

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

Ecole Nationale Polytechnique



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

Projet de fin d'études

Pour l'obtention du diplôme d'ingénieur d'état en Hydraulique

Thème :

Couplage Analyse en Composantes Principales Neuro-flou pour la
modélisation de l'énergie électrique de la STEP de Boumerdes

Soraya ABDELAZIZ

Sous la direction de Mr. Abdelmalek BERMAD Professeur ENP

Composition du Jury :

Président	Mr. E.BENYOUSSEF	Professeur	ENP
Promoteur	Mr. A.BERMAD	Professeur	ENP
Examineurs	Mr. A.LEFKIR	Docteur	ENSTP
	Mme. A.HAMRICHE	Doctorante	ENP

ENP 2016

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

A la mémoire de mon grand-père

Je dédie ce travail à ma petite sœur Sonia en lui souhaitant la guérison

REMERCIEMENTS :

Je tiens tout d'abord à exprimer mes remerciements les plus chaleureux au professeur Abdelmalek BERMAD qui m'a encadrée dans la réalisation de ce mémoire et m'a fait bénéficier de ses conseils éclairés avec une disponibilité et une patience qui m'ont grandement facilité ma tâche.

De même, je remercie vivement madame A.HAMRICHE et monsieur A.LEFKIR pour l'aide permanente qu'ils m'ont apportée chaque fois que je me suis heurtée à des difficultés dans mon travail.

Je n'oublierai pas non plus d'exprimer ma reconnaissance à l'ensemble du jury qui a bien voulu étudier mon mémoire et l'évaluer avec toute la compétence qui est la sienne.

A cette occasion, je mentionne aussi avec une particulière affection mon père, monsieur Abdelaziz Antar, ma mère, Abdelaziz Fella, mon frère, Fouzi et ma sœur Radhia et son mari Hamza qui, indirectement mais efficacement m'ont apporté leur soutien moral et ont contribué à me soutenir dès l'entame de mon mémoire.

Je remercie également le reste de ma famille avec à leur tête mon oncle Aiouache Salah dont l'aide généreuse et le soutien ne se sont jamais démentis de même que sa femme Latifa, mon autre oncle Abdelaziz Mohamed et son épouse, avec une mention particulière à mon adorable grand-mère Habiba.

Soraya

ملخص

إن تنقية المياه المستعملة تعتبر وسيلة ضرورية للحد من ارتفاع تركيز المواد الملوثة. المشاكل البيئية التي يسببها صرف المياه من محطات التنقية وكذا ارادة التحسين والحفاظ على جودة الوسط المستقبل تستلزم رفع أداء المحطات من ناحيتي الجودة والفعالية.

الهدف من هذه الدراسة هو تصميم الطاقة المستهلكة من طرف محطة معالجة مياه الصرف بومرداس، وذلك بالإعتماد على ثلاثة نماذج مختلفة: نموذج تحليل المكون الرئيسي، نموذج الذكاء الإصطناعي و التقارن للتحليل المكون الرئيسي و الذكاء الإصطناعي ثم القيام بالإنحدار الخطي بين الطاقة المحاكاة بكل واحد من هذه النماذج و الطاقة الخام من أجل إستنتاج أي هذه النماذج أفضل.

الكلمات المفتاحية : معالجة مياه الصرف الصحي، تحليل المكون الأساس، الذكاء الإصطناعي، التقارن للتحليل المكون الأساسي و الذكاء الإصطناعي، تصميم الطاقة، الانحدار الخطي.

Abstract :

Now days, water purification is essential to face the tough consequences of the existence of polluting substances in nature. Taking care of environmental problems caused by waste water treatment plant (WWTP) rejections and the will to preserve Receivers Lake and streams has lead us to look for a way to improve the efficiency of WWTP economically and environmentally. The purpose of our work is the modelization of the energy of the WWTP Boumerdes using three different models: PCA , Neuro- Fuzzy and Neuro- Fuzzy PCA coupling then make a linear regression between energy simulated by each of these three models and raw power to decide on the quality of the model.

keyword : Wastewater treatment , principal component analysis (PCA) , Neur – fuzzy, coupling PCA Neuro- Fuzzy , energy modelization, linear regression.

Résumé

L'épuration des eaux usées est aujourd'hui indispensable pour faire face à l'augmentation des concentrations substances polluantes. La prise de conscience des problèmes environnementaux causés par les rejets des stations d'épuration et la volonté d'améliorer et de préserver la qualité des milieux récepteurs a entraîné l'obligation d'augmenter les performances des stations d'épuration, tant en terme de qualité de traitement qu'en terme de fiabilité. L'objet de notre travail, est la modélisation de l'énergie de la STEP de Boumerdes par trois différents modèles : ACP, Neuro-floue et couplage ACP Neuro-floue ensuite faire une régression linéaire entre l'énergie simulée par chacun de ces trois modèles et l'énergie brute afin de décider sur la qualité du modèle.

Mots clés : Epuration des eaux usées, analyse en composantes principales (ACP), Neur-floue, couplage ACP Neuro-floue, modélisation de l'énergie, régression linéaire.

Sommaire

Liste des figures	
Liste des tableaux	
Introduction général.....	16
Chapitre I. Généralités sur le fonctionnement des stations d'épuration	
I. Introduction	18
I.2.1. Les eaux usées industrielles.....	18
I.2.2. Les eaux usées domestiques.....	18
I.2.3. Les eaux usées pluviales.....	19
I.2.4. Les eaux usées agricoles.....	19
I.3.1 Les paramètres physico-chimiques.....	19
I.3.1.1 La température.....	19
I.3.1.2. Le Potentiel Hydrogène (pH).....	19
I.3.1.3. L'oxygène dissous (OD).....	20
I.3.1.4. Matières en suspension (MES).....	20
I.3.1.5. La conductivité électrique (CE).....	20
I.3.2. Les paramètres chimiques.....	20
I.3.2.1. Demande chimique en oxygène (DCO).....	20
I.3.2.2. Demande biochimique en oxygène (DBO5).....	20
I.3.2.3. Azote Kjeldahl (NTK).....	21
I.3.2.4. Phosphore Total (PT).....	21
I.3.3. Paramètres toxiques.....	21
I.3.3.1. Métaux lourds.....	21
I.3.3.2 Tensioactifs.....	21
I.3.3.3. Hydrocarbures.....	22
I.3.4. Les micro-organismes.....	22
I.3.4.1. Les bactéries.....	22
I.3.4.2. Les virus.....	22
I.3.4.3. Les parasites.....	22
I.3.4.4. Les champignons microscopiques.....	22
I.4. Les critères de la pollution organique.....	23

I.4.1. L'équivalent habitant (EH).....	23
I.5.1. Sur le milieu naturel.....	23
I.5.1.1. Océan.....	23
I.5.1.2. Eaux superficielles.....	23
I.5.1.3. Eaux souterraines.....	23
I.5.2. Sur la santé de l'homme.....	23
I.6.1. Les différents réseaux de collecte.....	24
I.6.1.1. Les réseaux unitaires.....	25
I.6.1.2. Les réseaux séparatifs.....	25
I.6.2. L'entretien du système de collecte.....	26
I.7. Le fonctionnement d'une station d'épuration.....	26
I.7.1. Définition.....	27
I.7.2. Les étapes d'épuration.....	27
I.7.2.1. Prétraitement.....	27
I.7.2.2. Traitement primaire.....	28
I.7.2.3. Traitement secondaire.....	29
I.7.2.4. Traitement tertiaire.....	31
I.7.2.5. Traitement des boues.....	32

Chapitre II. Théorie de l'Analyse en Composantes Principales

II.1. Introduction.....	36
II.2. Analyse en composantes principales.....	36
II.2.1. Définitions.....	36
II.2.2. Objectifs.....	39
II.2.3. Domaine d'application.....	40
II.2.4. Cadre.....	40
II.2.5. Procédé d'application de l'ACP.....	40
II.2.6. Résultats d'application.....	41
II.2.7. Analyse des résultats.....	41
II.2.8. La représentation graphique.....	43
II.2.9. Influence de différentes transformations des variables.....	44
II.3. Présentation du logiciel XLSTAT.....	48
II.3.1. Caractéristiques du logiciel.....	48

II.3.2. La démarche à suivre pour faire une régression avec XLSTAT.....	51
---	----

Chapitre III. Théorie de la Neuro-floue

III.1. Introduction.....	54
III.2. Identification floue.....	54
III.3. Définitions de base.....	55
III.3.1. Les réseaux de neurones.....	55
III.3.2. La logique floue.....	55
III.3.3. Les sous-ensembles flous.....	55
III.3.4. La fonction d'appartenance.....	56
III.3.5. Les variables linguistiques.....	57
III.3.6. Le raisonnement en logique floue.....	57
III.3.7. Système d'inférence flou (SIF).....	57
III.3.7.1. La fuzzification.....	58
III.3.7.2. La défuzzification.....	58
III.3.8. Synoptique d'un système flou.....	58
III.3.9. Type des systèmes d'inférence.....	59
III.3.9.1. Système d'inférence de type Mamdani (linguistique).....	59
III.3.9.2. Système d'inférence de type Takagi Sugeno Kang (TSK).....	59
III.4. Système hybride neuro-flou [28].....	61
III.4.1. Emergence des systèmes hybrides pour la prédiction.....	61
III.4.2. Architectures neuro-floues.....	63
III.4.2.1. L'outil hybride ANFIS.....	64
III.4.2.2. Modélisation.....	66
III.4.3. Améliorations potentielles d'un système ANFIS.....	67
III.4.3.1. Architecture du réseau.....	67
III.4.3.2. Base d'apprentissage.....	68
III.4.3.3. Algorithme d'apprentissage.....	69
III.4.4. Les critères d'évaluation d'un modèle.....	69
III.4.4.1. Le critère de Nash Sutcliffe (Nash et Sutcliffe, 1970).....	70
III.4.4.2. L'erreur quadratique moyenne (RMSE: Root Mean Square Error).....	70
III.4.4.3. Percent bias PBIAS.....	71
III.4.4.4. Le coefficient de détermination.....	71

Chapitre IV. Présentation de la STEP de Boumerdes

IV.1. Introduction.....	74
IV.2. Localisation de la STEP de Boumerdes.....	74
IV.3. Caractéristiques techniques.....	75
IV.4. Paramètres de pollution de conception.....	75
IV.5. Les étapes du procédé d'épuration.....	76
IV.5.1. Principe du traitement.....	76
IV.5.1.1. Un traitement au fil de l'eau.....	77
IV.5.1.2. Traitement au fil de boues.....	80
IV.6. Statistiques et critiques des données.....	81
IV.6.1. Statistiques descriptives.....	81
IV.6.2. Evolution des paramètres de pollution dans le temps.....	83
IV.6.3. Détermination des valeurs singulières.....	95
IV.6.3.1. Test des horsains de Grubbs et Beck.....	96
IV.6.3.2. Résultat du test pour la STEP de Bomerdes.....	96

Chapitre V. Modélisation par ACP et Neuro-flou

V.1. Introduction.....	100
V.2. Procédure de l'Analyse en Composantes Principales.....	100
V.2.1. Données utilisées.....	100
V.3. APPLICATION.....	100
V.3.1. Application de l'ACP sur les paramètres de pollution d'entrée (Eau brute).....	100
V.3.1.1. Pas de temps journalier.....	100
V.3.1.2. Pas de temps mensuel.....	105
V.3.2. Application de l'ACP sur les paramètres de l'eau (Entrée/Sortie).....	109
V.3.2.1. Pas de temps journalier.....	109
V.3.2.2. Pas de temps mensuel.....	111
V.4. Modélisation de la STEP de Boumerdes.....	112
Introduction.....	112
V.4.1. La modélisation par ACP.....	113
V.4.1.1. Application de l'ACP.....	113

V.4.1.2. Modèle issu de l'ACP	116
V.4.2. Modélisation par Neuro-Floue (Modèle ANFIS).....	121
V.4.2.1. Elaboration du modèle Neuro-floue.....	121
V.4.3. Couplage ACP Neuro-floue.....	126
V.5. Conclusion	131
Conclusion générale	132
Bibliographie.....	133

Liste des figures

Chapitre I :

Figure I. 1 : Représentation schématique d'un tensioactif.....	22
Figure I. 2 : Réseau unitaire.....	25
Figure I. 3 : Réseau séparatif.....	25
Figure I. 4 : Schéma représentant les étapes du traitement d'eau usée.....	26
Figure I. 5 : Schéma représentant le prétraitement de l'eau usée.....	28
Figure I. 6 : Schéma qui illustre le principe de fonctionnement d'un décanteur.....	29
Figure I. 7 : Schéma qui illustre l'étape du traitement biologique secondaire.....	30
Figure I. 8 : L'épaississement des boues.....	32
Figure I. 9 : Schéma illustrant la déshydratation par centrifugation.....	33
Figure I. 10 : Lits de séchage.....	34

Chapitre II :

Figure II. 1: Interface de XLSTAT.....	48
Figure II. 2. Lancement d'une analyse en composantes principales.....	49
Figure II. 3 : Insertion du tableau d'observations.....	50
Figure II. 4 : Choix des sorties.....	50
Figure II. 5 : Validation.....	51
Figure II. 6 : Lancement d'une régression linéaire.....	51
Figure II. 7 : Insertion des variables et des individus.....	52

Chapitre III :

Figure III. 1 : Fonction d'appartenance caractérisant les sous-ensembles des variables d'entrées et de sorties.....	56
Figure III. 2 : Partitionnement de l'espace des entrées.....	68

Chapitre IV :

Figure IV. 1 : STEP de Boumerdes (Google Earth).....	74
Figure IV. 2 : Vue d'ensemble des ouvrages de la STEP de Boumerdes.....	75
Figure IV. 3 : Organigramme d'exploitation de la STEP de Boumerdes.....	76
Figure IV. 4 : Bassin d'orage de la STEP de Boumerdes.....	77
Figure IV. 5 : Dégrilleur du type à champ courbe.....	78
Figure IV. 6 : Dessableur-Déshuileur.....	78
Figure IV. 7 : Bassin d'aération.....	79
Figure IV. 8 : Décanteur-Clarificateur.....	79

Figure IV. 9 : Bassin de chloration.....	80
Figure IV. 10 : Presse à bande.....	81
Figure IV. 11 : Evolution journalière du PH de l'eau brute.....	83
Figure IV. 12 : Evolution journalière du PH de l'eau épurée.....	83
Figure IV. 13 : Variation journalière de la température brute.....	84
Figure IV. 14 : Variation journalière de la température de l'eau épurée.....	84
Figure IV. 15 : Evolution journalière de la concentration en MES à l'entrée de la STEP.....	85
Figure IV. 16 : Evolution journalière de la concentration en MES à la sortie de la STEP.....	85
Figure IV. 17 : Evolution de la Demande Biologique en Oxygène à l'entrée de la STEP.....	86
Figure IV. 18 : Evolution de la Demande Biologique en Oxygène à la sortie de STEP.....	87
Figure IV. 19 : Evolution de la Demande Chimique en Oxygène à la sortie de la STEP.....	88
Figure IV. 20 : Evolution de la Demande Chimique en Oxygène à la sortie de la STEP.....	88
Figure IV. 21 : Evolution journalière de l'ammonium (NH ₄ ⁺) à l'entrée de la STEP.....	89
Figure IV. 22 : Evolution journalière de l'ammonium (NH ₄ ⁺) à la sortie de la STEP.....	89
Figure IV. 23 : Evolution journalière du Nitrite (NO ₂ ⁻) à l'entrée de la STEP.....	90
Figure IV. 24 : Evolution du Nitrite (NO ₂ ⁻) à la sortie de la STEP.....	90
Figure IV. 25 : Evolution du Nitrate (NO ₃ ⁻) à l'entrée de la STEP.....	91
Figure IV. 26 : Evolution du Nitrate (NO ₃ ⁻) à la sortie de la STEP.....	91
Figure IV. 27 : Evolution du NTK à l'entrée de la STEP.....	92
Figure IV. 28 : Evolution du NTK à la sortie de la STEP.....	92
Figure IV. 29 : Evolution de PO ₄ -3 à l'entrée de la STEP.....	93
Figure IV. 30 : Evolution de PO ₄ -3 à la sortie de la STEP.....	93
Figure IV. 31 : Evolution journalière du débit dans la STEP.....	94
Figure IV. 32 : Evolution journalière de l'énergie dans la STEP.....	95

Chapitre V :

Figure V. 1 : Paramètres de pollution de l'eau brute à l'échelle journalière (1X2).....	103
Figure V. 2 : Paramètres de pollution de l'eau brute à l'échelle journalière (1X3).....	104
Figure V. 3 : Paramètres de pollution de l'eau brute à l'échelle mensuelle (1X2).....	107
Figure V. 4 : Paramètres de pollution de l'eau brute à l'échelle mensuelle (1X3).....	108
Figure V. 5 : Paramètres de pollution (Entrée-sortie) à l'échelle journalière (1X2).....	111
Figure V. 6 : Paramètres de pollution (Entrée – sortie) à l'échelle mensuelle (1X2).....	112
Figure V. 7 : Régression linéaire (Energie brute- Energie simulée par ACP) / calage.....	119
Figure V. 8 : Régression linéaire (Energie brute- Energie simulée par ACP) / validation....	120

Figure V. 9 : Icône Neuro-floue dans MATLAB.....	122
Figure V. 10 : Interface de l'application Neuro-fuzzy design toolbox.....	122
Figure V. 11 : Interface du choix de la fonction d'appartenance et du nombre de règles floues.....	123
Figure V. 12 : Aperçu du programme Neuro-floue.....	123
Figure V. 13 : Régression linéaire (Energie brute-Energie simulée par NF) / calage.....	124
Figure V. 14 : Régression linéaire (Energie brute- Energie simulée par NF)/ validation.....	125
Figure V. 15 : Régression linéaire (Energie brute-Energie simulée par couplage) /calage...	129
Figure V. 16 : Régression linéaire (Energie brute- Energie simulée par couplage)/ validation.....	129

Liste des tableaux

Chapitre IV :

Tableau IV. 1 : Paramètres de pollution de conception de la station de Boumerdes.....	75
Tableau IV. 2 : Statistique descriptives des paramètres de pollution de la STEP de Boumerdes.....	82
Tableau IV. 3 : Résultats du test de Grubbs et Beck pour la STEP de Boumerdes.....	96
Tableau IV. 4 : Les valeurs singulières détectées.....	97

Chapitre V :

Tableau V. 1 : Codification des variables (Eau brute).....	101
Tableau V. 2 : Matrice de corrélation des paramètres de l'eau brute (journalier).....	101
Tableau V. 3 : Test de sphéricité de Bartlett et K-M-O de l'eau brute (journalier).....	102
Tableau V. 4 : Variance totale expliquée (paramètres de l'eau brute-journalier).....	102
Tableau V. 5 : Matrice de corrélation entre les CP' et les variables de l'eau brute.....	103
Tableau V. 6 : Matrice de corrélation des paramètres de l'eau brute (mensuel).....	105
Tableau V. 7 : Test de sphéricité de Bartlett et K-M-O de l'eau brute (mensuel).....	106
Tableau V. 8 : Variance totale expliquée (paramètres de l'eau brute-mensuel).....	106
Tableau V. 9 : Matrice de corrélation entre les CP' et les variables de l'eau brute-échelle mensuelle.....	107
Tableau V. 10 : Codification des variables (Eau brute- Eau épurée).....	109
Tableau V. 11 : Test de sphéricité de Bartlett et K-M-O (Eau brute –Eau épurée / journalier).....	109
Tableau V. 12 : Variance totale expliquée (Eau brute – Eau épurée / journalier).....	110
Tableau V. 13 : Matrice de corrélation entre les CP' et les variables (Eau brute-Eau épurée / journalier).....	110
Tableau V. 14 : Codification des variables de rabattement.....	114
Tableau V. 15 : Test de sphéricité de Bartlett et K-M-O (rabattement des variables).....	114
Tableau V. 16 : Variance totale expliquée (Rabattement des variables).....	115
Tableau V. 17 : Matrice des cosinus carrés des variables.....	115
Tableau V. 18 : Matrice des vecteurs propres.....	117
Tableau V. 19 : Résultats des critères de validation- Simulation par ACP (calage/validation).....	120

Tableau V. 20 : Résultats des critères de validation- Simulation par Neuro-floue (calage/validation).....	126
Tableau V. 21 : Résultats des critères de validation- Simulation par couplage ACP Neuro- floue (calage/validation).....	130

Introduction générale

Le traitement des eaux usées s'effectue au niveau des stations d'épuration. De nos jours on s'intéresse de plus en plus à ce domaine, car cela permet de nettoyer les eaux usées (eaux usées ménagères, industrielles et parfois pluviales) afin de rendre une eau propre au milieu naturel.

Le plus grand souci des gérants de station d'épuration est de chercher comment optimiser le fonctionnement de cette dernière, par exemple, comment assurer le meilleur fonctionnement de la STEP en tenant compte de l'aspect économique et temporel.

Nous avons essayé à travers ce mémoire de modéliser l'énergie de la station d'épuration de Boumerdes en utilisant trois modèles différents : l'ACP, la méthode Neuro-floue et le couplage ACP Neuro-floue dans le but de choisir le meilleur modèle qui nous permettra d'optimiser l'énergie de cette STEP.

On a entamé ce mémoire par une introduction générale où on a mis l'accent sur l'importance à accorder au sujet de l'épuration des eaux usées ainsi qu'à la protection du milieu récepteur des effets désastreux engendré par la pollution de cette ressource.

Le premier chapitre n'est qu'une synthèse des différents agents et paramètres qui entrent en jeu dans l'analyse de l'eau avant et après son traitement et une décortication des différentes étapes du procédé de l'épuration des eaux usées ainsi que les formes que peut prendre une boue suite à son traitement.

On a expliqué dans le deuxième chapitre les bases mathématiques de la méthode de l'analyse en composantes principales (ACP).

Dans le troisième chapitre on a parlé de la théorie Neuro-floue qui permet d'intégrer dans un seul concept les avantages des deux domaines : Logique floue et Réseaux de neurones.

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

L'élaboration des trois modèles, la présentation des résultats obtenus et leur interprétation ont fait l'objet de ce dernier chapitre.

Enfin, une conclusion générale, dans laquelle on a fait valoir les résultats obtenus de recherche, tout en précisant la méthode suivie et l'objectif principal de cette étude par ce travail.

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

I.1. Introduction

L'eau est une ressource précieuse qui nous est offerte par Dieu. Il nous est donc impératif de la préserver et de bien la gérer afin qu'elle ne soit pas gaspillée ou menacée. On doit donc faire face aux problèmes qui la menacent de manière économiquement efficace et durable.

La dépollution des eaux usées est devenue un impératif pour nos sociétés modernes. En effet, le développement des activités humaines s'accompagne inévitablement d'une production croissante de rejets polluants. L'assainissement des eaux usées (ou dépollution) répond à deux préoccupations essentielles : préserver les ressources en eau, préserver le patrimoine naturel et la qualité de la vie.

I.2. Les différents types d'eaux usées

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

I.2.1. Les eaux usées industrielles

Presque toutes les eaux de rejets résultant d'une utilisation de l'eau autre que domestique qualifiées de rejets industriels. Cette définition concerne les rejets des usines ainsi que les rejets d'activités artisanales ou commerciales. Les eaux industrielles peuvent contenir des produits toxiques, des métaux lourds, des polluants organiques et des hydrocarbures. [31]

I.2.2. Les eaux usées domestiques

Constituant généralement l'essentiel de la pollution, elles se composent :

- Des eaux vannes d'évacuation des toilettes.
- Des eaux ménagères d'évacuation des cuisines, salles de bain.

Les déchets présents dans ces eaux souillées sont constitués par des matières organiques dégradables et des matières minérales. Ces substances sont sous forme dissoute ou en suspension. [31]

I.2.3. Les eaux usées pluviales

Ce sont les eaux pluviales ou de ruissellement urbain (lavage des chaussées, des marchés et des rues). Les eaux de service public contiennent des matières solides, graisses et huiles de lavages des chaussées. [31]

I.2.4. 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.

I.3. Les différentes caractéristiques des eaux usées

I.3.1 Les paramètres physico-chimiques

I.3.1.1 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. [32]

Les propriétés de décantation : quand la température augmente, l'indice de décantation augmente car la densité des floccs diminue. [32]

La toxicité : l'effet toxique des substances chimiques est plus grand à température élevée. [33]

La demande biologique en oxygène : une augmentation de la température conduit à une augmentation de la consommation de l'O₂ dissous. [33]

I.3.1.2. Le Potentiel Hydrogène (pH)

Le PH (potentiel Hydrogène) mesure la concentration en ions H⁺ de l'eau, il traduit ainsi la balance entre acide et base sur une échelle de 0 à 14.

Il influence la plupart des mécanismes chimiques et biologiques dans les eaux, son rôle est capital pour la croissance des micro-organismes qui ont généralement un pH optimum variant de 6,5 à 7,5. Lorsque le pH est inférieur à 5 ou supérieur à 8,5, la croissance des micro-organismes est directement affectée. [34]

I.3.1.3. 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.

I.3.1.4. Matières en suspension (MES)

Les matières en suspension (MES) sont des matières non dissoutes qui comportent à la fois des éléments minéraux (MM) et organiques (MVS), elles donnent à l'eau une apparence trouble et souvent un mauvais goût et une mauvaise odeur.

Par leurs effet obscurcissant, les MES présentes dans les eaux diminuent l'efficacité de la photosynthèse qui contribue à l'aération de l'eau, les organismes vivants peuvent alors manquer d'oxygène ce qui baisse la productivité du milieu récepteur.[35]

I.3.1.5. 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).

I.3.2. Les paramètres chimiques

I.3.2.1. Demande chimique en oxygène (DCO)

La demande chimique en oxygène (DCO) représente la quantité d'oxygène qui permettra l'oxydation de toutes les substances organiques présentes dans l'eau. Elle a un pouvoir auto-épuration qui lui permet d'être un des paramètres de la qualité d'une eau.

I.3.2.2. Demande biochimique en oxygène (DBO₅)

Quantité d'oxygène consommée (exprimée en mgO₂/L) dans les conditions d'essai (incubation à 20 °C et à l'obscurité après un laps de temps de 5 jours) pour assurer par voie bactérienne l'oxydation biologique des matières organiques dites biodégradables.

I.3.2.3. 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]

I.3.2.4. 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).[2]

I.3.3. Paramètres toxiques

I.3.3.1. Métaux lourds

On appelle métaux lourds les éléments métalliques naturels dont la masse volumique dépasse 5g/cm³. Ceux-ci sont présents le plus souvent dans l'environnement sous forme de traces : mercure, plomb, cadmium, cuivre, arsenic, nickel, zinc, cobalt, manganèse. Les plus toxiques d'entre eux sont le plomb, le cadmium et le mercure.[3]

I.3.3.2 Tensioactifs

Une molécule tensioactive se définit par la présence conjointe d'une partie hydrophile et d'une partie hydrophobe, qui lui donne son caractère amphiphile (Figure I.1). C'est l'antagonisme de cette structure qui lui confère ses propriétés particulières [20]

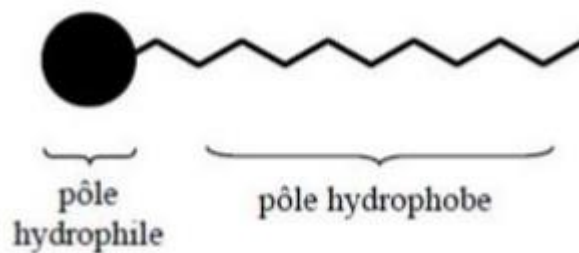


Figure I. 1 : Représentation schématique d'un tensioactif

I.3.3.3. Hydrocarbures

La production et la consommation d'hydrocarbures connaissent un développement extraordinaire. On a malheureusement omis jusqu'à présent d'accorder une attention suffisante à la récupération ou à l'élimination des hydrocarbures contenus dans les eaux résiduaires et on les retrouve pour une bonne part, dans le milieu naturel. Les origines de leur présence dans les eaux sont multiples : rejets de produits pétroliers (garage, industrie pétrolière, navigation...), effluents des usines à gaz, fumées dont les particules sont entraînées par les eaux météoriques, etc...[4]

I.3.4. Les micro-organismes

I.3.4.1. Les bactéries

Parmi les micro-organismes présents dans les eaux usées on trouve les bactéries responsables de troubles digestifs, de troubles respiratoires ou encore de surinfection des plaies. [8]

I.3.4.2. Les virus

Il y'a des centaines de virus qui peuvent contaminés les eaux usées ainsi que les boues, mais les plus représentées sont de gastro-entérites ainsi que le virus de l'hépatite A.

I.3.4.3. Les parasites

Tels que les œufs vers intestinaux qui peuvent se trouver dans l'eau ou dans les boues stabilisées.

I.3.4.4. Les champignons microscopiques

Ils sont aussi identifiés dans les eaux, les boues et l'air ambiant de station d'épuration. Ces champignons peuvent être responsables d'irritations ou d'allergie.

I.4. Les critères de la pollution organique

I.4.1. L'équivalent habitant (EH)

Un EH est défini par la pollution de l'eau générée par un individu en une journée. La quantité de pollution journalière produite par un individu est estimée à 57g de matières oxydables (MO), 90g de matières en suspension (MES), 15g de matières azotées (MA), et 4g de matières phosphorées (MP). Enfin, la concentration des germes est généralement de l'ordre de 1 à 10 milliards de germes pour 100 ml. [5]

I.5. Les impacts des eaux usées

Les eaux usées n'ont pas seulement un impact sur le milieu naturel mais aussi la santé de l'homme.

I.5.1. Sur le milieu naturel

I.5.1.1. Océan

En général, les eaux rejetées dans la mer ne sont pas épurées ou le traitement effectué n'est pas suffisant, ce qui cause la pollution de ces sites et provoque leur contamination par des bactéries, des virus ou parasites issues des eaux usées.

I.5.1.2. Eaux superficielles

Il arrive que les eaux usées soient déversées directement vers le milieu naturel. La présence excessive du phosphate provoque la diminution de l'oxygène dans les milieux aquatiques et par conséquent la mort des poissons.

I.5.1.3. Eaux souterraines

La qualité de l'eau des nappes phréatiques peut être dégradée par les eaux usées, si l'étanchéité de la station d'épuration ou de la lagune est défectueuse ou lorsque le système d'assainissement non collectif présente des dysfonctionnements.[6]

I.5.2. Sur la santé de l'homme

L'eau, ressource naturelle indispensable à la vie, est aussi devenue, de manière directe ou indirecte, la première cause de mortalité et de maladie au monde. L'inégalité dans la répartition des ressources en eau associée à la dégradation de la qualité de l'eau engendrent de

grands problèmes de santé. Ainsi, dans les pays en développement, 80 % des maladies sont dues à l'eau, un Africain sur deux souffre d'une maladie hydrique.[6]

I.6. La collecte des eaux usées

Tout immeuble d'habitation doit être raccordé à un réseau d'assainissement collectif ou doit avoir son propre système d'assainissement propre à lui pour collecter les eaux usées ensuite les traiter avant de les rejeter dans le milieu naturel.

I.6.1. Les différents réseaux de collecte

L'assainissement des eaux usées est l'un des enjeux actuels pour garder un environnement sain et assurer un approvisionnement en eau potable pour les années à venir. A ce titre il existe deux types d'assainissement selon le type d'habitation et les choix par la collectivité : [36]

- L'assainissement collectif.
- L'assainissement individuel.

L'assainissement collectif : est assuré par un réseau de collecte (égouts) qui recueillent les eaux usées et les acheminent vers les stations d'épurations.

L'assainissement individuel : (dit « assainissement autonome ») pour des habitations ou de petits groupes d'habitations trop éloignés pour être relié aux réseaux collectifs. Une fosse septique récupère les eaux usées.

Le système d'assainissement peut présenter plusieurs variantes de conception, on distingue : [36]

- Le système unitaire.
- Le système séparatif.

I.6.1.1. Les réseaux unitaires

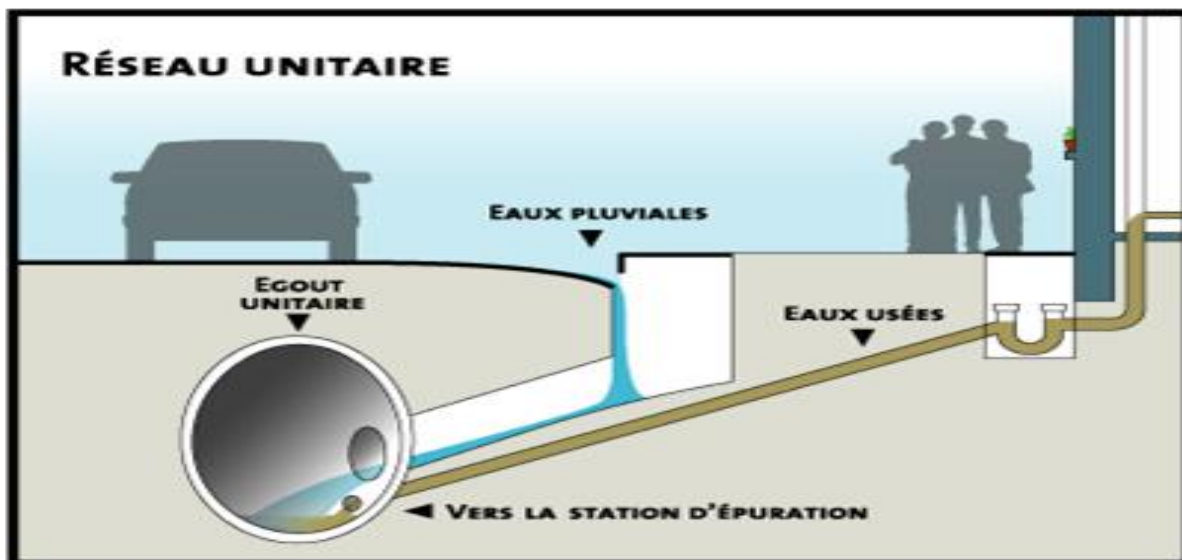


Figure I. 2 : Réseau unitaire

Dans ce système, un seul collecteur assure à la fois le transport des eaux usées et des eaux pluviales jusqu'à la station d'épuration. Il est de conception simple, mais il ne permet pas de maîtriser la nature de l'effluent arrivant à la station d'épuration, selon la pluviométrie le débit est très variable et l'effluent « eaux usées » est dilué par le débit pluvial.

I.6.1.2. Les réseaux séparatifs

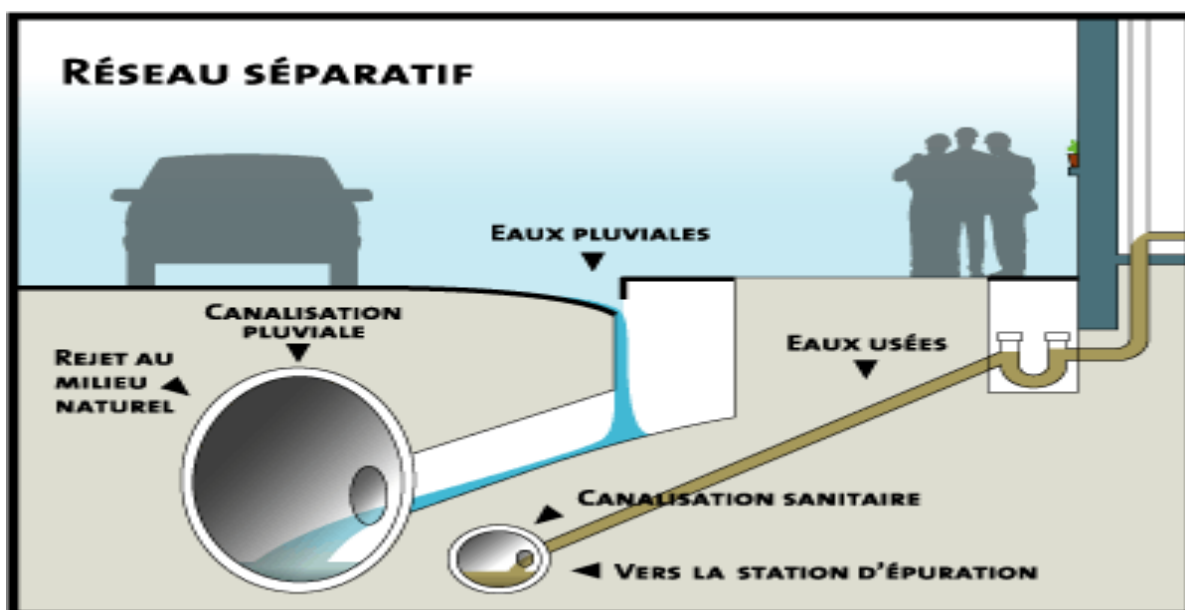


Figure I. 3 : Réseau séparatif

Pour la collecte et le transport des eaux, deux réseaux distincts ont été mis en place :

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

- L'un pour évacuer les eaux de pluie.
- L'autre pour recueillir les eaux usées.

Dans ce système, le réseau « Eaux usées » sera dirigé vers la station d'épuration, par contre, le réseau « Eaux pluviales » rejoint directement la rivière : les eaux de pluie ne seront pas épurées.

Ce système permet de mieux maîtriser la qualité et la quantité d'effluent arrivant à la station d'épuration.

I.6.2. L'entretien du système de collecte

Les canalisations transportent en permanence des eaux chargées de pollution. Il est inévitable qu'elles s'encrassent. Elles sont également menacées par l'érosion et la corrosion. Elles font donc l'objet de visites et de curages périodiques. L'entretien des ouvrages annexes comprend notamment le curage des bouches d'égout, l'entretien des bassins de retenue, des déversoirs d'orage, des postes de relèvement des eaux usées et des branchements. Une exploitation efficace du réseau suppose un travail d'entretien rigoureux et permanent du personnel qualifié du service d'assainissement.[7]

I.7. Le fonctionnement d'une station d'épuration

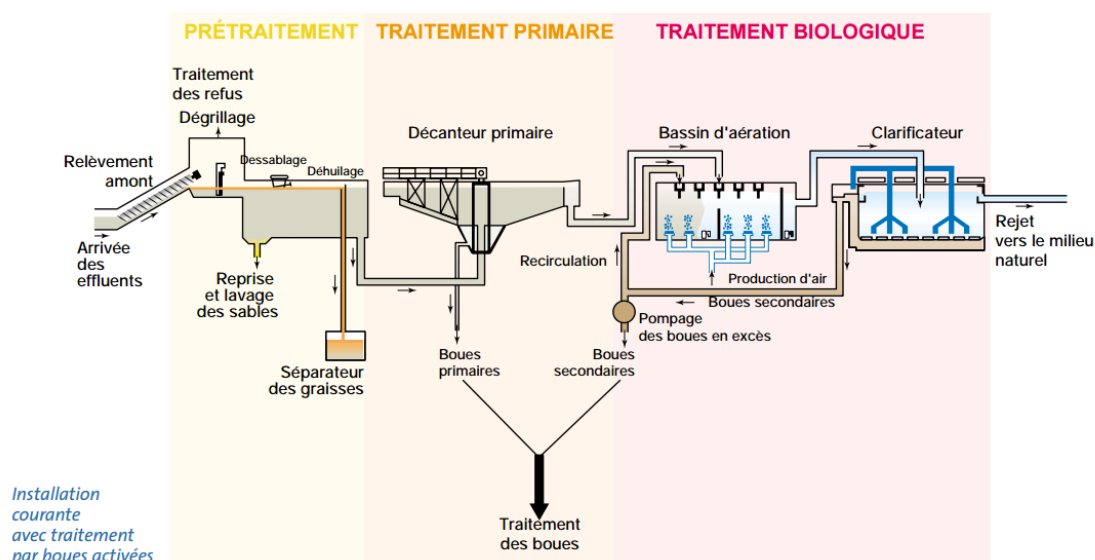


Figure I. 4 : Schéma représentant les étapes du traitement d'eau usée

I.7.1. 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. [37]

I.7.2. Les étapes d'épuration

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 :

- Prétraitements
- Traitement primaire
- Traitement secondaire
- Traitement tertiaire
- Traitement des boues

I.7.2.1. Prétraitement

Le prétraitement a pour objectif l'extraction des matières les plus grossières (brindilles, feuilles, tissus,...) et des éléments susceptibles de gêner les étapes ultérieures du traitement.

Il passe par trois étapes :

I.7.2.1.1. 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.

I.7.2.1.2. 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.

I.7.2.1.3. 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.

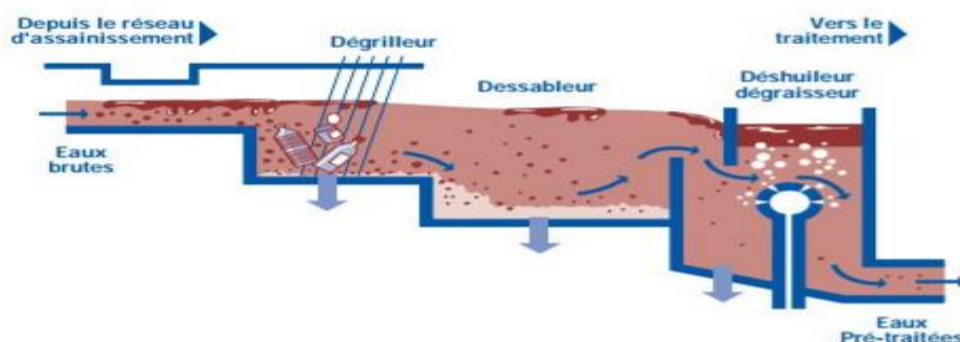


Figure I. 5 : Schéma représentant le prétraitement de l'eau usée

I.7.2.2. Traitement primaire

Le traitement primaire permet d'éliminer jusqu'à 70% environ de matières minérales et organiques en suspension et les déposer au fond, elles forment les boues primaires qui sont récupérées et traitées dans des épaisseurs.

Il fait appel à différents procédés physiques : [22]

- **La flottation**

La flottation vise à séparer les phases solides des phases liquides par la poussée d'Archimède. En flottation naturelle, les floccs de faible densité remontent librement à la surface. La flottation assistée s'obtient par injection d'air. [22]

- **La décantation (physique)**

Le principe de séparation solide-liquide est la pesanteur, les matières en suspension ou colloïdales tendent à se séparer du liquide par sédimentation. [23]

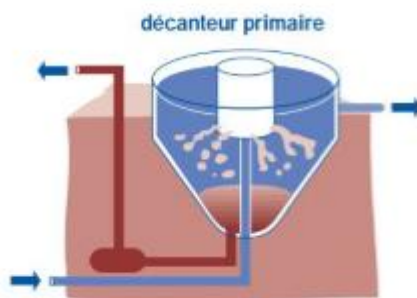


Figure I. 6 : Schéma qui illustre le principe de fonctionnement d'un décanteur

- **La décantation associée à l'utilisation d'un coagulant- flocculant (voie physicochimique)**

Le principe est ici de favoriser l'agrégation des molécules en suspension grâce aux techniques de coagulation et de floculation de façon à augmenter la sédimentation grâce à l'obtention de floccs plus gros. [23]

- **La filtration**

C'est le passage du mélange liquide-solide à travers un milieu poreux (filtre) qui retient les solides et laisse passer les liquides. [22]

I.7.2.3. Traitement secondaire

Dans cette étape la pollution est dégradée grâce aux microorganismes présents dans l'effluent de façon naturelle. Ces microorganismes sont nourris de la pollution organique présente et l'oxygène est fourni par l'air des turbines.

La clarification permet de séparer par décantation l'eau épurée des boues secondaires issues du traitement biologique. Cette décantation se fait dans des ouvrages spéciaux, le plus souvent circulaires, appelées clarificateurs ou décanteurs secondaires.[9]

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

Une partie des boues secondaires est évacuée en aval vers le traitement des boues, l'autre partie est recyclée vers le bassin d'aération pour maintenir la masse biologique nécessaire au fonctionnement de l'installation.[9]

On utilise typiquement :

- **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). [22]

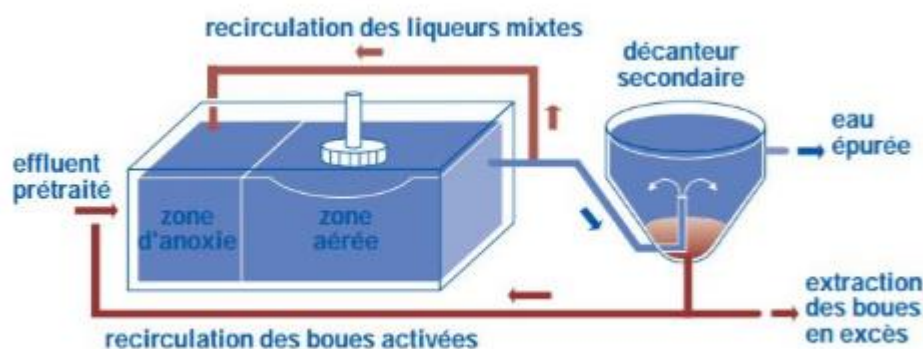


Figure I. 7 : Schéma qui illustre l'étape du traitement biologique secondaire

I.7.2.4. Traitement tertiaire

Si le milieu récepteur exige une épuration meilleure, il est nécessaire d'effectuer des traitements complémentaires tels que : la désinfestation 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.

Différentes méthodes de nature variée peuvent alors être utilisées :

- **Biologiques**

Pour l'élimination de l'azote (processus de nitrification-dénitrification) et Biologiques du phosphore (déphosphatation).

- **Physico-chimiques**

Pour la précipitation du phosphore (coagulation-décantation) ou l'élimination des dernières matières en suspension (filtration sur lits de sable, tamis métalliques ou charbon actif). L'élimination de l'azote et du phosphore par voie biologique ou chimique évite la prolifération de végétaux dans les corps d'eau (les lacs, les étangs ou les rivières) et protège la vie aquatique.

- **Chimiques**

Pour la désinfection dans le traitement final des effluents. L'élimination des risques de contamination bactériologique ou virale se fait souvent par oxydation en utilisant des agents tels que le chlore et l'ozone. Les procédés chimiques sont utilisés aussi pour l'adoucissement de l'eau et pour la neutralisation en agissant sur le pH.

- **Radioactifs**

Pour les opérations de désinfection de l'eau telle que les rayonnements ultraviolets qui irradiant les cellules vivantes indésirables permettant d'éliminer les risques de contamination due aux bactéries et virus. Suivant la quantité d'énergie UV reçue, elles sont soit stérilisées (effet bactériostatique) soit détruites (effet bactéricide).

I.7.2.5. Traitement des boues

Le traitement d'un mètre cube d'eaux usées produit de 350 à 400 grammes de boues. Ces boues, généralement très liquides, contiennent une forte proportion de matières organiques. Elles sont donc très susceptibles de causer des nuisances.

I.7.2.5.1. L'Épaississement



Figure I. 8 : L'épaississement des boues

Afin d'effectuer la première étape du traitement des boues qui est l'épaississement, on ajoute quelques flocculants organiques ou minéraux (chaux, sel de fer...) afin de faciliter la séparation des phases liquides des phases solides des boues.

L'épaississement consiste à séparer les phases solides des phases liquides par gravitation, il existe aussi une autre technique appelée flottation basée sur l'injection de gaz dans les boues et par différence de densité on peut séparer les solides du liquides.

Une fois à la sortie, les boues sont liquides de 4 à 6%.

I.7.2.5.2. La déshydratation

Une fois que les boues primaires, secondaires ou même tertiaires sont mélangées, on leur applique la déshydratation afin de leur donner une siccité allant de 15 à 40%.

Il existe deux types de déshydratation :

- **La déshydratation mécanique**

Elle s'opère par centrifugation ou par filtration.

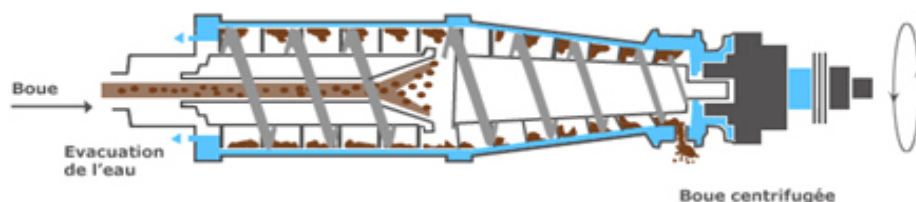


Figure I. 9 : Schéma illustrant la déshydratation par centrifugation

La centrifugation consiste à séparer l'eau des boues épaissies par la force centrifuge développée dans un cylindre tournant à grande vitesse. En sortie, les boues sont pâteuses avec une siccité de 18 à 20 % pour la première génération d'équipements, et de 20 à 25 % de siccité pour la seconde. Pendant longtemps, cette technique a surtout concerné les stations de plus de 10.000 EH ; aujourd'hui des solutions existent pour les plus petites.[10]

La filtration par filtres à bandes consiste en une compression et un cisaillement des boues entre deux toiles. Les premiers modèles (à basse et moyenne pression) ne permettaient d'atteindre que 15 à 17% de siccité. Les modèles plus récents (à haute pression) permettent d'atteindre jusqu'à 18 à 20%. En sortie, les boues se présentent sous forme de petites plaques.[10]

- **La déshydratation géomembrane**

C'est une technique récente qui permet de déshydrater les boues jusqu'à 15 à 25% de siccité par l'usage des tubes aux pores minuscules qui permettent à l'eau de passer petit à petit. En fin d'opération les tubes sont ouverts et les boues sont expédiées vers une autre destination.

I.7.2.5.3. Le séchage

Le séchage des boues permet l'évaporation de l'eau et donc une déshydratation quasi-totale des boues. Il existe deux façons de séchage

- **Le séchage thermique**

Cette méthode est très énergivore car elle repose sur le chauffage des parois mais elle permet d'atteindre jusqu'à 90 à 95% siccité.

- **Lits de séchage**

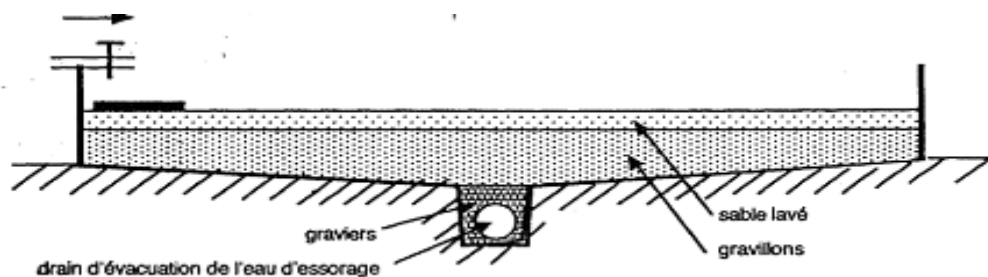


Figure I. 10 : Lits de séchage

Ce procédé permet la déshydratation des boues naturellement en utilisant les rayons de soleil.

Sur des surfaces drainantes composées de graviers et de sables on dépose les boues issues du traitement primaire, secondaire et tertiaire. À l'aide des rayons de soleil le phénomène d'évaporation est accéléré et en sortie les boues ont une siccité d'environ 35 à 40%.

Chapitre II : Théorie de l'Analyse en Composantes Principales

II.1. Introduction

L'Analyse en Composante Principale (ACP) est une méthode d'analyse des données dite méthode factorielle. Cette méthode permet la présentation géométrique d'un tableau rectangulaire composé de p variables et de n individus.

Les données sur lesquelles on applique une ACP sont obtenues soit d'une observation d'un échantillon ou de toute une population.

A partir de la représentation graphique, on peut définir une liaison linéaire entre certaines variables formant un groupe entre elles et donc définir les variables qui sont fortement corrélées ou au contraire, celles aussi qui ne sont pas corrélées entre elles.

La projection des individus sur des axes factoriels permet de voir les individus qui se ressemblent et ceux qui se distinguent.

Enfin, comme pour toute méthode descriptive, réaliser une ACP n'est pas une fin en soi. 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.[24]

II.2. Analyse en composantes principales

II.2.1. Définitions

L'Analyse en Composantes Principales (ACP) ou méthode factorielle est définie comme étant une méthode descriptive multidimensionnelle. Elle permet à partir d'un tableau rectangulaire de données comportant m variables et n observations d'obtenir des représentations géométriques de ces variables et observations.

La représentation géométrique des observations permet de voir s'il existe une structure décrivant l'ensemble des données d'un côté. De l'autre côté, la représentation géométrique des variables permet d'étudier les structures de liaisons linéaires sur l'ensemble des variables considérées.

Grâce à la représentation géométrique des observations et celle des variables, on pourrait voir qu'elles sont ces dernières qui se ressemblent et celles qui se distinguent des autres.

Chapitre II. Théorie de l'Analyse en Composante Principales

- Notion d'individu et de caractère

Individu

L'individu est défini comme étant un élément de la population étudié. Il peut être une année d'observation ou une autre unité de temps.

Un individu est défini dans l'espace R^P , par le vecteur :

$$X_i = \{X_{ij} / j = 1 \text{ à } P\} \quad (\text{II.1})$$

Le terme X_{ij} est un nombre réel qui représente la mesure de la variable X_j relevée sur l'individu i . [17]

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. [17]

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

$$X_j = \{X_{ij} / i = 1 \text{ à } N\} \quad (\text{II.2})$$

- 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 :

$$[X] = \begin{pmatrix} X_{11} & X_{12} & \cdot & \cdot & X_{1j} & \cdot & \cdot & X_{1P} \\ X_{21} & X_{22} & \cdot & \cdot & X_{2j} & \cdot & \cdot & X_{2P} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ X_{i1} & X_{i2} & \cdot & \cdot & X_{ij} & \cdot & \cdot & X_{iP} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ X_{N1} & X_{N2} & \cdot & \cdot & X_{Nj} & \cdot & \cdot & X_{NP} \end{pmatrix} \quad (\text{II.3})$$

Une fois qu'on a la matrice des données, on pourrait définir les paramètres statistiques pour chaque variable :

Chapitre II. Théorie de l'Analyse en Composante Principales

1. La moyenne

$$\overline{X_j} = \frac{1}{N} \sum_{i=1}^N X_{ij} \quad (\text{II.4})$$

$\overline{X_j}$: Moyenne de la $j^{\text{ième}}$ variable.

2. Ecart type

$$S_j = \left[\frac{1}{N} \sum_{i=1}^N (X_{ij} - \overline{X_j})^2 \right]^{1/2} \quad (\text{II.5})$$

S_j = Ecart type de la $j^{\text{ième}}$ variable.

N : Nombre d'individus.

3. Le coefficient de covariance

$$\text{Cov}(X_j, X_k) = \frac{1}{N} \sum_{i=1}^N (X_{ij} - \overline{X_j}) * (X_{ik} - \overline{X_k}) \quad (\text{II.6})$$

4. Le coefficient de corrélation

$$\text{Cor}(X_j, X_k) = \frac{\text{Cov}(X_j, X_k)}{S_j * S_k} = \frac{\sum_{i=1}^N (X_{ij} - \overline{X_j}) * (X_{ik} - \overline{X_k})}{\left[\sum_{i=1}^N (X_{ij} - \overline{X_j})^2 * \sum_{i=1}^N (X_{ik} - \overline{X_k})^2 \right]^{1/2}} \quad (\text{II.7})$$

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

Avec : CP, composantes principales.

Celles-ci (notées ici C_j) sont calculées par détermination de P relations linéaires entre les variable X_j , tel que :

Chapitre II. Théorie de l'Analyse en Composante Principales

- ✓ 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.

II.2.3. Domaine d'application

Les applications sont très nombreuses. Il y a en fait deux façons d'utiliser l'ACP :

- Soit pour l'étude d'une population donnée en cherchant à déterminer la typologie des individus et des variables. Par exemple, en hydrologie, l'ACP est utilisé dans la critique, la reconstitution et/ou la cartographie de données hydrologique, et aussi en analyses de séries chronologiques provenant de réseaux de mesures hydrométéorologiques et hydrogéologiques, et enfin l'élaboration de méthodologie de simulation et de prévision des variables climatologiques aléatoires et cycliques par l'analyse en composantes principales.[18]
- Soit pour réduire les dimensions des données sans perte importante d'information, par exemple en hydrologie, l'ACP intervient dans l'élaboration des bilans énergétiques en expliquant un phénomène précis, en réduisant la quantité de variables qui peuvent l'expliquer et en évitant la redondance (telle que le phénomène de l'évapotranspiration qui peut être influencé par de nombreuses variables telles que la température, l'insolation, la vitesse du vent, l'humidité...etc, l'ACP permet de prendre l'évapotranspiration comme variable à expliquer et trouver les variables explicatives de celle-ci et d'éliminer celles qui n'interviennent pas dans le phénomène).

II.2.4. Cadre

Le tableau sur lequel on applique une analyse en composantes principales, est constitué de « N » individus qui doivent être de même nature et qui sont numérotés de (1 à n).

D'autre part les « P » variables peuvent ne pas être homogènes et avoir des unités de mesure différentes.

II.2.5. 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).

Chapitre II. Théorie de l'Analyse en Composante Principales

2. Calcul des paramètres statistiques :(moyenne, écart type).
3. Calcul des variables centrées réduites

$$Y_j = \frac{X_j - \bar{X}_j}{\sigma_{X_j}}$$

4. Calcul des coefficients de corrélation :

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

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 $V_P(j)$: 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.

II.2.6. 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.

II.2.7. Analyse des résultats

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

Chapitre II. Théorie de l'Analyse en Composante Principales

- **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 :

- ✓ Misérable en dessous de 0,5
- ✓ Médiocre entre 0,5 et 0,6
- ✓ Moyen entre 0,6 et 0,7
- ✓ Bien entre 0,7 et 0,8
- ✓ Très bien entre 0,8 et 0,9
- ✓ Et excellent au-delà de 0.9

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

- ✓ Si la signification (Sig.) tend vers 0.000 c'est très significatif.
- ✓ Inférieur à 0.05, significatif.
- ✓ Entre 0.05 et 0.10, acceptable.
- ✓ Au-dessus de 0.10, on rejette.

Toutefois le test est très sensible au nombre de cas, il est presque toujours significatif lorsque le nombre de cas est grand.

- **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. Graphiquement, on part des composants qui apportent le moins d'information (qui se trouvent à droite), on relie par une droite les points presque alignés et on ne retient que les axes qui sont au-dessus de cette ligne.

- **Comment interpréter les résultats ?**

C'est la phase la plus délicate de l'analyse. On donne un sens à un axe grâce à une recherche lexicale à partir des coordonnées des variables et des individus. Ce sont les éléments extrêmes qui concourent à l'élaboration des axes.

Si la variance expliquée est trop faible, on peut choisir d'exclure certaines variables. Pour choisir les variables à éliminer, on observe leur qualité de représentation : plus la valeur associée à la ligne «Extraction » est faible, moins la variable explique la variance.

Il faut également tenir compte du positionnement de chaque variable sur chaque axe.

II.2.8. La représentation graphique

Une fois qu'on aura les résultats numériques on passe à la représentation graphique

- La carte des variables

Dans la représentation graphique des variables, chaque variable a comme coordonnées (covariance, CP réduite)

$$Cov(X_j, C_k) = (\lambda)^{1/2} * V_k(j) \quad (II.9)$$

L'interprétation de la représentation graphique permet de voir que la distance d'une variable à l'origine des axes est égale à sa norme, le coefficient de corrélation est égal au cosinus de l'angle formé par deux variables, et aussi deux variables fortement corrélées, et éloignées l'une de l'autre possèdent une covariance élevée négative.

Si nous considérons une variable centrée réduite Y_j au lieu de la variable initiale X_j , les coordonnées de cette nouvelle variable dans le plan factoriel sont calculées par :

$$COV(Y_j, C_i) = COV\left(\frac{X_j}{S_j}, C_i\right) = Cor(X_j, C_i) \quad (\text{II.10})$$

C_i : Composante principale normée.

S_j : Ecart type de la $j^{\text{ème}}$ variable.

Comme la variance de Y_j est égale à 1, cela signifie que son point représentatif se trouve sur une sphère de rayon égal à 1. C'est pourquoi le cercle unité tracé sur chacun des plans étudiés s'appelle cercle de corrélation. La variable sera d'autant mieux expliquée, que son point représentatif se rapproche du cercle et inversement.[17]

- La carte des individus

Une fois qu'on choisira les axes principaux, on peut faire une représentation graphique des individus.

La représentation graphique de ces individus nous permet de voir ceux qui se ressemblent et ceux qui sont totalement différents et cela par rapport à leur inter distance dans l'espace R^P .

II.2.9. Influence de différentes transformations des variables

Grâce à la matrice de corrélation (variables centrées réduites) ou de covariance (variables brutes), on peut calculer les valeurs propres de ces dernières ainsi que les vecteurs propres pour enfin calculer les CP. Et cela après diagonalisation.

Or la valeur du coefficient de corrélation (ou de covariance) entre les variables X_j et X_k dépend essentiellement des couples d'observations X_{ij} et X_{ik} et par conséquent toute transformation appliquée aux p variables X_j modifiera les coefficients de corrélation (ou de covariance), et de ce fait la structure même des composantes principales. [18]

- Variable centrée

$$Y_j = X_j - \overline{X_j} \quad (\text{II.11})$$

Avec :

X_j : Variable initiale ;

$\overline{X_j}$: Moyenne de la variable initiale ;

Chapitre II. Théorie de l'Analyse en Composante Principales

Y_j : Variable transformée (centrée).

- Variable réduite

$$Y_j = \frac{X_j}{\sigma_{X_j}} \quad (\text{II.12})$$

Avec :

X_j : Variable initiale ;

σ_{X_j} : Ecart type de la variable initiale ;

Y_j : Variable transformée (réduite).

- Variable entrée réduite

$$Y_j = \frac{X_j - \overline{X_j}}{\sigma_{X_j}} \quad (\text{II.13})$$

Avec :

X_j : Variable initiale ;

$\overline{X_j}$: Moyenne de la variable initiale ;

σ_{X_j} : Ecart type de la variable initiale ;

Y_j : Variable transformée (centrée réduite).

On peut donc conclure que l'analyse en composantes principales peut s'appliquer sur quatre types de variables :

- ✓ ACP sur variables brutes (sans transformation).
- ✓ ACP sur variables centrées.
- ✓ ACP sur variables réduites.
- ✓ ACP sur variables centrées réduites.

Chapitre II. Théorie de l'Analyse en Composante Principales

Considérons la variable brute X et la variable centrée Y , et montrons que la covariance des deux variables brutes (X_j, X_k) est équivalente à celle des deux variables centrées respectives (Y_j, Y_k).

$$Y_j = X_j - \overline{X_j} \Rightarrow \overline{Y_j} = 0 \quad (\text{II.14})$$

$$Y_k = X_k - \overline{X_k} \Rightarrow \overline{Y_k} = 0 \quad (\text{II.15})$$

Or :

$$\text{Cov}(Y_j, Y_k) = \frac{1}{N} \sum_{i=1}^N (Y_{ij} - \overline{Y_j}) * (Y_{ik} - \overline{Y_k}) = \frac{1}{N} \sum_{i=1}^N Y_{ij} * Y_{ik} \quad (\text{II.16})$$

En remplaçant chaque variable par son expression nous obtenons :

$$\text{Cov}(Y_j, Y_k) = \frac{1}{N} \sum_{i=1}^N Y_{ij} * Y_{ik} = \frac{1}{N} \sum_{i=1}^N (X_{ij} - \overline{X_j}) * (X_{ik} - \overline{X_k}) = \text{Cov}(X_j, X_k) \quad (\text{II.17})$$

D'où :

$$\text{Cov}(Y_j, Y_k) = \text{Cov}(X_j, X_k) \quad (\text{II.18})$$

Le traitement des variables réduites et centrées réduites se fait de la manière :

Soient Y_j et Z_j les variables obtenues par les transformations précédentes, montrons que la covariance des deux variables réduites (Y_j, Y_k) est égale à celle des deux variables centrées réduites (Z_j, Z_k), qui n'est autre que la corrélation entre les deux variables brutes correspondantes.

$$Y_j = \frac{X_j}{\sigma_{X_j}} \Rightarrow \overline{Y_j} = \frac{\overline{X_j}}{\sigma_{X_j}} \quad (\text{II.19})$$

$$Y_k = \frac{X_k}{\sigma_{X_k}} \Rightarrow \overline{Y_k} = \frac{\overline{X_k}}{\sigma_{X_k}} \quad (\text{II.20})$$

$$Z_j = \frac{X_j - \overline{X_j}}{\sigma_{X_j}} \Rightarrow \overline{Z_j} = 0 \quad (\text{II.21})$$

$$Z_k = \frac{X_k - \bar{X}_k}{\sigma_{X_k}} \Rightarrow \bar{Z}_k = 0 \quad (\text{II.22})$$

Evaluons maintenant les covariances des variables (Y_j, Y_k) et (Z_j, Z_k) :

$$\begin{aligned} \text{Cov}(Y_j, Y_k) &= \frac{1}{N} \sum_{i=1}^N (Y_{ij} - \bar{Y}_j) * (Y_{ik} - \bar{Y}_k) = \frac{1}{N} \sum_{i=1}^N \left(\frac{X_{ij} - \bar{X}_j}{\sigma_{X_j}} - \frac{\bar{X}_j}{\sigma_{X_j}} \right) * \left(\frac{X_{ik} - \bar{X}_k}{\sigma_{X_k}} - \frac{\bar{X}_k}{\sigma_{X_k}} \right) \\ &= \frac{\frac{1}{N} \sum_{i=1}^N (X_{ij} - \bar{X}_j) * (X_{ik} - \bar{X}_k)}{\sigma_{X_j} * \sigma_{X_k}} \\ &= \frac{\text{Cov}(X_j, X_k)}{\sigma_{X_j} * \sigma_{X_k}} \\ &= \text{Cor}(X_j, X_k) \end{aligned}$$

Donc :

$$\text{Cov}(Y_j, Y_k) = \text{Cor}(X_j, X_k)$$

$$\begin{aligned} \text{Cov}(Z_j, Z_k) &= \frac{1}{N} \sum_{i=1}^N (Z_{ij} - \bar{Z}_j) * (Z_{ik} - \bar{Z}_k) = \frac{1}{N} \sum_{i=1}^N Z_{ij} * Z_{ik} = \frac{1}{N} \sum_{i=1}^N \left(\frac{X_{ij} - \bar{X}_j}{\sigma_{X_j}} \right) * \left(\frac{X_{ik} - \bar{X}_k}{\sigma_{X_k}} \right) \\ &= \frac{\frac{1}{N} \sum_{i=1}^N (X_{ij} - \bar{X}_j) * (X_{ik} - \bar{X}_k)}{\sigma_{X_j} * \sigma_{X_k}} \\ &= \frac{\text{Cov}(X_j, X_k)}{\sigma_{X_j} * \sigma_{X_k}} \\ &= \text{Cor}(X_j, X_k) \end{aligned}$$

Donc :

$$\text{Cov}(Z_j, Z_k) = \text{Cor}(X_j, X_k)$$

D'où :

$$\text{Cov}(Y_j, Y_k) = \text{Cov}(Z_j, Z_k) = \text{Cor}(X_j, X_k)$$

En conclusion nous aurons deux types de variantes à appliquer :

- ACP brute = ACP centrée.
- ACP centrée réduite (ou normée) = ACP réduite.

L'ACP brute peut être utilisée pour tous les types de variables homogènes c'est à dire possédant les mêmes unités de mesure alors que l'ACP normée peut être utilisée pour des variables homogènes et hétérogènes. En effet en centrant et en réduisant la variable, on élimine le problème de l'hétérogénéité de celle-ci. [18]

II.3. Présentation du logiciel XLSTAT

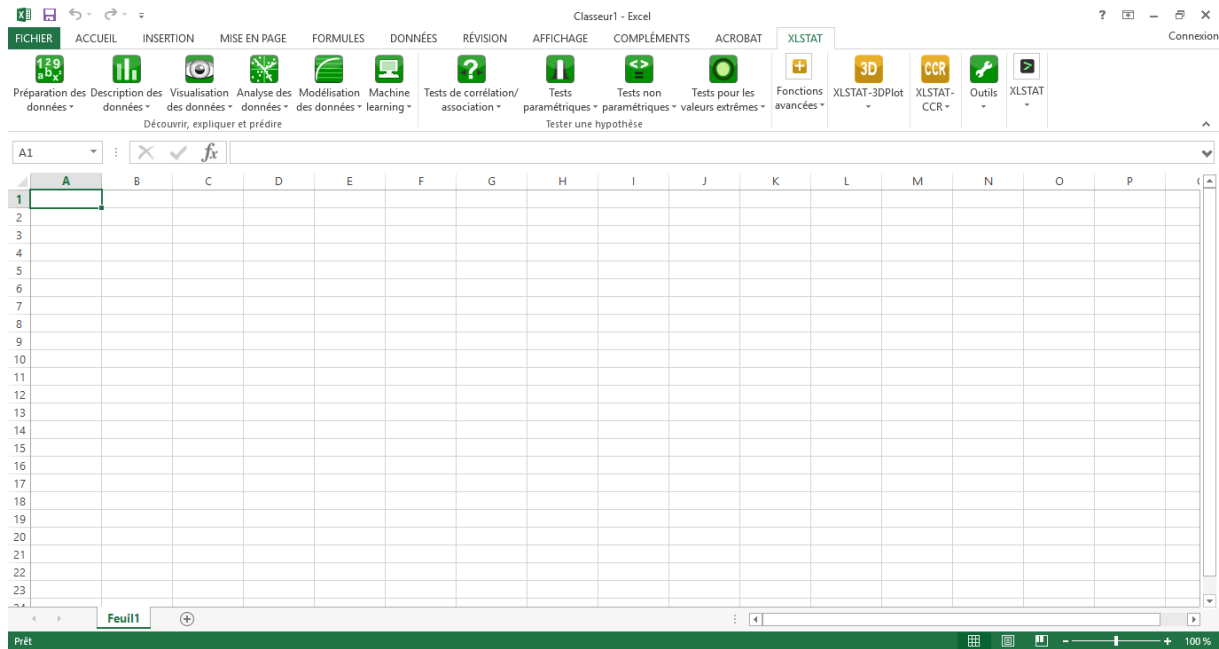


Figure II. 1: Interface de XLSTAT

XLSTAT est un logiciel d'analyse de données et de statistiques pour Microsoft Excel. Le développement de XLSTAT est depuis le début piloté par *Thierry Fahmy* fondateur d'addinsoft. Le logiciel est disponible sur internet depuis 1996.

II.3.1. Caractéristiques du logiciel

Le logiciel XLSTAT est un logiciel intégré à Excel, il suffit de lancer XLSTAT pour qu'une barre d'outils apparaisse, il utilise Excel comme interface de récupération de données et d'affichage de résultats.

Chapitre II. Théorie de l'Analyse en Composante Principales

XLSTAT nous permet d'utiliser les techniques de :

- ✓ Statistique ;
- ✓ Analyse de données ;
- ✓ Modélisation.

II.4.2. La démarche à suivre pour l'application de l'ACP sous XLSTAT

Lancer XLSTAT dans Excel > analyse des données > Analyse en composantes principales (ACP).

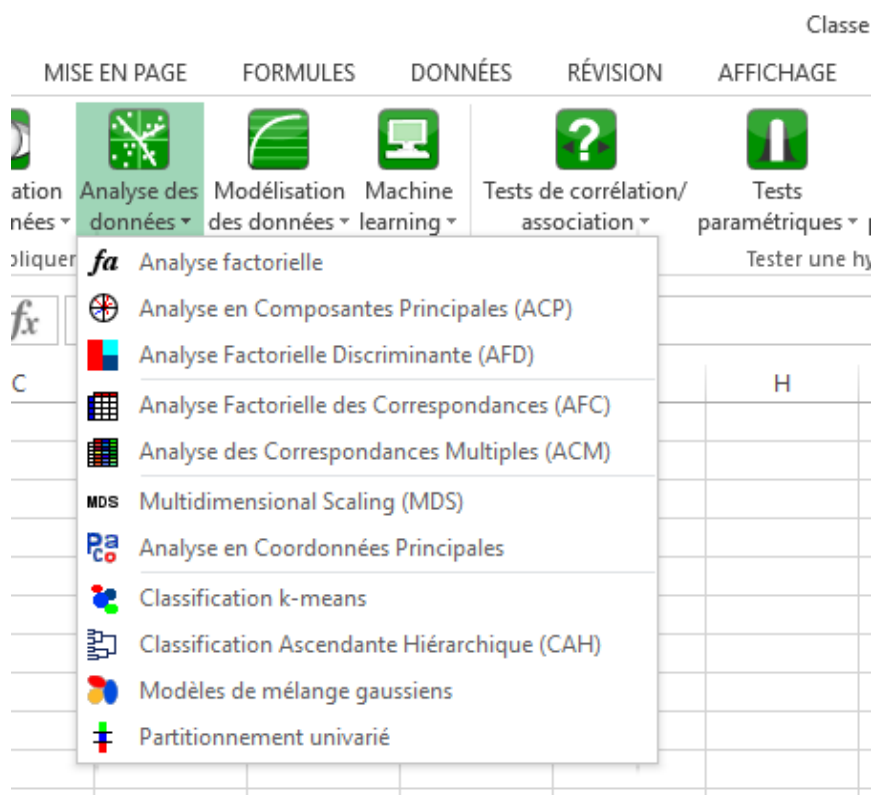


Figure II. 2. Lancement d'une analyse en composantes principales

Une boîte de dialogue apparaîtra :

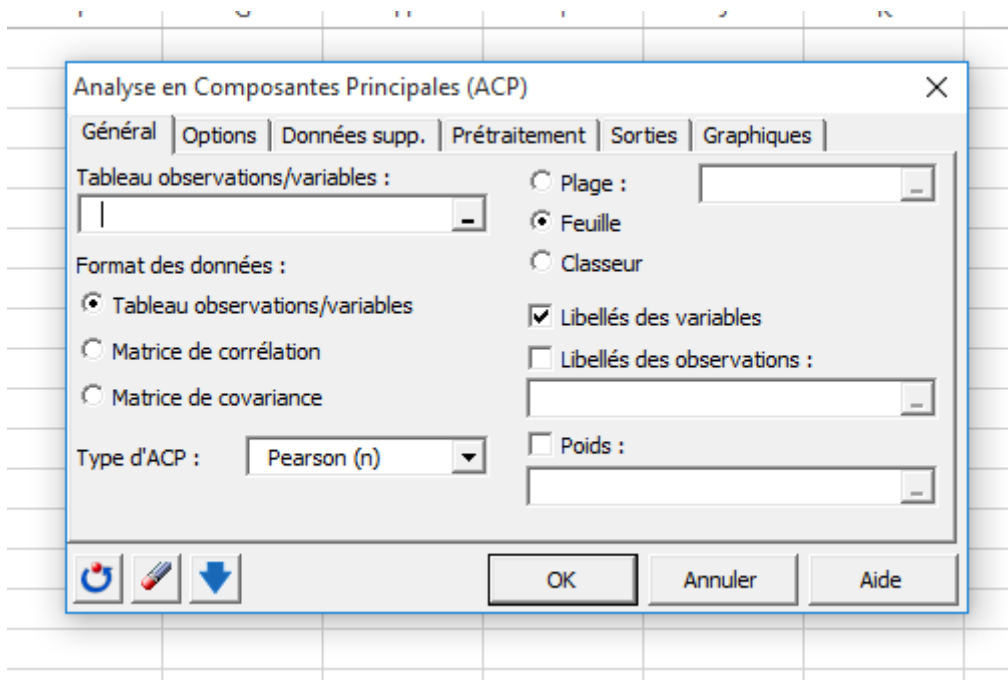


Figure II. 3 : Insertion du tableau d'observations

Dans l'onglet « Général », on sélectionne notre matrice [X] dans la case « tableau observations/variables »

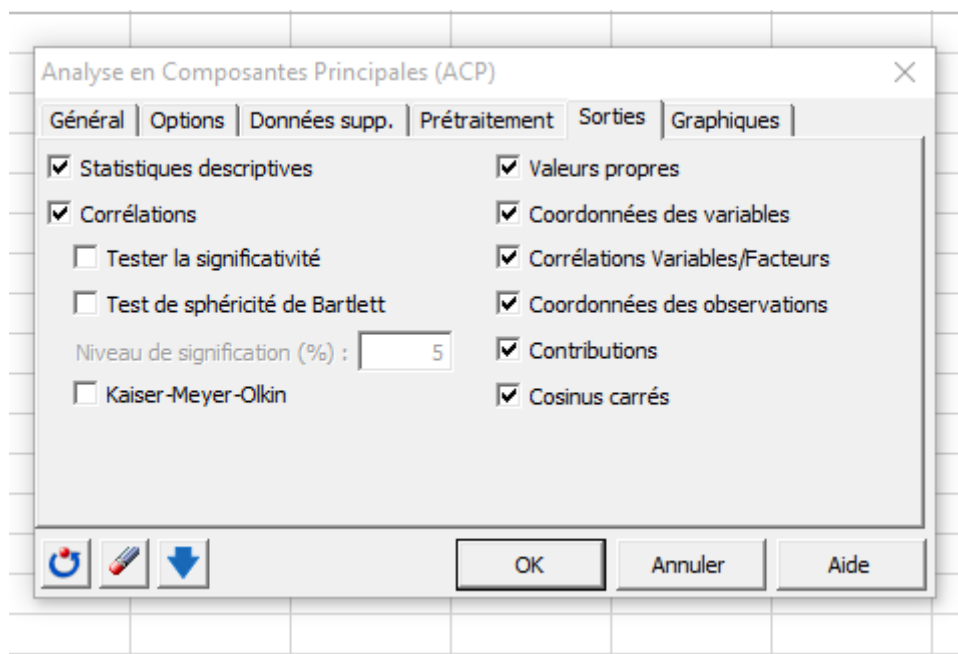


Figure II. 4 : Choix des sorties

Dans l'onglet « Sortie », il faut cocher les tests qu'on veut avoir comme résultats.

Dans notre cas, on sélectionne : tester la signification, test de sphéricité et Kaiser-Meyer-Olkin.

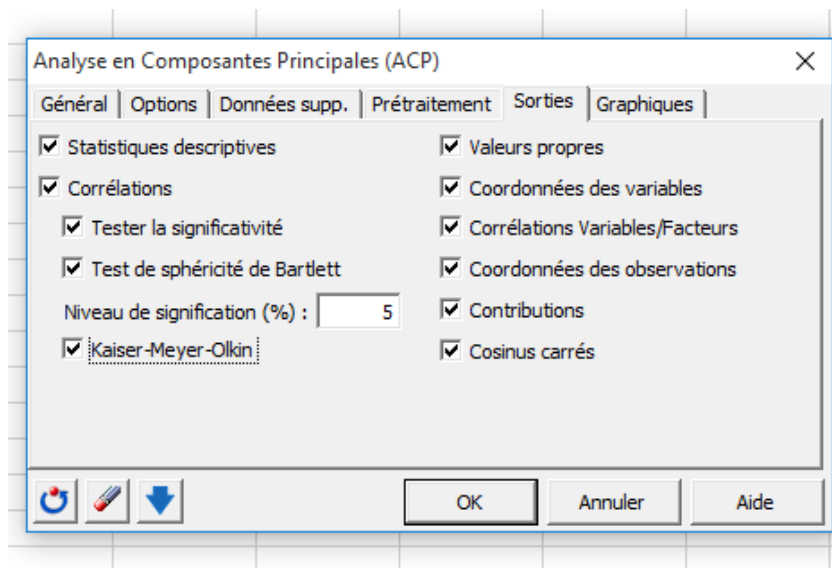


Figure II. 5 : Validation

Et enfin, on clique sur OK.

II.3.2. La démarche à suivre pour faire une régression avec XLSTAT

Une fois que XLSTAT lancé, dans la barre d'outils en haut il faut cliquer sur :

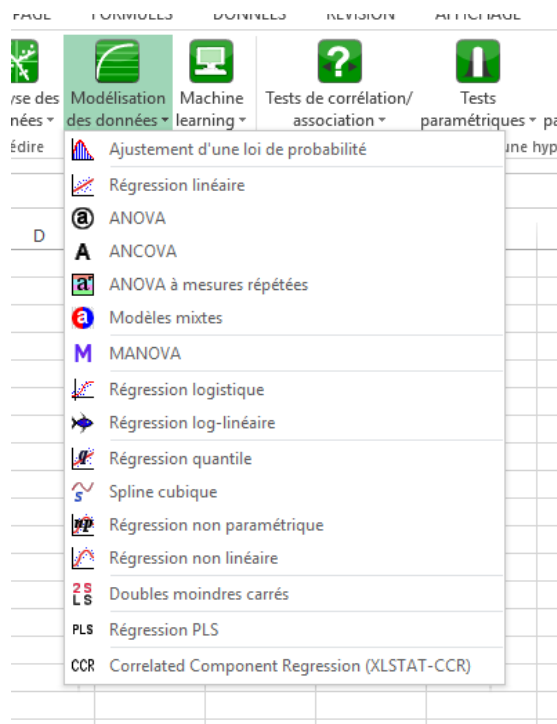


Figure II. 6 : Lancement d'une régression linéaire

Modélisation des données > Régression linéaire ... Une boîte de dialogue apparaîtra :

Dans la case « variables dépendantes » on sélectionne le tableau de la variable à expliquer.

Dans la case « variables explicatives » on selectionne la matrice des variables explicatives.

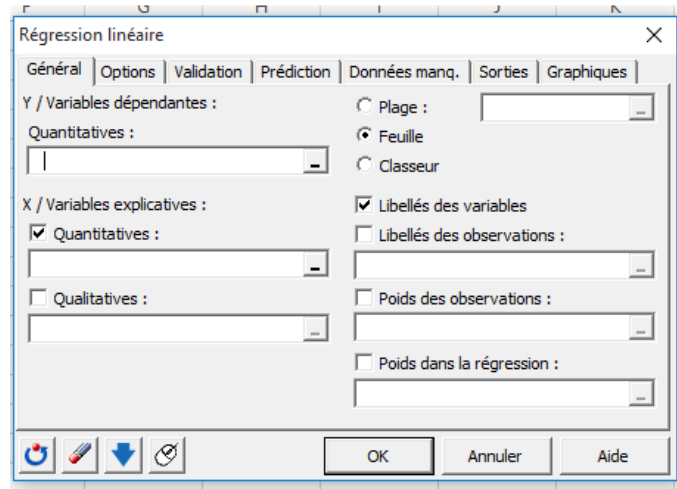


Figure II. 7 : Insertion des variables et des individus

Et on clique sur Ok.

Chapitre III : Théorie de la Neuro-floue

III.1. Introduction

Ces dernières années, le nombre et la variété des applications de la logique floue ont augmenté de façon significative, elle a connu un succès remarquable dans la description des systèmes complexes. Ces applications vont des produits de consommation à la reproduction des phénomènes naturels ...

Pour les phénomènes naturels la problématique de prédiction peut être formalisée de la manière suivante : étant donnée une situation caractérisée à l'instant « t » (entrée), l'outil de prédiction choisi doit prédire l'évolution de cette situation à l'instant « t+prévision » (sortie). Des développements actuels sur ce type de problématique montrent que les performances des neuro-flous dépassent celles d'autres méthodes aussi bien en terme de précision des résultats que d'efficacité d'apprentissage (Wang W.Q.et al.,2004) surtout pour les systèmes complexes à comportement non stationnaire et non linéaire.

Dans cet ensemble, notre travail traite plus spécifiquement du système ANFIS (Adaptative Neuro-Fuzzy Inference System) proposé par (Jang, 1993).

III.2. Identification floue

L'identification floue est un domaine extrêmement vaste, où chaque jour ou presque, une nouvelle approche est proposée. Certaines approches confinent à la théorie des réseaux de neurones.

L'idée générale d'un modèle neuro-flou est d'intégrer dans un seul concept les avantages des deux domaines : Logique floue et Réseaux de neurones.

- De la logique floue : Introduction de connaissances à priori afin de réduire l'espace des paramètres à optimiser.
- Des réseaux de neurones : utilisation de la capacité d'apprentissage afin d'améliorer les performances.

Ainsi, nous sommes dans l'obligation de présenter dans ce qui suit plusieurs définitions qui nous aiderons à la compréhension du système ANFIS.

III.3. Définitions de base

III.3.1. Les réseaux de neurones

Les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit. Toute structure hiérarchique du réseau est évidemment un réseau, elle est dotée d'une fonction particulière.

Chaque neurone des différents réseaux reçoit un nombre variable d'entrées en provenance de neurones amont. A chacune de ces entrées est associé un poids w représentatif de la force de connexion. Chaque processeur élémentaire est doté d'une sortie unique, qui se ramifie ensuite pour alimenter un nombre variable de neurones aval.

III.3.2. La logique floue

La logique floue est une extension de la logique booléenne par Lofti Zadeh en 1965, donc elle représente une généralisation des ensembles classiques. En introduisant la notion de degré dans la vérification d'une condition, permettant ainsi d'avoir un état que vrai ou faux. La logique floue confère une flexibilité appréciable aux raisonnements de l'utilisateur, ce qui rend possible la prise en compte des imprécisions et des incertitudes.

- **Imprécisions** désigne les connaissances qui ne sont pas perçues ou définies nettement.
- **Incertitudes** désigne les connaissances dont la validité est sujette à question.

Un des intérêts de la logique floue pour formaliser le raisonnement humain est que les règles sont énoncées en langage naturel.

III.3.3. Les sous-ensembles flous

La logique floue repose sur la théorie des ensembles flous, qui est une généralisation de la théorie des ensembles classiques [Zadeh, 1965]. Par abus de langage, nous utiliserons indifféremment les termes sous-ensembles flous et ensembles flous.

Soit x un ensemble. Un sous-ensemble flou A de x est caractérisé par une fonction d'appartenance. $F^A = x \rightarrow [0, 1]$. La fonction d'appartenance est notée μ_A .

Cette fonction d'appartenance est l'équivalent de la fonction caractéristique d'un ensemble classique.

Le sous-ensemble flou est défini sur un univers de discours X . L'univers de discours est l'ensemble des valeurs que peut prendre la variable floue x est $\mu_A(x)$ est le degré d'appartenance de l'élément x à l'ensemble flou A . [19]€

III.3.4. La fonction d'appartenance

La fonction d'appartenance représente le degré d'appartenance d'une valeur à un état donné, elle peut avoir différentes formes (en triangle, en trapèze...). Afin de mieux expliciter ces deux dernières définitions nous allons introduire l'exemple suivant :

Exemple :

Nous allons définir des fonctions d'appartenance pour chaque sous-ensemble flou des deux variables d'étude :

- Entrée : Débit . Sous ensembles : faible, moyen, fort.
- Sortie : Concentration en MES. Sous ensembles : faible, moyen, fort.

La forme de la fonction d'appartenance est choisie arbitrairement en suivant les conseils de l'expert ou en faisant des études statistiques : forme sigmoïde, tangente hyperbolique, exponentielle, gaussienne, triangulaire, trapézoïdale... [26]

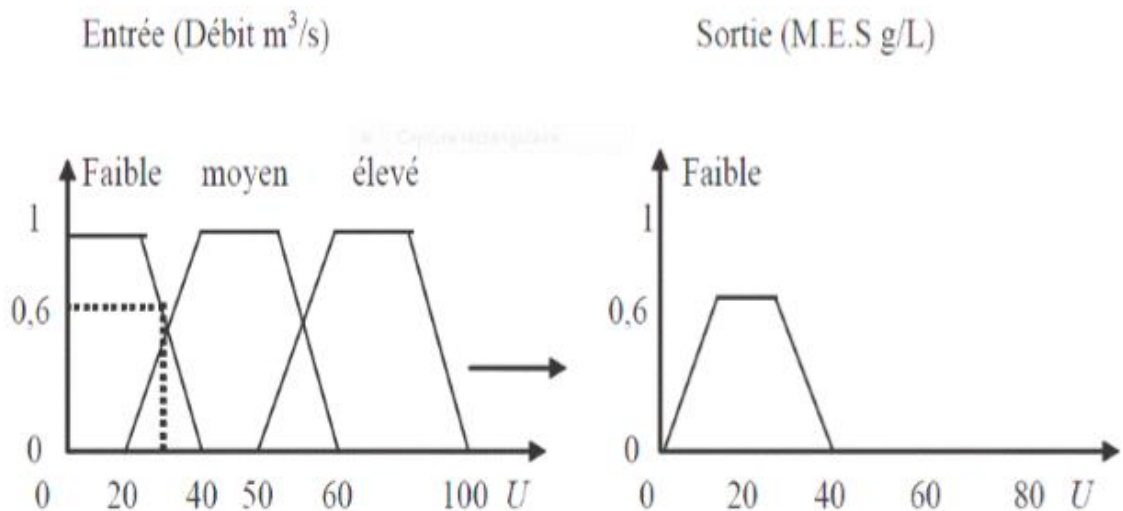


Figure III. 1 : Fonction d'appartenance caractérisant les sous-ensembles des variables d'entrées et de sorties.

III.3.5. Les variables linguistiques

Soit x une variable (débit solide), X la plage de valeur de la variable et A_x un ensemble fini ou infini de sous-ensembles flous. Une variable linguistique correspond au triplet (x, X, A_x) .

III.3.6. Le raisonnement en logique floue

En logique classique, les raisonnements sont de la forme :

$$\begin{cases} \text{Si } p \text{ alors } q \\ p \text{ vrai alors } q \text{ vrai} \end{cases}$$

En logique floue, le raisonnement flou, également appelé raisonnement approximatif, se base sur des règles floues qui exprimées en langage naturel en utilisant les variables linguistiques.

Une règle floue aura cette forme :

Si $x \in A$ et $y \in B$ alors $z \in C$, avec A, B et C des ensembles flous.

Par exemple :

Si (le débit est élevé), alors (la concentration des MES est élevée).

La variable « débit » appartient à l'ensemble flou « élevé » à un degré qui dépend du degré de validité de la prémisse, autrement dit du degré d'appartenance de la variable « débit » à l'ensemble flou « élevé ». L'idée sous-jacente est que plus les propositions en prémisse sont vérifiées, plus l'action préconisée pour les sorties doit être respectée.

III.3.7. Système d'inférence flou (SIF)

Une simple règle floue ne suffit pas à elle seule de représenter un problème complexe. Un ensemble de règles doivent être établies pour bien décrire notre système et simplifier son utilisation. L'association de ces ensembles de règles avec des mécanismes de décision permettent de déduire de nouvelles connaissances, elles constituent des systèmes d'inférences floues.

Un système d'inférence flou est essentiellement composé de deux blocs fonctionnels différents :

- Une base de connaissance constituée d'une base de règles floues et d'une base de données définissant les fonctions d'appartenance des ensembles flous.
- Un mécanisme d'inférence flou qui détermine la sortie du système sous la forme d'un ensemble flou.

On y ajoute deux blocs :

- Une interface de fuzzification qui transforme des entrées ponctuelles en ensembles flous
- Une interface de défuzzification qui transforme la sortie floue en sortie ponctuelle.

III.3.7.1. La fuzzification

C'est la première étape du traitement flou, les variables d'entrées se trouvent caractérisées à l'aide des sous-ensembles flous. La fuzzification est l'étape où les variables d'entrées « physiques » subissent une conversion artificielle et se transforment en des variables « linguistique ou floues ».

III.3.7.2. La défuzzification

La dernière étape est celle de la défuzzification, elle transforme la partie floue résultant de l'inférence en une valeur numérique. Le but de cette étape est de transformer l'ensemble flou en valeurs non floues et de trouver une sortie précise correspondant le mieux à la caractérisation floue.

La technique la plus utilisée est celle du centre de gravité. Dans notre cas la valeur simulée (concentration en MES) correspond à la projection du centre de gravité (défuzzification centroïde) de la surface d'inférence sur l'abscisse de la série des concentrations observées (Lefkir et al, 2006).

III.3.8. Synoptique d'un système flou

- Les inputs ;
- La fuzzification elle correspond aux deux variables linguistiques ;

- Le moteur d'inférence est constitué du choix des opérateurs flous;
- La base de connaissances floues est l'ensemble des règles floues;
- La défuzzification;
- Les outputs correspondent à la décision finale ;

Ainsi, la logique floue permet de mettre en place des systèmes d'inférence dont les décisions sont sans discontinuités, flexibles et non linéaires, plus proche du comportement humain que ne l'est la logique classique.

III.3.9. Type des systèmes d'inférence

Les systèmes flous sont classés selon la nature de leur structure. Il existe deux types de système d'inférence flou :

III.3.9.1. Système d'inférence de type Mamdani (linguistique)

Les modèles de Mamdani permettent une description linguistique du système par une base des règles floues de la forme :

$$Si x_1 est A_1 et x_2 est A_2 et \dots et x_n est A_n$$

$$Alors y_1 est B_1 et y_2 est B_2 et \dots et y_m est B_m$$

Où x_i et y_j sont respectivement les variables d'entrées et de sorties linguistiques et A_i et B_j sont des ensembles linguistiques flous. Ces modèles nécessitent un grand nombre de règles pour approcher le comportement global d'un système : ils possèdent une capacité de représentation limitée. Les modèles linguistiques ne permettent pas d'intégrer directement des connaissances objectives sur le système qui ne sont pas exprimées par des sous-ensembles flous. Ainsi des connaissances exprimées sous forme analytiques décrivant la structure physique du système ne peuvent pas être pris en compte directement. Sugeno et ses collaborateurs ont proposé une classe de modèles appelés modèles (TSK) qui permet de représenter ce type d'information [27].

III.3.9.2. Système d'inférence de type Takagi Sugeno Kang (TSK)

Le système d'inférence de Sugeno est identique à celui de Mamdani, tous deux sont construits à partir d'une base de règles. Mais ils diffèrent sur la partie défuzzification. En effet,

Chapitre III. Théorie de la Neuro-Floue

Sugeno propose de remplacer la défuzzification des variables de sortie par une combinaison linéaire (polynomiale ou constante) des variables d'entrée.

Soit un système à r entrées et une sortie, le modèle TSK est constitué d'une base de règles de la forme :

Si x_1 est A_1^{i1} est x_2 est A_2^{i2} et ... et x_r est A_r^{ir}

Alors $y = f_i = (x_1, x_2, \dots, x_r)$

Où x_1, x_2, \dots, x_r sont les variables numériques d'entrée du modèle flou et f_i est une fonction numérique de l'univers d'entrée. Chaque règle représente un modèle local sur une région floue d'entrée, ou sur un sous-espace d'entrée. Dans chaque région, le modèle flou est défini par la fonction qui relie les entrées à la sortie numérique. Le même modèle global est constitué par interpolation des modèles locaux.

Dans la littérature, les f_i sont souvent des fonctions linéaires, soit :

$$f_i = (x_1, x_2, \dots, x_r) = b_{i0} + b_{i1} x_1 + b_{i2} x_2 + b_{ir} x_r$$

Où $b_{i0}, b_{i1}, \dots, b_{ir}$ sont des coefficients constants. Chaque règle peut être vue comme un modèle local, linéaire selon les variables d'entrées x_1, x_2, \dots, x_r .

Ces modèles permettent d'approcher le comportement d'un système complexe par une collection de modèles locaux. Ils ont une capacité de représentation importante. En effet, le nombre de règles nécessaires pour approcher un système à un certain degré de précision est généralement réduit.

Le paramétrage d'un SIF de type Sugeno se limite généralement à la définition des éléments suivants :

- Le nombre de fonctions d'appartenance sur chaque domaine d'entrée.
- Le placement de ces fonctions d'appartenance.
- La valeur numérique des conclusions.

III.4. Système hybride neuro-flou [28]

Nous abordons maintenant une méthode d'optimisation du modèle TSK qui est l'apprentissage par approche neuro-floue. Le système étudié est le système ANFIS, dans tout ce qui suit nous le détaillerons du point de vue fonctionnement (algorithmes d'apprentissage...).

III.4.1. Emergence des systèmes hybrides pour la prédiction

L'utilisation des réseaux de neurones pour la prédiction date de 1964 par le chinois Hu. Il utilisa le réseau adaptatif de Widrow afin de réaliser des prédictions climatologiques. Le manque d'algorithmes d'apprentissage limita la poursuite de ce type d'étude.

- Etape 1 : La première application des réseaux de neurones pour la prédiction s'est faite en 1987. Par la suite, les réseaux de neurones non bouclés se montrèrent plus performants que les modèles classiques d'auto régression pour la prédiction de séries temporelles non linéaires.
- Etape2 : Amélioration de l'apprentissage et du paramétrage des réseaux de neurones où beaucoup de facteurs sont mis en jeu. Dès les années 90, beaucoup de développements sont menés pour améliorer la précision des prédictions réalisées par des réseaux de neurones tout en diminuant la complexité des modèles et le temps de calcul.
- Etape3 : Les réseaux récurrents : Afin de prendre en compte de manière explicite le temps, les architectures de réseaux récurrents ont été développés et comparés aux autres techniques de prédiction de série temporelles non linéaires. Les résultats montrent que ces réseaux de neurones présentent de meilleurs performances que celles des méthodes classiques et plus encore, que celles des réseaux non bouclés.
- Etape4 : Les réseaux de neurones ont été utilisés avec succès pour supporter l'activité de prédiction. Cependant, certains auteurs restent sceptiques :
 - L'optimisation d'un réseau de neurone est considéré comme un art plus qu'une science.
 - Les réseaux de neurones sont une boîte noire et il n'est pas possible d'explicitier et d'analyser les relations entre entrées et sorties.

Chapitre III. Théorie de la Neuro-Floue

Afin de pallier cette faiblesse des réseaux de neurones, des travaux insistent sur l'intérêt de systèmes combinant l'aspect connexionniste des réseaux de neurones à des techniques de raisonnement.

Ainsi, les systèmes neuro-flous ont vu le jour pour répondre à ces exigences. L'utilisation conjointe des réseaux de neurones et de la logique floue, permet de tirer les avantages des deux méthodes : les capacités d'apprentissage de la première et la lisibilité et la souplesse de la seconde.

Pour résumer l'apport du neuro-flou, le tableau suivant regroupe les avantages et inconvénients de la logique floue et des réseaux de neurones.

Réseaux de neurones	Logique floue
Avantages	
Modèle à base de données	Pas de modèles mathématiques
Pas de règles linguistiques	Intéprétabilité
Algorithmes d'apprentissage	...
Inconvénients	
Modèle boîte noire	Ajustement par tâtonnement
...	Pas d'apprentissage
...	Nécessité de posséder les règles

Une définition des systèmes neuro-flous est donnée dans (Nauck et al., 1997) et reprise dans (Rococeanu, 2006) selon laquelle :

«Les systèmes neuro-flous sont des systèmes flous formés par un algorithme d'apprentissage inspiré de la théorie des réseaux de neurones. La technique d'apprentissage opère en fonction de l'information locale et produit uniquement des changements locaux dans le système flou d'origine» (Otilia, 2008).

Donc, l'intérêt est de construire un système prédictif qui s'appuie sur l'intégration des réseaux de neurones et des systèmes d'inférence floue (SIF) à cause de leur complémentarité. Les SIF exploitent des règles linguistiques du type SI-ALORS traduisant une connaissance sur la dynamique d'un système. Cependant, un SIF n'est pas à même d'apprendre : les règles

doivent être formulées ce qui est parfois difficile...Ainsi, une solution consiste à utiliser les propriétés des RN afin d'apprendre la structure floue et d'en adapter les paramètres en conséquence.

Très généralement, les réseaux neuro-flous remplacent les différentes couches cachées des réseaux de neurone par des règles floues (i.e règles linguistiques). Ils utilisent ensuite des algorithmes d'apprentissage pour définir et optimiser ces paramètres.

III.4.2. Architectures neuro-floues

Divers associations des méthodes et architectures neuro-floues ont été développées depuis 1988.

(Racoceanu, 2006) a recensé plusieurs architectures de systèmes neuro-flous. La figure ci-dessous montre les différentes architectures des systèmes neuro-flous. FALCON, GARIC interprètent la règle floue avec une structure à 5 couches, NEFCLASS interprète la règle floue avec une structure à 3 couches et ANFIS interprète la règle floue avec une structure à 6 couches.

- FALCON et GARIC

Il s'agit de modèles à 5 couches, utilisant la fuzzyfication en entrée et la defuzzyfication en sortie. Ceci correspond à l'interprétation juste de la technique de Mamdani. La précision accrue des résultats provoque une lenteur dans l'exécution du système. Ce modèle est rarement utilisé en pratique mais il reste le meilleur pour la commande.

- NEFCLASS

Modèle utilisé généralement en classification, il est constitué de 3 couches : une couche d'entrée avec les fonctions d'appartenance, une couche cachée représentée par des règles et une couche de sortie définissant les classes.

Ce modèle est facile à mettre en application, il évite l'étape de défuzzyfication, tout en étant précis dans le résultat final, avec une rapidité bien supérieure aux autres architectures.

- NEFCON

Modèle formé de 3 couches. Une couche cachée formée par des règles, une couche d'entrée incluant les noeuds d'entrée avec les sous-ensembles flous d'antécédentes et une couche de sortie avec un nœud de sortie et les sous-ensembles des conséquents.

- NEFPROX (Neuro Fuzzy function apPROXimator)

Modèle obtenu par l'association des deux architectures NEFCLASS et NEFCON, il est utilisé dans différentes applications comme la classification et l'approximation de fonctions. NEFCLASS utilise un algorithme supervisé pour définir les règles floues, le NEFCON utilise un algorithme d'apprentissage non supervisé avec le calcul de l'erreur de sortie. Les deux modèles emploient la rétropropagation afin de définir les sous-ensembles flous. Comparé au modèle ANFIS, NEFPROX est beaucoup plus rapide, mais ANFIS donne de meilleurs résultats en approximation. Le NEFPROX est le premier système interprétable et lisible, dédié à l'approximation de fonction. Néanmoins, ses résultats en classification restent moins bons que ceux donnés par le NEFCLASS.

- ANFIS (Adaptative Network based Fuzzy Inference System)

ANFIS représente un système à inférence floue mis en application dans le cadre des réseaux adaptatifs. Il utilise la procédure d'apprentissage hybride. Cette architecture affine les règles floues obtenues par des experts humains pour décrire le comportement d'entrée-sortie d'un système complexe. Ce modèle donne de très bons résultats en poursuite de trajectoire, approximation non linéaire, commande dynamique et traitement de signal.

- SANFIS

Une autre structure connexionniste basée sur la fonction de base (FBF), a été développée récemment. Trois types de réseaux SANFIS peuvent être identifiés.

Pour ces 3, les entrées sont les mêmes mais les sorties sont différentes. Ils sont utilisés selon le type d'applications.

III.4.2.1. L'outil hybride ANFIS

Le système ANFIS est une classe de réseau adaptatif. Il peut être vu comme un réseau de neurones non bouclé pour lequel chaque couche est un composant d'un système neuro-flou et, à ce titre, c'est un "approximateur" universel. Il est utilisé dans différentes applications de prédictions.

Un système ANFIS réalise une approximation linéaire de la variable de sortie en décomposant l'espace des entrées en différents espace flous.

Chapitre III. Théorie de la Neuro-Floue

Dans ce qui suit, nous nous intéresserons à décrire l'architecture du système ANFIS et le mécanisme de son système d'inférence. Nous allons utiliser pour cela la figure ci-dessous :

Le réseau possède deux entrées et un système à 5 couches. Chaque couche comprend différents nœuds.

- Couche1 (Layer1): La première couche permet une fuzzification des variables X et Y. La sortie du nœud k de la couche 1 $x_{1,k}$, désigne la fonction d'appartenance du nœud ξ (présent en indice). Les paramètres utilisés dans les fonctions d'activations (classiquement, il s'agit de gaussiennes ou de sigmoïdes) sont appelés paramètres "prémises". La valeur alors obtenue $\mu_{A1}(x)$ représente le degré d'appartenance de la valeur X à l'ensemble A1.
- Couche2 (Layer2) : Chaque nœud correspond à une T-Norme floue (l'opérateur TNorme permet de réaliser l'équivalent d'un "ET" booléen). Il reçoit la sortie des nœuds de fuzzification et calcule sa valeur de sortie grâce à l'opérateur produit) (Cet opérateur est généralement utilisé mais il en existe d'autres : max, min...).

$$x_{2,k} = \prod_{(1,j) \in (2,k)} x_{1,j}$$

- Couche3 (Layer3) : Cette couche normalise les résultats fournis par la couche précédente. Les résultats obtenus représentent le degré d'implication de la valeur dans le résultat final.

$$x_{3,k} = \frac{x_{2,k}}{\sum_{(2,j) \in Ant(3,k)} x_{2,k}}$$

- Couche4 (Layer4) : Chaque nœud de cette couche est relié aux entrées initiales. On calcule le résultat en fonction de son entrée et d'une combinaison linéaire du premier ordre des entrées initiales (Approche de TAKAGI-SUGENO).

$$x_{4,k} = x_{3,k} [m_{k0} + m_{k1}X + m_{k2}Y]$$

Les paramètres $\{m_{k0}, m_{k1}, m_{k2}\}$ sont appelés paramètres "conséquents"

- Couche5 (Layer5) : La couche de sortie est constituée d'un unique neurone. Où n est le nombre de règles floues.

$$x_{5,k} = \sum_{i=1}^n x_{4,i}$$

En supposant des ensembles flous du type gaussien, le réseau ANFIS décrit ci avant comporte 14 paramètres devant être optimisés (8 inhérents aux fonctions gaussiennes et 6 à la linéarisation des sorties des règles). A cette fin, une base d'apprentissage est nécessaire. L'apprentissage consiste en la correction des paramètres (prémises et conséquents) du réseau afin de généraliser une fonction de transfert (inconnue à priori) entre les entrées et la sortie du réseau. Celle-ci constituée d'un ensemble de couples 'entrées/sortie' connus. Le déploiement des algorithmes d'apprentissage sur cette base de données permet de construire une fonction d'approximation de la sortie à partir des nouveaux vecteurs d'entrée.

En général, des algorithmes hybrides sont utilisés. Plus particulièrement, des travaux montrent que la combinaison de l'algorithme de descente du gradient et de l'algorithme LSE (Least Square Estimate : estimation des moindres carrés) permet de réduire la complexité de l'apprentissage et d'en améliorer l'efficacité.

III.4.2.2. Modélisation

Le problème de modélisation floue ou de construction de modèles flous peut être vu comme un processus d'identification de système. Ce processus est composé de trois phases :

- Phase d'identification de la structure du modèle

Cette phase se base sur la détermination de la structure de la base des règles, du nombre de symboles et de la structure des conclusions des règles.

- Phase d'identification des paramètres du modèle

L'identification des paramètres d'un modèle se fait à partir d'un ensemble de données appelé ensemble d'apprentissage. Cette phase permet l'identification des paramètres d'entrée, des paramètres de sortie et de la base des règles ou de la relation floue sous-jacente.

- Phase de validation du modèle

Cette phase sert à tester la validité du modèle d'abord par rapport à l'ensemble d'apprentissage, puis sur de nouvelles données de test.

III.4.3. Améliorations potentielles d'un système ANFIS

Les réseaux neuro-flous sont connus pour leur rapidité et précision. Nous présenterons ci-dessous un inventaire des leviers permettant d'améliorer les performances des réseaux neuro-flou. [28]

III.4.3.1. Architecture du réseau

Une des plus importantes étapes pour la génération de la structure des réseaux neuroflous est l'établissement des règles d'inférence. En utilisant un certain mécanisme d'inférence, les règles sont définies comme combinaisons des fonctions d'appartenance pour chaque variable d'entrée. Les variables d'entrée sont divisées en un nombre limité de fonctions d'appartenance (et leurs combinaisons mènent à beaucoup de règles d'inférence différentes).

- *Structure du réseau*

(Rui et al., 1995), a montré qu'il n'y a pas de structure du réseau idéal a priori. Celle-ci est à définir au cas par cas selon le problème. Concernant le réseau neuro-flou ANFIS, définir une architecture consiste à choisir le nombre de fonctions d'appartenance, leur type (gaussienne, sigmoïde...) et le nombre de règles floues utilisées. Aussi, un compromis simplicité/ généralisation doit être trouvé : il n'est pas concevable de construire un système ANFIS trop complexe, la base d'apprentissage étant, elle, limitée.

- *Partitionnement de l'espace des entrées*

Dans le même cadre, il existe aussi des discussions sur les méthodes de partitionnement de l'espace des entrées (grid, tree and scatter prtioning).

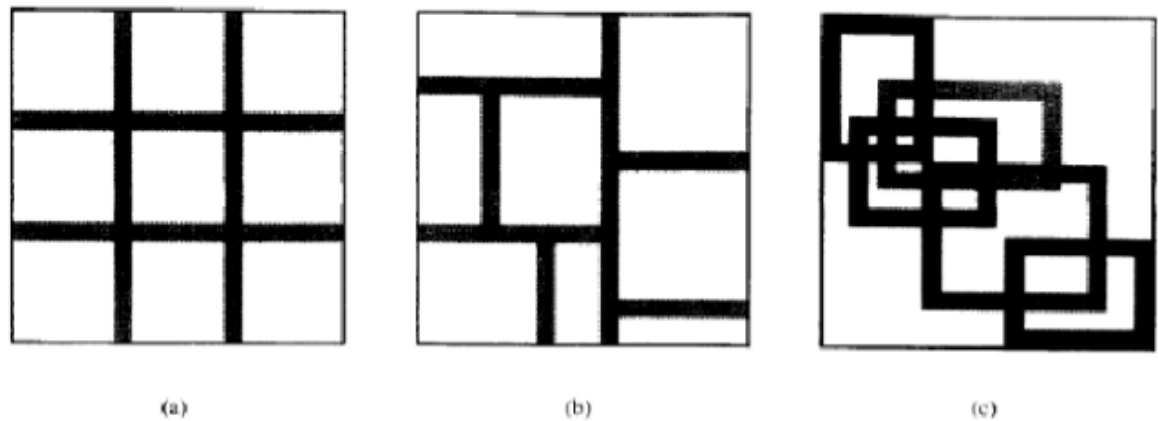


Figure III. 2 : Partitionnement de l'espace des entrées

La figure (a) illustre la « grid partition » dans un espace des entrées bidimensionnel. Cette stratégie a besoin seulement d'un nombre restreint de fonctions d'appartenance (MF) pour chaque entrée. Cependant, elle rencontre des problèmes quand nous avons un nombre modérément grand d'entrées. Par exemple, un modèle flou avec 10 entrées et 02 MF's sur chaque entrée aurait en conséquence $2^{10}=1024$ règles de type si-alors ce qui est prohibitivement grand. Ce problème, habituellement désigné sous le nom du problème de la dimensionnalité, peut être allégé par les autres stratégies de partage de l'espace des entrées.

La figure (b) montre une partition arborescente « tree partition », dans laquelle chaque région peut être univoquement indiquée par un arbre de décision. Cette partition soulage le problème d'une augmentation exponentielle du nombre de règles. Cependant, davantage de MF's pour chaque entrée sont nécessaires pour définir les régions floues, et ces MF's ne correspondent plus à des significations linguistiques claires telles que 'petit' ou 'grand'.

La figure (c) montre un sous ensemble de l'espace d'entrée qui caractérise une région de l'occurrence possible de l'entrée, scatter partition approche peut également limiter le nombre de règles à une quantité raisonnable.

III.4.3.2. Base d'apprentissage

L'ensemble d'apprentissage disponible impacte bien entendu la qualité de généralisation du système ANFIS. Il est nécessaire de disposer d'une base d'apprentissage de taille 5 fois supérieure au nombre de paramètres à optimiser (prémises et conséquents).

Cependant, au-delà de la taille de cet ensemble d'apprentissage, c'est également la pertinence des données qu'il contient qui se révèle critique.

Le numéro d'époques choisit doit être assigné pour surmonter le problème de paramétrisation (overfitting et underfitting), elle comprend une propagation en avant et une rétro-propagation. Durant la propagation en avant, les variables d'entrée servent à déterminer les sorties des neurones couche par couche, permettant de déterminer les valeurs des paramètres des conséquents. Durant la phase de rétro-propagation, l'algorithme de rétropropagation d'erreur est appliqué pour optimiser les poids des différentes couches. Les époques dans l'apprentissage dépendent également de l'erreur d'apprentissage.

Le processus d'apprentissage se termine lorsque l'erreur d'apprentissage est réduite à un niveau de tolérance indiqué.

III.4.3.3. Algorithme d'apprentissage

L'apprentissage consiste en la correction des paramètres (prémises et conséquents) du réseau afin de généraliser une fonction de transfert (inconnue à priori) entre les entrées et la sortie du réseau. A cette fin, une base d'apprentissage est nécessaire, elle sera constituée des entrées/sorties. Le déploiement des algorithmes d'apprentissage sur cette base de données permet de construire une fonction d'approximation (appelée aussi prédiction) de la sortie désirée à partir des nouveaux vecteurs d'entrée.

En général, des algorithmes hybrides sont utilisés. Une combinaison entre l'algorithme de descente du gradient qui permet l'estimation des poids du réseau et de l'algorithme d'estimation des moindres carrés qui permet d'estimer des paramètres des conséquents réduisent la complexité de l'apprentissage est réduite et l'efficacité se trouve améliorer (Riedmiller et al., 1993) proposent de modifier l'algorithme d'apprentissage afin d'en accélérer la convergence. (Yaam et al., 2000) proposent d'optimiser l'initialisation des poids d'apprentissage. La modification des pas d'apprentissage peut aussi être envisagée.

III.4.4. Les critères d'évaluation d'un modèle

Pour analyser la performance et la fiabilité d'un modèle, plusieurs critères d'évaluation (validation) ont été développés. Ils peuvent être utilisés aussi bien pour la calibration qu'en validation.

Les principaux critères objectifs utilisés sont :

III.4.4.1. Le critère de Nash Sutcliffe (Nash et Sutcliffe, 1970)

Il donne une indication sur la robustesse (pouvoir prédictif) des modèles utilisés. Ainsi, il est utilisé comme un outil d'aide à la décision.

Le critère est le suivant :

$$NASH = 1 - \frac{\sum_{t=1}^T (Q_0^t - Q_m^t)^2}{\sum_{t=1}^T (Q_0^t - \overline{Q_0})^2}$$

Avec : Q_0 représente les valeurs observées ;

Q_m sont les valeurs simulées est Q_0^t est l'observé à l'instant t ;

$\overline{Q_0}$ est la moyenne des valeurs observées.

La valeur du NASH varie entre « -∞ » à « 1 » et est généralement présenté en pourcentage.

- Une efficacité de NASH=1 correspond à une modélisation parfaite où les simulées = les observées.
- Une efficacité de NASH=0 indique que les prédictions du modèle sont aussi précises que la moyenne des données observées.
- Une efficacité de NASH<0 indique que les valeurs simulées conduisent à de moins bons résultats et que la moyenne observée est meilleur prédicateur que le modèle.

III.4.4.2. L'erreur quadratique moyenne (RMSE: Root Mean Square Error)

La mesure de l'erreur quadratique moyenne est aussi un critère souvent utilisé pour mesurer l'écart entre la simulation et l'observation, en donnant notamment une idée de la dispersion entre les deux et cela dans l'unité des variables étudiées. Plus le RMSE sera faible et plus l'écart entre la simulation et les observations sera faible.

Le critère est le suivant :

$$RMSE = \sqrt{\frac{1}{n_0} \sum_{t=1}^T (Q_m^t - Q_0^t)^2}$$

Le RMSE est couramment utilisé pour d'écrire l'erreur statistique. La performance du modèle est retenue selon une valeur minimale du RMSE.

III.4.4.3. Percent bias PBIAS

Le pourcentage du bias mesure la tendance moyenne des données simulées à être plus grandes ou plus petites que leurs homologues observées. Le PBIAS prend sa valeur optimale quand il est égal à zéro.

Si le PBIAS <0 il indique que les valeurs simulées sont plus élevés que ceux observés, ainsi on peut dire qu'il y'a une surestimation des valeurs.

Si le PBIAS >0 il indique que les valeurs simulées sont inférieurs à ceux observés, ainsi on peut dire qu'il y'a une sous-estimation des valeurs par le modèle.

$$PBIAS = \frac{\sum_{t=1}^T (Q_0^t - Q_m^t) * 100}{\sum_{t=1}^T Q_0^t}$$

III.4.4.4. Le coefficient de détermination

Il est défini comme étant la valeur élevée au carré du coefficient de corrélation de Pearson. Donc il estime la dispersion entre la série des observées et des simulées, plus il sera important plus l'accord entre les deux séries de valeurs le sera aussi.

Il nous informe également sur le degré de corrélation entre deux variables à étudier (pour notre cas entre le débit liquide et le débit solide), si sa valeur est égale à « 0 » cela signifie qu'il n'y a aucune corrélation entre les deux variables.

La valeur maximale est de « 1 » et elle représente la valeur optimale.

Le R² est donné par la relation :

$$R^2 = \left[\frac{\sum_{t=1}^T (Q_m^t - \overline{Q_0}) * (Q_m^t - \overline{Q_m})}{\sqrt{\sum_{t=1}^T (Q_0^t - \overline{Q_0})^2} * \sqrt{\sum_{t=1}^T (Q_m^t - \overline{Q_m})^2}} \right]$$

Chapitre III. Théorie de la Neuro-Floue

L'inconvénient du R^2 est qu'il quantifie la dispersion des points "observés-simulés" et ne donne aucune information sur la qualité du modèle, donc même si les valeurs simulées sont sous-estimées ou surestimées mais que leur dispersion est identique à celle des observées, on aura un R^2 proche de « 1 ». Comme solution à ce problème on définit ci-dessous le coefficient de détermination pondéré.

Chapitre IV : Présentation de la STEP de Boumerdes

IV.1. Introduction

La station d'épuration de Boumerdes a pour objectif l'épuration des eaux usées à faible charge massique, c'est une station à boues activées assurant l'épuration d'un débit journalier de 15000 m³/j. La station fonctionne à aération prolongée ($C_m = 0.076$ Kg DBO5 / Kg MVS/Jour).

Elle a été mise en service en 2006 et est gérée par l'Office National d'Assainissement (ONA).

IV.2. Localisation de la STEP de Boumerdes

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 IV. 1 : STEP de Boumerdes (Google Earth)



Figure IV. 2 : Vue d'ensemble des ouvrages de la STEP de Boumerdes

IV.3. Caractéristiques techniques

L'eau à traiter possède les caractéristiques suivantes :

- Volume journalier15 000 m³/j
- Débit moyen 24h 625 m³/h
- Débit de pointe temps sec 1063 m³/h
- Débit de pointe temps de pluie 1944 m³/h

IV.4. Paramètres de pollution de conception

La station d'épuration a été conçue pour épurer les eaux usées et les rejeter selon les caractéristiques suivantes :

Tableau IV. 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

IV.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 :

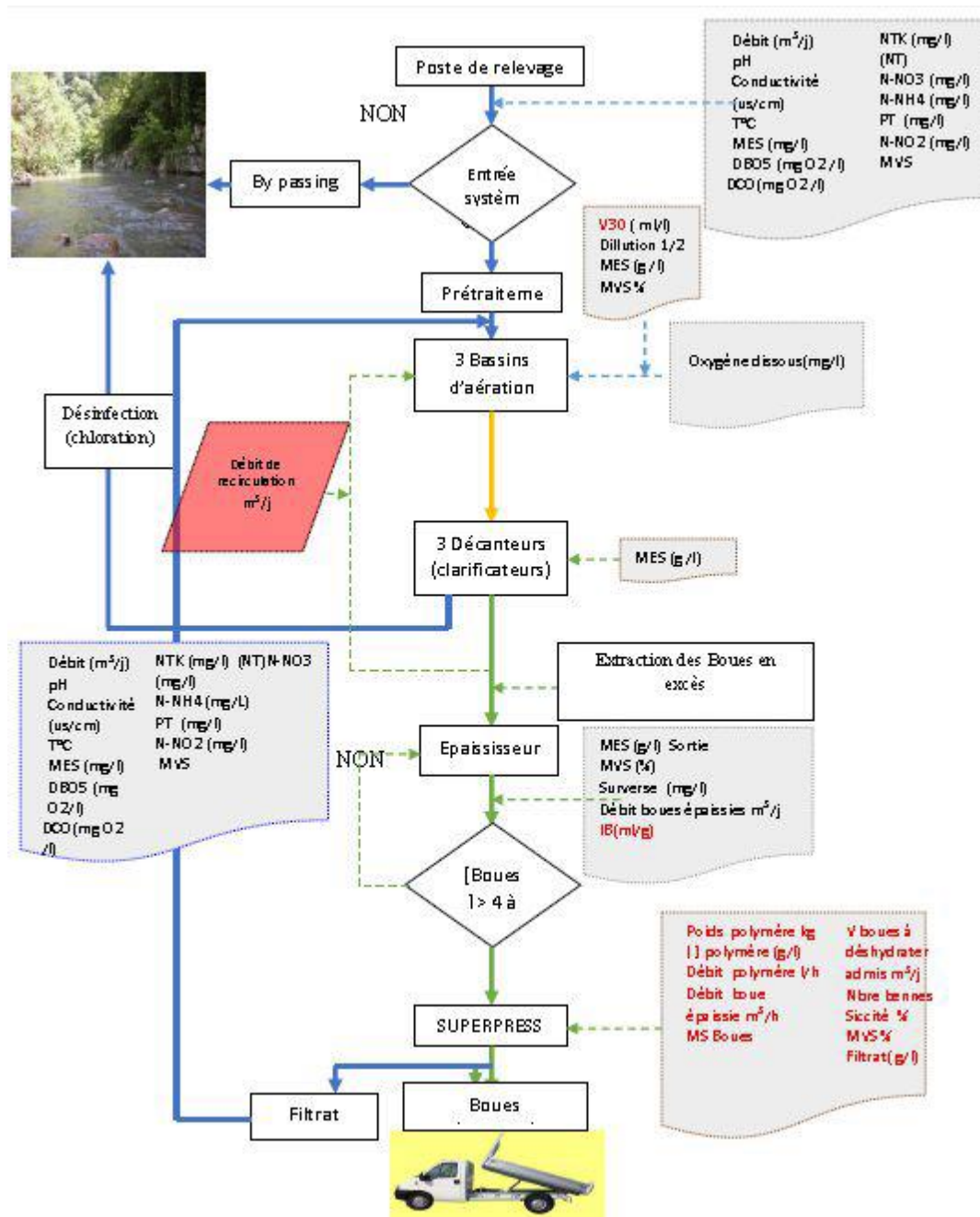


Figure IV. 3 : Organigramme d'exploitation de la STEP de Boumerdes

IV.5.1. Principe du traitement

La STEP de Boumerdes comporte deux chaînes de traitement :

IV.5.1.1. Un traitement au fil de l'eau :

L'eau qui arrive passe par le prétraitement, le traitement biologique à boues activées, décantation et désinfection.

a. Arrivée des eaux, By-Pass et bassin d'orage :

L'alimentation générale de la STEP en eaux usées se fait soit par pompage ou par gravité. Une fois que l'eau est entrée dans la STEP, on fait un comptage d'eau brute.

Cette eau est divisée en deux parties, une partie est pompée directement dans le canal d'alimentation du prétraitement et une partie est dirigée par gravité vers la bache de relèvement et ce par deux conduites de diamètre 600 mm. Ce poste de relèvement permet :

- L'alimentation du prétraitement en aval.
- Le By-pass total des installations par manœuvre d'un batardeau contrôlant l'entrée générale de la station.



Figure IV. 4 : Bassin d'orage de la STEP de Boumerdes

b. Le prétraitement

- Dégrillage :

Le dégrillage se fait par un dégrilleur automatique du type à champ courbe avec espacement entre barreaux de 20 mm. Les déchets cumulés sont évacués par convoyeur à vis. Dans le cas où le dégrilleur automatique est en maintenance, la station utilise un dégrilleur de secours à champ droit à nettoyage manuel afin d'effectuer un By-Pass. Ce dernier a un espacement entre barreaux identique au dégrilleur automatique.



Figure IV. 5 : Dégrilleur du type à champ courbe

- Dessaleur-Déshuileur

C'est un ouvrage circulaire de forme cylindro-conique. Il permet d'éliminer les sables de dimensions supérieures à 150 mm par décantation ainsi qu'une grande partie des matières flottantes (graisses, écumes, ect.) dans la partie supérieure de l'ouvrage.



Figure IV. 6 : Dessableur-Déshuileur

Chapitre IV : Présentation de la STEP de Boumerdes

c. 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 IV. 7 : Bassin d'aération

- 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 IV. 8 : Décanteur-Clarificateur

Chapitre IV : Présentation de la STEP de Boumerdes

d. Traitement final

- Désinfection

Une fois que l'eau est clarifiée, elle est dirigée vers un ouvrage en béton comportant un certain nombre de canaux en chicanes. Un premier canal est utilisé pour la mesure le débit d'eau épurée. Afin d'assurer un contact prolongé entre l'eau à désinfecter et l'eau chlorée, une série de canaux en chicanes est utilisée.

Une fois épurée l'eau est rejetée dans le milieu naturel.



Figure IV. 9 : Bassin de chloration

IV.5.1.2. Traitement au fil de boues

a. Epaissement des boues

L'épaississeur n'est pas le dépotoir final du traitement de l'eau mais l'ouvrage intermédiaire indispensable entre la chaîne de traitement d'eau et la chaîne de traitement des boues.

Le diamètre de l'épaississeur de la STEP est de 13m, la charge massique est de 30 kg/m²/j.

La concentration de sortie moyenne est de 20 g/l minimum avec un volume à transférer en déshydratation de 792 m³.

b. Déshydratation mécanique des boues bandes presseuses

Avant que les boues épaissies ne soient envoyées vers une bande presseuse SUPERPRESS, elles sont d'abord floculées et afin d'améliorer la filtrabilité des boues, l'ajout de polymère en faible quantité est nécessaire.

Chapitre IV : Présentation de la STEP de Boumerdes

La largeur de bande de la SUPERPRESS est de 2 m, la masse journalière de boues produites est de 4 t/j de boues sèches.



Figure IV. 10 : Presse à bande

IV.6. Statistiques et critiques des données

IV.6.1. Statistiques descriptives

L'étude a été faite sur une série de données journalières des paramètres de la pollution à l'entrée et à la sortie, le débit et l'énergie totale, mise à notre disposition par la STEP de Boumerdes.

Les données utilisées s'étalent sur une durée de 10 ans (du 1 Janvier 2006 jusqu'au 29 Février 2016).

Afin d'effectuer notre étude, nous avons procédé à une analyse statistique des paramètres retenus, ainsi qu'à leurs variations en fonction du temps.

Les résultats obtenus des calculs des paramètres (moyenne, écart type, minimum et maximum) sont résumés dans le tableau suivant :

Tableau IV. 2 : Statistique descriptives des paramètres de pollution de la STEP de Boumerdes

	Eau brute				Eau épurée			
	Moyenne	Ecart Type	Min	Max	Moyenne	Ecart type	Min	Max
PH	7.32	0.14	6.71	8.08	7.28	0.16	6.39	8.03
T (°C)	20.40	3.81	7.55	35.93	20.32	3.99	2.43	29.90
MES (mg/l)	240.23	137.31	24	2400	13.53	8.21	1	160
DBO5 (mg/l)	248.64	156.38	20	2500	10.69	8.28	1	55
DCO (mg/l)	473.16	251.46	110	3264	42.17	24.84	1	173
NH4 (mg/l)	30.98	18.49	0.80	126.54	5.40	8.25	0	60.10
NO2 (mg/l)	0.20	0.28	0	2.70	0.44	0.96	0	9.40
NO3 (mg/l)	4.62	9.50	0	145	35.05	156.90	0	1529.15
NTK (mg/l)	48.98	21.97	1.50	218.96	14.48	32.33	0.05	60.1
PO4 (mg/l)	6.88	5.40	0.35	64.44	3.18	2.58	0.10	22.40
Débit (m ³ /j)	11991.33	4186.17	0	22491	-	-	-	-
Energie (Kwh)	3932.84	1082.98	0	8355	-	-	-	-

IV.6.2. Evolution des paramètres de pollution dans le temps

Afin d'avoir une autre vision de l'évolution des paramètres de pollution dans le temps, nous avons fait une représentation graphique de chaque paramètre. Les résultats sont présentés dans les figures suivantes :

- Evolution du PH

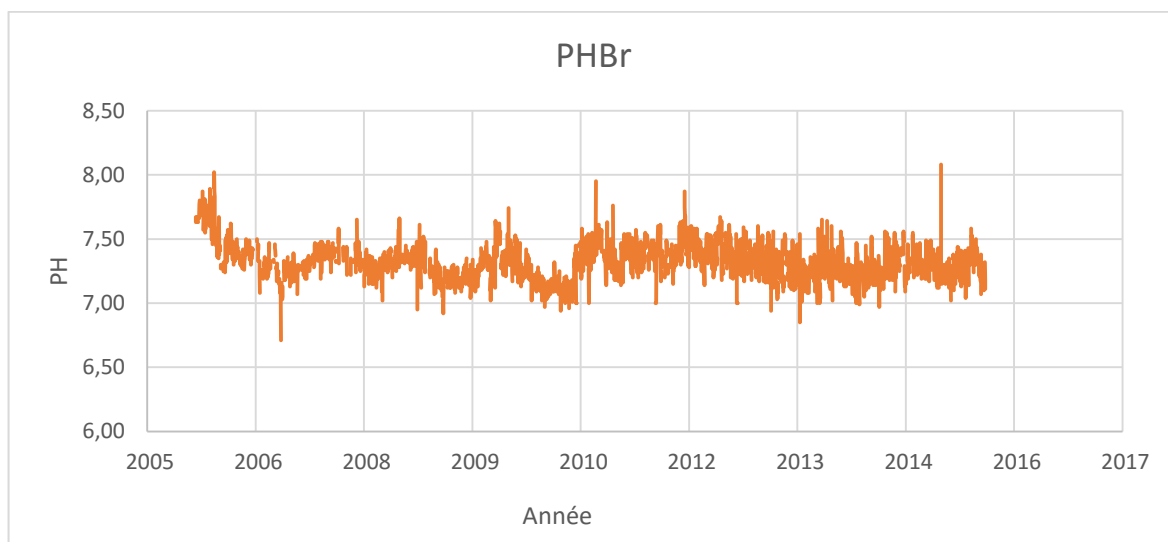


Figure IV. 11 : Evolution journalière du PH de l'eau brute

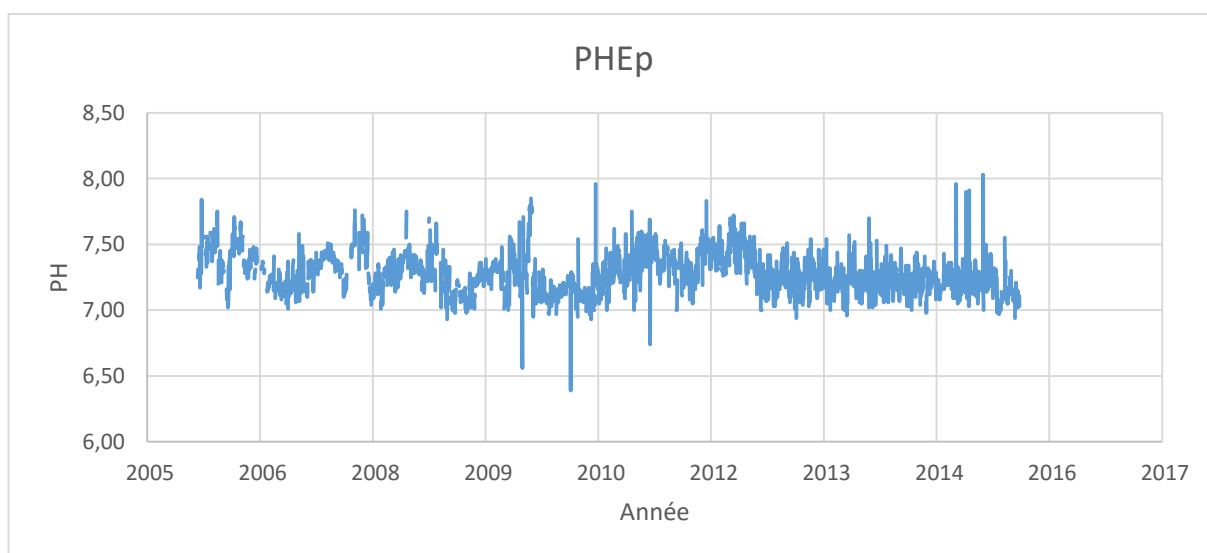


Figure IV. 12 : Evolution journalière du PH de l'eau épurée

Les deux figures ci-dessus montrent que le PH de l'eau brute ou de l'eau épurée varie entre 6.4 et 8.2 Avec une moyenne respectivement de 7.32 et 7.28

Chapitre IV : Présentation de la STEP de Boumerdes

En comparant l'évolution du PH à l'entrée et à la sortie de la STEP, on remarque des fluctuations insignifiantes de ce paramètre, par conséquent celui-ci n'apportera aucune information pertinente pour la problématique étudiée, il ne sera pas pris en considération dans la suite de l'analyse des paramètres de la STEP.

- Evolution de la température

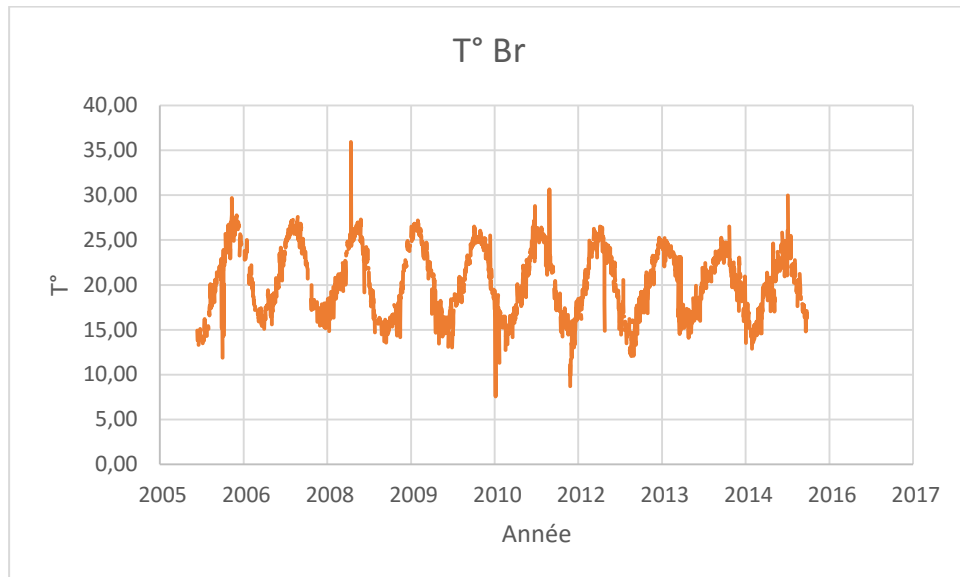


Figure IV. 13 : Variation journalière de la température brute

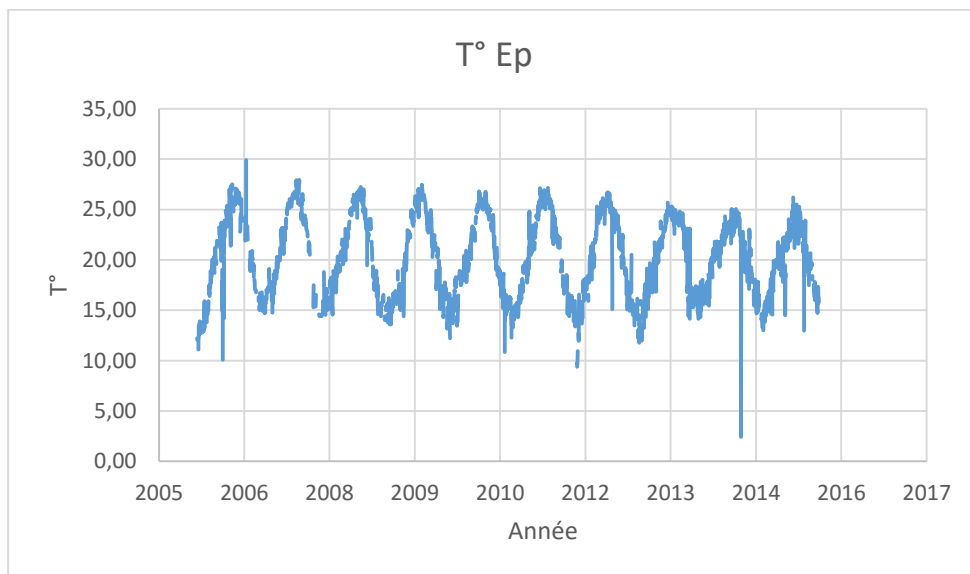


Figure IV. 14 : Variation journalière de la température de l'eau épurée

Chapitre IV : Présentation de la STEP de Boumerdes

D'après la représentation graphique de la température à l'entrée et à la sortie, on remarque que cette dernière oscille entre 12 et 28 °C à l'exception de certains cas où elle atteint un minimum de 2.5°C et un maximum de 35.9 °C.

Donc, on peut dire que la température de l'eau brute et celle de l'eau épurée est cyclique et est due aux variations saisonnières. Ceci nous permet de retirer ce paramètre de notre analyse car il n'influera pas sur la problématique étudiée.

- Evolution des Matières En Suspension (MES)

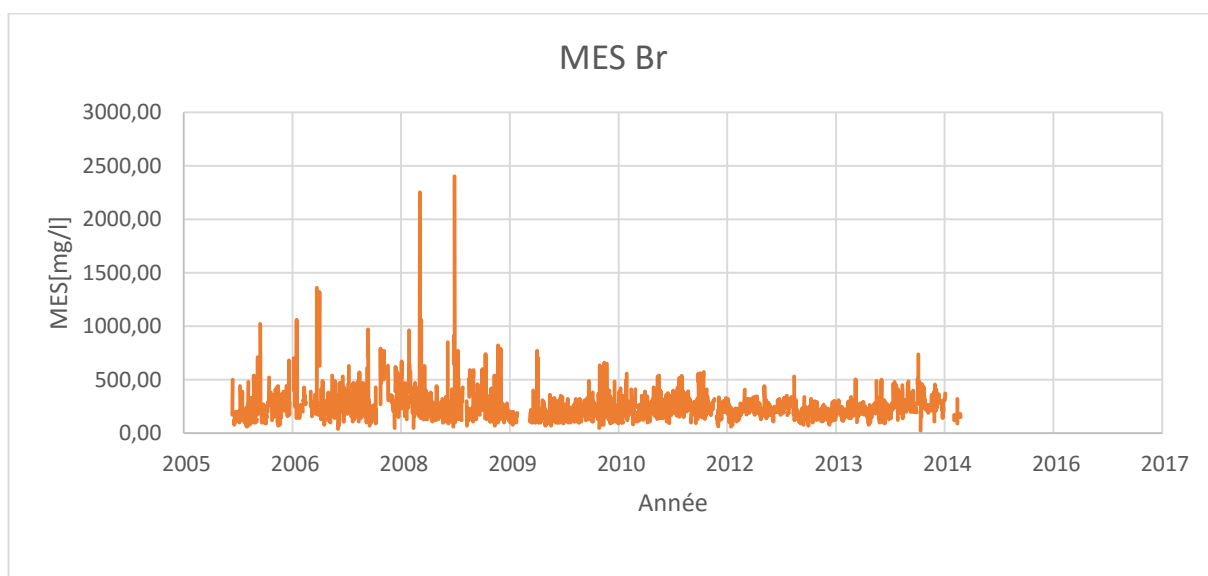


Figure IV. 15 : Evolution journalière de la concentration en MES à l'entrée de la STEP

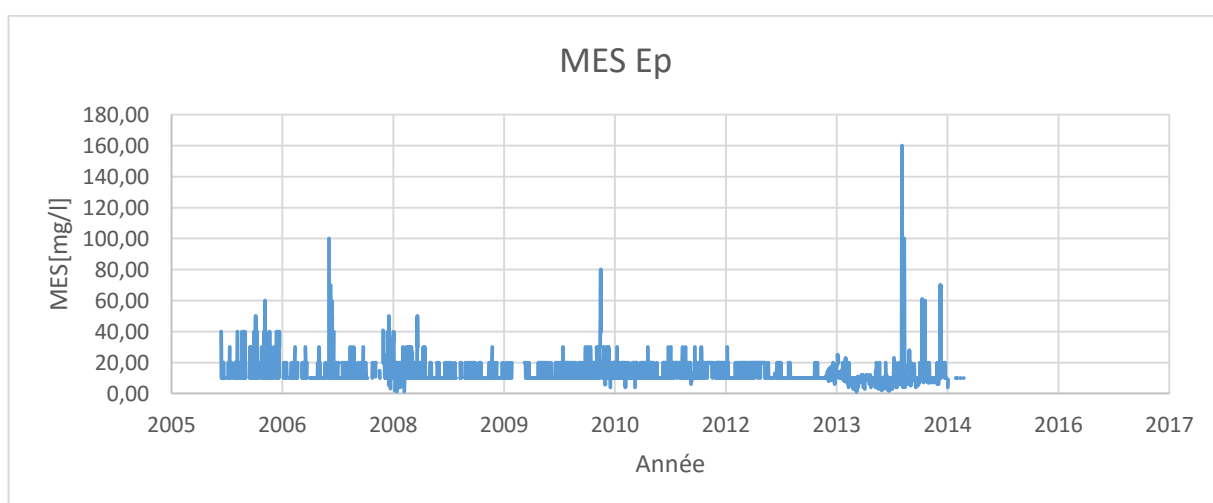


Figure IV. 16 : Evolution journalière de la concentration en MES à la sortie de la STEP

Chapitre IV : Présentation de la STEP de Boumerdes

A l'entrée de la STEP, les MES ont une concentration moyenne de 240.23 mg/l. Les valeurs oscillent entre 100 et 500 mg/l ce qui dépasse la concentration admissible par la STEP (350 mg/l).

La concentration des MES atteint parfois des valeurs très élevées (2400 mg/l), ceci peut être expliqué 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 inférieures à la norme de rejet (30mg/l), avec exception de certains cas où les valeurs sont supérieures à cette dernière, ceci est dû à des arrêts accidentels du fonctionnement.

A partir du 2013, on remarque quelques pics de la concentration des MES dans les eaux épurées qui peuvent atteindre même (160 mg/l) qui est une valeur très élevée par rapport à la norme du rejet. Ceci peut être expliqué par les forts débits qui arrivent à la station à partir de 2013 car cette dernière fonctionne en surcharge.

- Evolution de la Demande Biologique en Oxygène (DBO5)

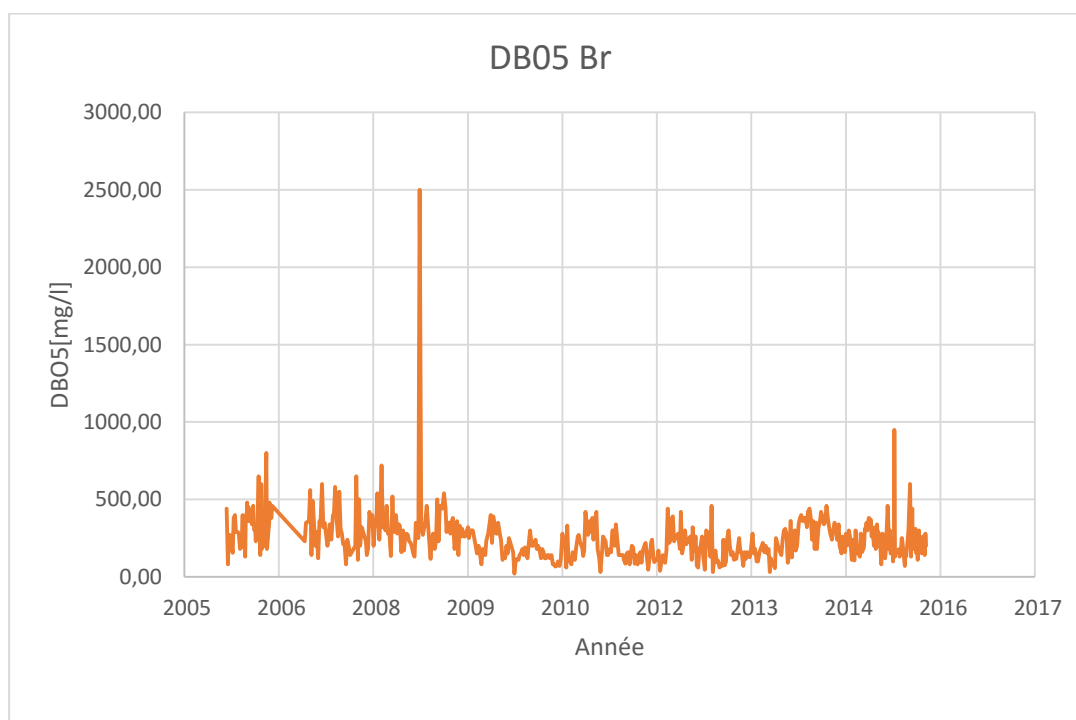


Figure IV. 17 : Evolution de la Demande Biologique en Oxygène à l'entrée de la STEP

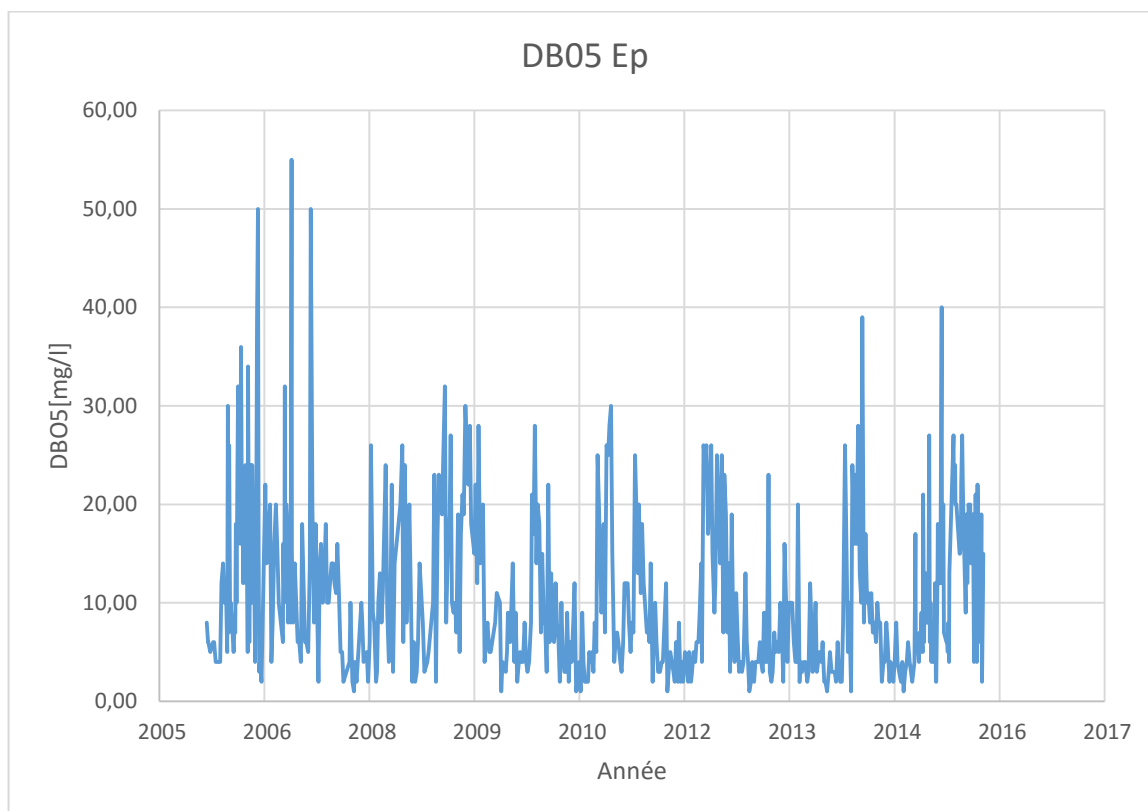


Figure IV. 18 : Evolution de la Demande Biologique en Oxygène à la sortie de STEP

A l'entrée de la STEP, la concentration de la Demande Biologique en Oxygène oscille entre 20 et 950 mg/l ce qui dépasse en majorité largement la concentration admissible par la STEP (270 mg/l) et avec une moyenne de 248.64 mg/l.

En 2008 une valeur douteuse de 2500 mg/l a été enregistrée, un test statistique sera appliqué pour vérifier la fiabilité de cette donnée.

A la sortie de la STEP, la concentration de la DBO5 est 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.

Chapitre IV : Présentation de la STEP de Boumerdes

- Evolution de la Demande Chimique en Oxygène (DCO)

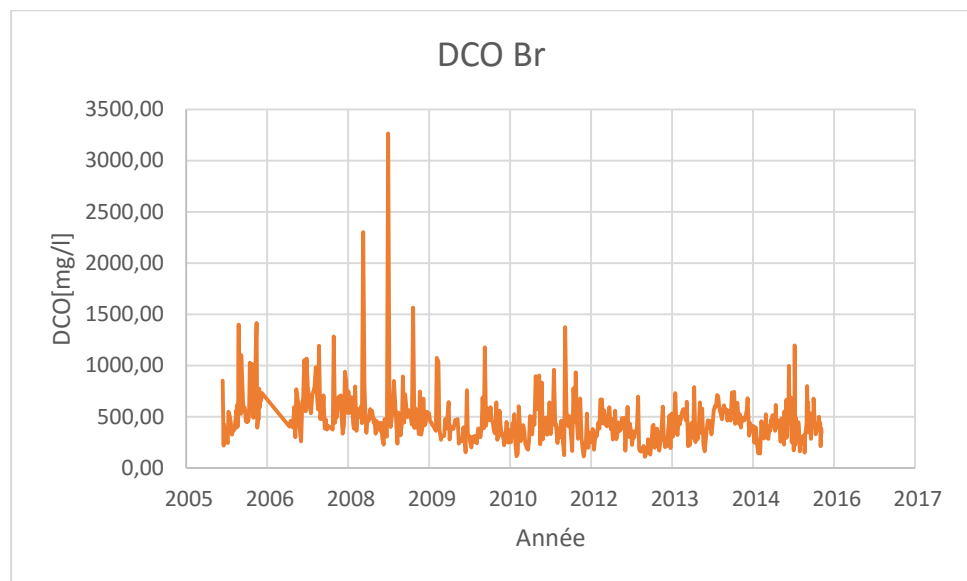


Figure IV. 19 : Evolution de la Demande Chimique en Oxygène à la sortie de la STEP

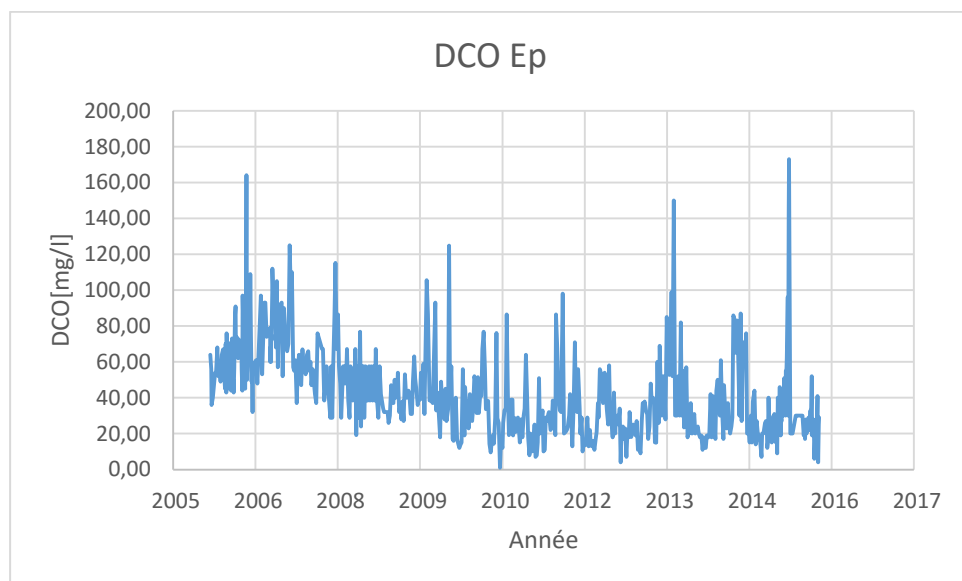


Figure IV. 20 : Evolution de la Demande Chimique en Oxygène à la sortie de la STEP

A l'entrée de la STEP, la concentration de la Demande Chimique en Oxygène (DCO) oscille entre 110 et 1500 mg/l, avec une moyenne de 473,16 mg/l.

Chapitre IV : Présentation de la STEP de Boumerdes

On remarque quelques valeurs douteuses qui dépassent largement la concentration admissible de la STEP (750 mg/l) et qui ont été enregistrées en 2008. Afin de vérifier la fiabilité de ces valeurs, un test statistique sera appliqué.

A la sortie de la STEP, la concentration de la DCO est généralement inférieure à la norme de rejet (90 mg/l) [29] ou bien (120 mg/l) [30].

On remarque aussi quelques exceptions dans les années 2006, 2013 et 2015 où les valeurs atteignent les 180 mg/l. Ceci est dû à une forte concentration en DCO qui ne peut être métabolisée par les bactéries (pollution industrielle).

- Evolution de l'ammonium (NH_4^+)

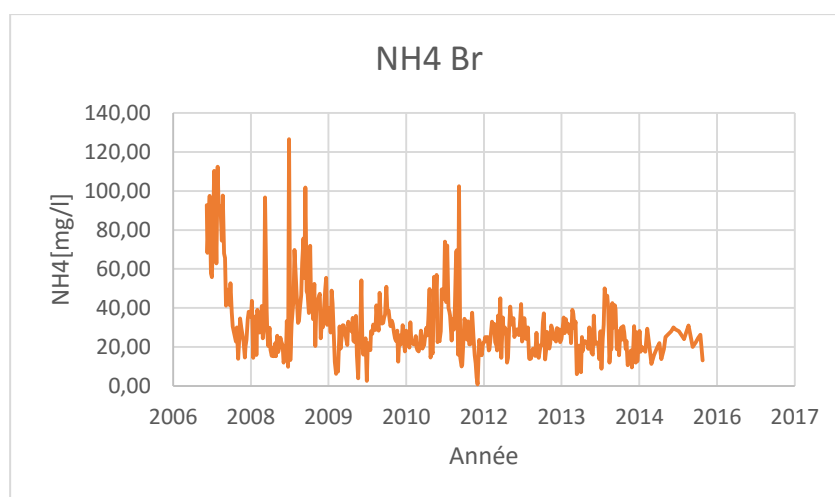


Figure IV. 21 : Evolution journalière de l'ammonium (NH_4^+) à l'entrée de la STEP

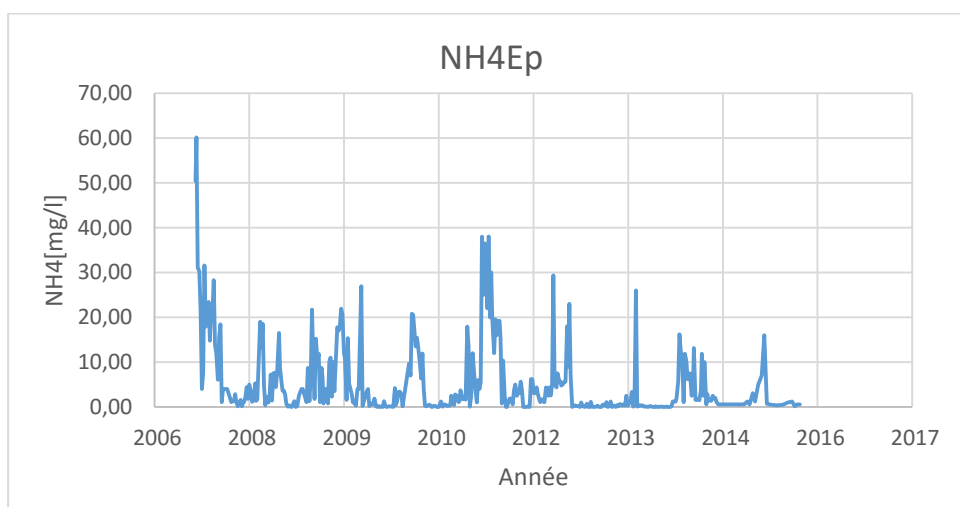


Figure IV. 22 : Evolution journalière de l'ammonium (NH_4^+) à la sortie de la STEP

Chapitre IV : Présentation de la STEP de Boumerdes

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, ces variations sont fonction de la température et du temps de séjour dans le réseau de collecte.

A la sortie de la STEP, la concentration d'ammonium varie entre 0 et 60.1 mg/l avec une moyenne de 5.4 mg/l qui est supérieure à la norme du rejet fixée par l'OMS à 5 mg/l. Ceci est dû à de fortes charges organiques qui entraînent nécessairement un degré de nitrification plus faible.

- Evolution du Nitrite (NO_2^-)

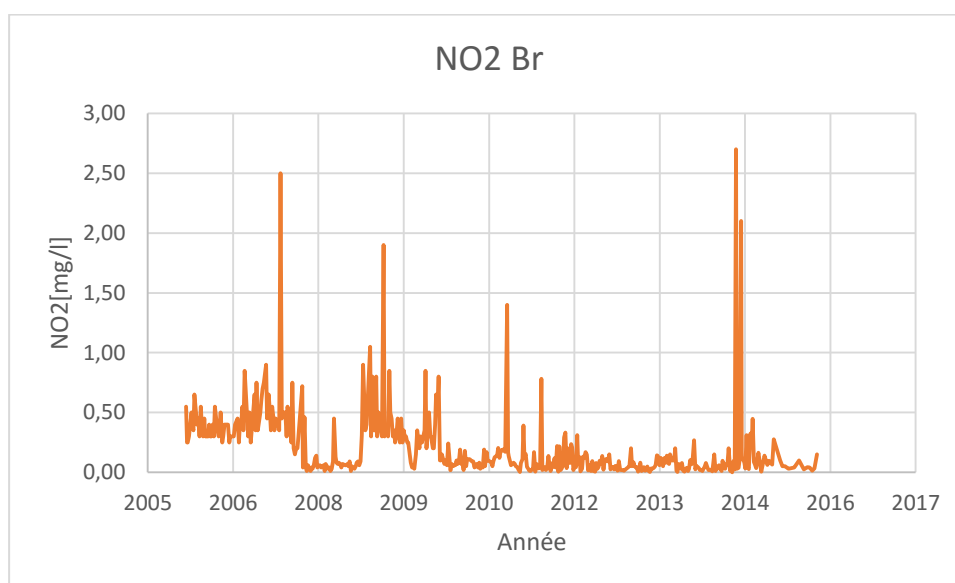


Figure IV. 23 : Evolution journalière du Nitrite (NO_2^-) à l'entrée de la STEP

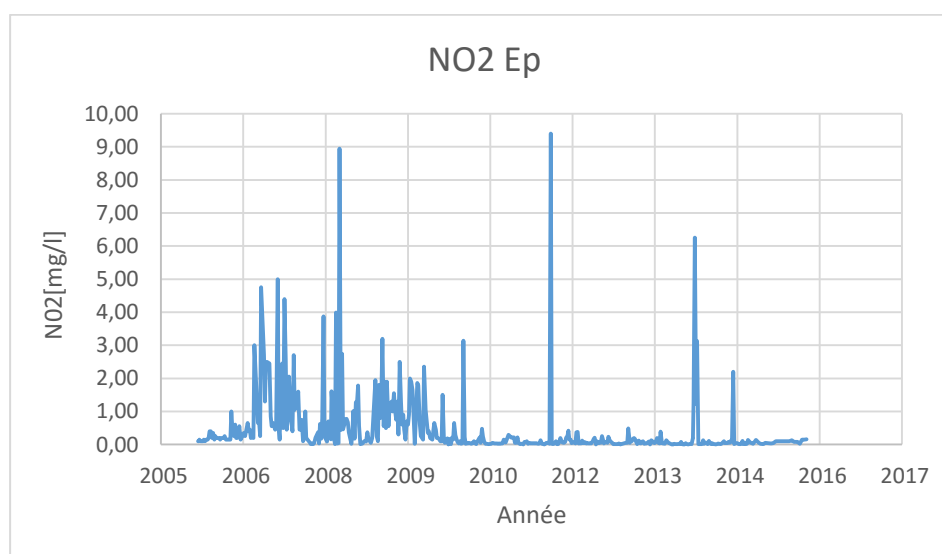


Figure IV. 24 : Evolution du Nitrite (NO_2^-) à la sortie de la STEP

Chapitre IV : Présentation de la STEP de Boumerdes

Les concentrations du Nitrite dans l'eau brute varient entre 0 et 1 mg/l avec quelques augmentations brutes dans les années suivantes : 2007, 2009, 2011 et 2014.

Les concentrations du Nitrite dans l'eau épurée varient entre 0 et 5 mg/l avec trois cas exceptionnels enregistrés lors des années : 2008, 2011, et 2013 où les valeurs atteignent les 9 mg/l.

Cela peut être expliqué soit par la forme du nitrite qui est instable durant le processus de nitrification, soit par l'importance des bactéries Nitrosomonas par rapport aux bactéries Nitrobacter ou bien elle est dû à une oxydation incomplète de l'ammonium.

- Variation du Nitrate (NO_3^-)

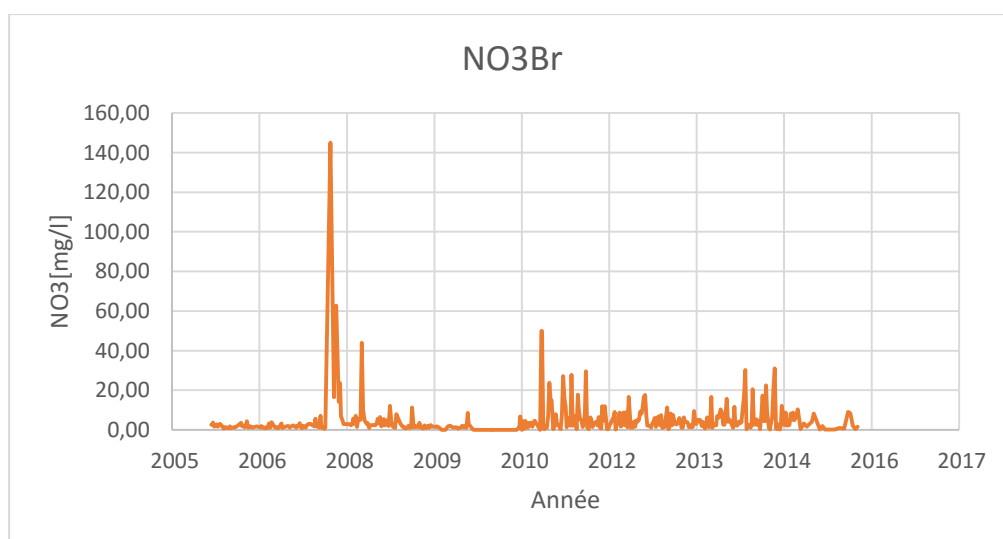


Figure IV. 25 : Evolution du Nitrate (NO_3^-) à l'entrée de la STEP

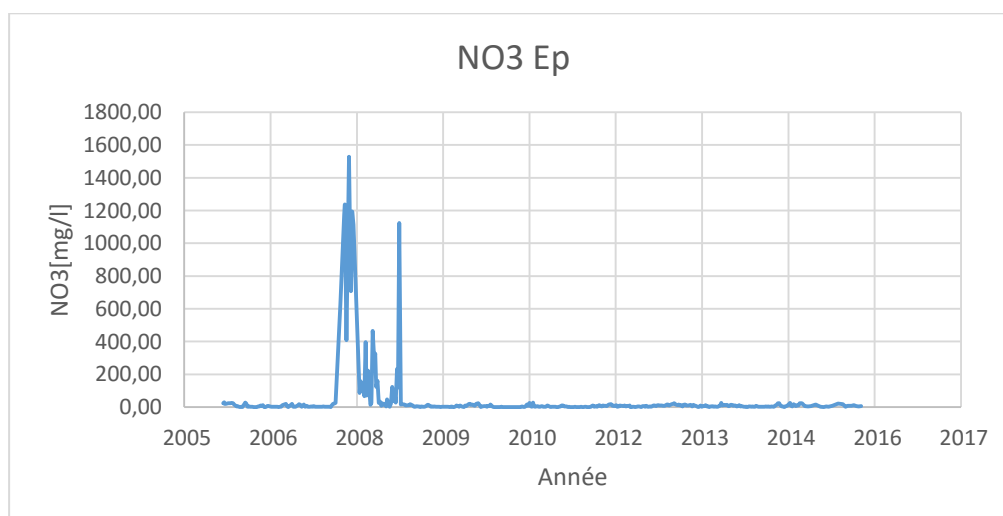


Figure IV. 26 : Evolution du Nitrate (NO_3^-) à la sortie de la STEP

Chapitre IV : Présentation de la STEP de Boumerdes

A l'entrée de la STEP, les concentrations du NO_3^- 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 NO_3^- oscillent entre 0 et 30 mg/l où elles dépassent les normes de rejet de l'OMS (10 mg/l).

La période 2007, 2008 est une période douteuse car les valeurs du NO_3^- atteignent (1600 mg/l). Ceci peut être expliqué par un problème de manque de maîtrise du temps d'aération.

- Evolution de l'azote Kjeldahl NTK

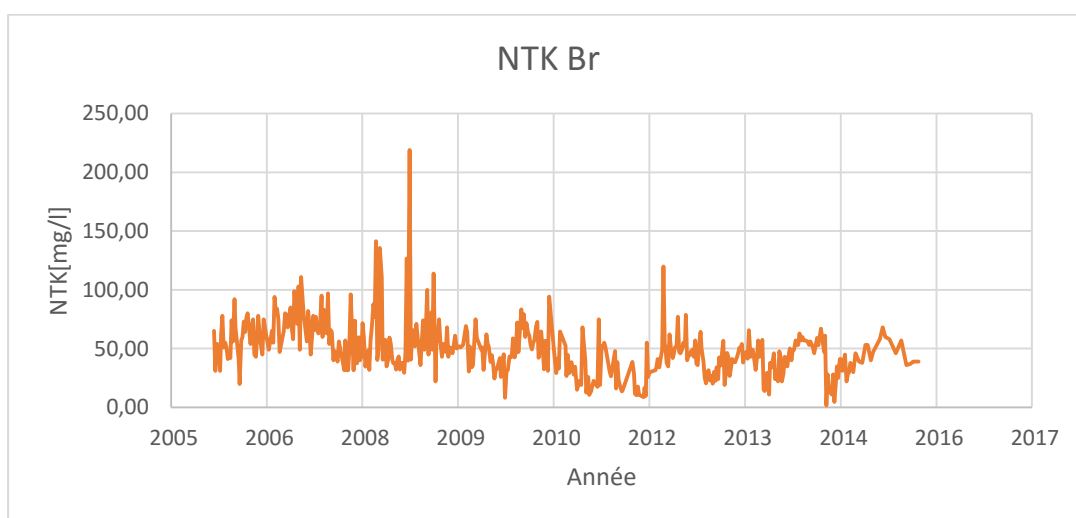


Figure IV. 27 : Evolution du NTK à l'entrée de la STEP

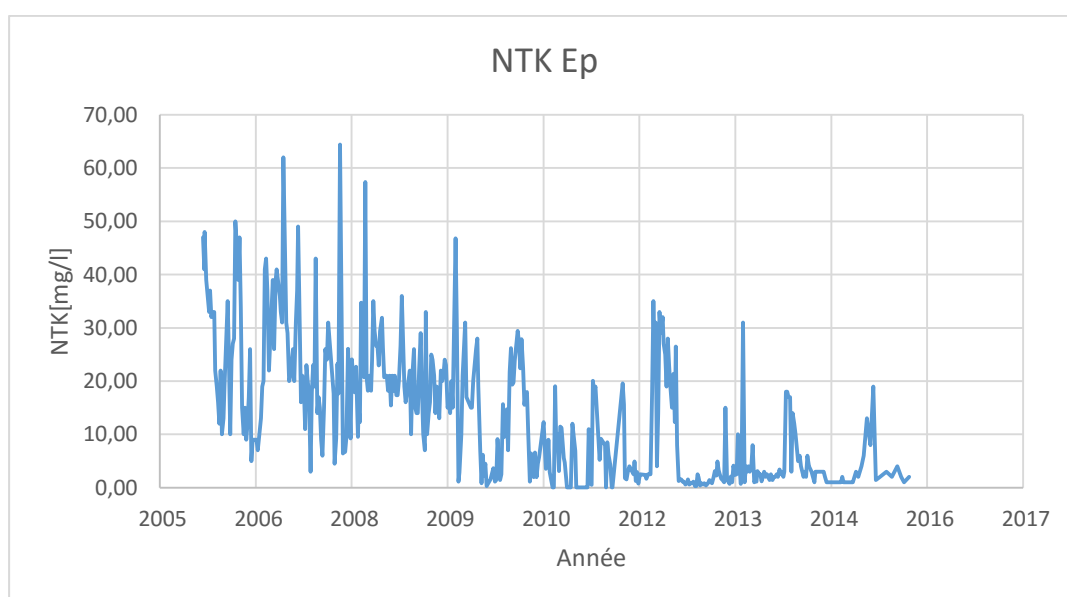


Figure IV. 28 : Evolution du NTK à la sortie de la STEP

Chapitre IV : Présentation de la STEP de Boumerdes

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).

- Evolution de l'ortho-phosphate (PO_4^{-3})

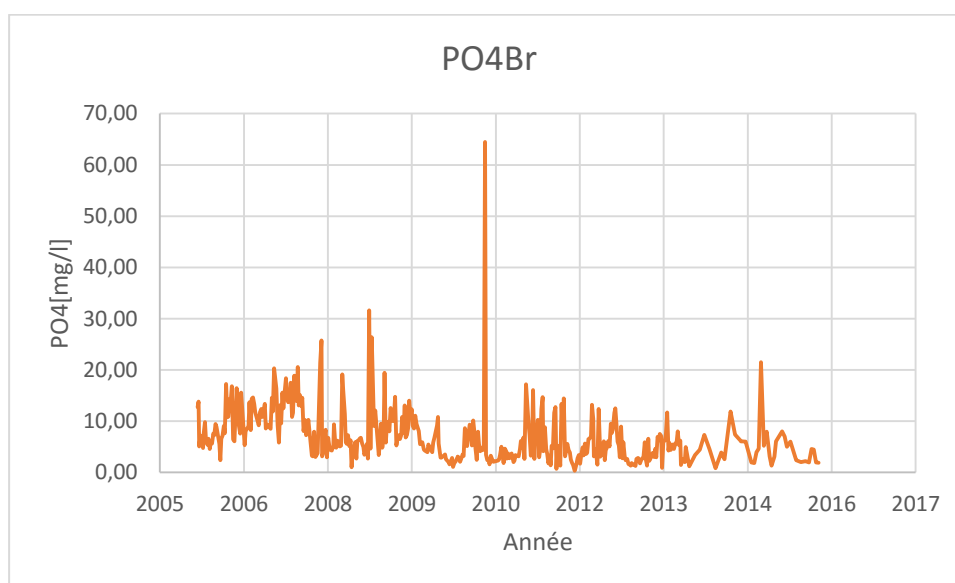


Figure IV. 29 : Evolution de PO_4^{-3} à l'entrée de la STEP

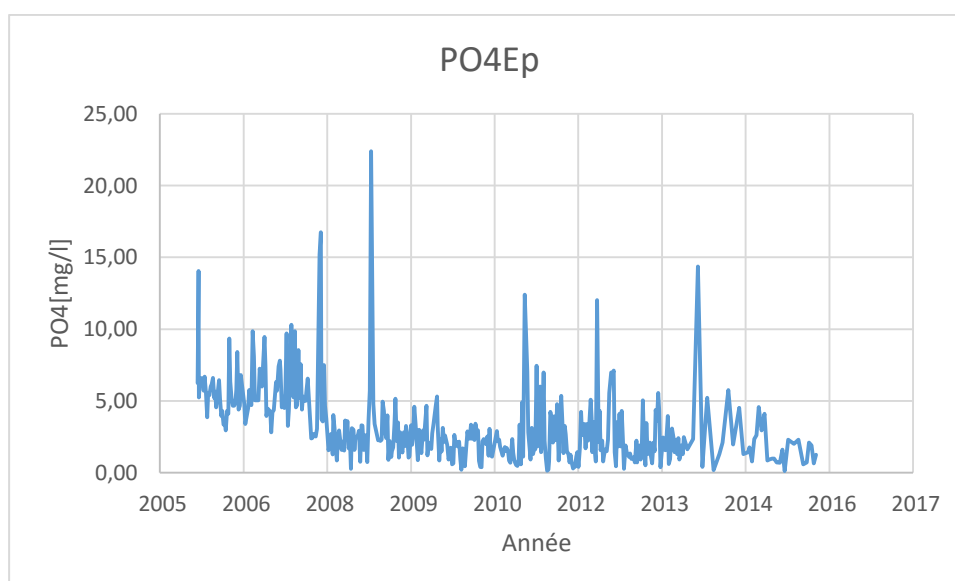


Figure IV. 30 : Evolution de PO_4^{-3} à la sortie de la STEP

Chapitre IV : Présentation de la STEP de Boumerdes

A l'entrée de la STEP, les concentrations du PO_4^{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.

- Evolution de débit journalier

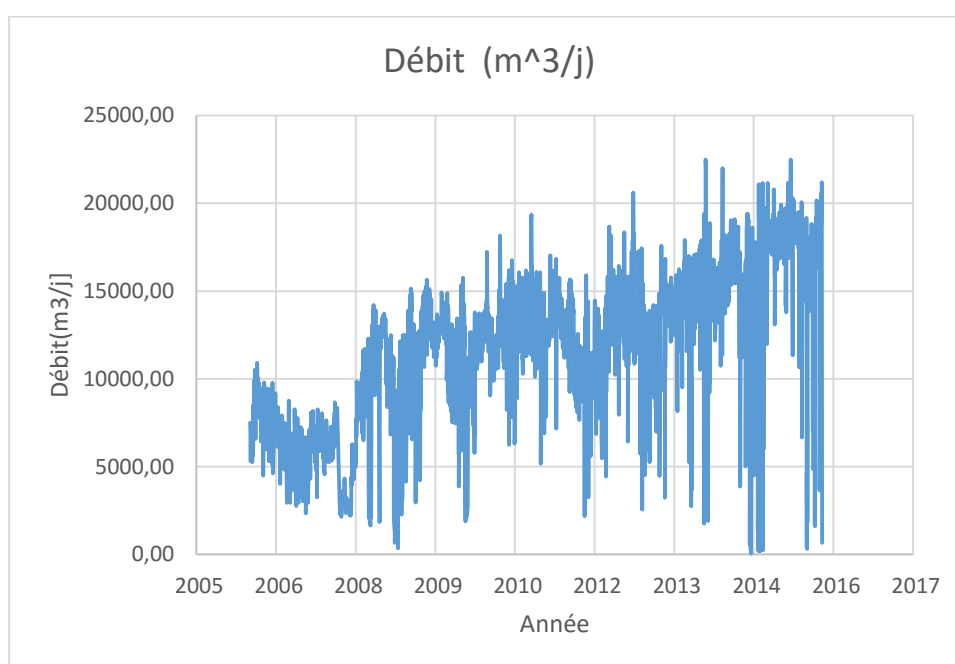


Figure IV. 31 : Evolution journalière du débit dans la STEP

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.

- Evolution de l'énergie

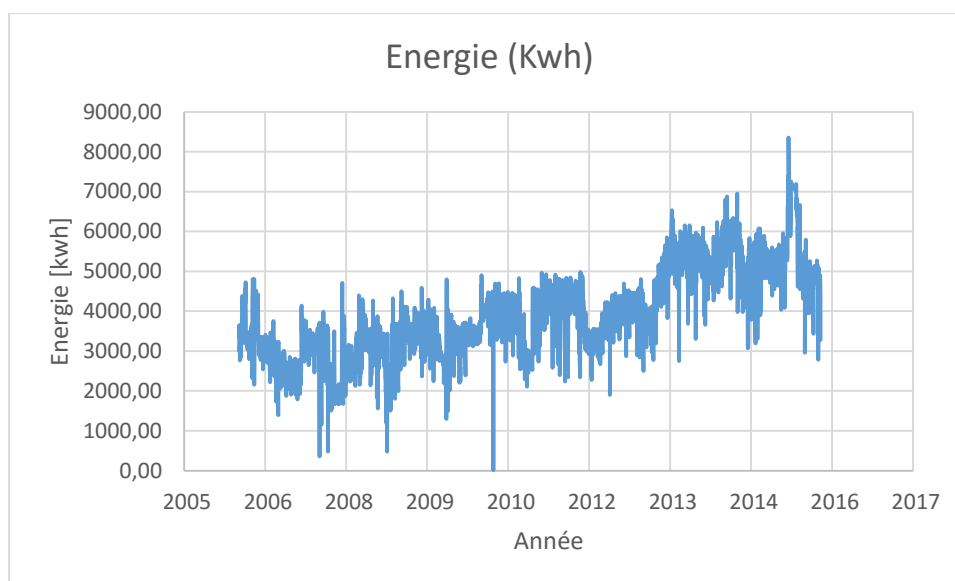


Figure IV. 32 : Evolution journalière de l'énergie dans la STEP

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.

IV.6.3. Détermination des valeurs singulières

Certaines valeurs douteuses enregistrées, peuvent avoir de mauvaises conséquences et peuvent même fausser toute l'étude statistique. Ceci peut être dû à une erreur lors de la saisie des données.

Afin d'éliminer ces valeurs, un test statistique est primordial pour la suite de toute analyse possible sur la série.

Parmi ces tests statistiques, nous avons utilisé est celui « Grubbs et Beck » qui permet de détecter les données douteuses et aberrantes et décider si on doit les éliminer ou les conserver.

IV.6.3.1. Test des horsains de Grubbs et Beck

Les valeurs singulières appelées horsains, sont déduites en calculant les statistiques suivantes :

$$X_{\text{high}} = \exp (x+K_n *S_x)$$

$$X_{\text{low}} = \exp (x-K_n *S_x)$$

Avec: x et S_x sont respectivement, la moyenne et l'écart type des logarithmes naturels des éléments constituant l'échantillon.

$$K_n = 3.62201 + 6.28446*n^{1/4} - 2.49835*n^{1/2} + 0.491436 * n^{3/4} - 0.037911 *n$$

N : taille de l'échantillon

Les valeurs de la série à utiliser dans l'étude doivent appartenir à l'intervalle $[X_{\text{low}}, X_{\text{high}}]$

Si une valeur n 'appartient pas à cet intervalle, on dira que c'est une valeur horsain.

IV.6.3.2. Résultat du test pour la STEP de Bomerdes

Le tableau suivant résume les résultats obtenus lors de l'aaplication de ce test sur la série de données de la STEP de Boumerdes.

Tableau IV. 3 : Résultats du test de Grubbs et Beck pour la STEP de Boumerdes

Variables	Eau Brute						
	x	Sx	n	Kn	Xh	Xl	
PH	1.99	0.019	2416	2.64	7.69	6.95	
T°	3	0.193	2416	2.63	33.35	12.03	
MES	5.37	0.454	2243	3.68	1147.26	40.41	
DBO5	5.37	0.551	537	10.44	67961.17	0.68	
DCO	6.06	0.446	552	10.88	54592.38	3.34	
NH4	3.87	0.557	407	10.54	17066.66	0.135	
NO2	-2.29	1.254	470	10.50	53307.75	0.00	
NO3	0.53	1.8	464	10.51	282017037	0.00	
NTK	3.78	0.503	442	10.87	10469.09	0.18	
PO4	1.69	0.712	426	10.53	9836.4	0.0029	

Chapitre IV : Présentation de la STEP de Boumerdes

		Eau Epurée					
PH		1.98	0.022	2413	2.65	7.7	6.86
T°		2.99	0.208	2412	2.66	34.61	11.44
MES		13.53	0.446	2248	3.65	150.12	1.47
DBO5		2.07	0.814	561	10.41	37783.18	0.00
DCO		3.57	0.613	581	10.38	20745.89	0.06
NH4		0.29	2.072	401	10.55	4148221152	0.00
NO2		-2.15	1.646	470	10.50	3797175.93	0.00
NO3		1.53	1.722	466	10.51	335810611	0.00
NTK		1.83	1.548	444	10.52	75103082.7	0.00
PO4		0.86	0.820	427	10.53	13398.57	0.00
Débit		9.3	0.508	3374	-3.95	81634.97	1464.57
Energie		8.24	0.308	3518	-5.04	17830	798.40

Après application du test Grubbs et Beck sur la série de données au pas de temps journalier, on a pu détecter Les valeurs singulières suivantes :

Tableau IV. 4 : Les valeurs singulières détectées

Eaux brutes	PH	T°	MES
	6.71	35.9	1320
	6.85	7.55	2250
	7.95	8.7	2400
	7.87		24
	8.08		
Eaux épurées	7.96	10.1	160
	9.91	2.43	
	8.03	9.4	
	6.56		
	6.39		
	6.74		

Chapitre IV : Présentation de la STEP de Boumerdes

Pour les paramètres de pollution, on peut déduire que les valeurs qui sont hors intervalle cité ci-dessous seront éliminées.

MES < 100 mg/l ou MES > 1000 mg/l

DBO5 < 100 mg/l ou DBO > 800 mg/l

DCO > 1100 mg/l

NH4 > 100 mg/l

NO3 > 100 mg/l

Pour le débit, les valeurs nulles représentent les jours où le débitmètre était en panne. Par contre les valeurs inférieures à 2000 m³/j représentent les jours où un By-Pass a été effectué au niveau de la STEP.

Les valeurs supérieures à 15000 m³/j représentent les années où la station a commencé à travailler en surcharge.

Et enfin, pour l'énergie les valeurs inférieures à 800 Kwh sont expliquées par des coupures d'électricité ou bien des pannes de fonctionnement au niveau de la STEP. Les valeurs moyennes sont expliquées par des jours d'inondations au niveau de la STEP.

Les valeurs supérieures à 5000 kwh représentent les années où la station a commencé à travailler en surcharge.

Chapitre V : Modélisation par ACP et Neuro-floue

V.1. Introduction

La modélisation par régression est une technique statistique permettant de modéliser la relation qui existe entre des variables explicatives (notées X_i) et une variable à expliquer (notée Y).

Pour cela trois modèles vont être utilisés dans notre étude :

- Régression par ACP
- Régression par Neuro-Floue
- Régression par couplage ACP Neuro-floue

V.2. Procédure de l'Analyse en Composantes Principales

V.2.1. Données utilisées

On dispose d'une série de données brutes des paramètres de pollution en (mg/l) : MES, DBO5, DCO, NH_4 , NO_3^- , NO_2^- , PO_4^- , NTK des eaux brutes et épurées (Entrée-Sortie), du débit de l'effluent (m^3/j) ainsi que de l'énergie totale (Kwh) consommée durant le processus d'épuration, de Janvier 2006 au 29 Février 2016 de la STEP de Boumerdes.

V.3. APPLICATION

V.3.1. Application de l'ACP sur les paramètres de pollution d'entrée (Eau brute)

V.3.1.1. Pas de temps journalier

➤ Présentation des données

Une analyse en composantes principales (ACP) a été effectuée sur le tableau des données journalières, qui contient 8 variables et 310 observations de l'eau brute, après avoir éliminé toutes les données non concomitantes.

Pour faciliter la manipulation des variables, on affecte à chacune d'elles un code qui l'identifie, la liste des variables avec leurs codes correspondants est présentée dans le tableau suivant :

Tableau V. 1 : Codification des variables (Eau brute)

Variable	Code
MES (Eau brute)	MESBr
DBO5 (Eau brute)	DBO5Br
DCO (Eau brute)	DCOBr
NH ₄ ⁺ (Eau brute)	NH4Br
NO ₂ ⁻ (Eau brute)	NO2Br
NO ₃ ⁻ (Eau brute)	NO3Br
NTK (Eau brute)	NTKBr
PO ₄ ⁻³ (Eau brute)	PO4Br

➤ Résultats de l'ACP

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

- Matrice de corrélation

Tableau V. 2 : Matrice de corrélation des paramètres de l'eau brute (journalier)

Corrélation	MESBr	DBO5Br	DCOBr	NH4Br	NO2Br	NO3Br	NTKBr	PO4Br
MESBr	1							
DBO5Br	0,648	1						
DCOBr	0,657	0,691	1					
NH4Br	0,314	0,424	0,529	1				
NO2Br	0,029	0,132	0,130	0,362	1			
NO3Br	0,133	0,038	0,031	-0,032	0,023	1		
NTKBr	0,463	0,571	0,589	0,515	0,157	0,001	1	
PO4Br	0,394	0,361	0,456	0,430	0,269	0,012	0,467	1

Nous allons nous intéresser à la matrice de corrélation. Il ne suffit pas d'avoir un tableau croisant des variables quantitatives et des individus pour appliquer une ACP ; encore faut-il que ce tableau vérifie au préalable certains critères.

Le calcul du déterminant se fait automatiquement sous XLSTAT. Il suffit, lorsque nous voulons faire une ACP, de le préciser dans les sorties. Il en est de même pour le test de sphéricité de Bartlett et l'indice K-M-O.

La matrice de corrélation a pour déterminant (0,05), on déduit qu'elle n'est pas singulière.

Chapitre V. Modélisation par ACP et Neuro-Floue

Pour le test de sphéricité de Bartlett et l'indice K-M-O, les résultats obtenus sous XLSTAT sont résumés dans le tableau ci-dessous :

Tableau V. 3 : Test de sphéricité de Bartlett et K-M-O de l'eau brute (journalier)

Mesure de précision de l'échantillonnage de Kaiser-Mayer-Olkin		0,834
Test de sphéricité de Bartlett	Khi-deux approximé	878,767
	ddl	28
	Signification de Bartlett	0,000

On remarque que la signification de Bartlett tend vers (0.000), et le KMO globale de la matrice se situe dans la fourchette «très bien» d'après l'échelle de valeur pour K-M-O

La matrice de corrélation est une matrice symétrique et diagonale, elle permet après diagonalisation de calculer les valeurs propres (en tenant compte des 8 variables), ainsi que la contribution à la variance en terme de pourcentage. Les résultats obtenus sont résumés dans le tableau suivant :

- Variance totale expliquée par les valeurs propres

Tableau V. 4 : Variance totale expliquée (paramètres de l'eau brute-journalier)

Composante	Valeurs propres initiales			Extraction des carrés des facteurs		
	Total	% de la variance	% cumulés	Total	% de la variance	% cumulés
1	3,602	45,020	45,020	3,602	45,020	45,020
2	1,177	14,706	59,726	1,177	14,706	59,726
3	1,005	12,567	72,293	1,005	12,567	72,293
4	0,628	7,846	80,139			
5	0,568	7,104	87,243			
6	0,438	5,476	92,719			
7	0,306	3,819	96,538			
8	0,277	3,462	100,000			

L'analyse de variance montre qu'à partir de (8) variables réelles constituant la matrice des données, on aboutit à trois (3) variables fictives, dites composantes principales restituant près de 72% de l'information globale, les autres composantes contribuent très faiblement à l'explication de la variance totale, alors leur rejet n'entraîne qu'une perte d'informations minime.

- Matrice de corrélation entre les CP' et les variables

Tableau V. 5 : Matrice de corrélation entre les CP' et les variables de l'eau brute

Variables	CP'		
	1	2	3
DCOBr	0,736	0,030	0,007
DBO5Br	0,656	0,059	0,007
NTKBr	0,612	0,000	0,011
MESBr	0,562	0,173	0,000
NH4Br	0,495	0,155	0,000
PO4Br	0,443	0,059	0,004
NO2Br	0,094	0,577	0,134
NO3Br	0,003	0,124	0,842

A partir de ce tableau, on remarque qu'il y a une bonne corrélation entre les variables (DCOBr, DBO5Br, NTKBr, MESBr) et la première composante d'une part et la variable NO2Br avec la deuxième et la troisième composante d'autre part.

Pour bien expliquer les constats tirés à partir du tableau (V.5), on passe aux représentations graphiques en traçant les cercles de corrélations dans les deux plans factoriels (1×2) et (1×3).

- Cercles de corrélation

Cercle (1) : Projection des variables sur le plan factoriel (1×2) :

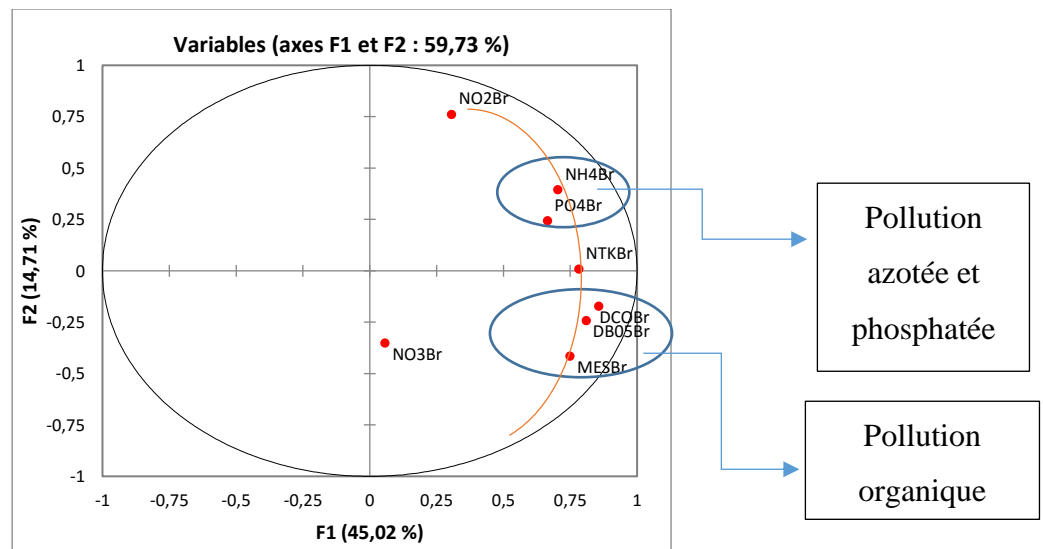


Figure V. 1 : Paramètres de pollution de l'eau brute à l'échelle journalière (1X2)

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

Chapitre V. Modélisation par ACP et Neuro-Floue

(PO4Br) caractérisant une pollution dissoute, l'autre formé par (DCOBr), (DBO5Br), (MESBr) caractérisant une pollution organique. On déduit alors que le plan factoriel formé par les axes (1 et 2) donne une information sur l'origine de la pollution.

En outre, on remarque que l'axe (1) définit un gradient de pollution décroissant de droite à gauche, et que les paramètres de la pollution organique 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.

Cercle (2) : Projection des variables sur le plan factoriel (1×3) :

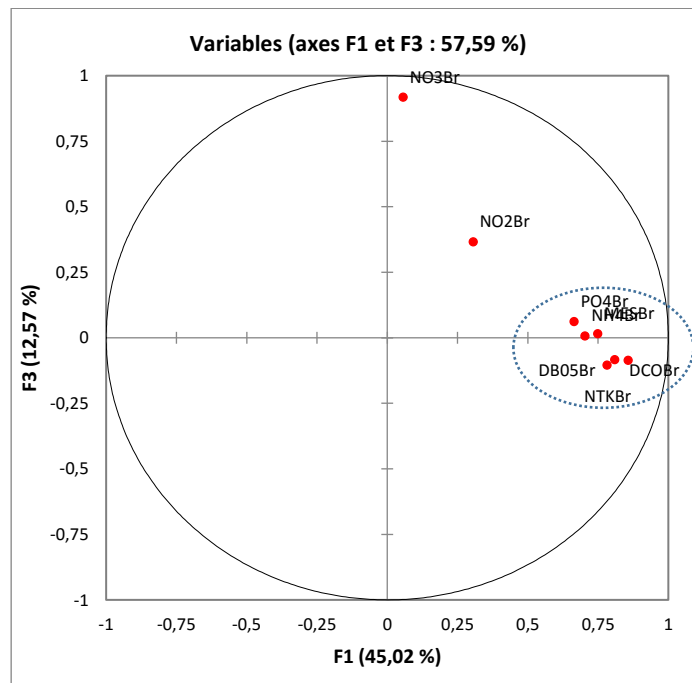


Figure V. 2 : Paramètres de pollution de l'eau brute à l'échelle journalière (1X3)

En examinant le cercle (2), on constate un regroupement de variables composées de : DBO5Br, DCOBr, NTKBr, MESBr, NH4Br et PO4Br autour de l'axe (1), celui-ci permet de former des ratios en combinant ces différents paramètres, par exemple le ratio DCO/NTK influence la cinétique de nitrification et donc le degré d'abattement de l'azote global. Plus ce ratio est faible, plus la vitesse de nitrification est élevée.

Le ratio NTK/DBO est lié à la cinétique de dénitrification. Une valeur faible de ce ratio désigne l'efficacité du procédé d'élimination de l'azote organique donc le taux de nitrification sera important.

Chapitre V. Modélisation par ACP et Neuro-Floue

Le ratio DBO5/NH4/PO4 (C/N/P) exprime l'équilibre nutritionnel, le métabolisme bactérien s'accompagne de besoins azotés sous forme d'azote ammoniacal et de besoins phosphorés sous forme d'orthophosphates dans des proportions bien précises.

En outre, on remarque que les paramètres NO2Br et NO3Br sont éloignés de l'axe (1) ces derniers représentent des formes oxydées de l'azote difficiles à maîtriser.

- Conclusion

Les résultats de l'analyse des paramètres de l'eau brute, au pas de temps journalier, ont montré que la pollution de l'eau qui arrive à la STEP est organique caractérisant un effluent domestique.

Le regroupement des variables autour d'un axe permet de former de nouveaux indicateurs de pollution sous forme de ratios, ces derniers permettent d'identifier les insuffisances et d'améliorer le rendement de différentes installations du traitement des eaux usées.

V.3.1.2. Pas de temps mensuel

- Présentation des données

On dispose d'un tableau de données mensuelles des paramètres de pollution à l'entrée de la STEP (Eau brute) : « 103 observations × 8 variables », après élimination des données non concomitantes.

- Résultats de l'ACP : les résultats de l'analyse sont présentés ci-dessous :

- Matrice de corrélation :

Tableau V. 6 : Matrice de corrélation des paramètres de l'eau brute (mensuel)

Corrélation	MES Br	DBO5 Br	DCO Br	NH4 Br	NO2 Br	NO3 Br	NTK Br	PO4 Br
MES Br	1							
DBO5 Br	0,555	1						
DCO Br	0,606	0,696	1					
NH4 Br	0,156	0,341	0,553	1				
NO2 Br	-0,088	0,131	0,131	0,499	1			
NO3 Br	0,134	-0,007	-0,010	-0,052	0,261	1		
NTK Br	0,344	0,605	0,634	0,570	0,214	-0,096	1	
PO4 Br	0,291	0,342	0,576	0,590	0,344	-0,062	0,482	1

Chapitre V. Modélisation par ACP et Neuro-Floue

La matrice de corrélation a pour déterminant (0,05), nous déduisons qu'elle est singulière. Pour le test de sphéricité de Bartlett et l'indice K-M-O, les résultats obtenus sous XLSTAT sont résumés dans le tableau ci-dessous :

Tableau V. 7 : Test de sphéricité de Bartlett et K-M-O de l'eau brute (mensuel)

Mesure de précision de l'échantillonnage de Kaiser-Mayer-Olkin		0,749
Test de sphéricité de Bartlett	Khi-deux approximé	351,988
	ddl	28
	Signification de Bartlett	0,000

On remarque que la signification de Bartlett tend vers (0.000), et le KMO globale de la matrice se situe dans la fourchette «bien».

L'étape suivante consiste à diagonaliser la matrice de corrélation, pour pouvoir calculer les valeurs propres.

- Variance totale expliquée par les valeurs propres

Tableau V. 8 : Variance totale expliquée (paramètres de l'eau brute-mensuel)

Composante	Valeurs propres initiales			Extraction des carrés des facteurs		
	Total	% de la variance	% cumulés	Total	% de la variance	% cumulés
1	3,588	44,847	44,847	3,588	44,847	44,847
2	1,422	17,777	62,625	1,422	17,777	62,625
3	1,148	14,356	76,980	1,148	14,356	76,980
4	0,589	7,358	84,338			
5	0,422	5,279	89,617			
6	0,351	4,385	94,001			
7	0,292	3,644	97,645			
8	0,188	2,355	100,000			

En examinant l'apport de chaque composante, on remarque qu'avec trois (3) CP seulement, on peut expliquer 77 % de la variance totale.

L'étape suivante consiste à déterminer la corrélation entre les composantes principales centrées réduites (CP') et les variables centrées réduites des données brutes.

- Matrice de corrélation entre les CP' et les variables

Tableau V. 9 : Matrice de corrélation entre les CP' et les variables de l'eau brute-échelle mensuelle

Variables	CP'		
	1	2	3
DCO Br	0,779	0,051	0,001
NTK Br	0,648	0,001	0,029
DBO5 Br	0,587	0,112	0,015
NH4 Br	0,549	0,198	0,044
PO4 Br	0,537	0,060	0,030
MES Br	0,351	0,297	0,140
NO2 Br	0,137	0,642	0,037
NO3 Br	0,000	0,062	0,852

En analysant la matrice de corrélation entre les trois (3) CP' et les variables centrées réduites des variables brutes, on remarque que : Les variables DCOBr, DBO5Br, NTKBr sont bien corrélées avec la première composante.

La variable NO2Br est corrélée avec la deuxième composante, alors que la variable NO3Br est bien corrélée avec la troisième composante.

Pour visualiser les résultats tirés à partir de la matrice de corrélation entre les CP' et les variables, on passe aux représentations graphiques en traçant les différents cercles de corrélation.

- Cercles de corrélation

Cercle (1) : Projection des variables sur le plan factoriel (1x2) :

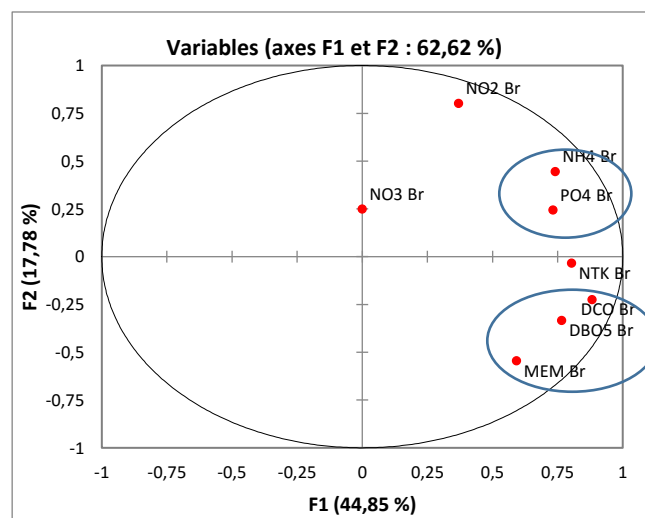


Figure V. 3 : Paramètres de pollution de l'eau brute à l'échelle mensuelle (1X2)

Chapitre V. Modélisation par ACP et Neuro-Floue

Le cercle de corrélation formé par les axes principaux (1 et 2), montre toujours la formation des deux groupes opposés par rapport au premier axe. L'un formé par les composés azotés et phosphatés, l'autre caractérisant une pollution organique. Ce qui explique toujours l'information sur l'origine de la pollution.

Cercle (2) : Projection des variables sur le plan factoriel (1×3) :

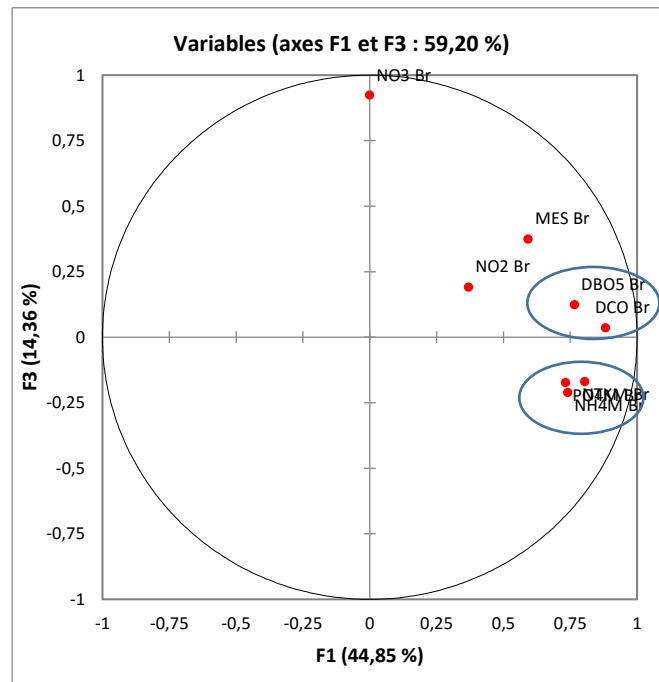


Figure V. 4 : Paramètres de pollution de l'eau brute à l'échelle mensuelle (1X3)

En analysant le cercle de corrélation formé par les deux axes principaux (1 et 3), on remarque que les variables NH4Br, PO4Br et NTKBr sont très proches formant même un ratio NH4/NTK/PO4 qui exprime l'équilibre entre les besoins azotés sous forme d'azote ammoniacal accompagné de l'azote de Kjeldahl et les besoins phosphorés sous forme d'ortho-phosphates dans des proportions bien précises.

On remarque aussi que les variables DCOBr et DBO5Br sont proches formant un ratio DCO/DBO5 qui permet d'évaluer la biodégradabilité de la matière organique d'un effluent donné. En effet, une valeur faible du ratio DCO/DBO5 implique la présence d'une grande proportion de matières biodégradables ce qui permet d'envisager un traitement biologique.

Conclusion : L'analyse des paramètres de l'eau brute au pas de temps mensuel, a permis d'obtenir de nouveaux indicateurs de pollution sous forme de ratios qui synthétisent d'avantage les différents paramètres, et permettent une meilleure appréciation de la pollution.

V.3.2. Application de l'ACP sur les paramètres de l'eau (Entrée/Sortie)

V.3.2.1. Pas de temps journalier

➤ Présentation des données

Une analyse en composantes principales (ACP) a été effectuée sur le tableau des données journalières, qui contient 16 variables et 291 observations de l'eau brute et épurée, après avoir éliminé toutes les données non concomitantes.

On affecte à chacune de ces variables un code qui l'identifie.

Tableau V. 10 : Codification des variables (Eau brute- Eau épurée)

Variable	Code
Paramètre eau brute	Paramètre Br
Paramètre eau épurée	Paramètre Ep

➤ Résultats de l'ACP

- Matrice de corrélation

La matrice de corrélation a pour déterminant (0,05), nous déduisons qu'elle est singulière. et Pour le test de sphéricité de Bartlett et l'indice K-M-O, les résultats obtenus sous XLSTAT sont résumés dans le tableau ci-dessous :

Tableau V. 11 : Test de sphéricité de Bartlett et K-M-O (Eau brute –Eau épurée / journalier)

Mesure de précision de l'échantillonnage de Kaiser-Mayer-Olkin		0,728
Test de sphéricité de Bartlett	Khi-deux approximé	1126,509
	ddl	120
	Signification de Bartlett	0,000

On remarque que la signification de Bartlett tend vers (0.000), et le KMO globale de la matrice se situe dans la fourchette «bien». L'étape suivante consiste à diagonaliser la matrice de corrélation, pour pouvoir calculer les valeurs propres.

Chapitre V. Modélisation par ACP et Neuro-Floue

- Variance expliquée par les valeurs propres

Tableau V. 12 : Variance totale expliquée (Eau brute – Eau épurée / journalier)

Composante	Valeurs propres initiales			Extraction des carrés des facteurs		
	Total	% de la variance	% cumulés	Total	% de la variance	% cumulés
1	3,873	24,204	24,204	3,873	24,204	24,204
2	1,864	11,651	35,855	1,864	11,651	35,855
3	1,383	8,645	44,500	1,383	8,645	44,500
4	1,213	7,582	52,082			
5	1,118	6,989	59,071			
6	0,976	6,098	65,169			
7	0,861	5,381	70,549			
8	0,802	5,010	75,559			
9	0,732	4,576	80,135			
10	0,650	4,065	84,200			
11	0,595	3,720	87,920			
12	0,514	3,213	91,133			
13	0,466	2,915	94,048			
14	0,393	2,458	96,505			
15	0,305	1,903	98,409			
16	0,255	1,591	100,000			

- Corrélation entre les CP' et les variables centrées réduites

**Tableau V. 13 : Matrice de corrélation entre les CP' et les variables (Eau brute-
Eau épurée / journalier)**

Variables	CP'		
	1	2	3
NH4Br	0,590	0,058	0,000
NTKBr	0,484	0,007	0,000
DCOBr	0,470	0,027	0,019
PO4Br	0,440	0,006	0,245
DB05Br	0,385	0,001	0,009
NH4Ep	0,378	0,144	0,076
PO4Ep	0,216	0,007	0,087
NO2Br	0,194	0,017	0,018
DB05Ep	0,169	0,207	0,226
MESBr	0,155	0,231	0,062
DCOEp	0,138	0,006	0,274
NO2Ep	0,130	0,001	0,003
NTKEp	0,066	0,188	0,163
MESEp	0,044	0,002	0,047
NO3Ep	0,009	0,499	0,052
NO3Br	0,005	0,463	0,101

- Cercle de corrélation

Cercle (1) : Projection des variables sur le plan factoriel (1×2) :

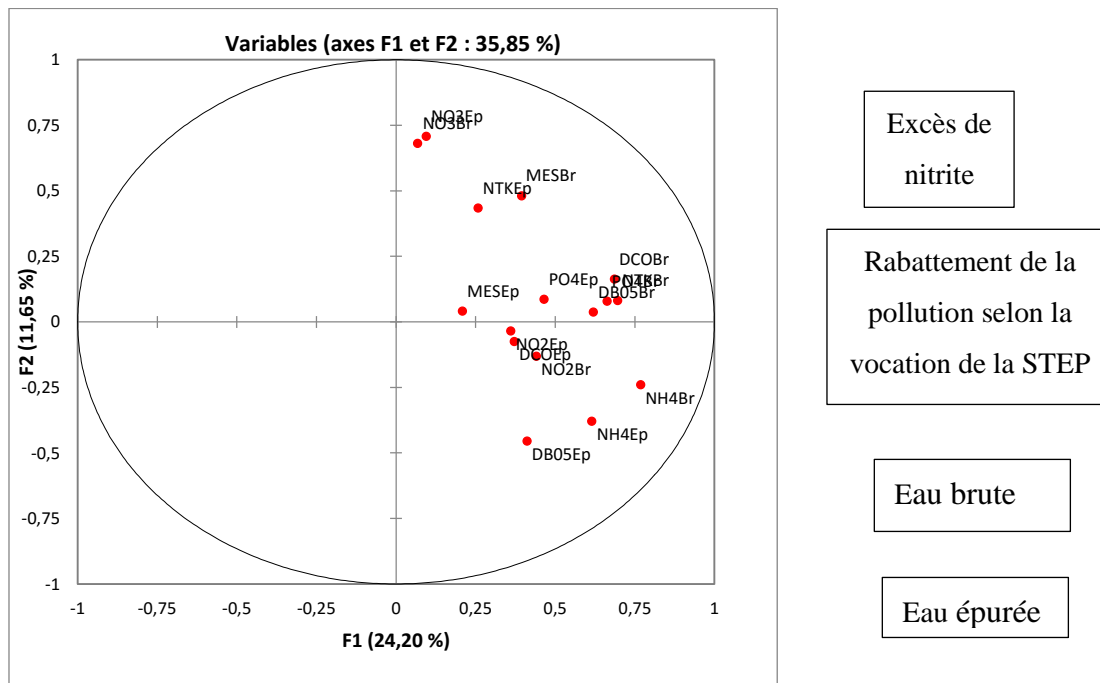


Figure V. 5 : Paramètres de pollution (Entrée-sortie) à l'échelle journalière (1X2)

La projection des variables sur le plan factoriel (1×2) montre que les paramètres de pollution de l'eau brute et épurée sont disposés sur deux arcs fictifs respectivement, tel que Les équidistances entre ces deux arcs fictifs définissent le rabattement de la pollution selon la vocation de la STEP, contrairement au paramètre NO2 dont le processus n'est pas bien maîtrisé vu qu'il se présente en excès.

V.3.2.2. Pas de temps mensuel

- Présentation des données

On dispose d'un tableau de données mensuelles «101 observations × 16 variables». Après application de l'ACP et la vérification de trois critères, on obtient les mêmes résultats qu'à l'échelle journalière avec de légères différences. On se limite à montrer le cercle de corrélation dans le plan factoriel formé par les axes principaux (1 et 2).

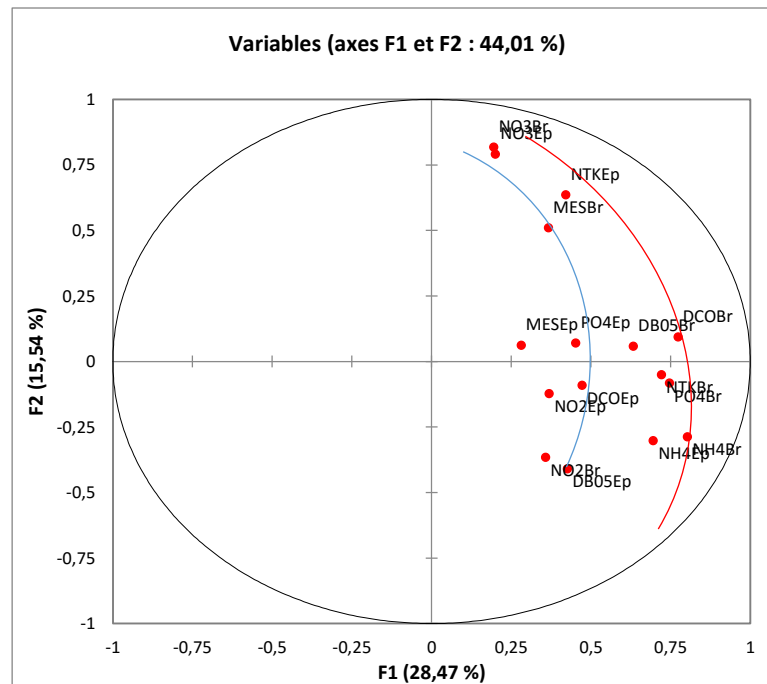


Figure V. 6 : Paramètres de pollution (Entrée – sortie) à l'échelle mensuelle (1X2)

On remarque une disposition des paramètres de l'eau brute et épurée sur deux arcs fictifs respectivement, avec des exceptions pour NO₂ qui sont en excès.

➤ Conclusion

Les résultats de l'analyse des paramètres de l'eau brute et épurée, au pas de temps journalier et mensuel, ont montré que le manque de maîtrise au niveau de la filière biologique conduit à un dépassement des normes de rejets du paramètre NO₂ à la sortie de la STEP, et ont permis de mettre en évidence les faibles performances du processus de dénitrification et l'absence de la zone anaérobie.

V.4. Modélisation de la STEP de Boumerdes

Introduction

La STEP de Boumerdes est une STEP à boues activées, son but principal est la dégradation de la charge polluante mais la difficulté réside dans la détermination de la dose optimale à injecter en fonction des caractéristiques de l'eau brute afin d'éliminer cette pollution. Un mauvais contrôle de ce procédé peut entraîner une augmentation importante des coûts de fonctionnement et le non-respect des objectifs de qualité en sortie de la station de traitement.

Dans le but de déterminer l'énergie à déployer, trois différents modèles ont été établis soit : modèle par ACP, modèle ANFIS et modèle couplage ACP Neuro-floue.

Chapitre V. Modélisation par ACP et Neuro-Floue

Dans l'ensemble des cas traités, la série de données sera divisée en deux parties. Une partie réservée pour le calage (apprentissage) du modèle et qui contient 318 observations au pas de temps journalier et l'autre pour la validation, contenant 51 observations au pas de temps journalier aussi et qui représente les données de l'année 2013. Le choix de cette année a été fait à partir du fait qu'elle représente l'année de transition, autrement dit l'année de passage au travail en surcharge de la STEP.

Le choix des paramètres s'est fait comme suit, par ordre d'élimination on a :

NO₂, NO₃ : ont la forme la plus oxydée de l'azote, leurs apparition dépend de la disparition des autres paramètres azotés. Ces derniers peuvent ne pas être représentatifs de la quantité d'énergie consommée donc on les a éliminés.

NTK (NH₄, N_{org}) : engendre plusieurs formes azotées et comme l'augmentation de la concentration d'une de ces formes n'est pas forcément relié à la concentration des autres, la valeur de NTK peut aussi ne pas être représentative, donc on l'a également éliminée.

PO₄ : la STEP de Boumerdes est une STEP à boues activées, les bactéries qui dégradent la matière organique ne sont pas forcément adaptées pour l'élimination de la pollution phosphorée. Ce paramètre aussi n'a pas été pris en considération.

Donc les (4) autres paramètres qui restent: MES (mg/l), DBO₅ (mg/l), DCO (mg/l) et NH₄ (mg/l) sont ceux qui influent sur l'énergie déployée.

Une fois que le choix des paramètres a été fait, on passe au calcul du rabattement de chacun d'eux comme suit :

$$\text{Rabattement MES} = [\text{MES}]_{\text{entrée}} - [\text{MES}]_{\text{sortie}} \quad (\text{mg/l})$$

$$\text{Rabattement DBO}_5 = [\text{DBO}_5]_{\text{entrée}} - [\text{DBO}_5]_{\text{sortie}} \quad (\text{mg/l})$$

$$\text{Rabattement DCO} = [\text{DCO}]_{\text{entrée}} - [\text{DCO}]_{\text{sortie}} \quad (\text{mg/l})$$

$$\text{Rabattement NH}_4 = [\text{NH}_4]_{\text{entrée}} - [\text{NH}_4]_{\text{sortie}} \quad (\text{mg/l})$$

V.4.1. La modélisation par ACP

V.4.1.1. Application de l'ACP

L'Analyse en Composantes Principales (ACP) est une méthode statistique qui permet de réduire la dimension d'une matrice de données. En effet, elle transforme un premier jeu de

Chapitre V. Modélisation par ACP et Neuro-Floue

données en un second jeu de plus petite dimension composé de nouvelles variables qui sont des combinaisons linéaires des variables originelles. [38]

➤ Présentation des données

Dans un premier lieu, on doit centrer et réduire notre matrice de données (calage et validation) pour ce faire, on prendra la moyenne et l'écart type de la partie calage uniquement car on suppose qu'on ne possède pas les données de la partie validation.

Une analyse en composante principale (ACP) a été effectuée seulement aux données de la partie calage de la STEP.

On affecte à chacune de ces variables un code qui l'identifie.

Tableau V. 14 : Codification des variables de rabatement

Variable	Code
Rabattement MES	RMES
Rabattement DBO5	RDBO5
Rabattement DCO	RDCO
Rabattement NH4	RNH4

➤ Vérification des critères d'application de l'ACP

Pour le test de sphéricité de Bartlett et l'indice K-M-O, les résultats obtenus sous XLSTAT sont résumés dans le tableau ci-dessous :

Tableau V. 15 : Test de sphéricité de Bartlett et K-M-O (rabatement des variables)

Mesure de précision de l'échantillonnage de Kaiser-Mayer-Olkin		0,686
Test de sphéricité de Bartlett	Khi-deux approximé	253,654
	ddl	10
	Signification de Bartlett	0,000

On remarque que la signification de Bartlett tend vers (0.000), et le KMO globale de la matrice se situe dans la fourchette «moyen».

L'étape suivante consiste à diagonaliser la matrice de corrélation, pour pouvoir calculer les valeurs propres.

- Variance expliquée par les valeurs propres

Tableau V. 16 : Variance totale expliquée (Rabattement des variables)

Composante	Valeurs propres initiales			Extraction des carrés des facteurs		
	Total	% de la variance	%cumulés	Total	% de la variance	% cumulés
1	2.182	43.647	43.647	2.182	43.647	43.647
2	0.963	19.260	62.907	0.963	19.260	62.907
3	0.833	16.664	79.571	0.833	16.664	79.571
4	0.588	11.752	91.323			
5	0.434	8.677	100.00			

L'analyse de variance montre qu'à partir de (5) variables réelles constituant la matrice des données, on aboutit à trois (3) variables fictives (celles dont la valeur propre est proche ou supérieure à 1), dites composantes principales restituant près de 79,5% de l'information globale, les autres composantes contribuent très faiblement à l'explication de la variance totale, alors leur rejet n'entraîne qu'une perte d'informations minime.

- Matrice des cosinus carrés des variables

Tableau V. 17 : Matrice des cosinus carrés des variables

	F1	F2	F3	F4	F5
RMES	0,448	0,011	0,332	0,160	0,049
RDBO5	0,414	0,340	0,001	0,180	0,065
RDCO	0,693	0,019	0,001	0,000	0,287
RNH4	0,376	0,003	0,499	0,091	0,031
Energie	0,251	0,590	0,000	0,156	0,003

A partir de la matrice des cosinus carrés des variables on peut déduire que :

L'axe F1 explique respectivement le rabattement des MES, DBO5 et DCO.

L'axe F2 explique l'énergie.

L'axe F3 explique le rabattement de NH4.

V.4.1.2. Modèle issu de l'ACP

➤ Matrice de passage

Dans le but d'obtenir la matrice de passage, on a procédé comme suit :

- ✓ A partir des résultats de l'ACP effectuée sur les cinq paramètres (MES, DBO5, DCO, NH4 et Energie) et depuis le tableau des coordonnées des observations, on a effectué à l'aide de XLSTAT une régression multiple entre chaque axe (F1, F2, F3, F4 et F5) et les données centrées réduites (MES, DBO5, DCO, NH4 et Energie) et cela axe par axe.
- ✓ Une fois qu'on obtiendra l'équation de chaque axe en fonction des paramètres centrés réduits, on aura une matrice dite A.
- ✓ A l'aide du logiciel Matlab, on calculera l'inverse de A et qui doit être égale à sa transposée.
- ✓ L'inverse de la matrice A est notre matrice de passage.

Les résultats obtenus de cette démarche sont les suivants :

- ✓ Les équations obtenues à partir de la régression multiple de chaque axe sont les suivantes :

$$F1 = 0,453 * RMES_{CR} + 0,436 * RDBO5_{CR} + 0,564 * RDCO_{CR} + 0,415 * RNH4_{CR} + 0,339 * Energie_{CR}$$

$$F2 = 0,107 * RMES_{CR} - 0,595 * RDBO5_{CR} - 0,14 * RDCO_{CR} + 0,056 * RNH4_{CR} + 0,784 * Energie_{CR}$$

$$F3 = -0,632 * RMES_{CR} - 0,041 * RDBO5_{CR} - 0,026 * RDCO_{CR} + 0,774 * RNH4_{CR} - 0,005 * Energie_{CR}$$

$$F4 = -0,522 * RMES_{CR} + 0,554 * RDBO5_{CR} - 0,028 * RDCO_{CR} - 0,394 * RNH4_{CR} + 0,515 * Energie_{CR}$$

$$F5 = -0,335 * RMES_{CR} - 0,386 * RDBO5_{CR} + 0,814 * RDCO_{CR} - 0,266 * RNH4_{CR} - 0,082 * Energie_{CR}$$

Chapitre V. Modélisation par ACP et Neuro-Floue

✓ La matrice A est donc la suivante :

	F1	F2	F3	F4	F5
RMES	0,453	0,436	0,564	0,415	0,339
RDBO5	0,107	-0,595	-0,14	0,056	0,784
RDCO	-0,632	-0,041	-0,026	0,774	-0,005
RNH4	-0,522	0,554	-0,028	-0,394	0,515
Energie	-0,335	-0,386	0,814	-0,266	-0,082

✓ L'inverse de la matrice A :

Les résultats obtenus par MATLAB sont les suivants

	F1	F2	F3	F4	F5
RMES	0,454	0,107	-0,631	-0,521	-0,334
RDBO5	0,435	-0,593	-0,041	0,554	-0,386
RDCO	0,563	-0,140	-0,027	-0,029	0,813
RNH4	0,415	0,057	0,773	-0,394	-0,266
Energie	0,339	0,782	-0,005	0,515	-0,083

✓ On voit bien que l'inverse de la matrice A est égale sa transposée.

On remarque aussi que la matrice de passage obtenue n'est autre que la matrice des vecteurs propres obtenue par ACP.

Tableau V. 18 : Matrice des vecteurs propres

	F1	F2	F3	F4	F5
RMES	0,453	0,108	-0,632	-0,522	-0,334
RDBO5	0,435	-0,594	-0,042	0,554	-0,386
RDCO	0,564	-0,140	-0,027	-0,028	0,813
RNH4	0,415	0,056	0,774	-0,394	-0,266
Energie	0,339	0,783	-0,005	0,515	-0,083

Maintenant on passe à l'élaboration de notre modèle ACP.

Depuis la matrice de passage on a écrit les F_i (F1, F2 et F3) en fonction des X_i (RMES, RDBO5, RDCO, RNH4, Energie).

$$F1 = 0,453 \cdot RMES + 0,435 \cdot RDBO5 + 0,564 \cdot RDCO + 0,415 \cdot RNH4 + 0,339 \cdot E$$

$$F2 = 0,108 \cdot RMES - 0,594 \cdot RDBO5 - 0,14 \cdot RDCO + 0,056 \cdot RNH4 + 0,783 \cdot E$$

Chapitre V. Modélisation par ACP et Neuro-Floue

$$F3 = -0,632*RMSE - 0,042*RDB5 - 0,027*RDCO + 0,774*RNH4 - 0,005*E$$

On a pris seulement les trois premières composantes car avec ces composantes on a une explication de l'information de 79%

Ensuite on a écrit les X_i en fonction des F_i (F_1 , F_2 et F_3). Entre autre la variable à expliquer l'énergie.

$$RMES = 0,453*F1 + 0,108*F2 - 0,632*F3$$

$$RDBO5 = 0,435*F1 - 0,594*F2 - 0,042*F3$$

$$RDCO = 0,564*F1 - 0,14*F2 - 0,027*F3$$

$$RNH4 = 0,415*F1 + 0,056*F2 + 0,774*F3$$

$$E = 0,339*F1 + 0,783*F2 - 0,005*F3$$

En remplaçant F_1 , F_2 et F_3 avec leurs équations on obtiendra l'équation de l'énergie en fonction des (5) variables.

$$E_{SIM} = 0,24*RMES_{CR} - 0,32*RDBO5_{CR} + 0,082*RDCO_{CR} + 0,18*RNH4_{CR} + 0,73*Energie_{CR}$$

➤ Modélisation de l'énergie

Une fois qu'on a obtenu le modèle de l'ACP, on remplacera les variables (RMES, RDBO5, RDCO, RNH4 et Energie) par leurs valeurs dans la partie calage et validation. On obtiendra par la suite l'énergie simulée par ACP centrée réduite.

Afin de faire une régression linéaire entre l'énergie brute observée et l'énergie brute simulée des deux parties calage et validation, on doit faire le passage de l'énergie simulée centrée réduite à l'énergie simulée brute. Pour ce faire on multiplie l'énergie simulée centrée réduite par l'écart type de la partie calage et on ajoute la moyenne de la partie calage (moy= 0,37 et Ect= 0,17)

➤ Régression linéaire

Une fois qu'on a calculé l'énergie simulée par ACP, on représente le nuage de points de l'énergie simulée brute en fonction de l'énergie observée brute et on trace la courbe de tendance.

Les résultats obtenus sont représentés dans les figures suivantes :

➤ *Représentation graphique des résultats la partie calage :*

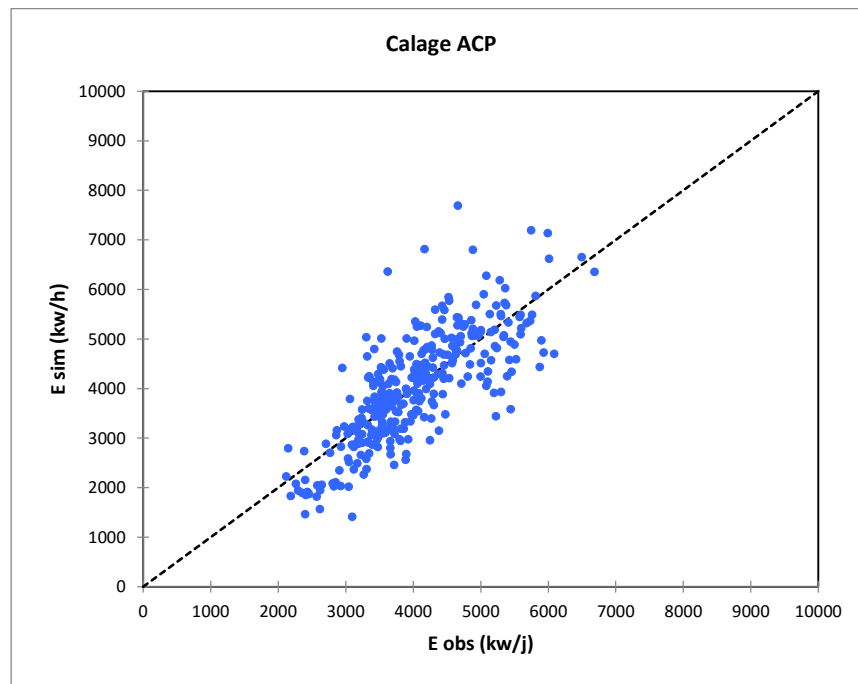


Figure V. 7 : Régression linéaire (Energie brute- Energie simulée par ACP) / calage

Le passage de l'énergie simulée par ACP centrée réduite à l'énergie simulée brute nous a permis d'étendre notre échantillon ce qui rend la visualisation plus simple.

D'après la représentation graphique des résultats obtenus pour la partie calage on voit que les valeurs simulées s'étalent sur un intervalle de]1500, 8000[donc on remarque une surestimation de l'énergie. Ceci peut être dû à la qualité des données car dans la partie calage on possède des données des années où la station a commencé à travailler en surcharge.

On remarque aussi qu'une partie des points est proche la première bissectrice, par contre, pas mal de points valeurs sont visiblement très loin de cette dernière.

➤ **Représentation graphique des résultats de la partie validation :**

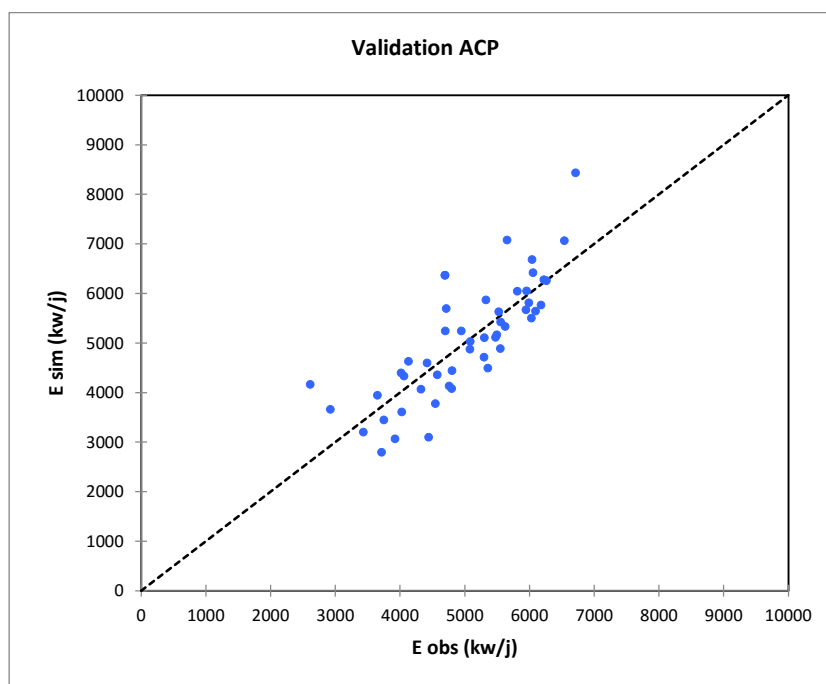


Figure V. 8 : Régression linéaire (Energie brute- Energie simulée par ACP) / validation

Pour la représentation graphique des résultats obtenus pour la partie validation on peut voir que l'échantillon est bien étendu. Une grande partie des points sont proches de la première bissectrice mais on voit quand même qu'il y a quelques points qui sont loin de cette dernière, donc on peut juger que le modèle obtenu par ACP est de qualité moyenne.

Dans le but de confirmer notre jugement sur la qualité du modèle, on s'est proposé de calculer numériquement le coefficient de corrélation (R^2) ainsi que (PBIAS) pour les deux parties : calage et validation.

Les résultats obtenus sont représentés dans le tableau suivant :

Tableau V. 19 : Résultats des critères de validation- Simulation par ACP (calage/validation)

	Calage	Validation
R^2	60,6%	63,86%
PBIAS	-1,09	-6,4

➤ *Interprétation des résultats de la partie calage :*

Une fois qu'on a calculé le coefficient de corrélation (R^2) on a obtenu une valeur de (60,6%) et donc la corrélation entre l'énergie observée et l'énergie simulée par ACP est moyenne.

Le $PBIAS < 0$ ceci montre que certaines valeurs simulées sont supérieures aux valeurs observées.

On peut conclure que la qualité du modèle issu de l'ACP pour la partie calage est moyenne et ceci confirme l'interprétation de la distribution du nuage de points autour de la courbe de tendance.

➤ *Interprétation des résultats de la partie validation :*

Le coefficient de corrélation est de (63,86%) pour la partie validation, on remarque une légère augmentation par rapport au R^2 obtenu dans la partie calage et donc la corrélation entre l'énergie observée brute et l'énergie simulée brute est moyenne.

Le $PBIAS$ est toujours inférieur à zéro et ceci montre que dans les valeurs de la partie validation, certaines sont surestimées.

On peut conclure que la modélisation par ACP nous donne une simulation de l'énergie de qualité moyenne.

V.4.2. Modélisation par Neuro-Floue (Modèle ANFIS)

Introduction :

Le modèle ANFIS (Système d'inférence flou à base de réseaux de neurones adaptifs) est un modèle qui combine les techniques floues et neuronales en formant un réseau à apprentissage supervisé, et donc il est appliqué durant la phase calage et testé en période de validation.

V.4.2.1. Elaboration du modèle Neuro-floue

Pour ce faire, on a appliqué le modèle sur des données centrées réduites ensuite on a fait le chemin inverse pour obtenir l'énergie simulée brute.

Pour générer un système d'inférence flou, c'est-à-dire notre modèle, nous avons fait usage de l'outil de génération de systèmes neuro-flou dans Matlab « The Neuro-Fuzzy Design Toolbox ».

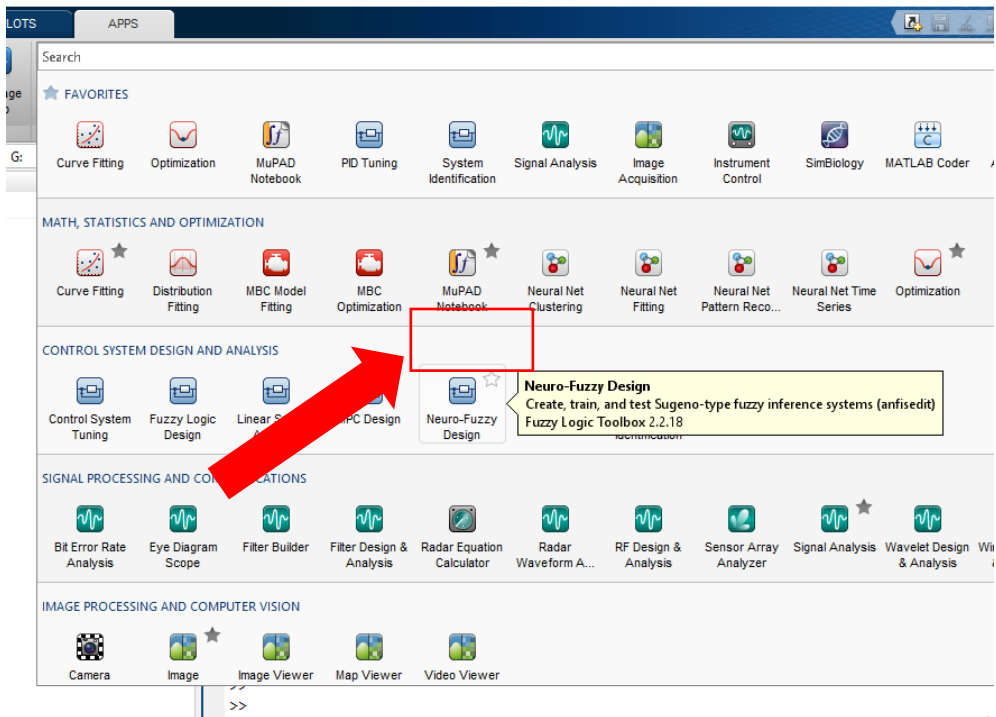


Figure V. 9 : Icône Neuro-floue dans MATLAB

On fait les jeux de données de calage et de validation, et on choisit une fonction d'appartenance et le nombre de règles floues.

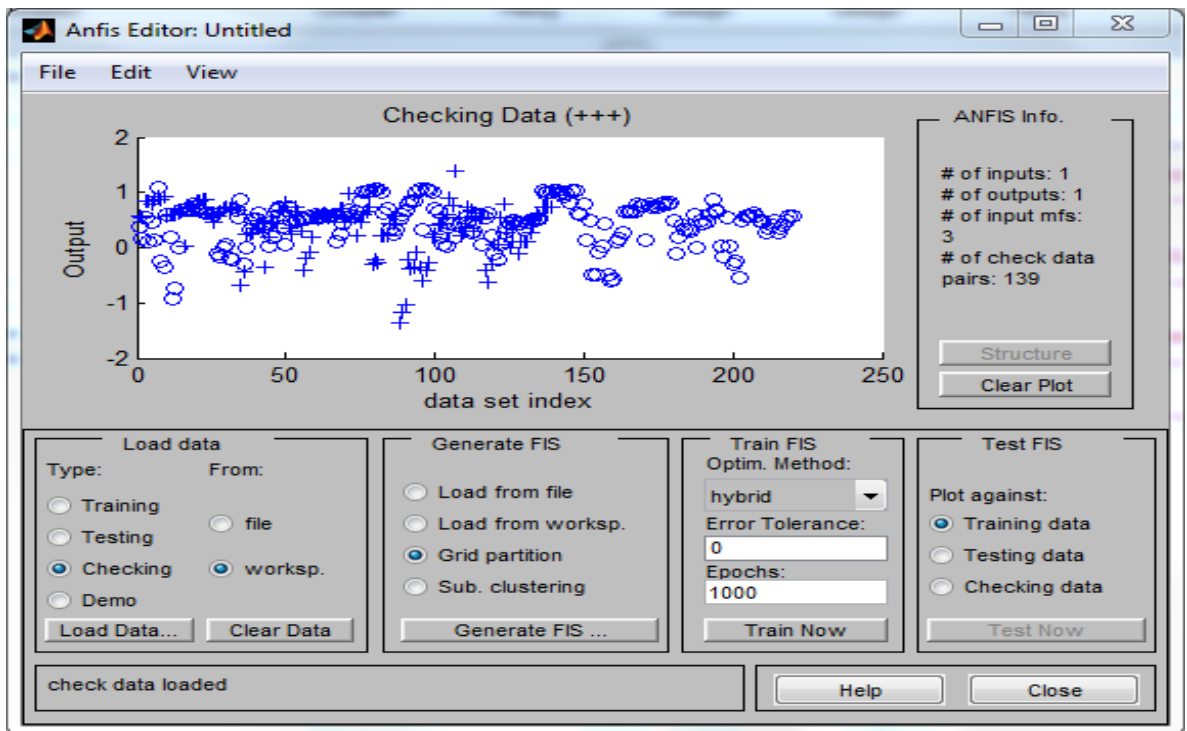


Figure V. 10 : Interface de l'application Neuro-fuzzy design toolbox

Après avoir complété l'apprentissage du modèle, on procède à son utilisation dans un programme qui a comme sorties les deux colonnes de l'énergie simulée par Neuro-floue centrée réduite de la partie calage ainsi que celle de la partie validation.

Afin de revenir à des données de l'énergie simulée brute on multiplie les résultats obtenus dans la partie du calage ainsi que la validation par l'écart type de la partie calage et on ajoute la moyenne de la partie calage.

➤ Régression linéaire

Une fois qu'on a obtenu les résultats de l'énergie simulée par Neuro-floue, on a fait une représentation graphique du nuage de points de l'énergie brute observée (kwh) et l'énergie brute simulée par Neuro-floue de la partie calage et la partie validation.

Les résultats sont présentés dans les figures suivantes :

➤ *Représentation graphique des résultats de la partie calage :*

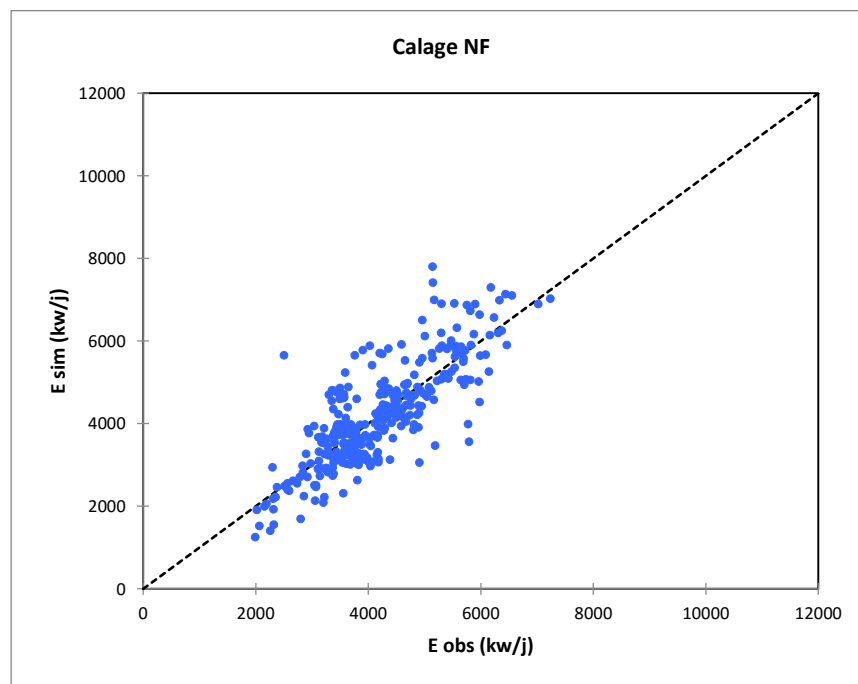


Figure V. 13 : Régression linéaire (Energie brute-Energie simulée par NF) / calage

D'après la représentation graphique du nuage de points de l'énergie brute simulée par Neuro-floue en fonction de l'énergie brute observée, on remarque que pas mal de points sont presque superposés sur la première bissectrice cela montre que l'écart entre l'énergie observée et l'énergie simulée est très petit.

On remarque aussi que certains points sont très loin de la première bissectrice et donc on peut dire que ces points n'ont pas été bien simulés et que l'écart entre l'énergie observé et l'énergie simulée est assez important.

➤ *Présentation graphique des résultats de la partie validation :*

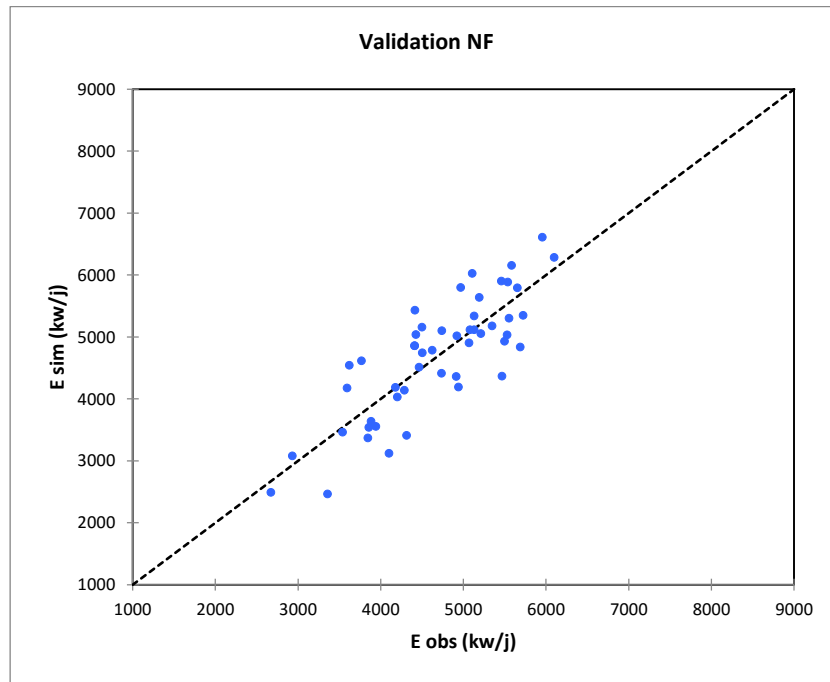


Figure V. 14 : Régression linéaire (Energie brute- Energie simulée par NF) / validation

D'après la représentation graphique des résultats de la modélisation par Neuro-floue de la partie validation, on voit que les points sont proches de la première bissectrice et donc on peut dire que le modèle Neuro-floue a bien simulé l'énergie.

On remarque aussi qu'il n'y a plus de surestimation de l'énergie, ceci peut être expliqué par le fait que les systèmes Neuro-floue sont connus pour leur réduction des erreurs.

Dans le but de confirmer notre jugement sur la qualité du modèle, on s'est proposé de calculer numériquement le coefficient de corrélation (R^2) ainsi que (PBIAS) pour les deux parties : calage et validation.

Les résultats obtenus sont représentés dans le tableau suivant :

Tableau V. 20 : Résultats des critères de validation- Simulation par Neuro-floue (calage/validation)

	Calage	Validation
R ²	66%	68%
PBIAS	-5,02	1,1

➤ **Interprétation des résultats obtenus dans la partie calage :**

Une fois qu'on a calculé le coefficient de corrélation (R²) on a obtenu une valeur de (66%) et donc la corrélation entre l'énergie observée et l'énergie simulée par Neuro-floue est moyenne avec une légère amélioration par rapport à celle obtenue par ACP.

On remarque que la valeur du PBIAS<0 ceci montre que le modèle surestime l'énergie.

On peut conclure que la qualité du modèle Neuro-floue pour la partie calage est moyenne mais il s'est amélioré par rapport au modèle issu de l'ACP et ceci confirme l'interprétation de la distribution du nuage de points autour de la courbe de tendance.

➤ **Interprétation des résultats de la partie validation :**

Le coefficient de corrélation est de (68%) pour la partie validation, on remarque une augmentation par rapport au R² obtenu dans la partie calage et donc la corrélation entre l'énergie observée et l'énergie simulée est meilleure. Par contre le PBIAS>0 donc le modèle sous-estime l'énergie.

On peut conclure que la modélisation par Neuro-floue nous donne une bonne simulation de l'énergie si on ne prend pas en considération la surestimation de certaines valeurs.

V.4.3. Couplage ACP Neuro-floue

On a vu précédemment que le modèle Neuro-floue a donné de meilleurs résultats par rapport au modèle issu de l'ACP et afin d'essayer d'améliorer davantage les résultats on a pensé au couplage ACP Neuro-floue.

D'après les résultats de l'application de l'ACP, on remarque que l'énergie est bien représentée sur l'axe F2. Donc on prendra l'axe F2 comme variable à expliquer et les autres axes (F1, F3, F4 et F5) comme variables explicatives.

Chapitre V. Modélisation par ACP et Neuro-Floue

Afin d'obtenir notre matrice de données (calage et validations) on a procédé comme suit :

- ✓ Les données de la parties calage sont ceux obtenues par ACP (matrice des coordonnées des observations)
- ✓ Les données de la partie validation ont été calculées à partir de la matrice de passage :

$$F1 = 0,453*RMSE + 0,435*RDBO5 + 0,564*RDCO + 0,415*RNH4$$

$$F3 = -0,632*RMES - 0,042*RDBO5 - 0,027*RDCO + 0,774*RNH4$$

$$F4 = -0,522*RMES + 0,554*RDBO5 - 0,028*RDCO - 0,394*RNH4$$

$$F5 = -0,334*RMES - 0,386*RDBO5 + 0,813*RDCO - 0,266*RNH4$$

Pour les (4) axes on a pris $E=0$ et qui est égale à l'énergie moyenne car on travaille avec des données centrées réduites.

$$F2 = 0,108*RMES - 0,594*RDBO5 - 0,140*RDCO + 0,056*RNH4 + 0,783*E$$

Comme l'énergie est représentée par l'axe (2) donc on l'a pris en considération dans l'équation du F2.

En remplaçant RMES, RDBO5, RDCO, RNH4 et l'énergie par leurs valeurs on obtiendra la matrice des données de la partie validation.

Ensuite on passe à la simulation de l'axe F2 par le modèle Neuro-floue.

On génère un système d'inférence flou à l'aide de MATLAB, c'est-à-dire notre modèle, On fait usage de l'outil de génération de systèmes neuro-flou « The Neuro-Fuzzy Design Toolbox ».

On fait les jeux de données de calage et de validation, et on choisit une fonction d'appartenance et le nombre de règles floues.

Après avoir complété l'apprentissage du modèle, on procède à son utilisation dans un programme qui a comme sorties les deux colonnes de F2 simulée par Neuro-floue de la partie calage ainsi que celle de la partie validation.

A partir de la matrice de passage, on obtient le modèle suivant qui nous permettra de simuler l'énergie à partir d'un couplage ACP Neuro-floue pour la partie calage ainsi que pour la partie validation :

$$E_{\text{sim}} = 0,339 * F1 + 0,783 * F2_{\text{sim}} - 0,005 * F3 + 0,515 * F4 - 0,083 * F5$$

Afin de faire une régression linéaire entre l'énergie brute observée et l'énergie brute simulée des deux parties calage et validation, on doit faire le passage de l'énergie simulée centrée réduite à l'énergie simulée brute. Pour ce faire on multiplie l'énergie simulée centrée réduite par l'écart type de la partie calage et on ajoute la moyenne de la partie calage (moy= 0,37 et Ect= 0,17).

➤ Régression linéaire

Une fois qu'on a obtenu les résultats de l'énergie simulée par couplage ACP Neuro-floue, on a fait une régression linéaire entre l'énergie brute observée et l'énergie brute simulée de la partie calage et la partie validation.

Les résultats sont présentés dans les figures suivantes :

➤ *Représentation graphique des résultats de la partie calage :*

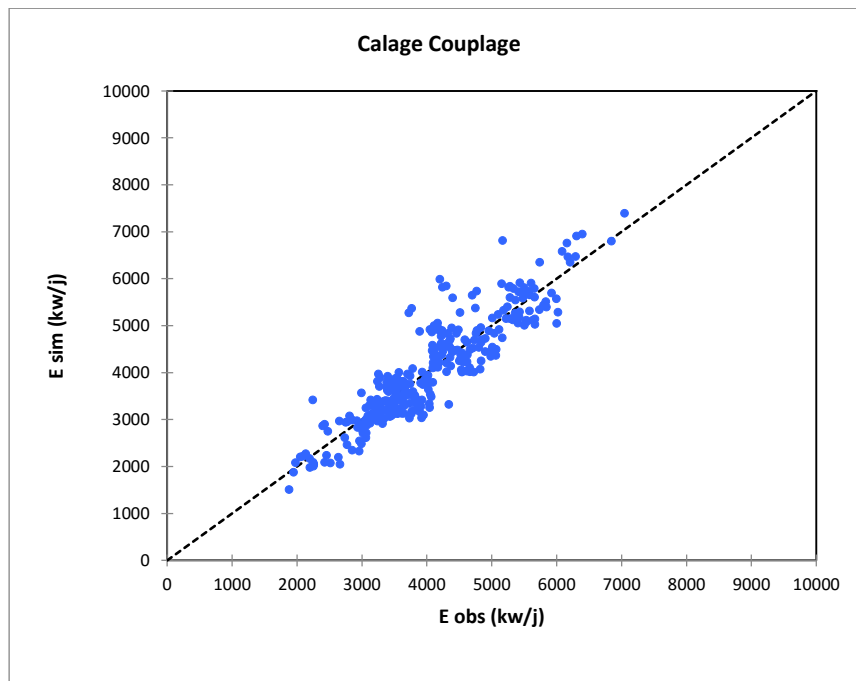


Figure V. 15 : Régression linéaire (Energie brute-Energie simulée par couplage) /calage

D'après la représentation graphique des résultats de la partie calage on voit que les valeurs simulées sont bien regroupées autour de la première bissectrice. Ceci montre que le

modèle couplage ACP Neuro-floue a permis de bien simuler l'énergie et que l'écart entre l'énergie simulée et l'énergie observée est très petit pour la majorité des points de l'échantillon.

D'après la représentation graphique, on peut déjà dire que le ce modèle est de très bonne qualité.

➤ **Représentation graphique des résultats de la partie validation :**

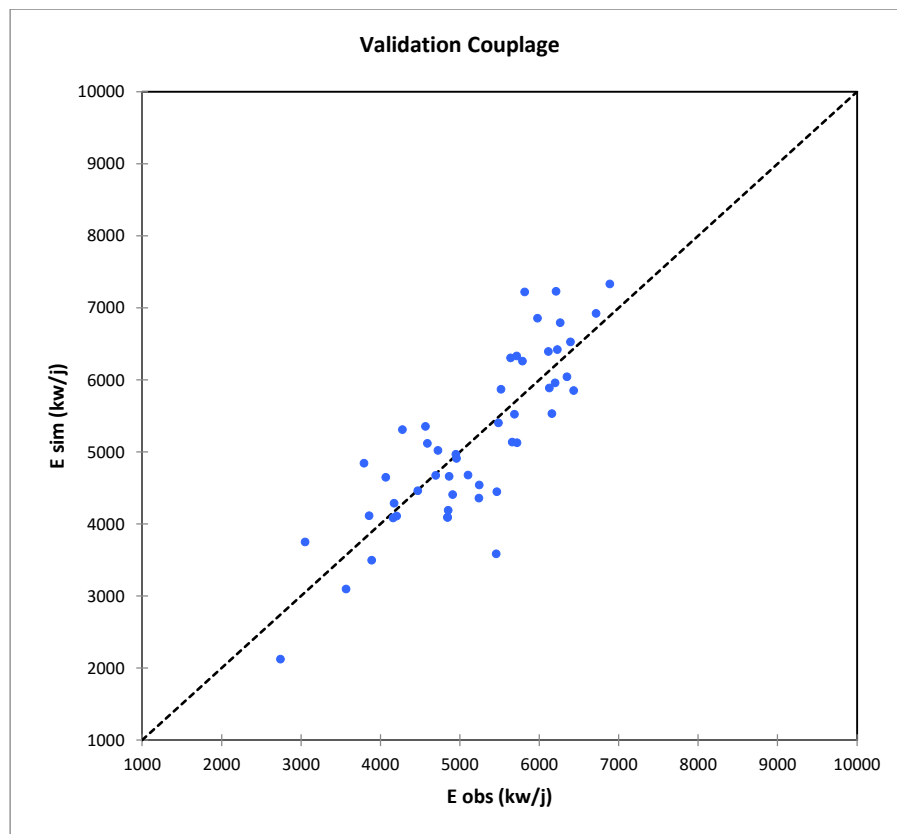


Figure V. 16 : Régression linéaire (Energie brute- Energie simulée par couplage)/ validation

D'après la représentation graphique des résultats de la modélisation par couplage ACP Neuro-floue de la partie validation, on voit que les points sont de plus en plus proches de la première bissectrice par rapport aux deux derniers modèles (ACP et Neuro-floue) et donc on peut dire que le modèle couplage ACP Neuro-floue nous a permis d'améliorer d'avantage les résultats.

Dans le but de confirmer notre jugement sur la qualité du modèle, on s'est proposé de calculer numériquement le coefficient de corrélation (R^2) ainsi que (PBIAS) pour les deux parties : calage et validation.

Les résultats obtenus sont représentés dans le tableau suivant :

Tableau V. 21 : Résultats des critères de validation- Simulation par couplage ACP Neuro-floue (calage/validation)

	Calage	Validation
R ²	82,43%	70%
PBIAS	-1,95	-26,3

➤ *Interprétation des résultats de la partie calage :*

Une fois qu'on a calculé le coefficient de corrélation (R²) on a obtenu une valeur de (82,43%) qui est une très bonne valeur pour le coefficient de corrélation et qui indique que les valeurs simulées par couplage ACP Neuro-floue sont très proches des valeurs observées de l'énergie.

Le PBIAS est toujours négatif et ceci montre que les certaines valeurs sont surestimées par le modèle couplage ACP Neuro-flou.

➤ *Interprétation des résultats de la partie validation*

Le R² de la validation a diminué par rapport au calage (70%) mais il est quand même meilleur par rapport aux deux autres modèles (ACP et Neuro-floue). Cette diminution peut être expliquée par le fait que le vecteur F2 a été calculé à partir de la matrice de passage et donc il est moins précis que celui de la partie calage qui a été obtenu directement de l'application de l'ACP.

Le PBIAS est négatif et donc ce modèle surestime certaines valeurs de la partie validation.

On peut conclure que le modèle couplage ACP Neuro-floue a donné les meilleurs résultats de la simulation de l'énergie.

V.5. Conclusion :

Après analyse des résultats obtenus par les trois modèles soit : ACP, Neuro-floue ou couplage ACP Neuro-floue, on peut se prononcer sur le modèle à adopté pour la modélisation de l'énergie. Le modèle couplage ACP Neuro-floue est celui à retenir car il permet de combiner entre l'avantage de l'ACP qui est la réduction de la dimension d'une matrice de données et l'avantage de la Neuro-floue qui est la grande robustesse dans la prédiction des phénomènes complexes (traitement biologique).

De manière générale le modèle couplage ACP Neuro-floue réussit à améliorer les résultats des deux modèles ACP et Neuro-floue et présente des résultats très satisfaisants.

On a remarqué aussi que pour les trois modèles, certaines valeurs de l'énergie sont surestimées ceci peut être expliqué par la qualité des données car on a des valeurs de l'énergie très grandes dans la partie calage que le modèle ne les rencontre pas dans la partie validation et cela peut influencer sur la qualité de ce dernier.

Conclusion générale

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

Afin d'atteindre ces objectifs, on a fait appel à la modélisation par trois différents modèles soit : modèle issu de l'ACP, modèle Neuro-flou et modèle couplage ACP Neuro-flou.

Avant d'arriver à l'étape de l'élaboration de ces 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 effectué une analyse en composante principale (ACP) sur les différents paramètres de pollution de l'eau brute ainsi que l'eau épurée et cela nous a permis de détecter un excès de nitrite.

L'épuration biologique est un phénomène complexe, et la meilleure méthode pour l'expliquer est de se baser sur des méthodes de création ou de simulation de l'intelligence dite intelligence artificielle tel que : le modèle Neuro-flou ou couplage ACP Neuro-flou.

Dans le but de réduire le nombre de paramètres à tenir compte dans la modélisation, on a éliminé certains paramètres qui peuvent influencer notre énergie déployée et on a juste gardé quatre paramètres pour la suite de notre travail soit : MES, DBO5, DCO et NH4 et notre raisonnement s'est porté sur le rabattement des concentrations de ces quatre paramètres.

Après avoir effectué une modélisation sur l'ensemble de données dont on possède et par les différents modèles cités ci-dessus, on est arrivé à la conclusion que modèle couplage ACP Neuro-flou est le plus performant car il nous a permis d'améliorer les résultats obtenus par les deux autres modèles, ACP et Neuro-flou.

Le modèle couplage ACP Neuro-flou permet de combiner entre l'avantage de l'ACP qui est la réduction de la dimension d'une matrice de données et l'avantage de la Neuro-floue qui est la grande robustesse dans la prédiction des phénomènes complexes (dans notre cas, le traitement biologique).

Pour finir, il est à noter que la station de Boumerdes a commencé à travailler en surcharge à partir de l'année 2013 et on a vu que cela a pu influencer certains de nos résultats car l'énergie a beaucoup augmenté à partir de cette année.

Bibliographique

- [1] Division de l'expertise technique Montréal : Détermination de l'azote total Kjeldahl (NTK) dans l'eau.
- [2] Commission de protection des eaux, du patrimoine, de l'environnement, du sous-sol et des chiroptères : Les différents paramètres physiques et chimiques des eaux et commentaires.
- [3] Lenntech : Métaux lourds.
- [4] Brémond R. et Vuichard R. (1973) Paramètres de la qualité des eaux, Ministère de la protection de la nature et de l'environnement, SPEPE, Paris, 179 p.
- [5] Thèse de doctorat, Mr METAHRI Mohamed Saïd : Elimination simultanée de la pollution azotée et phosphatée des eaux usées traitées, par des procédés mixtes. Cas de la STEP Est de la ville de Tizi-Ouzou.
- [6] WWF SEPANGUY : Eau et quotidien > Les menaces > 2 - Les eaux usées.
- [7] Le Centre d'Information de l'eau CIEAU: Le traitement des eaux usées.
- [8] INRS : Station d'épuration des eaux usées > Prévention des risques biologiques.
- [9] INRS : Le traitement des eaux usées.
- [10] ACTUENVIRONNEMENT.com : Traitement des boues en station d'épuration > déshydratation.
- [11] AgroParisThec : Analyse en Composante Principale. C. Duby, S. Robin.
- [12] Larousse
- [13] Thèse de projet de fin d'étude : Modélisation du transport solide par les modèles empirique et neuro-flou à différents pas de temps. BACHA Fella, 2012.
- [14] Franck Dernoncourt. La logique floue: entre raisonnement humain et intelligence artificielle. Mémoire de FCS2. Janvier 2011.
- [15] S.Zribi Boujelbene et al.,. Système à inférences floues pour la classification phonémique. 4th international conference: Sciences of Electronic, Technologie of Information and Telecommunications. 2007.
- [16] Arnaud MARTIN « L'analyse de données » Polycopié de cours ENSIETA - Réf. : 1463, Septembre 2004.

[17] HAMRICHE, « Contribution à l'étude et à la simulation des paramètres hydrométriques par l'analyse en composantes principales (ACP) », Mémoire de Projet de Fin d'Etude, ENP Alger, 1993.

[18] A. BERMAD, « Elaboration d'une méthodologie de simulation et de prévision des variables climatologiques aléatoires et cycliques par l'analyse en composantes principales ». Thèse de doctorat d'Etat, ENP 2004.

[19] Frank Dernoncourt. La logique floue : entre raisonnement humain et intelligence artificielle. Mmoire de FCS2. Janvier 2011.

[20] Ludovic LE MOAL «L'Analyse en Composantes Principales », 2002.

[22] Victor-Hugo « Grisales Palacio. Modélisation et commande floues de type takagi-sugeno appliquées à un bioprocédé de traitement des eaux usées » (2007) Toulouse.

[23] ATTAB Sarah. « Amélioration de la qualité microbiologique des eaux épurées par boues activées de la station d'épuration haoud berkaoui par l'utilisation d'un filtre à sable local » Ouargla, 2011.

[24] C. Duby, S. Robin « Analyse en Composantes Principales » AgroParisTech.

[25] Faby J.A., Brissaud F. (1997)., L'utilisation des eaux usées épurées en irrigation. Office International de l'Eau, 76 pages.

[26] A. Lefkir et al., Quantification du transport solide par la technique floue, application au barrage de Beni Amrane (Algérie). Revue des Sciences de l'Eau. 19(3) 247-257. 2006.

[27] S.Zribi Boujelbene et al., Système à inférences floues pour la classification phonémique. 4th international conference: Sciences of Electronic, Technologie of Information and Telecommunications. 2007.

[28] Otilia Elena Vasile. Contributon au prognostic de défaillances par réseaneuro-flou: maîtrise de l'erreur de prédiction. 2008.

[29] OMS.

[30] Journal Officiel.

[31] W.W.ECKENFELDER, « Gestion des eaux usées urbaines et industrielles », Techniques et documentation (Lavoisier), Paris 1982.

[32] Gaëlle Deronzier, Sylvie Schétrite, Yvan Racault, Jean-Pierre Canler, Alain Liénard, Alain Héduit, Philippe Duchène, « Traitement de l'azote dans les stations d'épuration biologique des petites collectivités », FNDAE n° 25, Document technique.

[33] René Moletta : l'eau, sa pollution, et son traitement.

[34] « 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.

[35] BELAHMADI Mohamed Seddik Oussama, « 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, 16 Juin 2004.

[36] A.BLANDIN et J.MERESSE, pôle environnement Laurent Burget-Mountain Riders. 41 cours du Dr long 69003 Lyon, Juin 2007.

[37] DESHAYES Matthieu, « 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 », Mémoire de Projet de Fin d'Etudes, Année universitaire 2007/2008 .INSA Strasbourg.

[38] Diagnostic de fonctionnement par analyse en composantes principales : application à une station de traitement des eaux usées. Yvon Tharrault. 29 Feb 2009.