur les côtes sud-est de la Tunisie. par asma Hamza, Mohamed)

Références bibliographiques

nedjmeddine Bradai, institue national scientifique et technique d'océanographie et d'écologie, 3029 sfax (Tunisie). Mar. Life 1994.

[13] CHEVALLIER. H, 2007 : Titre de livre : L'eau un enjeu pour demain. p 26, ETAT DES LIEUX ET PERSPECTIVES, Sang de la terre.

[14] Centre International De Formation (C.I.D.F.), Principes généraux de traitement des eaux, Lyonnaise des Eaux.

[15] <http://pravarini.free.fr>.

[16] Thèse de DOCTORAT SCIENCES THEME Contribution au Développement de Systèmes Multi capteurs Intelligents Dédiés à la Surveillance et au Contrôle de la Qualité des Eaux Propres, 2019.

[17] FILTRATION LENTE SUR SABLE ,1974 : PRESENTATION GENERALE) (<http://www.oieau.fr>).

[18] <http://m.actu-environnement.com>.

[19] 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.

[20] M. R. Zemouri, Contribution à la surveillance des systèmes de production à l'aide des réseaux de neurones dynamique, Application à la e-maintenance, Thèse de doctorat, Université de Franche-Comté, 2003.

[21] H. Hernández, Supervision et diagnostic des procédés de production d'eau potable, Thèse de doctorat, L'INSA de Toulouse 2006.

[22] https://www.oieau.org/ReFEA/fiches/AnalyseEau/Physico_chimie_PresGen.htm.

[23] Armani mohammed.pdf ,1/8/2020).(<http://dlibrary.univ-boumerdes.dz>).

[24] M. Combacau, "Commande et surveillance des systèmes à événements discrets complexes : Application aux ateliers flexibles", Thèse de doctorat, Université Paul Sabatier, Toulouse, France1991.

- [25] D. Racoceanu, "Contribution à la surveillance des systèmes de production en utilisant les techniques de l'intelligence artificielle", Thèse d'HDR, Université de Franche-Comté de Besançon, France, N° d'ordre : 140, 2006.
- [26] W. Schon, K. Odeh, T. Denoeux, F. Fotoohi, Maîtrise des risques dans le domaine de l'eau potable, In Actes du 12ème Colloque National de Sécurité de Fonctionnement, Laboratoire SIME Système Intelligents pour la Maîtrise de l'Eau, Montpellier, France, pages 695-701, March 2000.
- [27] N. Valentin, Construction d'un capteur logiciel pour le contrôle automatique des procédés de coagulation entrainement d'eau potable, Thèse de doctorat, Laboratoire des Eaux, UTC, 2000.
- [28] SURVEILLANCE DE LA QUALITÉ DE L'EAU DANS LE RÉSEAU DE DISTRIBUTION ,14/2/2005.<http://fcm.ca>
- [29] M. Ladjal, "Contribution au développement de systèmes de surveillance innovants dédiés au contrôle de la qualité des eaux potables", Thèse de doctorat, Université de M'sila, Algérie, 2013.
- [30] K.COULIBALY, étude de la qualité physico-chimique et bactériologique de l'eau des puits de certains quartiers du district de Bamako, UNIVERSITÉ DE BAMAKO, Avril 2005.
- [31]I. Manssouri, Calcul De L'indice De Qualité De L'eau (IQE) Pour L'évaluation De La Qualité Physico-Chimique Des Eaux Superficielles De L'Oued Moulouya (NE, Maroc), European Scientific Journal January 2020.
- [32] Paul Viola and Michael Jones. Robust real-time object detection. In Second international work shop on statistical and computation theories of vision, Vancouver, Canada, July 13 2001.
- [33] www.polytech.univ-tours.fr; Projet de Fin d'études SVR avec Boosting pour la prévision à long terme, 2011-2012.
- [34] M.H.D.J.K.N. Ahuja, transactions on pattern analysis and machine intelligence (01_2002).

Références bibliographiques

[35] Mme Benyahia souad "application de la classe des methods d'apprentissage statistique SVM",2012.

[36] THEME : Proposé et dirigé par : Dr. DJOUANE lotfi Réalisé par : MOHAMED SALEM lehibbN° D'ordre: 2016/IMI17/89/385PROMOTION: JUIN 2016Prédiction de la Fréquence de Resonance d'une AntenneMicrobande Avec une Ouverture dans le Plan de Masse.

[37] Redouane LEKHAL, application des SVM pour la reconnaissance d'extrasystoles, mémoire soutenu publiquement le 15/06/2015.

[38] Robert G. Mogull, Second-Semester Applied Statistics, Kendall/Hunt Publishing Company, 2004.

[39] Roger Koenker et G. Basset, 'Regression quantiles', *Econometrica*, 1978, p.33-50.

[40] <http://wikistat.fr/>, 20 mai 2016.

[41] Alex J. Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14(3) :199_222, 2004.

[42] J. Mercer. Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 209 :415_446, 1909.

[43] Voir le Workshop « Kernel Learning : Automatic Selection of Optimal Kernels » à l'url : http://www.cs.nyu.edu/learning_kernels/ comme point d'entrée vers d'autres références sur ce sujet.

[44] A generalized water quality index for taiwan, SM Liou, SL Lo, SH wang- Environmental monitoring and assessment, 2004-Springer.

[45] Water quality assessment in shrimp culture using an analytical hierarchical process, JJ Carbajal-Hernandez, LP Sanchez-Fernandez... *Ecological...*, 2013-Elsevier.

Références bibliographiques

- [46] Analysis of the British Columbia water quality index for watershed managers, PA Zandbergen, K.J Hall-water quality research journal, 1998-iwapolline.com.
- [47] Utilization of the water quality index method as a classification tool, H Boyacioglu-Environmental monitoring and assessment, 2010-Springer.
- [48] Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors, NM Gazzaz, MK Yusoff, AZ Aris, H Juahir.....-Marine pollution..., 2012-Elsevier.
- [49] Etude de la qualité physicochimique et de la contamination métallique des eaux de surface du bassin versant de Beht (Maroc), L Ferdaous, B Mohammed, B Nouzha...-European scientific... , 2015-researchgate .net.
- [50] Brown et al, 1970 ; Brown et al, 1972 ; Chatterji et Raziuddin, 2002 ; Yidana et Yidana, 2010.
- [51] Norme Algérienne de qualité des eaux, 2011.
- [52] Dilmi, S., & **Ladjal, M.** (2021). A novel approach for water quality classification based on the integration of deep learning and feature extraction techniques. *Chemometrics and Intelligent Laboratory Systems*, 214, 104329. <https://doi.org/10.1016/j.chemolab.2021.104329>
<https://www.sciencedirect.com/science/article/pii/S0169743921000976>
- [53] Djerioui, M., Bouamar, **M.**, **Ladjal, M.**, & Zerguine, A. (2018b). Chlorine Soft Sensor Based on Extreme Learning Machine for Water Quality Monitoring. *Arabian Journal for Science and Engineering*, 44(3), 2033-2044. <https://doi.org/10.1007/s13369-018-3253-8> <https://link.springer.com/article/10.1007/s13369-018-3253-8>
- [54] Djerioui, M., Brik, Y., **Ladjal, M.**, & Attallah, B. (2019). Neighborhood Component Analysis and Support Vector Machines for Heart Disease Prediction. *Ingénierie des systèmes d'information*, 24(6), 591-595. <https://doi.org/10.18280/isi.240605>

[55] **Mohamed LADJAL**, Mohamed Assam OUALI, Mohamed DJERIOUI, Optimization of SVM parameters with hybrid PCA-PSO methods for water quality monitoring, 2020 International Conference on Electrical Engineering (ICEE) September 25-27, 2020, Istanbul, Turkey. <https://ieeexplore.ieee.org/document/9249881>

[56] **Mohamed LADJAL**, Mohamed Imed KHELIL, Application of machine learning techniques for predicting the WQI for water quality monitoring: a case study in Algeria, ONLINE 9th INTERNATIONAL CONFERENCE ON APPLIED ANALYSIS AND MATHEMATICAL MODELLING (ICAAMM2021), June 11 -13, 2021, Istanbul, TURKEY, Id: 9213 (Accepted for communication). <https://ntmsci.com/Conferences/DES/Participants>

[57] Mohamed Imed KHELIL, **Mohamed LADJAL**, Hybrid Predictive Models for Water Quality Assessment Based on Water Quality Index Using ANN, LSSVM and multivariate statistical Methods, ONLINE 9th INTERNATIONAL CONFERENCE ON APPLIED ANALYSIS AND MATHEMATICAL MODELLING (ICAAMM2021), June 11 -13, 2021, Istanbul, TURKEY, Id: 9246 (Accepted for communication).

[58] **M. Ladjal**, M. Bouamar M. Djerioui, Application of ANN and SVM multiclass models used for multi-sensor monitoring and measurement of surface water quality. The Mediterranean Journal of Measurement and Control, Vol. 12, No. 3, 2016. pp. 627-635

[59] M. Bouamar and **M. Ladjal**, Performance evaluation of three pattern classification techniques used for water quality monitoring. International Journal of Computational Intelligence and Applications. 11(2) (2012) 1250013-1 1250013-14.

Résumé :

L'évaluation de la qualité de l'eau par des indices avec plusieurs types de fonctions d'agrégation est appliquée depuis de nombreuses années. Cependant, l'utilisation de ces indices doit toujours être liée à la zone dans laquelle la masse d'eau est influencée, le climat de la région, la géologie, les activités développées autour de celle-ci et son utilisation par la population, afin qu'il soit possible d'identifier quelles variables physiques, chimiques et biologiques doivent être mesurées. En ce sens, les techniques de l'intelligence artificielle, sont désormais pertinentes et importantes pour la création d'indices qui englobent les variables les plus diverses pour chaque usage spécifique de l'eau. Il est important de souligner que de telles méthodes permettraient une plus grande flexibilité dans la sélection des variables, rendant les indices de qualité de l'eau plus affirmés, efficaces et plus faciles à comprendre et à manipuler. En effet, et pour de nombreuses raisons, il est souvent indispensable de disposer un modèle d'aide à la décision, dont la fonction consiste à fournir à l'utilisateur une information complète, telle qu'un paramètre physico-chimique non mesurable en continu (ou à un coût prohibitif), sous forme d'un indice global unique de la qualité de l'eau. Cet indice est ensuite utilisé comme une information dans la procédure de surveillance et le contrôle d'eau potable.

Mots-clés : Eau de surface, Indice de qualité de l'eau, Régression, SVM, Surveillance, Tilesdit.

Abstract :

For many years, indices with various sorts of aggregation functions have been used to assess water quality. However, the use of these indices should always be linked to the area in which the water body is influenced, the region's climate, the geology, the activities developed around it, and the population's use of it, in order to identify which physical, chemical, and biological variables need to be measured. Artificial intelligence approaches, in this sense, are now relevant and crucial for the production of indices that include the most diversified characteristics for each individual use of water. It's worth noting that such methodologies would

allow for more flexibility in variable selection, making water quality indexes more aggressive, efficient, and simple to comprehend and control. Indeed, having a decision support model, whose function is to provide the user with complete information, such as a physicochemical parameter that cannot be continuously measured (or at prohibitive cost), in the form of a single global index of water quality, is often necessary for a variety of reasons. This index is then used as data in the drinking water monitoring and control procedures.

Keywords: Surface water, Water quality index, Regression, SVM, Monitoring, Tilesdit.

ملخص:

لسنوات عديدة ، تم استخدام مؤشرات ذات أنواع مختلفة من وظائف التجميع لتقييم جودة المياه. ومع ذلك ، يجب أن يكون استخدام هذه المؤشرات دائماً مرتبطاً بالمنطقة التي يتأثر فيها الجسم المائي ، ومناخ المنطقة ، والجيولوجيا ، والأنشطة التي تم تطويرها حولها ، واستخدام السكان لها ، من أجل تحديد المواد الفيزيائية والكيميائية. والمتغيرات البيولوجية يجب قياسها. أصبحت مناهج الذكاء الاصطناعي ، بهذا المعنى ، ذات صلة وحاسمة الآن لإنتاج المؤشرات التي تشمل الخصائص الأكثر تنوعاً لكل استخدام فردي للمياه. تجدر الإشارة إلى أن مثل هذه المنهجيات ستسمح بمزيد من المرونة في الاختيار المتغير ، مما يجعل مؤشرات جودة المياه أكثر قوة وكفاءة وسهولة في الفهم والتحكم. في الواقع ، إن وجود نموذج دعم القرار ، الذي تتمثل وظيفته في تزويد المستخدم بمعلومات كاملة ، مثل المعلمة الفيزيائية والكيميائية التي لا يمكن قياسها باستمرار (أو بتكلفة باهظة) ، في شكل مؤشر عالمي واحد لجودة المياه ، غالباً ما يكون ضروري لمجموعة متنوعة من الأسباب. ثم يتم استخدام هذا المؤشر كبيانات في إجراءات مراقبة مياه الشرب والتحكم فيها.

الكلمات الرئيسية: المياه السطحية ، مؤشر جودة المياه ، الانحدار ، SVM ، الرصد ، تيليزديت.