

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Ecole Nationale Polytechnique



LABORATOIRE MATERIALS DE GENIE CIVIL
ET ENVIRONNEMENT



Département Hydraulique
Laboratoire de génie civil

Mémoire de projet de fin d'études
Pour l'obtention du diplôme d'ingénieur d'état en Hydraulique

**Clustérisation et cumul des données
d'observations pour l'optimisation de la
consommation énergétique dans les STEP à
boue activée.**

Présenté et soutenu publiquement le 23/06/2018 par :

Mr.FERRAOUN Rafik
Mr.BENSALEM Ouail

Composition du Jury :

Président	Mr. EH.BENYOUSSEF	Professeur	ENP
Promoteur	Mr. A.BERMAD	Professeur	ENP
Co-Encadreur	Mr. R. OULEBSIR	Docteur	USTHB
Examineurs	Mr. A.LEFKIR	MCA	ENSTP
	Mr. S.LAHLAH	Chef département	ONA

ENP 2018

E.N.P. 10, Avenue Hassen Badi, BP 182 El-Harrach, Alger 16200

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Ecole Nationale Polytechnique



LABORATOIRE MATERIAUX DE GENIE CIVIL
ET ENVIRONNEMENT



Département Hydraulique
Laboratoire de génie civil

Mémoire de projet de fin d'études
Pour l'obtention du diplôme d'ingénieur d'état en Hydraulique

**Clustérisation et cumul des données
d'observations pour l'optimisation de la
consommation énergétique dans les STEP à
boue activée.**

Présenté et soutenu publiquement le 23/06/2018 par :

Mr.FERRAOUN Rafik
Mr.BENSALEM Ouail

Composition du Jury :

Président	Mr. EH.BENYOUSSEF	Professeur	ENP
Promoteur	Mr. A.BERMAD	Professeur	ENP
Co-Encadreur	Mr. R. OULEBSIR	Doctorant	USTHB
Examineurs	Mr. A.LEFKIR	MCA	ENSTP
	Mr. S.LAHLAH	Chef département	ONA

ENP 2018

E.N.P. 10, Avenue Hassen Badi, BP 182 El-Harrach, Alger 16200

ملخص:

محطة المعالجة عبارة عن منشأة مصممة لتنقية مياه الصرف المنزلية أو الصناعية ومياه الأمطار قبل تصريفها في البيئة الطبيعية. الغرض من العلاج هو فصل المياه عن المواد غير المرغوب فيها للوسط المتلقي

الهدف من عملنا هو وضع نموذج لتحسين الطاقة الكهربائية للمحطتين بومرداس وتيبازة عن طريق التعلم العميق القائم على ثلاثة مقاربات إحصائية: "المنهج الهرمي التحليلي" و " تحليل المكون الرئيسي" و "طريقة ك-المتوسط"

الكلمات المفتاحية: التعلم العميق، تحليل المكون الرئيسي، الطريقة الهرمية التحليلية، طريقة ك-المتوسط

Abstract :

Waste water treatment plants is an installation designed to purify domestic or industrial wastewater and rainwater before discharge into the natural environment. The purpose of the treatment is to separate the water from the undesirable substances for the receiving medium.

The object of our work is the elaboration of a model of optimization of the electric energy of the two STEP Boumerdes and Tipaza by the Deep Learning based on three statistical approaches: "The analytical method hierarchical (AHP)" combined to "Principal Component Analysis (PCA)" and "The K-Means Method".

Key words: Deep Learning, Principal Component Analysis (PCA), Analytical Hierarchical Method (AHP), The K-Means Method

Résumé :

La **station d'épuration** est une installation destinée à épurer les eaux usées domestiques ou industrielles et les eaux pluviales avant le rejet dans le milieu naturel. Le but du traitement est de séparer l'eau des substances indésirables pour le milieu récepteur.

L'objet de notre travail est l'élaboration d'un modèle d'optimisation de l'énergie électrique des deux STEP Boumerdes et Tipaza par le Deep Learning en se basant sur trois approches statistiques : « La méthode hiérarchique analytique (AHP) » combinée à « L'analyses en composantes principales (ACP) » ainsi que « La méthode des K-Means ».

Mots clé : Deep Learning, Analyse en composante principale (ACP), La méthode hiérarchique analytique (AHP), La méthode des K-Means

Remerciements

Nous tenons tout d'abord à remercier Dieu le tout puissant et miséricordieux, qui nous a donné la force et la patience d'accomplir ce Modeste travail.

En second lieu, nous tenons à remercier notre encadreur Mr BERMAD Abdelmalek pour ses précieux conseils et son aide durant toute la période du travail.

Nos remerciements s'étendent également à Mr A. LAFKIR et R. OULEBSIR pour leurs bonnes explications qui nous ont éclairé le chemin de la recherche et leur collaboration avec nous dans l'accomplissement de ce travail.

Nos vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner notre travail et de l'enrichir par leurs propositions.

Nous tenons à exprimer nos sincères remerciements à tous les professeurs qui nous ont enseigné et qui par leurs compétences nous ont soutenu dans la poursuite de nos études.

Enfin, nous tenons également à remercier toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail.

Je dédie ce travail à

A mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études,

A mes chers frères, Yacine et Anis, pour leur appui et leur encouragement,

A mes très chers neveux Wassim et Tina,

A toute ma famille et particulièrement ma tante Lila et son mari Hamid pour leur soutien tout au long de mon parcours universitaire.

A tous mes amis de l'école préparatoire de Annaba, de l'école Nationale Polytechnique ainsi qu'à toute la promo de l'hydraulique avec qui j'ai passé des années inoubliables.

Que ce travail soit l'accomplissement de vos vœux tant allégués, et le fruit de votre soutien infailible,

Merci d'être toujours là pour moi.

RAFIK

Dédicace

*A celle qui m'a consacré tout son cœur, tout son Espoir qui se compte en termes de soupirs infinis, Mon Ange qui, jamais ne se plaint, ne Demande de contrepartie, à part que je sois comme elle a tant espéré. A celle qui n'a jamais Cessé de prier Dieu pour que tous mes pas soient couronnés de succès et mes vœux soient Exaucés : **Ma mère...***

*A celui qui s'est incliné pour ma droiture, qui s'est donné tant de mal pour ma dignité, Qui m'a appris que mon aise ne se réalise qu'à travers tant d'efforts fournis de ma part, qui M'a fait connaître que le savoir n'est complet qu'en lui adjoignant la morale. A celui qui a été Et restera mon soutien inépuisable me permettant de franchir tous les obstacles passés et Futurs. A celui qui m'a illuminé mon chemin : **Mon Père...***

***A ma grande- mère :** Que Dieu la protège et la prête en bonne santé et longue vie.*

***A mes frères** Ziad et Ismail ainsi qu'à **ma sœur** Nesrine, qui ont toujours été présents pour moi.*

***A mes amis** Yacine, kheir-Eddine, Mohammed, Rafik, Abderrahmane, Aissa et Fouad avec lesquels j'ai passé des moments inoubliables.*

***A mes cousins et cousines** que j'adore énormément, ZEROUAL Mohammed El-Amine, ZEROUAL Noor El-Imen et sans oublier SANFOURA Aida.*

***A toute** la promotion Hydraulique 2015-2018.*

A tous ceux-là, je dédie ce fruit de mes efforts.

Wael

Table des matières

Liste des tableaux

Liste des figures

Introduction générale.....	10
1. Généralités sur le fonctionnement des stations d'épuration	14
1.1 Introduction	14
1.2 Définition des eaux usées	14
1.3 Les différents types d'eaux usées.....	14
1.3.1 Les eaux usées domestiques	14
1.3.2 Les eaux usées industrielles.....	14
1.3.1 Les eaux usées agricoles.....	15
1.3.2 Les eaux de ruissellement.....	15
1.4 Les différentes caractéristiques des eaux usées.....	15
1.4.1 Les paramètres physico-chimiques.....	15
1.4.2 Les paramètres chimiques	16
1.5 Le fonctionnement d'une station d'épuration.....	17
1.5.1 Définition.....	17
1.5.2 Les étapes d'épuration.....	17
2. METHODES STATISTIQUE D'ANALYSE DES DONNEES	21
2.1 INTRODUCTION.....	21
2.2 Analyse en Composantes Principales (ACP)	21
2.2.1 DESCRIPTION DE LA METHODE.....	21
2.2.2 Objectifs	23
2.2.3 Application de l'ACP	23
2.2.4 Procédé d'application de l'ACP	24
2.2.5 Résultats d'application	24
2.2.6 Analyse des résultats	24
2.3 Méthode des K-means	25
2.3.1 DEFINITION	25
2.3.2 PRINCIPE DE LA METHODE	26
2.3.3 Algorithme.....	26
2.3.4 Avantages de K-means :.....	27
2.3.5 Inconvénients :	27
2.4 Processus d'Analyse Hiérarchique (AHP) :	27
2.4.1 Définition.....	27
2.4.2 Principe.....	27
2.4.3 Procédé d'application de l'AHP	27
3. RESEAUX DE NEURONES ARTIFICIELS	33
3.1 INTRODUCTION.....	33

3.2	LES RESEAUX DE NEURONES ARTIFICIELS	33
3.2.1	NOTION DE NEURONE	33
3.2.2	DEFINITION	34
3.2.3	ARCHITECTURE.....	34
3.2.4	Apprentissage	36
3.3	Deep learning	41
3.3.1	Comment fonctionne le Deep Learning.....	41
3.3.2	Quelle est la différence entre l'apprentissage automatique et le Deep Learning ?	42
4.	Présentation des STEP de Boumerdes et Tipaza	44
4.1	STEP de Boumerdes.....	44
4.1.1	INTRODUCTION.....	44
4.1.2	Localisation de la STEP	44
4.1.3	Caractéristiques Techniques.....	44
4.1.4	Paramètres de pollution de conception.....	45
4.1.5	Les étapes du procédé d'épuration :	45
4.1.6	Statistiques et critiques des données :	47
4.2	STEP Tipaza (station de Chenoua).....	53
4.2.1	Localisation	53
4.2.2	Caractéristiques Techniques.....	54
4.2.3	Description des procédés d'épuration au niveau de la station.....	54
4.2.4	Statistiques et critiques des données.....	56
5.	Modélisation de la consommation d'énergie	64
5.1	Introduction	64
5.2	APPLICATION DE L'ACP	64
5.2.1	STEP DE BOUMERDES.....	64
5.2.2	TIPAZA.....	65
5.2.3	DONNEES CUMULEES	66
5.3	Clustering par K-means.....	67
5.3.1	Regroupement par K-means.....	68
5.3.2	Résultat du Clustering	69
5.4	Elaboration du modèle d'optimisation de la consommation de l'énergie	73
5.4.1	Première approche.....	74
5.4.2	Deuxième approche.....	88
5.4.3	Calcul du gain d'énergie.....	98
5.4.4	Conclusion.....	100
	Bibliographie.....	102

Liste des tableaux

Tableau 2-1: Echelle de jugement	29
Tableau 2-2: La matrice de jugement relatif à un niveau.....	29
Tableau 2-3: Indice aléatoire élaboré par Saaty.	30
Tableau 3-1 : Fonctions d'activation.	40
Tableau 4-1: Paramètres de pollution de conception de la station de Boumerdes	45
Tableau 4-2: Statistiques descriptives des paramètres de pollution de la STEP Boumerdes ...	47
Tableau 4-3: Paramètres de pollution de conception de la station de Tipaza	54
Tableau 4-4: Statistiques descriptives des paramètres de pollution de la STEP Tipaza	56
Tableau 5-1 : Matrice de corrélation entre les CP' et les variables de l'eau brute journalière.	64
Tableau 5-2: Corrélation entre les CP' et les variables (Paramètres de l'eau brute-journalier).	65
Tableau 5-3: Corrélation entre les CP' et les variables (Paramètres de l'eau brute-journalier).	66
Tableau 5-4: Tableau récapitulatif des modèles	73
Tableau 5-5: Classement et Poids attribués à chaque paramètre sélectionné pour la STEP de Boumerdes.....	76
Tableau 5-6: Classement et Poids attribués à chaque paramètre sélectionné pour la STEP de Tipaza.....	76
Tableau 5-7: Classement et Poids attribués à chaque paramètre sélectionné pour les deux STEP.....	76
Tableau 5-8: Critères et code de l'AHP.	88
Tableau 5-9: Matrice de jugement de l'indicateur CPAS	89
Tableau 5-10: Matrice de jugement du rendement épuratoire.	89
Tableau 5-11: Matrice de jugement rejet dans la nature.	89
Tableau 5-12: Matrice de jugement du niveau 1	89
Tableau 5-13: Matrice de jugement du niveau 3.....	91
Tableau 5-14: Matrice de jugement du niveau 5.....	94
Tableau 5-15: Matrice de jugement du niveau 7.....	96
Tableau 5-16: Résultats des modèles	98
Tableau 5-17: Calcul d'économie annuelle d'énergie de Boumerdes.....	99
Tableau 5-18: Calcul d'économie annuelle d'énergie de Tipaza	99

Liste des figures

Figure 1.1: Schéma représentant les étapes du traitement d'eau usée	17
Figure 1.2: Schéma représentant le prétraitement de l'eau usée.	18
Figure 1.3: Schéma qui illustre l'étape du traitement biologique secondaire	19
Figure 2.1: Principe du regroupement par K-Means.....	26
Figure 2.2: Structure hiérarchique d'un problème selon la méthode AHP.....	28
Figure 3.1: Réseau d'une seule couche qui contient R éléments d'entrée et S neurones	35
Figure 3.2: Réseau de neurones multicouches..	35
Figure 3.3: Connexion entre deux neurones.....	39
Figure 3.4: Schéma d'un signal d'entrée et fonction d'activation	39
Figure 4.1: Vue d'ensemble des ouvrages de la STEP de Boumerdes par Google Earth.....	44
Figure 4.2 : Organigramme d'exploitation de la STEP de Boumerdes.....	45
Figure 4.3 : Bassin d'aération..	46
Figure 4.4 : Décanteur-Clarificateur..	46
Figure 4.5: Evolution journalière du PH de l'eau brute	47
Figure 4.6: Evolution journalière du PH de l'eau épurée.....	47
Figure 4.7: Variation journalière de la température de l'eau brute	48
Figure 4.8: Variation journalière de la température de l'eau épurée.....	48
Figure 4.9: Evolution journalière de la concentration en MES à l'entrée de la STEP.....	48
Figure 4.11 : Evolution de la Demande Biologique en Oxygène à l'entrée de STEP.....	49
Figure 4.12 : Evolution de la Demande Biologique en Oxygène à la sortie de STEP.....	49
Figure 4.13 : Evolution de la Demande Chimique en Oxygène à l'entrée de la STEP.....	49
Figure 4.14 : Evolution de la Demande Chimique en Oxygène à la sortie de la STEP.....	49
Figure 4-15 : Evolution journalière de l'ammonium (NH ₄ ⁺) à l'entrée de la STEP.....	50
Figure 4-16 : Evolution journalière de l'ammonium (NH ₄ ⁺) à la sortie de la STEP.....	50
Figure 4.17 : Evolution journalière du Nitrite (NO ₂ ⁻) à l'entrée de la STEP	50
Figure 4.18 : Evolution journalière du Nitrite (NO ₂ ⁻) à la sortie de la STEP	50
Figure 4.19: Evolution du Nitrate(NO ₃ ⁻) à l'entrée de la STEP.....	51
Figure 4.20: Evolution du Nitrate (NO ₃ ⁻) à la sortie de la STEP	51
Figure 4.21 : Evolution du NTK à l'entrée de la STEP	51
Figure 4.22 : Evolution du NTK à la sortie de la STEP.....	51
Figure 4.23: Evolution de l'ortho-phosphate (PO ₄ ⁻³) à l'entrée de la STEP.....	51
Figure 4.24: Evolution de l'ortho-phosphate (PO ₄ ⁻³) à la sortie de laSTEP.....	51
Figure 4.25: Evolution journalière du débit dans la STEP.....	52
Figure 4.26: Evolution journalière de l'énergie dans la STEP.....	52
Figure 4.27: Evolution du débit recirculé.....	53
Figure 4.28: Photo de la STEP par Google Earth	54
Figure 4.29: Organigramme d'exploitation de la STEP de Tipaza..	55
Figure 4.30: Evolution journalière du PH de l'eau brute	55
Figure 4.31: Evolution journalière du PH de l'eau épurée.....	55
Figure 4.32: Variation journalière de la température de l'eau brute	55
Figure 4.33: Variation journalière de la température de l'eau épurée.....	55
Figure 4.34: Evolution journalière de la concentration en MES à l'entrée de la STEP.....	55
Figure 4.35: Evolution journalière de la concentration en MES à la sortie de la STEP	55
Figure 4.36: Evolution de la Demande Biologique en Oxygène à la sortie de la STEP	58
Figure 4.37: Evolution de la Demande Biologique en Oxygène à l'entrée de la STEP.....	58
Figure 4.38: Evolution de la Demande Chimique en Oxygène à la sortie de la STEP	58

Figure 4.39: Evolution de la Demande Chimique en Oxygène à l'entrée de la STEP.....	58
Figure 4.40: Evolution du Nitrite (NO ₂ -) à la sortie de la STEP.....	59
Figure 4.41: Evolution du Nitrite (NO ₂ -) à l'entrée de la STEP.....	59
Figure 4.42: Evolution du Nitrate (NO ₃ -) à la sortie de la STEP.....	60
Figure 4.43: Evolution du Nitrate (NO ₃ -) à l'entrée de la STEP.....	60
Figure 4.45: Evolution de PO ₄ -3 à l'entrée de la STEP.....	60
Figure 4.46: Evolution de PO ₄ -3 à la sortie de la STEP.....	60
Figure 4.47: Evolution du débit d'entré journalier de la STEP.....	61
Figure 4.48: Evolution de l'énergie totale consommée à la STEP.....	61
Figure 4.49: Evolution des Boues de recirculation à la STEP.....	62
Figure 5.1: Paramètres de pollution de l'eau brute à l'échelle journalière [1×2].....	65
Figure 5.2 :Paramètres de pollution de l'eau brute à l'échelle journalière [1×2].....	66
Figure 5.3: Paramètres de pollution de l'eau brute à l'échelle journalière [1×2].....	67
Figure 5.4: Variation de la variance intra-classe en fonction du nombre de classe.....	68
Figure 5.5: Regroupement par K-means de MESBr.....	69
Figure 5.6: Regroupement par K-means de DBO5Br.....	70
Figure 5.7: Regroupement par K-means de DCOBr.....	70
Figure 5.8: Regroupement par K-means de NH ₄ Br.....	71
Figure 5.9: Regroupement par K-means de Débit.....	71
Figure 5.10: Regroupement par K-means de l'énergie.....	72
Figure 5.11: Regroupement par K-means sur les données de Tipaza.....	79
Figure 5.12: Rendement Globale en fonction de l'indice de pollution.....	80
Figure 5.13 : Validation.....	81
Figure 5.14 : Calibration.....	81
Figure 5.15: Comparaison de l'énergie réelle et simulée de la STEP de Tipaza.....	81
Figure 5.16: Regroupement par K-Means pour les données de Boumerdes.....	82
Figure 5.17 : Rendement Globale en fonction de l'indice de pollution.....	83
Figure 5.18 : Calibration.....	82
Figure 5.19 : Validation.....	84
Figure 5.20: Comparaison de l'énergie réelle et simuléede la STEP de Boumerdes.....	84
Figure 5.21: Regroupement par K-Means pour les données cumulées.....	85
Figure 5.22: Rendement Globale en fonction de l'indice de pollution.....	85
Figure 5.23 : Calibration.....	85
Figure 5.24 : Validation.....	85
Figure 5.25: Comparaison de l'énergie réelle et simulée des deux STEP.....	87
Figure 5.26: regroupement par K-Means.....	90
Figure 5.27: Comparaison entre énergie simulée et mesuré pour saaty=1.....	91
Figure 5.28: regroupement par K-Means.....	92
Figure 5.29: Validation.....	93
Figure 5.30: Calibration.....	93
Figure 5.31: Comparaison entre énergie simulée et mesuré pour saaty = 3.....	93
Figure 5.32: regroupement par K-Means.....	94
Figure 5.33: Validation.....	95
Figure 5.34: Calibration.....	95
Figure 5.37: Comparaison entre énergie simulée et mesuré pour saaty=5.....	95
Figure 5.38: regroupement par K-Means.....	96
Figure 5.39: Calibration.....	97
Figure 5.40: Validation.....	97
Figure 5.41: Comparaison entre énergie simulée et mesuré pour saty = 5.....	97

Figure 5.42: Simulation du gain d'énergie	98
Figure 5.43: Histogramme du gain d'énergie annuelle de Boumerdes.....	99
Figure 5.44 :Histogramme du gain d'énergie annuelle de Tipaza.....	99

Introduction générale

L'eau constitue un élément essentiel dans la vie et l'activité humaine. C'est une composante majeure des mondes minéral et organique. Dans le monde présent, l'eau participe à toutes les activités quotidiennes notamment, domestiques, industrielles et agricoles ce qui la rend un élément récepteur exposé à tous les genres de pollution. Le phénomène de la pollution contribue de façon considérable à la limitation des ressources en eau potable.

L'eau est menacée dans sa qualité et dans sa quantité. La gestion optimale d'une station d'épuration et la maîtrise des procédés et étapes de traitement reste un but et un objectif à atteindre afin de protéger au mieux l'environnement et la santé publique.

L'épuration s'applique principalement aux eaux usées, pour les eaux pluviales on parle plutôt de traitement du rejet urbain en temps de pluie. L'épuration peut être assurée à l'échelle individuelle (assainissement autonome ou individuel), semi-collective ou plus généralement à l'échelle collective (à l'aval d'un réseau d'assainissement dans une station d'épuration).

La problématique observée dans les stations d'épuration est la forte consommation énergétique dans le processus d'épuration qui est due. Pour cela, une optimisation énergétique est recommandée, et cela tout en préservant l'environnement.

L'optimisation énergétique est un enjeu de premier ordre pour gérer les stations d'épuration, mais pour être efficace, cette optimisation doit s'insérer dans une planification systémique et à long terme.

L'objectif de ce travail est de mettre en place un modèle d'optimisation de la consommation en énergie électrique des deux STEP de Boumerdes et de Tipaza dans le but d'aider les gérants de celles-ci dans leurs prises de décisions optimales du temps d'aération du bassin biologique. Pour cela une approche de cumul des données des stations a été adoptée puis combinée au Deep learning afin d'obtenir le meilleur modèle possible.

Le Deep Learning est un nouveau domaine de recherche sur l'apprentissage automatique, qui a été introduit dans le but de rapprocher l'apprentissage automatique de l'un de ses objectifs initiaux : l'intelligence artificielle. Ce domaine a révolutionné l'apprentissage automatique dans les dernières années.

Ce mémoire est scindé en 5 chapitres :

Le premier chapitre s'intéresse à la synthèse des différents agents et paramètres qui entrent en jeu dans l'analyse de l'eau avant et après traitement et à la présentation des différentes filières d'épuration des eaux usées.

La théorie de l'analyse en Composantes Principales (ACP) et des K-Means, ainsi que la méthode de Hiérarchie analytique (AHP) fera l'objet de notre deuxième chapitre.

Le troisième chapitre sera consacré à la présentation du deep learning en général et des réseaux de neurones en particulier.

Le quatrième chapitre a été consacré à la présentation des STEP de Boumerdes et Tipaza, aux critiques et aux interprétations des données recueillies.

L'élaboration et l'application de notre modèle a été expliquée dans le cinquième chapitre, les différents résultats présentés et détaillés.

Enfin, une conclusion générale, dans laquelle on a fait valoir notre étude de recherche, tout en précisant la méthode suivie et l'objectif principal de cette étude.

Chapitre 1 : Généralités sur le fonctionnement des stations d'épuration

1. Généralités sur le fonctionnement des stations d'épuration

1.1 Introduction

L'eau est l'une des ressources fondamentales de notre planète et nous devons absolument faire tout notre possible pour la protéger.

Les eaux résiduaires englobent les eaux usées des ménages ainsi que l'eau utilisée dans les procédés industriels. Lorsqu'elles ne sont ni traitées ni purifiées, elles polluent les mers et les rivières avec les conséquences négatives inévitables pour l'environnement.

Afin de remédier à ce problème de pollution et ainsi contribuer à la protection de notre environnement, on procède au traitement de ces eaux par l'intermédiaire d'ouvrages spécifiques appelés Station d'épuration (STEP).

Le processus d'épuration des eaux usées se fait en trois étapes principales : primaire, secondaire et tertiaire. Au cours du traitement, secondaire, la consommation énergétique est très importante et requiert une analyse approfondie dans le but de notre étude.

Dans ce chapitre, on va définir en premier les différents types d'eaux usées ainsi que leurs caractéristiques, puis on présentera les STEP ainsi que leur fonctionnement.

1.2 Définition des eaux usées

Ramade (2000) définit les eaux usées comme étant des eaux ayant été utilisées pour des usages domestiques, industriels ou même agricole, constituant donc un effluent pollué et qui sont rejetées dans un émissaire d'égout. (Ramade, 2000).

Les eaux usées regroupent les eaux usées domestiques (les eaux vannes et les eaux ménagères), les eaux de ruissellement et les effluents industriels (eaux usées des usines). (Baumont, (2004)).

1.3 Les différents types d'eaux usées

Les eaux usées se distinguent en quatre catégories : les eaux domestiques, les eaux industrielles, les eaux agricoles et les eaux pluviales.

1.3.1 Les eaux usées domestiques

Elles proviennent des habitations, elles sont généralement véhiculées par le réseau d'assainissement jusqu'à la station d'épuration.

Ces eaux se caractérisent par leurs fortes teneurs en matière organique, en sels minéraux (azotes, phosphore), en détergent et en germes fécaux.

Les principales sources de ces eaux sont les eaux de cuisine, les eaux de buanderie, Eaux de vannes.(www.oieau.fr/eaudoc/integral/reuinter.htm, 2017)

1.3.2 Les eaux usées industrielles

Tous les rejets résultant d'une utilisation de l'eau autre que domestique sont qualifiés de rejets industriels. Cette définition concerne les rejets des usines, mais aussi les rejets d'activités artisanales ou commerciales : blanchisserie, restaurant, laboratoire d'analyses médicales, etc.(Edline, 1979).Les eaux industrielles peuvent contenir des produits toxiques, des métaux lourds, des polluants organiques et des hydrocarbures.

1.3.1 Les eaux usées agricoles

Ce sont des eaux qui proviennent de l'agriculture et contiennent des fertilisants, pesticides et autres produits chimiques. Elles se caractérisent par des fortes concentrations en composés azotés et phosphorés.(W.W.ECKENFELDER, 1982)

1.3.2 Les eaux de ruissellement

Ce sont essentiellement les eaux de pluie et de lavage des chaussées. Les eaux de pluie sont caractérisées par un débit fortement variable, présentant des valeurs moyennes à fortes variations saisonnières à l'intérieur desquelles la répartition des débits est aléatoire.

-La pollution entraînée est maximale en début de précipitation. Elle correspond au lavage des toits et chaussées. Elle décroît ensuite fortement en cas de débits persistants.

-Les polluants sont en majorité des matières en suspension d'origine minérale. On y trouve aussi des polluants de l'atmosphère (poussières, oxyde d'azote NO_x, oxyde de soufre SO_x, du plomb....).(www.oieau.fr/eaudoc/integral/reuinter.htm, 2017)

1.4 Les différentes caractéristiques des eaux usées

1.4.1 Les paramètres physico-chimiques

L'estimation de la qualité physico-chimique d'une eau ne peut s'effectuer par la mesure d'un seul, mais d'un ensemble de paramètres de natures diverses. Des résultats anormaux dans le contrôle de paramètres permettent de déceler et d'évaluer les niveaux de pollution. La pollution de l'eau est fonction des substances dissoutes susceptibles d'être nuisibles qu'elle contient et, dont la plupart, ne sont décelables qu'à l'analyse.

1.4.1.1 La température

La température de l'eau joue un rôle important par exemple en ce qui concerne la solubilité des sels et des gaz, c'est un paramètre qui a une influence sur :

L'activité de la biomasse nitrifiante. (Gaëlle Deronzier)

Les propriétés de décantation : quand la température augmente, l'indice de décantation augmente car la densité des floccs diminue. (Gaëlle Deronzier)

La toxicité : l'effet toxique des substances chimiques est plus grand à température élevée. (Moletta)

La demande biologique en oxygène : une augmentation de la température conduit à une augmentation de la consommation de l'O₂ dissous. (Moletta).

1.4.1.2 Le Potentiel Hydrogène (pH)

Le pH est une mesure de l'acidité de l'eau c'est à dire de la concentration en ions d'hydrogène (H⁺). L'échelle des pH s'étend en pratique de 0 (très acide) à 14 (très alcalin) ; la valeur médiane 7 correspond à une solution neutre à 25°C. Le pH d'une eau naturelle peut varier de 4 à 10 en fonction de la nature acide ou basique des terrains traversés. Des pH faibles (eaux acides) augmentent notamment le risque de présence de métaux sous une forme ionique plus toxique. Des pH élevés augmentent les concentrations d'ammoniac, toxique pour les poissons. On admet généralement qu'un pH naturel situé entre 6,5 et 8,5 caractérise des eaux où la vie se développe de manière optimale

1.4.1.3 L'oxygène dissous (OD)

Les concentrations en oxygène dissous constituent l'un des plus importants paramètres de qualité des eaux pour la vie aquatique. L'oxygène dissous dans les eaux de surface provient essentiellement de l'atmosphère et de l'activité photosynthétique des algues et des plantes aquatiques. Cette concentration en oxygène dissous est également fonction de la vitesse d'appauvrissement du milieu en oxygène par l'activité des organismes aquatiques et les processus d'oxydation et de décomposition de la matière organique présente dans l'eau.

1.4.1.4 Matières en suspension (MES)

Les matières en suspension sont des matières non solubilisées, de nature organique et minérale, les premières sont généralement volatiles, elles peuvent être traitées par dégradation biologique ; les dernières de natures minérales essentiellement inertes, peuvent être éliminées en grande partie par sédimentation. (Gaid, 2005).

1.4.1.5 La conductivité électrique (CE)

La conductivité est une mesure de la capacité d'une solution à laisser passer un courant électrique. Elle fournit une indication précise sur la teneur en sels dissous (salinité de l'eau). La conductivité s'exprime en micro Siemens par centimètre. La mesure de la conductivité permet d'évaluer la minéralisation globale de l'eau. (REJSEK, 2002).

1.4.2 Les paramètres chimiques

1.4.2.1 Demande chimique en oxygène (DCO)

La DCO est la quantité d'oxygène nécessaire pour oxyder les matières organiques y compris les matières biodégradables et non biodégradables par voie chimique. Vu la simplicité de mesure de DCO et sa précision, il s'est avéré nécessaire de développer des corrélations entre la DBO5 et la DCO ainsi le rapport DCO/ DBO5 des eaux usées urbaines est proche de 2 mg/l (GUERREE, 1978), le rapport DCO/ DBO5 des effluents domestiques est de 1,9 à 2,5 (Hamdani.A, 2002).

1.4.2.2 Demande biochimique en oxygène (DBO5)

Sa détermination consiste à mesurer la quantité totale de l'oxygène consommée par des processus biochimiques, au cours de l'oxydation des matières organiques dans un échantillon donné dans les conditions de l'essai (incubation à 20°C et à l'obscurité).

La DBO a été standardisée en DBO5, mesurée au bout de 5 jours, considérée comme une période significative du processus global de biodégradation qui prend des semaines. Ce paramètre constitue un bon indicateur de la teneur en matières organiques biodégradables d'une eau, et il donne une indication indirecte de l'activité bactérienne. (BELAHMADI Mohamed Seddik Oussama, 2004).

1.4.2.3 Phosphore Total (PT)

Il y'a deux formes de phosphore qui contribuent à la pollution des eaux, le premier est le phosphore organique venant des matières vivantes et le second est le phosphore minéral qui représente 50% à 90% de la totalité du phosphore présent dans les eaux urbaines.

PT = Porganique + Pminéral Le rejet de matières phosphorées est d'environ 4 grammes par habitant et par jour. La quantité de PT s'exprime en mg/l (milligrammes par litre).

(Commission de protection des eaux, du patrimoine, de l'environnement, du sous-sol et des).

1.4.2.4 Azote Kjeldahl (NTK)

L'azote total Kjeldahl (NTK) est une appellation qui désigne la somme de l'azote ammoniacal et de l'azote organique. L'azote qui se retrouve sous forme oxydée, tel que les nitrites ou les nitrates par exemple, n'est pas mesuré par cette technique. Les composés azotés mesurés par cette méthode proviennent principalement de la dégradation bactérienne des composés organiques provenant de l'azote. L'industrie alimentaire, certaines industries de traitement des viandes non comestibles, les procédés de nettoyages industriels et l'épandage d'engrais sont aussi des sources importantes d'azote dans l'environnement.

1.5 Le fonctionnement d'une station d'épuration

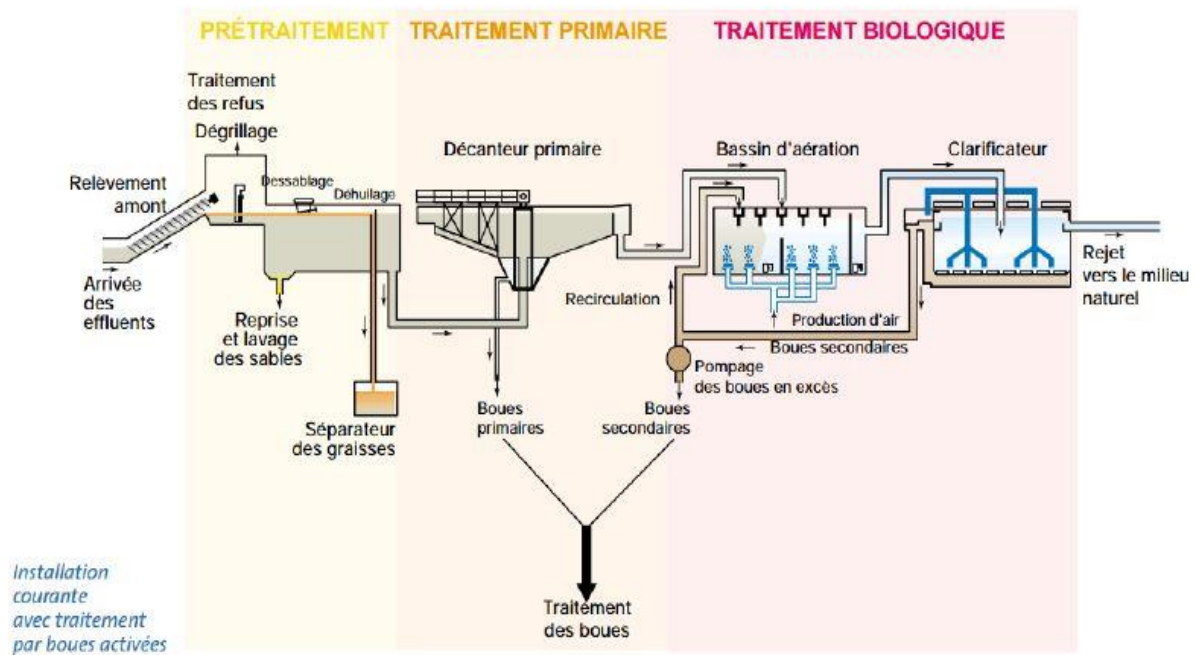


Figure 1.1: Schéma représentant les étapes du traitement d'eau usée. (ABDELAZIZ, 2016)

1.5.1 Définition

Une station d'épuration est une usine de dépollution des eaux usées avant leur rejet au milieu naturel, en général dans une rivière ou cours d'eau. Elle est installée généralement à l'extrémité d'un réseau de collecte, sur l'émissaire principal, juste en amont de la sortie des eaux vers le milieu naturel.

Elle ressemble une succession de dispositifs, empruntés tour à tour par les eaux usées. Chacun de ces dispositifs est conçu pour extraire un ou plusieurs polluants contenus dans ces eaux usées.

La capacité de la station étant limitée, des ouvrages de dérivation vers le milieu naturel « déversoir d'orage » sont répartis sur le réseau afin de réduire le débit arrivant à la station. En conséquence, en cas de fortes pluies, une partie de l'eau véhiculée par le réseau rejoint directement la rivière. (Matthieu., 2007/2008 .).

1.5.2 Les étapes d'épuration

Dans le but d'épurer les eaux usées plusieurs procédés et dispositifs sont utilisés :

Une station de relevage est nécessaire quand la pente du terrain ne permet pas un écoulement de l'eau par la gravité.

Une fois que les eaux usées sont collectées, elles passent par les cinq grandes étapes du traitement qui sont :

1.5.2.1 Prétraitement

C'est un processus physique et mécanique dont le but est d'extraire de l'eau des éléments dont la taille, le pouvoir abrasif et la masse pourraient endommager le matériel ou perturber la suite du protocole d'épuration. Il passe par trois étapes :

1.5.2.1.1 Dégrillage

A l'arrivée des eaux usées, elles passent dans un tamis où les déchets volumineux sont retenus (papiers, feuilles, matières plastiques...) et l'eau est pompée vers la station d'épuration.

1.5.2.1.2 Dessablage

La présence des sables et des graviers dans les eaux à traiter peut endommager les installations, alors des bassins sont conçus spécialement pour qu'ils se déposent au fond ensuite ils sont récupérés par le raclage vers une fosse de collecte.

1.5.2.1.3 Dégraissage-Déshuilage

Afin de remonter les huiles et les graisses en surface, on injecte de fines bulles d'air ensuite on les racle plus facilement.

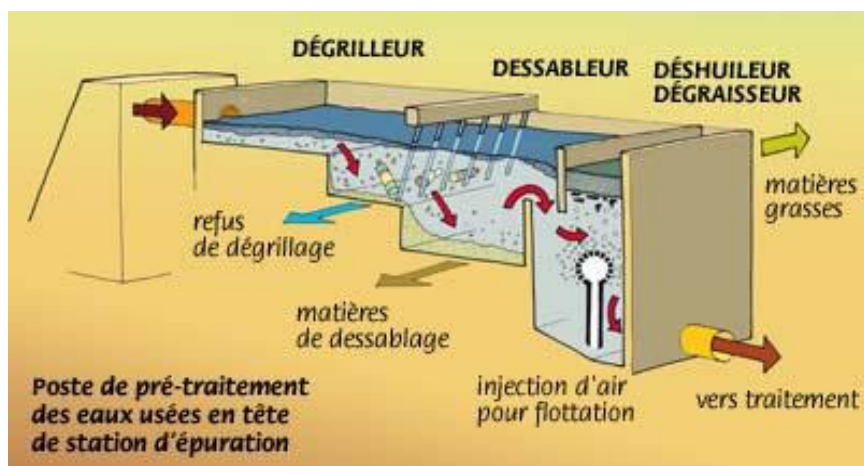


Figure 1.2: Schéma représentant le prétraitement de l'eau usée. (memoireonline, 2017)

Après le prétraitement, 3 étapes de traitement sont fondamentales qui sont le traitement primaire, secondaire et tertiaire. On mettra l'accent sur le traitement secondaire étant donné que notre problématique de consommation d'énergie est située à son niveau.

1.5.2.2 Traitement primaire

Le traitement primaire élimine plus de la moitié des matières en suspension (MES) et constitue une pré-épuration non négligeable quoique insuffisante pour garantir la qualité du rejet en milieu naturel. (Y.libes)

Le processus principal du traitement primaire est la décantation, il existe deux types

De décantation : (Matthieu., 2007/2008 .)

- Décantation physique.
- Décantation avec réactifs chimiques.

1.5.2.3 Traitement secondaire

Généralement le taux d'épuration exigé pour le rejet des effluents dans le milieu naturel n'est pas atteint par une simple épuration préliminaire et primaire.

Les traitements secondaires, conçus à l'origine essentiellement pour l'élimination de la pollution carbonée et des matières en suspension, ont pour objet de poursuivre l'épuration de l'effluent provenant du décanteur primaire ou issu d'un prétraitement.

On distingue deux classes de procédés du traitement secondaire en fonction de la qualité de l'effluent :

- **Les traitements physico-chimiques**

Consistent à transformer chimiquement, à l'aide de réactifs, les éléments polluants non touchés par les traitements biologiques (matières non biodégradables).

- **Les traitements biologiques**

Ils sont appliqués aux matières organiques en utilisant des cultures de microorganismes (notamment bactéries) reproduisent le processus de l'autoépuration naturelle dans des bassins adaptés à ce propos (les bioréacteurs). Plus récemment, le traitement biologique de l'azote a été intégré à cette étape, et de la même manière, le traitement des phosphates commence aussi à y être intégré.

Les impuretés sont alors digérées par des êtres vivants microscopiques et transformées en boues. La culture des bactéries se fait soit en milieu aéré (aérobie), soit en absence d'oxygène (anoxie). On distingue aussi les cultures fixées (lits bactériens, disques biologiques), et les cultures libres (lagunage aéré, boues activées). (Victor-Hugo, 2007).

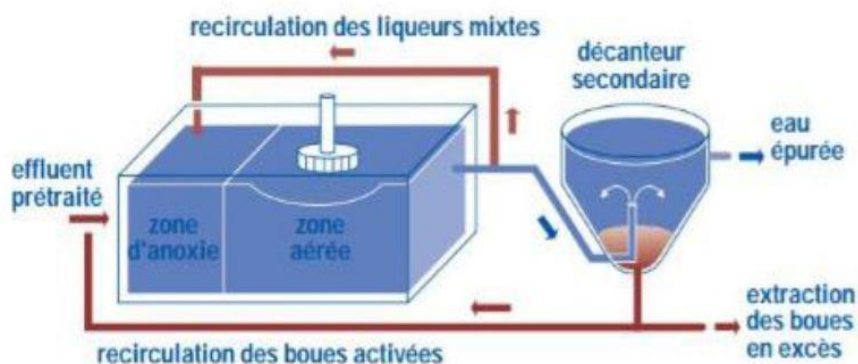


Figure 1.3: Schéma qui illustre l'étape du traitement biologique secondaire. (memoireonline, 2017)

1.5.2.4 Traitement tertiaire

Si le milieu récepteur exige une épuration meilleure, il est nécessaire d'effectuer des traitements complémentaires tels que : la désinfection par le chlore ou l'élimination de l'azote et du phosphore.

Les usines doivent avoir des stations d'épurations locales, car les déchets retenus lors du dégrillage, déshuilage... ainsi que les boues doivent subir des traitements spécifiques avant d'être rejetés dans le milieu naturel.

Conclusion

Dans ce premier chapitre, une idée globale du fonctionnement des stations d'épuration a été donnée puis l'accent a été mis sur les différents paramètres importants pour la modélisation et l'optimisation de la consommation énergétique des STEP. Une critique des données sera ensuite menée pour la fiabiliser les données recueillies.

CHAPITRE 2 : METHODES STATISTIQUES D'ANALYSE DES DONNEES

2. METHODES STATISTIQUE D'ANALYSE DES DONNEES

2.1 INTRODUCTION

Le Clustering est défini comme étant la tâche qui consiste à regrouper, de façon non supervisée, un ensemble d'objets ou plus largement de données, de telle manière que les objets d'un même groupe (appelé cluster) sont plus proches (au sens d'un critère de similarité choisie) les uns aux autres que celles des autres groupes(Clusters). Il s'agit d'une tâche principale dans la fouille exploratoire de données, et une technique d'analyse statistique des données très utilisée dans de nombreux domaines. L'idée est donc de découvrir des groupes au sein des données, de façon automatique (CHAMROUKHI, 2013). Dans ce cadre plusieurs méthodes ont été développées, la plus populaire est celle Des k moyennes (K-means), elle doit sa popularité à sa simplicité et sa capacité de traiter de larges ensembles de données.(Kogan, 2007).

L'analyse des données est une famille de méthodes statistiques dont les principales caractéristiques sont d'être multidimensionnelles et descriptives, ces méthodes ont commencé à être développées dans les années 50 poussées par le développement de l'informatique et du stockage des données qui depuis n'a cessé de croître.

L'analyse multicritère AHP est un processus de hiérarchie analytique qui a été développé afin d'aider les décideurs à faire un choix judicieux au niveau d'un problème faisant intervenir plusieurs critères.

Dans ce chapitre, on va présenter quelques-unes des principales méthodes de l'analyse de données : l'Analyse en Composantes Principales, théorie des K-means ainsi que le processus d'analyse hiérarchique, leurs notions de base ainsi que leurs objectifs.

2.2 Analyse en Composantes Principales (ACP)

2.2.1 DESCRIPTION DE LA METHODE

L'Analyse en Composantes principales (ACP) fait partie du groupe des méthodes descriptives multidimensionnelles appelées méthodes factorielles.

L'ACP propose, à partir d'un tableau rectangulaire de données comportant les valeurs de p variables quantitatives pour n unités (appelées aussi individus), des représentations géométriques de ces unités et de ces variables. Ces données peuvent être issues d'une procédure d'échantillonnage ou bien de l'observation d'une population toute entière.

L'ACP servira à mieux connaître les données sur lesquelles on travaille, à détecter éventuellement des valeurs suspectes, et aidera à formuler des hypothèses qu'il faudra étudier à l'aide de modèles et d'études statistiques inférentielles. (C. Duby, juillet 2006).

2.2.1.1 DEFINITIONS

2.2.1.1.1 Notions d'individu et de caractère

- **Individu**

L'individu peut désigner selon les cas : une année d'observations ou une autre unité de temps. L'ensemble des individus peut provenir d'un échantillonnage dans une population ou, il peut être de la population toute entière.

L'individu "i" est décrit par le vecteur appartenant à R^P :

$$X_i = \{X_{ij} / j = 1 \text{ à } P\}$$

Le terme X_{ij} est un nombre réel qui représente la mesure de la variable X_j sur l'individu i .

- **Caractère**

Sur un individu, on relève un certain nombre de caractères, dits aussi variables, désignant en général un paramètre intervenant dans un phénomène complexe à étudier.

Le caractère (ou variable) "j" est décrit par le vecteur de R^N :

$$X_j = \{X_{ij} / i = 1 \text{ à } N\}$$

Ainsi si l'ensemble des individus doit être homogène, l'ensemble des variables peut être hétérogène. (HAMRICHE, 1993).

2.2.1.1.2 Définition algébrique

L'analyse en composantes principales se fait sur une matrice de P variables et N individus. Cette dernière est dite matrice de données, qui résulte du croisement « $N \times P$ ».

La matrice de données est notée par $[X]$ et elle est décrite comme suit :

$$[X] = (a_{i,j})_{1 \leq i \leq n, 1 \leq j \leq p}$$

A partir de cette matrice des données brutes, on procède au calcul des paramètres statistiques :

1. La moyenne

$$\overline{X_j} = \frac{1}{N} \sum_{i=1}^N X_{ij}$$

$\overline{X_j}$: Moyenne de la j ème variable X

2. Ecart type

$$S_j = \left[\frac{1}{N} \sum_{i=1}^N (X_{ij} - \overline{X_j})^2 \right]^{1/2}$$

S_j = Ecart type de la j ème variable.

N : Nombre d'individus.

3. Le coefficient de covariance

$$Cov(X_j, X_k) = \frac{1}{N} \sum_{i=1}^N (X_{ij} - \overline{X_j}) * (X_{ik} - \overline{X_{ik}})$$

4. Le coefficient de corrélation

$$Cor(X_j, X_k) = \frac{Cov(X_j, X_k)}{S_j * S_k} = \frac{\sum_{i=1}^N (X_{ij} - \bar{X}_j) * (X_{ik} - \bar{X}_k)}{\left[\sum_{i=1}^N (X_{ij} - \bar{X}_j)^2 * \sum_{i=1}^N (X_{ik} - \bar{X}_k)^2 \right]^{1/2}}$$

Une fois ces paramètres statistiques calculés, on remplace la matrice de données [X] par une nouvelle matrice définie par le croisement (individus x CP).

Avec : CP, composantes principales

Calculer les composantes principales notées Cj revient à déterminer P relations linéaires entre les variables Xj :

$$C_j = a_{j0} + a_{j,1} X_1 + \dots + a_{jN} X_N \quad 1 \leq j \leq N$$

Avec : Cj : j^{ème} Composante Principale.

Xj : Vecteur variable initiale.

a_{jk} : Coefficient du système.

Notons au passage que les termes a_{j0} désignent le vecteur permettant la translation de l'origine de l'ancien repère vers le centre de gravité du nuage de points. Un centrage des données initiales annule les coefficients a_{j0}. (HAMRICHE, 1993).

2.2.1.1.3 Définition géométrique

L'Analyse en Composantes Principales est puissante par son support géométrique :

La méthode consiste à rechercher un premier axe qui soit le plus près possible de tous les points au sens des moindres carrés : tel que la somme des moindres carrés des distances des N points à cet axe soit minimale, ou encore la projection de ces derniers sur cet axe ait une dispersion maximale. Cet axe est appelé « axe factoriel ».

Un second axe est obtenu après projection des N points sur un hyperplan orthogonal au premier axe, tel que la dispersion des projections des N points sur celui-ci soit toujours maximale, et le processus se réitère P fois. (HAMRICHE, 1993).

Ainsi un nouveau système d'axe est obtenu. Ce dernier est défini par des nouvelles variables appelées composantes principales.

2.2.2 Objectifs

L'Analyse en Composantes Principales a pour objectifs :

- Avoir une corrélation entre les variables grâce à une représentation des données brutes et fournir des outils simples et lisibles.
- Essayer d'examiner les liens existants entre les variables afin de faciliter leur interprétation.
- Essayer de réduire au maximum le nombre de dimensions étudiées tout en gardant l'ensemble original et en utilisant les relations détectées entre les variables.

2.2.3 Application de l'ACP

Il existe plusieurs applications pour l'Analyse en Composantes Principales, parmi lesquelles :

- L'étude et la visualisation des corrélations entre les variables, afin éventuellement de limiter le nombre de variables à mesurer par la suite.

- L'obtention de facteurs non corrélés qui sont des combinaisons linéaires des variables de départ, afin d'utiliser ces facteurs dans des méthodes de modélisation telles que la régression linéaire, et l'analyse discriminante.
- La visualisation des observations dans un espace à deux ou trois dimensions, afin d'identifier des groupes homogènes d'observations, ou au contraire des observations atypiques.

2.2.4 Procédé d'application de l'ACP

Afin d'appliquer une Analyse en Composantes Principales, il faut procéder comme suit :

1. Formation de la matrice des données brutes [X] (composée par le croisement de N individus x P variables).
2. Calcul des paramètres statistiques :(moyenne, écart type).
3. Calcul des variables centrées réduites
4. Calcul des coefficients de corrélation
5. Calcul des valeurs propres et des vecteurs propres de la matrice de corrélation obtenue.
6. Calcul des composantes principales, avec :
CP = matrice des variables centrées réduites (NxP) x matrice des vecteurs propres (NxN).
7. Calcul des composantes principales centrées réduites :

$$CP'(i,j) = \frac{CP(i,j)}{\sqrt{VP(j)}}$$

Avec VP (j) : la valeur propre correspondant à l'individu j.

8. Calcul de la matrice de corrélation entre CP 'et XCR (cor (CP', XCR)).
9. Tracé des cercles de corrélation

2.2.5 Résultats d'application

L'Analyse en composantes principales met à notre disposition trois résultats nécessaires à notre interprétation, et qui sont :

- Un tableau de vecteurs et valeurs propres.
- Un tableau de corrélation des individus avec les axes factoriels.
- Un tableau de corrélation des variables avec les axes principaux.

2.2.6 Analyse des résultats

Les résultats obtenus par l'ACP permettent de répondre aux questions suivantes :

- **Les données sont-elles en factorisation ?**

Dans un premier temps, il convient d'observer la matrice de corrélation. Si plusieurs variables sont corrélées (> 0.5), la factorisation est possible. Sinon, la factorisation n'a pas de sens et n'est donc pas conseillée.

Pour tester la corrélation parfaite, il faut calculer le déterminant de la matrice. Notons que nous recherchons un déterminant petit, ce qui constitue un bon indice de l'existence de patrons de corrélations entre les variables, mais non égal à zéro.

Dans un deuxième temps, il faut observer l'indice de KMO (Kaiser-Meyer-Olkin) qui doit tendre vers 1. Il indique jusqu'à quel point l'ensemble de variables retenu est un ensemble cohérent et permet de constituer une ou des mesures adéquates de concepts.

Pour juger de l'indice de KMO, on peut utiliser l'échelle suivante :

Inacceptable	<0.5
Médiocre	0.5 < KMO < 0.6
Moyen	0.6 < KMO < 0.7
Bien	0.7 < KMO < 0.8
Très Bien	0.8 < KMO < 0.9
Excellent	>0.9

Enfin, on utilise le test de sphéricité de Bartlett, dont l'échelle qui peut être utilisée est :

Très significatif	→0
Significatif	< 0.05
Acceptable	0.05 < I < 0.1
Rejeté	>0.1

- **Combien de facteurs retenir ?**

Trois règles sont applicables :

1ère règle : La règle de Kaiser qui veut qu'on ne retienne que les facteurs aux valeurs propres supérieures à 1.

2ème règle : On choisit le nombre d'axes en fonction de la restitution minimale d'informations que l'on souhaite.

3ème règle : Le « Scree-test » ou test du coude. On observe le graphique des valeurs propres et on ne retient que les valeurs qui se trouvent à gauche du point d'inflexion.

2.3 Méthode des K-means

2.3.1 DEFINITION

L'algorithme K-means mis au point par McQueen en 1967, est l'un des plus simples Algorithmes d'apprentissage non supervisé, appelée algorithme des centres mobiles.

Il attribue chaque point dans un cluster dont le centre (centroïde) est le plus proche. Le centre est la moyenne de tous les points dans le cluster, ses coordonnées sont la moyenne arithmétique pour chaque dimension séparément de tous les points dans le cluster c'est à dire chaque cluster est représentée par son centre de gravité. (Likas A., 2003).

Étant donné un ensemble de points (x_1, x_2, \dots, x_n) , on cherche à partitionner les n points en k ensembles $S = \{S_1, S_2, \dots, S_k\}$ ($k \leq n$) en minimisant la distance entre les points à l'intérieur de chaque partition :

$$\arg \min_S \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

Où μ_i est le barycentre des points dans S_i .

2.3.2 PRINCIPE DE LA METHODE

On suppose qu'il existe K classes distinctes. On commence par désigner K centres de classes μ_1, \dots, μ_k parmi les individus. Ces centres peuvent être soit choisis par l'utilisateur pour leur "représentativité", soit désignés aléatoirement. On réalise ensuite itérativement les deux étapes suivantes :

- Pour chaque individu qui n'est pas un centre de classe, on regarde quel est le centre de classe le plus proche. On définit ainsi K classes C_1, \dots, C_k , où $C_i = \{\text{ensemble des points les plus proches du centre } \mu_i\}$.
- Dans chaque nouvelle classe C_i , on définit le nouveau centre de classe μ_i comme étant le barycentre des points de C_i .

L'algorithme s'arrête suivant un critère d'arrêt fixé par l'utilisateur qui peut être choisi parmi les suivants : soit le nombre limite d'itérations est atteint, soit l'algorithme a convergé, c'est-à-dire qu'entre deux itérations les classes formées restent les mêmes, soit l'algorithme a "presque" convergé, c'est-à-dire que l'inertie intra-classe ne s'améliore quasiment plus entre deux itérations.

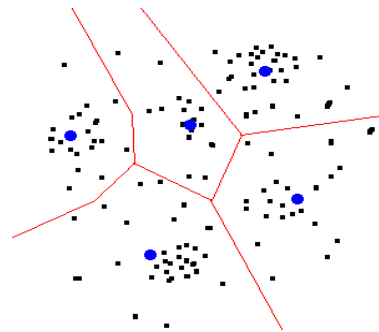


Figure 1.1: Principe du regroupement par K-Means (Belhabib Abdelkader, 2012)

2.3.3 Algorithme

Entrée

Ensemble de N données, noté par x

Nombre de groupes souhaité, noté par k

Sortie

Une partition de K groupes $\{C_1, C_2, \dots, C_k\}$

Début

1) Initialisation aléatoire des centres C_k ;

Répéter

2) Affectation : générer une nouvelle partition en assignant chaque objet au groupe dont le centre est le plus proche :

$$x_i \in C_k \text{ si } \forall_j |x_i - \mu_k| = \min |x_i - \mu_j|$$

Avec μ_k le centre de la classe K ;

3) Représentation : Calculer les centres associés à la nouvelle partition ;

$$\mu_k = \frac{1}{N} \sum_{x \in C_k} x_i$$

Jusqu'à convergence de l'algorithme vers une partition stable ;

Fin.

2.3.4 Avantages de K-means :

Nous pouvons citer quelques avantages de K-means :

- L'avantage de ces algorithmes est avant tout leur grande simplicité.
- Tend à réduire l'erreur quadratique.
- Applicable à des données de grandes tailles.

2.3.5 Inconvénients :

- Le nombre de classe doit être fixé au départ.
- Ne détecte pas les données bruitées.
- Le résultat dépend de tirage initial des centres des classes.
- Les clusters sont construits par rapports à des objets inexistantes (les milieux)
- N'est pas applicable en présence d'attributs qui ne sont pas du type intervalle

(S . P . Bradley, 1998).

2.4 Processus d'Analyse Hiérarchique (AHP) :

2.4.1 Définition

L'analyse multicritère AHP est un processus de hiérarchie analytique qui a été développé par Saaty (1982) afin d'aider les décideurs à faire un choix judicieux au niveau d'un problème faisant intervenir plusieurs critères. Bruce et al. (1989) la définisse comme suit :

Analyse (Analytics) : la méthode AHP est un outil de prise de décision. En utilisant les mathématiques il est plus facile de comprendre et de décrire les choix à faire. Toutes les méthodes de description des décisions sont analytiques. La méthode est facile à mettre en œuvre malgré la complexité de sa structure mathématique.

Hiérarchie (Hierarchy) : la méthode AHP organise le problème de la prise de décision selon des niveaux basés sur la compréhension de la situation par les décideurs. En résumant le problème sous forme de niveaux : objectif, critères, sous-critères, alternatives, le décideur peut analyser en profondeur des petits ensembles de décisions.

Procédé (Process) : pour chaque problème de prise de décision on a besoin d'un procédé d'apprentissage, de débat et de révision de priorités.(Durand., juillet 2005).

2.4.2 Principe

La méthode AHP est simple : elle est basée sur des comparaisons binaires entre les éléments de la hiérarchie, ensuite elle transpose les résultats dans des matrices de comparaison à partir desquelles on extrait des vecteurs de priorité. En dernière étape, elle calcule les scores des alternatives pour ces vecteurs de priorités qui font l'objet d'analyse.

La hiérarchie de la méthode AHP est souvent composée de trois niveaux : le premier niveau contient l'objectif, le deuxième niveau comprend les critères et les sous-critères et le troisième niveau contient les alternatives.

2.4.3 Procédé d'application de l'AHP

La procédure d'agrégation selon la méthode AHP comporte cinq étapes :

Etape 1 : Décomposer le problème complexe en une structure hiérarchique.

Etape 2 : Comparaison par paire des éléments. (Matrice de jugement)

Etape 3 : Déterminer les priorités. (Les poids)

Etape 4 : Évaluer la cohérence des jugements.

Etape 5 : Calculer le « poids » final.

2.4.3.1 Décomposer le problème complexe en une structure hiérarchique

On commence par décortiquer un système complexe en structure hiérarchique et ceci en déterminant les éléments du problème des plus généraux aux plus spécifiques, on procédera par la suite au classement de ces éléments par niveaux.

Dans le cas général la hiérarchisation comporte trois niveaux sous forme de pyramide, le plus haut représente l'objectif à atteindre, juste en-dessous on retrouve les critères et les sous-critères et c'est les éléments qui affectent la prise de décision et enfin comme troisième niveau et en bas de la pyramide on retrouve les alternatives (Bruce L.G, 1989).

En résumé cette première étape consiste à :

- Définir l'objectif cible (niveau 0).
- Définir les critères et les sous-critères de décision ou d'analyse (niveau 1).
- Définir les alternatives (Niveau 2).

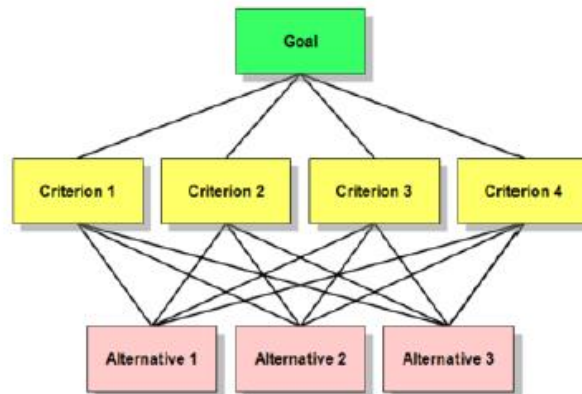


Figure 22.2: Structure hiérarchique d'un problème selon la méthode AHP. (Souar, 2017).

2.4.3.2 Comparaison par paires des éléments

Une fois la pyramide construite, on doit émettre des priorités pour chaque élément inscrit dans chaque niveau de la pyramide afin d'en déterminer l'importance de chaque élément.

Pour ce faire, l'évaluation de l'importance relative de chaque élément par rapport à un autre élément du même niveau se fait en comparant deux à deux ces derniers contribuant ainsi à la réalisation du niveau supérieur adjacent.

Il faut donc choisir une échelle de valeurs pour spécifier le degré d'importance d'un élément par rapport à un autre. L'échelle de valeur (1 à 9) est adoptée.(L.G.Vargas, 1991).

Cette échelle « échelle de Saaty » de valeur permet au décideur d'incorporer sa subjectivité, de contribuer par son expérience et ses connaissances lors de la détermination des importances relatives des éléments pour chaque niveau. Le choix de cette échelle repose sur notre capacité naturelle à établir des distinctions dans la force des relations qui existent entre les éléments.

Saaty propose une évaluation sur 5 niveaux selon le tableau suivant :

Tableau 22-1: Echelle de jugement (L.G.Vargas, 1991)

Description	Valeur numérique
Les deux éléments ont une importance égale	1
L'élément en ligne est un peu plus important que celui en colonne	3
L'élément en ligne est plus important que celui en colonne	5
L'élément en ligne est beaucoup plus important que celui en colonne	7
L'élément en ligne est absolument plus important que celui en colonne	9
L'élément en ligne est un peu moins important que celui en colonne	1/3
L'élément en ligne est moins important que celui en colonne	1/5
L'élément en ligne est beaucoup moins important que celui en colonne	1/7
L'élément en ligne est absolument moins important que celui en colonne	1/9
Dans certain cas, si nous avons des difficultés pour évaluer, on pourra affiner en rajoutant des valeurs intermédiaires.	2, 4, 6, 8

Ainsi, à partir de cette échelle de valeurs, nous complétons une matrice de jugement M, à chaque niveau. Le nombre de matrices dépend du nombre d'éléments présents à chaque niveau. Le rang de la matrice de chaque niveau dépend du nombre d'éléments présents au niveau inférieur.

Pour un niveau on note la matrice carrée de jugement $B = [b_{ji}] m \times m$ où m est le nombre de critères. b_{ji} : le coefficient d'importance relative du critère C_j par rapport au critère C_i .

Le tableau suivant présente la matrice carrée de jugement :

Tableau 2-2: La matrice de jugement relatif à un niveau. (L.G.Vargas, 1991).

Critères	C_1	C_2	C_3	...	C_m
C_1	1	b_{12}	b_{13}	...	b_{1m}
C_2	$1/b_{12}$	1	b_{23}	...	b_{2m}
C_3	$1/b_{13}$	$1/b_{23}$	1	...	b_{3m}
...	1	...
C_m	$1/b_{1m}$	$1/b_{2m}$	$1/b_{3m}$...	1

2.4.3.3 Déterminer les priorités

Une fois ces matrices comparatives créées, une mesure relative est dérivée de ces différents éléments. La mesure relative d'un élément à un niveau donné (critère fils), en lien avec l'élément du niveau adjacent plus élevé (critère père), constitue le poids du critère fils associé à ce niveau.

Pour définir ces poids, Saaty a proposé une méthode basée sur la détermination des valeurs propres λ de la matrice des jugements.

En effet, la détermination des poids à un niveau donné, caractérisé par une matrice des jugements M est effectuée comme suit :

- Résoudre l'équation $\det [M - \lambda I] = 0$ et déterminer les valeurs propres de la matrice M . Avec \det (déterminant).
 - Calculer le vecteur propre V relatif à la plus grande valeur propre : λ_{\max} .
- Ainsi, les valeurs des composants du vecteur V représentent les poids des éléments du niveau en question.

Saaty a aussi montré que la valeur propre maximale λ_{\max} du problème de recherche des valeurs propres était réelle et supérieure à la taille de la matrice et en ajoutant la condition que la somme des poids doit être égale à 1, il a démontré que les valeurs des poids peuvent être obtenues par une autre technique qui consiste à :

- Normaliser la matrice de jugement par colonne en divisant chaque élément d'une colonne par la somme des éléments de cette même colonne.
- Calculer la moyenne arithmétique des nombres sur chaque ligne de la matrice normalisée obtenue précédemment. Ainsi, la valeur de la moyenne obtenue pour une ligne correspond au poids de ce critère relativement à ce niveau.

2.4.3.4 Evaluer la cohérence des jugements

A ce stade, nous avons donc le « poids » de chacun des critères du niveau ciblé. La méthode AHP propose alors d'en valider la fiabilité du résultat en calculant un index de cohérence.

En effet, l'incohérence potentielle peut venir du fait qu'un critère peut ne pas être jugé de façon cohérente vis à vis d'un ou plusieurs autres. Cet indice va donc nous permettre de détecter des défauts importants dans notre calcul et notre évaluation.

L'indice de cohérence CI d'une matrice $n \times n$ est :

$$RC = \frac{IC}{RI}$$

Avec :

- **RI, Indice aléatoire élaboré par Saaty** : Il indique le niveau de fiabilité d'une même évaluation effectuée à plusieurs reprises.

Tableau 2-3: Indice aléatoire élaboré par Saaty. (T.L, 1984)

n	1	2	3	4	5	6	7	8	9	10
R.I	0	0	0,58	0,9	1,12	1,24	1,32	1,41	1,45	1,49

- **IC, Indice de Cohérence** : il représente le niveau de fiabilité de notre jugement.
- **RC, Ratio de Cohérence** : En divisant IC par RI , on compare donc la fiabilité réelle avec une fiabilité théorique.

Au plus ce ratio est proche de 0, au plus notre évaluation est cohérente. Saaty donne une valeur au plus de 10% d'erreur pour pouvoir accepter l'évaluation. Le cas contraire, il faudra refaire l'évaluation.

L'indice de cohérence se calcule de la manière suivante :

$$IC = (\lambda_{\max} - n) / (n - 1)$$

- λ_{\max} : Valeur propre maximale.
- n : taille de la matrice

2.4.3.5 Calculer le « poids » final

Une fois l'ensemble des poids des critères par niveau déterminé, on va calculer le poids final en multipliant simplement les poids des critères du niveau le plus bas par les poids des critères correspondants des niveaux supérieurs.

Remarque : Les poids mesurés dans chaque niveau peuvent être utilisés directement dans ce même niveau sans avoir besoin de calculer un poids final du niveau supérieur.

Conclusion

Dans ce chapitre, les techniques statistiques utilisées par la suite ont été détaillées et expliquées. Leurs résultats seront utilisés dans le cadre d'un choix optimal d'expériences pour une modélisation avec Deep Learning. La théorie de cette modélisation est détaillée dans le chapitre suivant.

CHAPITRE 3 : RESEAUX DE NEURONES ARTIFICIELS

3. RESEAUX DE NEURONES ARTIFICIELS

3.1 INTRODUCTION

Les réseaux de neurones ont commencé à se développer aux alentours des années 40, avec les travaux de Warren McCulloch and Walter Pitts. Si leur approche était principalement orientée vers des applications mathématiques, d'autres scientifiques se sont penchés sur la question d'un point de vue plus théorique.

Les réseaux de neurones artificiels (RNA) constituent une nouvelle méthode d'approximation de systèmes complexes, particulièrement utile lorsque ces systèmes sont difficiles à modéliser à l'aide des méthodes statistiques classiques. Une classe de modèles de RNA appelée perceptrons multicouches (PMC) a, ces dernières années, été privilégiée pour la prévision de phénomènes hydrologiques. La théorie et le langage connexionniste restent, malgré cette percée, encore peu connus de la communauté des hydrologues.

Dans ce chapitre, on va présenter l'aspect théorique des réseaux de neurones artificiels, leur architecture ainsi que ses objectifs et ses applications.

3.2 LES RESEAUX DE NEURONES ARTIFICIELS

3.2.1 NOTION DE NEURONE

Un neurone est un objet mathématique qui fut à l'origine introduit, entre autres choses, pour modéliser le fonctionnement du cerveau humain dans le cadre d'études de la cognition. En interconnectant plusieurs neurones, nous formons alors un réseau de neurones.

Un neurone est donc l'unité élémentaire de calcul d'un réseau de neurones. Son principe général est de retourner une information en sortie à partir de plusieurs informations en entrée. L'information entrante peut être, par exemple, issue de l'information sortante d'autres neurones dans le cadre d'un réseau. Plus précisément, nous notons $x_1, \dots, x_n \in \mathbb{R}$ les informations entrantes et, pour chaque $i \in \{1, \dots, n\}$, nous associons un poids $w_i \in \mathbb{R}$ à x_i . Contrairement aux poids que nous avons utilisés en cours, il faut noter que les w_i sont ici des nombres réels potentiellement négatifs et dont la somme ne vaut pas nécessairement 1. Nous introduisons également un poids $w_0 \in \mathbb{R}$, appelé coefficient de biais, associé à une information virtuelle $x_0 = -1$ dont le rôle sera précisé ultérieurement. L'information traitée par le neurone n'est pas l'ensemble des x_i mais leur moyenne pondérée.

$$\bar{x} = \sum_{i=0}^n w_i x_i = \sum_{i=1}^n w_i x_i - w_0$$

Le rôle d'un neurone est de fournir une réponse y entre 0 et 1 à partir de \bar{x} . Nous utilisons pour cela une fonction $g : \mathbb{R} \rightarrow [0, 1]$, appelée fonction d'activation. Lorsque la réponse est proche de 1, nous dirons que le neurone est actif et pour une réponse proche de 0, le neurone sera dit inactif. Nous avons donc la sortie du neurone qui est définie par :

$$y = g(\bar{x}) = g\left(\sum_{i=1}^n w_i x_i - w_0\right)$$

Il faut noter que si g est une fonction linéaire des informations entrantes, ce modèle correspond à la régression linéaire. Bien entendu, les réseaux de neurones deviennent plus intéressants quand g n'est pas une fonction linéaire. (Data_Mining).

3.2.2 DEFINITION

Les réseaux de neurones artificiels constituent une technique d'apprentissage et de traitement automatique qui s'inspire du fonctionnement du cerveau humain. Les réseaux de neurones artificiels peuvent être définis comme une structure de traitement parallèle massif constituée de plusieurs unités très simples (que l'on appelle des neurones), qui sont capables de stocker des connaissances expérimentales et de les rendre utilisables.

Les réseaux de neurones artificiels sont similaires aux réseaux de neurones biologiques sur plusieurs points :

1. Les neurones sont des éléments simples et fortement connectés (même si les neurones artificiels sont beaucoup plus simples).
2. Les connexions entre les neurones permettent de déterminer la fonction du réseau. Ces connexions, connues sous le nom de « poids » (poids synaptiques), servent à stocker les connaissances acquises.
3. Les connaissances sont extraites de leur environnement via un processus d'apprentissage.(mathworks, 2018).

3.2.3 ARCHITECTURE

L'architecture (également appelée topologie) du réseau désigne la disposition des neurones dans ce réseau. Les neurones sont organisés en couches, le réseau de neurones peut donc être constitué d'une ou plusieurs couches de neurones.

Chaque neurone reçoit un ensemble d'entrées multipliées par leur interconnexion (poids), qui sont ensuite additionnées et calculées par une fonction de transfert (ou fonction d'activation), avant d'être transmises à la couche suivante ou de former la sortie du réseau.

3.2.3.1 RESEAUX DE NEURONES MONOCOUCHE

Le premier réseau de neurones conçu est le perceptron mono-couche. Les neurones ne sont pas, à proprement parlé, en réseau mais ils sont considérés comme un ensemble. Ici, nous considérons n variables d'entrée $x_1, \dots, x_n \in \mathbb{R}$. Le perceptron mono-couche est composé de p neurones, chacun étant connecté à toutes les variables d'entrée. Globalement, ce réseau dispose donc de n entrées et de p sorties et il peut être représenté comme sur la figure suivante :

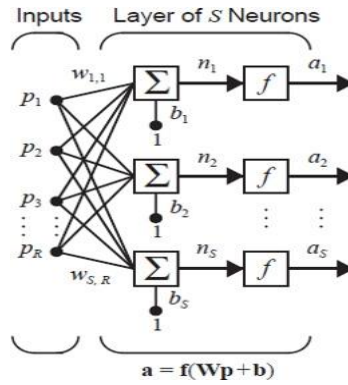


Figure 3.1: Réseau d'une seule couche, qui contient R éléments d'entrée et S neurones. (mathworks, 2018).

3.2.3.2 RESEAUX DE NEURONES MULTICOUCHES

Un réseau peut avoir plusieurs couches. Chaque couche possède une matrice de poids \mathbf{W} , un biais vecteur \mathbf{b} et une sortie vectorielle \mathbf{Un} . Pour distinguer entre les matrices de poids, les vecteurs de sortie, etc., pour chacune de ces couches dans les chiffres, le numéro du calque est ajouté en exposant à la variable d'intérêt. Vous pouvez voir l'utilisation de cette couche notation dans le réseau de trois-couche illustré ci-après et dans les équations au bas de la figure suivante :

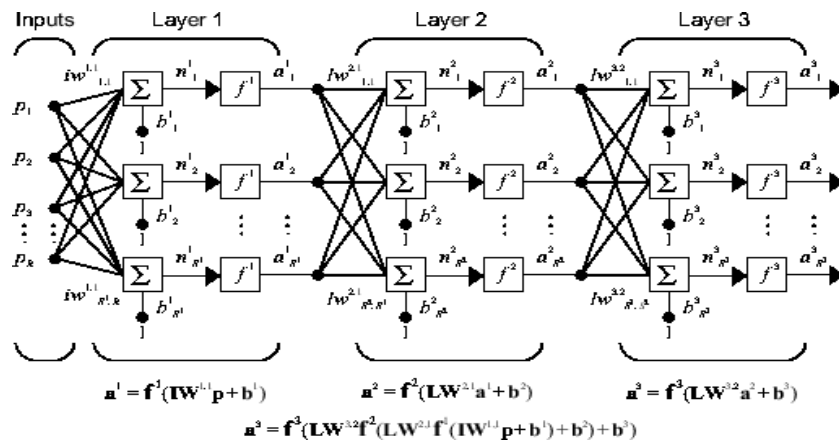


Figure 3.2: Réseau de neurones multicouches. (mathworks, 2018).

Le réseau indiqué ci-dessus a R^1 entrées, S^1 neurones dans la première couche, S^2 neurones dans la deuxième couche, etc. Il est commun pour les différentes couches d'avoir différents nombres de neurones. Une entrée constante 1 est alimentée à la polarisation pour chaque neurone.

A Noter que les sorties de chaque couche intermédiaire sont entrées à la couche suivante. Ainsi, la couche 2 peut être analysée comme une couche réseau avec S^1 entrées, S^2 les neurones et un $S^2 \times S^1$ matrice de poids \mathbf{W}^2 . L'entrée de couche 2 est \mathbf{Un}^1 ; la sortie est \mathbf{Un}^2 . Maintenant que tous les vecteurs et les matrices de la couche 2 ont été identifiés, elle peut être traitée comme un réseau simple couche. Cette approche peut être prise avec n'importe quelle couche du réseau.

Les couches d'un réseau multicouche jouent des rôles différents. Une couche qui produit la sortie réseau est appelée une *couche de sortie*. Tous les autres calques sont appelés *cachés les*

couches. Les trois couches du réseau illustré précédemment à une sortie (couche 3) et deux des couches cachées (couche 2 et couche 1). Certains auteurs considèrent la couche d'entrées comme une quatrième couche.

L'architecture d'un réseau multicouche avec un vecteur d'entrée unique peut être spécifiée avec la notation $R - S^1 - S^2 - \dots - S^M$, où le nombre d'éléments du vecteur d'entrée et le nombre de neurones dans chaque couche sont spécifiés. (mathworks, 2018).

3.2.4 Apprentissage

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

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

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

Dans le cas des réseaux de neurones artificiels, on ajoute souvent à la description du modèle l'algorithme d'apprentissage. Le modèle sans apprentissage présente en effet peu d'intérêt.

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

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

3.2.4.1 Apprentissage Supervisé

L'apprentissage supervisé désigne un type d'algorithme d'apprentissage automatique qui utilise un ensemble de données connues (appelé ensemble de données d'apprentissage) pour réaliser des prévisions. L'ensemble de données d'apprentissage contient des données d'entrée et des valeurs de réponses. À partir de ces éléments, l'objectif de l'algorithme d'apprentissage supervisé consiste à créer un modèle capable de réaliser des prévisions sur les valeurs de réponse pour un nouvel ensemble de données. Un ensemble de données de test est souvent utilisé pour valider le modèle. Les ensembles de données d'apprentissage plus volumineux produisent généralement des modèles à la puissance prédictive plus élevée qui peuvent facilement s'appliquer aux nouveaux ensembles de données.

L'apprentissage supervisé inclut deux catégories d'algorithmes :

- Classification : pour les valeurs de réponse catégoriques, où les données peuvent être divisées en « classes » spécifiques.
- Régression : pour les valeurs de réponse continue.(mathworks, 2018)

3.2.4.2 Apprentissage non supervisé

L'apprentissage non supervisé est un type d'algorithme d'apprentissage automatique utilisé pour tirer des conclusions à partir de jeux de données composés de données d'entrée sans réponses catégorisées.

La méthode d'apprentissage non supervisé la plus commune est le partitionnement de données, qui est utilisé pour effectuer une analyse exploratoire des données afin de trouver des modèles cachés ou des regroupements dans les données. Les clusters sont conçus au moyen d'une mesure de similarité définie par des métriques telles que la distance euclidienne ou la distance probabiliste.

Les algorithmes de Clustering communs comprennent :

- **Regroupement hiérarchique** : permet de développer une hiérarchie de clusters à plusieurs niveaux en créant un arbre de clusters
- **Partitionnement k-means** : partitionne les données dans k clusters distincts selon la distance au barycentre d'un cluster
- **Modèle de mélange gaussien** : modélise les clusters comme mélange de composants suivant la loi normale multivariée
- **Carte auto adaptative** : utilise les réseaux neuronaux qui apprennent la topologie et la répartition des données
- **Modèle de Markov caché** : utilise les données observées pour récupérer la séquence d'états

Les méthodes d'apprentissage non supervisé sont utilisées en bio-informatique pour les analyses de séquences et les regroupements génétiques ; en data Mining pour l'extraction de séquences et de modèles ; en imagerie médicale pour les segmentations d'images ; et en vision par ordinateur pour la reconnaissance d'objets.(mathworks, 2018).

3.2.4.3 FONCTION D'ERREUR

Une fonction de coût est une mesure de la « qualité » d'un réseau de neurones par rapport à son échantillon d'apprentissage donné et à la production attendue. Cela peut aussi dépendre de variables telles que les poids et les biais.

Une fonction de coût est une valeur unique, et non un vecteur, car elle évalue la qualité de l'ensemble du réseau de neurones.

Plus précisément, une fonction de coût est de la forme

$C(W, B, S_r, E_r)$

Où W est le poids de notre réseau neuronal, B est le biais de notre réseau neuronal, S_r est l'entrée d'un seul échantillon d'apprentissage, et E_r est la sortie désirée de cet échantillon d'apprentissage. Notez que cette fonction peut aussi potentiellement dépendre de y_{ij} et z_{ij} pour n'importe quel neurone j dans la couche i , parce que ces valeurs dépendent de W , B et S_r .

En backpropagation, la fonction de coût est utilisée pour calculer l'erreur de notre couche de sortie, δ^L , via

$$\delta_j^L = \frac{\partial C}{\partial a_j^L} \sigma'(z_j^i)$$

Qui peut aussi être écrit en tant que vecteur via

$$\delta^L = \nabla_a C \odot \sigma'(z^i)$$

Pour utiliser les fonctions d'activations, deux critères importants doivent être vérifiés :

- La fonction de coût C doit pouvoir être écrite en tant que moyenne.

$$C = \frac{1}{n} \sum_x C_x$$

- La fonction de coût C ne doit pas dépendre des valeurs d'activation d'un réseau de neurones en plus des valeurs de sortie.

Les fonctions de couts sont de plusieurs types, citons quelques-uns :

- **Coût quadratique** : Aussi connue sous le nom d'erreur quadratique moyenne, de maximum de vraisemblance et d'erreur quadratique, elle est définie comme :

$$C_{MST}(W, B, S^r, E^r) = 0.5 \sum_i (a_j^L - E_j^r)^2$$

Où W est le poids de notre réseau neuronal, B est le biais de notre réseau neuronal, S_r est l'entrée d'un seul échantillon d'apprentissage, et E_r est la sortie désirée de cet échantillon d'apprentissage, a_{ij} représente la valeur d'activation du j ème neurone dans la couche i .

- **Coût de l'entropie croisée** : Aussi connu sous le nom de log-vraisemblance négative de Bernoulli et d'entropie binaire. (stats.stackexchange, 2018).

$$C_{CE}(W, B, S^r, E^r) = - \sum_j [E_j^r \ln a_j^L + (1 - E_j^r) \ln (1 - a_j^L)]$$

3.2.4.4 FONCTION D'ACTIVATION

3.2.4.4.1 DEFINITION

Les réseaux de neurones biologiques ont inspiré le développement de réseaux de neurones artificiels. Un neurone typique a une structure physique qui consiste en un corps cellulaire, un

axone qui envoie des messages à d'autres neurones, et des dendrites qui reçoivent des signaux ou des informations d'autres neurones.

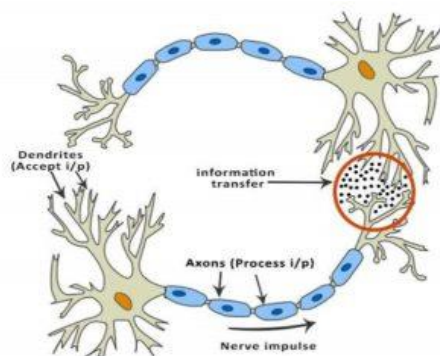


Figure 3.3: Connexion entre deux neurones. (deep-learning, 2018).

Dans l'image ci-dessus, le cercle rouge indique la région où les deux neurones communiquent. Le neurone reçoit des signaux d'autres neurones à travers les dendrites. Le poids (force) associé à une dendrite, appelé poids synaptique, est multiplié par le signal entrant. Les signaux provenant des dendrites sont accumulés dans le corps cellulaire, et si la force du signal résultant est supérieure à un certain seuil, le neurone passe le message à l'axone. Sinon, le signal est tué par le neurone et n'est plus propagé.

La fonction d'activation prend la décision de passer ou non le signal. Dans ce cas, il s'agit d'une fonction d'étape simple avec un seul paramètre - le seuil. Maintenant, quand nous apprenons quelque chose de nouveau (ou désapprendre quelque chose), le seuil et les poids synaptiques de certains neurones changent. Cela crée de nouvelles connexions entre les neurones, ce qui permet au cerveau d'apprendre de nouvelles choses.

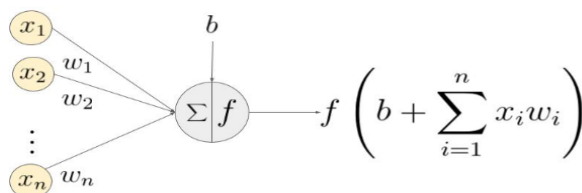


Figure 3.4: Schéma d'un signal d'entrée et fonction d'activation. (deep-learning, 2018).

Dans la figure ci-dessus, (x_1, x_2, \dots, x_n) est le vecteur de signal qui se multiplie avec les poids (w_1, w_2, \dots, w_n) . Ceci est suivi par l'accumulation (c'est-à-dire la sommation + l'addition du biais b). Enfin, une fonction d'activation f est appliquée à cette somme.

A noter que les poids (w_1, w_2, \dots, w_n) et le biais b transforment linéairement le signal d'entrée. L'activation, d'autre part, transforme le signal non-linéairement et c'est cette non-linéarité qui nous permet d'apprendre des transformations arbitrairement complexes entre l'entrée et la sortie.

Au fil des années, diverses fonctions ont été utilisées, et il reste un domaine de recherche actif pour trouver une fonction d'activation qui rend le réseau neuronal plus performant et plus rapide. (deep-learning, 2018).

3.2.4.4.2 TYPES DE FONCTION D'ACTIVATION

Plusieurs types de fonctions d'activation existent, parmi lesquelles :

Tableau 3-1 : Fonctions d'activation (deep-learning, 2018).

Nom	Type	Expression
Fonction d'activation linéaire	Linéaire	$f(x) = x$
Sigmoïde	Non linéaire	$\sigma(x) = \frac{1}{1 + e^{-x}}$
Tangente hyperbolique (Tanh)	Non linéaire	$\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1}$
Rectified Linear Unit (ReLU)	Non linéaire	$f(x) = \max(0, x)$

La fonction ReLU est la fonction d'activation principale qu'on utilisera dans la partie modélisation, car elle a permis d'accroître les performances de nos réseaux de neurones d'une façon significative.

3.2.4.5 Rétropropagation (Backpropagation)

La rétropropagation du gradient de l'erreur (ou backpropagation) est un algorithme d'optimisation permettant d'ajuster les paramètres d'un réseau de neurones multicouches pour mettre en correspondance des entrées et des sorties référencées dans une base d'apprentissage.

Pour pouvoir entraîner ces systèmes, il faut savoir comment ajuster les paramètres de chaque couche de neurones. La rétropropagation permet de calculer le gradient de l'erreur pour chaque neurone, de la dernière couche vers la première. Le calcul de ce gradient se fait par la méthode de rétropropagation. Cela permet de corriger les erreurs selon l'importance des éléments qui ont justement participé à la réalisation de ces erreurs. Ainsi, les poids synaptiques qui contribuent à engendrer une erreur importante se verront modifiés de manière plus significative que les poids qui ont engendré une erreur marginale. Moyennant quelques précautions lors de l'apprentissage, les procédures d'optimisation finissent par aboutir à une configuration stable, généralement un extremum local, au sein du réseau de neurones.

Dans les réseaux de neurones artificiels, les poids sont mis à jour en utilisant une méthode appelée Backpropagation. Les dérivées partielles de la fonction d'erreur avec les poids sont utilisées pour mettre à jour les poids. Dans un sens, l'erreur est rétropropagée dans le réseau en utilisant ces dérivés. Ceci est fait de manière itérative et après de nombreuses itérations, l'erreur atteint une valeur minimale, et la dérivée de la perte devient nulle.

L'optimisation basée sur le gradient stochastique revêt une importance pratique fondamentale dans de nombreux domaines scientifiques et ingénierie. De nombreux problèmes dans ces domaines peuvent être exprimés comme l'optimisation de certains paramètres paramétrés scalaires nécessitant une maximisation ou une minimisation par rapport à ses paramètres. (Adam, 2015).

3.3 Deep learning

L'apprentissage profond (en anglais Deep learning, Deep structured learning, hierarchical learning) est un ensemble de méthodes d'apprentissage automatique tentant de modéliser avec un haut niveau d'abstraction des données grâce à des architectures articulées de différentes transformations non linéaires. Ces techniques ont permis des progrès importants et rapides dans de nombreux domaines. Ces progrès ont suscité des investissements privés, universitaires et publics importants, notamment de la part de Google. (Schmidhuber, 2015).

L'apprentissage profond utilise des couches cachées de réseaux de neurones artificiels, des « Machines de Boltzmann restreintes », et des séries de calculs propositionnels complexes. Les algorithmes d'apprentissage profond s'opposent aux algorithmes d'apprentissage peu profonds du fait du nombre de transformations réalisées sur les données entre la couche d'entrée et la couche de sortie, où une transformation correspond à une unité de traitement définie par des poids et des seuils.

Le concept d'apprentissage profond prend forme dans les années 2010, avec la convergence de trois facteurs :

- des réseaux de neurones artificiels multicouches (eux-mêmes issus entre autres du concept de perceptron, datant de la fin des années 1950)
- des algorithmes d'analyse discriminante et apprenants (dont l'émergence remonte aux années 1980).
- Des machines dont la puissance de traitement permet de traiter des données massives. (Collobert, 2011).

3.3.1 Comment fonctionne le Deep Learning

La plupart des méthodes d'apprentissage en profondeur utilisent des architectures de réseaux neuronaux, ce qui explique pourquoi les modèles d'apprentissage profond sont souvent appelés réseaux de neurones profonds.

Le terme « profond » fait généralement référence au nombre de couches cachées dans le réseau de neurones. Les réseaux neuronaux traditionnels ne contiennent que deux ou trois couches cachées, tandis que les réseaux profonds peuvent en contenir jusqu'à 150.

Les modèles d'apprentissage en profondeur sont formés en utilisant de grands ensembles de données étiquetées et des architectures de réseau neuronal qui apprennent les caractéristiques directement à partir des données sans avoir besoin d'extraire manuellement les caractéristiques.

3.3.2 Quelle est la différence entre l'apprentissage automatique et le Deep Learning ?

L'apprentissage profond est une forme spécialisée d'apprentissage automatique. Un workflow d'apprentissage automatique commence par l'extraction manuelle des caractéristiques pertinentes à partir des images. Les entités sont ensuite utilisées pour créer un modèle qui catégorise les objets dans l'image. Avec un workflow d'apprentissage en profondeur, les fonctionnalités pertinentes sont automatiquement extraites des images. En outre, l'apprentissage en profondeur effectue un « apprentissage de bout en bout » : un réseau reçoit des données brutes et une tâche à effectuer, telle que la classification, et apprend à le faire automatiquement.

Une autre différence clé est l'échelle des algorithmes d'apprentissage profond avec des données, alors que l'apprentissage superficiel converge. L'apprentissage superficiel fait référence à des méthodes d'apprentissage automatique qui se stabilisent à un certain niveau de performance lorsque vous ajoutez d'autres exemples et des données d'apprentissage au réseau.

Un avantage clé des réseaux d'apprentissage en profondeur est qu'ils continuent souvent à s'améliorer à mesure que la taille de vos données augmente.

Conclusion

A travers ce chapitre, la théorie des modèles utilisés dans le cadre de cette étude a été mise en avant concernant le Deep Learning. Une étude statistique en amont de ces modèles sera menée dans le chapitre suivant afin de déterminer les modèles les plus optimaux en vue de la résolution de la problématique posée.

CHAPITRE 4 : Présentation des STEP de Boumerdes et Tipaza

4. Présentation des STEP de Boumerdes et Tipaza

4.1 STEP de Boumerdes

4.1.1 INTRODUCTION

La station d'épuration des eaux usées de Boumerdes est une station à boues activées à faible charge massique fonctionnant à aération prolongée. Elle assure l'épuration d'un débit journalier de 15000 m³/j. Elle a été mise en service en 2006 et est gérée par l'Office National d'Assainissement (ONA).

La station d'épuration des eaux usées de Tipaza, quant à elle, est une station à boues activées. Son débit journalier est de 11200 m³/j. Elle a été mise en service en 2008 et est actuellement gérée par SEAAL.

4.1.2 Localisation de la STEP

La STEP de Boumerdes est située en zone côtière à 50 Km à l'Est d'Alger, sur le territoire du chef-lieu de la wilaya, soit la commune de Boumerdes.

Le terrain d'assiette du projet se situe à 1.5 Kilomètre au Sud-Est de l'agglomération de Boumerdes, en contre bas de la Route Nationale 24 sur la rive gauche de l'Oued Tatareg et à proximité de la confluence de celui-ci avec l'Oued Safsaf.

Cette station occupe une superficie de 3,11 hectares et une capacité de traitement de 75000 Equivalent Habitant (EH), elle est destinée à épurer les eaux usées domestiques de la ville de Boumerdes et des communes voisines telles que Tidjellabine et Corso.

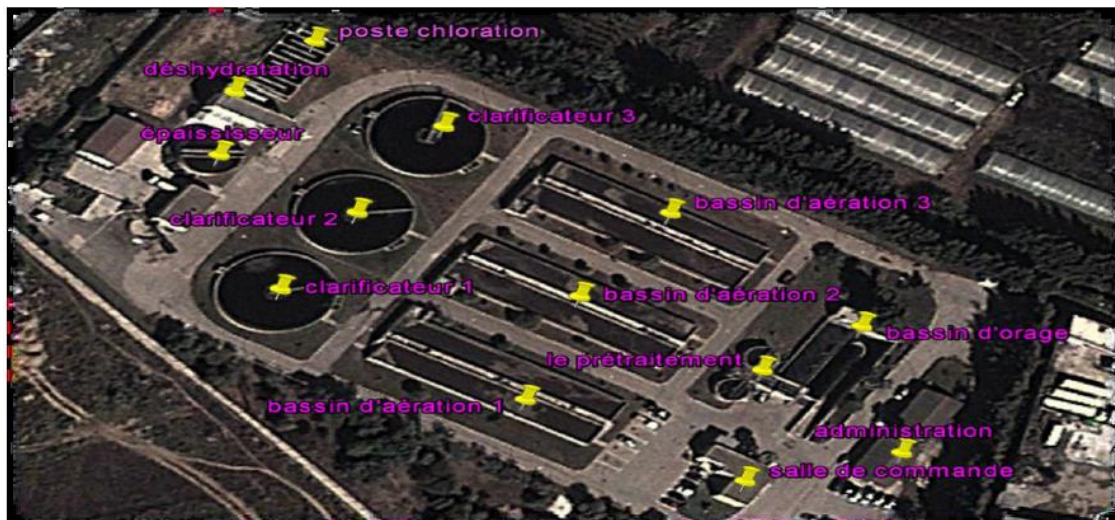


Figure 4.1: Vue d'ensemble des ouvrages de la STEP de Boumerdes par Google Earth.

4.1.3 Caractéristiques Techniques

La station de Boumerdes est du type « boues activées à faible charge massique à aération prolongée » ($C_m = 0.076 \text{ kg DBO/kg MVS/Jour}$).

L'eau à traiter possède les caractéristiques suivantes :

- Volume journalier..... 15000 m³/j

- Débit moyen 24 h..... 625 m3/h
- Débit de pointe temps sec 1063 m3/h
- Débit de pointe temps de pluie..... 1944 m3/h

4.1.4 Paramètres de pollution de conception

Le tableau suivant donne les concentrations admissibles des paramètres de pollution.

Tableau 4-1: Paramètres de pollution de conception de la station de Boumerdes

	Entrée	Sortie	Abattement (%)
DBO5 (mg/l)	270	30	88.89
DCO (mg/l)	750	90	88.00
MES (mg/l)	350	30	91.43

4.1.5 Les étapes du procédé d'épuration :

Les eaux usées arrivent à la station par gravité ou par pompage, et passent par différentes étapes de traitement qui sont schématisées dans la Synoptique suivante :

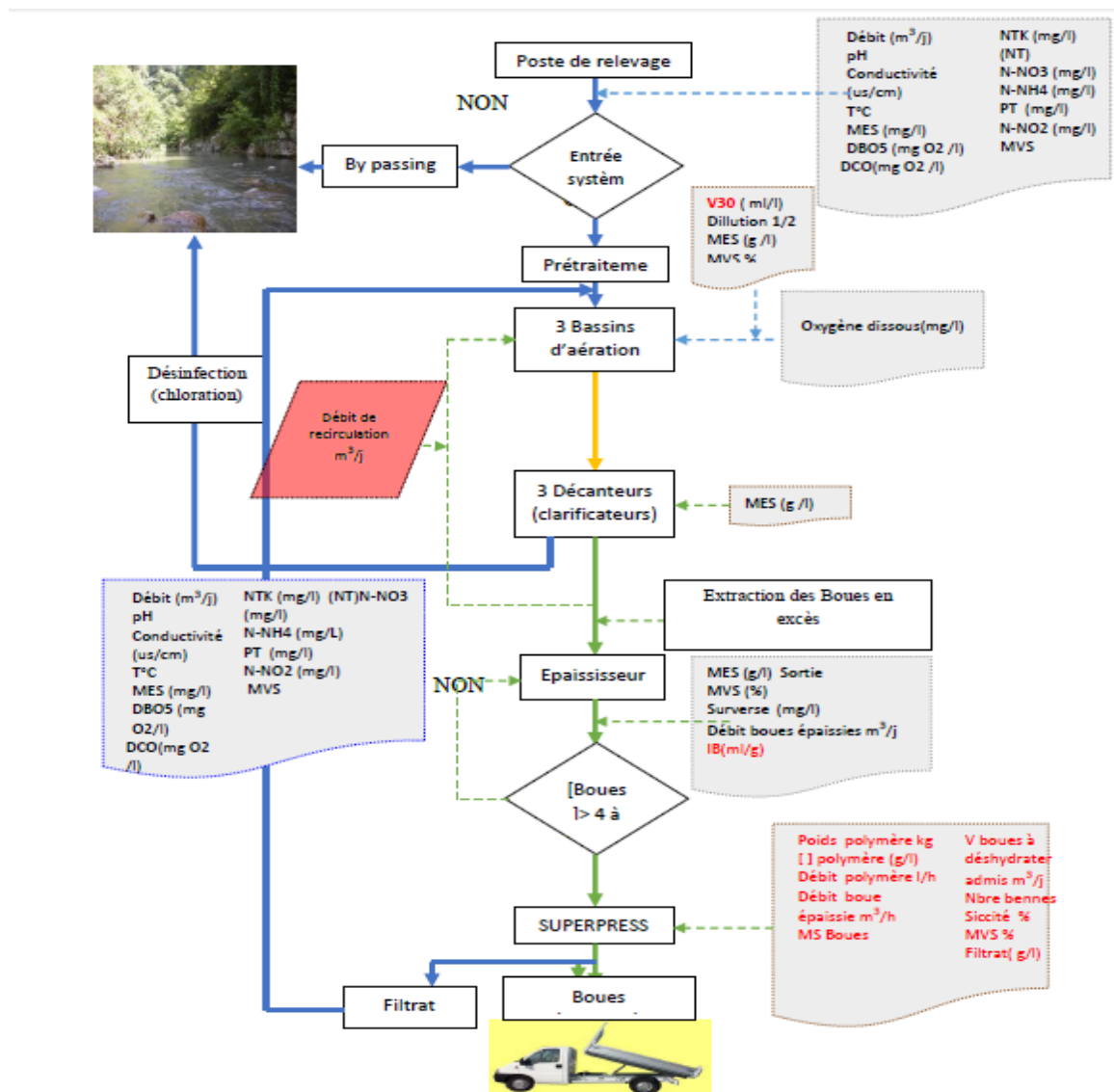


Figure 4.2 : Organigramme d'exploitation de la STEP de Boumerdes. (LAREDJ.Nesrine, 2013).

4.1.5.1 Principe du traitement

La STEP de Boumerdes comporte deux chaînes de traitement :

Traitement au fil de l'eau : le prétraitement, le traitement biologique à boues activées, décantation et désinfection.

Traitement au fil de boues : épaissement, déshydratation et lit de séchage.

Procédé biologique

- Bassins d'aération

Les eaux prétraitées sont dirigées vers 3 bassins d'aération munis de turbines type <<actirotor>>, aérateur à axe vertical et à vitesse lente, permettant l'aération prolongée de la culture bactérienne à l'origine du traitement. Les bassins reçoivent d'autre part la <<liqueur mixte >> constituée par la recirculation des boues provenant des clarificateurs finaux.

Caractéristiques : Nombre de bassins = 3, volume unitaire= 3600 m³, profondeur= 4.5m, nombre d'aérateurs = 9.



Figure 4.3 : Bassin d'aération. (DJAOUI.Toufik, 2016).

- Clarification finale et ouvrage de recirculation

Les eaux sortantes des ouvrages d'aération sont dirigées vers 3 clarificateurs de diamètre 24 m, les boues décantées au fond de chaque ouvrage sont dirigées à l'aide d'un racleur vers un puits central de collecte. Elles sont reprises par une tuyauterie les acheminant vers la bache de recirculation attenante aux bassins d'aération sous le nom de la liqueur mixte et une partie est extraite pour être envoyée vers l'épaississeur. Les eaux décantées sont recueillies par surverse dans une rigole périphérique pour être dirigée vers la désinfection et le canal de comptage.



Figure 44.4 : Décanteur-Clarificateur. (DJAOUI.Toufik, 2016).

4.1.6 Statistiques et critiques des données :

4.1.6.1 Statistiques Descriptives

La STEP de Boumerdes a mis à notre disposition une série de données journalières de mesure de différents paramètres à l'entrée et à la sortie durant la période (2006 - 2016).

Dans le cadre de ce travail, nous avons procédé à une analyse générale de ces données : une critique visuelle des graphiques de variation des paramètres en fonction du temps,

Les paramètres pris en considération sont résumés dans le tableau suivant :

Tableau 44-2: Statistiques descriptives des paramètres de pollution de la STEP Boumerdes

	Eau Brute				Eau Epurée			
	Min	Max	Moyenne	Ecart-type	Min	Max	Moyenne	Ecart-type
pH	6.71	8.08	7.32	0.14	6.39	8.03	7.27	0.16
T (°C)	7.55	35.93	20.35	3.82	2.43	29.9	20.27	4
MES [mg/l]	24	2400	240.18	137.31	1	160	13.53	8.21
DB05[mg/l]	20	2500	248.64	156.38	1	55	10.69	8.28
DCO [mg/l]	110	3264	472.98	251.27	1	173	42.14	24.83
NH4 [mg/l]	0.8	126.54	30.98	18.49	0.001	60.1	5.37	8.24
NO2 [mg/l]	0	2.7	0.20	0.28	0.001	9.4	0.44	0.96
NO3 [mg/l]	0	145.	4.62	9.50	0.001	1529.15	34.98	156.73
NTK [mg/l]	1.5	218.96	48.98	21.97	0.050	644	14.48	32.33
PO4 [mg/l]	0.35	64.44	6.88	5.40	0.1	22.4	3.18	2.58
Débit [m3/j]	42	22491	11991.3	4186.17	/	/	/	/
Energie [KWh]	11	8355	3932.84	1082.98	/	/	/	/
RB [m3/j]	0	27170	16497	4885.05	/	/	/	/

4.1.6.2 Evolution des paramètres de pollution dans le temps

Un autre moyen d'avoir une idée générale sur l'ensemble des données recueillies est d'étudier l'évolution des différents paramètres dans le temps, les résultats sont présentés dans les figures suivantes :

4.1.6.2.1 Evolution du PH

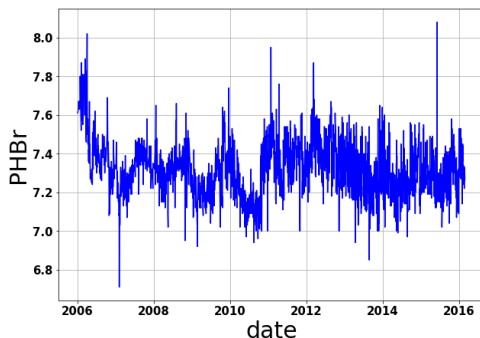


Figure 4.5: Evolution journalière du PH de l'eau brute

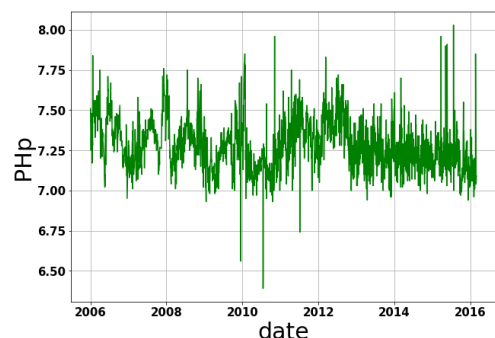


Figure 4.6: Evolution journalière du PH de l'eau

Graphiquement, on remarque que l'évolution du pH au cours du temps n'est pas significative, cette dernière oscillant autour d'une moyenne de 7.32 et 7.27, un minimum de 6.71 et 6.39 et un maximum de 8.08 et 8.03, respectivement à l'entrée et à la sortie de la station.

4.1.6.2.2 Evolution de la température

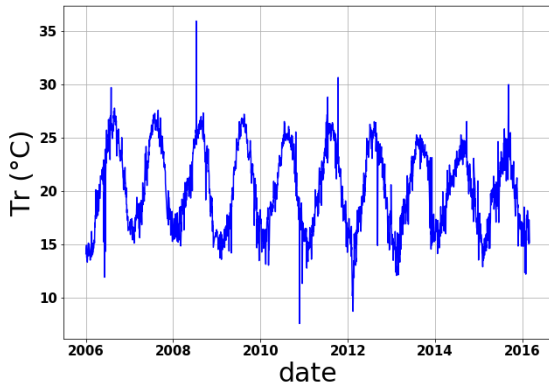


Figure 4.7: Variation journalière de la température de l'eau brute

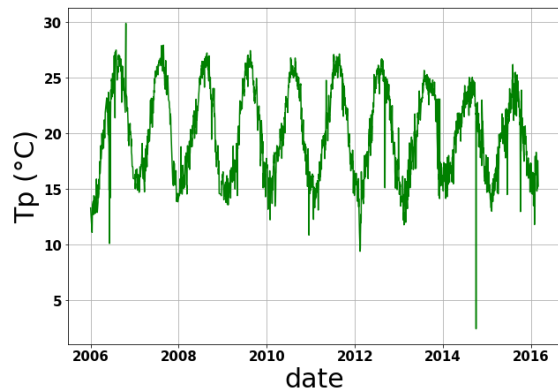


Figure 4-8 : Variation journalière de la température de l'eau épurée

On voit bien que l'évolution de la température au cours du temps se fait d'une façon cyclique, avec une moyenne de 20.35°C et 20.27°C, un minimum de 7.55°C et 2.43°C et un maximum de 35.93°C et 29.9°C, respectivement à l'entrée et à la sortie de la station.

4.1.6.2.3 Evolution des Matières En Suspension (MES)

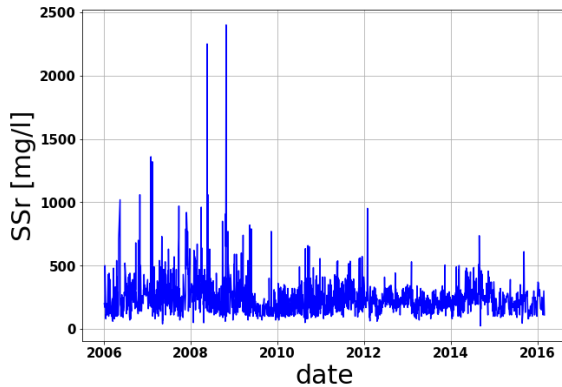


Figure 4.9 : Evolution journalière de la concentration en MES à l'entrée de la STEP

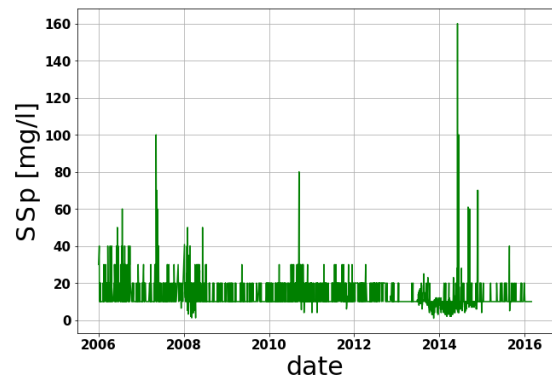


Figure 4.10 : Evolution journalière de la concentration en MES à la sortie de la STEP

A l'entrée de la STEP, les MES ont une concentration moyenne de 240.18 mg/l. Généralement, les valeurs oscillent entre 100 et 500 mg/l ce qui dépasse la concentration admissible par la STEP (350 mg/l).

On enregistre aussi des pics de MES atteignant 2400 mg/l que l'on peut expliquer par des phénomènes exceptionnels, comme l'augmentation du flux de MES par temps de pluie qui provient majoritairement de la remise en suspension des dépôts accumulés.

A la sortie de la STEP, la majorité des valeurs enregistrées sont comprise entre 10 et 20 mg/l, on retient aussi quelques exceptions où la concentration dépasse la norme de rejet (30mg/l).

4.1.6.2.4 Evolution de la Demande Biologique en Oxygène (DBO5)

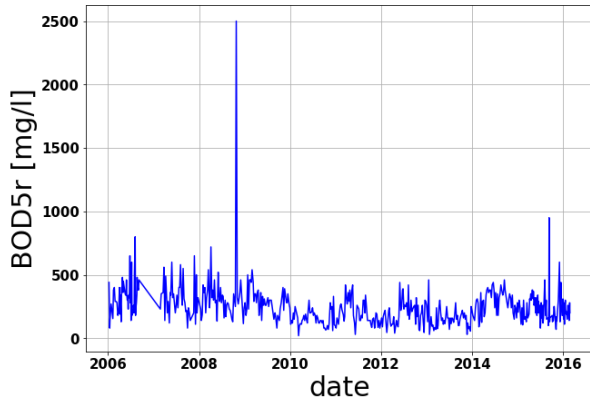


Figure 4.11 : Evolution de la Demande Biologique en Oxygène à l'entrée de la STEP

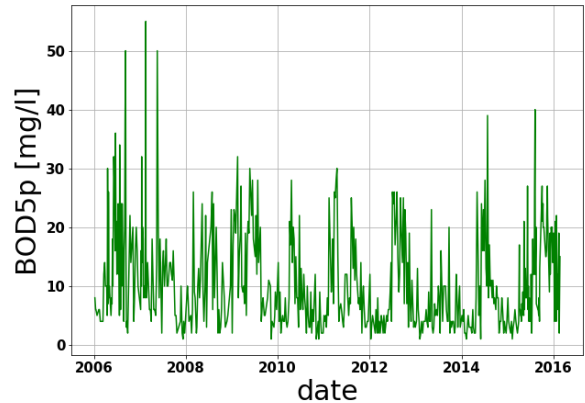


Figure 4.12 : Evolution de la Demande Biologique en Oxygène à la sortie de STEP

A l'entrée de la STEP, la concentration de la Demande Biologique en Oxygène oscille principalement entre 20 et 700 mg/l avec une moyenne de 248.64 mg/l. On souligne aussi le pic de 2500 mg/l enregistré en 2008.

A la sortie de la STEP, la concentration de la DBO5 oscille principalement entre 5 mg/l et 30 mg/l et généralement inférieure à la norme de rejet (30 mg/l) avec quelques exceptions dans les années 2006, 2007 et 2013, 2015 qui peuvent s'expliquer soit par une charge organique entrante supérieure à la charge nominale dont les microorganismes ne peuvent pas tout dégrader, soit par un problème au niveau du bassin d'aération ou encore le travail en surcharge de la STEP à partir de 2013.

4.1.6.2.5 Evolution de la Demande Chimique en Oxygène (DCO)

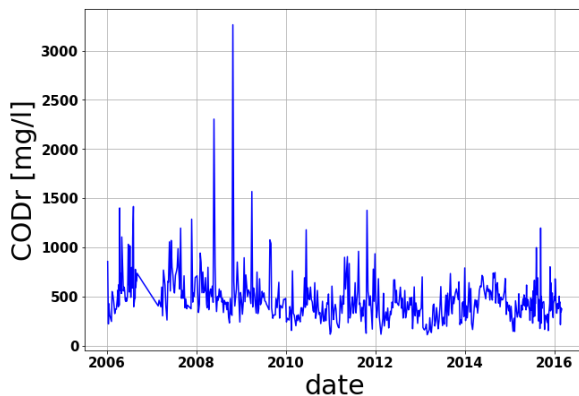


Figure 4.13 : Evolution de la Demande Chimique en Oxygène à l'entrée de la STEP

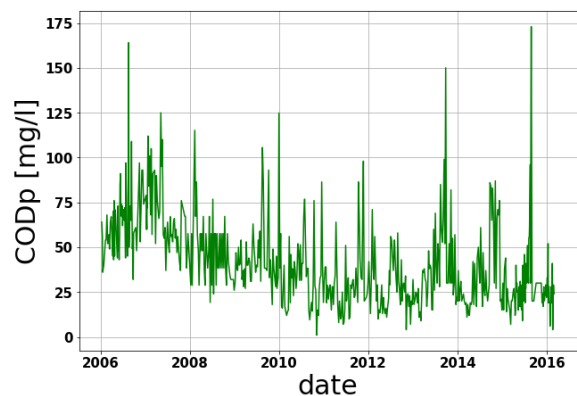


Figure 4. 14 : Evolution de la Demande Chimique en Oxygène à la sortie de la STEP

Graphiquement, on remarque que l'évolution de le DCO au cours du oscille autour d'une moyenne de 472.98mg/l et 42.14mg/l, un minimum de 110 mg/l et 1mg/l et un maximum de 3264 mg/l et 173 mg/l, respectivement à l'entrée et à la sortie de la station.

4.1.6.2.6 Evolution de l'ammonium (NH₄⁺)

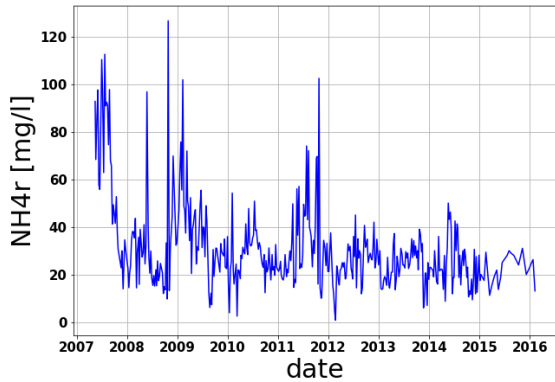


Figure 4-15 : Evolution journalière de l'ammonium (NH₄⁺) à l'entrée de la STEP

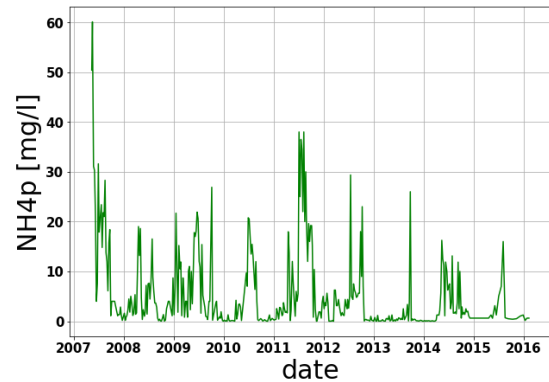


Figure 4.16 : Evolution journalière de l'ammonium (NH₄⁺) à la sortie de la STEP

A l'entrée de la STEP, la concentration d'ammonium oscille entre 0.8 mg/l et 126.5 mg/l avec une moyenne de 31 mg/l

A la sortie de la STEP, la concentration d'ammonium varie entre 0 et 60.1 mg/l avec une moyenne de 5.37 mg/l qui est supérieure à la norme du rejet (5 mg/l). Ceci est dû à de fortes charges organiques entraînant un degré de nitrification plus faible.

4.1.6.2.7 Evolution du Nitrite (NO₂⁻)

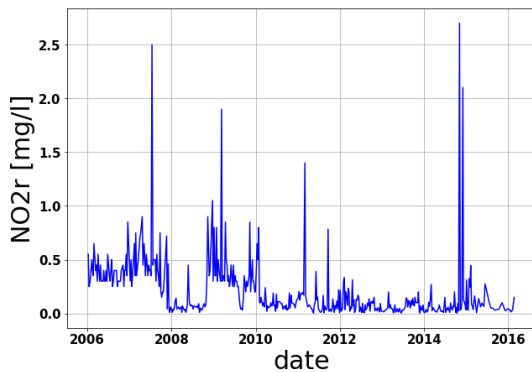


Figure 4.17 : Evolution journalière du Nitrite (NO₂⁻) à l'entrée de la STEP

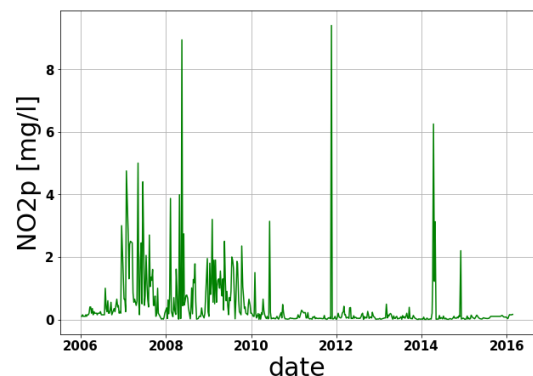


Figure 4.18 : Evolution journalière du Nitrite (NO₂⁻) à la sortie de la STEP

A l'entrée de la STEP, Les concentrations du Nitrite varient entre 0 et 1 mg/l avec quelques pics occasionnels atteignant jusqu'à 2.7 mg/l en 2015.

A la sortie de la STEP, Les concentrations du Nitrite varient entre 0 et 4 mg/l avec quelques pics atteignant jusqu'à 9.4mg/l en 2011.

4.1.6.2.8 Variation du Nitrate (NO₃⁻)

A l'entrée de la STEP, les concentrations du NO₃⁻ oscillent entre 0 et 40 mg/l avec une moyenne de 4.62 mg/l à l'exception de certains cas où la valeur atteint les 145 mg/l en 2007.

A la sortie de la STEP, les concentrations du NO₃⁻ oscillent entre 0 et 30 mg/l où elles dépassent les normes de rejet de l'OMS (10 mg/l).

On remarque aussi des pics atteignant jusqu'à (1600 mg/l) à la sortie de la station qui peuvent être expliqués par un problème de manque de maîtrise du temps d'aération.

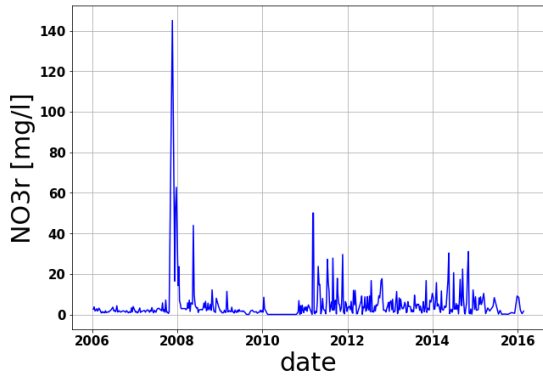


Figure 4.19: Evolution du Nitrate à l'entrée de la STEP

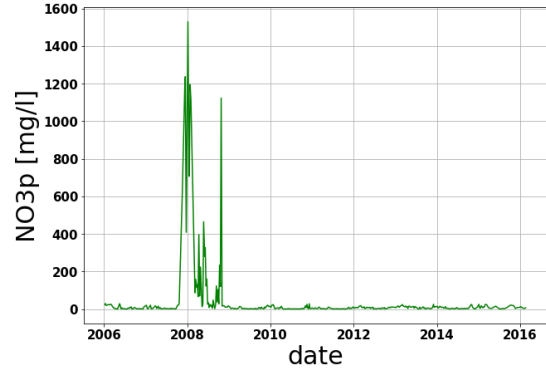


Figure 4.20 : Evolution du Nitrate (NO3-) à la sortie de la STEP

4.1.6.2.9 Evolution de l'azote Kjeldahl NTK

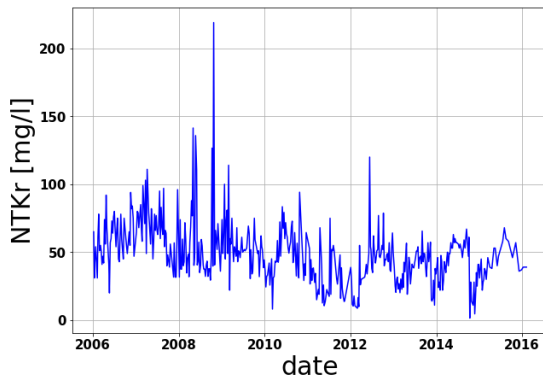


Figure 4.21 : Evolution du NTK à l'entrée de la STEP

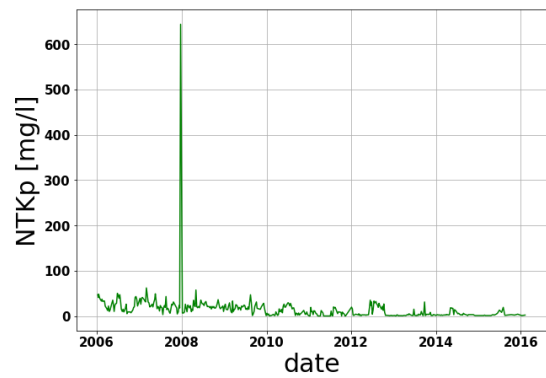


Figure 4.22 : Evolution du NTK à la sortie de la STEP

A l'entrée de la STEP, les concentrations de l'Azote Kjeldahl oscillent entre 0 et 135.8 mg/l, avec une moyenne de 49 mg/l.

Une valeur de 218.9 mg/l un peu plus élevée que les autres valeurs a été enregistrée en 2008.

A la sortie de la STEP, les concentrations de l'Azote Kjeldahl sont en majorité inférieures à la norme du rejet (40 mg/l).

4.1.6.2.10 Evolution de l'ortho-phosphate (PO4³⁻)

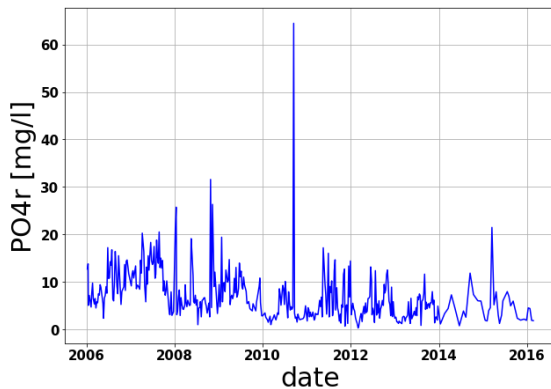


Figure 4-23 : Evolution de PO4-3 à l'entrée de la STEP

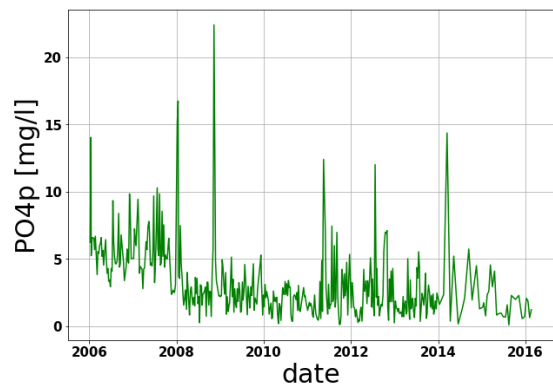


Figure 4-24 : Evolution de PO4-3 à la sortie de la STEP

A l'entrée de la STEP, les concentrations du PO4-3 oscillent entre 0.35 et 31.6 mg/l, avec une moyenne de 6.88 mg/l.

En 2010, un pic de 64,44 mg/l a été enregistré A la sortie de la STEP, les concentrations enregistrées varient entre 0.10 et 15.21 mg/l. Les valeurs enregistrées sont en majorité inférieures à la norme du rejet (10mg/l).

En 2008, un pic de 22.4 mg/l a été enregistré.

Ces valeurs peuvent être expliquées par un faible pourcentage de déphosphatation.

4.1.6.2.11 Evolution de débit journalier

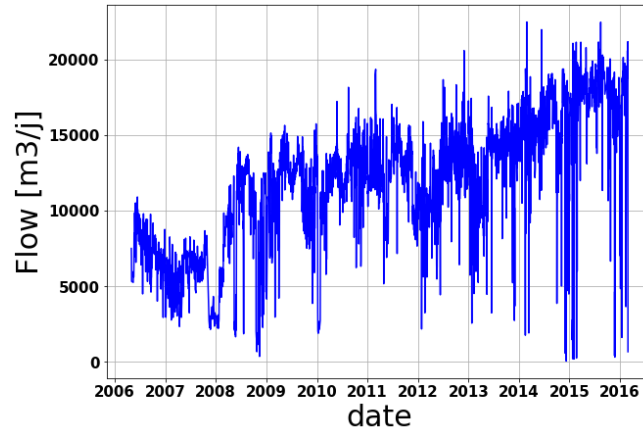


Figure 4.25 : Evolution journalière du débit dans la STEP

Le débit de l’effluent d’entrée varie entre 0 et 15000 m3/j pendant la période 2006, 2013 avec quelques élévations sont dues aux fortes précipitations, mais à partir du 2013 les valeurs du débit atteignent (22491 m3/j) et donc on peut dire que la station à partir de 2013 travaille en surcharge ce qui explique certaines valeurs très élevées des effluents à l’entrée ou à la sortie de la STEP.

Les valeurs qui dépassent la capacité nominale, on peut aussi les expliquer par de fortes précipitations.

4.1.6.2.12 Evolution de la consommation d’énergie

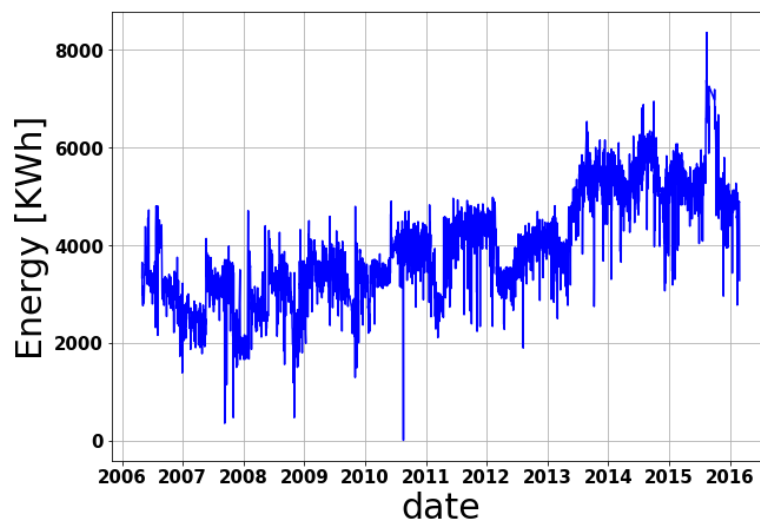


Figure 4.26: Evolution journalière de l’énergie dans la STEP

Les valeurs de l’énergie totale consommée par la STEP de Boumerdes durant le processus d’épuration varient entre 0 et 5000 KWh durant la période 2006, 2013. A partir du 2013 on

remarque une augmentation dans les valeurs de l'énergie qui arrivent à atteindre les 8000 kWh, cette augmentation est expliquée par le travail en surcharge de la station puisque les débits qui arrivent à la station à partir du 2013 sont bien plus élevés que le débit pour lequel elle a été mise en service et donc elle a besoin de plus d'énergie pour dégrader la pollution. Les valeurs nulles enregistrées sont dues à des coupures d'électricité.

4.1.6.2.13 Evolution du débit recirculé

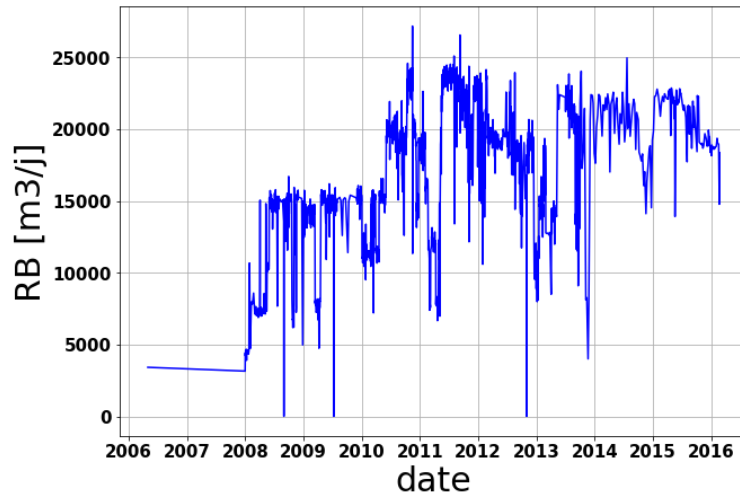


Figure 4.27: Evolution du débit recirculé

Le débit recirculé dans la STEP de Boumerdes augmente progressivement avec le temps avec une moyenne de 16497 m³/j et un maximum de 27170 m³/j, cette augmentation est expliquée par le fait de la surcharge du travail de la station à partir de 2013.

4.2 STEP Tipaza (station de Chenoua)

4.2.1 Localisation

La station d'épuration des eaux usées de la ville de Tipaza est administrativement localisée dans la daïra et la wilaya de Tipaza. Elle est géographiquement située dans la région de Chenoua à environ 70 kilomètres à l'ouest de la wilaya d'Alger.

Mise en service en janvier 2008, la STEP a pour missions de collecter et d'épurer les eaux usées domestiques des communes de Tipaza et de Nador.

Après la réalisation et la mise en service de la station d'épuration de Hadjout et suite à la réhabilitation de la STEP de Koléa, le projet de la STEP de la ville de Tipaza a été lancé.

Cette réalisation s'inscrit dans le prolongement des nombreuses actions menées par la ville de Tipaza en faveur du développement durable.

Depuis 2008, la station d'épuration de Tipaza est certifiée ISO 14001 /2004.



Figure 44.28: Photo de la STEP par Google Earth

4.2.2 Caractéristiques Techniques

La station fonctionne à faible charge massique de 0.075 kg DBO5/ kg MVS à 70 000 EH (Equivalent Habitant) suivant le processus d'une aération prolongée à boues activées.

Les effluents sont de type urbain et parviennent à cette unité par l'intermédiaire d'un réseau de type unitaire, les eaux sont collectées dans une bache équipée de pompes immergées.

- **Charge hydraulique :**

En termes de charge hydraulique les caractéristiques de la station sont :

- Débit moyen rejeté.....11200 m3/j.
- Débit moyen horaires467 m3/h.
- Débit de pointe au temps sec.....803 m3/h.

- **Charges polluantes :**

En termes de charge polluante à l'entrée et à la sortie de la station, les caractéristiques de l'eau brute et épurée sont représentées dans le tableau ci-dessous.

Tableau 4-3: Paramètres de pollution de conception de la station de Tipaza

	Entrée	Sortie	Abattement (%)
DBO5 (mg/l)	250	30	88
DCO (mg/l)	700	90	87
MES (mg/l)	300	30	90

4.2.3 Description des procédés d'épuration au niveau de la station

La STEP de Tipaza comporte deux chaînes de traitement :

- **Traitement au fil de l'eau :** le prétraitement, le traitement biologique à boues activées, décantation et désinfection
- **Traitement au fil de boues :** épaissement, déshydratation et lit de séchage

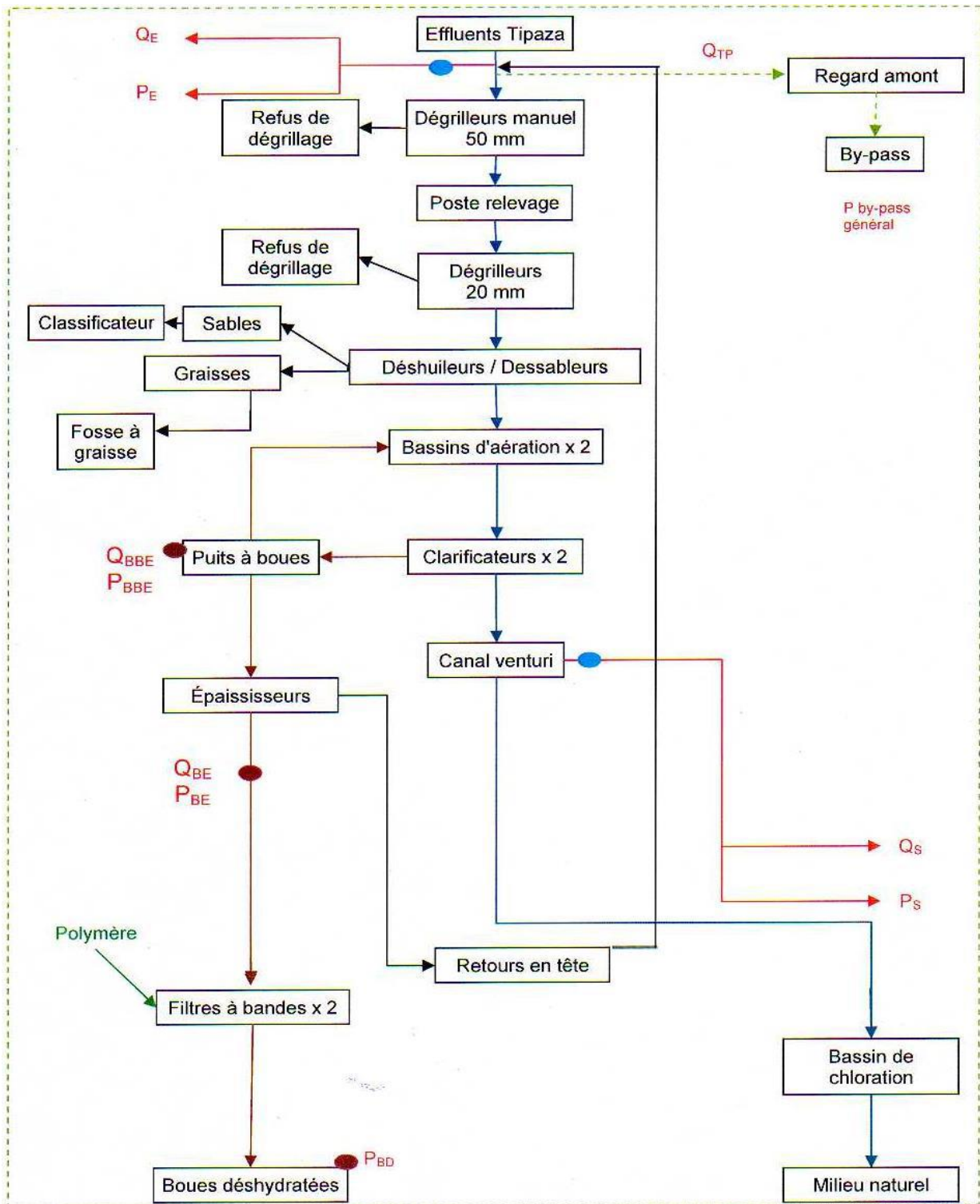


Figure 4. 29 : Organigramme d'exploitation de la STEP de Tipaza.
(MELIANI.Zouhir_SERIKMA.Khaled, 2014).

4.2.4 Statistiques et critiques des données

4.2.4.1 Statistiques Descriptives

La STEP de Tipaza a mis à notre disposition une série de données journalières des paramètres de la pollution à l'entrée et à la sortie, le débit, l'énergie consommée, les boues recirculées, durant le processus d'épuration pour la période de Janvier 2008 à Décembre 2012 avec un certain nombre d'observations NaN pour chaque paramètre comme l'indique le tableau ci-dessous.

Tableau 44-4: Statistiques descriptives des paramètres de pollution de la STEP Tipaza

	Eau Brute				Eau Epurée			
	Min	Max	Moyenne	Ecart-type	Min	Max	Moyenne	Ecart-type
pH	6,19	8,5	7,69	0,20	6,36	8,81	7,79	0,24
T (°C)	9,7	28,5	20,09	4,29	9,6	29,7	20,14	4,65
MES [mg/l]	2,64	2460	320,68	285,36	1	30	9,27	4,66
DB05 [mg/l]	30,32	490	163,07	78,58	0,5	23	5,05	3,80
DCO [mg/l]	71	850	338,06	138,98	4	88,3	33,35	17,60
NH4 [mg/l]	4,5	56,5	23,04	8,74	0,1	9,5	1,25	1,65
NO2 [mg/l]	0,005	14	0,50	1,68	0	4	0,16	0,45
NO3 [mg/l]	0,006	41	1,83	3,20	0,002	24,8	8,19	5,39
PO4 [mg/l]	0,5	32,4	8,39	5,01	0,11	39,7	7,79	6,15
Débit [m3/j]	843	12936	5326,1	2138,4	/	/	/	/
Energie [KWh]	1657	7833	3538,8	1150,14	/	/	/	/
RB [m3/j]	160	9920	4975,14	1451,83	/	/	/	/

4.2.4.2 Evolution des paramètres de pollution dans le temps

Un autre moyen d'avoir une idée générale sur l'ensemble des données recueillies est d'étudier l'évolution des différents paramètres dans le temps, les résultats sont présentés dans les figures suivantes :

4.2.4.2.1 Evolution du PH

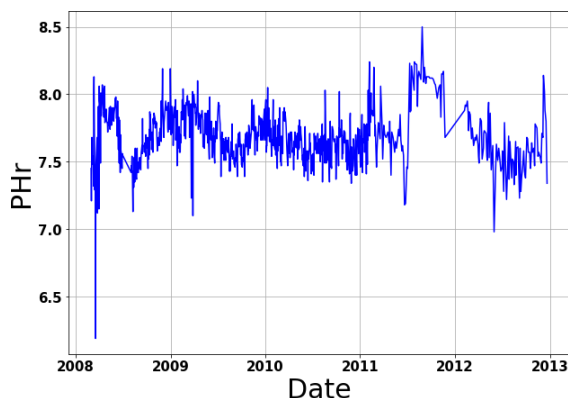


Figure 4-30 : Evolution journalière du Ph de l'eau brute.

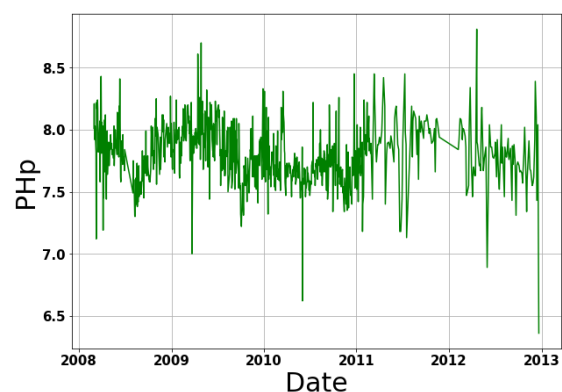


Figure 4-31 : Evolution journalière du Ph de l'eau épurée.

Les deux figures ci-dessus montrent que le PH de l'eau brute ou de l'eau épurée varie entre 6.7 et 8.3 Avec une moyenne respectivement de 7.69 et 7.79

4.2.4.2.2 Evolution de la température

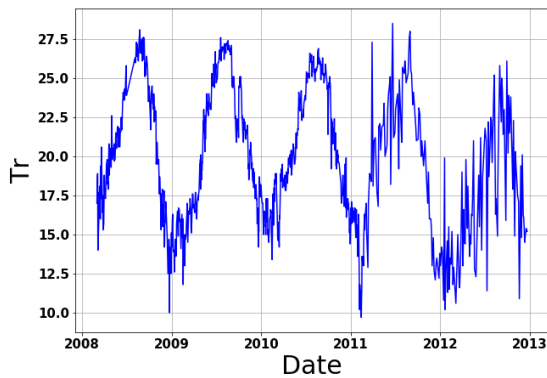


Figure 4-32 : Variation journalière de la température de l'eau brute

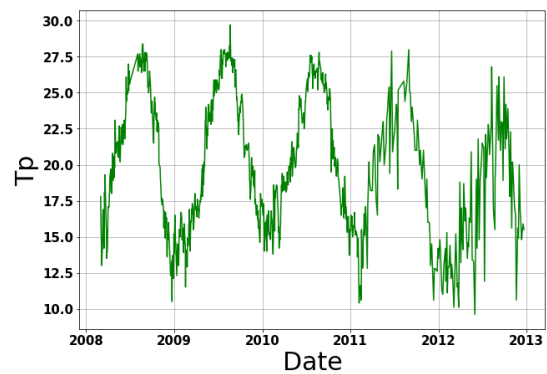


Figure 4-33 : Variation journalière de la température de l'eau épurée.

On remarque que la variation de la température est cyclique à l'entrée comme à la sortie, et oscille entre 11 et 27 °C à l'exception de certains cas où elle atteint un minimum de 9.6°C et un maximum de 29.7 °C.

On déduit que la température de l'eau brute et celle épurée est bien cyclique, celle-ci est tout simplement due aux variations saisonnières.

4.2.4.2.3 Evolution des Matières En Suspension (MES)

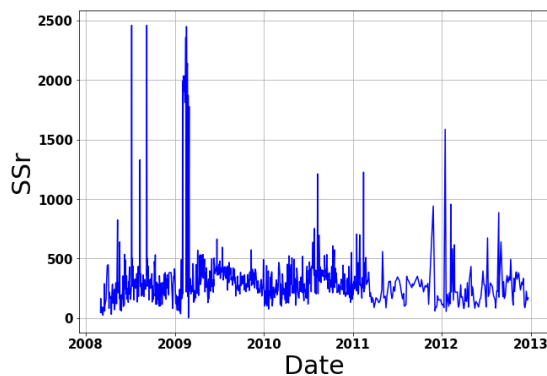


Figure 4-34 : Evolution journalière de la concentration en MES à l'entrée de la STEP

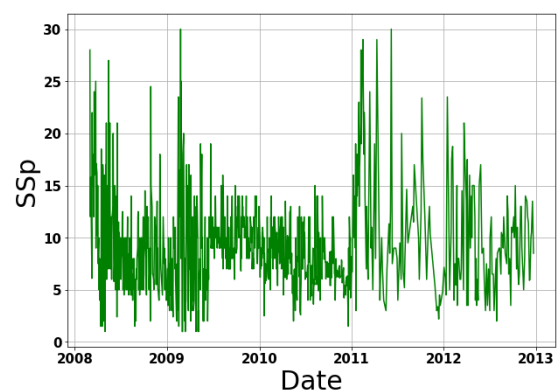


Figure 4-35 : Evolution journalière de la concentration en MES à la sortie de la STEP

La moyenne de la concentration de MES des eaux brutes à l'entrée de la STEP est de 320.68 mg/l. Celle-ci oscillent majoritairement entre 2.64 et 1000 mg/l ce qui dépasse la concentration admissible par la STEP. Certaines valeurs atteignant les 2460 mg/l sont enregistrées et peuvent s'expliquer par des phénomènes exceptionnels, comme l'augmentation du flux de MES par temps de pluie qui provient majoritairement de la remise en suspension des dépôts accumulés.

A la sortie de la STEP, les valeurs enregistrées sont inférieures à la norme de rejet (30mg/l) avec une moyenne de 9.24 mg/l.

4.2.4.2.4 Evolution de la DBO5

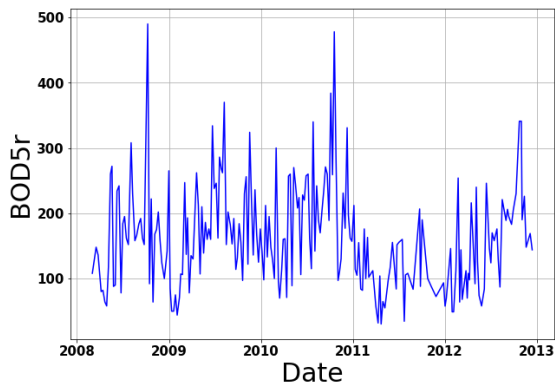


Figure 4.36 : Evolution de la Demande Biologique en Oxygène à l'entrée de la STEP.

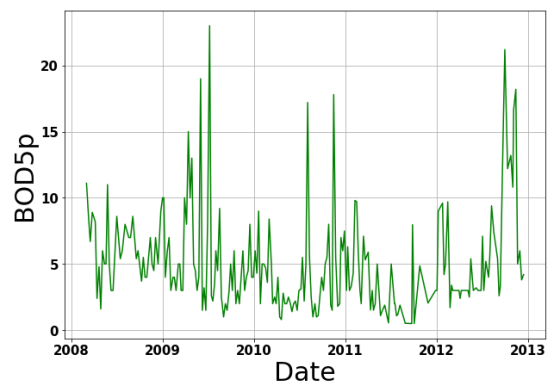


Figure 4.37: Evolution de la Demande Biologique en Oxygène à la sortie de la STEP

La concentration de la demande biologique en oxygène (DBO5) de l'eau brute oscille entre 30.32 et 490 mg/l avec une moyenne de 163.07 mg/l.

La DBO5 de l'eau épurée est inférieure à la norme de rejet, avec une moyenne de 5.05 mg/l

4.2.4.2.5 Evolution de la DCO

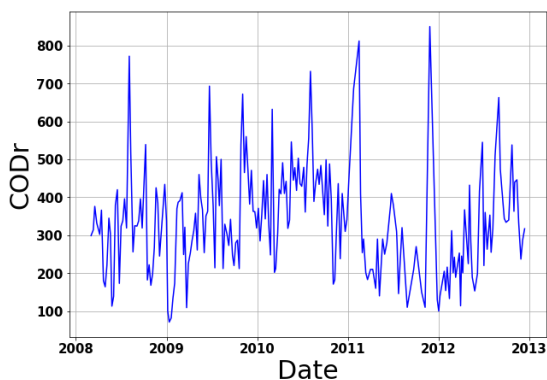


Figure 4.38: Evolution de la Demande Chimique en Oxygène à l'entrée de la STEP.

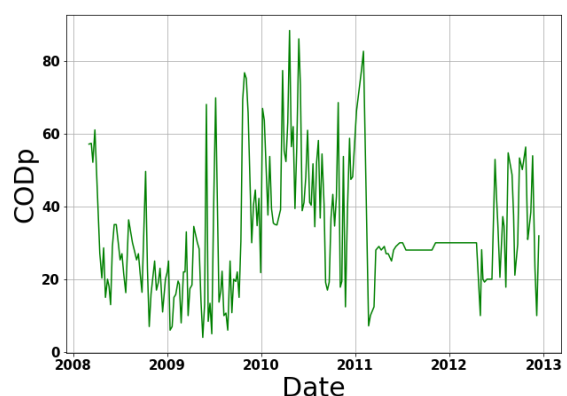


Figure 4.39: Evolution de la Demande Chimique en Oxygène à la sortie de la

La concentration de la Demande Chimique en Oxygène (DCO) de l'eau brute oscille entre 71 et 850 mg/l, avec une moyenne de 485 mg/l avec des dépassements de la concentration admissible par la STEP qui peuvent s'expliquer par les temps de pluie

La DCO de l'eau épurée est inférieure à la norme de rejet qui est de 90 mg/l, avec une moyenne de 33.35mg/l.

4.2.4.2.6 Evolution de l'ammonium (NH4 +)

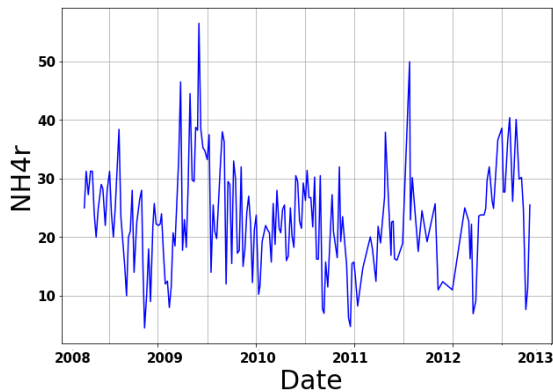


Figure 4-39 : Evolution de l'ammonium (NH4 +) à l'entrée de la STEP.

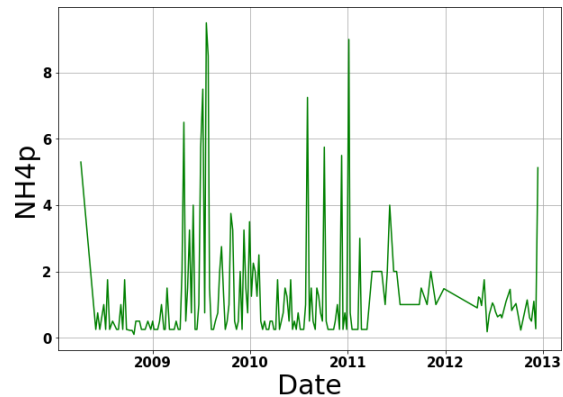


Figure 4-40 : Evolution de l'ammonium (NH4 +) à la sortie de la STEP

Les concentrations d'ammonium Brute oscillent entre 4.5 mg/l et 56.5 mg/l avec une moyenne de 23.04 mg/l, ces variations sont fonction de la température et du temps de séjour dans le réseau de collecte.

A la sortie de la STEP, les concentrations d'ammonium sont variables, allant de 0.1 mg/l à 9.5 mg/l avec une moyenne de 1.63 mg/l.

4.2.4.2.7 Evolution du Nitrite (NO2 -)

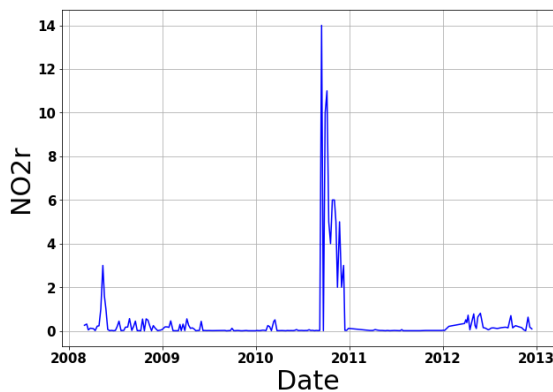


Figure 4.41: Evolution du Nitrite (NO2 -) à l'entrée de la STEP

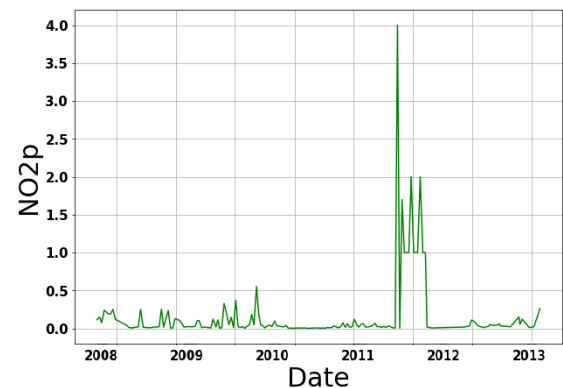


Figure 4.42: Evolution du Nitrite (NO2 -) à la sortie de la STEP

Les concentrations en nitrite brute varient entre 0,005 et 5 mg/l, avec des augmentations brusques en 2010 atteignant 14mg/l qui peuvent être expliqués par la nitrification dans le réseau d'assainissement.

A la sortie de la STEP, les concentrations en nitrites varient de 0 à 2 mg/l, avec un cas exceptionnel enregistré en 2010 atteignant 4mg/l, le Nitrite est en excès par rapport à la concentration de l'entrée, cela peut s'expliquer par sa forme qui est instable durant le processus de nitrification.

4.2.4.2.8 Variation du Nitrate (NO₃⁻)

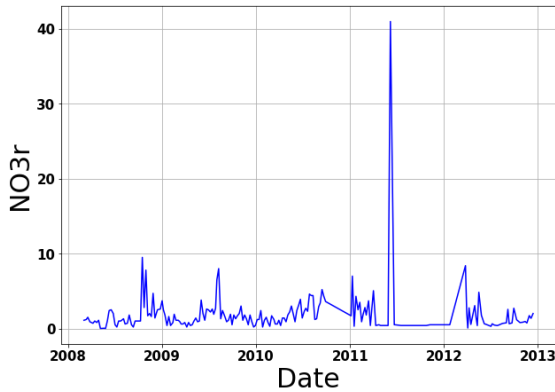


Figure 4.43 : Evolution du Nitrate (NO₃⁻) à l'entrée de la STEP

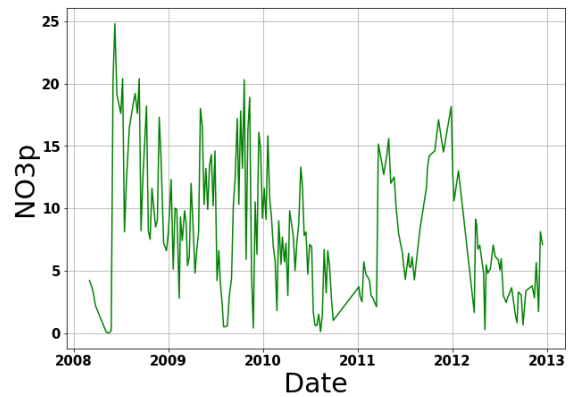


Figure 4.44 : Evolution du Nitrate (NO₃⁻) à la sortie de la STEP

Les concentrations de nitrate brute oscillent entre 0.006 à 9.5 mg/l avec une moyenne de 1.83 mg/l à l'exception d'un cas où elles atteignent un pic de 41 mg/l en 2011. Les concentrations enregistrées à la sortie de la STEP oscillent entre 0.002 et 24.8 mg/l.

4.2.4.2.9 Evolution de l'ortho-phosphate

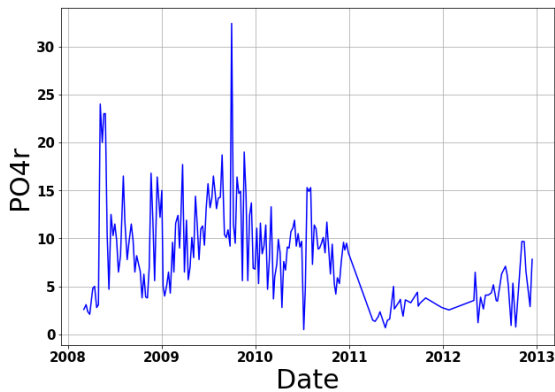


Figure 4.45 : Evolution de PO₄-3 à l'entrée de la STEP

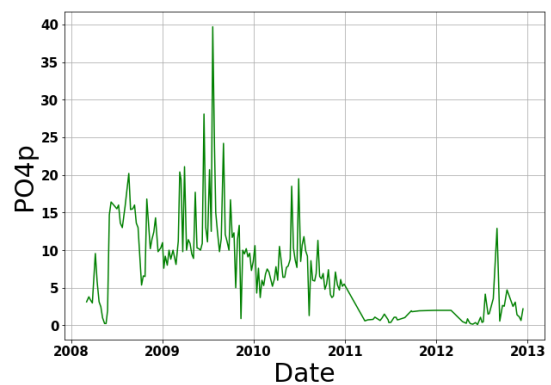


Figure 4.46 : Evolution de PO₄-3 à la sortie de la STEP

Les concentrations de l'ortho-phosphate brute (PO₄-3) oscillent entre 0.5 et 32.4 mg/l, à, avec une moyenne de 8.39 mg/l. A la sortie, les concentrations enregistrées varient généralement entre 0.11 et 39.7 mg/l, avec une moyenne de 7.79 mg/l. Ces valeurs sont inférieures à la norme qui est de 10 mg/l.

4.2.4.2.10 Evolution du Débit entrant

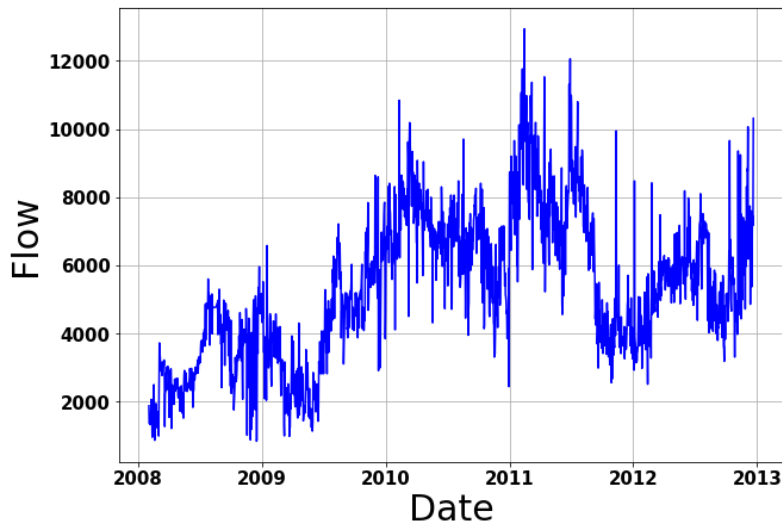


Figure 44.47 : Evolution du débit d'entré journalier de la STEP

Le débit de l'effluent varie entre 843 et 12936 m³ avec une moyenne de 5326.1 m³/j, on remarque certains dépassements de la capacité nominale, cela est dû à des précipitations intenses.

4.2.4.2.11 Evolution de la consommation d'énergie

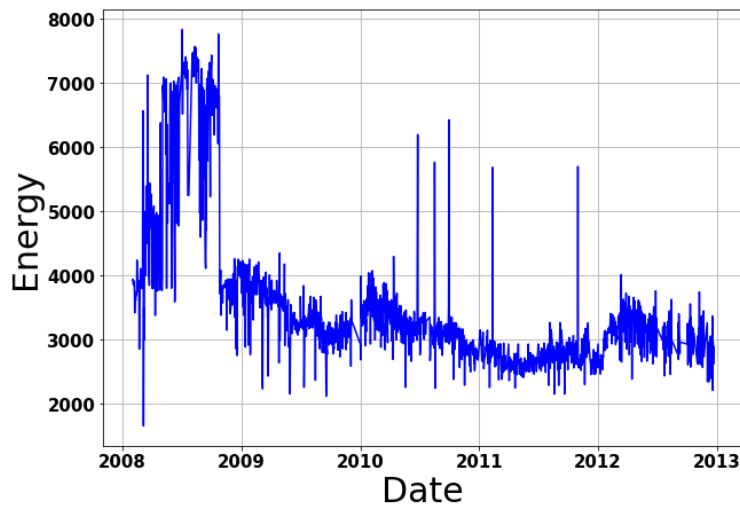


Figure 44.48 : Evolution de l'énergie totale consommée à la STEP

L'énergie totale consommée durant le processus d'épuration des eaux de la STEP de Tipaza oscille entre 1655 et 7833 KWh avec une moyenne de 4330 KWh,

Les valeurs nulles enregistrées sont enlevées car elles sont causées par des coupures d'électricité.

4.2.4.2.12 Evolution du débit de recirculé

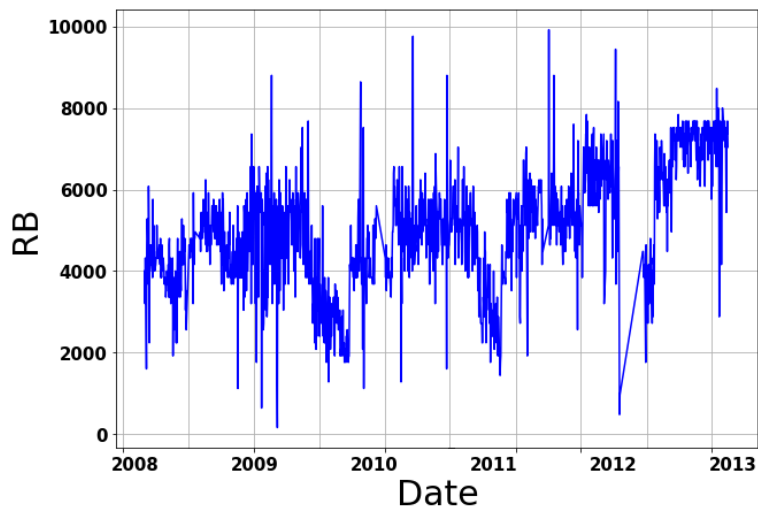


Figure 44.49 : Evolution des Boues de recirculation à la STEP

Les boues de recirculation varient entre 160 et 9920 m³/j avec une moyenne de 4975,14 m³/j.

Conclusion

Dans ce chapitre, nous avons présenté la zone d'étude a été vue, ainsi qu'une discussion sur l'évolution des paramètres mesurés pour les deux stations. La modélisation sera abordée au chapitre suivant et qui visera l'économie de la consommation de l'énergie pour ces STEP.

CHAPITRE 5 : Modélisation de la consommation d'énergie

5. Modélisation de la consommation d'énergie

5.1 Introduction

Les réseaux de neurones profonds (Deep learning) ont prouvé leur efficacité sur tout un ensemble de tâches complexes. L'une d'elles est la régression entre variables.

Dans ce chapitre, le but étant l'optimisation de la consommation d'énergie pour les deux stations d'épuration de Boumerdes et de Tipaza, Pour ce faire, plusieurs filtres ont été adoptés dans l'objectif d'obtenir les meilleures expériences, puis ces dernières seront utilisées dans un modèle réseau de neurones dans le cadre du Deep learning.

Le modèle a été développé par Keras, librairie open source incorporée dans le langage Python 3.6, en utilisant la plateforme de développement Anaconda.

5.2 APPLICATION DE L'ACP

- **Données utilisées**

On dispose d'une série de données brutes des paramètres de pollution en (mg/l) : MES, DBO5, DCO, NH4, NO3-, NO2-, PO4- , NTK de l'eau brute de janvier 2006 à décembre 2016 (*STEP de Boumerdes*), et de mars 2008 à décembre 2012 (*STEP de Tipaza*).

5.2.1 STEP DE BOUMERDES

Une analyse en composantes principales (ACP) a été effectuée sur le tableau des données journalières, qui contient 8 variables et 274 observations de l'eau brute, après avoir éliminé toutes les données non concomitantes.

Les résultats de l'analyse sont présentés ci-dessous :

Tableau 5-1 : Matrice de corrélation entre les CP' et les variables de l'eau brute journalière. (ABDELAZIZ.Soraya, 2016)

Correlation	F1	F2
MESBr	0,736	0,030
DB05Br	0,656	0,059
NTKBr	0,612	0,000
DCOBr	0,562	0,173
NH4Br	0,495	0,155
NO2Br	0,443	0,059
NO3Br	0,094	0,577
PO4Br	0,003	0,124

• Cercle de corrélations

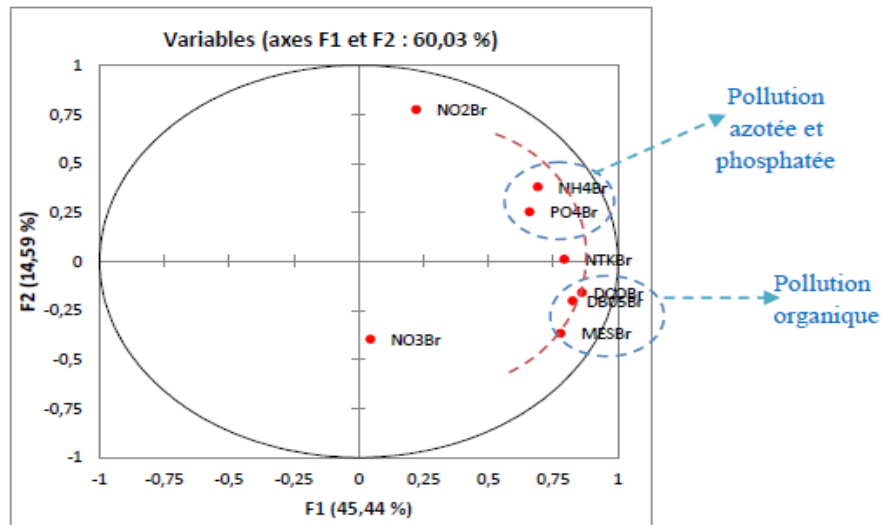


Figure 5.1: Paramètres de pollution de l'eau brute à l'échelle journalière [1×2]. (ABDELAZIZ.Soraya, 2016)

Le cercle de corrélation formé par les axes principaux (1 et 2), montre deux groupes opposés par rapport au premier axe. L'un formé par les composés azotés (NH4Br) et phosphatés (PO4Br) caractérisant une pollution dissoute, l'autre formé par (DCOBr), (DBO5Br), (MESBr) caractérisant une pollution organique.

5.2.2 TIPAZA

Une analyse en composantes principales (ACP) a été effectuée sur le tableau des données journalières, qui contient 7 variables et 161 observations de l'eau brute, après avoir éliminé toutes les données non concomitantes.

Les résultats de l'analyse sont présentés ci-dessous :

Tableau 5-2: Corrélation entre les CP' et les variables (Paramètres de l'eau brute journalière). (MELIANI.Zouhir_SERIKMA.Khaled, 2014)

corrélation	CP'	
	1	2
DBO5Br	0,853	0,117
DCOBr	0,820	0,040
MESBr	0,694	-0,044
NH4Br	0,665	-0,069
PO4Br	0,607	-0,168
NO2Br	0,011	0,734
NO3Br	0,052	0,719

• Cercle de corrélations

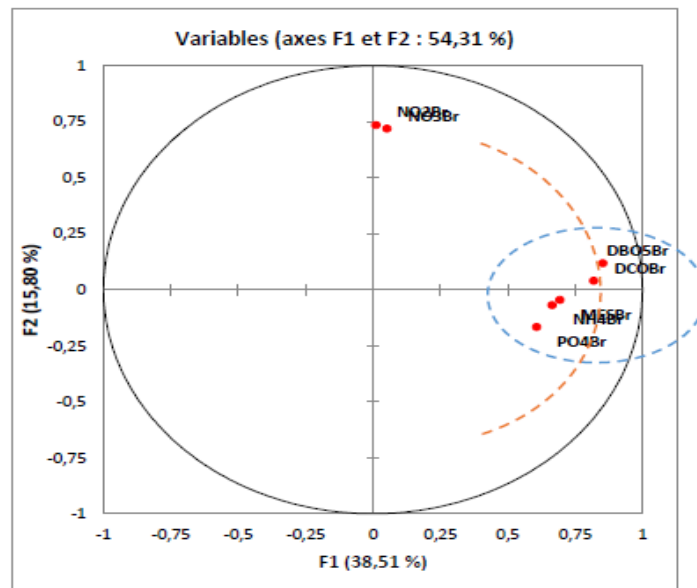


Figure 5.2 : Paramètres de pollution de l'eau brute à l'échelle journalière [1×2].
(MELIANI.Zouhir_SERIKMA.Khaled, 2014)

En examinant le cercle, on remarque que l'axe (1) définit un gradient de pollution décroissant de droite à gauche, et que les paramètres de la pollution organique (DCOBr, DBO5Br) sont disposés sur un arc fictif ayant un gradient de pollution élevé par rapport aux autres paramètres, ceci s'explique par le caractère domestique ou urbain de l'effluent arrivant à la STEP.

5.2.3 DONNEES CUMULEES

Une analyse en composantes principales (ACP) a été effectuée sur le tableau des données journalières, qui contient 7 variables et 492 observations de l'eau brute dont 331 de la STEP de Boumerdes et 161 de la STEP de Tipaza, après avoir éliminé toutes les données non concomitantes.

Les résultats de l'analyse sont présentés ci-dessous :

Tableau 5-3: Corrélations entre les CP' et les variables (Paramètres de l'eau brute journalière).

Correlation	F1	F2
MESBr	0.691	0.123
DBO5Br	0.824	0.102
DCOBr	0.868	0.050
NH4Br	0.689	-0.166
NO2Br	0.072	-0.256
NO3Br	0.082	0.916
PO4Br	0.608	-0.253

• **Cercle de corrélation**

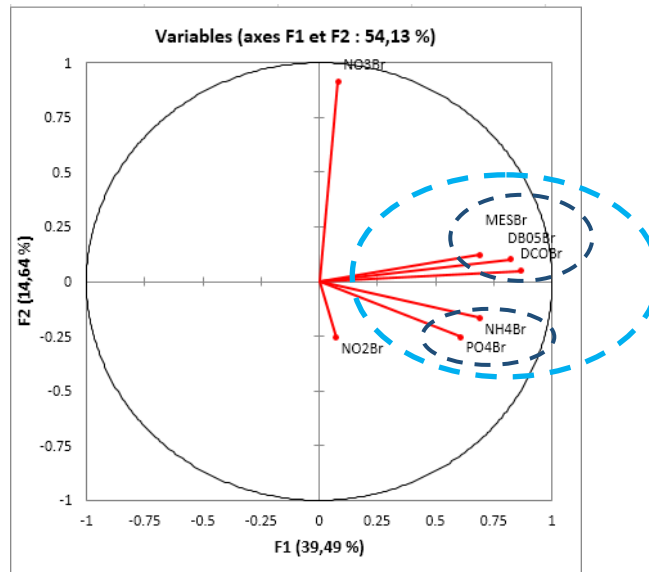


Figure 5.3: Paramètres de pollution de l'eau brute à l'échelle journalière [1×2].

D'après la figure ci-dessus, on remarque la formation de deux groupes opposés par rapport à l'axe F1, le premier constituant la pollution organique (MES, DBO5, DCO) et le deuxième constituant la pollution azotée (NH4+, PO4).

On peut dire qu'à partir de l'analyse des paramètres de l'eau brute au pas de temps journalier, que la pollution de l'eau qui arrive aux deux STEP de Tipaza et de Boumerdes est organique caractérisant un effluent domestique.

Ceci nous amène à conclure que l'utilisation du cumul de données, dans le cadre de notre travail est justifiée. Une confirmation sera faite à travers K-Means dans l'étape ci-dessous.

5.3 Clustering par K-means

Dans cette partie, la compatibilité des données des STEP de Boumerdes et de Tipaza sera discutée par l'utilisation de l'algorithme de regroupement K-means, afin de dégager les relations éventuelles dans les données cumulées de ces deux stations et ceci pour les différentes variables mesurées.

Le k-means est un algorithme de Clustering, en d'autres termes il permet de réaliser des analyses non supervisées, d'identifier un pattern au sein des données et de regrouper les individus ayant des caractéristiques similaires.

L'inconvénient de cette méthode est qu'elle ne permet pas de découvrir le nombre optimal de classes. Le résultat est dépendant du tirage initial et l'algorithme est sensible aux valeurs extrêmes (en particulier lors du calcul du barycentre).

Pour y remédier à ce problème, on adopte la méthode du coude (Elbow Method).

L'idée de cette méthode du coude est d'exécuter k-means Clustering sur l'ensemble de données pour une gamme de valeurs de k (disons k de 2 à 10 dans les exemples ci-dessous), et pour chaque valeur de k calculer la variance intra-classe.

Avant de se lancer dans l'analyse des données des stations d'épuration de Boumerdes et de Tipaza, il faut tout d'abord déterminer le nombre de cluster décrivant ces données.

On applique l'algorithme de la méthode du coude pour notre série de données journalière composé de 317 observations et 17 variables. Pour cela on utilise XLSTAT.

On obtient le graphe suivant :

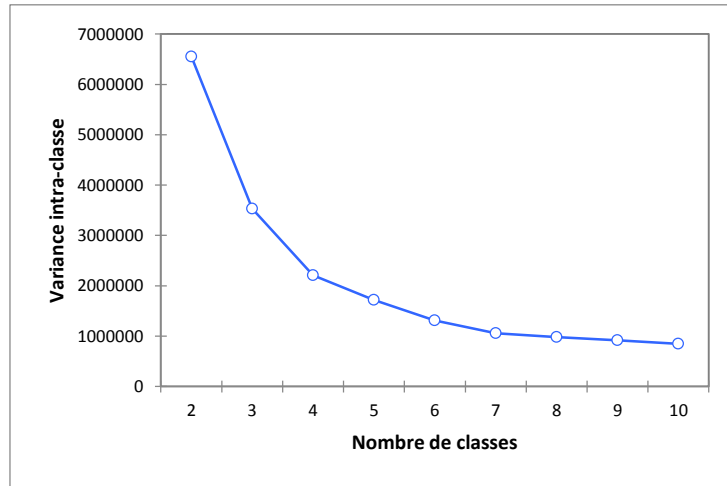


Figure 5.4: Variation de la variance intra-classe en fonction du nombre de classe

Graphiquement, le changement de pente le plus brusque se situe à $k = 4$, et donc on peut déduire que le nombre de classe optimal pour cette étude est de 4 classes.

5.3.1 Regroupement par K-means

On applique ainsi l'algorithme des K-means pour un nombre de classe $k = 4$ classes
Application l'algorithme :

Etape 1 : Initialisation

On tire aléatoirement 4 individus. Ces individus correspondent aux centres initiaux des 4 classes.

Etape 2 :

On calcule la distance entre les individus et chaque centre. Plusieurs métriques existent pour définir la proximité entre 2 individus. La méthode "classique" se base sur la distance euclidienne.

Etape 3 :

On affecte chaque individu au centre le plus proche.

Etape 4 :

On calcule les centres de gravité des groupes qui deviennent les nouveaux centres

Boucle itérative :

On recommence les étapes 2, 3 et 4 tant que les individus sont réaffectés à de nouveaux groupes après une itération.

5.3.2 Résultat du Clustering

Après avoir appliqué l'algorithme des K-means sur les données brutes à savoir MES, DBO5, DCO, NH4+, Débit et l'Energie, les résultats obtenus sont les suivants :

Pour MES

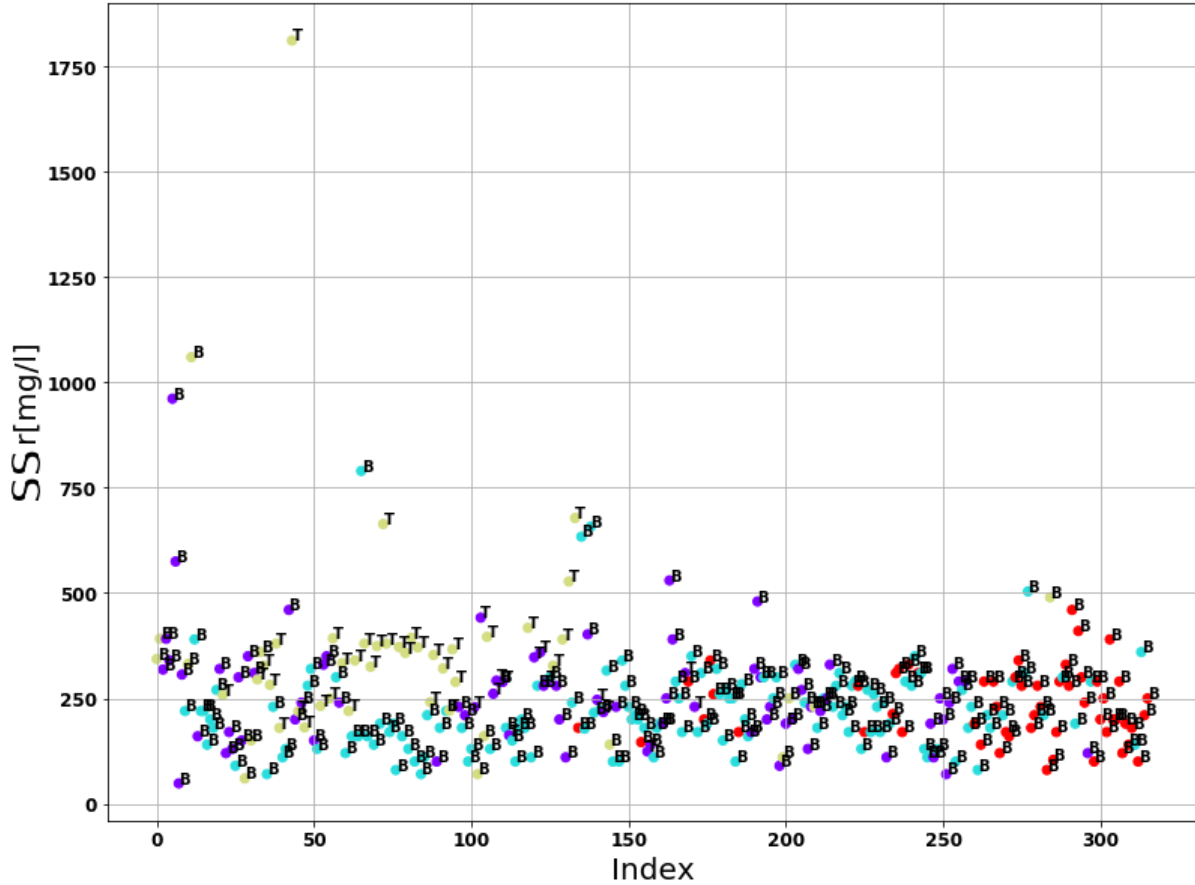


Figure 5.5: Regroupement par K-means de MESBr

Pour DBO5

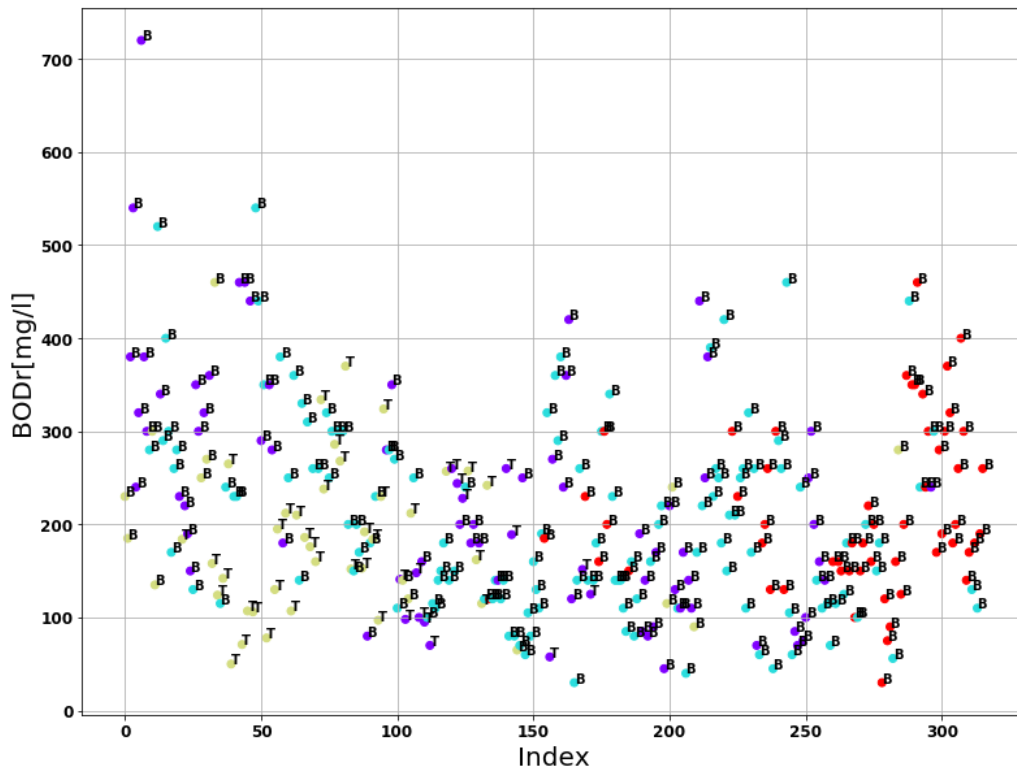


Figure 5.6: Regroupement par K-means de DBO5Br

Pour DCO

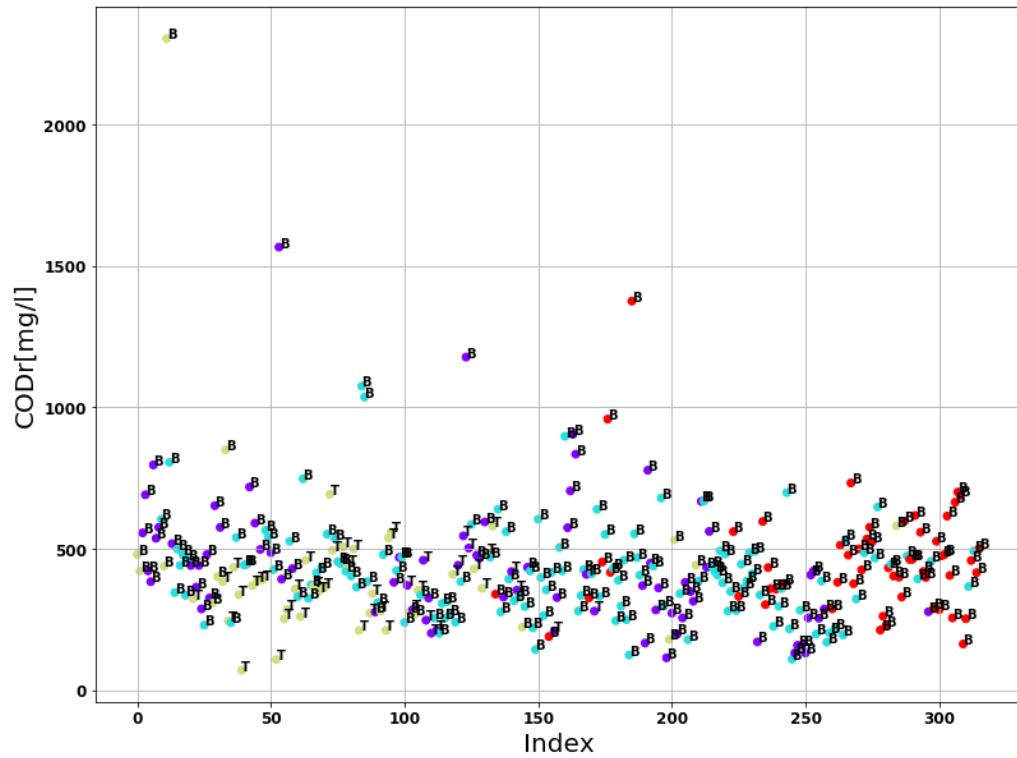


Figure 5.7: Regroupement par K-means de DCOBr

Pour NH4

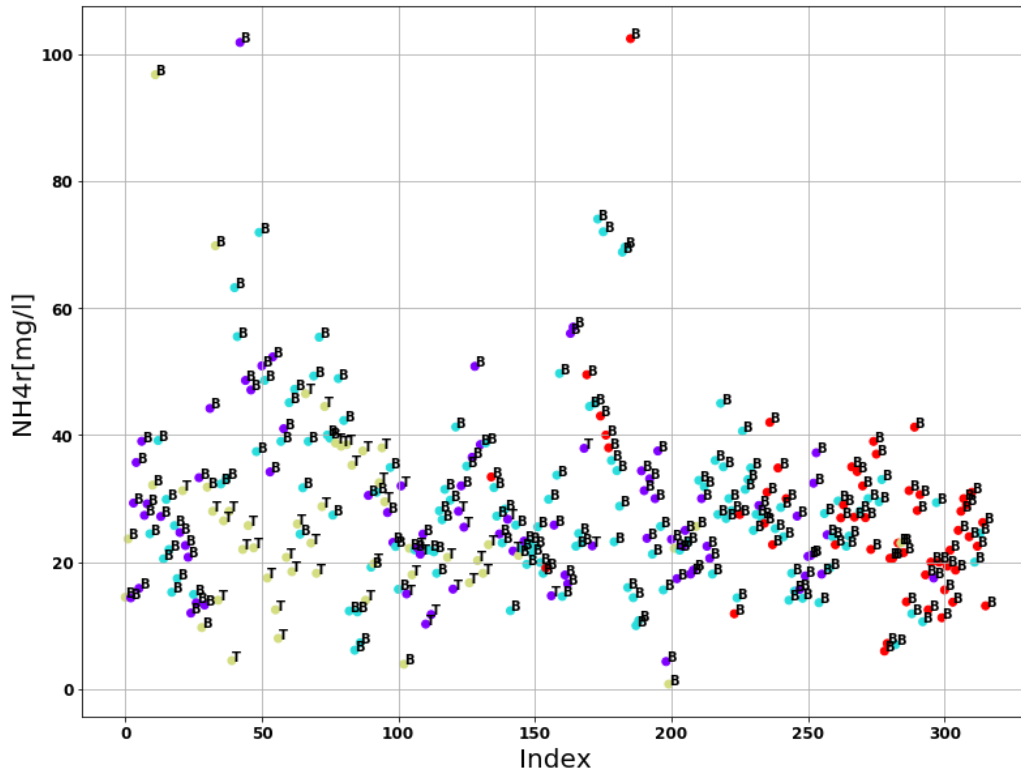


Figure 5.8: Regroupement par K-means de NH4Br

Pour Débit

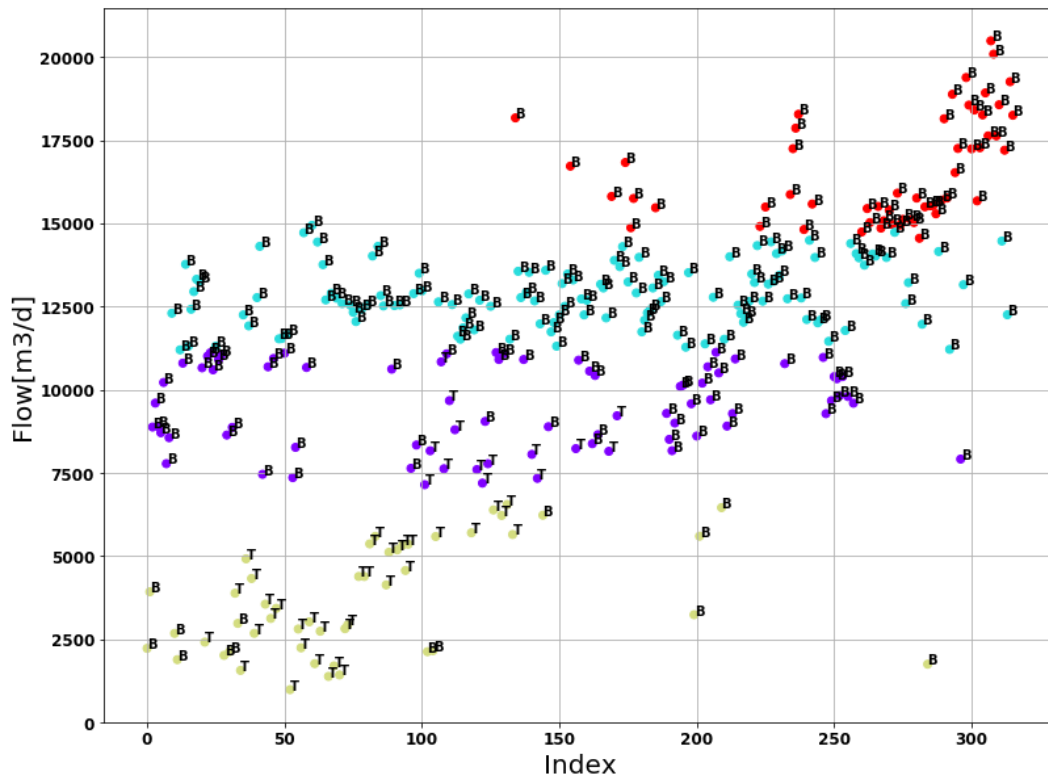


Figure 5.9: Regroupement par K-means de Débit

Pour Energie

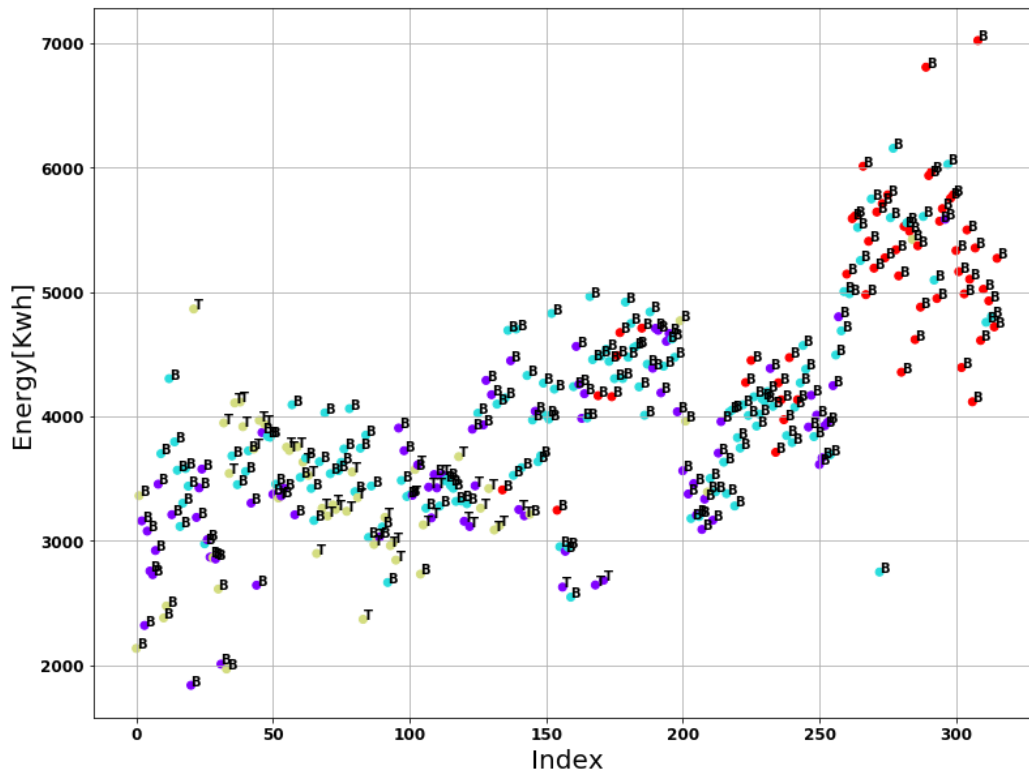


Figure 5.10: Regroupement par K-means de l'énergie

Interprétation des résultats

D'après les regroupements des observations des graphes précédents, on remarque bien que les classes 2 et 3 sont communes pour les deux stations, contenant 14 et 36 observations pour la classe 2 et 65 et 14 observations pour la classe 3, respectivement pour les STEP de Boumerdes et de Tipaza.

Les classes 1 et 4 quant à elles, contiennent respectivement 56 et 131 observations de la STEP de Boumerdes. Les deux ne contenant aucune donnée de la STEP Tipaza.

Le contraste des proportions entre les deux stations est justifié par la différence du nombre d'observations mises à disposition pour cette étude entre les deux STEP.

Conclusion

La combinaison des résultats de l'ACP et des K-Means nous permet l'utilisation du cumul des données des deux STEP, et ceci dans l'optique de la conception d'un modèle d'optimisation de la consommation de l'énergie des deux STEP.

5.4 Elaboration du modèle d'optimisation de la consommation de l'énergie

Le but recherché à travers cette partie sera de concevoir un modèle d'optimisation de la consommation énergétique des stations d'épuration de Boumerdes et de Tipaza. Le modèle sera basé sur l'historique des observations des deux STEP de la période entre 2006 et 2016 pour Boumerdes et de 2008 jusqu'à 2012 pour Tipaza.

Pour cela, deux approches différentes sont utilisées en amont du modèle : la première étant empirique basée sur un filtre comprenant deux sélections, et la deuxième basée sur l'AHP.

L'environnement utilisé pour le développement de ce modèle est la distribution Anaconda utilisant le langage Python, ainsi que les bibliothèques nécessaires (Keras, Pandas ...).

Récapitulatif des modèles

Les modèles utilisés par la suite auront les configurations suivantes :

Tableau 5-4: Tableau récapitulatif des modèles

Model	Input	ANN Architecture	Number of H.L	Number of neurons for each H.L	Total data	Training Data	Test Data
1	Q, ABG, RB, T, WQI	FFBP	4	40	24	16	8
2	Q,ABG,RB, T,WQI	FFBP	7	50	113	75	38
3	Q, ABG, RB, T,WQI	FFBP	4	40	131	86	45
4	Q, ABG,RB, T	FFBP	9	60	266	155	111
5	Q, ABG, RB, T	FFBP	9	60	182	120	62
6	Q, ABG,RB, T	FFBP	9	60	175	116	59
7	Q, ABG, RB, T	FFBP	9	60	167	110	57

Avec:

Q : Débit d'entrée (m³/j)

ABG : Abattement Global

T : Température (°C)

RB : Débit recirculée (m³/j)

WQI: Water Quality Index

FFBP: Feed Forward Back Propagation (propagation en Avant).

H.L: Hidden Layers (Couches cachées).

5.4.1 Première approche

5.4.1.1 Formulation de la problématique

Pour concevoir un modèle performant d'optimisation, il faut connaître les différents paramètres ayant une influence sur la variation de l'énergie électrique. Ce dernier a été établi à partir des expériences jugées bonnes en matière de fonctionnement de la STEP.

L'approche utilisée dans ce modèle est basée sur les Réseaux de neurones multi couches qui a déjà été présentée dans le chapitre III.

Les paramètres qui entrent dans la composition du modèle sont :

A la sortie :

- E : énergie consommée par les aérateurs.

A l'entrée :

- $X1$: Water Quality Index (WQI).
- $X2$: l'abattement de la pollution.
- $X3$: débit recirculé.
- $X4$: Débit entrant dans la STEP.
- $X5$: la température.

Le choix de ces paramètres permet de prendre en considération 4 effets importants :

- La qualité de l'effluent à travers les paramètres de pollution à l'entrée de la STEP.
- L'effet de la quantité à travers le débit entrant et du débit recirculé.
- L'effet du climat à travers la température.
- Le respect de la norme de rejet à travers l'abattement de la pollution.

5.4.1.1.1 Indice de Pollution :

Si exprimer la quantité d'une eau est évident (donner son volume en m³) exprimer sa qualité est plus complexe et subjectif.

C'est là qu'intervient l'idée d'indice de qualité. Celui-ci permettra de donner une valeur à la qualité générale de l'eau selon son utilisation et donc de comparer des eaux de qualités différentes.

L'indice de qualité de l'eau calculé à partir de différents paramètres caractérisant l'effluent de la STEP vis-à-vis d'une certaine utilisation (lavage industriel, baignade, eau potable...), permet de comparer et de hiérarchiser des eaux différentes en vue de leur utilisation et ceci à travers un seul indice.

Nous avons recherché un indice de pollution pour pouvoir modéliser la consommation en énergie électrique des aérateurs et nous avons pris en considération l'indice de qualité.(Abassi, 2012).

5.4.1.1.1 Les étapes d'élaboration d'un indice de pollution

Les quatre étapes suivantes sont le plus souvent associées à l'élaboration de tout indice de pollution :

- a. Sélection des paramètres.
- b. Transformation des paramètres des différentes unités et dimensions à une échelle commune.
- c. Attribution d'un poids à chaque paramètre.
- d. Agrégation des sous-indices pour obtenir un score d'indice final.

Application de ces étapes pour le cas étudié :

a) La sélection des paramètres

On va garder les mêmes paramètres sélectionnés dans l'indice précédent : DBO5, DCO, MES

b) Transformation des paramètres des différentes unités et dimensions à une échelle commune :

Dans notre cas, les paramètres choisis (DBO, DCO, MES) ont déjà la même unité (mg/l).

c) Attribuer un poids à chaque paramètre

Après avoir sélectionné les paramètres qu'on doit utiliser dans le calcul de l'indice de pollution, on cherche à les pondérer selon leur degré d'importance.

L'étape principale pour la détermination du poids de chaque paramètre est de faire un classement par apport à leurs degrés d'importance. Dans la majorité des études effectuées sur les différents indices, le classement est fait par des experts dans les domaines étudiés.

N'ayant pas cette compétence, nous nous sommes tournés vers l'ACP (Analyse en Composantes Principales), méthode statistique qui permet de hiérarchiser les paramètres et de chiffrer leur importance les uns par rapport aux autres.

Application de la méthode d'ACP

Afin d'obtenir un classement des quatre paramètres sélectionnés, on a fait appel à l'ACP qui nous permettra de mesurer l'importance de chaque paramètre dans le phénomène étudié.

L'application de cette méthode sur les données de la STEP de Boumerdes, permet d'obtenir le degré de participation de chaque paramètre, et qui se traduit par la quantité d'information extraite de chaque donnée après l'application de l'ACP.

Vu le grand nombre de données à manipuler, nous avons utilisé le logiciel « XLSTAT ».

Le système de notation (classement) prend en considération la corrélation des paramètres avec les composantes principales "*Fi*". Les paramètres de pollution qui sont corrélés avec la première composante seront mieux classés que les paramètres corrélés avec la deuxième composante. Les paramètres corrélés à la même composante seront classés entre eux de la même manière.

Le classement et le poids attribué à chaque paramètre de pollution est comme suit :

Tableau 5-5: Classement et Poids attribués à chaque paramètre sélectionné pour la STEP de Boumerdes

variable	le classement	poids temporaire	poids totale
DCO _b	1	1	0,54
MES _b	2	0.5	0,27
DBO5 _b	3	0,33	0,19

Tableau 5-6: Classement et Poids attribués à chaque paramètre sélectionné pour la STEP de Tipaza

variable	le classement	poids temporaire	poids totale
DCO _b	1	1	0,54
DBO5 _b	2	0.5	0,27
MES _b	3	0,33	0,19

Tableau 5-7: Classement et Poids attribués à chaque paramètre sélectionné pour les deux STEP

variable	le classement	poids temporaire	poids totale
DCO _b	1	1	0,54
DBO5 _b	2	0.5	0,27
MES _b	3	0,33	0,19

d) **Agrégation des sous-indices pour obtenir un score d'indice final**

$$WQI = \prod_{i=1}^n (I_i)^{W_i}$$

$$\sum_i^n W_i = 1$$

W_i : Le poids associé à chaque paramètre sélectionné.

I_i : Le sous-indice de qualité associé à chaque paramètre sélectionné.

La détermination des sous-indices de pollution

Les sous-indices sont déterminés de manière linéaire :

$$I_{DBO5} = 1 - \left(\frac{DBO5_{br} - 30}{1000} \right)$$

$$I_{DCO} = 1 - \left(\frac{DCO_{br} - 90}{1100} \right)$$

$$I_{MES} = 1 - \left(\frac{MES_{br} - 30}{1000} \right)$$

$I_i=1$: tous les paramètres polluants sont inférieurs aux normes de limite de pollution en vigueur.

$I_i=0$: l'eau est très polluée

Elaboration de l'indice final

Après avoir déterminé les sous-indices, on détermine le WQI (Water Quality Indice) via la formule suivante :

$$WQI = \prod_{i=1}^n (I_i)^{W_i}$$

$$WQI = I_{DBO5}^{W1} * I_{DCO}^{W2} * I_{MES}^{W3}$$

WQI=1 : la pollution est inférieure à la norme (eau presque potable).

WQI=0 : l'eau est très polluée.

Pour passer d'un indice de qualité à un indice de pollution on utilise la relation suivante :

$$IP = 1 - WQI.$$

Ip=1 : l'eau est très polluée.

Ip=0 : la pollution est inférieure à la norme (eau presque potable).

5.4.1.1.2 L'abattement de la pollution (AB)

L'abattement de la pollution (AB) de l'effluent représente la charge de pollution éliminée par le processus d'épuration.

$$AB_{DBO5} = \frac{\text{charge } DBO5_b - \text{charge } DBO5_e}{\text{charge nominale } DBO5}$$

$$AB_{DCO} = \frac{\text{charge } DCO_b - \text{charge } DCO_e}{\text{charge nominale } DCO}$$

$$AB_{MES} = \frac{\text{charge } MES_b - \text{charge } MES_e}{\text{charge nominale } MES}$$

Nb : l'indice 'b' concerne les données brutes et l'indice 'r' concerne les données épurées.

Afin de n'avoir qu'une seule valeur nous avons pris la norme du vecteur qui englobe ces trois équations :

$$AB_{totale} = \sqrt{\frac{AB_{DBO5}^2 + AB_{DCO}^2 + AB_{MES}^2}{3}}$$

5.4.1.1.3 Le débit recirculé (DR)

Pour assurer un traitement biologique efficace, il faut maintenir une concentration constante (Voisine de 5g/l) dans le bassin d'aération. Pour cela, il est nécessaire de recirculer une partie des boues qui sont extraites du décanteur secondaire vers l'entrée du bassin d'aération (boues recirculées), donc c'est un paramètre essentiel dans l'optimisation de l'aération.

5.4.1.1.4 Le débit entrant à la STEP (D)

La prise en considération du débit entrant dans les deux STEP comme paramètre du modèle d'optimisation est justifié par la très grande fluctuation de ce paramètre durant la période d'étude.

5.4.1.1.5 La température (T)

En effet, le temps d'aération est fonction de la température.

La diminution de l'activité métabolique des micro-organismes diminue avec la baisse de la température du système, ainsi une fluctuation du temps d'aération du bassin biologique existe du fait de l'effet cyclique de la température.

Mais les questions de comment déterminer ces bonnes expériences ? Et quel est le critère de ce jugement ? Restent posées.

5.4.1.2 Rendements épuratoires

On définit les rendements comme suit :

$$RMES: \text{Rendement épuratoire } MES = 1 - \left(\frac{MESe}{MESbr} \right)$$

$$RDBO5: \text{Rendement épuratoire } DBO5 = 1 - \left(\frac{DBO5e}{DBO5br} \right)$$

$$RDICO: \text{Rendement épuratoire } DCO = 1 - \left(\frac{DCOe}{DCObr} \right)$$

$$RNMES: \text{Rendement } MES \text{ lié à la norme} = 1 - \left(\frac{30}{MESbr} \right)$$

$$RNDBO5: \text{Rendement } DBO5 \text{ lié à la norme} = 1 - \left(\frac{30}{DBO5br} \right)$$

$$RNDCO : \text{Rendement } DCO \text{ lié à la norme} = 1 - \left(\frac{9030}{DCO5br} \right)$$

On définit aussi :

$$RG : \text{rendement global épuratoire} = \text{le module du vecteur : } \begin{cases} -RDBO5 \\ -RDICO \\ -RMES \end{cases}$$

$$RNG : \text{rendement global lié à la norme} = \text{le module du vecteur : } \begin{cases} -RNDBO5 \\ -RNDCO \\ -RNMES \end{cases}$$

5.4.1.3 La sélection des données

Après avoir classé toutes les expériences des deux stations, on obtient 3 matrices à partir desquelles une sélection d'expériences jugées bonnes a été faite, et ceci en passant par deux étapes :

5.4.1.3.1 Première Sélection :

Cette sélection est basée sur la détection des expériences où le rendement épuratoire de la station est proche du rendement optimal, autrement dit on choisit un intervalle, et on sélectionne toutes les expériences qui se trouvent à l'intérieur.

5.4.1.3.2 Deuxième Sélection :

Cette 2^{ème} sélection est basée sur l'énergie consommée par les aérateurs, on détecte les journées où on a le même indice de pollution et on choisit la ligne qui correspond à la consommation énergétique la plus faible.

5.4.1.4 Résultats

Afin de rendre l'étude basée sur les données des deux STEP de Boumerdes et de Tipaza la plus homogène possible, étant donné que le débit entrant et l'énergie consommée pour la station de Boumerdes sont supérieurs aux débits entrant et l'énergie consommée par la station de Tipaza, on considèrera le rapport E/Q couplé au WQI comme premiers critères de sélection des meilleures expériences par l'utilisation de l'algorithme des K-means.

Une sélection des meilleures classes selon le rapport E/Q et le WQI sera ensuite prise en compte pour la suite de la procédure de cette approche de modélisation.

5.4.1.4.1 STEP de Tipaza

- **K-means**

Avant d'appliquer l'algorithme des K-Means sur la série de données de la station de Tipaza, les valeurs non numériques (les valeurs NaN) ont été supprimées.

La série de données obtenue contient 163 observations. Pour chacune de ces observations, on a calculé le WQI par les formules présentées précédemment, ainsi que le rapport de l'énergie sur le débit E_Q [KW/m³].

Le résultat des K-Means est présenté dans la figure suivante :

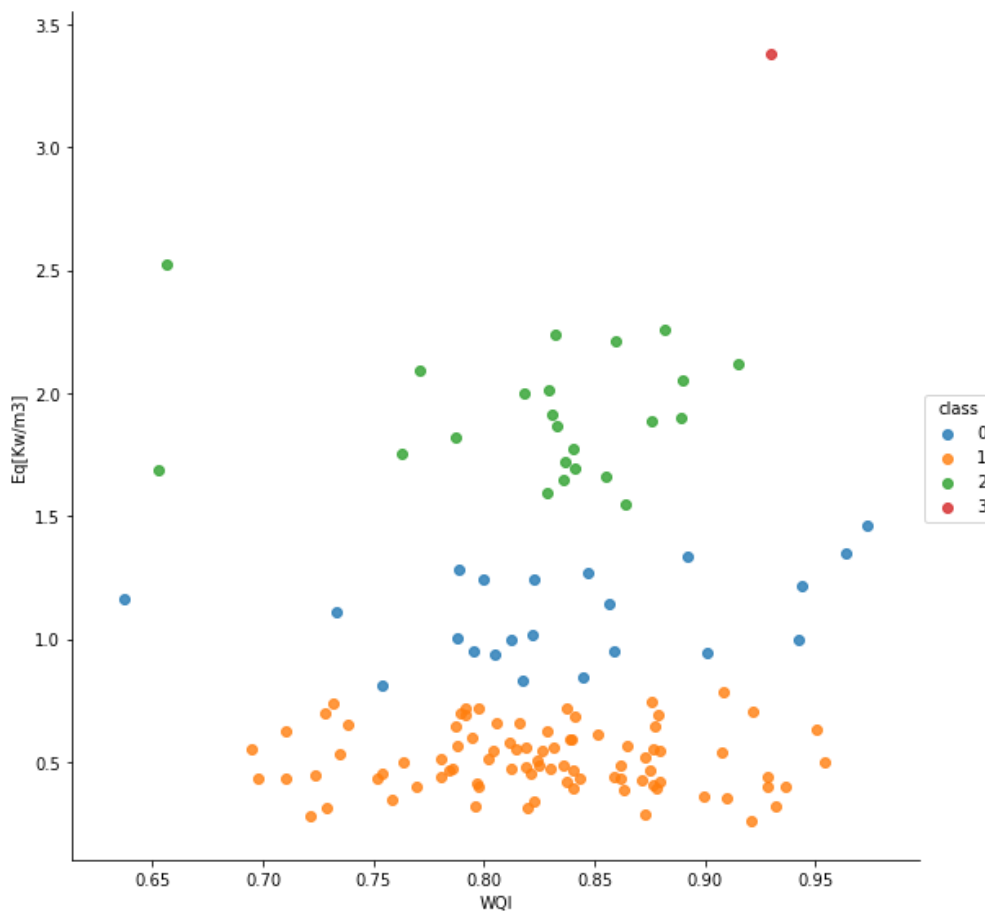


Figure 5.11: Regroupement par K-means sur les données de Tipaza

D'après cette figure, on remarque bien que le regroupement se fait en quatre classes, qui sont regroupées autour des centres de classes suivants :

- **Class 0** : [0.79201548 , 1.11979926]
- **Class 1** : [0.79436568 , 1.91941423]
- **Class 2** : [0.85979943 , 3.59189271]
- **Class 3** : [0.78463057 , 0.52189254]

Pour la suite de l'étude, on a considéré les deux premières classes (le seuil de la sélection étant $E_Q = 1.4$)

Le nombre d'observations qu'on a obtenu après la sélection par K-Means est de 128 observations.

Ces dernières observations seront ensuite introduites dans la première sélection du filtre cité plus haut.

- **Première Sélection :**

La première sélection du filtre a été appliquée sur les observations obtenues, les résultats sont les suivants :

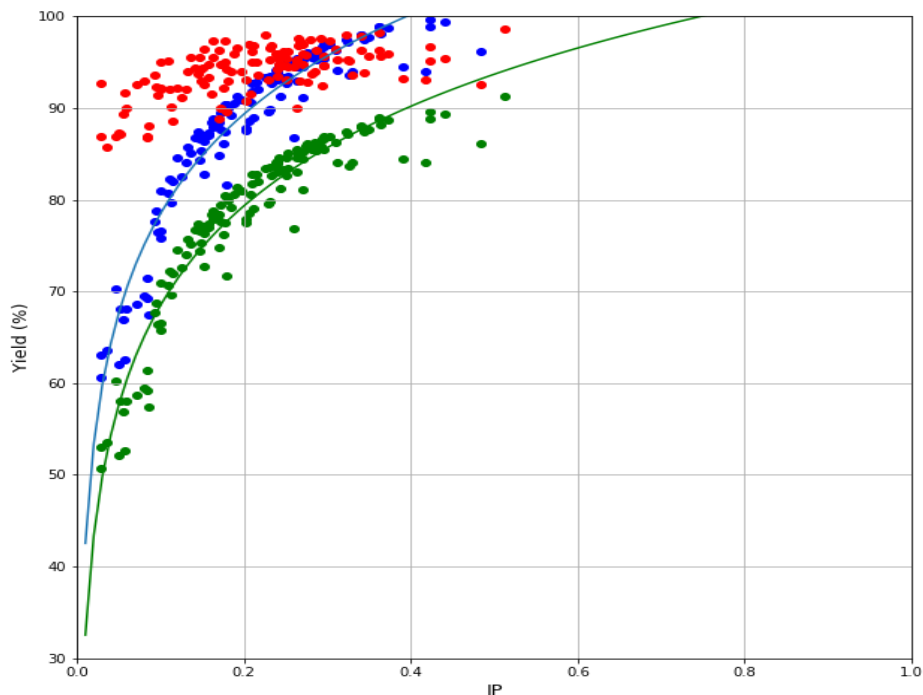


Figure 5.12: Rendement Globale en fonction de l'indice de pollution

La figure ci-dessus représente le rendement global (Rouge) de 2008 à 2012, les deux courbes en (Vert) et en (Blue) représentant respectivement le rendement global lié à la norme et le rendement global lié à la norme auquel on a ajouté une marge de (10%), en fonction de l'indice de pollution.

Les points rouges situés entre les deux courbes, au nombre de 26 valeurs, jugés pertinents, seront sélectionnés pour la deuxième sélection du filtre.

- **Deuxième Sélection :**

On obtient, après application de la deuxième sélection du filtre, un minimum de 24 valeurs, qui seront utilisées par la suite pour l'élaboration de notre modèle d'optimisation.

• **Modélisation**

A partir des résultats obtenus de la deuxième sélection, On procède à la conception du modèle de régression entre les paramètres d'entrée et sortie. Après plusieurs essais (variation sur le nombre de couches, nombre de neurones, types de fonction d'erreur et d'activation...), Le modèle ANN le plus performant et adapté pour ce problème est composé d'une couche d'entrée, 4 couches cachées de 40 neurones chacune et une couche de sortie.

La série de données a été décomposée en deux groupes, le premier contenant 66% est destiné à l'apprentissage du modèle, et 33% pour la validation des résultats du modèle.

Les résultats obtenus sont les suivants :

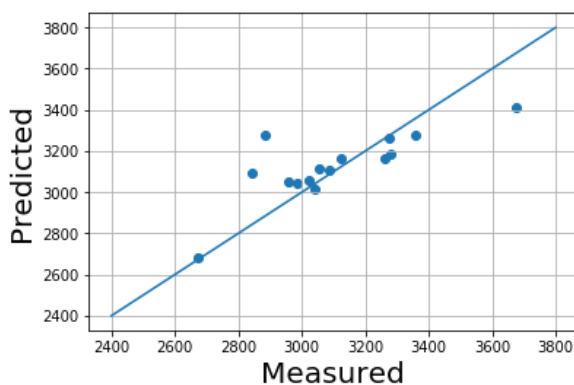


Figure 5.13 : Calibration

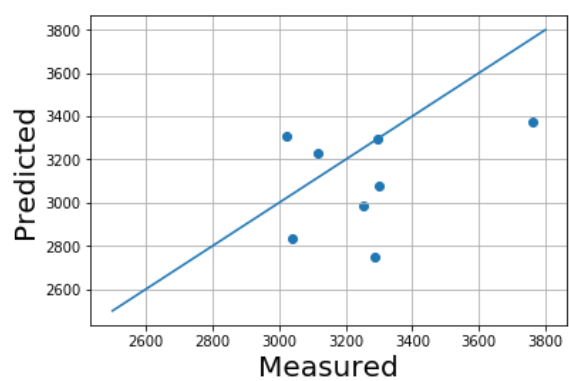


Figure 5.14 : Validation

	R2_Score	RMSE (KWh)
Calibration	0.686	576.27
Validation	0.349	794.05

Comparaison de l'énergie réelle et simulée pour la validation

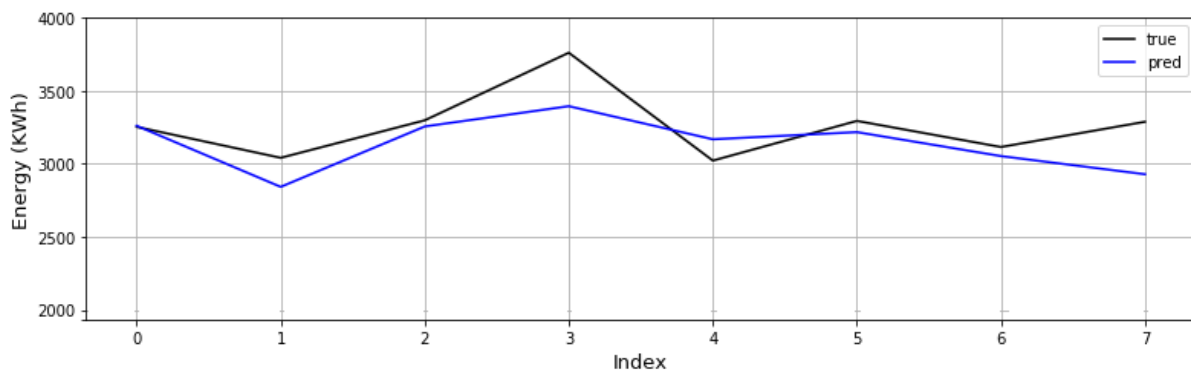


Figure 5.15: Comparaison de l'énergie réelle et simulée pour la STEP de Tipaza

Les résultats obtenus pour la STEP de Tipaza démontrent bien que le modèle est de mauvaise qualité à première vue. Ceci est justifié par le nombre très limitées d'observations surtout après l'application des deux filtres.

5.4.1.4.2 STEP de Boumerdes

- **K-Means**

Avant d'appliquer l'algorithme des K-Means sur la série de données de la station de Boumerdes, les valeurs non numériques (les valeurs NaN) ont été supprimées.

La série de données obtenue contient 399 observations. Pour chacune de ces observations, on a calculé le WQI par les formules présentées précédemment, ainsi que le rapport de l'énergie sur le débit E_Q .

Le résultat des K-Means est présenté dans la figure suivante :

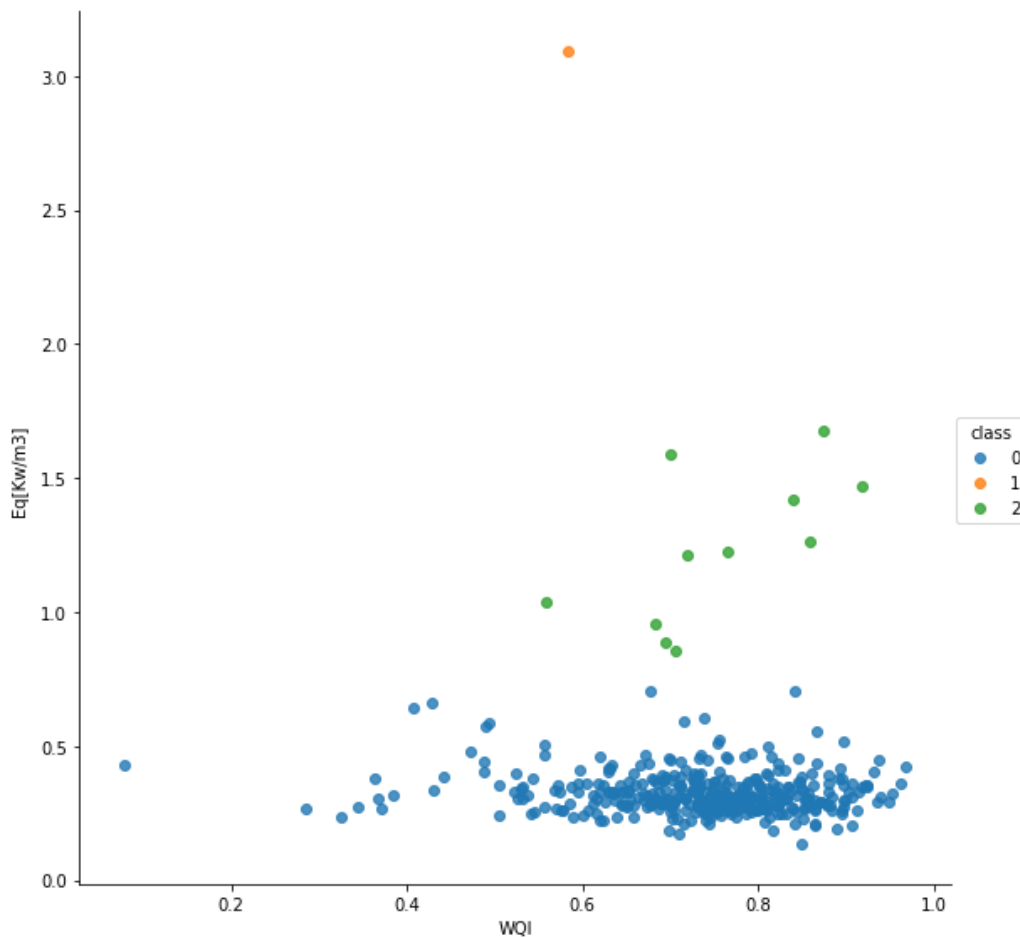


Figure 5.16: Regroupement par K-Means pour les données de Boumerdes

D'après cette figure, on remarque bien que le regroupement se fait en trois classes, qui sont regroupées autour des centres de classes suivants :

- **Class 0** : [0.7143872 ,0.32693394]
- **Class 1** : [0.8174328 ,1.23653446]
- **Class 2** : [0.5824137 ,3.09210151]

Pour la suite de l'étude, on a considéré le même seuil pour les deux STEP ($E_Q = 1.4$)

Le nombre d'observations qu'on a obtenu après la sélection par K-Means est de 395 observations.

Ces dernières observations seront ensuite introduites dans la première sélection du filtre cité plus haut.

- **Première Sélection :**

La première sélection du filtre a été appliquée sur les observations obtenues, les résultats sont les suivants :

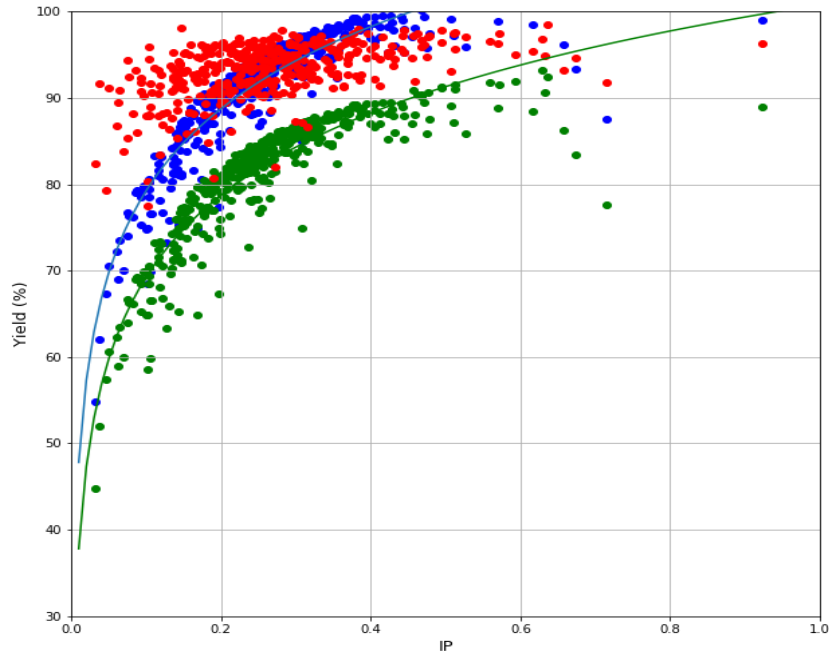


Figure 5.17 : Rendement Global en fonction de l'indice de pollution

La figure ci-dessus représente le rendement global (Rouge) de 2006 à 2016, les deux courbes en (Vert) et en (Blue) représentant respectivement le rendement global lié à la norme et le rendement global lié à la norme auquel on a ajouté une marge de (10%), en fonction de l'indice de pollution.

Les points rouges situés entre les deux courbes, au nombre de 140 valeurs, jugés pertinents, seront sélectionnés pour la deuxième sélection du filtre.

- **Deuxième Sélection :**

On obtient, après application de la deuxième sélection du filtre énergie minimum 113 valeurs, qui seront utilisées par la suite pour l'élaboration de notre modèle d'optimisation.

- **Modélisation**

A partir des résultats obtenus de la deuxième sélection, On procède à la conception du modèle de régression entre les paramètres d'entrée et sortie. Après plusieurs essais (variation sur le nombre de couches, nombre de neurones, types de fonction d'erreur et d'activation...), Le modèle ANN le plus performant et adapté pour ce problème est composé d'une couche d'entrée, 7 couches cachées de 50 neurones chacune et une couche de sortie.

La série de données a été décomposé en deux groupes, le premier contenant 66% est destiné à l'apprentissage du modèle, et 33% pour la validation des résultats du modèle.

Les résultats obtenus sont les suivants :

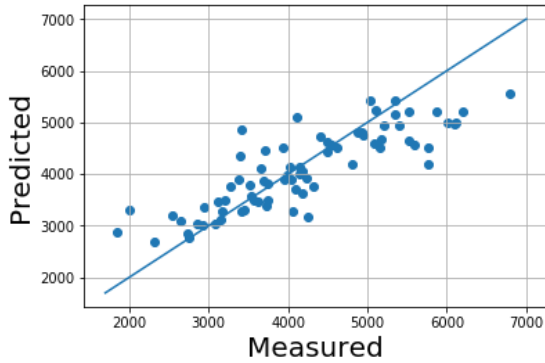


Figure 5-18 : Calibration

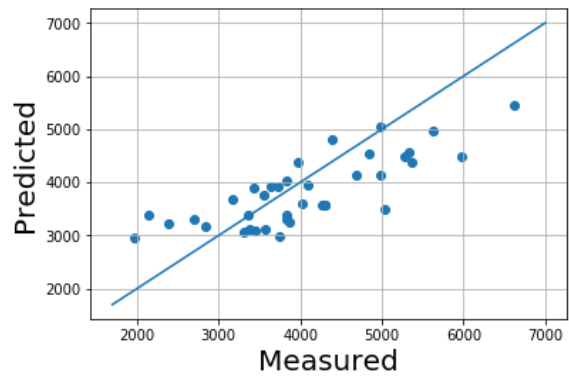


Figure 5.19 : Validation

	R2_Score	RMSE (KWh)
Calibration	0.728	534.15
Validation	0.506	713.12

Comparaison de l'énergie réelle et simulée pour la validation

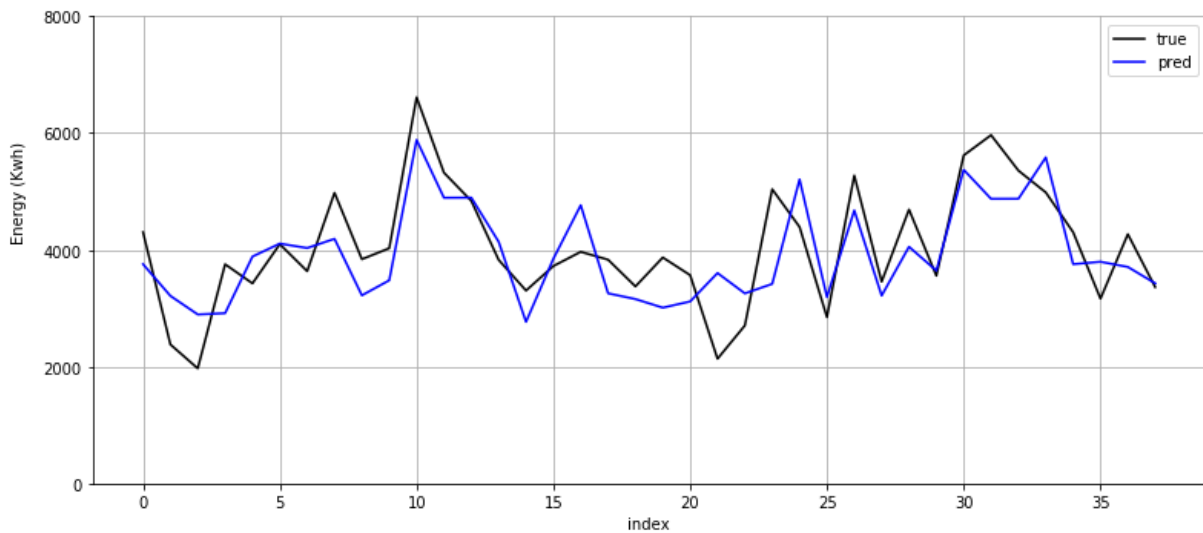


Figure 5.20: Comparaison de l'énergie réelle et simulée pour la STEP de Boumerdes

Les résultats obtenus pour la STEP de Boumerdes démontre bien que le modèle est de qualité moyenne. de meilleurs résultats pourraient être obtenus si les échantillons d'apprentissage et de validation étaient plus grands.

5.4.1.4.3 Données cumulées des STEP de Boumerdes et Tipaza

- **K-means**

Avant d'appliquer l'algorithme des K-Means sur la série des données cumulée, les valeurs non numériques (les valeurs NaN) ont été supprimées.

La série de données obtenue contient 562 observations dont 399 sont de Boumerdes et 163 de Tipaza. Pour chacune de ces observations, on a calculé le WQI par les formules présentées précédemment, ainsi que le rapport de l'énergie sur le débit E_Q .

Le résultat des K-Means est présenté dans la figure suivante :

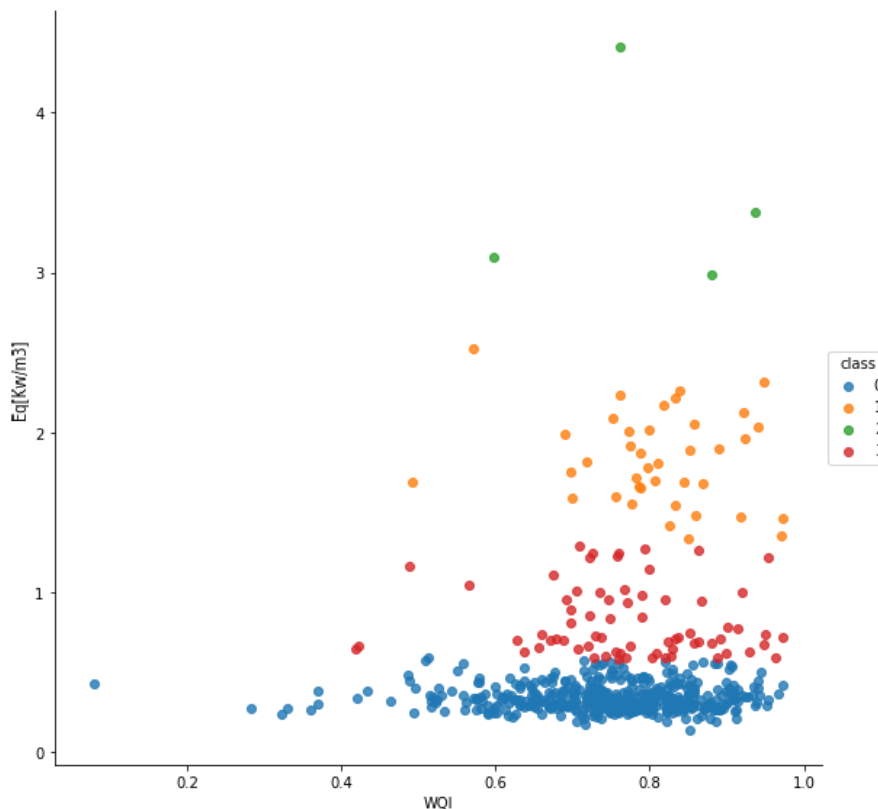


Figure 5.21: Regroupement par K-Means pour les données cumulées

D'après cette figure, on remarque bien que le regroupement se fait en quatre classes, qui sont regroupées autour des centres de classes suivants :

- **Class 0:** [0.74350405,0.34029535]
- **Class 1:** [0.8100155,1.82773924]
- **Class 2:** [0.79437073,3.46694491]
- **Class 3:** [0.77120696,0.81577405]

Pour la suite de l'étude, on a considéré le même seuil ($E/Q = 1.4$), Le nombre d'observations qu'on a obtenu après la sélection par K-Means est de 521 observations, dont 394 sont issues de Boumerdes et 127 de Tipaza.

Ces dernières observations seront ensuite introduites dans la première sélection du filtre cité plus haut.

- **Première Sélection :**

La première sélection du filtre a été appliquée sur les observations obtenues, les résultats sont les suivants :

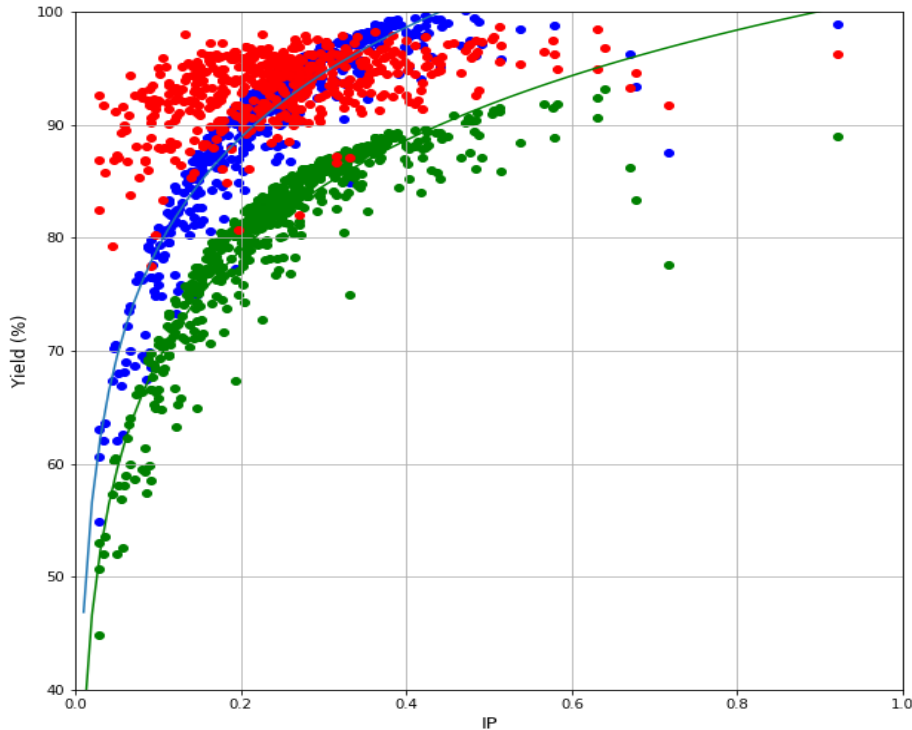


Figure 5.22: Rendement Globale en fonction de l'indice de pollution

La figure ci-dessus représente le rendement global (Rouge) de 2006 à 2016, les deux courbes en (Vert) et en (Blue) représentant respectivement le rendement global lié à la norme et le rendement global lié à la norme auquel on a ajouté une marge de (10%), en fonction de l'indice de pollution.

Les points rouges situés entre les deux courbes, sont au nombre de 168 valeurs dans lesquelles 148 de Boumerdes et 20 de Tipaza, jugés pertinents, seront sélectionnés pour la deuxième sélection du filtre.

- **Deuxième Sélection :**

On obtient, après application de la deuxième sélection du filtre énergie minimum 131 valeurs dans lesquelles 114 de Boumerdes et 17 de Tipaza. Celles-ci seront utilisées par la suite pour l'élaboration de notre modèle d'optimisation.

- **Modélisation**

A partir des résultats obtenus de la deuxième sélection, on procède à la conception du modèle de régression entre les paramètres d'entrée et sortie. Après plusieurs essais (variation sur le nombre de couches, nombre de neurones, types de fonction d'erreur et d'activation...), Le modèle ANN le plus performant et adapté pour ce problème est composé d'une couche d'entrée, 4 couches cachées de 40 neurones chacune et une couche de sortie.

La série de données a été décomposée en deux groupes, le premier contenant 66% est destiné à l'apprentissage du modèle, et 33% pour la validation des résultats du modèle.

Les résultats obtenus sont les suivants :

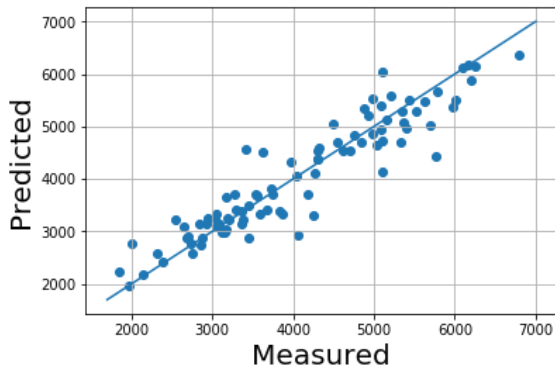


Figure 5-23 : Calibration

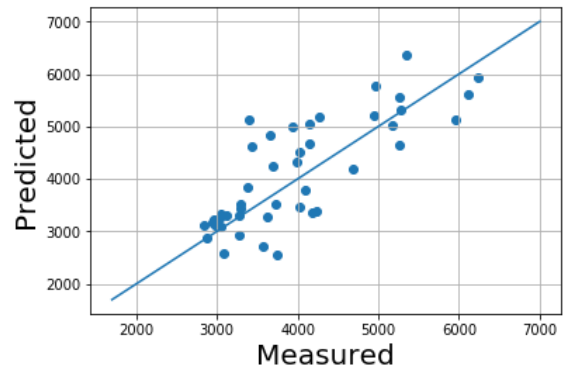


Figure 5-24 : Validation

	R2_Score	RMSE (KWh)
Calibration	0.833	435.44
Validation	0.601	651.27

Comparaison de l'énergie réelle et simulée pour la validation :

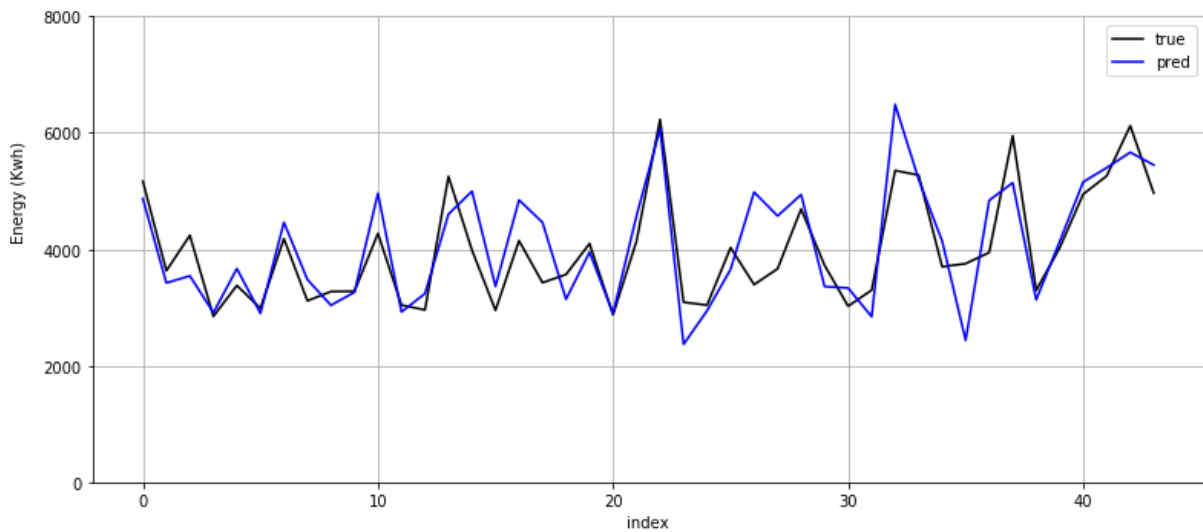


Figure 5.25: Comparaison de l'énergie réelle et simulée pour les données cumulées.

On remarque que, pour le cumul des données des deux STEP, les résultats de la modélisation se sont bien améliorés et ce que ce soit pour l'échantillon d'apprentissage ou celui de la validation, le modèle obtenu génère des valeurs très proches comparé aux valeurs mesurées.

A noter que le R2_score de l'apprentissage est assez important comparé à celui de l'échantillon de validation, ceci revient au fait que les données sélectionnées contiennent un bruit qui a été bien appris par notre modèle, ce phénomène est appelé le surapprentissage (Overfitting).

5.4.2 Deuxième approche

5.4.2.1 Application de la méthode de l'AHP

Afin d'appliquer la méthode de l'AHP on doit suivre les étapes suivantes :

- **Etape 1** : Construction de la hiérarchie.

Une fois la structure faite, nous développons pour chaque niveau les étapes qui suivront.

- **Etape 2** : Effectuer les comparaisons par paires ;

Pour chaque niveau de la hiérarchie faire la comparaison deux à deux des critères et des sous-critères puis faire la comparaison entre plusieurs expériences poids.

-donner pour chaque paire une note de 1 à 9 en suivant l'échelle de Saaty 1984.

- **Etape 3** : Déterminer les priorités ;

Une fois les matrices établies, et à l'aide de la méthode AHP, on détermine les priorités relatives pour chaque niveau de la hiérarchie.

- **Etape 4** : Évaluer la cohérence des jugements.
- **Etape 5** : Synthèses des priorités. (A. AIT-ABED, 2014).

5.4.2.2 Construction de la hiérarchie :

Nous avons construit quatre niveaux hiérarchiques à savoir :

- Niveau 0 : C'est l'objectif de cette analyse qui est l'optimisation du fonctionnement de la station d'épuration des eaux usées (STEP).
- Niveau 1 : Représente les critères établis à l'aide de la bibliographie et en se référant aux experts et aux gérants des stations d'épuration.

Ces critères seront comparés deux à deux par rapport à l'objectif principal. Les critères sont résumés dans le tableau suivant avec leurs codes :

Tableau 5-8: Critères et code de l'AHP.

CODE	CRITERE
CPAS	La Charge Polluante Admise à la STEP
RE	Le Rendement Epuratoire
RDE	Les Rejets Dans l'Environnement
E	La Consommation d'Energie

- Niveau 2 : Ce niveau compare les sous-critères qui sont les données recueillies de la STEP (MES, DBO5, DCO, NH4) par rapport aux critères (niveau supérieur).
- Niveau 3 : C'est le niveau le plus bas de la hiérarchie, il contient les alternatives, il représente les différentes expériences comparées par rapport aux expériences de référence. Nous détaillerons par la suite les niveaux 2 et 3 avec la procédure utilisée afin de parvenir à la notation finale.

Pour notre cas les matrices des critères de l'AHP se présentent comme suit :

Tableau 5-9: Matrice de jugement de l'indicateur CPAS

	MESb	DBOb	DCOb	NH4b
MESb	1	1	1	3
DBOb	1	1	1	3
DCOb	1	1	1	3
NH4b	1/3	1/3	1/3	1

Tableau 5-10: Matrice de jugement du rendement épuratoire.

	R(MES)	R(DBO)	R(DCO)	R(NH4)
R(MES)	1	3	7	9
R(DBO)	1/3	1	5	7
R(DCO)	1/7	1/5	1	3
R(NH4)	1/9	1/7	1/3	1

Tableau 5-11: Matrice de jugement rejet dans la nature.

	Be x	N° de Rendement Respectés
Be x	1	1
N° de Rendement Respectés	1	1

Pour plus d'information sur la procédure suivie pour l'obtention de ces résultats voir le PFE (Souar, 2017).

5.4.2.3 Modélisation

Dans cette partie on cherchera à modéliser la consommation énergétique en se basant sur les résultats de l'AHP, pour cela plusieurs variantes ont été exploitées.

Pour toutes ces variantes, le même modèle a été considéré, celui-ci est composé d'une couche d'entrée, 9 couches cachées contenant chacune 60 neurones et d'une couche de sortie.

5.4.2.3.1 Modélisation la consommation énergétique pour une échelle de Saaty = 1

Dans cette expérience tous les critères ont la même importance. Les résultats obtenus sont les suivants :

Tableau 5-12: Matrice de jugement du niveau 1

	charge polluante	Rendement epuratoire	Les rejets dans l'environnement	consommation Energétique
charge polluante	1	1	1	1
Rendement epuratoire	1	1	1	1
Les rejets dans l'environnement	1	1	1	1
consommation Energétique	1	1	1	1

- **K-means**

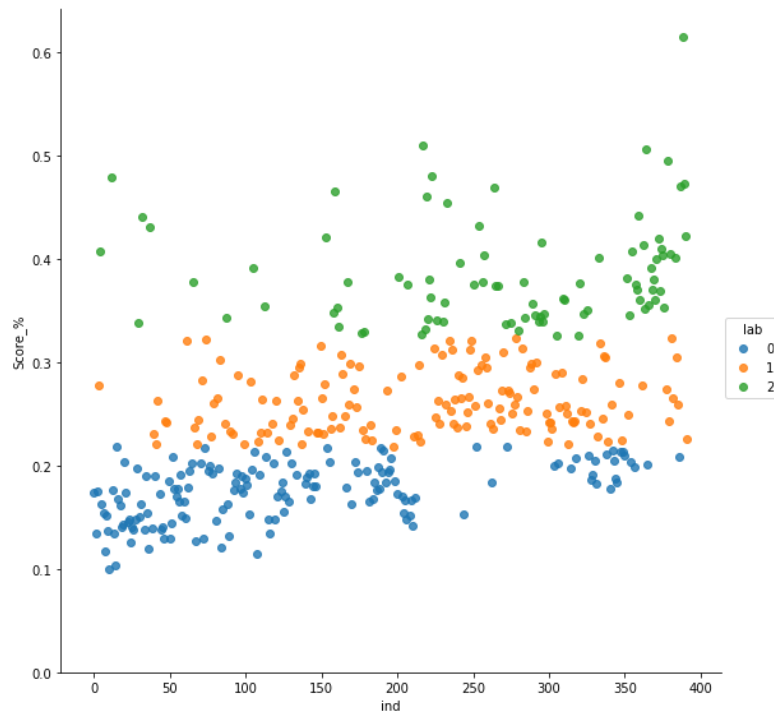


Figure 5.26: regroupement par K-Means pour Saaty=1

On obtient 3 classes avec les centroïdes suivants :

Classe 0 : 0.174968

Classe 1 : 0.261884

Classe 2 : 0.387177

Pour la suite de la modélisation, on a considéré les classes 1 et 2, car elles possèdent un bon score final.

Les résultats sont les suivants :

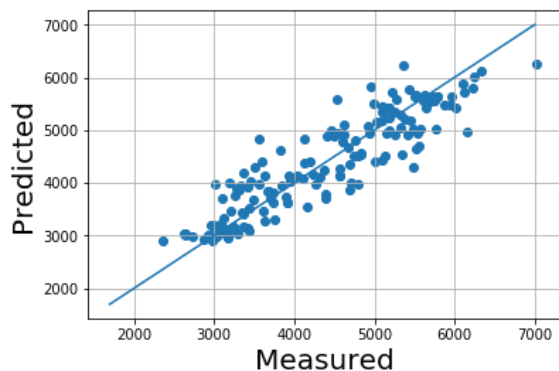


Figure 5-27 : Calibration

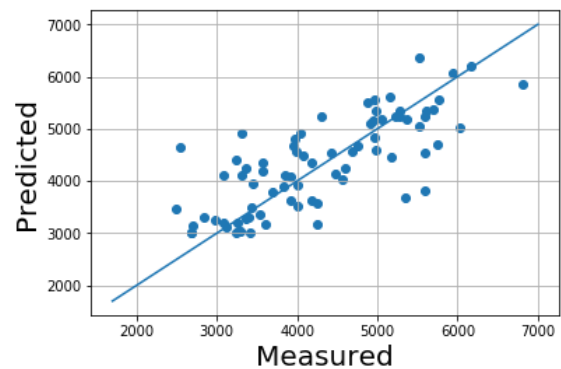


Figure 5-28 : Validation

	R2_Score	RMSE (KWh)
Calibration	0.82	438.46
Validation	0.57	665.24

Comparaison graphique des résultats pour la validation

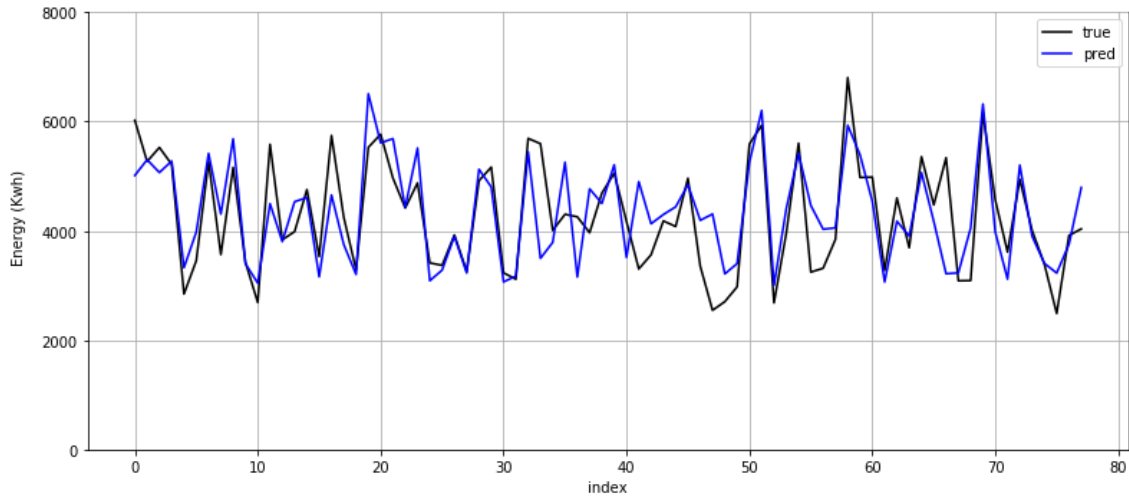


Figure 5.29 : Comparaison entre énergie simulée et mesuré pour Saaty=1

Pour une échelle de Saaty de 1, c'est-à-dire les poids de l'AHP équitables (tous les paramètres ont une pondération de 1), l'erreur absolue moyenne des prédictions est de l'ordre de 665.24 KWh ce qui représente 16.4% de la moyenne de la consommation énergétique. Le R2 Score entre l'énergie simulée et réelle étant estimé à 0.57 (Validation). Ainsi que le R2 Score de la partie calibration est de 0.82.

De là on peut déduire que le modèle obtenu est de qualité moyenne, tout en soulignant la présence de phénomène de surapprentissage (overfitting).

5.4.2.3.2 Modélisation de la consommation énergétique pour une échelle de Saaty = 3

Dans cette expérience la consommation énergétique a été peu privilégiée par rapport aux autres critères entrant dans le calcul de l'AHP. Le poids attribué à la consommation énergétique est 3 fois plus important que pour les autres termes.

Tableau 5-13: Matrice de jugement du niveau 3

	charge polluante	Rendement epuratoire	Les rejets dans l'environnement	consommation Energétique
charge polluante	1	1	1	1/3
Rendement epuratoire	1	1	1	1/3
Les rejets dans l'environnement	1	1	1	1/3
consommation Energétique	3	3	3	1

- **K-means**

On obtient 3 classes avec les centroïde suivants :

Classe 0 : 0.166892

Classe 1 : 0.291956

Classe 2 : 0.456077

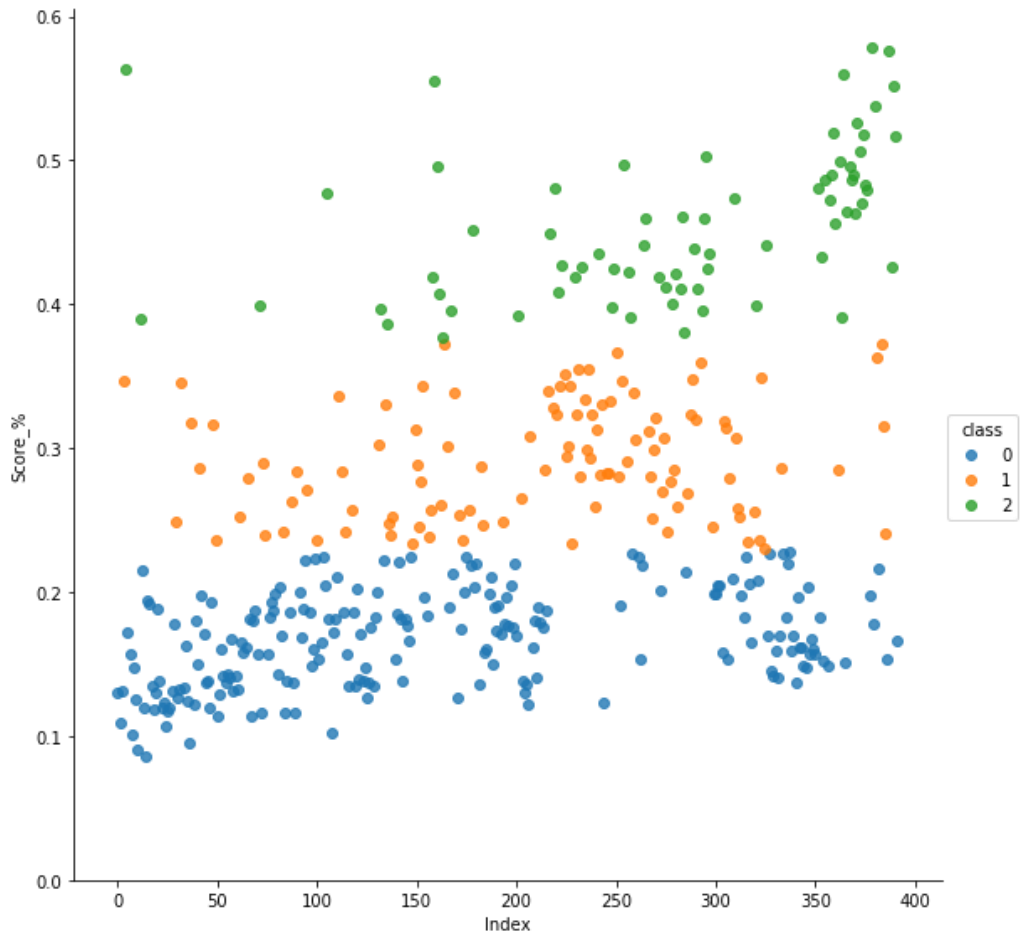


Figure 5.30 : regroupement par K-Means pour Saaty= 3

Pour la suite de la modélisation, on a considéré les classes 1 et 2, car elles possèdent un bon score final.

Les résultats sont les suivants :

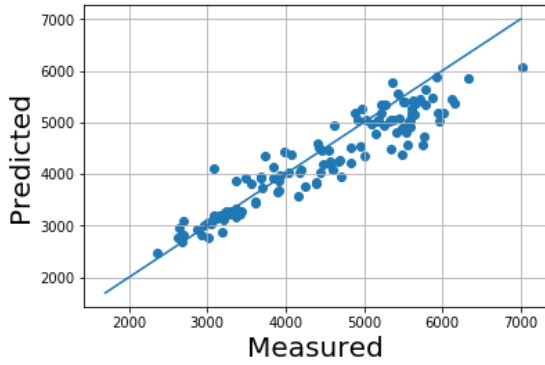


Figure 5.31 : Calibration

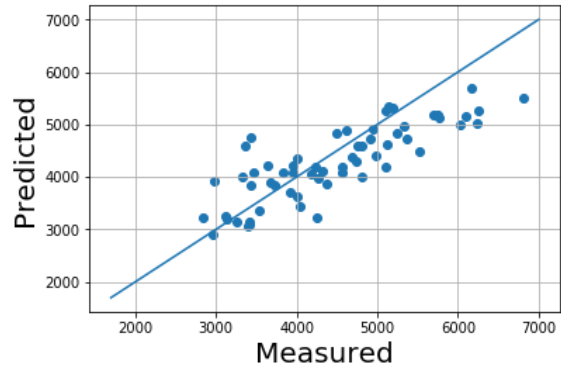


Figure 5.32: Validation

	R2_Score	RMSE (KWh)
Calibration	0.849	415.375
Validation	0.665	560.1018

Comparaison graphique des résultats pour la validation

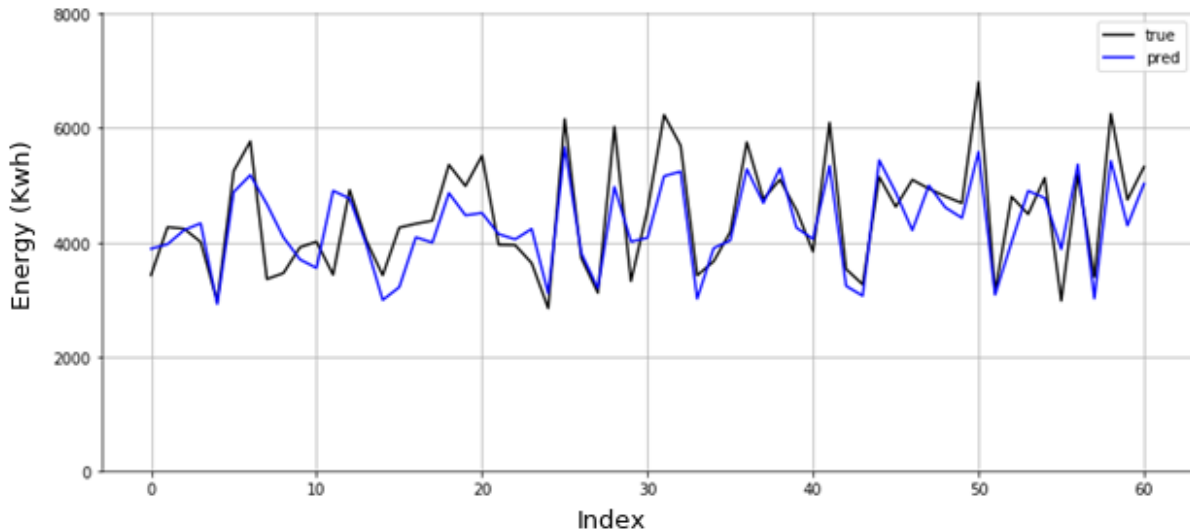


Figure 5.33 : Comparaison entre énergie simulée et mesuré pour Saaty=3

Pour une échelle de Saaty de 3, l'erreur absolue moyenne des prédictions est de l'ordre de 560.10KWh ce qui représente 13.8% de la moyenne de la consommation énergétique. Le R2 Score entre l'énergie simulée et réelle étant estimé à 0.66 (Validation). Ainsi que le R2 Score de la partie calibration est de 0.849.

On remarque bien que des améliorations au niveau des résultats ont été apporté par rapport au modèle précédent, ainsi qu'une diminution du surapprentissage (diminution du bruit).Le modèle quant à lui reste toujours de qualité moyenne.

5.4.2.3.3 Modélisation de l'énergie électrique pour une échelle de Saaty = 5

Dans cette expérience la consommation énergétique a été bien privilégiée par rapport aux autres critères entrant dans le calcul de l'AHP. Le poids attribué à la consommation énergétique est 5 fois plus important que pour les autres termes.

Tableau 5-14: Matrice de jugement du niveau 5

	charge polluante	Rendement epuratoire	Les rejets dans l'environnement	consommation Energétique
charge polluante	1	1	1	1/5
Rendement epuratoire	1	1	1	1/5
Les rejets dans l'environnement	1	1	1	1/5
consommation Energétique	5	5	5	1

- **K-Means**

On obtient 3 classes avec les centroïde suivants :

- Classe 0** : 0.303874
- Classe 1** : 0.500851
- Classe 2** : 0.154952

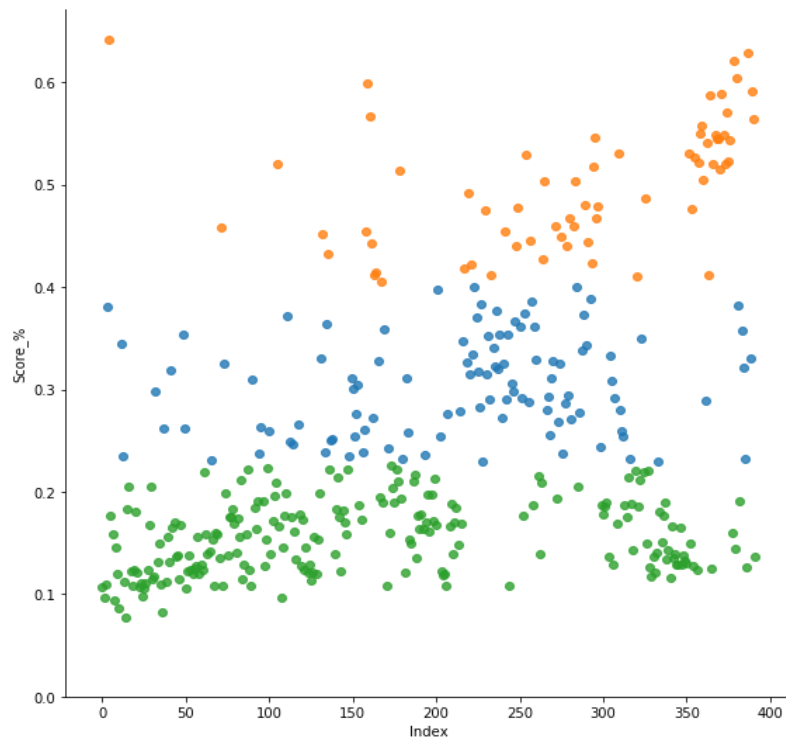


Figure 5.34 : regroupement par K-Means pour Saaty = 5

Pour la suite de la modélisation, on a considéré les classes 1 et 2, car elles possèdent un bon score final.

Les résultats sont les suivants :

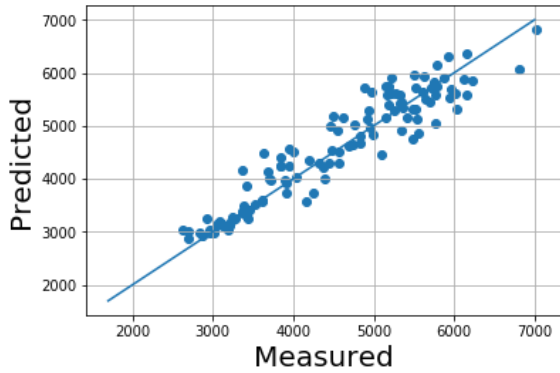


Figure 5.35 : Calibration

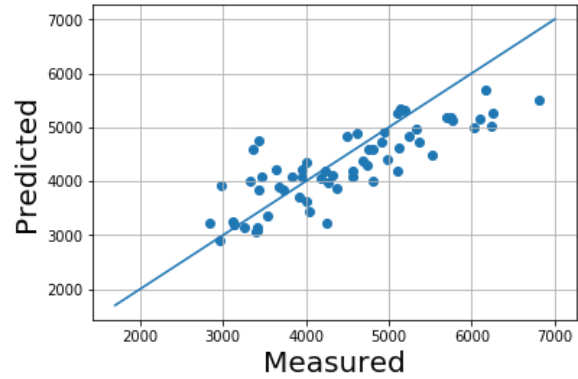


Figure 5.36 : Validation

	R2_Score	RMSE (KWh)
Calibration	0.887	358.282
Validation	0.758	473.372

Comparaison graphique des résultats pour la validation

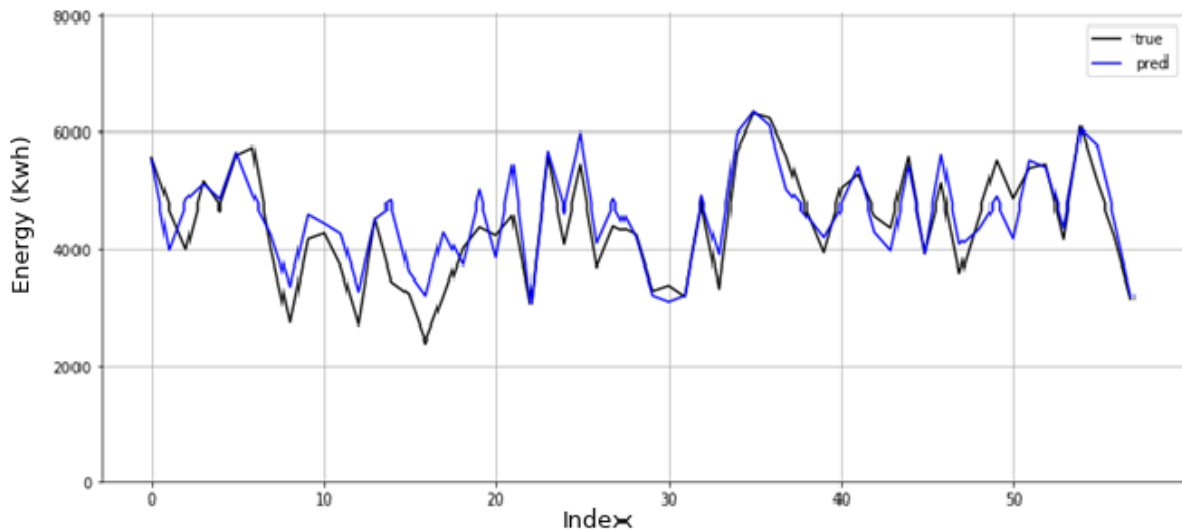


Figure 5.37 : Comparaison entre énergie simulée et mesuré pour Saaty = 5

Pour une échelle de Saaty de 5, l'erreur absolue moyenne des prédictions est de l'ordre de 473.372KWh ce qui représente 11.6% de la moyenne de la consommation énergétique. Le R2 Score entre l'énergie simulée et réelle étant estimé à 0.758 (Validation). Ainsi que le R2 Score de la partie calibration est de 0.887.

On constate qu'au fur et à mesure de la modification de l'échelle de Saaty les résultats du modèle s'améliorent d'une façon notable par rapport aux modèles précédents. en témoigne l'augmentation du R2_Score pour la validation et aussi pour la calibration (diminution du bruit), ainsi la diminution de l'erreur moyenne absolue.

5.4.2.3.4 Modélisation de la consommation énergétique pour une échelle de Saaty = 7

Dans cette expérience la consommation énergétique a été très privilégiée par rapport aux autres critères entrant dans le calcul de l'AHP. Le poids attribué à l'énergie est 7 fois plus important que pour les autres termes.

Tableau 5-15: Matrice de jugement du niveau 7

	charge polluante	Rendement epuratoire	Les rejets dans l'environnement	consommation Energétique
charge polluante	1	1	1	1/7
Rendement epuratoire	1	1	1	1/7
Les rejets dans l'environnement	1	1	1	1/7
consommation Energétique	7	7	7	1

- **K-means**

On obtient 3 classes avec les centroïde suivants :

Classe 0 : 0.319113

Classe 1 : 0.148614

Classe 2 : 0.536654

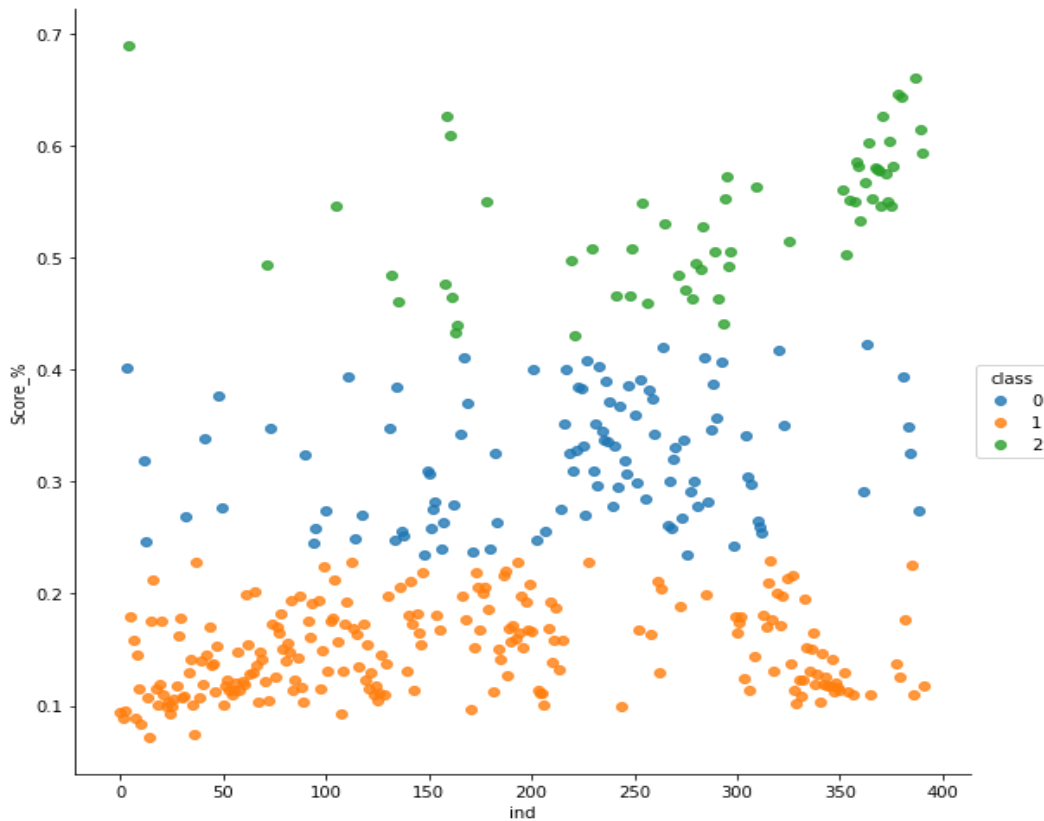


Figure 5.38: regroupement par K-Means pour Saaty=7

Pour la suite de la modélisation, on a considéré les classes 1 et 2, car elles possèdent un bon score final.

Les résultats sont les suivants :

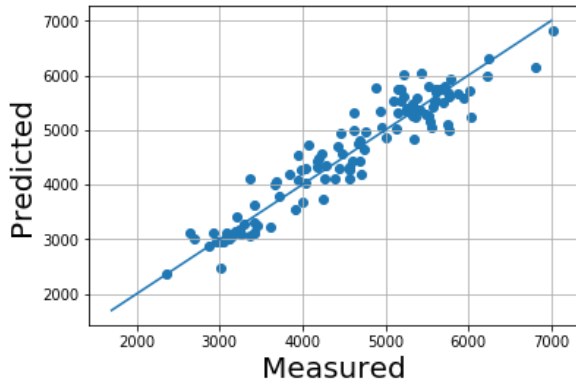


Figure 5.39 : Calibration

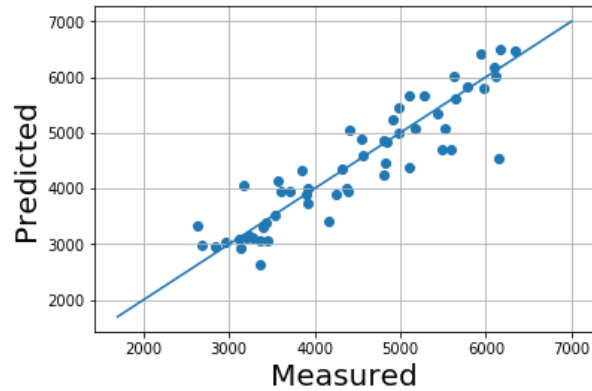


Figure 5.40: Validation

	R2_Score	RMSE (KWh)
Calibration	0.9	330.65
Validation	0.83	435.69

Comparaison graphique des résultats pour la validation

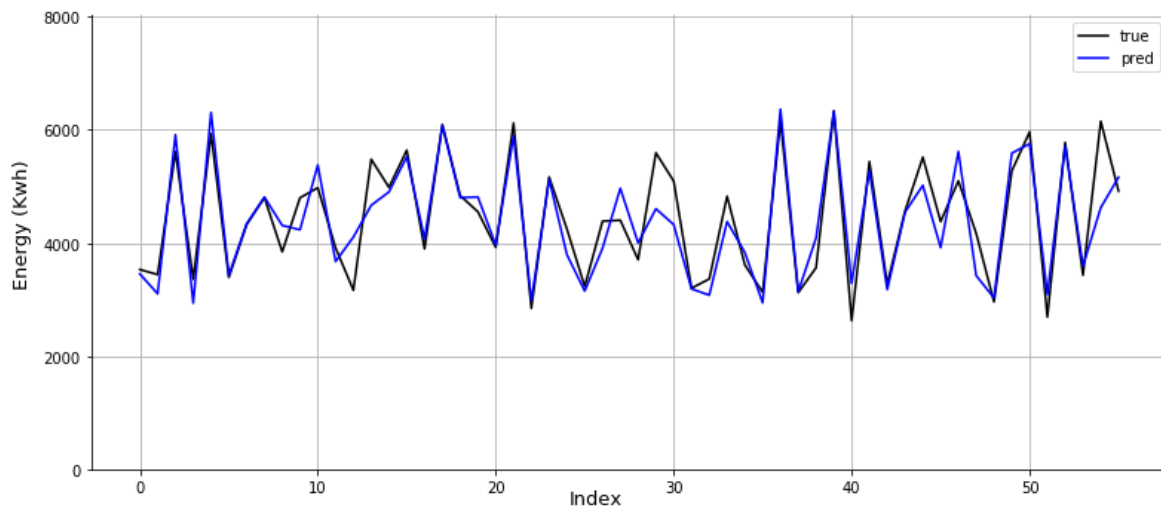


Figure 5.41: Comparaison entre énergie simulée et mesuré pour Saaty=7

Pour une échelle de Saaty de 7, l'erreur absolue moyenne des prédictions est de l'ordre de 435.69KWh ce qui représente 10% de la moyenne de la consommation énergétique. Le R2 Score entre l'énergie simulée et réelle étant estimé à 0.83 (Validation). Ainsi que le R2_Score de la partie calibration est de 0.9.

Les résultats obtenus démontrent les hautes performances obtenues par ce dernier modèle par rapport aux trois modèles élaborés précédemment, ceci en termes de qualité des paramètres statistiques considérés.

Le mérite de cette nette amélioration revient à la sélection faite par l'algorithme des K-means sur les résultats de l'AHP. En effet, les regroupements obtenus décrivent un caractère de plus

en plus précis, ce qui a permis une réduction très significative du bruit interne, et ainsi une élimination du surapprentissage.

Ce dernier modèle sera choisi pour le calcul du gain d'énergie pour les deux STEP, étant donné la pertinence des résultats obtenus.

Récapitulatif des résultats des modèles :

Tableau 5-16: Résultats des modèles

Model	RMES Training Data (Kwh)	RMES Test Data (Kwh)	Score Training Data	Score Test Data
1	576.27	794.05	0.686	0.349
2	534.15	713.12	0.728	0.506
3	435.44	651.27	0.833	0.601
4	438.46	665.24	0.82	0.57
5	415.37	650.1	0.849	0.665
6	358.28	473.37	0.887	0.758
7	330.65	435.69	0.9	0.83

5.4.3 Calcul du gain d'énergie

Dans cette partie nous allons exploiter le modèle choisi afin de déterminer le gain en matière d'énergie électrique durant les années « 2008-2016 » pour Boumerdes et 2008-2012 pour Tipaza.

Les résultats de la simulation de l'ensemble des journées non sélectionnées sont illustrés dans le graphe suivant :

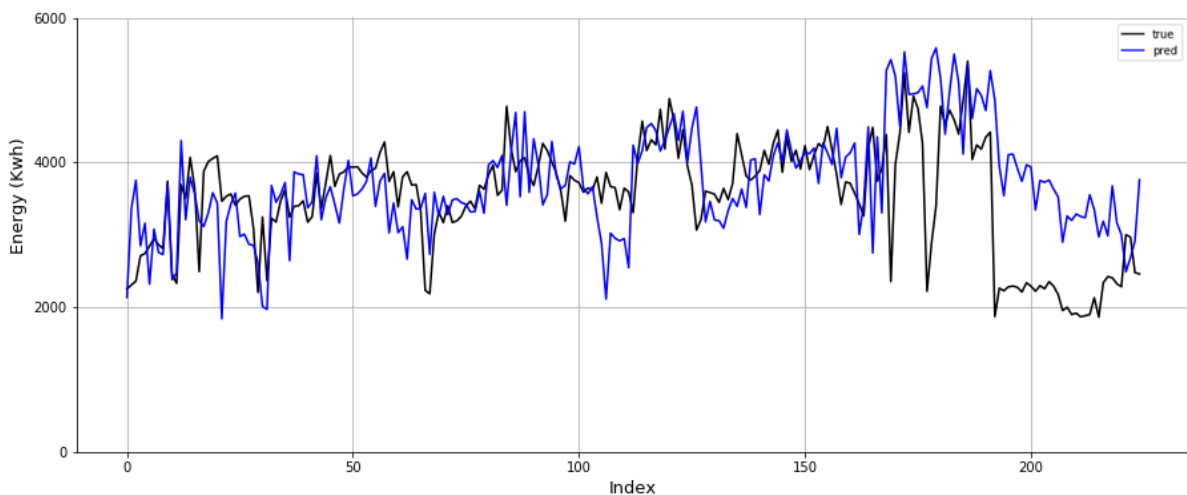


Figure 5.42 : Simulation du gain d'énergie

La figure ci-dessus représente la consommation énergétique réelle consommée par les aérateurs (en noir), et celle du gain simulé par notre modèle d'optimisation (en bleu).

L'économie d'énergie totale annuelle pour les STEP de Boumerdes et de Tipaza est représentée dans le tableau et l'histogramme suivant :

- **Boumerdes**

Tableau 5-17: Calcul du gain annuel d'énergie de Boumerdes

Année	2008	2009	2010	2011	2012	2013	2014	2015	2016
Gain (Kw)	120000	60000	240000	200000	290000	295000	330000	300000	290000

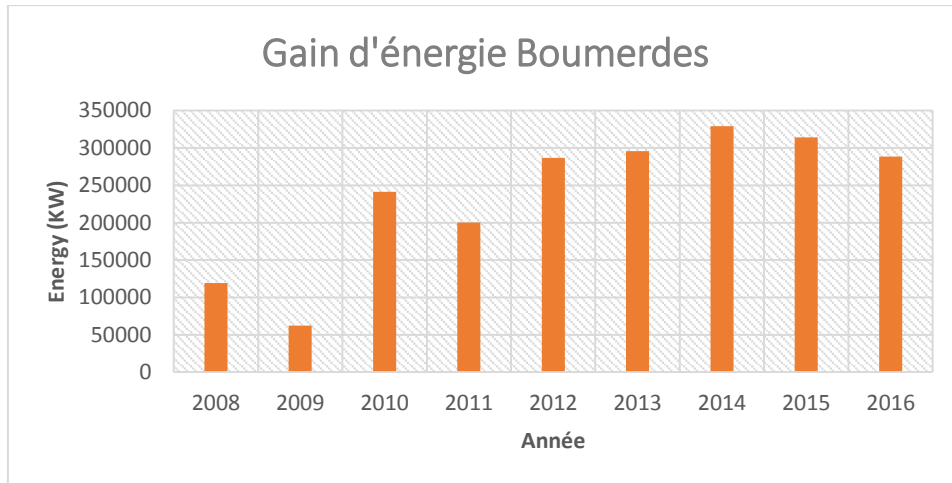


Figure 5.43 : Histogramme du gain d'énergie annuelle de Boumerdes

- **Tipaza**

Tableau 5-18: Calcul du gain annuel d'énergie de Tipaza

Année	2008	2009	2010	2011	2012
Gain (Kw)	-40000	-70000	-45000	-170000	175000

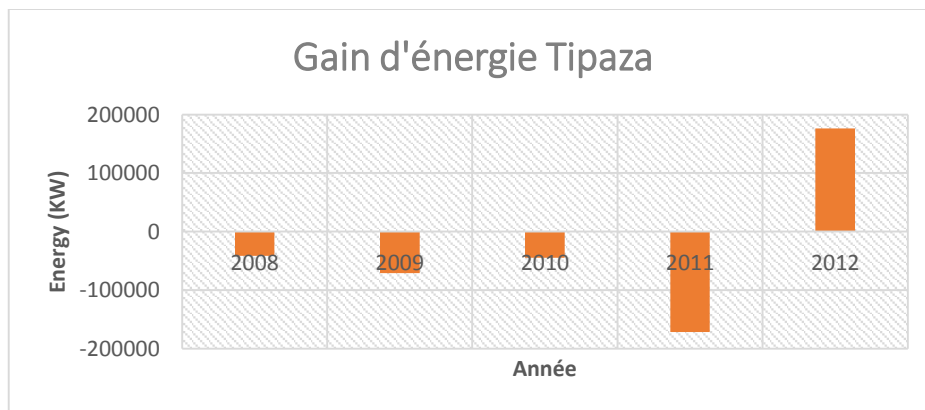


Figure 5.44 : Histogramme du gain d'énergie annuelle de Tipaza

D'après ces résultats, on voit bien que pour la STEP de Boumerdes le modèle d'optimisation conçu est très efficace, le gain d'énergie calculé étant important.

L'année où l'économie d'énergie est la plus importante est 2014 avec un gain de 330000 KW, qui est justifié par la surexploitation de la STEP à partir de 2013.

On remarque aussi que notre modèle a surestimé l'énergie pour la STEP de Tipaza, ceci s'explique par le fait que ce dernier est surtout calibré pour la STEP de Boumerdes étant donné la différence du nombre d'observations entre les deux STEP.

5.4.4 Conclusion

Dans ce chapitre, on a utilisé de deux approches statistiques qui sont le K-Means et l'ACP afin de justifier le cumul des données pour l'élaboration du modèle. Les résultats obtenus nous ont permis d'affirmer que le cumul des données est viable.

Par la suite, deux approches différentes ont été utilisées dans le but d'une sélection optimale des meilleures expériences : la première, empirique est basée sur K-Means et le filtre à deux sélections. Les résultats de cette première approche confirment l'utilité du cumul des données par apport à la considération individuelle de chaque STEP.

La deuxième approche, quant à elle, est basée sur l'AHP, pour des différentes échelles de Saaty. Le Clustering K-means a ensuite été appliqué sur ces différentes variantes pour enfin aboutir à une modélisation de la consommation énergétique. Les résultats de cette approche montrent que, plus le poids de la consommation énergétique dans l'AHP est important, plus le modèle devient meilleur.

Finalement, le dernier modèle a été choisi pour le calcul du gain d'énergie. Ce dernier montre que, pour la STEP de Boumerdes, une importante économie d'énergie est possible, à l'inverse de la STEP de Tipaza à cause de la différence du nombre d'observations entre les deux STEP.

Conclusion générale

Le but de ce travail est de porter un jugement sur la consommation de l'énergie dans les stations d'épuration de Boumerdes et de Tipaza. La bonne gestion des procédés épuratoires entraîne à une bonne estimation des coûts et un respect des objectifs de qualité à la sortie de la station.

La conception de modèles de régressions basées sur la nouvelle approche du Deep Learning utilisant les réseaux de neurones artificiels a été établie et expliqué en détail.

A l'amont des modèles, on a représenté l'évolution de chaque effluent au fil du temps afin de pouvoir faire une critique des données d'une part. D'autre part on a testé et critiqué à travers deux techniques statistiques qui sont K-Means et l'analyse en composantes principales dans le but de justifier l'utilisation du cumul des données des deux STEP pour l'élaboration de nos modèles.

Les premier modèle d'optimisation était établi à partir de la matrice des données sélectionnées, cette dernière est le fruit de deux étapes de sélection, une première sélection par rapport au rapprochement du rendement épuratoire du rendement optimal et une deuxième par rapport à une consommation énergétique minimale. Ceci a été appliqué pour la STEP de Boumerdes seul, puis celle de Tipaza puis le cumul des deux STEP

Les seconds modèles élaborés sont le fruit de l'application de la méthode AHP sur la série de mesures des différents paramètres caractérisant le fonctionnement de la STEP, cette méthode passe par quatre grandes étapes dont la plus importante est celle des différentes comparaisons par paires, une étape dans laquelle on fait intervenir la méthode ACP, pour remplacer en quelque sorte l'avis d'expert en ce qui concerne le jugement de l'importance et la participation dans le phénomène d'un paramètre par rapport à un autre.

Différents logiciels et outils de programmation ont été utilisés lors de ce travail, tel que le logiciel XLSTAT pour l'application de la méthode ACP, l'outil de calcul Matlab pour la méthode AHP, et le Langage de programmation PYTHON avec ses bibliothèques (Keras, Pandas...) pour le K-Means et Le Deep Learning.

L'AHP nous a permis d'obtenir des résultats très satisfaisants au niveau des modèles, spécialement pour une échelle de Saaty de 7, pour laquelle le modèle a été choisi pour le calcul du gain d'énergie. Les résultats montrent une meilleure économie d'énergie pour la STEP de Boumerdes comparé à celle de Tipaza, pour des raisons qu'on a citées plus haut.

Il est donc à dire que notre étude, nous a permis de franchir un pas très important dans la recherche des modèles d'optimisation de la consommation en énergie électrique des stations d'épuration des eaux usées, qui reste un domaine vierge qui n'a pas encore livré tous ses secrets.

Bibliographie

- (2017). Récupéré sur memoireonline: <https://www.memoireonline.com/04/10/3292/Contribution--lelaboration-de-la-politique-dassainissement-des-eaux-usees-domestiques-par-.html>
- (2017). Récupéré sur memoireonline: <https://www.memoireonline.com/11/13/7935/Traitement-des-eaux-usees-urbaines-par-boues-activees-au-niveau-de-la-ville-de-Bordj-Bou-Arzeridj.html>
- A. AIT-ABED, S. (2014). *Élaboration d'un Dashboard pour la gestion de la STEP de Boumerdes par l'analyse multicritère*. projet de fin d'étude en vue de l'obtention du diplôme d'ingénieur d'état en Hydraulique,ENPA.
- Abassi, T. (2012). *WATER QUALITY INDICES*.
- ABDELAZIZ, S. (2016). *Couplage Analyse en Composantes Principales Neuro-flou pour la modélisation de l'énergie électrique de la STEP de Boumerdes*.
- ABDELAZIZ.Soraya. (2016). *Couplage Analyse en Composantes Principales Neuro-flou pour la modélisation de l'énergie électrique de la STEP de Boumerdes*.
- Adam. (2015). *A Method for Stochastic Optimization arXiv:1412.6980 Conference paper at the 3rd International Conference for Learning Representations*. san diego.
- Baumont, S. &. ((2004)). *Réutilisation des eaux usées, risques sanitaires et faisabilité en Île-de-France*. France: Rapport ORS, 220 pages.
- BELAHMADI Mohamed Seddik Oussama, «. ». (2004, juin 16). étude de la biodégradation du 2.4-dichlorophénol par le micro-biote des effluents d'entrée et de sortie de la station d'épuration des eaux usées d'Ibn Ziad. *Mémoire de Magister en Microbiologie appliquée et biotechnologie microbienne, université Mentouri*. constantine, algerie.
- Belhabib Abdelkader, L. O. (2012). développement d'une application à base de l'algorithme de classification k-means'', Mémoire de fin d'études pour l'obtention du diplôme de Licence en Informatique. université Abou bakr belkaid, Tlemcen .
- Bruce L.G, W. E. (1989). *The Analytic Hierarchy Process : Applications and Studies*.Springer-Verlag, New York.
- C. Duby, S. R. (juillet 2006). *Support de cours Analyse en Composantes Principales*. AgroParisTech.
- CHAMROUKHI, F. (2013). *Classification supervisée : Analyse discriminante, Licence 2 Sciences Pour l'Ingénieur, Université du Sud Toulon*.
- Collobert, R. (2011). *Deep learning for efficient discriminative parsing*.AISTATS.
- Commission de protection des eaux, du patrimoine, de l'environnement, du sous-sol et des. (s.d.).
- Data_Mining. (s.d.). Support de cours Université Paul Sabatier : M2 SE .
- deep-learning. (2018, février). Récupéré sur <https://www.learnopencv.com/understanding-activation-functions-in-deep-learning>
- déshydratation, T. d. (2017, décembre). *ACTUENVIRONNEMENT* . Récupéré sur ACTUENVIRONNEMENT.com

- DJAOUI.Toufik. (2016). *Optimisation du débit de boue à recirculée pour la STEP de Boumerdes*.
- Duchène, P. (1994, octobre 21). Dysfonctionnement biologiques dans les stations 'épuration a boues activées. *Actes du colloque*. Lyon.
- Durand., C. (juillet 2005). *L'analyse factorielle et l'analyse de fidélité*.
- Edline, F. (1979). *L'épuration biologique des eaux résiduaires,306 pages*. CEBEDOC édition Paris.
- Gaëlle Deronzier, S. S.-P. (s.d.). *Traitement de l'azote dans les stations d'épuration biologique des petites collectivités* . FNDAE n° 25, Document technique.
- Gaid, A. (2005). Epuration biologique des eaux usées urbaines'. *publication international N° 1247 05/84*.
- GUERREE, H. &. (1978). *'Les eaux usées dans les agglomérations urbaines et rurales,tome 2,le traitement, EYROLLES*.
- Hamdani.A. (2002). *Caractérisation et essais de traitement des effluents d'une industrie laitière : aspects microbiologiques et physico- chimiques*. Thèse de doctorat, Faculté des Sciences d'El Jadida, Maroc.
- HAMRICHE, A. (1993). *Contribution à l'étude et à la simulation des paramètres hydrométrique par ACP*. Mémoire de Projet de Fin d'Etude,ENP Alger.
- INRS. (s.d.). *Station d'épuration des eaux usées > Prévention des risques biologiques*.
- J.MERESSE, A. e. (Juin 2007). *pôle environnement Laurent Burget-Mountain Riders. 41cours du Dr long 69003*. Lyon.
- Kogan, J. (2007). *Introduction to Clustering Large and High-Dimensional Data*. Cambridge University Press.
- L.G.Vargas, T. (1991). *Prediction, projection and forecasting* .
- LAREDJ.Nesrine. (2013). *Application de l'Analyse en Composantes pour le diagnostic du fonctionnement de la station d'épuration de Boumerdess*.
- Likas A., V. M. (2003). he global k-means clustering algorithm, *Pattern Recognition*, 36, pp. 451-461.,.
- mathworks. (2018, janvier). Récupéré sur www.mathworks.com/discovery/les-reseaux-de-neurones
- mathworks. (2018, fivrier). Récupéré sur www.mathworks.com/help/nnet/ug/neural-network-architectures
- mathworks. (2018, février). Récupéré sur <https://www.mathworks.com/discovery/apprentissage.html>
- Matthieu., D. (2007/2008 .). *Guide pour l'établissement des Plans d'Assurance de la Qualité dans le cadre de la réalisation des stations d'épuration de type boues activées en lots séparés*. INSA Strasbourg: Mémoire de Projet de Fin d'Etudes.
- MELIANI.Zouhir_SERIKMA.Khaled. (2014). *Caractérisation des paramètres de décision pour le diagnostic des STEP par l'Analyse en Composantes Principales (ACP) Cas des STEP de Boumerdes et Tipaza*.
- Mining, D. (s.d.). Support de cours Université Paul Sabatier : M2 SE .

- Moletta, R. (s.d.). *l'eau, sa pollution, et son traitement*.
- (s.d.). *Qualité physico-chimique et chimique des eaux de surface* . Cadre général Fiche 2. Institut Bruxellois pour la gestion de l'environnement/observatoire des données de l'environnement.
- Ramade, F. (2000). *Dictionnaire encyclopédique des pollutions*, 689 pages. Ediscience international Paris.
- S . P . Bradley, U. F. (1998). Scaling clustering algorithms to large databases In knowledge Discovery and Data Mining pages 9-15.
- Schmidhuber, J. (2015). *Deep learning in neural networks: An overview*, *Neural Networks*, 61, 85-117.
- Soraya, A. (2016). *Couplage Analyse en Composantes Principales Neuro-flou pour la modélisation de l'énergie électrique de la STEP de Boumerdes*.
- Souar, B. (2017). *Clustérisation par KPI pour l'optimisation du procédé d'épuration par boue activée*.
- stats.stackexchange. (2018, février). Récupéré sur <https://stats.stackexchange.com/questions/154879/a-list-of-cost-functions-used-in-neural-networks-alongside-applications>
- T.L, S. (1984). *Décider face à la complexité : une approche analytique multicritère d'aide à la décision*. Paris: traduit par Dahan L.
- Touzet, C. (s.d.). *LES RESEAUX DE NEURONES ARTIFICIELS, INTRODUCTION AU CONNEXIONNISME: COURS, EXERCICES ET TRAVAUX PRATIQUES. EC2, 1992, Collection de l'EERIE*, N. Giambiasi.
- Victor-Hugo, G. P. (2007). *Modélisation et commande floues de type takagi-sugeno appliquées à un bioprocédé de traitement des eaux usées*. Toulouse.
- W.W.ECKENFELDER. (1982). *Gestion des eaux usées urbaines et industrielles*. Paris: Techniques et documentation (Lavoisier).
- www.oieau.fr/eaudoc/integral/reuinter.htm. (2017, Decembre).
- Y.libes. (s.d.). *les eaux usées urbaines et leur épuration*.
- Activated sludge models ASM1, ASM2, ASM2d and ASM3*. IWA Publishing, 2000.
- Water Quality Indices*. Elsevier, 2012.

Deep Learning A Practitioner's Approach. O'Reilly Media, Inc, 2017.

R.M Brown, N.I. McClelland, R.A. Deininger, and R.G. Tozer. A water quality index do we dare? *Water Sewage Works*, 1970.

Benoit Chachuat, Nicolas Roche, and MA Latifi. Long-term optimal aeration strategies for small-size alternating activated sludge treatment plants. *Chemical Engineering and Processing: Process Intensification*, 2005.

Sylvie Gillot and Jean-marc Choubert. Biodegradable organic matter in domestic wastewaters: comparison of selected fractionation techniques. *Water Science and Technology*, 2010.

Rainier Hreiz, Abderrazak M. Latifi, and Nicolas Roche. Optimal design and operation of activated sludge processes: State-of-the-art. *The Chemical Engineering Journal*, 2015.

Gernaey Krist V., van Loosdrecht Mark C.M., Henze Mogens, Lind Morten, and Jørgensen Sten B. Activated sludge wastewater treatment plant modelling and simulation: state of the art. *Environmental Modelling & Software*, 2004.

H. Melcer, P.L. Dold, R. M. Jones, C. M. Bye, I. Takacs, H. D. Stensel, A. W. Wilson, Sun P., and Bury S. Methods for wastewater characterization in activated sludge modeling. *Water Environment Research Foundation (WERF)*, 2003.

Erika Pittoors, Yaping Guo, and Stijn W. H. Van Hulle. Modeling dissolved oxygen concentration for optimizing aeration systems and reducing oxygen consumption in activated sludge processes: A review. *Chemical Engineering Communications*, 2014.

R.Maachou, A. Lefkir, A. Bermad, T. Djaoui, and A. Khouider. Statistical analysis of pollution parameters in activated sludge process. *Desalination and Water Treatment*, 2016.

R.Maachou, A. Lefkir, A. Bermad, and A. Khouider. Factorization of physicochemical parameters of activated sludge process using the principal component analysis. *Desalination and Water Treatment*, 2015.

R.Maachou, A. Lefkir, A. Bermad, and A. Khouider. Modeling of activated sludge process using artificial neuro-fuzzy-inference system (anfis). *Desalination and Water Treatment*, 2015.

LeCun Yann, Bengio Yoshua, and Hinton Geoffrey. Deep learning. *Nature*, 2015.