

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



LABORATOIRE MATERIAUX DE GENIE CIVIL
ET ENVIRONNEMENT



Département Hydraulique
Laboratoire de Matériaux de Génie Civil et Environnement

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

**Optimisation de la consommation énergétique
des aérateurs du processus d'épuration par
boues activées par le Gradient Boosting
Cas de la STEP Béni Messous**

Mr. BENNACEUR Sofiane

Sous la direction de Mr. Abdelmalek BERMAD

Professeur

et Mr. Rafik OULEBSIR

Doctorant

Présenté et soutenu publiquement le 24/06/2019 par :

Composition du Jury :

| | | |
|------------|------------------------|--------------------|
| Président | Mr. BENYOUSSEF EL HADI | Professeur ENP |
| Promoteur | Mr. A. BERMAD | Professeur ENP |
| Promoteur | Mr. R. OULEBSIR | Doctorant USTHB |
| Examineurs | Mr. A. LEFKIR | MCA ENSTP |
| Examineurs | Mr. M. HASBAIA | Professeur U.MSILA |
| Examineurs | Mr. H. BOUTAGHANE | MCA U.ANNABA |

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



LABORATOIRE MATERIAUX DE GENIE CIVIL
ET ENVIRONNEMENT



Département Hydraulique
Laboratoire de Matériaux de Génie Civil et Environnement

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

**Optimisation de la consommation énergétique
des aérateurs du processus d'épuration par
boues activées par le Gradient Boosting
Cas de la STEP Béni Messous**

Mr. BENNACEUR Sofiane

Sous la direction de Mr. Abdelmalek BERMAD

Professeur

et Mr. Rafik OULEBSIR

Doctorant

Présenté et soutenu publiquement le 24/06/2019 par :

Composition du Jury :

| | | |
|------------|------------------------|---------------------|
| Président | Mr. BENYOUSSEF EL HADI | Professeur ENP |
| Promoteur | Mr. A. BERMAD | Professeur ENP |
| Promoteur | Mr. R. OULEBSIR | Doctorant USTHB |
| Examineurs | Mr. A. LEFKIR | MCA ENSTP |
| Examineurs | Mr. M. HASBAIA | Professeur U. MSILA |
| Examineurs | Mr. H. BOUTAGHANE | MCA U. ANNABA |

Dédicaces

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

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

A celui qui m'a illuminé mon chemin : Mon Père... Une pensée pour mes mes grandes- mères et grands- pères « Qu'Allah l'accueille en son vaste paradis »

A mes chers frères, Hicham , Adel, Raouf ainsi qu'à ma sœur Ikram , qui ont toujours été présents pour moi.

A toute ma famille pour leur soutien tout au long de ma carrière.

A tous mes amis avec qui j'ai passé des années inoubliables.

A tous ceux et toutes celles qui me connaissent et qui m'aiment.

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

Merci d'être toujours là pour moi

Remerciements

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

En second lieu, je tiens à remercier mes encadreurs Mr BERMAD Abdelmalek et Mr OULEBSIR Rafik ainsi que à Mr A. LAFKIR pour leur disponibilité et leur précieux conseils durant toute la période du travail, pour leurs bonnes explications qui m'ont éclairé le chemin de la recherche et leur collaboration dans l'accomplissement de ce travail.

Mes vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à notre projet en acceptant d'examiner notre travail de projet de fin d'étude. Enfin, je remercie toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail.

ملخص:

يعتبر فهم آلية تكلفة الطاقة لمحطات معالجة مياه الصرف الصحي موضوع ذو أهمية بالغة لمسيرها بسبب ارتفاع تكاليف الطاقة والإمكانيات الكبيرة لها، حالياً يتم تحديد نماذج تكلفة الطاقة باستخدام الدوال اللوغارتمية، الأسية أو الخطية والتي تؤدي إلى نتائج غير دقيقة عندما تكون العلاقة بين المتغيرات غير خطية ومعقدة للغاية. في هذه الأطروحة نقترح منهجية جديدة تسعى إلى تحسين الاستهلاك الطاقوي بطريقة مثالية قائمة على خوارزميات التعلم الذاتي و علم تحليل البيانات. الهدف من هذا العمل هو تطوير نموذج يسعى لتقليل الاستهلاك الطاقوي بشكل أمثل في محطة تصفية مياه الصرف الصحي لبني مسوس، من خلال استعمال طريقة 'تعزيز التدرج' باستعمال طرق إحصائيات تتمثل في تحليل المكونات الأساسية و طريقة معدل التجميع

الكلمات المفتاحية: تعزيز التدرج، التعلم الآلي، تحليل البيانات، تحليل المكونات الرئيسية، طريقة معدل التجميع، الطاقة

Abstract:

Understanding the energy cost structure of wastewater treatment plants is a relevant topic for plant managers due to the high energy costs and significant saving potentials. Currently, energy cost models are generally generated using logarithmic, exponential or linear functions that could produce not accurate results when the relationship between variables is highly complex and non-linear. This thesis proposes a new methodology of optimization of energy consumption based on Machine Learning algorithms and Data- Science tools. The object of our work is the elaboration of a model of optimization of the electric energy of the Béni Messous WWTP by using Gradient Boosting method based on two statistical approaches: "Principal Component Analysis (PCA)" and "The K-Means Method".

Keywords: Gradient Boosting, Principal Component Analysis (PCA), K-Means, WWTP, Machine Learning, Energy, Optimization.

Résumé :

Comprendre le fonctionnement des structures énergivore des stations d'épuration d'eaux usées est un sujet pertinent pour les gestionnaires en raison des coûts élevés de l'énergie et des potentiels d'économie importants. Actuellement, les modèles de coûts énergétiques sont généralement générés à l'aide de fonctions logarithmiques, exponentielles ou linéaires qui pourraient produire des résultats non précis lorsque la relation entre les variables est très complexe et non linéaire. Cette thèse propose une nouvelle méthodologie d'optimisation de la consommation d'énergie basée sur des algorithmes d'apprentissage automatique et des outils de Data-Science. L'objet de notre travail est l'élaboration d'un modèle d'optimisation de l'énergie électrique de la station d'épuration de Béni Messous en utilisant la méthode du "gradient boosting" basée sur deux approches statistiques: "Analyse en composantes principales (ACP)" et "Méthode K-Means".

Mots clé : Gradient Boosting, Analyse en composante principale, K-Means, Optimisation, Apprentissage automatique, Energie, STEP.

Table des matières

Table des matières

Liste des tableaux

Table des figures

| | |
|---|-----------|
| Introduction Générale | 10 |
| 1 Généralités sur le fonctionnement des stations d'épuration | 12 |
| 1.1 Introduction | 13 |
| 1.2 Définition des eaux usées | 13 |
| 1.3 Origines des eaux usées : | 13 |
| 1.3.1 Les eaux usées domestiques | 13 |
| 1.3.2 Les eaux usées industrielles | 13 |
| 1.3.3 Les eaux usées agricoles | 14 |
| 1.3.4 Les eaux usées de ruissèlement | 14 |
| 1.4 Les différentes caractéristiques des eaux usées | 14 |
| 1.4.1 Les paramètres physiques | 14 |
| 1.4.2 Les paramètres chimiques | 15 |
| 1.4.3 Les Paramètres Bactériologiques | 17 |
| 1.5 Le fonctionnement d'une station d'épuration | 17 |
| 1.5.1 Les procédés d'épuration des eaux usées | 18 |
| 1.6 La consommation énergétique des STEP | 20 |
| 1.7 Indicateur de Performance | 21 |
| 1.7.1 Rendement épuratoire | 21 |
| 1.7.2 Abattement de la Pollution | 21 |
| 1.7.3 Water Quality Index | 22 |
| 2 Matériels et Méthodes | 23 |
| 2.1 Introduction | 24 |
| 2.2 Analyse en Composantes Principales (ACP) | 24 |
| 2.2.1 Description de la méthode | 24 |
| 2.2.2 Définition géométrique | 24 |
| 2.2.3 Objectifs de la Méthode | 24 |
| 2.2.4 Application de l'ACP | 25 |
| 2.2.5 Procédé d'application de l'ACP | 25 |
| 2.2.6 Résultats d'application | 25 |
| 2.3 Méthode des K-means | 25 |
| 2.3.1 Définition Clustering | 26 |
| 2.3.2 Définition Kmeans | 26 |
| 2.3.3 Principe de la Méthode | 26 |

| | |
|---|-----------|
| 2.4 Arbres de Décision | 29 |
| 2.4.1 Fonctionnement des Arbres de décision | 29 |
| 2.4.2 Random Forest | 31 |
| 2.4.3 Gradient Boosting | 32 |
| 3 Présentation des STEP de Béni Messous | 33 |
| 3.1 STEP de Béni Messous | 34 |
| 3.1.1 Caractéristiques techniques et données de base du constructeur | 35 |
| 3.2 Principe de traitement | 37 |
| 3.2.1 Principe de traitement | 37 |
| 3.3 Statistiques et critiques des données : | 39 |
| 3.3.1 Statistiques Descriptives | 39 |
| 3.3.2 Evolution des paramètres de pollution dans le temps | 39 |
| 3.3.3 Conclusion | 56 |
| 4 Résultats et Discussion | 57 |
| 4.1 Introduction | 58 |
| 4.2 Application de l'ACP : | 58 |
| 4.2.1 Données utilisées : STEP Beni Messous | 58 |
| 4.2.2 Application de l'ACP sur les paramètres d'entrée (Eau brute) : | 58 |
| 4.2.3 Application de l'ACP sur les paramètres de l'eau (Entré/Sortie) | 59 |
| 4.3 Clustering K means : | 61 |
| 4.3.1 Application de Kmeans sur nos données : | 61 |
| 4.3.2 Conclusion | 73 |
| 4.4 Modélisation avec Gradient Boosting | 74 |
| 4.4.1 Modélisation | 74 |
| 4.4.2 Application du modèle sur les données a optimiser | 79 |
| 4.5 Conclusion | 81 |
| 5 Conclusion Générale | 82 |
| Bibliographie | 83 |

Liste des tableaux

| | | |
|-----|---|----|
| 3.1 | Caractéristiques techniques de la station de Béni Messous | 35 |
| 3.2 | Concentrations admissibles des paramètres de pollution | 35 |
| 3.3 | Tableau récapitulatif des structure de traitement pour les filière eau et filière boues . | 37 |
| 3.4 | Statistique descriptive de paramètres mesurées dans la STEP de Beni Messous . . . | 39 |
| 4.1 | Matrice de corrélation entre les CP' et les variables de l'eau brute journalière | 58 |
| 4.2 | Matrice de corrélation entre les CP' et les variables de l'eau brute et épurée journalière. | 60 |
| 4.3 | Résultats de modèle pour les différents cas étudié | 78 |

Table des figures

| | | |
|------|---|----|
| 1.1 | Schéma représentant les étapes du traitement d'eau usée | 17 |
| 1.2 | Schéma représentant le prétraitement de l'eau usée | 19 |
| 1.3 | Schéma d'une station de traitement par boues activées moyenne charge (Guide d'assainissement) | 20 |
| 1.4 | Besoins en électricité d'une station de traitement par boues activées | 21 |
| 2.1 | Principe de la Méthode de clusterisation par Kmeans | 27 |
| 2.2 | Progression de l'algorithme kmeans pour $k = 3$ | 27 |
| 2.3 | Schéma Descriptif de l'algorithme Kmeans | 28 |
| 2.4 | Principe de classification des Arbres de Décision | 30 |
| 2.5 | Principe de classification des Arbres de Décision | 31 |
| 3.1 | Schéma représentatif de la structure du Chapitre 1 | 34 |
| 3.2 | : Schéma Synoptique de la STEP de Béni Messous | 36 |
| 3.3 | Turbine en fonctionnement | 38 |
| 3.4 | Décanteur-Clarificateur | 38 |
| 3.5 | Evolution journalière du débit entrant dans la STEP | 40 |
| 3.6 | Evolution journalière du débit sortant dans la STEP | 40 |
| 3.7 | Evolution journalière du débit By Pass dans la STEP | 41 |
| 3.8 | Evolution journalière du débit recirculé dans la STEP | 41 |
| 3.9 | Evolution journalière de la concentration en MES à l'entrée de la STEP | 42 |
| 3.10 | Evolution journalière de la concentration en MES à la sortie de la STEP | 42 |
| 3.11 | Evolution journalière de la Demande Biologique en Oxygène à l'entrée de la STEP | 43 |
| 3.12 | Evolution journalière de la Demande Biologique en Oxygène à la sortie de la STEP | 43 |
| 3.13 | Evolution journalière de la Demande Chimique en Oxygène à l'entrée de la STEP | 44 |
| 3.14 | Evolution journalière de la Demande Chimique en Oxygène à la sortie de la STEP | 44 |
| 3.15 | Evolution journalière l'azote global NGL à l'entrée de la STEP | 45 |
| 3.16 | Evolution journalière DE l'azote global NGL à la sortie de la STEP | 45 |
| 3.17 | Evolution journalière de NH_4^+ à l'entrée de la STEP | 46 |
| 3.18 | Evolution journalière de NH_4^+ à la sortie de la STEP | 46 |
| 3.19 | Evolution journalière de NO_3^- à l'entrée de la STEP | 47 |
| 3.20 | Evolution journalière de NO_3^- à la sortie de la STEP | 47 |
| 3.21 | Evolution journalière de Phosphore total à l'entrée de la STEP | 48 |
| 3.22 | Evolution journalière de Phosphore total à la sortie de la STEP | 48 |
| 3.23 | Evolution journalière de l'ortho-phosphate (PO_4^{3-}) I à l'entrée de la STEP | 49 |
| 3.24 | Evolution journalière de l'ortho-phosphate (PO_4^{3-}) à la sortie de la STEP | 49 |
| 3.25 | Evolution journalière de l'ortho-phosphate (PO_4^{3-}) à la sortie de la STEP | 50 |
| 3.26 | Evolution journalière de l'ortho-phosphate (PO_4^{3-}) à la sortie de la STEP | 50 |
| 3.27 | Evolution journalière de la conductivité de l'eau brute | 51 |
| 3.28 | Evolution journalière de la conductivité de l'eau épurée | 51 |
| 3.29 | Variation journalière de la température de l'eau brute | 52 |

| | | |
|------|--|----|
| 3.30 | Variation journalière de la température de l'eau épurée | 52 |
| 3.31 | Evolution journalière de l'énergie dans la STEP | 53 |
| 3.32 | Evolution journalière des sommes des heures (aération) dans la STEP | 53 |
| 3.33 | Evolution journalière de l'énergie dans la STEP | 54 |
| 3.34 | Matrice des graphiques de rendement | 55 |
| 3.35 | Matrice des graphiques de l'abattement | 56 |
| | | |
| 4.1 | Paramètres d'entrée et le temps de fonctionnement des aérateurs à l'échelle journalière [1×2] | 59 |
| 4.2 | Paramètres (Entrée-Sortie) et le temps de fonctionnement des aérateurs à l'échelle journalière [1×2] | 60 |
| 4.3 | Matrice des graphiques du Cas 1 | 62 |
| 4.4 | Représentation en 3D du Cas 1 | 63 |
| 4.5 | application de la méthode du coude et l'analyse silhouette sur le cas 1 | 63 |
| 4.6 | Matrice des graphiques en appliquant l'analyse K means du cas 1 | 64 |
| 4.7 | Matrice des graphiques du Cas 2 | 65 |
| 4.8 | Représentation en 3D du Cas 2 | 66 |
| 4.9 | application de la méthode du coude et l'analyse silhouette sur le cas 2 | 66 |
| 4.10 | Matrice des graphiques en appliquant l'analyse K means du cas 2 | 67 |
| 4.11 | Matrice des graphiques du Cas 3 | 68 |
| 4.12 | Représentation en 3D du Cas 3 | 69 |
| 4.13 | application de la méthode du coude et l'analyse silhouette sur le cas 3 | 69 |
| 4.14 | Matrice des graphiques en appliquant l'analyse K means cas 3 | 70 |
| 4.15 | Matrice des graphiques du Cas 4 | 71 |
| 4.16 | Représentation en 3D du Cas 4 | 72 |
| 4.17 | application de la méthode du coude et l'analyse silhouette sur le cas 4 | 72 |
| 4.18 | Matrice des graphiques en appliquant l'analyse K means cas 4 | 73 |
| 4.19 | Calibration cas 1 | 75 |
| 4.20 | validation cas 1 | 75 |
| 4.21 | Calibration cas 2 | 76 |
| 4.22 | Validation cas 2 | 76 |
| 4.23 | Calibration cas 3 | 77 |
| 4.24 | Validation cas 3 | 77 |
| 4.25 | Calibration cas 4 | 78 |
| 4.26 | Validation cas 4 | 78 |
| 4.27 | Comparaison entre les heures de fonctionnement réel et simulée | 79 |
| 4.28 | Economies d'énergie réalisée pendant sur la période 2008-2016 | 80 |
| 4.29 | Economies d'énergie réalisée par rapport aux valeur réel | 80 |
| 4.30 | Economie d'heure de fonctionnement et heure de fonctionnement réel par an | 81 |

Introduction Générale

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

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

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

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

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

Au cours des vingt dernières années, de nombreux articles sur l'économie d'énergie montrent le manque d'outils de gestion et d'optimisation de la consommation d'énergie dans les STEP. Aux États-Unis, 4% de l'énergie électrique est consommée par les stations d'épurations, alors qu'en Europe elle avoisine les 27 TWh / an. Les coûts énergétiques représentent environ 28% du coût des eaux usées.

L'objet de ce mémoire de fin d'études, est de développer un modèle d'optimisation de la consommation énergétique de la STEP de Béni Messous, en optimisant la quantité d'oxygène soluble, par le contrôle du temps des aérateurs, et ce afin d'aider les gérants de celle-ci, dans leurs prises de décisions optimales de temps d'aération de bassin biologique.

L'originalité de ce travail, réside dans l'introduction de techniques et de méthodes récentes (Data sciences - Machine Learning), et l'application de ces techniques permettra de proposer et de concevoir des solutions sous un nouvel angle autre que celles employées dans la plupart des ouvrages et articles scientifiques traitant du sujet abordée, ce qui permet d'accélérer le processus de modélisation et de mise en application.

Après une introduction générale mettant en relief, la problématique étudiée, les objectifs essentiels de ce travail, et la démarche adoptée pour l'optimisation de la consommation énergétique de la STEP de Béni Messous, par l'utilisation des outils de Data sciences - Machine Learning, pour une meilleure gestion de l'énergie consommée et une protection de la ressource en eau et de l'environnement.

Le présent mémoire est structuré en cinq chapitres de la manière suivante :

Dans le **premier chapitre**, il est question de quelques généralités sur le fonctionnement des stations d'épuration ainsi qu'une synthèse des différents agents et paramètres qui entrent en jeu dans l'analyse de l'eau avant et après son traitement et à la présentation des différentes filières

d'épuration des eaux usées.

Le **chapitre II** est consacré à la présentation de la fouille de données (Data Mining) qui représente les différentes techniques d'analyse de données, et qui comprend les théories de l'Analyse en Composantes Principales (ACP), et du K-Means, ainsi que la technique du Gradient Boosting qui entrent dans les techniques d'arbre de décision.

Dans le **chapitre III**, on retrouve la présentation de la STEP de Béni Messous objet de l'étude, ainsi que celle des bases de données, recueillies auprès de l'exploitant de la STEP à savoir, les paramètres de la pollution physico chimiques, hydrauliques et de consommation d'énergie. Une approche statistique et une attention particulière ont été portées à la critique et à l'interprétation de tous les paramètres recueillis.

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

Le **chapitre IV**, se focalise sur la mise en œuvre des techniques du Data Mining : l'Analyse en Composantes Principales, le K-means, ainsi que l'élaboration et l'application du modèle développé pour optimiser l'énergie consommée de la STEP. La fin du chapitre est dédiée aux résultats obtenus avec les différents modèles, et une confrontation de leurs performances, est donnée par la suite.

Le manuscrit se termine par une conclusion générale dans laquelle, on présente les principaux résultats tirés de ce travail, où il semble, que la technique du Gradient Boosting donne de meilleurs résultats et s'avère être un outil plus performant que les autres modèles mis en œuvre quant à l'optimisation de la consommation énergétique de la STEP.

Chapitre 1

Généralités sur le fonctionnement des stations d'épuration

1.1 Introduction

L'eau est de toutes les matières la plus importante pour l'existence de l'homme. Elle est indispensable pour la survie et pour le développement de la société moderne. Dès que le développement démographique, urbain et industriel atteint certains niveaux, que commenceront à apparaître les premiers problèmes de pollution. L'introduction de l'épuration des eaux usées dans les plans nationaux devient donc nécessaire et même prioritaire.

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

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

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

1.2 Définition des eaux usées

Les eaux usées sont les eaux polluées chargées de substance minérales et organiques, issues de nombreux facteurs : rejets domestiques divers, industriels, agricole, ruissellement, infiltration, ... etc. Constituant donc un effluent pollué et qui sont rejetées dans un émissaire d'égout. Cette pollution peut être quantifiée par des caractéristiques physiques (MES, odeur, ... etc.), chimiques (DCO, DBO5, Azote Kjeldahl, azote minéral, teneur en phosphate, teneur en matières toxiques, ... etc.) et biologiques (bactéries, virus, ... etc.).

La notion d'équivalent habitant EH, qui représente la charge moyenne rejetée par habitant et par jour, a été introduite de manière à caractériser globalement les rejets d'eaux usées.

1.3 Origines des eaux usées :

Les eaux usées proviennent de quatre sources principales : les eaux domestiques , les eaux industrielles, les eaux agricoles et les eaux de ruissellement :

1.3.1 Les eaux usées domestiques

Elles proviennent des différents usages domestiques de l'eau, elles sont généralement véhiculées par le réseau d'assainissement jusqu'à la station d'épuration. Ces eaux se caractérisent par leurs fortes teneurs en matière organique, en sels minéraux (azotes, phosphore), en détergent et en germes fécaux.

Les principales sources de ces eaux sont les eaux de cuisine, les eaux de buanderie, Eaux de vannes.

1.3.2 Les eaux usées industrielles

Tous les rejets résultant d'une utilisation de l'eau autre que domestique sont qualifiés de rejets industriels. Cette définition concerne les rejets des usines, mais aussi les rejets d'activités artisanales ou commerciales : blanchisserie, restaurant, laboratoire d'analyses médicales, etc. Leurs

caractéristiques varient d'une industrie à l'autre. En plus de matières organiques, azotées ou phosphorées .Elles peuvent contenir des produits toxiques, des métaux lourds, des polluants organiques et des hydrocarbures [4] .

1.3.3 Les eaux usées agricoles

Ce sont des eaux qui ont été polluées par des substances utilisées dans le domaine agricole qui contiennent des fertilisants, pesticides et autres produits chimiques, en présentant des risques pour l'environnement et plus particulièrement pour la qualité des eaux. Elles se caractérisent par des fortes concentrations en composés azotés et phosphorés [5] .

1.3.4 Les eaux usées de ruissèlement

Comprennent les eaux de pluies, eaux de lavages et eaux de drainage. Ces eaux sont polluées par les matières qu'elles entraînent en provenance des trottoirs et des chaussées (huiles, mazoutes, graisse, sables... etc.). Elles contiennent également de zinc, plomb, et cuivre. Les eaux de pluie sont caractérisées par un débit fortement variable, présentant des valeurs moyennes à fortes variations saisonnières à l'intérieur desquelles la répartition des débits est aléatoire.

1.4 Les différentes caractéristiques des eaux usées

Les eaux usées sont principalement caractérisées par plusieurs paramètres physiques, chimiques et bactériologiques, il faut définir des paramètres pour avoir des renseignements sur la composition et les caractéristiques qualitatives et quantitatives des eaux usées et de leurs impacts sur le milieu récepteur.

1.4.1 Les paramètres physiques

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

Les principales caractéristiques physiques des eaux usées domestiques sont :

- la couleur : due à la présence des matières organiques dissoutes ou colloïdes par des composés chimiques solubles qui sont colorés
- l'odeur : est due à une fermentation des matières organiques
- La turbidité : représente l'opacité d'un milieu trouble. C'est la réduction de la transparence d'un liquide due à la présence de matière non dissoutes. Elle est causée, dans les eaux, par la présence des matières en suspension (MES) fines, comme les argiles, les grains de silice et les micro-organismes. Une faible part de la turbidité peut être due également à la présence des matières colloïdales d'origine organiques ou minérale [1] .
- La température : La température est un facteur écologique important des milieux aqueux, plus stable que la température de l'air , c'est un paramètre qui a une influence sur la solubilité des sels et des gaz et l'activité microbiennes .Ce qui concerne l'activité de la biomasse nitrifiante .Elle joue un rôle important dans la nitrification et la dénitrification biologique. Par ailleurs, la température détermine le taux et la vitesse des réactions de dégradation biochimique. Plus la température est importante, plus les réactions sont rapides. La température

des eaux usées influent sur l'efficacité du procédé de traitement ,par exemple, la décantation est plus efficace à des températures élevées.

1.4.2 Les paramètres chimiques

Demande chimique en oxygène (DCO)

La demande chimique en oxygène (DCO) est la quantité d'oxygène nécessaire pour oxyder les matières organiques y compris les matières biodégradables et non biodégradables par voie chimique. La DCO inclut donc la DBO et, pour l'essentiel (90 %), est biodégradable ou séparable dans les installations biologiques d'épuration des eaux usées.

La DCO étant fonction des caractéristiques des matières présentes, de leurs proportions respectives, des possibilités de l'oxydation [9].

La DCO est la concentration, exprimée en mg.L-1, d'oxygène équivalente à la quantité de dichromates consommée par les matières dissoutes et en suspension lorsqu'on traite un échantillon d'eau avec cet oxydant dans des conditions définies par la norme. [1] .

Demande biochimique en oxygène (DBO5)

La demande biochimique en oxygène DBO, exprimée en mg d'oxygène par litre, permet l'évaluation des matières organiques biodégradables dans les eaux. Plus précisément, ce paramètre mesure la quantité d'oxygène nécessaire à la destruction des matières organiques grâce aux phénomènes d'oxydation par voie aérobie.

Pratiquement, elle devrait permettre d'apprécier la charge du milieu considéré en substances putrescibles, son pouvoir auto-épurateur et d'en déduire la charge maximale acceptable, principalement au niveau des traitements primaires des stations d'épuration. [9] .

la demande biochimique en oxygène après 5 jours (DBO5) a 20 C d'un échantillon est la quantité d'oxygène consommé par les microorganismes aérobies présents dans cet échantillon pour l'oxydation biochimique des composés organiques et/ou inorganiques[1] .

L'oxygène dissous (OD)

L'oxygène dissous est un composé essentiel de l'eau qui joue un rôle prépondérant dans l'autoépuration des charges polluantes et il conditionne les réactions biologiques qui ont lieu dans les écosystèmes aquatiques. La concentration en oxygène dans l'eau représente le bilan d'activités de production (photosynthèse) et de consommation (respiration).Elle 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 .Elle dépend de différents facteurs, dont la température,la salinite, la pression et la force ionique du milieu.La concentration en oxygène dissous est exprimée en mg O2 l-1 [1] .

Matières en suspension (MES)

Les matières en suspension sont des matières non solubilisées, de nature organique et minérale, les premières sont généralement volatiles, elles peuvent être traitées par dégradation biologique; les dernières de natures minérales essentiellement inertes, peuvent être éliminé en grande partie par sédimentation [6] .

Elles déterminent la turbidité de l'eau , permettent une bonne évaluation du degré de pollution d'une eau. ,diminuent la teneur en oxygène dissous et nuisent au développement de la vie aquatique

Le Phosphore

Le phosphore peut exister dans les eaux en solution ou en suspension, à l'état minéral ou organique. Les composés phosphorés qui, sans hydrolyse ou minéralisation, répondent au test spectrophotométrique sont considérés comme étant des orthophosphates. L'hydrolyse en milieu acide fait apparaître le phosphore hydrolysable et minéralisation, le phosphore organique. Chaque fraction (phosphore en solution ou en suspension) peut être séparé analytiquement en orthophosphates, phosphore hydrolysable et phosphore organique. Suivant les cas, la teneur en phosphates peut être exprimée en mg/L de PO_4 ou de P_2O_5 $1\text{mg/L } PO_4 = 0,747 \text{ mg/L } P_2O_5 = 0,326 \text{ mg/L } P$ [9]

L'azote

L'azote présent dans l'eau peut avoir un caractère organique ou minéral. C'est un nutriment essentiel à la croissance des microorganismes dans le traitement biologique des eaux usées.

L'azote organique est principalement constitué par des composés tels que des protéines, des polypeptides, des acides aminés, de l'urée. Le plus souvent ces produits ne se trouvent qu'à de très faibles concentrations. Quant à l'azote minéral (ammoniacque, nitrate, nitrite), il constitue la majeure partie de l'azote total [9].

il se présente sous quatre formes :

- L'azote organique : c'est une forme de l'azote qu'on ne peut pas mesurer directement, on effectue généralement l'analyse de l'azote total KJELDAHL (NTK) $NK = N_{org} + N - NH_4^+$
- L'azote ammoniacal ($N - NH_4^+$) traduit un processus d'ammonification de la matière organique azotée. Les ions ammoniums subissent une nitrification par action des bactéries nitrifiantes.
- L'azote nitreux ($N - NO_2^-$) provient d'une oxydation incomplète de l'azote ammoniacal ou par une réduction des nitrates par dénitrification. Les nitrites sont instables et sont rapidement transformés en nitrates.
- L'azote nitrique ($N - NO_3^-$) est produit par une nitrification de l'azote ammoniacal. Il joue un rôle important dans le développement des algues et participe au phénomène d'eutrophisation.

Le Potentiel Hydrogène (PH)

L'acidité, la neutralité ou l'alcalinité d'une solution aqueuse peut s'exprimer par la concentration en H_3O^+ (noté H^+ pour simplifier). De manière à faciliter cette expression, on utilise le logarithme décimal de l'inverse de la concentration en ion H^+ : c'est le pH.

Le pH d'une eau naturelle peut varier de 4 à 10 en fonction de la nature acide ou basique des terrains traversés. Des pH faibles (eaux acides) augmentent notamment le risque de présence de métaux sous une forme ionique plus toxique. Des PH élevés augmentent les concentrations d'ammoniac, toxique pour les poissons. On admet généralement qu'un pH naturel situé entre 6,5 et 8,5 caractérise des eaux où la vie se développe de manière optimale.

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 [9].

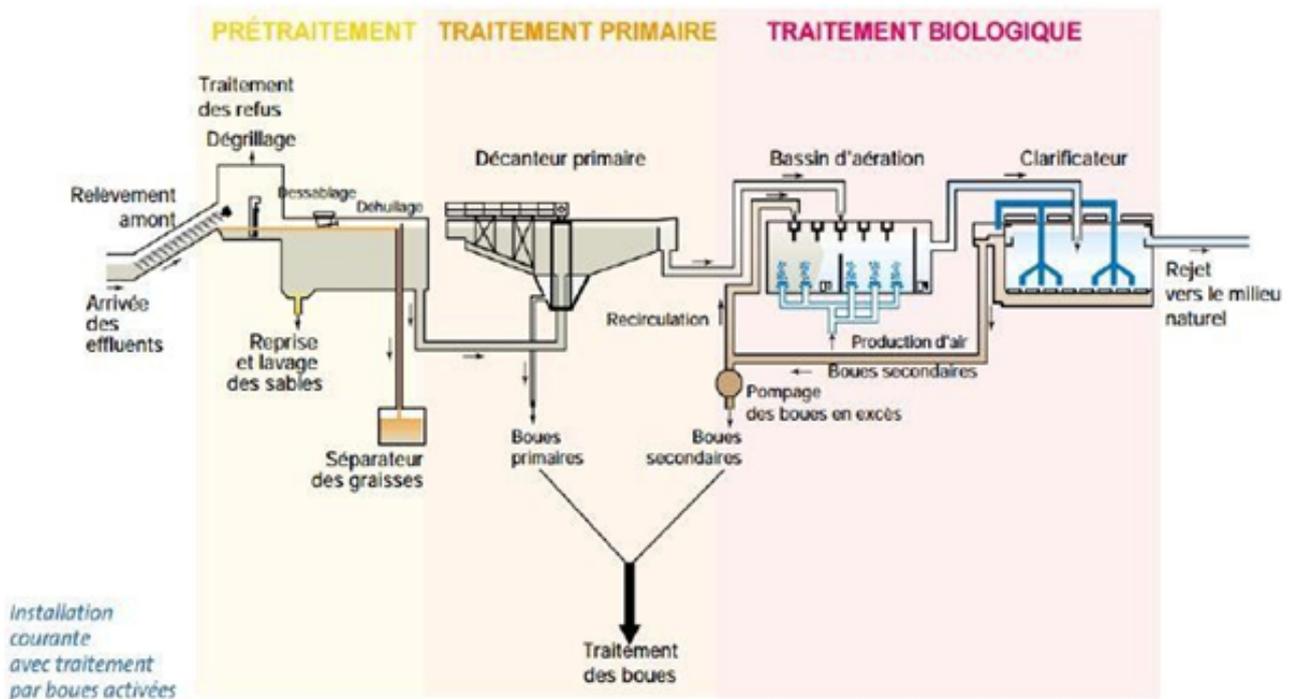


FIGURE 1.1 – Schéma représentant les étapes du traitement d'eau usée

1.4.3 Les Paramètres Bactériologiques

Les eaux usées contiennent différents types de micro-organismes (Bactéries, Protozoaires, Virus, Algues...ect). Les analyses microbiologiques réalisées en laboratoire ont pour but de détecter et évaluer la présence dans les eaux de microbes pathogènes dangereux pour l'homme : eau potable, eaux de baignade, ...

Ces analyses reposent sur la recherche dans les eaux de bactéries indicatrices de leur éventuelle contamination fécale, les *Escherichia coli* (*E. coli*) et les Entérocoques. Ces organismes, d'origine intestinale sont naturellement présents dans les déjections animales ou humaines qui via les déversements, eaux usées et épandages peuvent se retrouver dans l'eau. Les Entérocoques sont pathogènes de même que certains colibacilles. L'eau potable du robinet doit être exempte de la présence de ces bactéries. Une présence très importante de germes fécaux dans une eau indique une pollution fécale importante d'origine humaine ou animale en amont.

1.5 Le fonctionnement d'une station d'épuration

L'épuration consiste à éliminer les plus gros débris organiques ou minéraux, retirer les MES de densité suffisamment différente de l'eau tel que les grains de sables et les particules minérales, comme elle consiste éventuellement à éliminer les pollutions résiduelles qui pourraient être gênantes en aval (germes pathogènes, azote, phosphore...etc.) [7]

L'épuration des eaux usées a pour but de permettre le rejet des eaux sans risque majeur pour le milieu récepteur. 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 à 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 ligne de traitement complète des eaux résiduaires peut être schématiquement scindée en deux filières, La filière eau dans laquelle l'eau est débarrassée de tous les polluants avant son rejet dans le milieu naturel; La filière boue dans laquelle les résidus générés par la filière eau sont traités et déshydratés avant leur évacuation.

1.5.1 Les procédés d'épuration des eaux usées

Selon le degré d'élimination de la pollution et les procédés mise en œuvre, plusieurs niveaux de traitement sont définis les prétraitements le traitement primaire et le traitement secondaire. Dans certain cas des traitements tertiaires sont nécessaires, notamment lorsque l'eau épurée doit être rejetés en milieu particulièrement sensibles [3] Une fois que les eaux usées sont collectées, elles passent par les cinq grandes étapes du traitement qui sont :

Prétraitement

Le prétraitement est un processus physique et mécanique a pour objectif de séparer les matières les plus grossières et les éléments susceptibles de gêner les étapes ultérieures du traitement. Il comprend quatre principaux modes de traitement des eaux usées : le dégrillage, pour retenir les déchets volumineux, le dessablage, pour obtenir une meilleure décantation, le dégraissage et le déshuilage, pour éviter l'encrassement de la station par des corps gras

Dégrillage

Les eaux usées passent par des grilles de plus en plus fines où les matières volumineuses sont retenues qui pourraient perturber la suite du traitement. Les dégrilleurs sont équipés d'un dispositif de relevage et d'éjection automatique des déchets [6] .

Dessablage

Après le dégrillage, il reste encore dans l'eau des fragments qui peuvent décanter facilement, mais dont la dureté et la taille sont relativement importantes, supérieure à 0,2 mm de diamètre, et qui pourraient conduire à l'abrasion de certains éléments de la station et particulièrement les pompes, on élimine ces matériaux facilement décantables dans de petits bassins rectangulaires ou circulaires [16]

Dégraissage-Déshuilage

C'est le principe de la flottation qui est utilisé pour l'élimination des huiles, il est basé sur l'injection de fines bulles d'air dans le bassin de déshuilage, permettant de faire remonter rapidement les graisses en surface. Leur élimination se fait ensuite par raclage de la surface.

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

Traitement primaire

Il s'agit le plus souvent d'une décantation qui permet d'éliminer 50% à 60% MES décantables dans l'eau. L'utilisation de réactifs chimiques pour éliminer des particules plus fines constitue un traitement physico-chimique. Ce traitement permet donc essentiellement l'élimination de la pollution particulaire et d'une partie de la pollution organique sous forme particulaire (de l'ordre de 65 à 80% de la DCO avec un traitement physico-chimique) [9] .

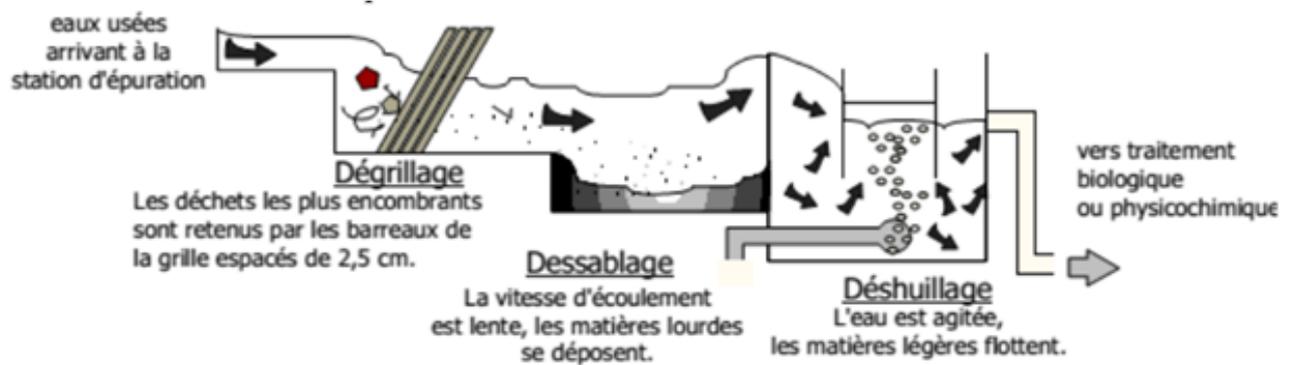


FIGURE 1.2 – Schéma représentant le prétraitement de l'eau usée

Traitement secondaire (biologique)

Le traitement secondaire est un traitement purement biologique des eaux usées et a pour objet de réduire la teneur en matière organique présentes dans ces eaux et leur dégradation biologique par les micro-organismes. Parmi les divers micro-organismes responsables de la dégradation on trouve les bactéries aérobies et les bactéries anaérobies. Cependant la vitesse de la dégradation des matières organiques est plus élevée au milieu aérobie, pour ce fait, les installations d'épuration biologique fonctionnent généralement en présence d'oxygène.

On distingue deux procédés pour le traitement biologique, Les procédés à cultures fixées (lits bactériens et disques biologiques) et Les procédés à culture libres (lagunage, boues activées).

Dans notre cas, on est intéressé par le procédé d'épuration par boues activées. En Algérie, ce procédé est le plus retenu aujourd'hui dans les stations d'épuration assurant un traitement secondaire. Les boues activées sont utilisées comme épuration biologique dans le traitement des eaux usées.

Les boues activées

Les boues activées constituent le traitement biologique aérobie le plus répandu. Le procédé consiste à provoquer le développement d'une culture bactérienne dispersée sous forme de flocons (boues activées), dans un bassin brassé et aéré (bassin d'aération) et alimenté en eau à épurer. Dans ce bassin, le brassage a pour but d'éviter les dépôts et d'homogénéiser le mélange des flocons bactériens et de l'eau usée (liqueur mixte); l'aération peut se faire à partir de l'oxygène de l'eau, du gaz enrichi en oxygène, voire même d'oxygène pur, a pour but de dissoudre ce gaz dans la liqueur mixte, afin de répondre aux besoins des bactéries épuratrices aérobies [26].

Après un temps de contact suffisant, la liqueur mixte est envoyée dans un clarificateur appelé parfois décanteur secondaire, destiné à séparer l'eau épurée des boues. Ces dernières sont recyclées dans le bassin d'aération pour y maintenir une concentration suffisante en bactéries épuratrices [23].

Traitement tertiaire

Appelées aussi les traitements complémentaires qui visent l'élimination de la pollution d'azotée et phosphatée ainsi que la pollution biologique des eaux usées domestiques, ayant déjà subi au préalable des traitements primaires et secondaires qui s'avèrent insuffisants pour arriver bout de ces polluants. Pour cela les traitements tertiaires d'imposent et deviennent plus que nécessaires, afin de garantir une meilleure protection des milieux naturels récepteurs.

Les traitements tertiaires souvent considérés comme facultatif ou complémentaire permettent d'affiner ou d'améliorer le traitement secondaire. De telles opérations sont nécessaires pour as-

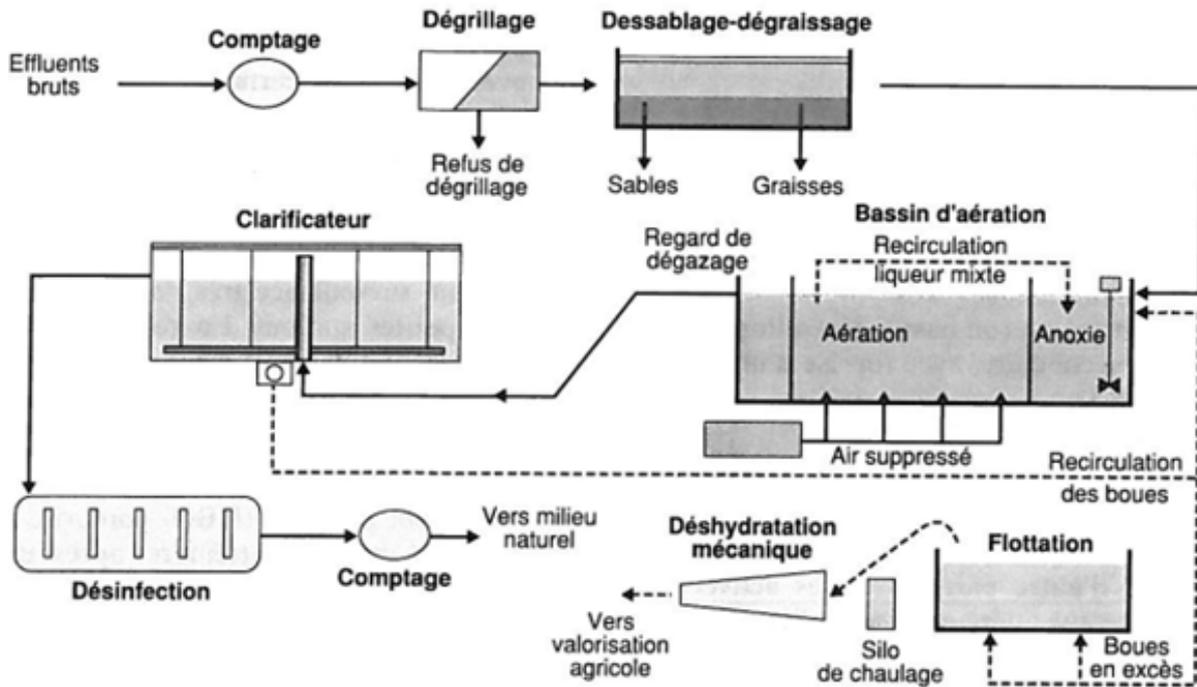


FIGURE 1.3 – Schéma d’une station de traitement par boues activées moyenne charge (Guide d’assainissement)

assurer une protection complémentaire de l’environnement récepteur ou une réutilisation de l’effluent en agriculture ou en industrie. Les traitements tertiaires visent à améliorer la qualité générale de l’eau.

Leur utilisation d’impose lorsque la nature des milieux récepteurs recevant l’eau dépolluée l’exige. On y distingue généralement les opérations suivantes :

- La nitrification-dénitrification et déphosphatation biologique ou mixte (biologique et physicochimique)
- La désinfection bactériologique et virologique : elle se fait par différentes méthodes soit par les rayonnement ultra-violet, la filtration, la chloration, l’ozonation ou les procédés chimiques.

1.6 La consommation énergétique des STEP

Les stations d’épuration sont des installations d’une grande complexité dans lesquelles la consommation d’énergie varie considérablement selon les procédés utilisés pour le traitement des eaux usées et des boues.

Dans le processus de boues activées, une grande quantité d’air est injectée dans un système d’aération bassin pour fournir de l’oxygène aux microorganismes en croissance et pour maintenir les solides en suspension . L’air est généralement fourni par des ventilateurs à volume élevé ou des compresseurs à basse pression à travers un système d’aération diffuse.

Au cours des vingt dernières années, de nombreux articles sur la consommation énergétique montrent le manque d’outils de gestion et d’optimisation de la consommation d’énergie dans Les STEP .Aux États-Unis, 4% de l’énergie électrique est consommée par les stations d’épurations [18] ,et en Europe les stations d’épuration consomment environ 27 TWh / an [29]. Les coûts énergétiques représentent environ 28% du coût des eaux usées.

Le processus d’aération dans le traitement secondaire est l’opération la plus énergivore. Dans la plupart des grandes STEP avec boues activées , l’aération prend environ 50 à 70% de la consom-

mation totale d'électricité, tandis que le traitement des boues consomme de 15 à 25% d'énergie, suivie d'une sédimentation secondaire comprenant des pompes de recirculation (15%).

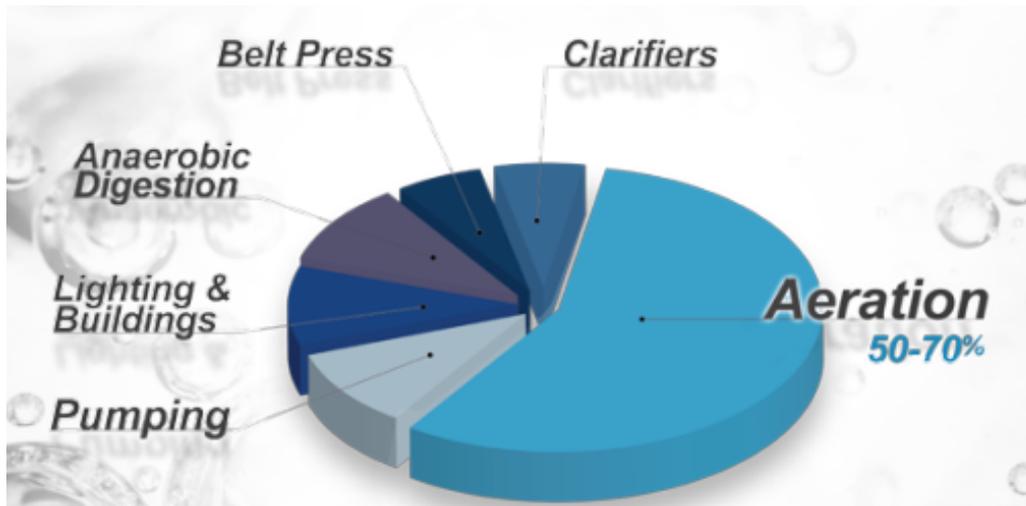


FIGURE 1.4 – Besoins en électricité d'une station de traitement par boues activées

1.7 Indicateur de Performance

Afin de juger les performances de la STEP a traiter les eaux usées plusieurs indicateur de performance (ou Key Performance Indicator (KPI)) ont été développée. Ceux ci pouvant donner une idée sur la performance de la STEP dans différents axes (économique, performance épuratoire, énergétique, ressources humaines ... etc). Dans le cadre de ce travail nous nous sommes limitée au KPI indiquant les performances épuratoires et aux KPI indicateur du niveau de pollution entrante a la STEP.

1.7.1 Rendement épuratoire

Le rendement épuratoire représente la performance de la STEP concernant le traitement d'un paramètres de pollution :

$$R_i = 1 - \frac{X_{is}}{X_{ie}}$$

- X_{ie} : Concentration entrante a la STEP du paramètre i (mg/l)
- X_{is} : Concentration sortante de la STEP du paramètre i (mg/l)
- T_i : Rendement épuratoire du paramètre i (%).

1.7.2 Abatement de la Pollution

l'abattement de la pollution représente la performance de la STEP a traiter une charge de pollution entrante :

$$AB_i = \frac{(X_{ie} - X_{is}) * Q_s}{\max(X_{ie}) * Q_{xie}}$$

- Q_s : Débit Sortant de la STEP (m^3)
- Q_{xie} : Débit correspondant a la valeur max de X_{ie} (m^3).
- PAB_i : Abbattement de la pollution du paramètres i (%).

1.7.3 Water Quality Index

Developpée par Brown et al [14] [10] , le Water quality index (WQI) est un indicateur du niveau de pollution d'une eau intégrant plusieurs paramètres mesurées a la fois, il a été developpée pour l'appréciation du niveau de pollution des eaux de surface. l'avantage de cette indicateur est qu'il intègre des poids qui sont assigné a chacuns des paramètres utilisée, et donc permet d'accordée plus d'importance a des paramètres jugée plus pertinent.

$$I_i = 1 - \frac{(X_{ie} - N_{is})}{(\max(X_{ie}) - N_{ie})}$$

$$WQI = \prod_{i=1}^N I_i^{w_i}$$

ou :

$$\sum_{i=1}^N w_i = 1$$

- I_i : Sous-Indices du Water Quality Index.
- WQI : Water Quality Index.
- w_i : Poids assigné a chaque sous indices.

Chapitre 2

Materiels et Méthodes

2.1 Introduction

La fouille de données (data mining) est une démarche qui a pour finalité la découverte de faits, de relations, de règles et de corrélations à la fois significatifs, insoupçonnés et pertinents, à partir de grands ensembles de données ou d'observations [Fayyad et al, 1996]. Forage de données, explorations de données ou fouilles de données, ce sont les traductions possibles du data mining en Français.

Le terme Data Mining désigne l'analyse de données depuis différentes perspectives et le fait de transformer ces données en informations utiles, en établissant des relations entre les données ou en repérant des caractéristiques (ou patterns), En employant des techniques, basées sur des concepts statistiques et algorithmiques qui permettent de traiter des données brutes de la STEP et de retourner de la connaissance nécessaire dans le but de délivrer la bonne information qui visent à l'optimisation de l'énergie dans la STEP étudiée .

Dans ce chapitre, on va présenter quelques-unes des principales méthodes de l'analyse de données : l'Analyse en Composantes Principales (ACP), la classification non supervises par K means ainsi que les forets aléatoires (ou Random forest)

2.2 Analyse en Composantes Principales (ACP)

2.2.1 Description de la méthode

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

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

L'ACP servira à mieux connaître les données sur lesquelles on travaille, à détecter éventuellement des valeurs suspectes, et aidera à formuler des hypothèses qu'il faudra étudier à l'aide de modèles et d'études statistiques inférentielle [2] .

2.2.2 Définition géométrique

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

Un second axe est obtenu après projection des N points sur un hyperplan orthogonal au premier axe, tel que la dispersion des projections des N points sur celui-ci soit toujours maximale, et le processus se réitère P fois [12].

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

2.2.3 Objectifs de la Méthode

L'Analyse en Composantes Principales a pour objectifs :

- Avoir une corrélation entre les variables grâce à une représentation des données brutes et fournir des outils simples et lisibles.

- Essayer d'examiner les liens existants entre les variables afin de faciliter leur interprétation.
- Essayer de réduire au maximum le nombre de dimensions étudiées tout en gardant l'ensemble original et en utilisant les relations détectées entre les variables.

2.2.4 Application de l'ACP

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

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

2.2.5 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).
2. Calcul des paramètres statistiques :(moyenne, écart type).
3. Calcul des variables centrées réduites.
4. Calcul des coefficients de corrélation.
5. Calcul des valeurs propres et des vecteurs propres de la matrice de corrélation obtenue.
6. Calcul des composantes principales, avec :
 $CP =$ matrice des variables centrées réduites (NxP) x matrice des vecteurs propres (NxN).
7. Calcul des composantes principales centrées réduites : Avec VP_j : la valeur propre correspondant à l'individu j.
8. Calcul de la matrice de corrélation entre CP' et XCR ($cor(CP', XCR)$).
9. Tracé des cercles de corrélation

2.2.6 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.

2.3 Méthode des K-means

La classification est une tâche appliquée dans la vie courante. Elle est utilisée pour expliquer les nouveaux phénomènes rencontrés en les comparant avec des concepts et des phénomènes connus et en essayant de rapprocher les caractéristiques le plus possible. C'est un sujet de recherche actif qui se place au cœur de data mining, ce qui justifie pleinement l'intérêt qui lui est porté.

Les techniques de classification sont réparties essentiellement en deux approches : la classification supervisée et la classification non supervisée (appelée aussi “clustering”).

L’objectif le plus simple de clustering est de répartir une population d’objets en groupes d’observations homogènes (classes), chaque classe étant bien différenciée des autres, pour former une partition.

Le clustering est un sujet abordé par plusieurs métiers. Ce fait lui donne l’avantage d’être utilisé par divers domaines d’application qui peuvent être catalogués suivant le but recherché selon trois grands types [22] :

- Extraction des connaissances : Dans ce type d’applications, on cherche généralement à donner une description sémantique aux groupes constitués afin de donner un sens à l’information dont on dispose.
- Réduction des bases de données : L’objectif visé est de réduire l’espace des données sur lequel on travaille en espérant que le problème qu’on cherche à résoudre devient beaucoup plus simple.
- Etude de comportement (profiling) : Le but est de détecter des sous populations qui partagent des caractéristiques proches, ce qui leur donne un certain comportement commun. Les systèmes de décision se basent souvent sur l’étude de comportement afin d’entreprendre des actions convenables à chacune des sous population.

2.3.1 Définition Clustering

Le clustering est une approche de classification dont les classes existent a posteriori. Au départ, on dispose d’un ensemble d’objets non étiquetés (dont la classe est inconnue). A partir de ces objets, l’idée est de parvenir à détecter des objets “similaires” afin de les regrouper dans des classes [22] .

Selon [30], le clustering consiste à diviser une population d’objets en sous-ensembles d’objets appelés clusters pour que tous les objets dans une même classe soient similaires et les objets de clusters distinctes soient dissimilaires.

Parmi les objectifs que le clustering doit permettre d’atteindre, citons la réduction des données et la prédiction basée sur les groupes.

2.3.2 Définition Kmeans

La méthode k-means est un algorithme de clustering Les algorithmes de clustering regroupent les données en fonction de leur disposition dans l’espace. Ils ne sont pas supervisés dans le sens qu’ils n’exigent aucune sorte de données indiquant à quelle groupe ou class ils appartienne. C’est le travail de l’algorithme de déduire les dites classes en se basant uniquement sur la géométrie des données.

2.3.3 Principe de la Méthode

L’algorithme de clustering k-means fonctionne comme suit. Tout d’abord, l’utilisateur doit choisir k , le nombre de classes (ou de groupes) .Il existe plusieurs méthodes pour connaître le nombre de classes nécessaires comme la méthode de coude (elbow method) ou (silhouette method) qui seront décrit plus loin.

Ensuite, nous plaçons au hasard k vecteurs d’entités, appelés centroïdes, dans l’espace. Nous calculons ensuite la distance entre chaque exemple x et chaque centroïde « c » en utilisant des fonctions de mesure des distances, la plus utilisées étant la distance euclidienne. Ensuite, nous assignerons chaque point x au centroïde le plus proche Pour chaque centroïde, nous calculons la

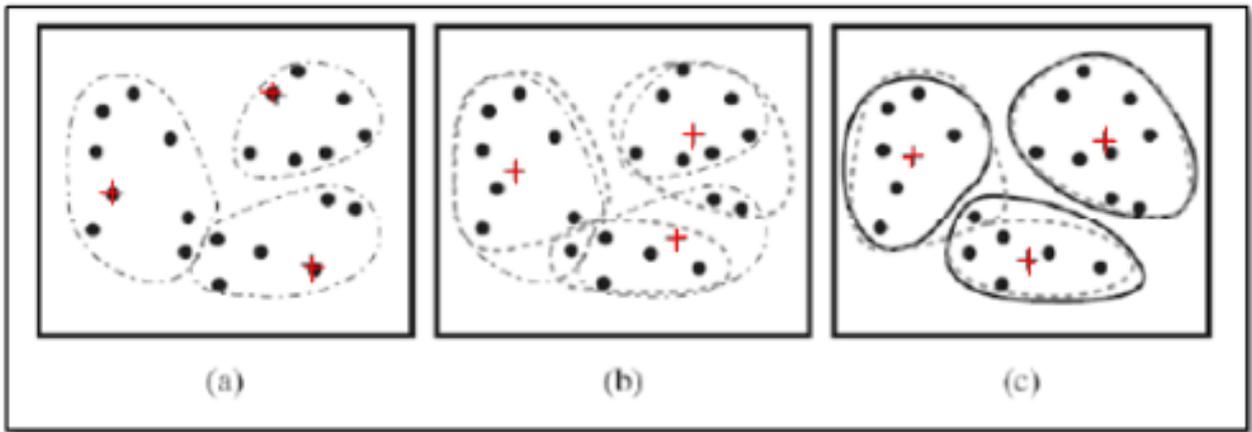


FIGURE 2.1 – Principe de la Méthode de clusterisation par Kmeans

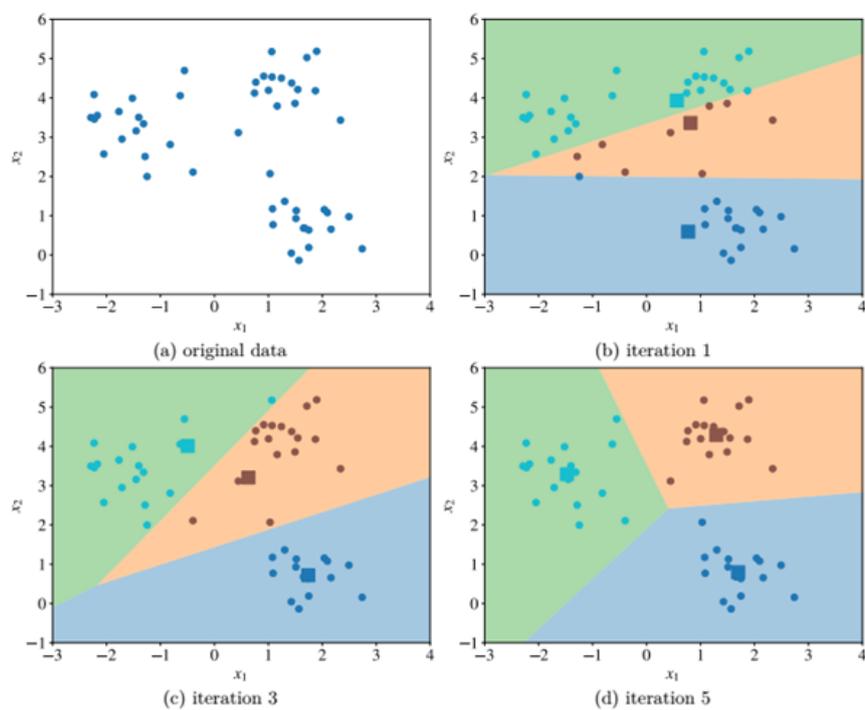


FIGURE 2.2 – Progression de l’algorithme kmeans pour $k = 3$

moyenne des positions des exemples assignés à celui-ci, qui sera la nouvelle valeur de la position du dit centroid.

Cette procédure sera réitéré jusqu’au ce que la position des centroides ne change pas. La position initiale des centroides influence les positions finales. Différentes positions initiales des centroides peuvent aboutir différents modèles.

Les différentes couleurs d’arrière-plan de la figure ci-dessous représentent les régions dans lesquelles tous les points appartiennent au même groupe. La valeur du nombre de clusters, est un hyperparamètre qui doit être choisis par l’utilisateur, quoique Certaines techniques existent pour la détermination du nombre de clusters adéquat.

Deux techniques sont majoritairement utilisées pour la détermination du nombre de clusters :

- Elbow method (ou technique du coude) : étant la méthode la plus utilisée, elle consiste à calculer une fonction représentant la densité de chaque clusters, et à calculer sa valeur pour plusieurs nombre de clusters croissant et ainsi déterminer le nombre de cluster adéquat en choisissant la valeur de k ou on observe un point d’infelxion de la courbe de la fonction

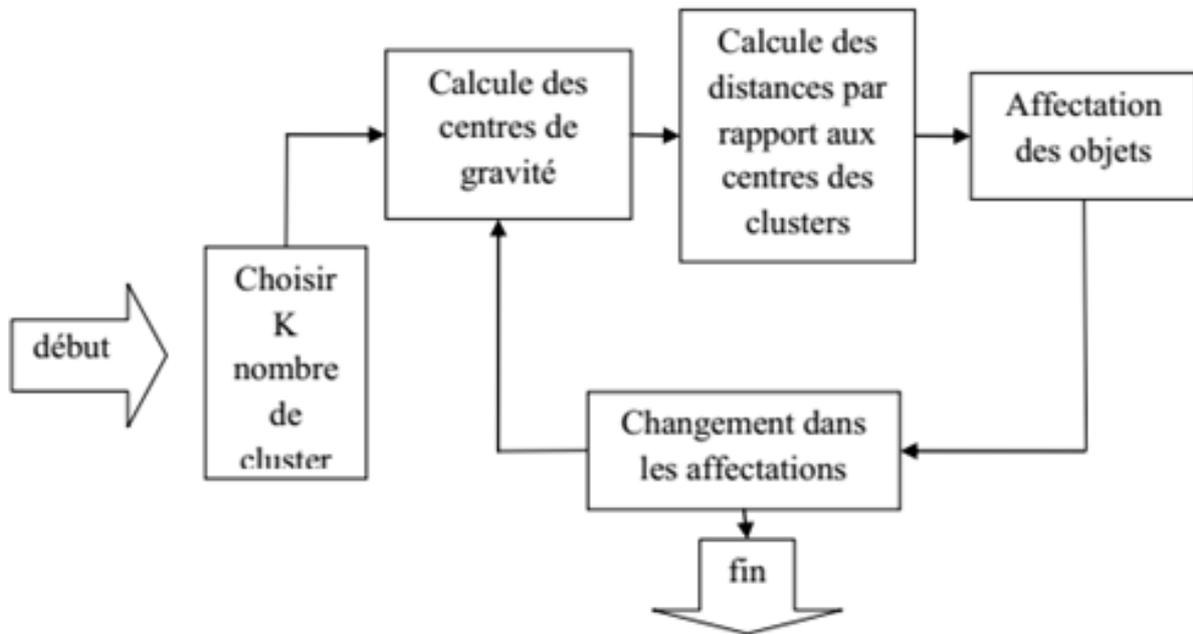


FIGURE 2.3 – Schéma Descriptif de l'algorithme Kmeans

densité utilisées en fonction du nombre de clusters.

- Silhouette Analysis : cette technique est utilisée pour étudier la distance de séparation entre les clusters. Le tracé de la silhouette affiche une mesure de la proximité de chaque point d'un cluster par rapport aux clusters voisins et permet ainsi d'évaluer les paramètres tels que le nombre de clusters de manière visuelle. Cette mesure varie entre $[-1, 1]$. Les coefficients de la silhouette ayant une valeur près de $+1$ indiquent que l'échantillon est loin des clusters voisins. Une valeur de 0 indique que l'échantillon est sur ou très proche de la limite de décision entre deux clusters voisins et des valeurs négatives indiquent que ces échantillons ont peut-être été affectés au mauvais cluster.

En plus des deux techniques décrite précédemment, un autre test permet de juger de la capacité de la techniques a clusterisé un échantillon. Le test statistique Hopkins permet de mesurer la tendance des Clusters d'un ensemble de données. Il appartient à la famille des tests d'échantillonnage clairsemé. Il s'agit d'un test d'hypothèse statistique dans lequel l'hypothèse nulle est que les données sont générées par un processus ponctuel et sont donc uniformément distribuées de manière aléatoire. Une valeur proche de 1 tend à indiquer que les données sont fortement groupées, les données aléatoires ont tendance à donner des valeurs voisines de $0,5$ et sont uniformément distribuées.

Avantages de K-means

Nous pouvons citer quelques avantages de K-means :

- L'avantage de ces algorithmes est avant tout leur grande simplicité.
- Tend à réduire l'erreur quadratique.
- Il s'adapte bien pour des populations de tailles importantes.

Inconvénients de K-means

- Le nombre de classes k , doit être spécifié à priori.

- Il est sensible aux objets isolés (bruits).
- Le résultat dépend de tirage initial des centres des classes.
- Les clusters sont construits par rapports à des objets inexistants (les centres).
- N'est pas applicable en présence d'attributs qui ne sont pas du type intervalle.
- Il n'est pas adapté pour découvrir des classes avec des structures non-convexes, et des classes de tailles et formes différentes.

2.4 Arbres de Décision

Un arbre de décision est l'une des techniques les plus simples et intuitives dans l'apprentissage automatique, basé sur le paradigme diviser pour régner. L'idée de base derrière les arbres de décision est de diviser l'espace en parcelles et d'adapter un modèle à une parcelle. Les arbres de décision ont un nombre de propriétés qui font d'eux un outil précieux, surtout quand il s'agit de faire l'analyse rapide d'un jeu de données ou d'élaborer un prototype de classifieur [15] :

- Modèle white box : le résultat est facile à conceptualiser, à visualiser et à interpréter.
- Ils nécessitent peu de préparation de données (e.g. normalisation, etc.).
- Le coût d'utilisation des arbres est logarithmique (classification d'une nouvelle donnée très rapide).
- Ils sont capables d'utiliser des données catégorielles et continues.
- Ils sont capables de gérer des problèmes multi-classes.
- Ils ont un bon comportement par rapport aux valeurs extrêmes (outliers).
- Ils gèrent bien les données manquantes.

2.4.1 Fonctionnement des Arbres de décision

On considère d'abord le problème de classement. Chaque élément x de la base de données est représenté par un vecteur multidimensionnel (x_1, x_2, \dots, x_n) correspondant à l'ensemble de variables descriptives du point [15]. Chaque nœud interne de l'arbre correspond à un test fait sur une des variables x_i [15] :

- Variable catégorielle : génère une branche (un descendant) par valeur de l'attribut.
- Variable numérique : test par intervalles (tranches) de valeurs.

Les feuilles de l'arbre spécifient les classes.

Une fois l'arbre construit, classer un nouvel candidat se fait par une descente dans l'arbre, de la racine vers une des feuilles (qui encode la décision ou la classe) [15]. A chaque niveau de la descente on passe un nœud intermédiaire où une variable x_i est testée pour décider du chemin (ou sous-arbre) à choisir pour continuer la descente [15].

Au départ, les points de base de l'apprentissage sont tous placés dans le nœud racine. cette variable est dite « variable cible ». La variable cible peut être catégorielle (problème de classement) ou valeur réelle (problème de régression). Chaque nœud est coupé (opération split) donnant naissance à plusieurs nœuds descendants. Un élément de la base d'apprentissage situé dans un nœud se retrouvera dans un seul de ses descendants.

L'arbre est construit par partition récursive de chaque nœud en fonction de la valeur de l'attribut testé à chaque itération (top-down induction). Le critère optimisé est l'homogénéité des descendants par rapport à la variable cible. La variable qui est testée dans un nœud sera celle qui maximise cette homogénéité [15].

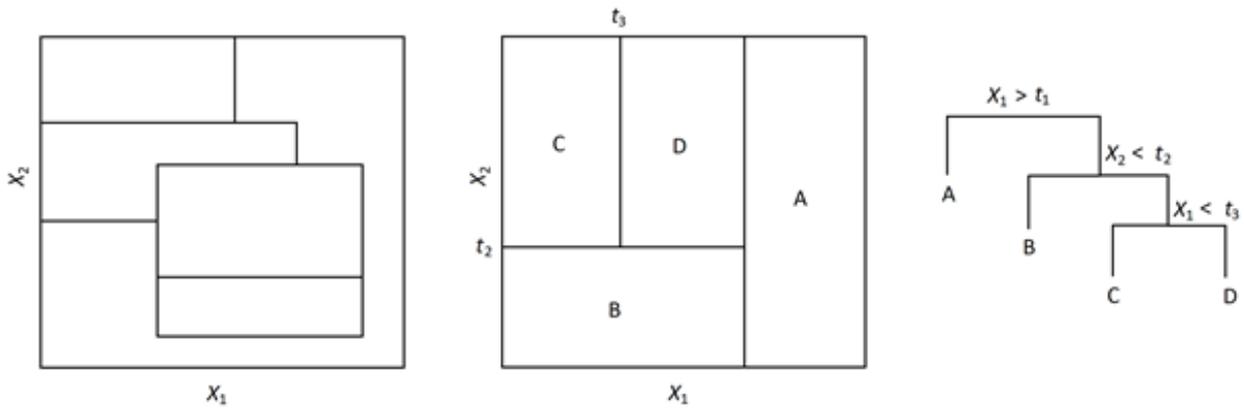


FIGURE 2.4 – Principe de classification des Arbres de Décision

Le processus s'arrête quand les éléments d'un nœud ont la même valeur pour la variable cible (homogénéité).

Il existe une implémentation alternative des arbres de décision, proposée par Breiman et al appelée CART, Les principales différences sont les suivantes [15] :

- CART pose seulement de questions-test binaires (arbres binaires).
- Fonctionne aussi pour des attributs aux valeurs continues.
- CART cherche tous les attributs et tous les seuils pour trouver celui qui donne la meilleure homogénéité du découpage.

Quand un nœud interne S est coupé sur l'attribut j , seuil a_j , il donne naissance à deux descendants :

- Sous-nœud gauche S_g ($p_g \approx |S_g|/|S|$), qui contient tous les éléments qui ont les valeurs de l'attribut $v_j \leq a_j$
- Sous-nœud droit S_d ($p_d \approx |S_d|/|S|$), qui contient tous les éléments qui ont les valeurs de l'attribut $v_j > a_j$

Soit $I(S)$ une fonction qui mesure l'impureté de S par rapport à la classe cible. CART étudie le changement de l'impureté par rapport au seuil et pour tous les attributs [15] :

- $E[I(S_g d)] = p_g I(S_g) + p_d I(S_d)$ ou $E[\cdot]$ est l'opérateur de moyenne statistique
- $\Delta I(S) = I(S) - E[I(S_g d)] = I(S) - p_g I(S_g) - p_d I(S_d)$

Le problème d'optimisation est le suivant :

$$\operatorname{argmax}_{j; a_j} \Delta I(S)$$

CART choisit donc l'attribut et le seuil qui maximisent la décroissance de l'impureté du nœud par rapport à la cible[15]. Pour un problème de régression on optimise le résidu quadratique moyen : minimise la variance moyenne des groupes.

Bien que les arbres de décision sont des outils assez robustes ils comportent tout de même des défauts :

- Parfois les arbres générés ne sont pas équilibrés (ce qui implique que le temps de parcours n'est plus logarithmique). Il est donc recommandé d'équilibrer la base de données avant la construction, pour éviter qu'une classe domine (en termes de nombre d'exemples d'apprentissage).

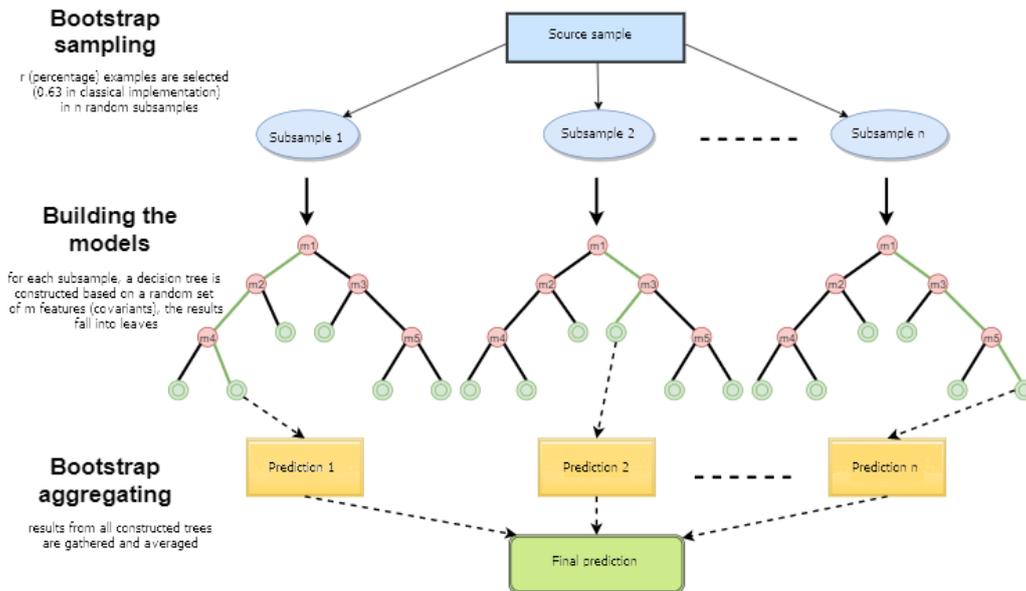


FIGURE 2.5 – Principe de classification des Arbres de Décision

- Sur-apprentissage : parfois les arbres générés sont trop complexes et généralisent mal (solution : élagage, le contrôle de la profondeur de l'arbre et de la taille des feuilles).
- Ils sont instables : des changements légers dans les données produisent des arbres très différents. Les changements des nœuds proches de la racine affectent beaucoup l'arbre résultant. On dit que les arbres produisent des estimateurs de variance élevée.

Le besoin de répondre à ce troisième problème, qui n'admet pas de solution par optimisation algorithmique, a conduit aux approches de type Bagging et « Forêts aléatoires » [15].

2.4.2 Random Forest

Le Bagging consiste à créer de nombreuses «copies» des données d'apprentissage (chaque copie est légèrement différente d'une autre), puis à appliquer un modèle à chaque copie pour obtenir plusieurs modèles dit « faibles », puis à les combiner [11]. L'approche dite « bagging » est derrière l'algorithme d'apprentissage de forêt aléatoire.

L'algorithme « vanilla bagging » fonctionne comme suit [11]. A partir d'un ensemble d'apprentissage, nous créons B échantillons aléatoires S_b (pour chaque $b = 1, \dots, B$) de l'ensemble d'apprentissage et construisons un modèle d'arbre de décision f en utilisant chaque échantillon S_b [11]. Pour échantillonner S pour certains b , nous faisons l'échantillonnage avec remplacement. Cela signifie que nous commençons avec un ensemble vide, puis prenons au hasard un exemple dans l'ensemble d'apprentissage et plaçons sa copie exacte sur S_b en conservant l'exemple d'origine dans l'ensemble d'apprentissage initial [11]. Nous continuons à choisir des exemples au hasard jusqu'à ce que $|S_b| = N$ [11].

Après la période d'apprentissage, nous avons B arbres de décision. La prédiction pour un nouvel exemple x est obtenue sous forme de moyenne des prédictions B :

$$y \leftarrow \hat{f}(x) = \frac{1}{B} \sum_{b=1}^B f_b(x)$$

L'algorithme des forêts aléatoire diffère de "vanilla bagging" d'une seule façon [11]. Il utilise un algorithme d'apprentissage d'arborescence modifié qui inspecte, à chaque division du processus d'apprentissage, un sous-ensemble aléatoire de fonctions [11].

La raison est d'éviter la corrélation entre les arbres : si une ou plusieurs paramètres d'entrée sont des prédicteurs très puissants pour la cible, elles seront sélectionnées pour fractionner des observations dans de nombreux arbres [11]. Cela entraînerait de nombreux arbres corrélés dans notre «forêt». Les prédicteurs corrélés ne peuvent pas aider à améliorer la précision des prévisions [11].

L'algorithme des forêts aléatoires est l'un des algorithmes d'apprentissage d'ensemble les plus largement utilisés [11]. La raison en est qu'en utilisant plusieurs échantillons du jeu de données original, nous réduisons la variance du modèle final [11].

2.4.3 Gradient Boosting

Un autre algorithme d'apprentissage d'ensemble efficace est le "gradient boosting" [11]. Pour construire un régresseur fort, nous commençons avec un modèle dit "constant" $f = f_0$ générée en utilisant l'algorithme des forêts aléatoires :

$$f = f_0(x) = \frac{1}{N} \sum_{i=1}^N y_i$$

Puis on calcule la différence entre la valeur réelle et estimée, et on affecte le résultat à une autre variable appelée résidu [11] :

$$\hat{y}_i \leftarrow y_i - f(x_i)$$

La prochaine étape est de construire un autre modèle f_1 en utilisant le résidu comme sortie [11] :

$$f = f_0 + \alpha f_1$$

Puis on recalcule les résidus du modèle f_1 , et on les réutilise pour construire un autre modèle f_2 [11]. Et on répète le processus jusqu'à ce que les arbres de chaque modèle se combinent [11].

En calculant les résidus, nous estimons dans quelle mesure la sortie ou la variable à modéliser de chaque exemple d'entraînement est prédite par le modèle f [11]. Nous formons ensuite un autre arbre pour corriger les erreurs du modèle actuel et ajoutons ce nouvel arbre au modèle existant avec un certain poids [11].

Par conséquent, chaque arbre supplémentaire ajouté au modèle corrige partiellement les erreurs commises par les arbres précédents jusqu'à ce que le nombre maximum d'arbres soit combinés [11].

Les trois principaux hyperparamètres sont le nombre d'arbres, le taux d'apprentissage et la profondeur des arbres. La profondeur des arbres influence également sur la vitesse d'entraînement et de prédiction.

On peut montrer que la formation sur les résidus optimise le modèle global f pour le critère d'erreur quadratique moyenne [11]. La différence avec le bagging est que le boosting réduit le biais (ou sous-ajustement) au lieu de la variance, et cela peut conduire à un sur-apprentissage, cependant, en optimisant la profondeur et le nombre d'arbres, cela peut être largement évité [11].

Chapitre 3

Présentation des STEP de Béni Messous

3.1 STEP de Béni Messous

La station d'épuration des eaux usées de Béni Messous est administrativement localisée dans la daïra d'Ain Bénian (Wilaya d'Alger) et a été mise en service en 2008. La STEP est une station à boues activées à moyenne charge avec une capacité épuratoire de 250 000 Eq.Hab, pour un débit moyen théorique de 50.400 m³/j. Une extension a été faite pour doubler sa capacité de traitement et compléter le traitement tertiaire à partir de 2016 (avec une capacité épuratoire de 500 000 Eq.Hab, pour un débit moyen théorique de 100.800 m³/j). Deux extensions ultérieures sont prévues, portant la capacité totale de la STEP de Béni Messous à 810.000 Eq.Hab à l'horizon 2025. Cette station a reçu en moyenne 33.400 m³/jour en 2013, et produit 8.800 tonnes de boues à 22.9 % de siccité. Les performances de traitement permettent de garantir une conformité du rejet de 100% sur 2013 et un rendement d'élimination de la pollution de 95%. La station d'épuration de Béni Messous est alimentée par quatre sous bassin versants : Sous bassin versant d'Ain Bénian, d'Oued Béni Messous, Staoueli et Ouled Fayet.

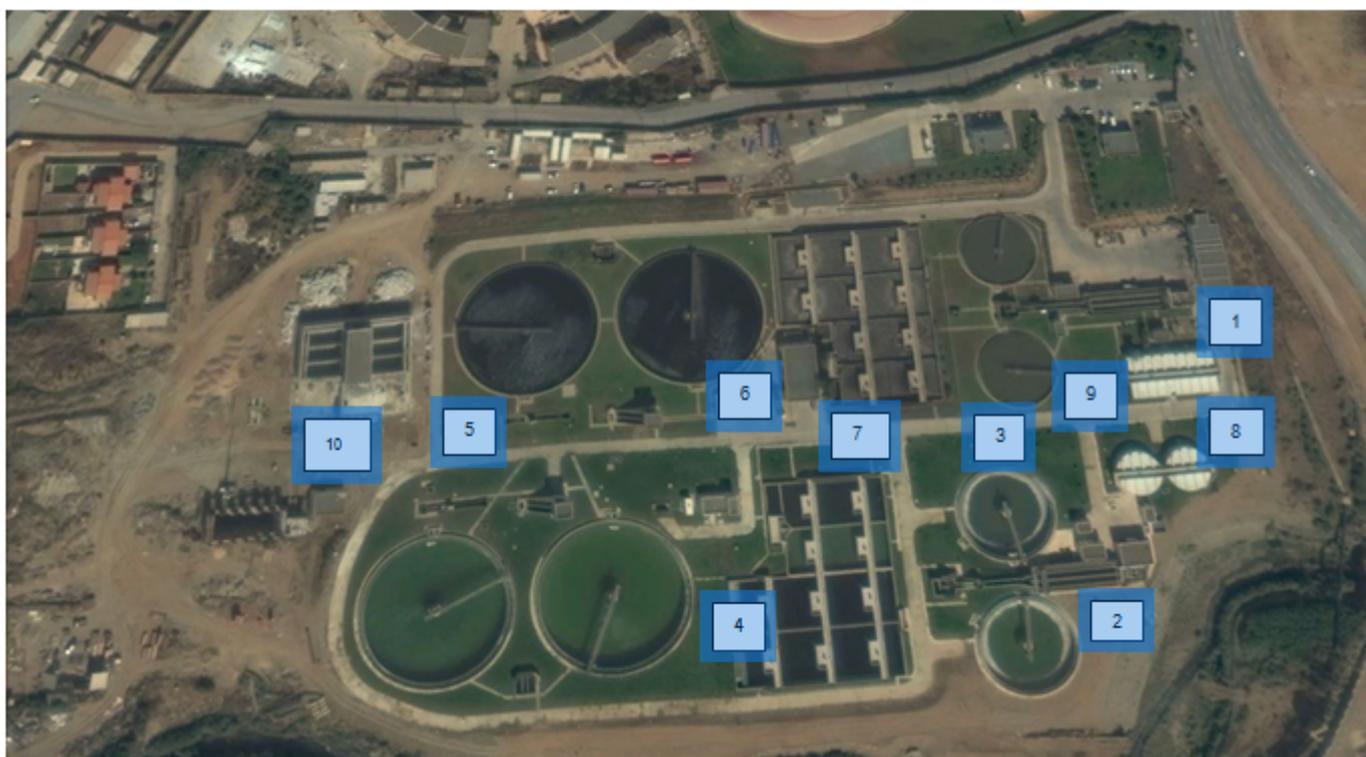


FIGURE 3.1 – Vue d'ensemble des ouvrages de la STEP de Béni Messous par Satellite

File d'eau

1. Déversoir d'orage / Dégrillage
2. Dessableur / déshuileur
3. Décanteur primaire : 4 décanteurs primaires.
4. Bassin d'aération : 4 bassins avec aérateurs de surface (Turbines)

File boues :

5. Clarificateur : 4 Clarificateurs

6. Déshydratation : 3 Filtres à bandes

7. Stabilisation : 4 bassins de stabilisation des boues.

8. Epaisseur : 4 épaisseurs hydrostatiques

Traitement de l'odeur

9. Bio filtres.

Désinfection :

10. Filtration sur sable et traitement UV.

3.1.1 Caractéristiques techniques et données de base du constructeur

station de Béni Messous a été dimensionnée pour fonctionner à moyenne charge, soit $C_m = 0,25$ à $0,3$ kg DBO5/kg MVS/jour, à capacité nominale. La connaissance réelle de la charge à traiter (inférieure à la charge nominale), permet de déterminer le taux de boue à maintenir dans le bassin d'aération. Dans le cas d'une réelle sous charge de la station (phase de démarrage par exemple), un fonctionnement à plus faible charge peut s'envisager.

Les caractéristiques de charges et de débits de la STEP sont résumées dans les tableaux suivants :

TABLE 3.1 – Caractéristiques techniques de la station de Béni Messous

| Paramètres | Entrée |
|--|--------|
| Débit nominal (m3/j) | 50 400 |
| Débit moyen horaire (m3/h) | 2 100 |
| Débit de pointe de temps sec (m3/h) | 2 750 |
| Débit de pointe de temps de pluie (m3/h) | 5 500 |

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

TABLE 3.2 – Concentrations admissibles des paramètres de pollution

| Paramètres | Entrée | | Sortie secondaire | | Sortie tertiaire | |
|------------|----------------------|------------|----------------------|------------|----------------------|------------|
| | Concentration (mg/L) | Flux (T/j) | Concentration (mg/L) | Flux (T/j) | Concentration (mg/L) | Flux (T/j) |
| MES | 297 | 30 | 3,0 | 5 | 0,5 | |
| DBO5 | 248 | 25,0 | 30 | 3,0 | 20 | 2 |
| DCO | 496 | 50,0 | 120 | 12,1 | 100 | 10 |

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 :

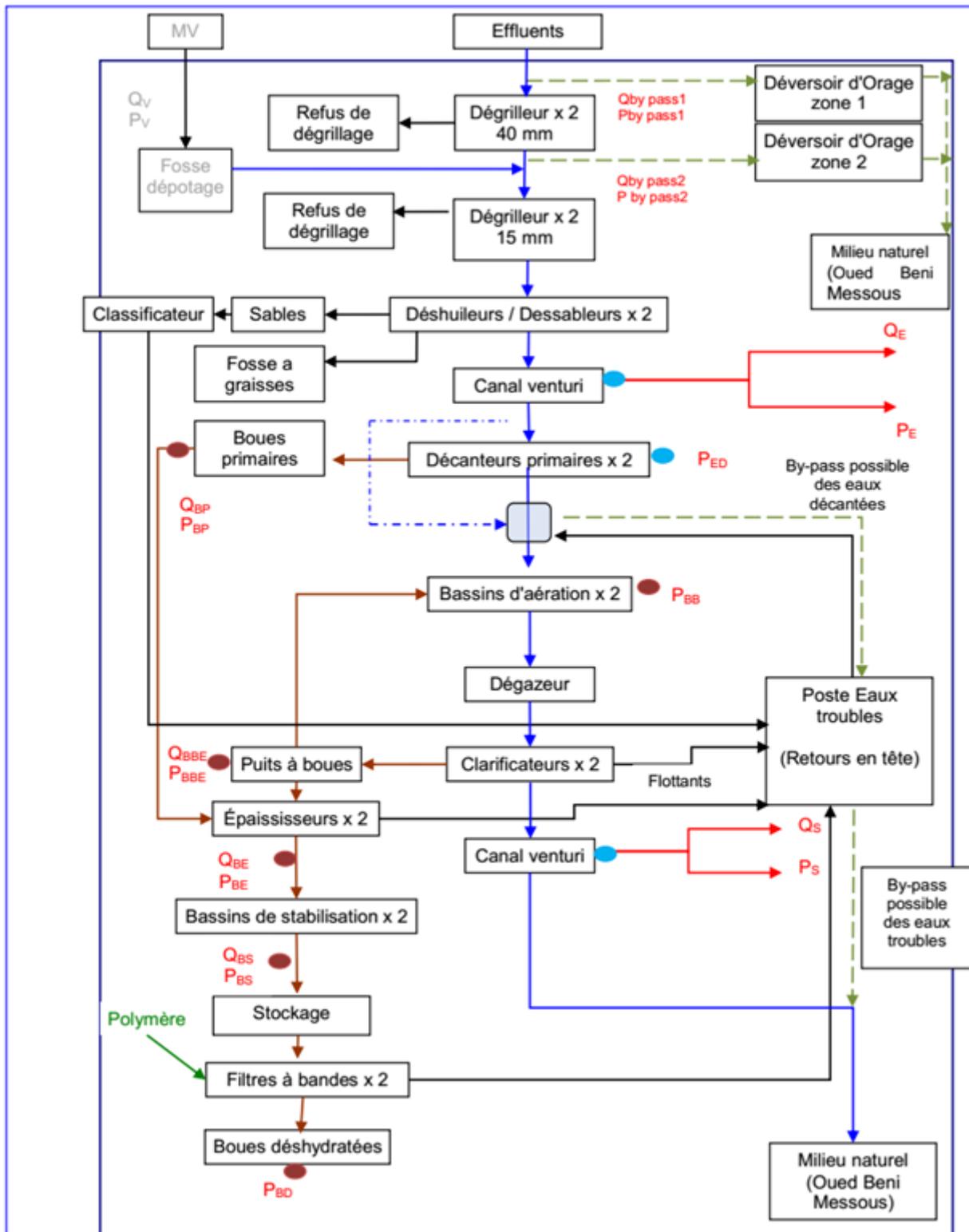


FIGURE 3.2 – : Schéma Synoptique de la STEP de Béni Messous

3.2 Principe de traitement

3.2.1 Principe de traitement

La STEP de Béni Messous comporte deux filières de traitement :

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

Le descriptif des filières et le procédé de traitement sont précisés ci-après.

TABLE 3.3 – Tableau récapitulatif des structure de traitement pour les filière eau et filière boues

| Ligne Eau | Ligne Boues |
|---|---|
| — Arrivée gravitaire équipée d'un déversoir d'orage · | — Poste de pompage des boues primaires · |
| — Pré-dégrillage x 2 de 40 mm · | — Poste de recirculation des boues biologiques · |
| — Dégrilleur x 2 de 15 mm · | — Poste d'extraction des boues biologiques en excès · |
| — Déshuileur / Désableur x 2 · | — Épaississeurs des boues x 2 · |
| — Réception des matières de vidange · | — Stabilisation aérobie x 2 · |
| — Canal de comptage de type Venturi (eaux brutes) · | — Bâche de stockage des boues à déshydrater · |
| — Décanteurs primaires x 2 · | — Filtres à bandes x 2 · |
| — Bassins d'aération x 2 · | — Aire de production des boues |
| — Répartiteur · | |
| — Clarificateurs x 2 · | |
| — Canal de comptage de type Venturi (eaux traitées) | |

Bassin d'aération

Les eaux prétraitées sont dirigées vers 2 bassins d'aération fonctionnant en parallèle : 2 x (55,5m x 15,5m) sur une profondeur de 4,58m, soit une capacité de 2 x 4 700 m³ = 9 400 m³. Chaque bassin est équipé de 3 turbines lentes de surface (soit 6 au total), d'apport spécifique unitaire brut de 130 kg O₂/h (donnée constructeur). Cette aération est assurée par 6 aérateurs de surface dont la fonction est double : aération et brassage des boues activées, afin d'éviter toute décantation de la biomasse. La régulation de l'aération s'effectue soit au moyen de sondes à oxygène (2 par bassin), implantées l'une en milieu de bassin, l'autre en sortie ou en mode tranche horaire. En sortie des deux bassins d'aération, la liqueur mixte (boue biologique /eau épurée) transite par un ouvrage rectangulaire de répartition, qui fait également office de dégazeur, avant l'alimentation des 2 clarificateurs. Le rôle du dégazeur est d'éviter le relargage de bulle d'air résiduelle dans le clarificateur qui pourrait empêcher la décantation finale des boues. La concentration de la boue doit être maintenue constante dans le bassin d'aération en tolérant une variation de +/- 0,5g/L (des concentrations excessives peuvent être pénalisantes pour l'aération d'une part et l'étape suivante de clarification d'autre part). La connaissance réelle de la charge à traiter (inférieure à la charge nominale), permet de déterminer le taux de boue à maintenir dans le bassin d'aération.



FIGURE 3.3 – Turbine en fonctionnement

Clarificateurs

Après la phase de contact entre l'eau décantée et la biomasse active dans le bassin d'aération, **Les eaux sortantes sont dirigées vers 2 clarificateurs circulaires** de diamètre unitaire 55m ,une surface unitaire d'un clarificateur de 2375m² ,une hauteur cylindrique de 3,95m et une volume unitaire de 9384m³ ,soit un volume total de 18 768m³ .Chaque clarificateur est équipé d'un pont raclé radial. La recirculation est assurée au moyen de deux vis d'Archimède (1 en secours) d'un débit max de 2 200m³/h. Le taux de recirculation maximum possible pour le débit moyen horaire est de 105% et de 40 % pour le débit de pointe temps de pluie. Ces équipements possèdent des variateurs de fréquence qui permettent de moduler le débit (débit minimum=1 500m³/h). Chaque vis dispose de sa propre fosse équipée de détecteurs niveau bas. Deux pompes d'extraction des boues en excès de débit unitaire 102m³/h sont implantées dans les fosses avec détecteur de niveau bas. Les boues en excès doivent être extraites le plus régulièrement possible vers l'épaississeur. Un contrôle du voile de boue présent dans les clarificateurs permet de vérifier l'évacuation constante des boues et d'éviter le passage en anaérobiose.



FIGURE 3.4 – Décanteur-Clarificateur

3.3 Statistiques et critiques des données :

3.3.1 Statistiques Descriptives

La STEP de Béni Messous a mis à notre disposition une série de données journalières de mesure de différents paramètres à l'entrée et à la sortie durant la période (2008 - 2016). Dans le cadre de ce travail, nous avons procédé à une analyse générale de ces données : une critique visuelle des graphiques de variation des paramètres en fonction du temps, avec un nombre d'observation de 3063 valeurs de 01/07/2008 à 18/11/2016. Les paramètres pris en considération sont résumés dans le tableau suivant :

TABLE 3.4 – Statistique descriptive de paramètres mesurées dans la STEP de Beni Messous

| | Eaux Brute | | | | Eaux Epurée | | | |
|----------------------|------------|-------|----------|------------|-------------|-------|----------|------------|
| | Min | Max | Moyenne | Ecart-type | Min | Max | Moyenne | Ecart-type |
| Débit m3/j | 0 | 72890 | 30609 | 10959.87 | 0 | 66270 | 27611.53 | 10432.91 |
| Débit By-Pass m3/j | 0 | 92680 | 3360.54 | 6911.6 | / | / | / | / |
| BR (m3/j) | 3000 | 86504 | 34942.47 | 6138.15 | / | / | / | / |
| pH | 6.35 | 8.63 | 7.67 | 0.26 | 6.3 | 8.69 | 7.69 | 0.28 |
| T °C | 6 | 29 | 16.26 | 4.55 | 7 | 27 | 16.03 | 4.34 |
| MES (mg/l) | 25 | 1351 | 221.52 | 152.72 | 1 | 91 | 9.52 | 7.2 |
| DBO5 (mg O2 /L) | 21 | 770 | 167.29 | 71.35 | 1.1 | 38.2 | 6.8 | 5.03 |
| DCO (mg O2 /L) | 29 | 960 | 276.39 | 108.52 | 5 | 96 | 24.61 | 13.47 |
| NGL (mg/L) (NT) | 11.8 | 283 | 41.67 | 19.96 | 9 | 146 | 29.4 | 13.09 |
| N-NO3 (mg/L) | 0.1 | 22 | 1.83 | 1.66 | 0.12 | 31 | 14.47 | 5.47 |
| N-NH4 (mg/L) | 0.1 | 58 | 21.11 | 7.42 | 0.01 | 19.7 | 1.91 | 3.35 |
| P-PO4 (mg/L) | 0.6 | 67 | 5.01 | 5.63 | 0.1 | 20 | 2.4 | 1.74 |
| PT (mg/L) | 1.2 | 13 | 4.33 | 1.89 | 0.4 | 19 | 2.7 | 2.09 |
| Conductivité (us/cm) | 504 | 2584 | 1517.3 | 228.7 | 717 | 2421 | 1431.3 | 203.4 |
| Energie Totale (kWh) | 266 | 21492 | 9818.39 | 1847.78 | / | / | / | / |
| Somme des heures | 8 | 129 | 71.94 | 18.25 | / | / | / | / |

3.3.2 Evolution des paramètres de pollution dans le temps

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

Evolution du Débit :

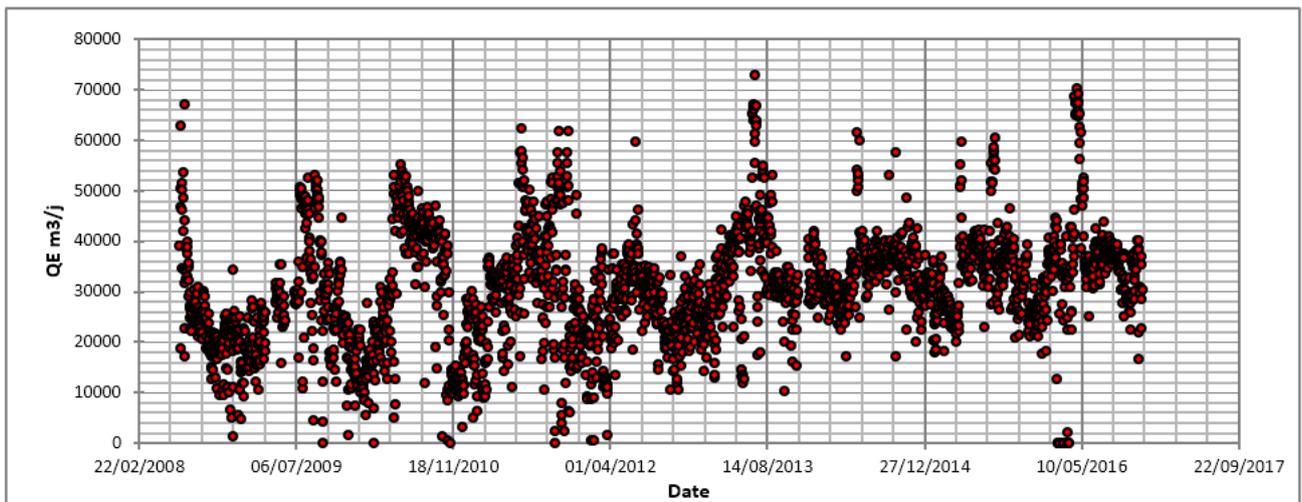


FIGURE 3.5 – Evolution journalière du débit entrant dans la STEP

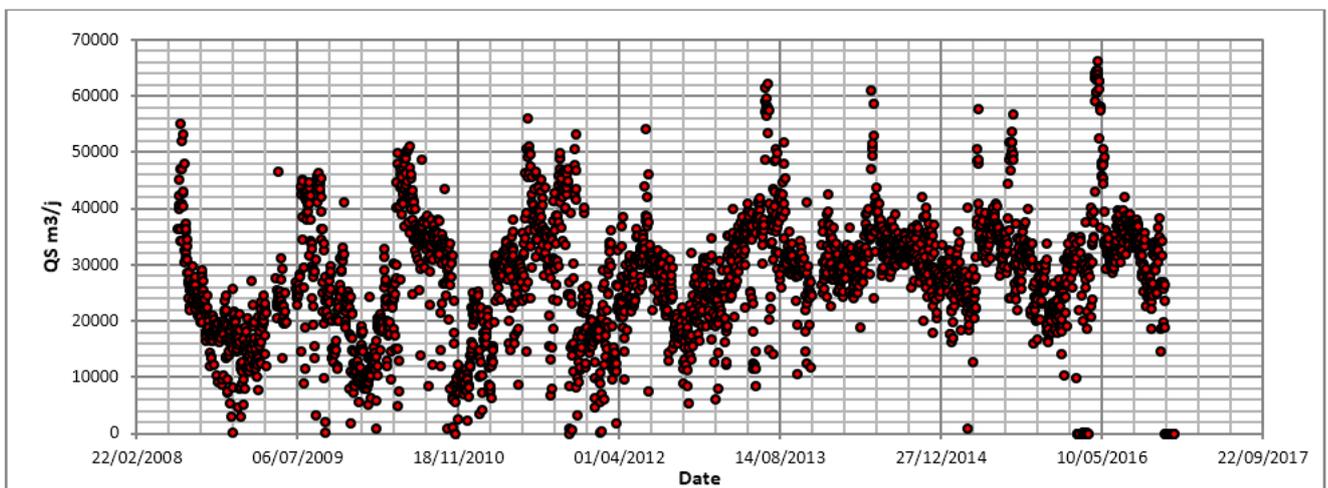


FIGURE 3.6 – Evolution journalière du débit sortant dans la STEP

Le débit entrant comprend une tendance croissante au fur et à mesure du temps, et varie entre 0 et 72890 m³/j (d'après l'analyse de données, on a trouvé que 133 valeurs sur 3063 sont supérieures à 50000 m³/j qui est le débit nominal de la STEP avant 2016) avec une moyenne de 30 609 m³/j.

On remarque une variation presque cyclique des débits, avec des valeurs extrêmes en période estivale (Avril - Aout) entre 2008 et 2013, elles sont probablement dues à la forte consommation domestique.

- Les débits dont les valeurs dépassent la capacité nominale, sont dirigés vers un By – Pass, et peuvent s'expliquer, soit par les fortes précipitations enregistrées, soit par les travaux sur le dalot ou par le dysfonctionnement de la STEP.
- L'évolution des débits entrants ressemble beaucoup à celle des débits sortants, avec une moyenne de 27 611 m³/j.

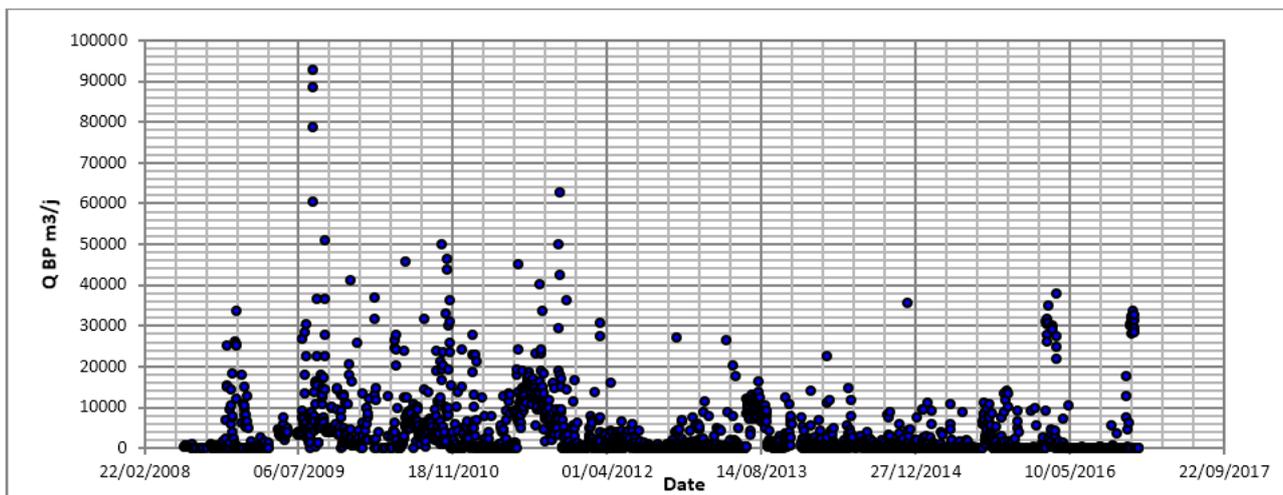


FIGURE 3.7 – Evolution journalière du débit By Pass dans la STEP

Pour le débit By passer :

- On remarque des valeurs max dépassant les 78000 m³/j.
- Du 19/08/2009 au 23/08/2009 le débit By Passer a atteint des pics variant entre 60390-92680 m³/j.
- On a observé des valeurs de débit by passer assez grande, depuis la mise en service de la STEP jusqu'à avril 2012, et puis une tendance à la baisse au fur et à mesure du temps.
- Des travaux sur le dalot T2 pendant le mois de mars 2016.

Evolution du débit recirculé

Le débit recirculé varie dans la période de 2008-2013 avec une stabilisation en 2013 -2016 avec une moyenne de 34942.6 m³/j et un maximum de (86504 m³/j) 60946 m³/j.

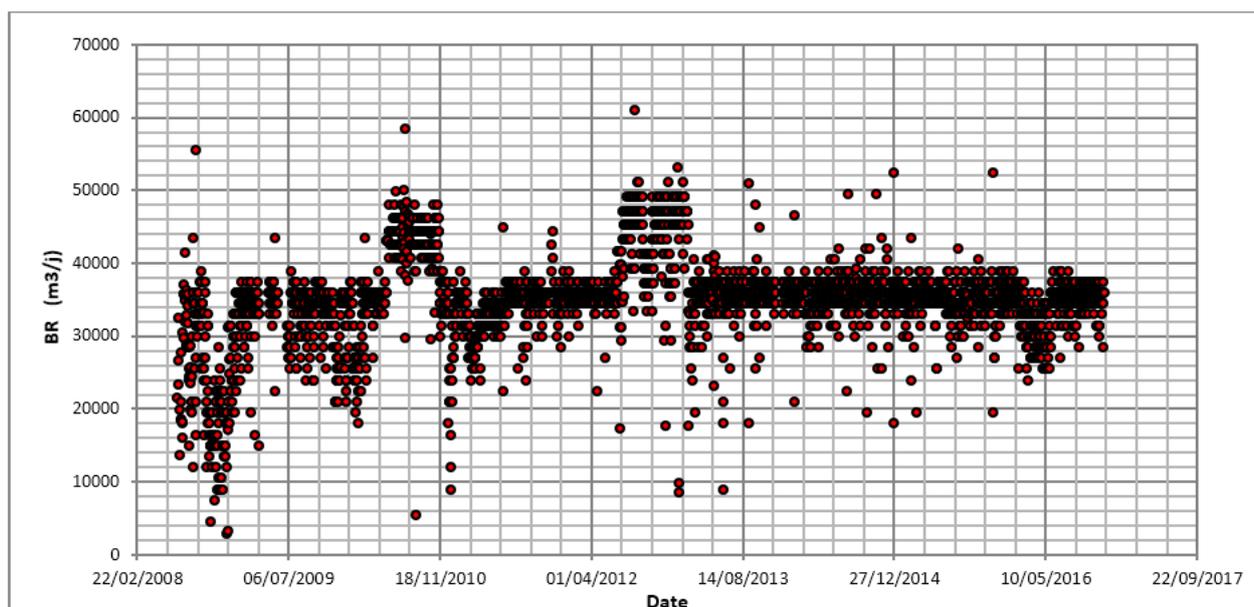


FIGURE 3.8 – Evolution journalière du débit recirculé dans la STEP

Evolution des Matières En Suspension (MES)

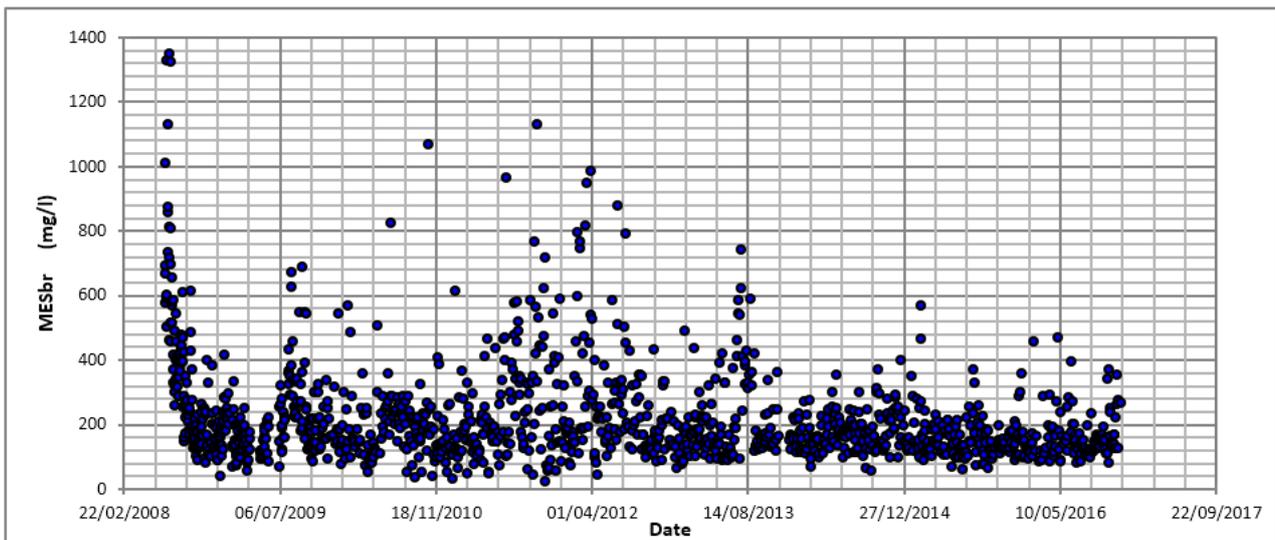


FIGURE 3.9 – Evolution journalière de la concentration en MES à l'entrée de la STEP

À l'entrée de la STEP, les valeurs des MES oscillent entre 50 et 400 mg/l, avec une concentration moyenne de 221,52 mg/l. On enregistre aussi des pics atteignant 1330 mg/l que l'on peut expliquer par des phénomènes exceptionnels, comme l'augmentation du flux de MES par temps de pluie qui provient majoritairement de la remise en suspension des dépôts accumulés. En outre, on peut aussi formuler les remarques suivantes :

- A partir du 01/09/2013, la concentration des MES brute s'est stabilisée dans le temps.
- Les valeurs des MES, ont une dispersion remarquable, avec des pics élevés, au début qui correspondent au début du fonctionnement de la STEP, et des valeurs dispersées verticalement à partir de Juillet 2009 à Aout 2013.

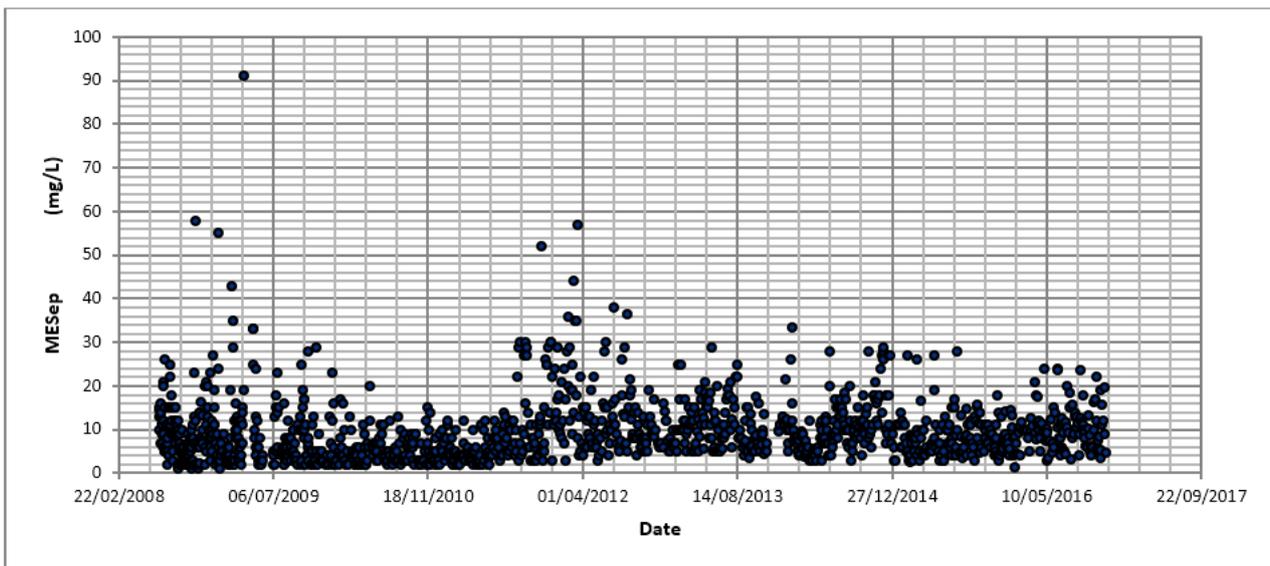


FIGURE 3.10 – Evolution journalière de la concentration en MES à la sortie de la STEP

À la sortie de la STEP, la majorité des valeurs des MES enregistrées, sont comprise entre 5 et 20 mg/l, avec une moyenne de 9,5 mg/l, avec quelques exceptions de 41 valeurs sur 1346 mesurées où la

concentration dépasse la norme de rejet (30mg/l), avec des valeurs max atteignant respectivement les 91 et 57 mg /l en 2009 et 2013. Il est à noter aussi que 891 valeurs sont inférieures à 10mg/l, ce qui dénote un rendement très élevé par rapport aux normes de rejets et aux valeurs brutes des MES, et par conséquent, une consommation élevée d'énergie.

Evolution de la Demande Biologique en Oxygène (DBO5)

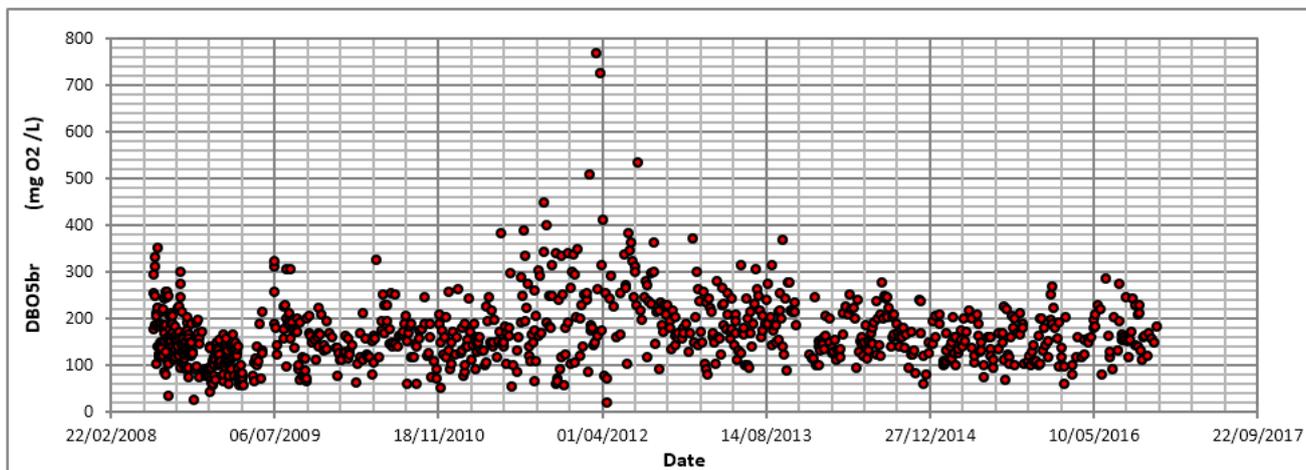


FIGURE 3.11 – Evolution journalière de la Demande Biologique en Oxygène à l'entrée de la STEP

A l'entrée de la STEP, la concentration de la Demande Biologique en Oxygène oscille principalement entre 50 et 500 mg/l avec une moyenne de 167,28 mg/l. On souligne aussi le pic de 770 mg/l enregistré le 14/03/2012.

On note aussi une évolution croissante de la DBO5, entre Mai 2011 et Mai 2013.

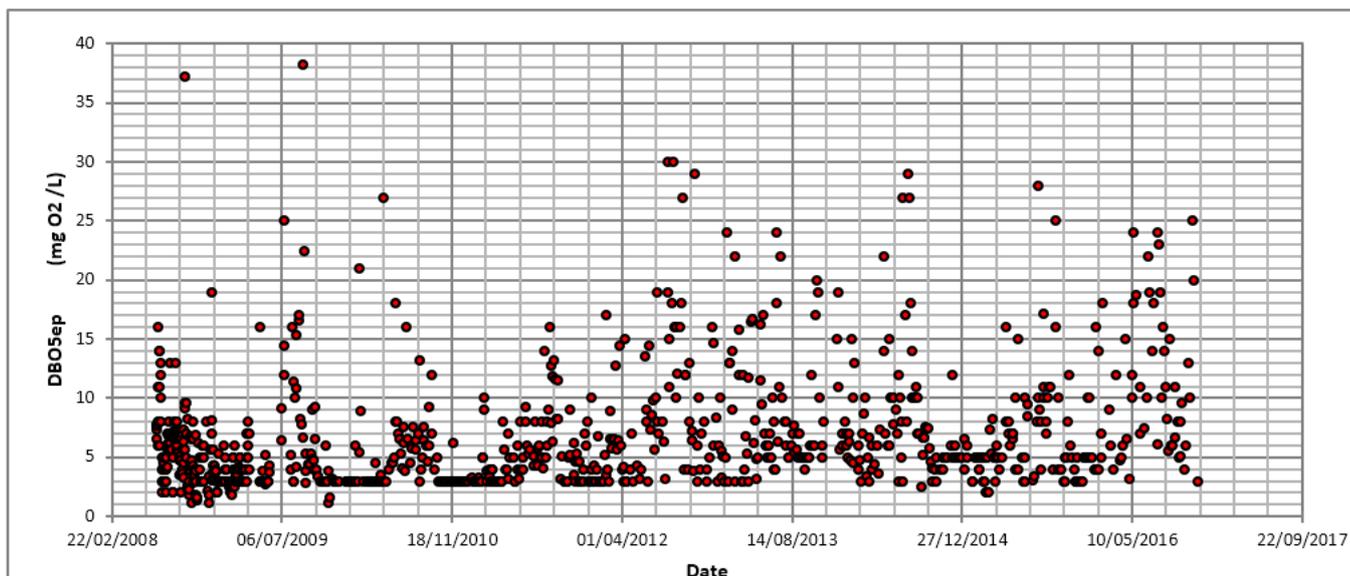


FIGURE 3.12 – Evolution journalière de la Demande Biologique en Oxygène à la sortie de la STEP

A la sortie de la STEP, la concentration de la DBO5 oscille principalement entre 3 mg/l et 20 mg/l, avec une moyenne de 6,8 mg/l, qui est une valeur vraiment inférieure à la norme de rejet (30 mg/l), ce qui peut expliquer la surconsommation d'énergie dans la STEP. On a noté aussi quelques valeurs exceptionnelles qui peuvent s'expliquer soit par une charge organique entrante

supérieure à la charge nominale dont les micro-organismes ne peuvent pas tout dégrader, soit par un problème au niveau du bassin d'aération ou encore des travaux ayant lieu au niveau de la STEP.

Evolution de la Demande Chimique en Oxygène (DCO)

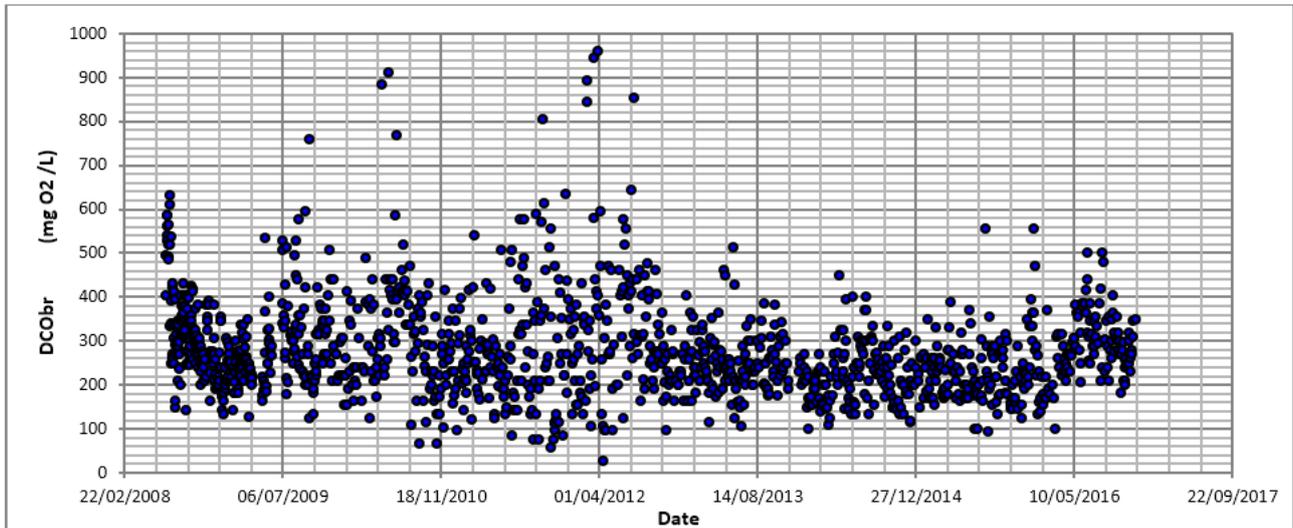


FIGURE 3.13 – Evolution journalière de la Demande Chimique en Oxygène à l'entrée de la STEP

La concentration de la Demande Chimique en Oxygène (DCO) oscille entre 70 et 600 mg/l, avec une moyenne de 276,4 mg/l. On remarque des valeurs exceptionnelles supérieures à 900 mg/l, et dépassant la concentration admissible par la STEP (496 mg/l) entre 2008 à 2013, avec une stabilité à partir de 2013.

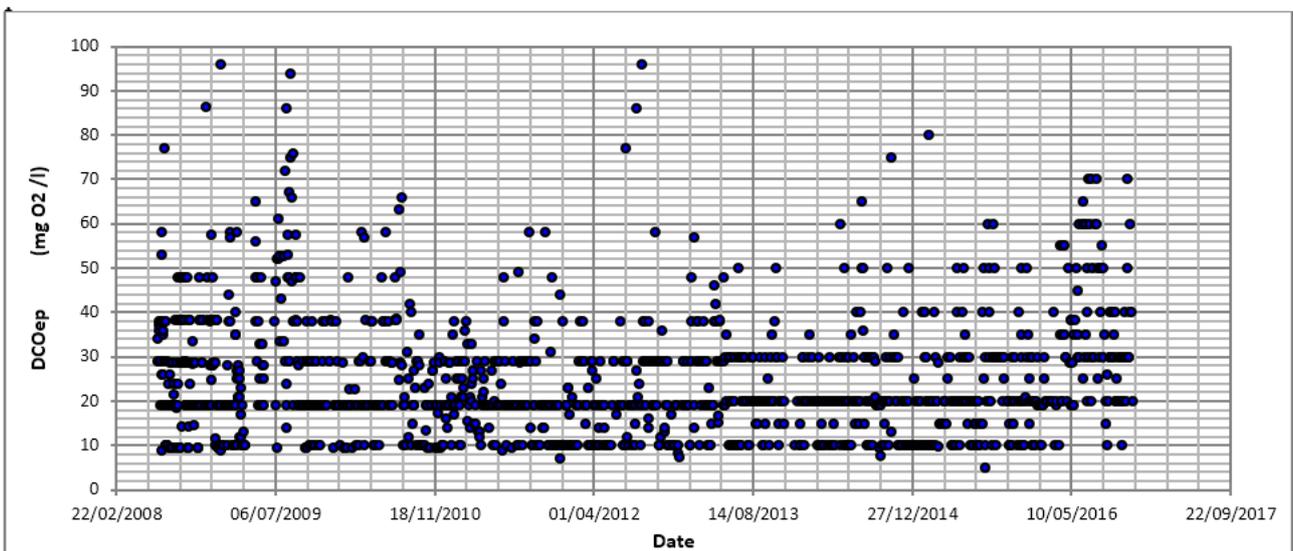


FIGURE 3.14 – Evolution journalière de la Demande Chimique en Oxygène à la sortie de la STEP

La DCO de l'eau épurée oscille entre 10 mg/l et 60 mg/l, avec une moyenne de 24,6 mg/l. La majorité des valeurs mesurées se situe dans la gamme de 10 - 70 mg/l, et est toujours inférieure à la norme de rejet qui est de 120 mg/l.

Evolution de l'azote globale (NGL)

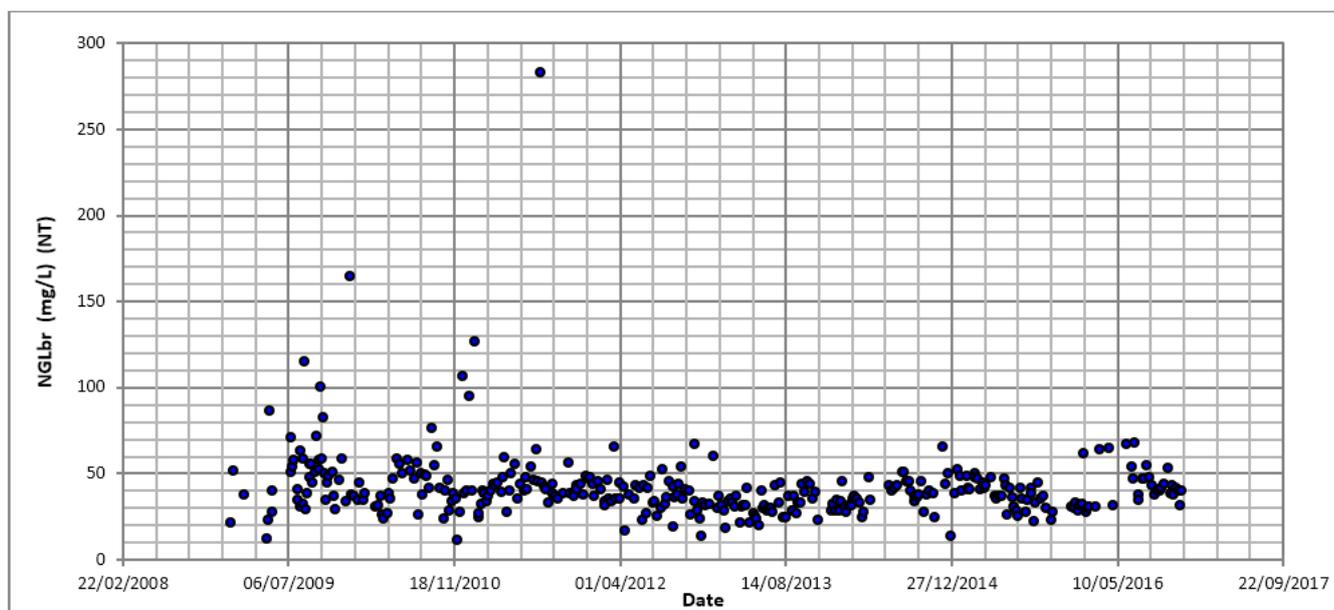


FIGURE 3.15 – Evolution journalière l'azote global NGL à l'entrée de la STEP

A l'entrée de la STEP, la concentration de l'azote globale oscille entre 11 et 100mg/l, avec une moyenne de 41,67 mg/l, et des valeurs max atteignant respectivement 165 mg/l et 283 mg/l en 2010 et 2011. Le NGL est relativement constant en fonction du temps.

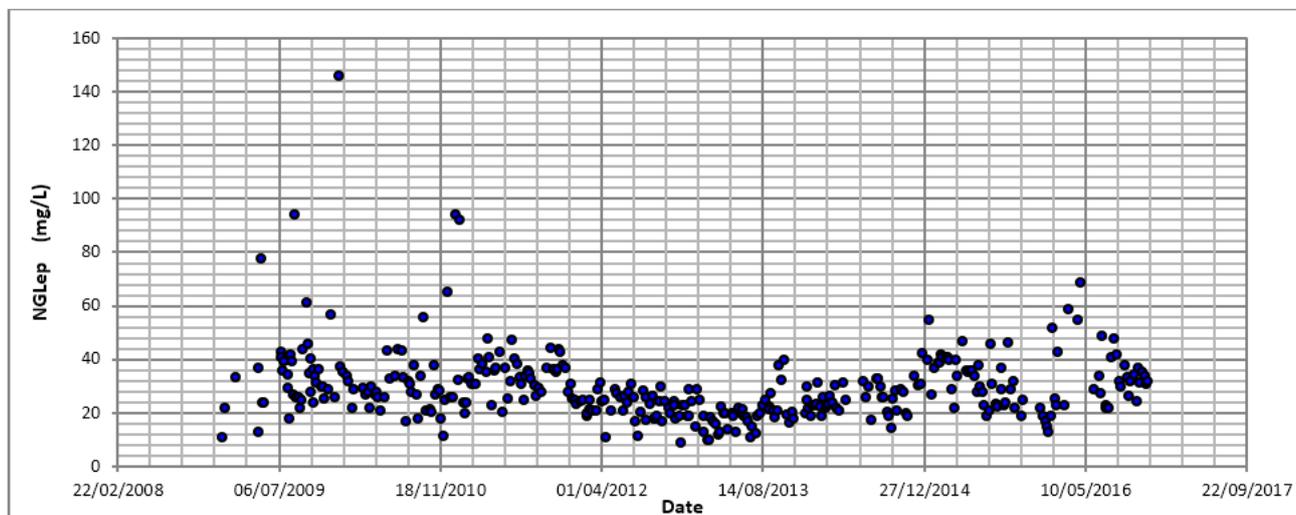


FIGURE 3.16 – Evolution journalière DE l'azote global NGL à la sortie de la STEP

A la sortie de la STEP, la concentration d'ammonium varie entre 10 mg/l et 70 mg/l avec une moyenne de 29,4 mg/l qui est supérieure à la norme du rejet (10 -15 mg/l). Ceci est dû à de fortes charges organiques entrainant un degré de nitrification plus faible, avec quelques pics remarquables comme en 2011 avec une valeur de 247mg/l.

Malheureusement pour ce paramètre, Il n y que 300 valeurs mesurées sur 3063 pour l'ensemble des autres paramètres, ce qui constitue pour nous un obstacle en termes de données pour la modélisation. Par conséquent, ce paramètre ne sera pas pris en considération dans la partie modélisation.

Evolution de l'ammonium (NH₄⁺)

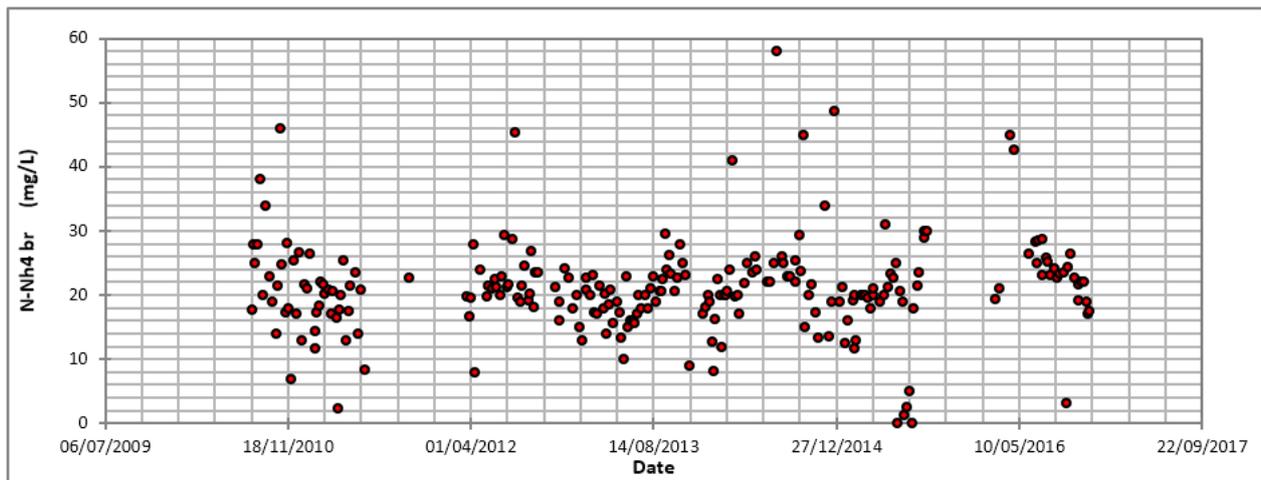


FIGURE 3.17 – Evolution journalière de NH₄⁺ à l'entrée de la STEP

A l'entrée de la STEP, La concentration d'ammonium brute (NH₄⁺) oscille entre 3 mg/l et 45 mg/l, avec une moyenne de 21.11 mg/l. On observe une dispersion assez élevée pour le NH₄, avec un pic de 58 mg/l enregistré en 2014 et quelques pics supérieurs à 40 mg/l durant toute la période d'observation. Ces variations sont fonction de la température et du temps de séjour dans le réseau de collecte.

Les bactéries utilisent l'ammonium (NH₃, NH₄⁺) dans le processus de nitrification, pour transformer la matière organique en nitrites (NO₂) puis en nitrates (NO₃).

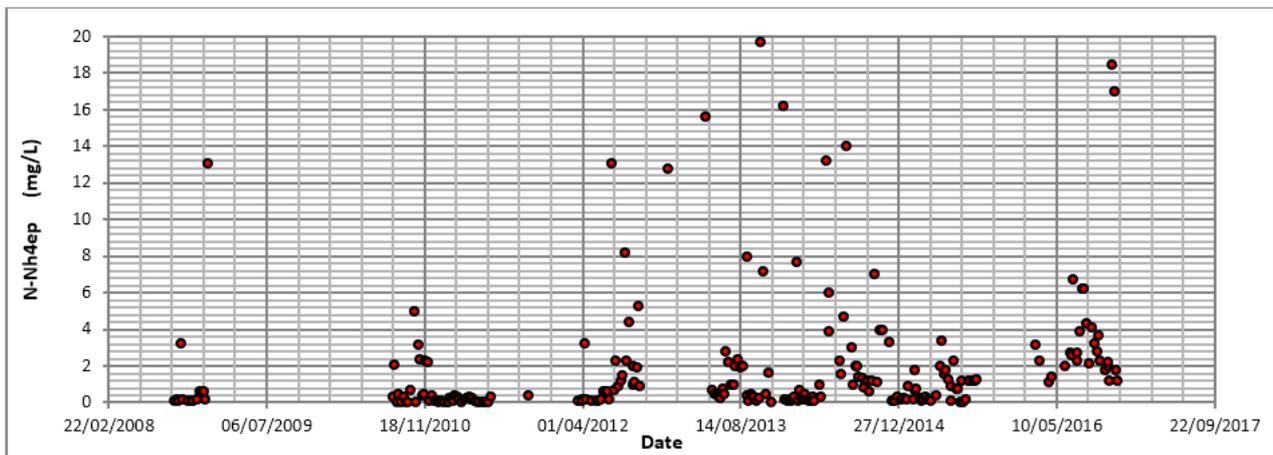


FIGURE 3.18 – Evolution journalière de NH₄⁺ à la sortie de la STEP

A la sortie de la STEP, la concentration d'ammonium est variable et dispersée, allant de 0,1 mg/l à 12 mg/l avec une moyenne de 11,18 mg/l. Graphiquement, on remarque une diminution de la concentration d'ammonium à la sortie par rapport l'entrée, à cause de la nitrification, afin d'éliminer le NH₄⁺ des eaux usées. On remarque aussi des pics qui varient de 12 mg/l à 19,7 mg/l à partir de 2012.

La même remarque que pour le paramètre précédent est appliquée pour l'ammonium, à savoir, l'élimination de ce paramètre dans la partie modélisation.

Variation du Nitrate (NO₃-)

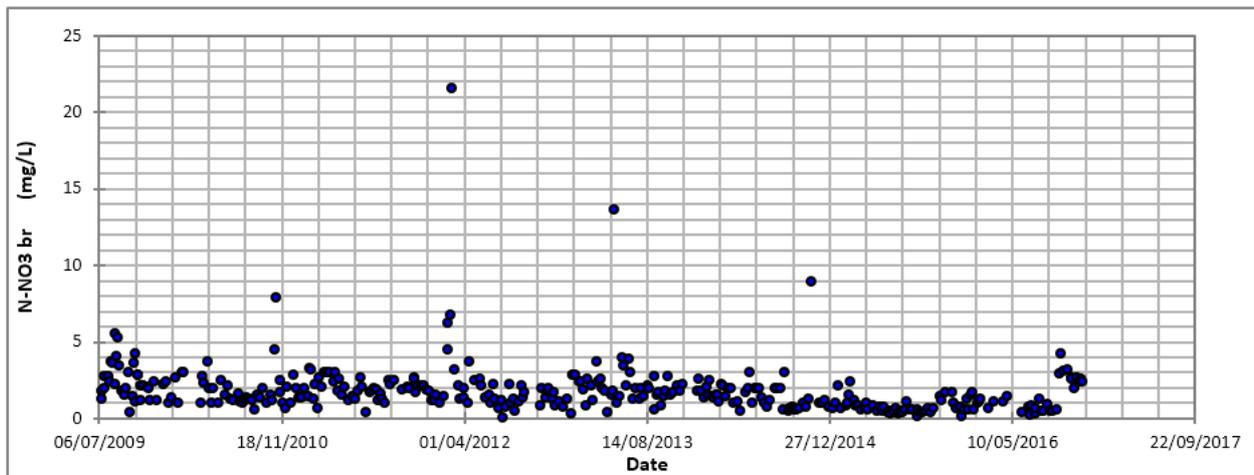


FIGURE 3.19 – Evolution journalière de NO₃- à l'entrée de la STEP

À l'entrée de la STEP, les concentrations du NO₃- oscillent entre 1 et 10 mg/l avec une moyenne de 1.83 mg/l, Le NO₃ reste lui aussi constant, mais on observe une légère tendance à la baisse à la fin de la période d'observation. Avec un pic de 21,6 mg/l en 2012.

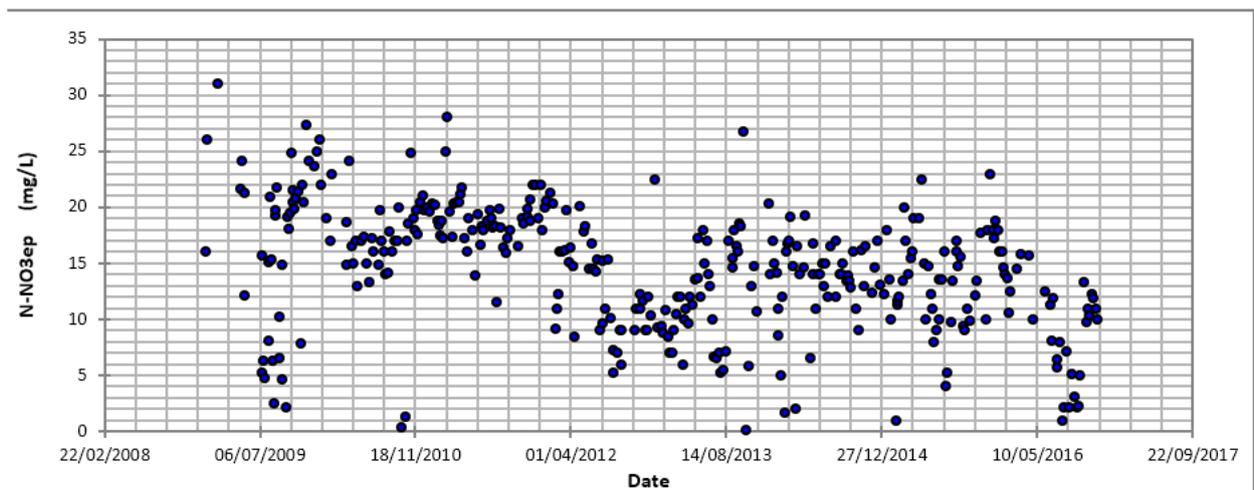


FIGURE 3.20 – Evolution journalière de NO₃- à la sortie de la STEP

À la sortie de la STEP, les concentrations du NO₃- oscillent entre 1 et 25 mg/l avec une moyenne de 14.47 mg/l. Graphiquement, on voit une légère tendance à la baisse au fur et à mesure du temps, avec une dispersion verticale.

Cette augmentation de concentration de nitrates (NO₃) est due à la nitrification et l'apparition de nitrates.

Evolution de Phosphore Total PT

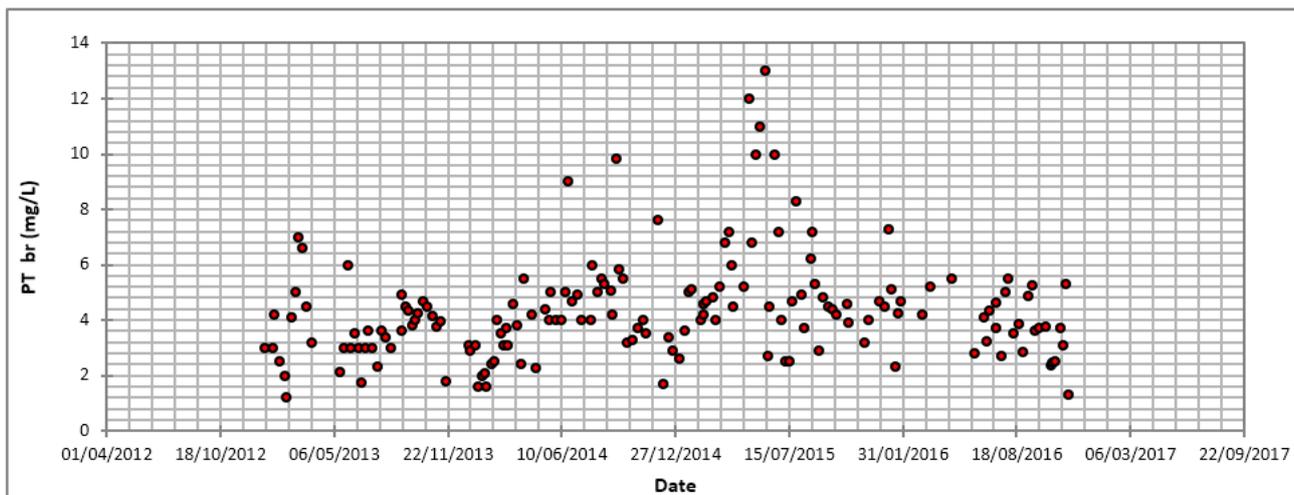


FIGURE 3.21 – Evolution journalière de Phosphore total à l'entrée de la STEP

Phosphore total : $PT = \text{Phosphore organique} + \text{Phosphore minéral}$. 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).

A l'entrée de la STEP, les concentrations du PT oscillent entre 1,2 et 12 mg/l, avec une moyenne de 4,33 mg/l et un pic de 13 mg/l. Le PT manque considérablement de mesures et on observe une dispersion de cette variable.

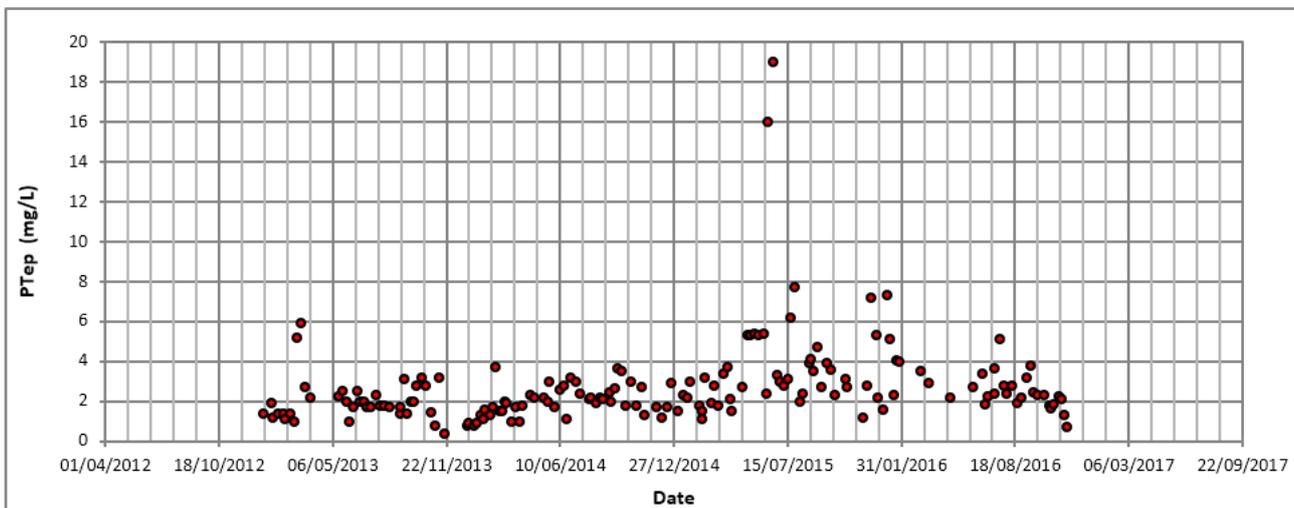


FIGURE 3.22 – Evolution journalière de Phosphore total à la sortie de la STEP

A la sortie de la STEP, les concentrations du PT oscillent entre 0,5 et 8 mg/l, avec une moyenne de 2,7 mg/l qui dépasse la norme de rejet (2mg/l), avec l'apparition de valeurs max de 16 à 19 mg/l en 2015.

Evolution de l'ortho-phosphate (PO4-3)

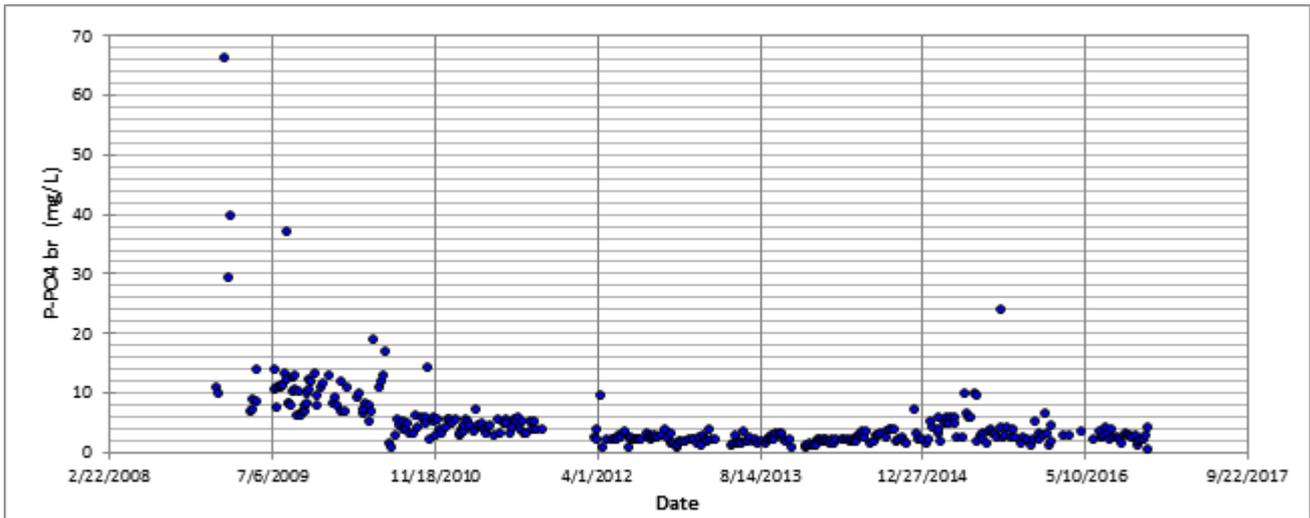


FIGURE 3.23 – Evolution journalière de l'ortho-phosphate (PO4-3) l à l'entrée de la STEP

A l'entrée de la STEP, les concentrations du (PO4-3) oscillent entre 0,6 et 25 mg/l, avec une moyenne de 5 mg/l, et une tendance à la baisse au fur et à mesure du temps et des pics de 66,57 mg/l et 40 mg/l en 2009, due probablement aux rejets d'eaux résiduares, au début du fonctionnement de la STEP.

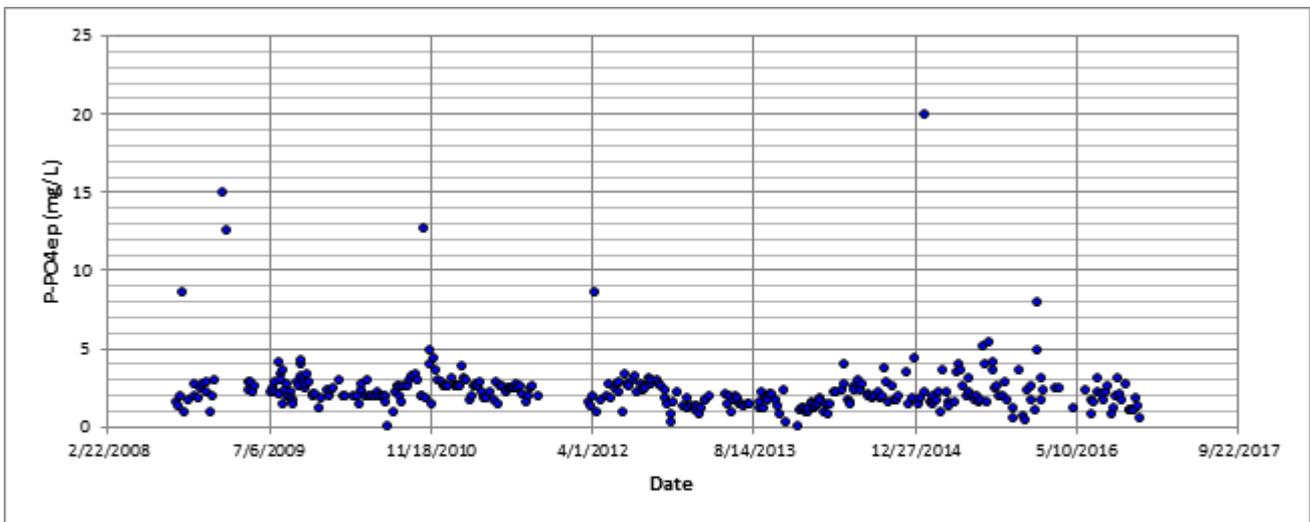


FIGURE 3.24 – Evolution journalière de l'ortho-phosphate (PO4-3) à la sortie de la STEP

A la sortie de la STEP, les concentrations du (PO4-3) oscillent entre 0,1 et 15 mg/l, avec une moyenne de 2,4 mg/l (5 mg/l à l'entrée), cette diminution est due à la déphosphatation biologique et chimique, en remarquant la dégradation de phosphate (dégradation de la MO).

Evolution de pH)

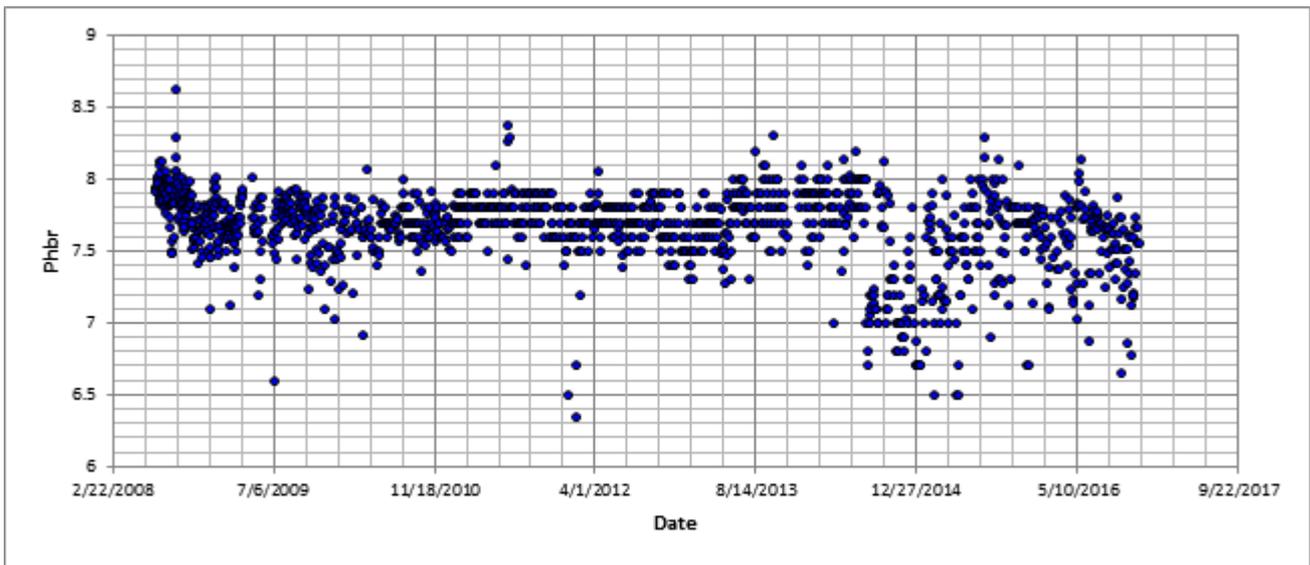


FIGURE 3.25 – Evolution journalière de l'ortho-phosphate (PO4-3) à la sortie de la STEP

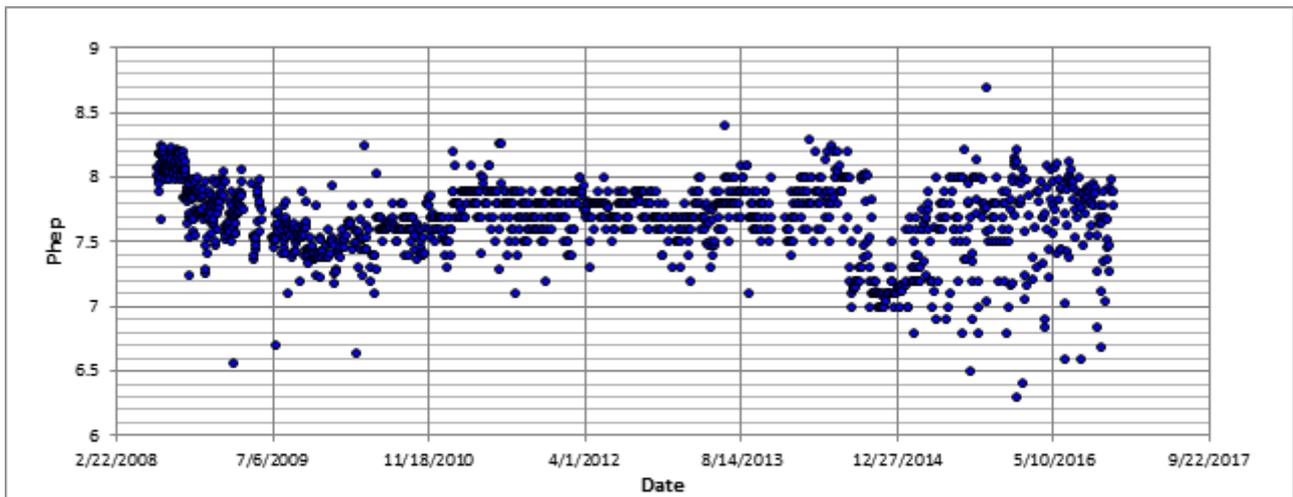


FIGURE 3.26 – Evolution journalière de l'ortho-phosphate (PO4-3) à la sortie de la STEP

En comparant l'évolution du PH à l'entrée et à la sortie de la STEP, on remarque des fluctuations insignifiantes. On voit une stabilisation entre la période de 2008 jusqu'à 2014, néanmoins une diminution est détectée entre 2014-2016. Le PH de l'eau brute varie légèrement et oscille entre 6,35 et 8,63 avec une moyenne de 7,67. Le PH de l'eau épurée varie légèrement et oscille entre 6,3 et 8,69 avec une moyenne de 7,68.

Les limites du pH pour la croissance et la reproduction bactérienne ont été étudiées par de nombreux auteurs. Les systèmes de bio-oxydation utilisés classiquement tolèrent une gamme de pH allant de 5 à 9 avec une zone optimale de 6 à 8. En milieu anaérobie, la gamme possible est même plus réduite avec un pH de 6,8 à 7,5 [13].

Variation de la conductivité

La conductivité électrique est elle aussi constante au fur et à mesure du temps, avec une moyenne de 1517 (us/cm), et des perturbations à partir de 2015 -2016.

La conductivité est un paramètre qui influe sur la croissance des bactéries. Il est important de la contrôler dans le bassin d'aération, d'autant plus que les bactéries sont sensibles à sa variation [13].

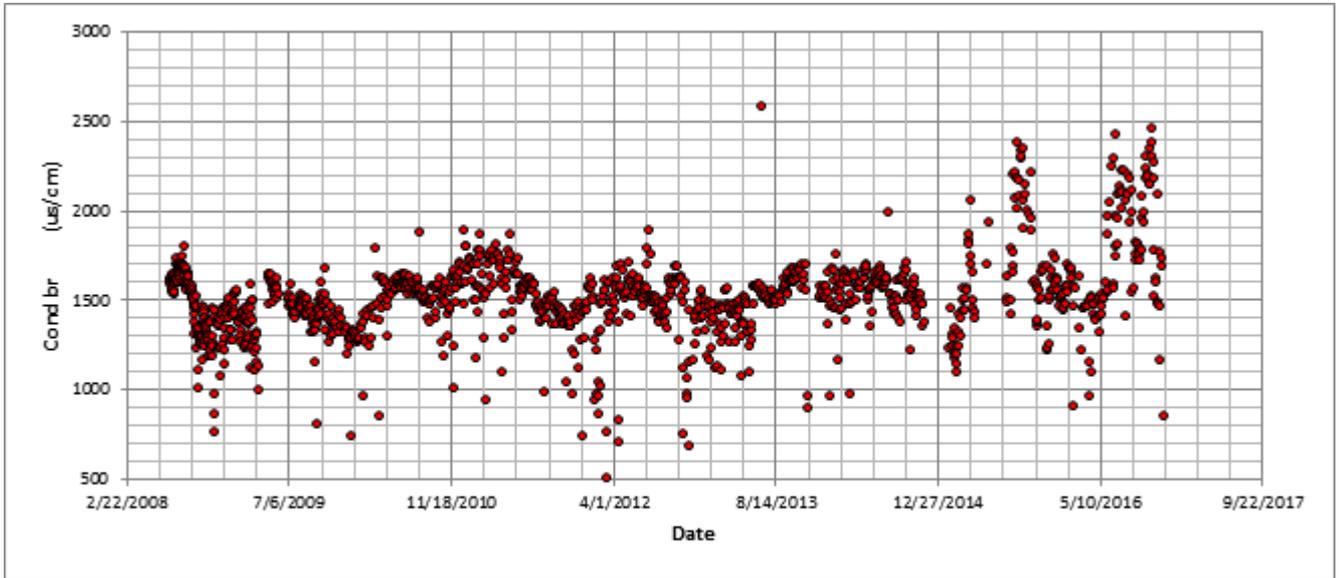


FIGURE 3.27 – Evolution journalière de la conductivité de l'eau brute

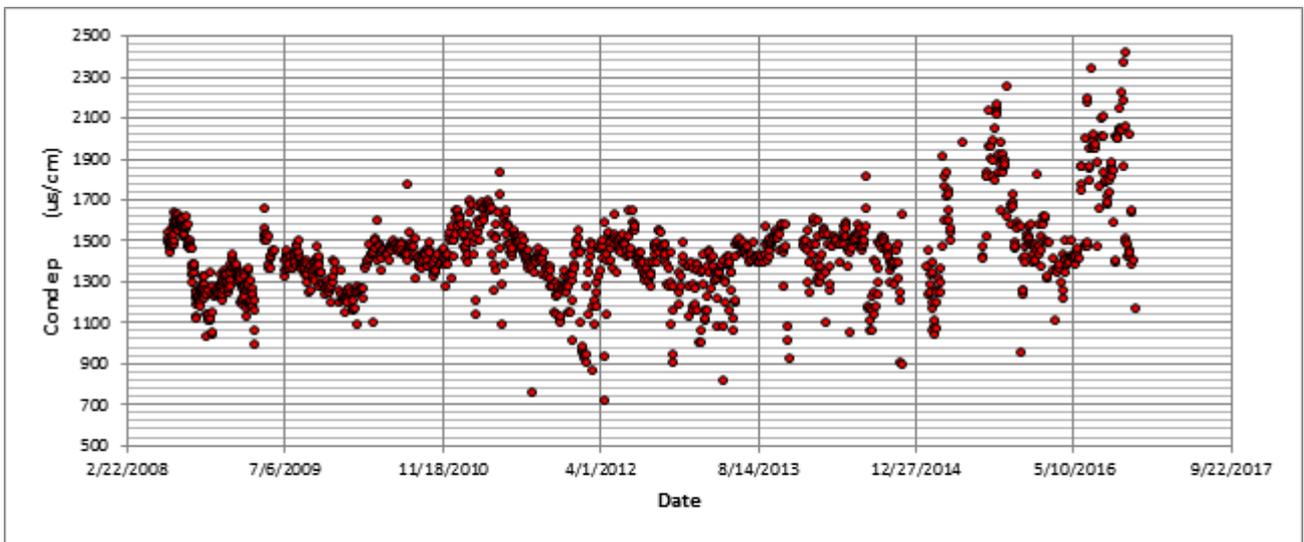


FIGURE 3.28 – Evolution journalière de la conductivité de l'eau épurée

Evolution de la température :

On remarque que la variation de la température est cyclique à l'entrée comme à la sortie, et oscille entre 10 et 27 °C, à l'exception de certains cas où elle atteint un minimum de 7°C et un maximum de 29 °C.

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

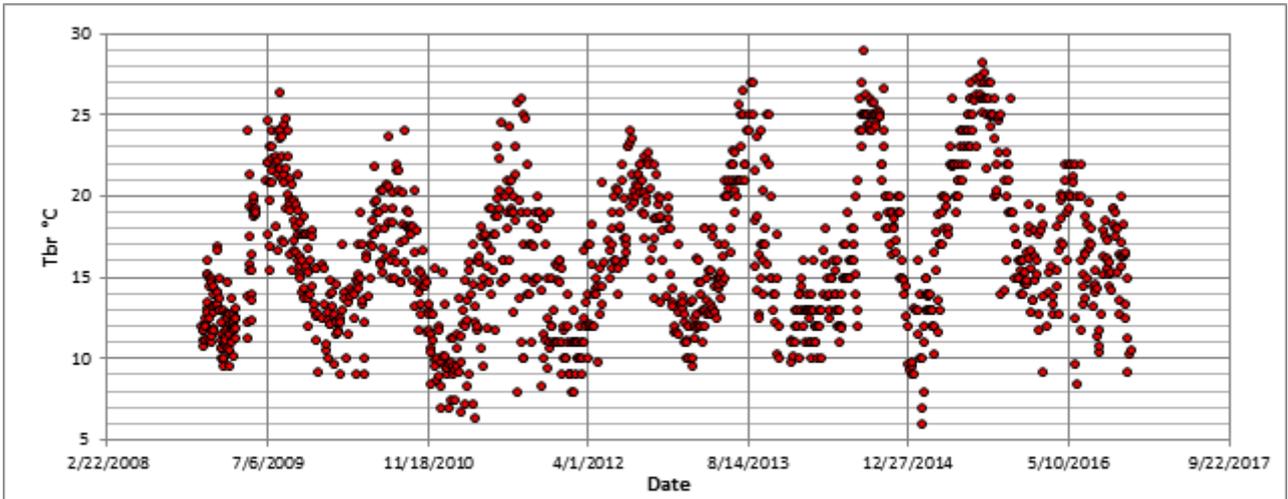


FIGURE 3.29 – Variation journalière de la température de l'eau brute

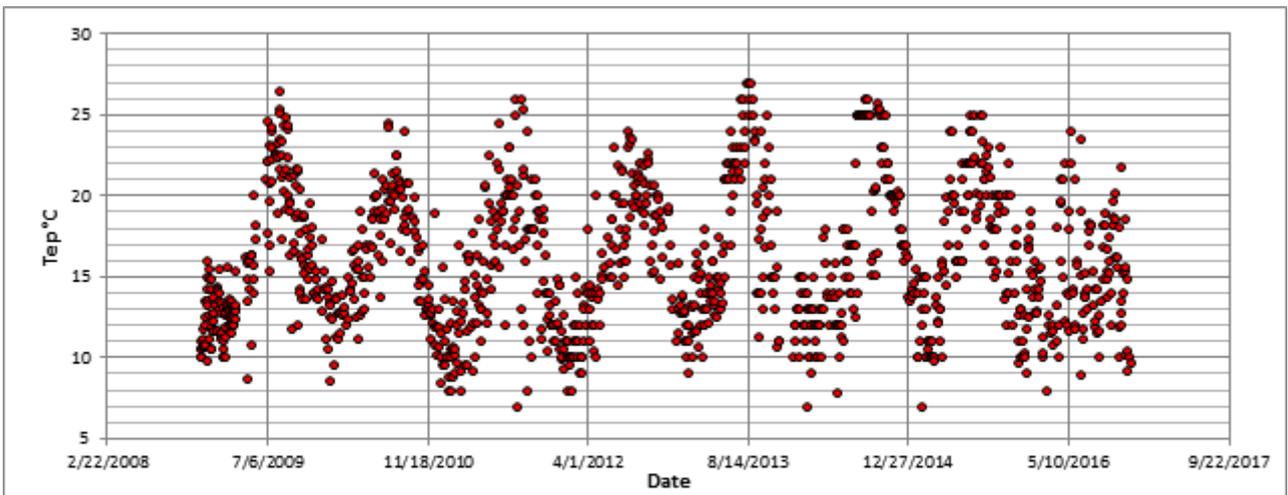


FIGURE 3.30 – Variation journalière de la température de l'eau épurée

Evolution de la consommation de l'énergie

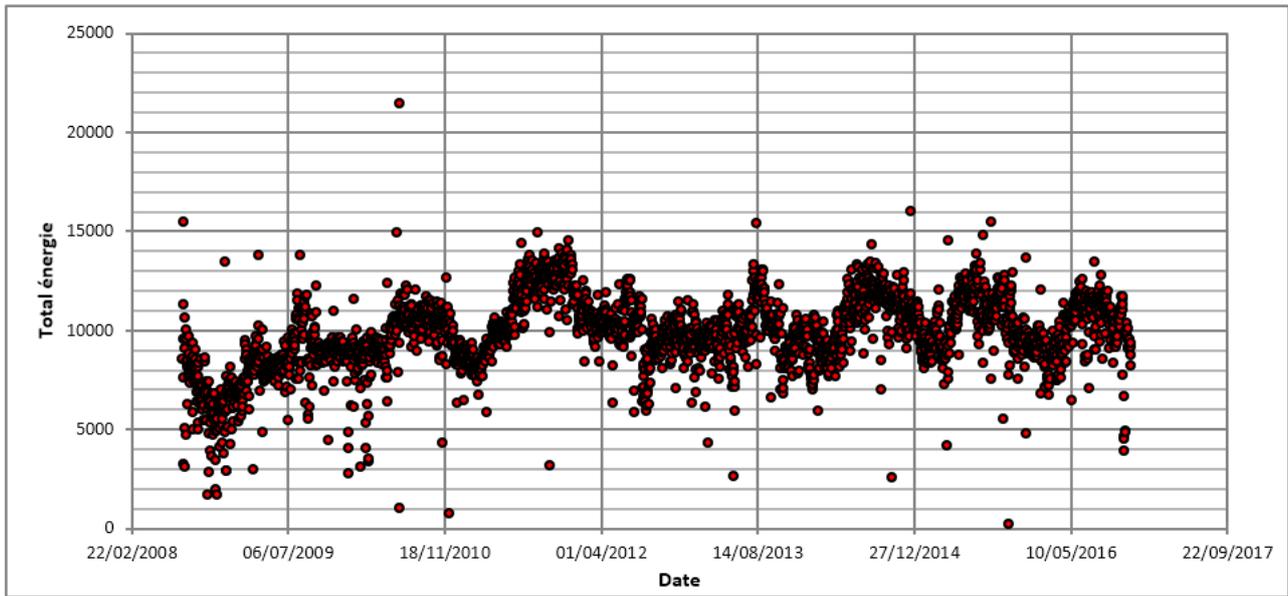


FIGURE 3.31 – Evolution journalière de l'énergie dans la STEP

L'énergie totale consommée durant le processus d'épuration des eaux de la STEP de Béni Mes-sous oscille entre 266 et 21493 KWh avec une moyenne de 9818.4 KWh.

On déduit que la consommation d'énergie a une forme légèrement cyclique qui suit les variations des débits entrants, celle-ci est tout simplement due aux variations saisonnières.

Variation de temps aération (Sommes des heures)

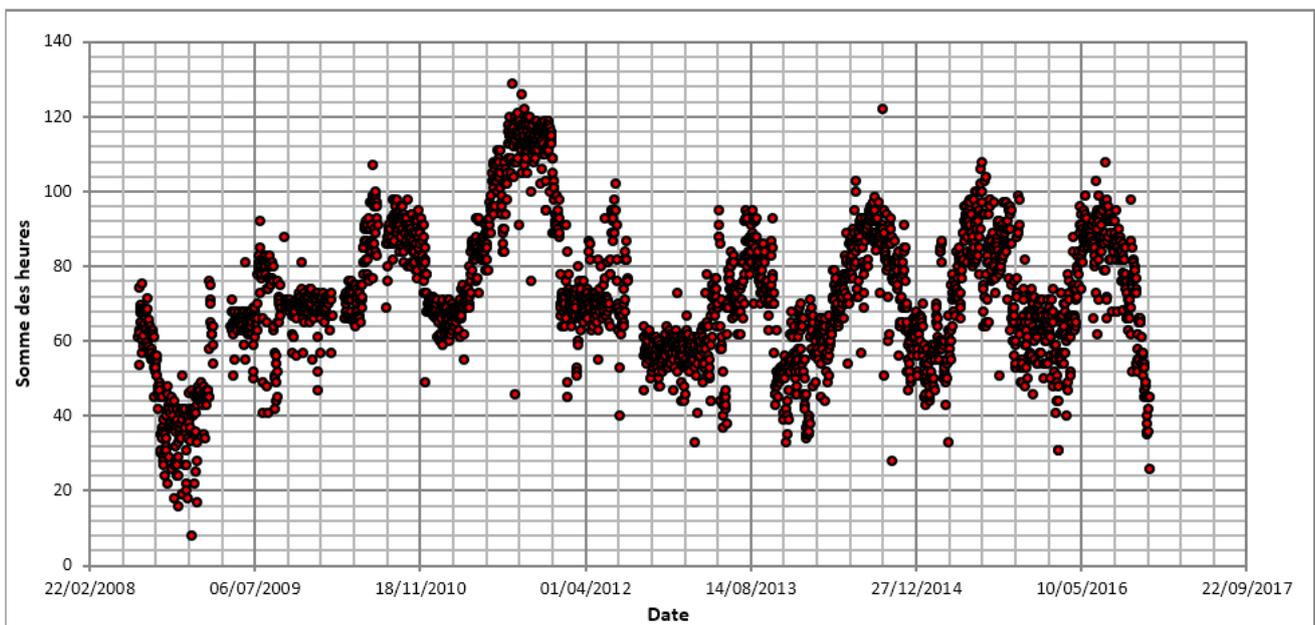


FIGURE 3.32 – Evolution journalière des sommes des heures (aération) dans la STEP

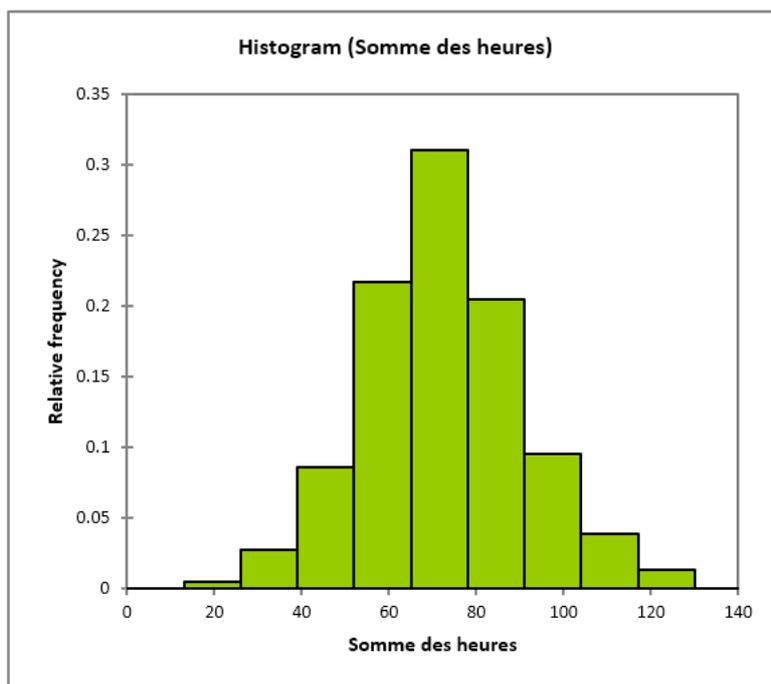


FIGURE 3.33 – Evolution journalière de l'énergie dans la STEP

La somme des heures d'aération (6 aérateurs) durant le processus d'épuration des eaux dans le bassin d'aération oscille entre 20 h et 120h avec une moyenne de 71,94 h.

Matrice des graphiques

La matrice des graphiques est une matrice de diagrammes de dispersion dont le but est de repérer des relations graphiques entre les différents paramètres observés au niveau de la STEP. Les paramètres utilisés sont : le débit entrant (QE), somme des heures par rapport au débit entrant (TH/QE), les abattements des MES, DBO5, DCO, et Global (ABMES, ABDBO5, ABDCO, ABG) et les rendements des mêmes paramètres (RMES, RDBO5, RDCO, RG). La matrice de diagramme de dispersion est symétrique.

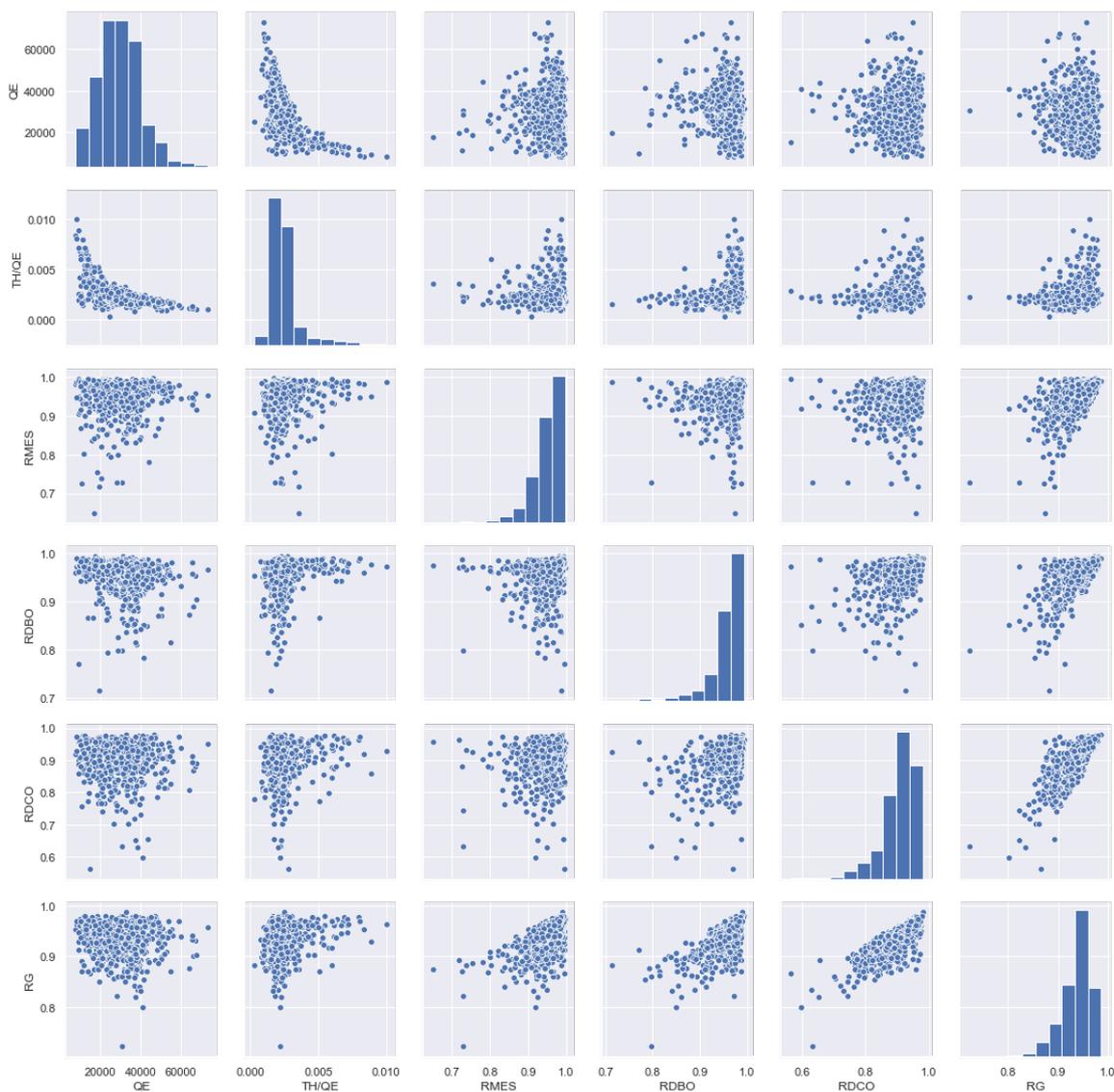


FIGURE 3.34 – Matrice des graphiques de rendement

On remarque sur ce diagramme des tendances logarithmiques entre les différents paramètres de rendement (MES, DCO, DBO5 et RG) en fonction de TH/QE ce qui nous a permis de justifier l'utilisation de ces paramètres en modélisation.

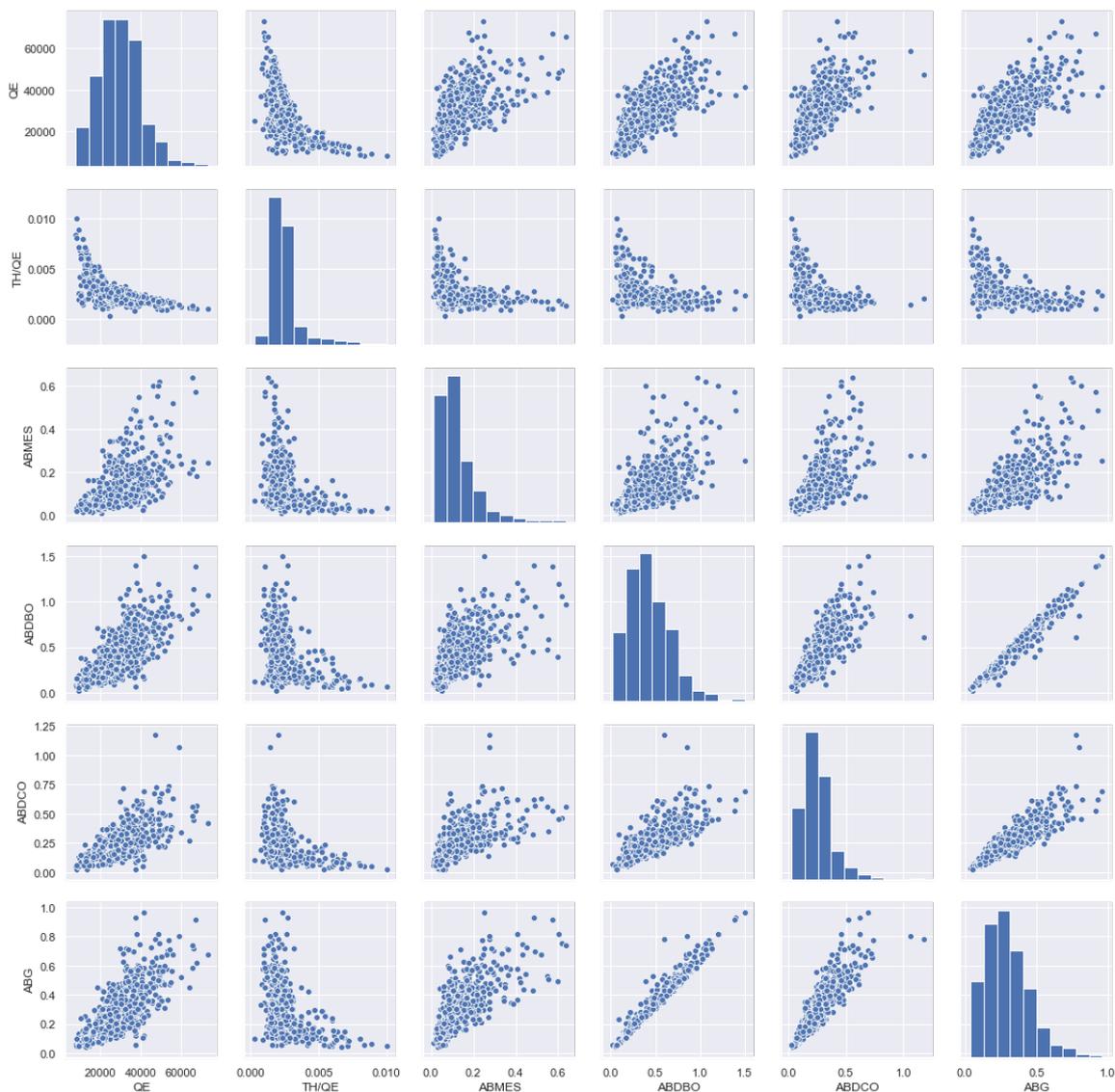


FIGURE 3.35 – Matrice des graphiques de l’abattement

On remarque aussi une tendance hyperbolique entre les paramètres de l’abattement des MES, DBO5, DCO et globale en fonction de TH/QE.

3.3.3 Conclusion

Dans ce chapitre, nous avons présenté la zone d’étude, ainsi qu’une discussion sur l’évolution des paramètres mesurés pour la STEP. La modélisation sera abordée au chapitre suivant et visera l’économie de la consommation de l’énergie pour la STEP.

Chapitre 4

Résultats et Discussion

4.1 Introduction

Dans ce chapitre, le but étant l'optimisation de l'énergie consommée par la STEP et plus particulièrement celle des aérateurs qui représente 45 à 70% de l'énergie électrique totale consommée. Pour cela, la démarche adoptée est la suivante : on a appliqué en premier lieu, l'ACP comme un outil statistique exploratoire pour connaître les paramètres qui influencent le processus à modéliser, ensuite plusieurs filtres ont été utilisés dans l'objectif d'obtenir les meilleures expériences observées au niveau de la STEP, la méthode utilisant le Clustering par K means, a été adoptée sur les différents paramètres pour repérer les groupements nécessaires à retenir pour la modélisation. Les résultats obtenus seront utilisés dans un modèle de Gradient Boosting, en optimisant les paramètres de ce dernier afin d'obtenir le modèle optimal pour la modélisation de la consommation de l'énergie électrique dans la STEP de Béni Messous.

4.2 Application de l'ACP :

L'ACP est un outil de statistique exploratoire qui ne permet pas généralement de tester des hypothèses. Cette technique présente l'avantage d'être exécutée, à différentes reprises, en éliminant ou rajoutant des observations ou des variables, à condition que ces manipulations soient justifiées dans les interprétations, dans notre cas on l'a appliqué aux données brute et épurée journalières de la série temporelle de la STEP de Béni Messous.

4.2.1 Données utilisées : STEP Beni Messous

On dispose d'une série de données journalières constituée de 3063 observations qui s'étale du 01/07/2008 au 18/11/2016. On dispose ainsi des mesures des différents paramètres physico chimiques de la STEP à l'entrée (brute) et à la sortie (épurée), ainsi que des débits d'entrée et de sortie, et de la somme des heures de fonctionnement des aérateurs, durant le processus d'épuration.

Il est à noter qu'il y a un manque de données remarquables de l'azote et du phosphate qui affectent l'interprétation des autres variables, par conséquent, ces deux paramètres ont été écartés de notre série d'étude.

4.2.2 Application de l'ACP sur les paramètres d'entrée (Eau brute) :

Une analyse en composantes principales (ACP) a été effectuée sur le tableau des données journalières, qui contient 8 variables (MES, DBO5, DCO, Total Hour (Somme des heures), T °C, Conductivité, QE, pH) de 701 observations de l'eau brute, après avoir éliminé toutes les données non concomitantes.

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

TABLE 4.1 – Matrice de corrélation entre les CP' et les variables de l'eau brute journalière

| | F1 | F2 |
|--------------|-------|--------|
| Total Hour | 0.481 | 0.575 |
| QE | 0.342 | 0.746 |
| pH | 0.063 | 0.125 |
| Conductivité | 0.105 | 0.056 |
| T °C | 0.335 | 0.667 |
| MES | 0.755 | -0.302 |
| DBO5 | 0.833 | -0.280 |
| DCO | 0.805 | -0.382 |

• Cercle de corrélations

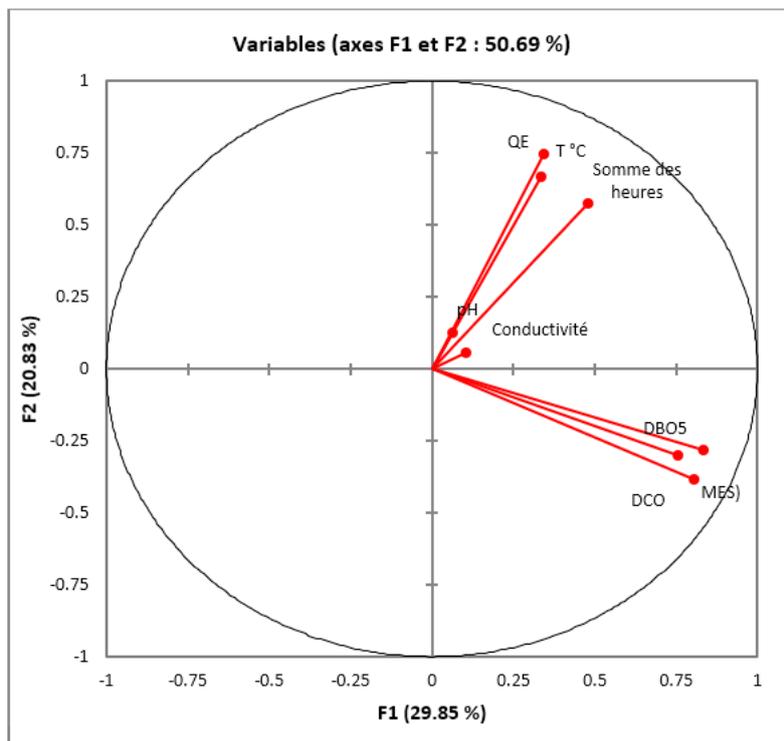


FIGURE 4.1 – Paramètres d’entrée et le temps de fonctionnement des aérateurs à l’échelle journalière [1x2]

En analysant la matrice de corrélation (entre les deux composantes principales) et aussi le cercle de corrélation (formé par les axes principaux (1 et 2)) ainsi que les variables centrées réduites des variables brutes, il s’avère que :

- Les variables DCO, DBO5, MES brutes sont bien corrélées avec la première composante et forment un groupe commun caractérisant la pollution organique, celui-ci est opposé au groupe formé par les variables QE, T°C et Somme des heures qui ont aussi une bonne corrélation avec la deuxième composante. En comparant les variables des 2 groupes, on remarque qu’elles sont orthogonales les unes par rapport aux autres, ce qui signifie leur non-corrélation (R proche de 0).
- Le pH et la conductivité ne reflètent pas une bonne information car ils sont proches de l’origine

4.2.3 Application de l’ACP sur les paramètres de l’eau (Entré/Sortie)

Une autre analyse en composantes principales (ACP) a été effectuée sur le tableau des données journalières, qui contient 11 variables (MESep, MESbr, DBO5br, DBO5ep, DCObr, DCOep, Total Hour (Somme des heures), Tep, Tbr, QE, QS) de 662 observations de l’eau brute et épurée, après avoir éliminé toutes les données non concomitantes.

Les résultats de cette analyse sont présentés ci-dessous :

TABLE 4.2 – Matrice de corrélation entre les CP' et les variables de l'eau brute et épurée journalière.

| | F1 | F2 |
|------------|--------|---------|
| QE | 0.7795 | -0.3735 |
| QS | 0.7807 | -0.3325 |
| Tbr | 0.7104 | -0.3082 |
| MESbr | 0.3633 | 0.7191 |
| DBO5br | 0.4462 | 0.7452 |
| DCObr | 0.3623 | 0.7903 |
| Tep | 0.7334 | -0.2380 |
| MESep | 0.1294 | 0.3439 |
| DBO5ep | 0.4725 | 0.0532 |
| DCOep | 0.2455 | 0.0968 |
| Total Hour | 0.5708 | -0.0515 |

• Cercle de corrélations

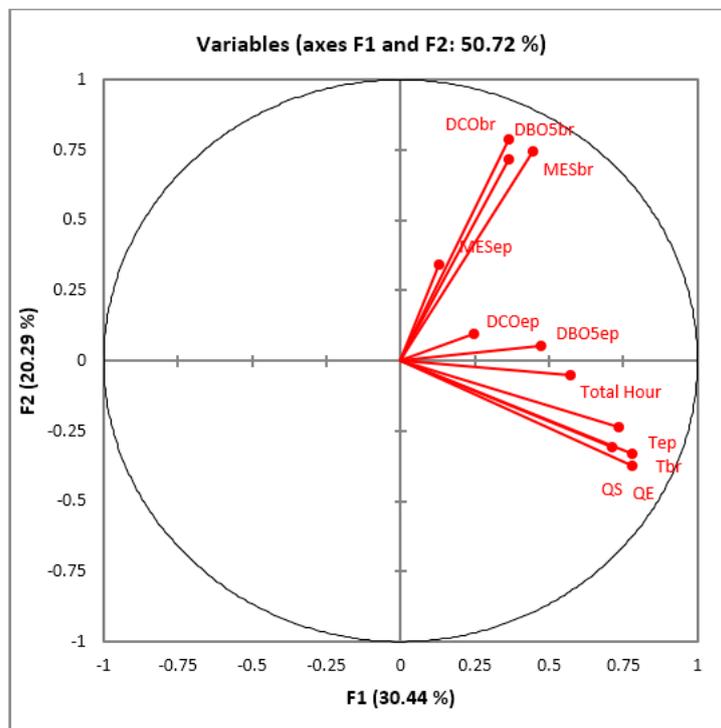


FIGURE 4.2 – Paramètres (Entrée-Sortie) et le temps de fonctionnement des aérateurs à l'échelle journalière [1×2]

En examinant ce cercle de corrélation, on peut dire que les variables centrées réduites QE, QS, Tep, Tbr sont bien corrélées avec la première composante et forment un groupe commun, de même, que la somme des heures (Total Hour) qui est proche du premier axe avec une corrélation satisfaisante.

Les variables MESbr, DBO5br et DCObr sont bien corrélées avec la deuxième composante, par contre MESep, DBO5ep et DCOep ne sont corrélées ni avec la première, ni avec la deuxième composante, elles caractérisent la pollution organique. On remarque que les paramètres de pollution brute forment un groupe commun, opposé au groupe formé par les paramètres épurés qui sont proches de l'origine et ne présentent aucune corrélation entre eux. On en déduit que le processus concernant l'abattement de la pollution organique n'est pas bien maîtrisé.

4.3 Clustering K means :

Dans cette partie, la compatibilité des données de la STEP de Beni Messous sera discutée par l'utilisation du clustering avec K means en Data Mining lors de l'exploration de données pour déceler des individus similaires et des relations éventuelles dans les données pour les différentes variables mesurées et calculées.

Le but de l'utilisation de l'algorithme de classification non supervisé (K-means) est d'identifier un pattern au sein des données, et de regrouper les individus ayant des caractéristiques similaires, dans ce cas on s'intéresse à distinguer les meilleures données qui appartiennent à une classe (ou des classes) optimales pour la modélisation.

L'inconvénient de cette méthode est qu'elle ne permet pas de découvrir le nombre optimal de classes. Le résultat dépend du tirage initial avec une sensibilité de l'algorithme aux valeurs extrêmes (en particulier lors du calcul du barycentre), pour y remédier, on adopte la méthode du coude (Elbow Method). L'idée de cette méthode est d'exécuter K-means Clustering sur l'ensemble de données pour une gamme de valeurs de K (K variant de 2 à 10 par exemple), et pour chaque valeur de K, calculer la variance intra-classe

4.3.1 Application de Kmeans sur nos données :

Pour concevoir un modèle performant d'optimisation, il faut connaître les différents paramètres ayant une influence sur la variation de la consommation de l'énergie électrique. Pour cela nous allons tester et choisir les données les plus pertinentes en appliquant le Clustering par K means.

Ce modèle sera établi à partir des expériences jugées bonnes en matière de fonctionnement de la STEP.

Notre série de données journalière composée de 739 observations et 27 variables (mesurées et calculées), les nouvelles variables calculées sont une combinaison des variables mesurées comme l'abattement, le rendement . . .

Les données utilisées doivent être normalisées avant l'analyse de K means. La normalisation est le processus de transformation des valeurs observées dans un intervalle standard de valeurs compris entre [0, 1].

Les paramètres qui entrent dans l'analyse K means sont :

- L'abattement de DBO, DCO, MES et l'abattement global.
- Le rendement de DBO, DCO, MES et le rendement global.
- La somme des heures de fonctionnement des aérateurs par rapport au débit entrant (TH/QE).
- WQI (Water Quality Index)

Avant l'analyse K means, il faut appliquer la méthode du coude (Elbow method) pour déterminer le nombre de clusters k optimal. L'analyse silhouette permettra d'avoir une représentation graphique succincte de la qualité de classification de chaque variable et de confirmer le nombre de k utilisé.

Le test de Hopkins a été effectué sur l'ensemble des données utilisées, chaque fois avant l'analyse de K means, les valeurs obtenues sont proches de 1 (>0.9), ce qui permet de dire que les données sont fortement groupées.

L'idée d'ajouter des variables brutes (DBO, DCO et MES) à l'analyse K means est justifiée par l'influence de ces dernières sur le comportement des classes trouvées, car de faibles valeurs de ces variables peuvent donner une information biaisée. De faibles concentrations de la DBO peuvent donner des rendements relativement faibles justifiant une consommation énergétique réduite.

CAS 1 : DBO

Dans cette étape, les variables utilisées sont : la concentration entrante de DBO5 et son rendement epuratoire, et la somme des heures de fonctionnement des aerateurs par le debit entrant TH/QE.

La methode du coude et l'analyse silhouette indiquent visuellement que le nombre de Clusters optimal est egale à 2 ($k=2$), de plus, le test de Hopkins est egale à 0.923, ce qui indique que les données sont fortement groupées.

La figure suivante montre les différentes relations entre les différentes variables utilisées :

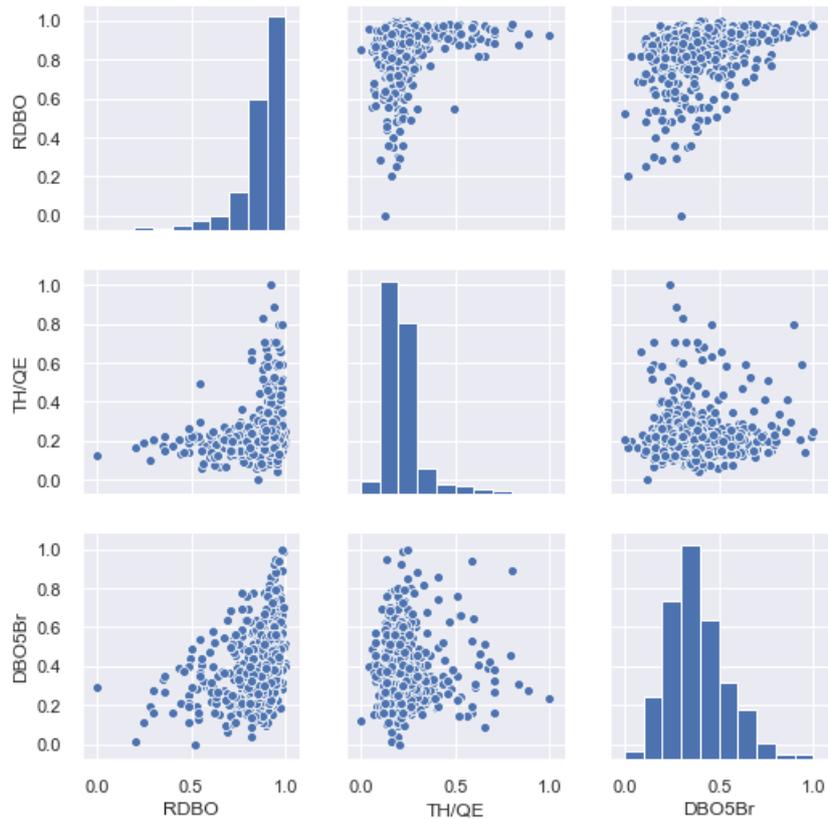


FIGURE 4.3 – Matrice des graphiques du Cas 1

On remarque une tendance logarithmique entre le rendement de la DBO5 avec TH/QE et DBO5Br.

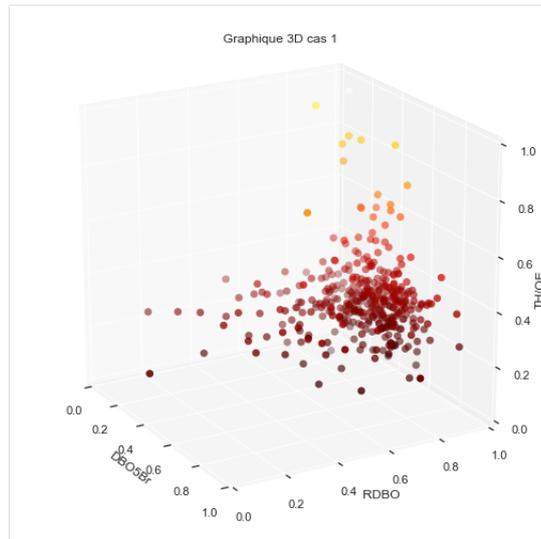
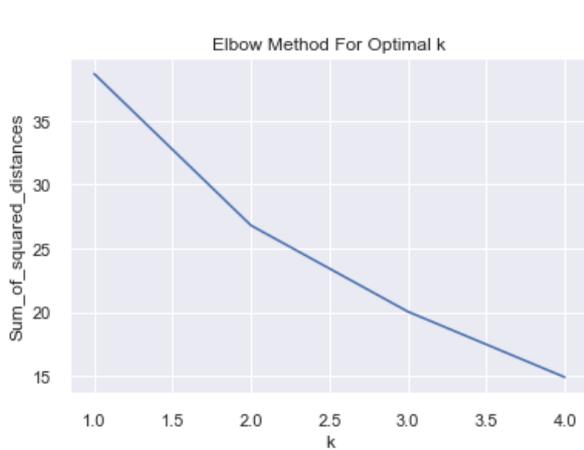
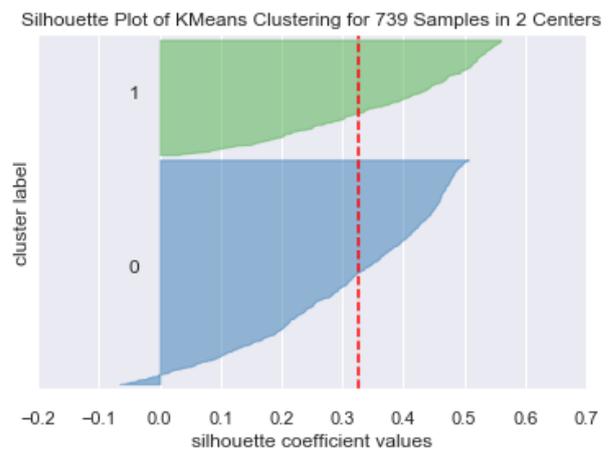


FIGURE 4.4 – Représentation en 3D du Cas 1

La représentation de la méthode de coude et L'analyse de la silhouette est présenté dans les figures suivantes :



(a) la methode de coude pour k optimal cas 1



(b) Représentation de l'analyse silhouette cas 1

FIGURE 4.5 – application de la méthode du coude et l'analyse silhouette sur le cas 1

L'analyse K means pour la 1er cas :

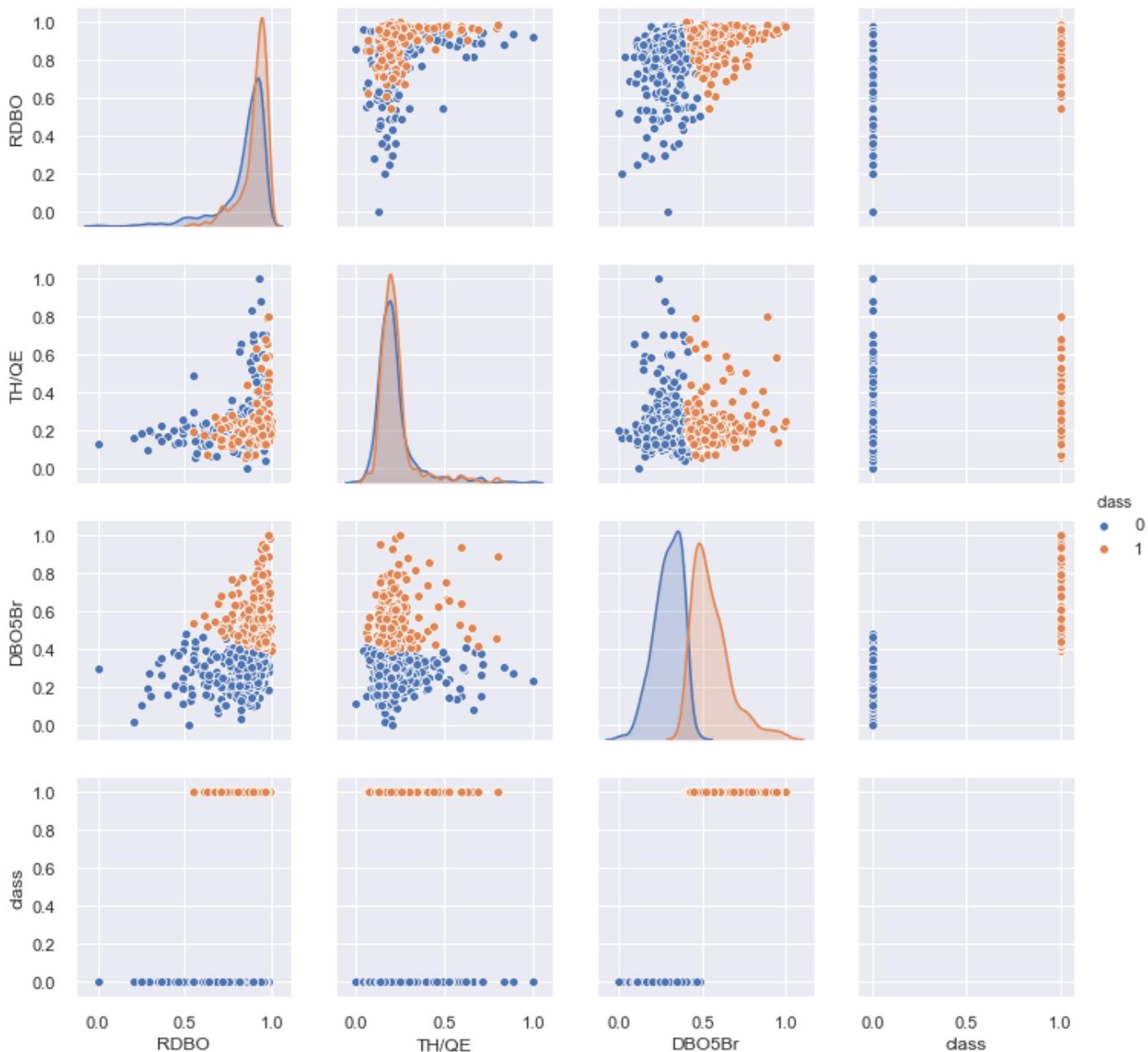


FIGURE 4.6 – Matrice des graphiques en appliquant l’analyse K means du cas 1

La figure ci-dessus met en relief l’existence de deux classes séparées par rapport à la concentration de DBO5Br. On remarque une tendance logarithmique du rendement par rapport à TH/QE pour la classe 1, et une grande variation pour la classe 0.

La classe 1 (en orange) contenant 249 observations, représentant les grandes valeurs de concentration de DBO5Br, justifiant l’énergie consommée, celle-ci sera utilisée dans la modélisation, alors que la classe 0 représentant les faibles valeurs de DBO5Br pour la même consommation énergétique sera rejetée .

CAS2 : DCO

Dans cette étape, les variables utilisées sont : la concentration entrante de DCO et son rendement epuratoire RDCO, et la somme des heures de fonctionnement des aerateurs par le debit entrant TH/QE.

La methode du coude et l'analyse silhouette indiquent visuellement que le nombre de Clusters optimal est egale à 2 ($k=2$), de plus, le test de Hopkins est egale à 0.908, ce qui indique que les données sont fortement groupées.

La figure suivante montre les différentes relations entre les différentes variables utilisées :

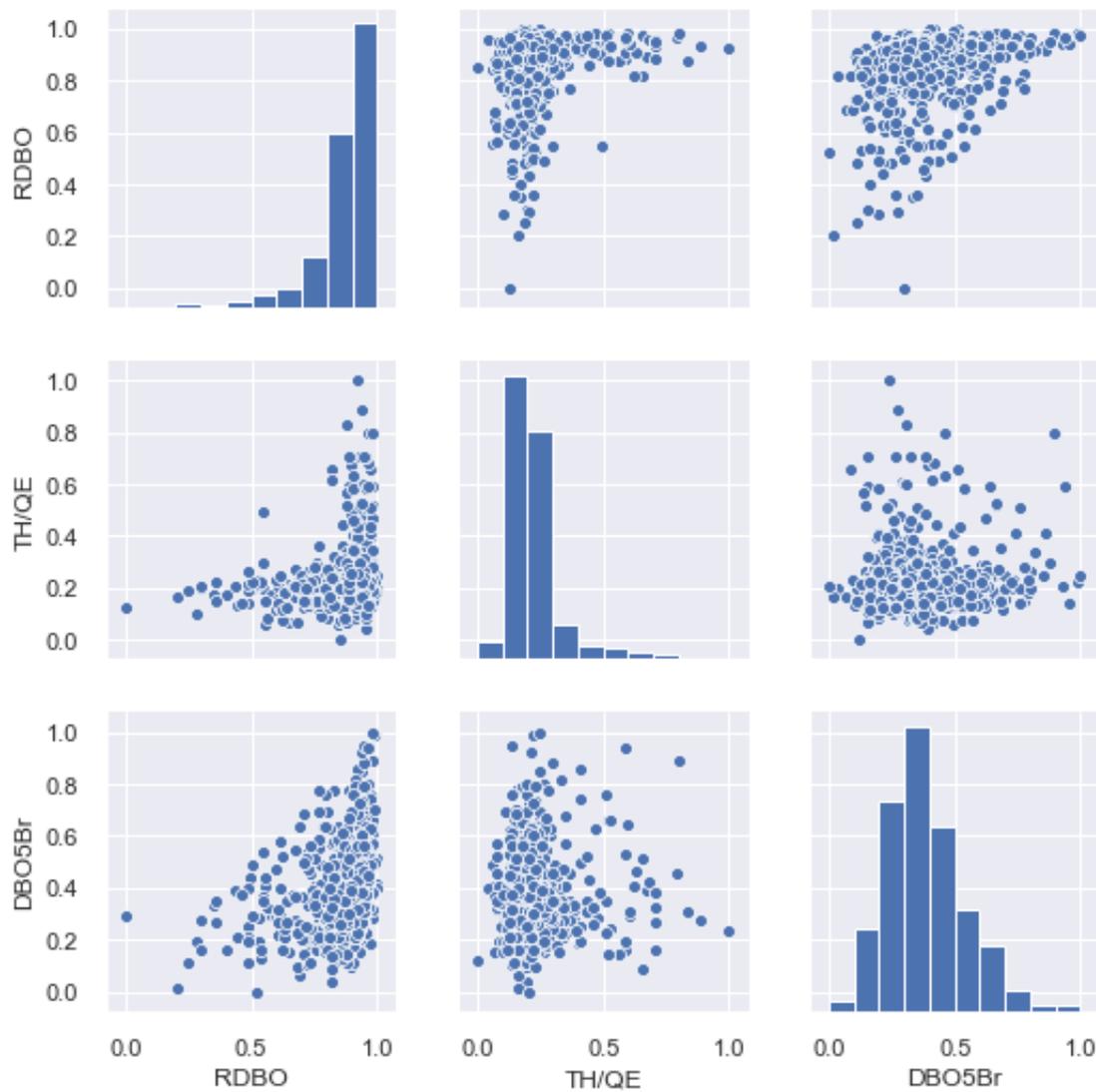


FIGURE 4.7 – Matrice des graphiques du Cas 2

On remarque une tendance logarithmique remarquable de RDCO avec DCOBr et TH/QE .

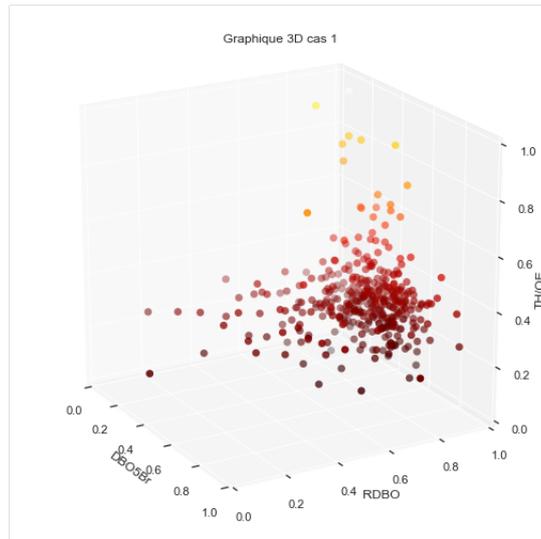
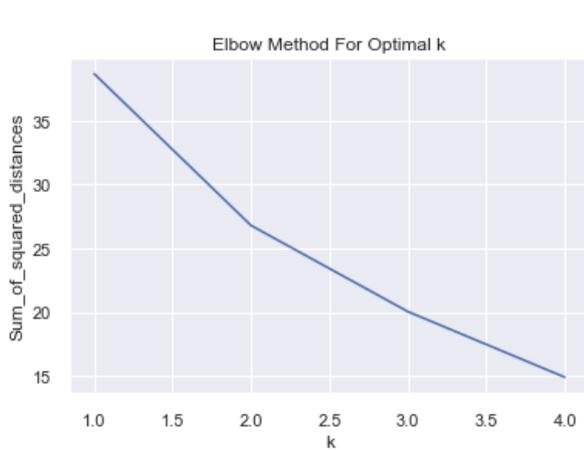
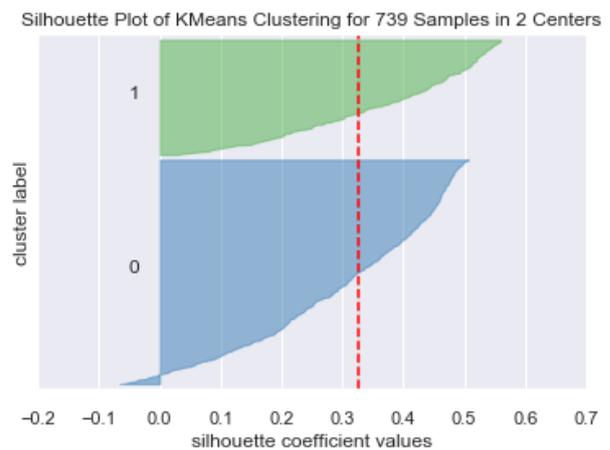


FIGURE 4.8 – Représentation en 3D du Cas 2

La représentation de la méthode de coude et L'analyse de la silhouette est exposée dans les figures suivantes :



(a) la methode de coude pour k optimal cas 2



(b) Représentation de l'analyse silhouette cas 2

FIGURE 4.9 – application de la méthode du coude et l'analyse silhouette sur le cas 2

L'analyse K means pour la 2eme cas :

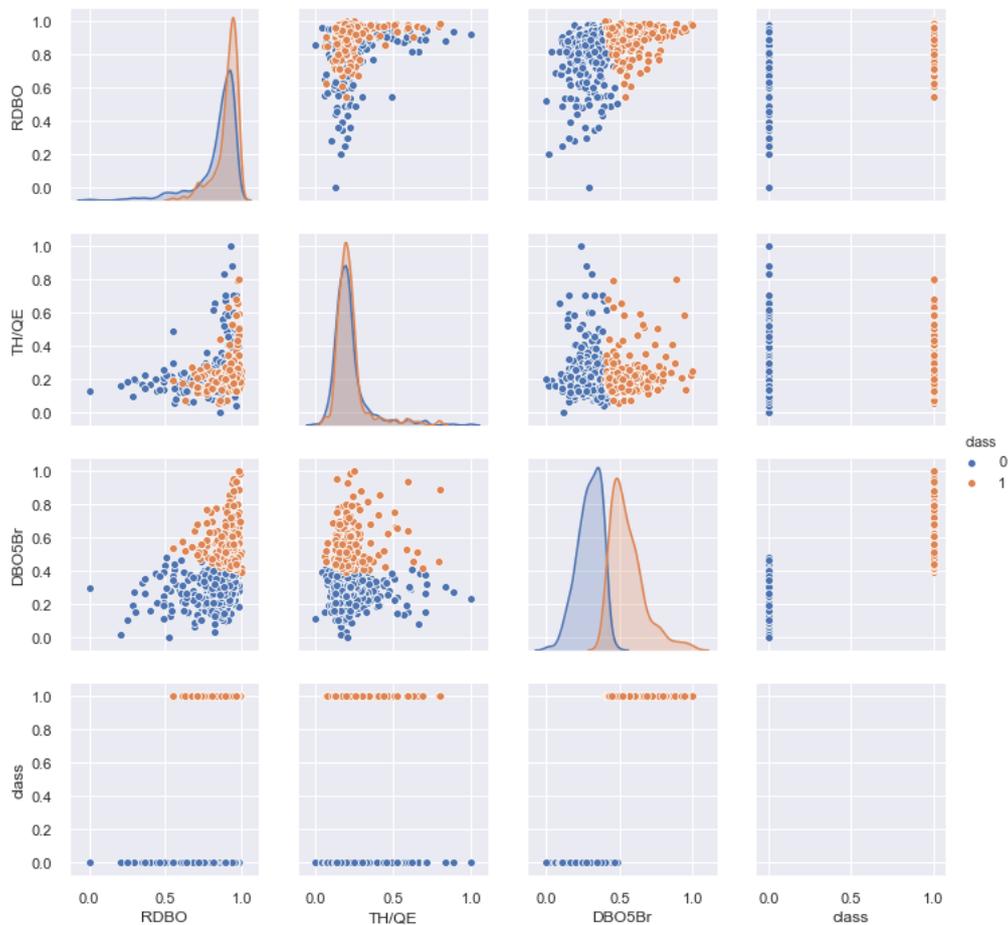


FIGURE 4.10 – Matrice des graphiques en appliquant l’analyse K means du cas 2

La figure ci-dessus met en relief l’existence de deux classes séparées par rapport au rendement.

La classe 0, représentant de GRANDES valeurs de RDCO (>90%), correspondant à une évolution variable des valeurs de TH/QE, celle-ci ne sera pas retenue dans la modélisation, ce qui justifie l’utilisation de la 1ere classe contenant les 215 observations pertinentes.

L’ajout des variables brutes à l’analyse K means, a une influence remarquable sur le regroupement des classes.

Cas 3 : MES

Dans cette etape, les variables utilisees sont : la concentration entrante de MES et son rendement epuratoire RMES, et la somme des heures de fonctionnement des aerateurs par le debit entrant TH/QE.

La methode du coude et l’analyse silhouette indiquent visuellement que le nombre de Clusters optimal est egale à 2 (k=2), de plus, le test de Hopkins est egale à 0.93, ce qui indique que les donnees sont fortement groupées.

La figure suivante montre les différentes relations entre les différentes variables utilisées :

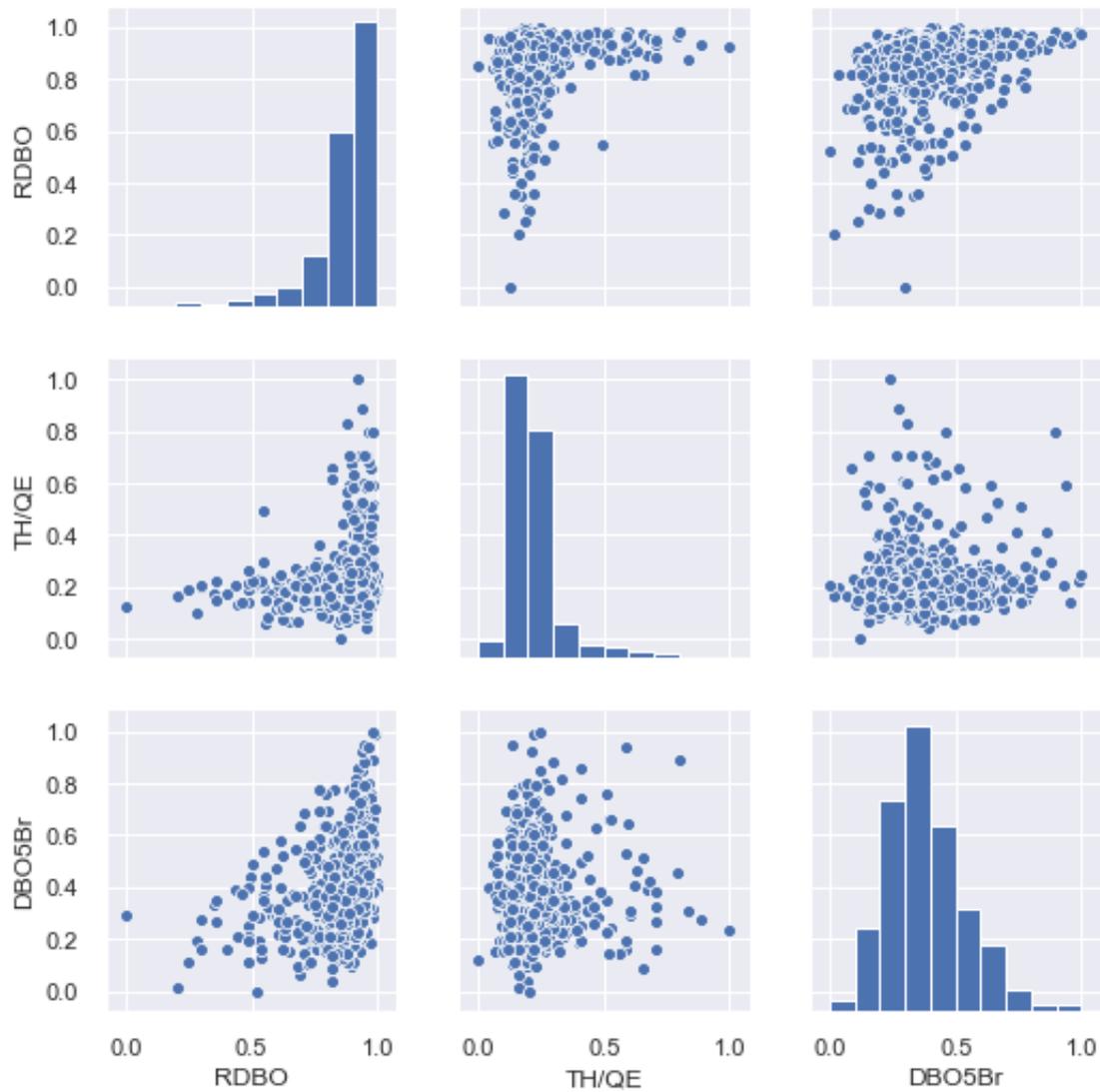


FIGURE 4.11 – Matrice des graphiques du Cas 3

On remarque une tendance logarithmique remarquable de RMES avec MESBr et TH/QE, aussi une tendance hyperbolique de TH/QE et MESBr.

La représentation des variables en 3D :

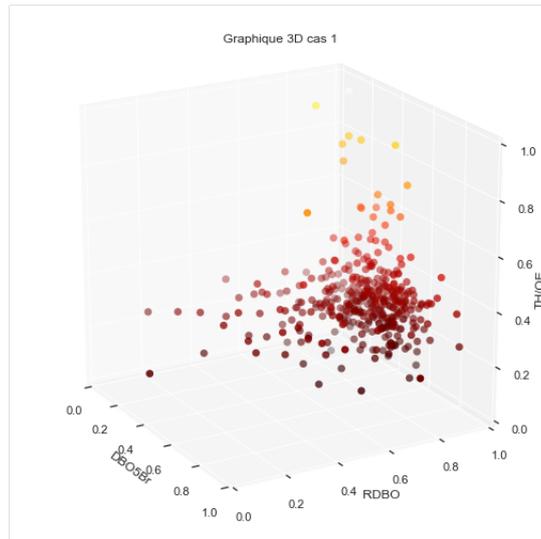
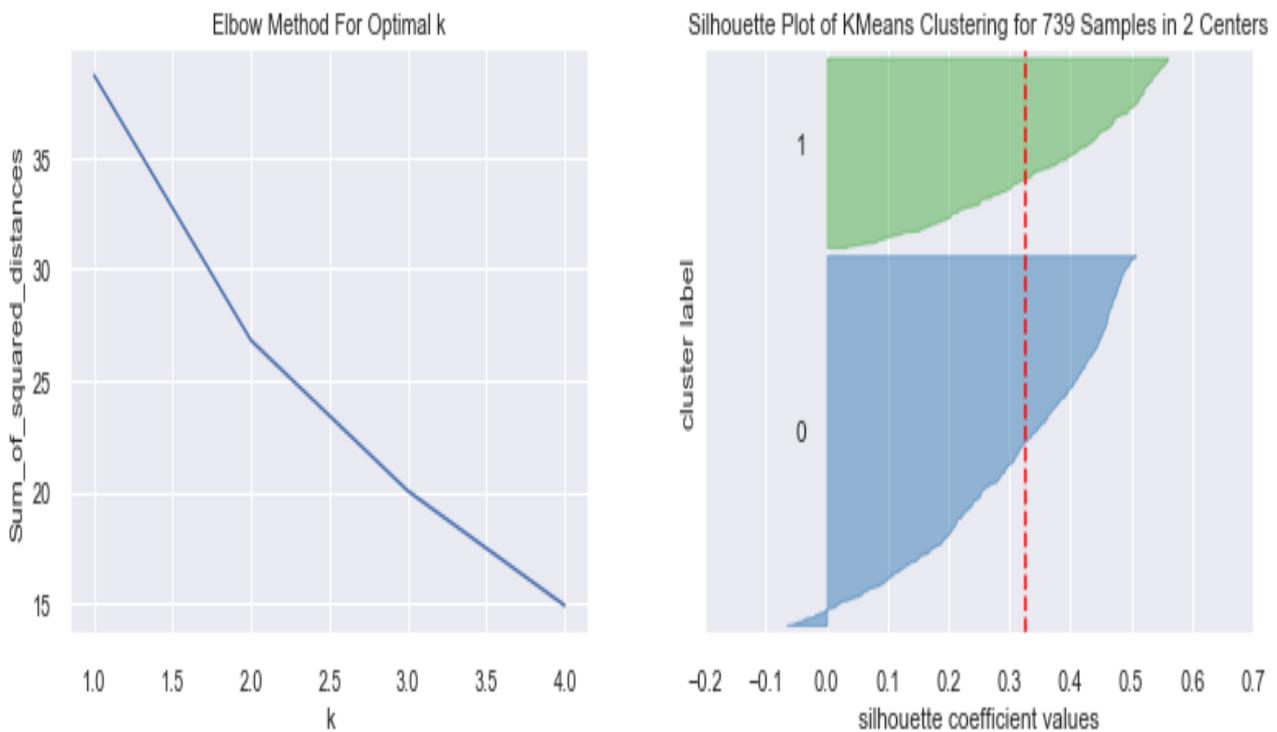


FIGURE 4.12 – Representation en 3D du Cas 3

La représentation de la méthode de coude et L'analyse de la silhouette est présenté dans les figures suivantes :



(a) la methode de coude pour k optimal cas 3

(b) Representaion de l'analyse silhouette cas 3

FIGURE 4.13 – application de la méthode du coude et l'analyse silhouette sur le cas 3

L'analyse K means pour le 3eme cas :

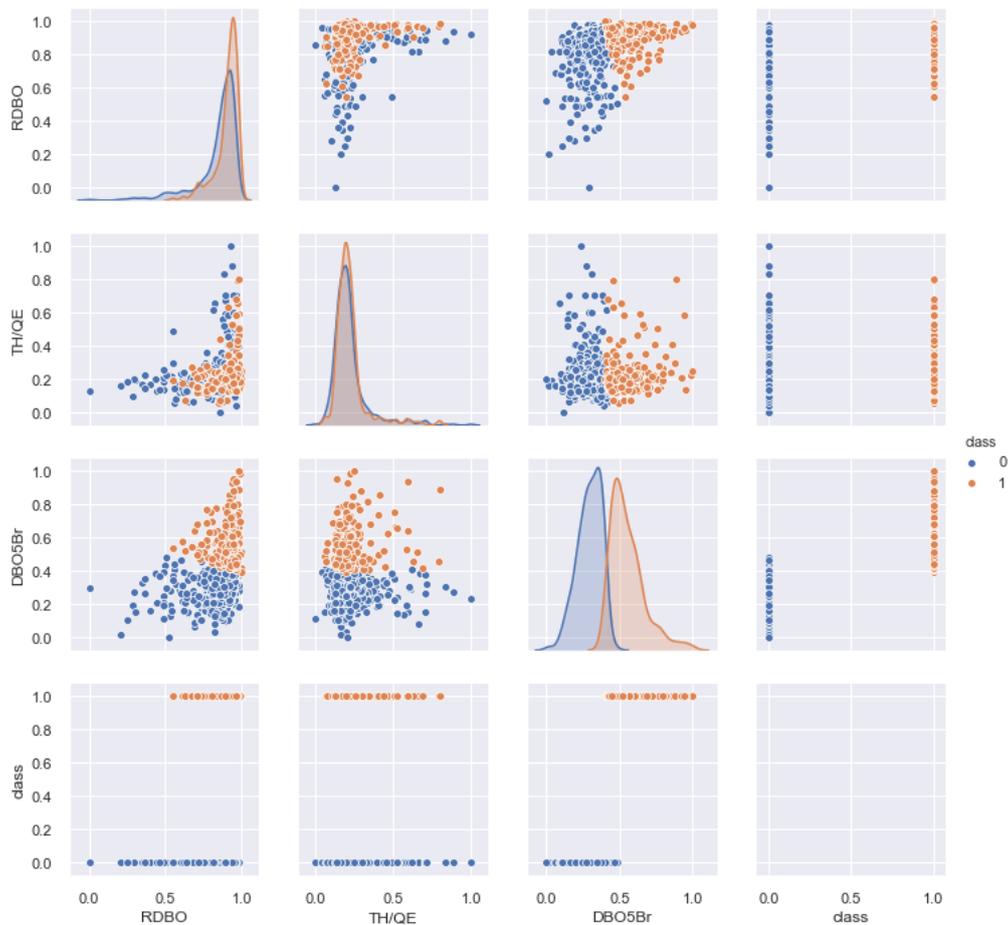


FIGURE 4.14 – Matrice des graphiques en appliquant l’analyse K means cas 3

La figure ci-dessus met en relief l’existence de deux classes séparées par rapport au MESBr.

La classe 0, représentant des valeurs faibles en concentration de MESBr correspondant à une évolution variable des valeurs de TH/QE, contrairement à la classe 1 contenant de grandes valeurs de concentrations de MESBr et un rendement élevé justifiant l’énergie consommée, ce qui justifie l’utilisation de la 1ere classe contenant les 154 observations pertinentes.

L’ajout des variables brutes à l’analyse K means, a une influence remarquable sur le regroupe-ment des classes.

Cas 4 : WQI ,RG

Dans cette étape, les variables utilisés sont : WQI (Water Quality Index), RG (Rendmeent Global) et TH/QE, la methode du coude avec l’analyse silhouette indiquent que le nombre de Clusters est égale a k=2 ou k=4. Le regroupe-ment des classes etant très difficile à interpreter pour k=4, le nombre de clusters choisi est k=2. La representation graphique indique la variabilité de chaque classe.

Le test de Hopkins est égale a 0.92, proche de 1 ce qui indique que les données sont fortement groupées.

La figure suivante montre les différentes relations entre les différentes variables utilisées :

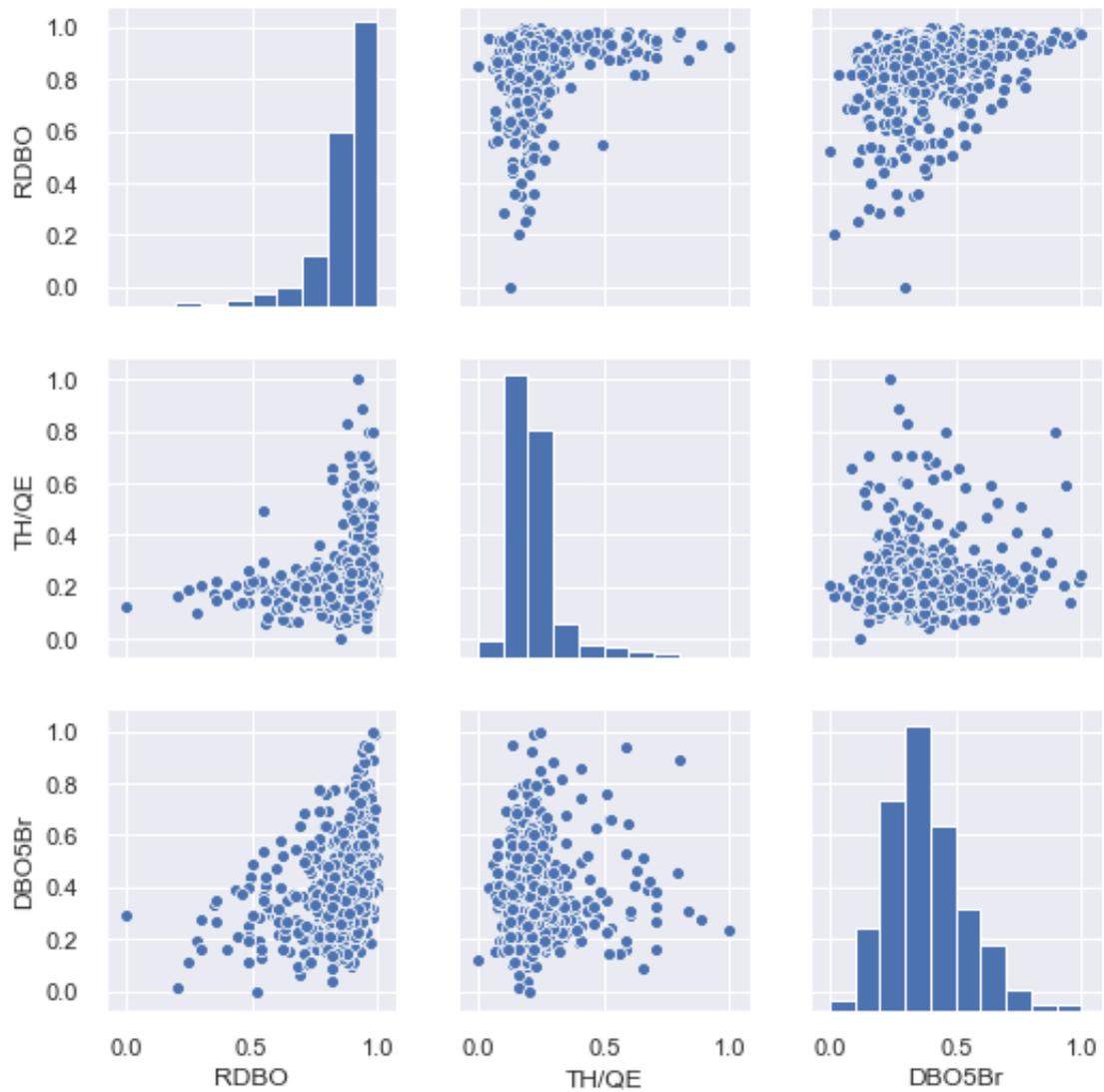


FIGURE 4.15 – Matrice des graphiques du Cas 4

On remarque une tendance logarithmique de RG avec TH/QE.
 La représentation des variables en 3D :

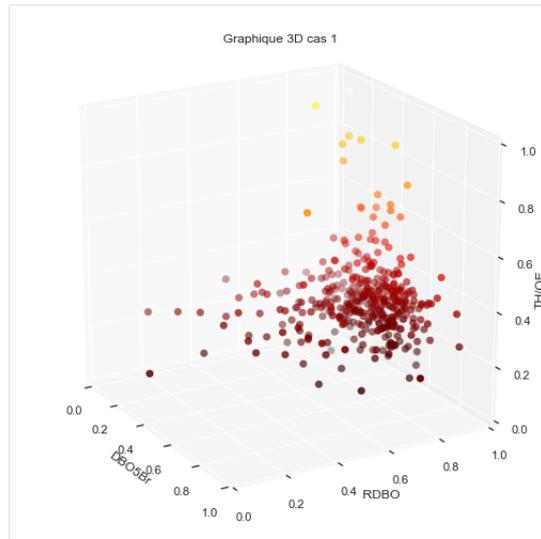
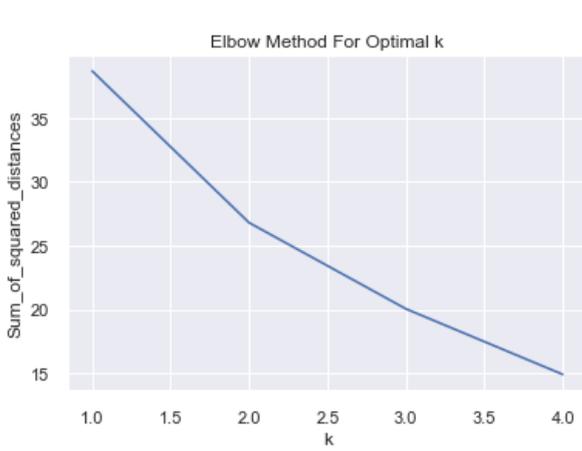
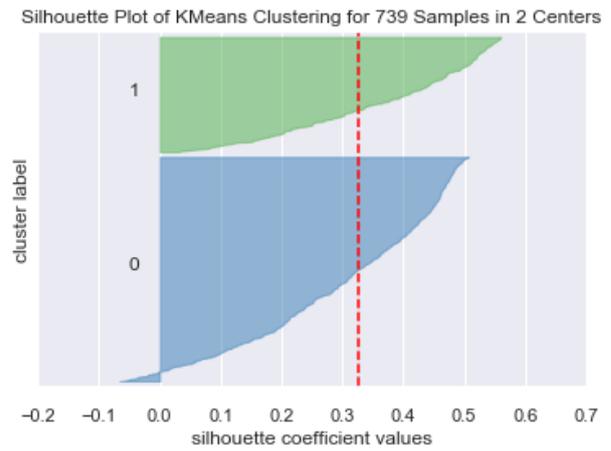


FIGURE 4.16 – Représentation en 3D du Cas 4



(a) la methode de coude pour k optimal cas 4



(b) Représentation de l'analyse silhouette cas 4

FIGURE 4.17 – application de la méthode du coude et l'analyse silhouette sur le cas 4

L'analyse K means pour le 4eme cas :

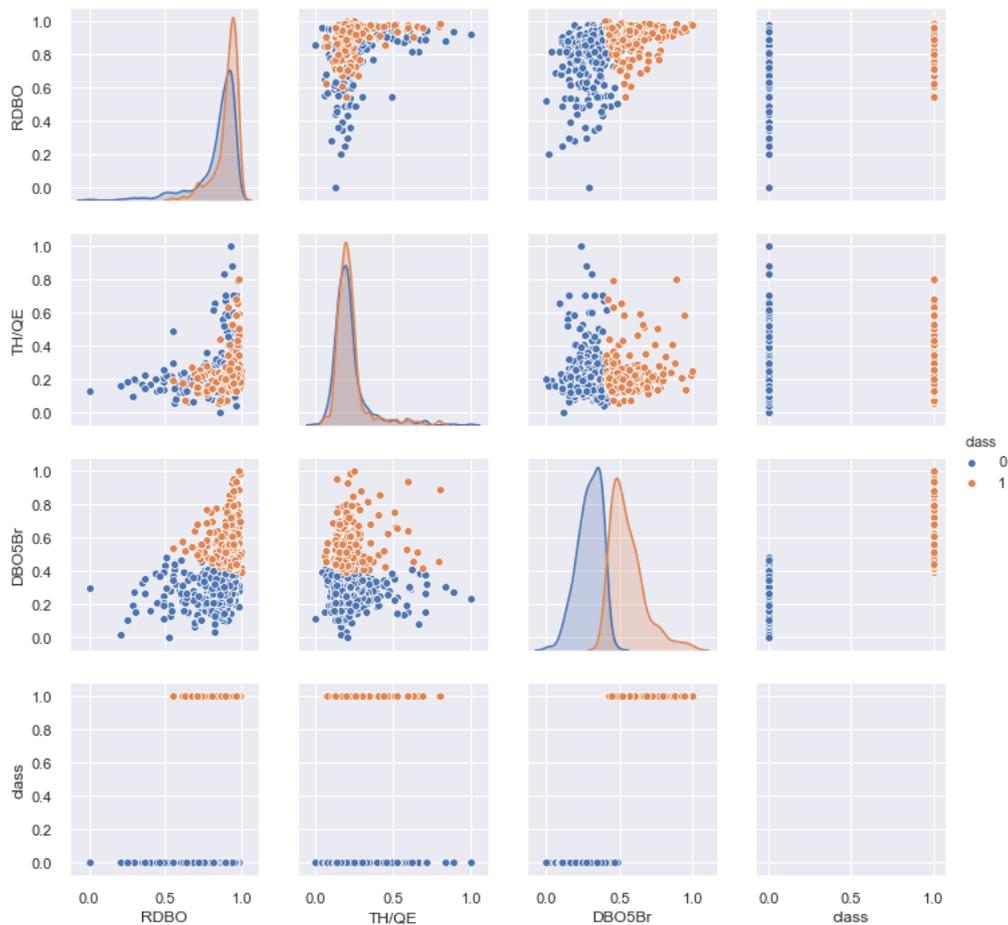


FIGURE 4.18 – Matrice des graphiques en appliquant l’analyse K means cas 4

La figure ci-dessus met en relief l’existence de deux classes séparées par rapport au WQI, ce qui confirme les résultats trouvés au 1er et 3eme cas. L’objectif est de choisir la classe qui contient de grandes valeurs de WQI (valeurs proche de 1) justifiant les valeurs de TH/QE.

La classe 1 (en orange) contenant 341 observations correspond à ces caractéristiques. L’ajout du WQI à l’analyse K means, a une influence remarquable sur le regroupement des classes.

4.3.2 Conclusion

L’analyse K means comme technique de Clustering à base geometrique pour la selection des donnees optimales par rapport a la consommation energetique, a donne des resultats significatifs concernant le regroupement des classes, pour chaque cas. Ceci permet de distinguer le comportement et les caracteristiques des données. En outre l’ajout d’un troisième paramètre à l’analyse K means, apporte une information utile pour l’interpretation.

4.4 Modélisation avec Gradient Boosting

A partir des résultats de l'analyse K means pour chaque cas, on procède à la conception du modèle. La première étape est de préparer les données pour que l'algorithme puisse mieux appréhender la structure de phénomène à modéliser.

La sélection des variables d'entrée du modèle est l'une des premières et importantes étapes de toute tâche d'apprentissage automatique.

On sélectionne les variables qui contribuent le plus à la consommation énergétique, les variables non pertinentes peuvent réduire la précision du modèle, et conduire vers un «sur-apprentissage» (Overfitting) du modèle.

Pour chaque cas étudié, le modèle aura pour paramètre d'entrée, un paramètre de pollution (DBO5Br, ou DCOBr, ou MESBr, ou WQI) son rendement respectif et le débit entrant (QE) et comme paramètre de sortie TH/QE.

La série de données a été décomposée en deux groupes, le premier contenant 80% est destiné à l'apprentissage du modèle, et 20% pour la validation des résultats du modèle.

Le Gradient Boosting est l'une des techniques les plus puissantes d'apprentissage automatique appliquée pour la construction des modèles prédictifs.

4.4.1 Modélisation

L'ensemble des 739 observations des différents paramètres retenus précédemment seront utilisés pour la modélisation.

Pour juger de la performance du modèle développé, le critère numérique retenu est le R2 couplé au critère graphique Q – Q Plot, que ce soit lors de l'apprentissage ou lors de la validation.

On a commencé par tester différentes combinaisons de valeurs d'hyper paramètre du modèle, pour les différents cas décrits précédemment. On en déduit que le nombre d'estimateurs est idéal pour l'intervalle [6-12], avec une profondeur maximal, de 4, et une valeur du coefficient d'apprentissage égale à 0.4 pour le bon fonctionnement de modèle développé, et si on augmente les valeurs des paramètres de profondeur maximal (> 4), et un nombre d'estimateurs élevé (> 13), on observera un sur apprentissage « overfitting » du modèle.

Cas 1 : DBO

Après plusieurs essais on a abouti aux résultats optimums suivants :

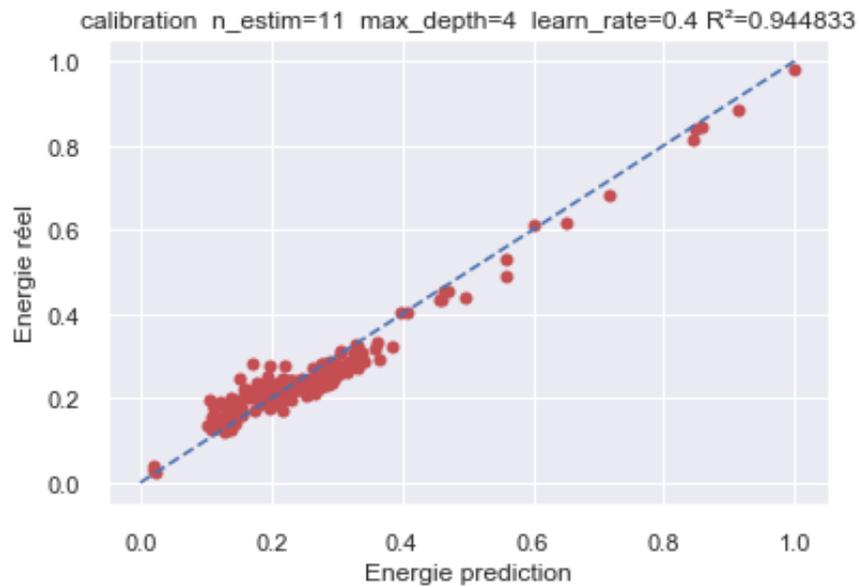


FIGURE 4.19 – Calibration cas 1

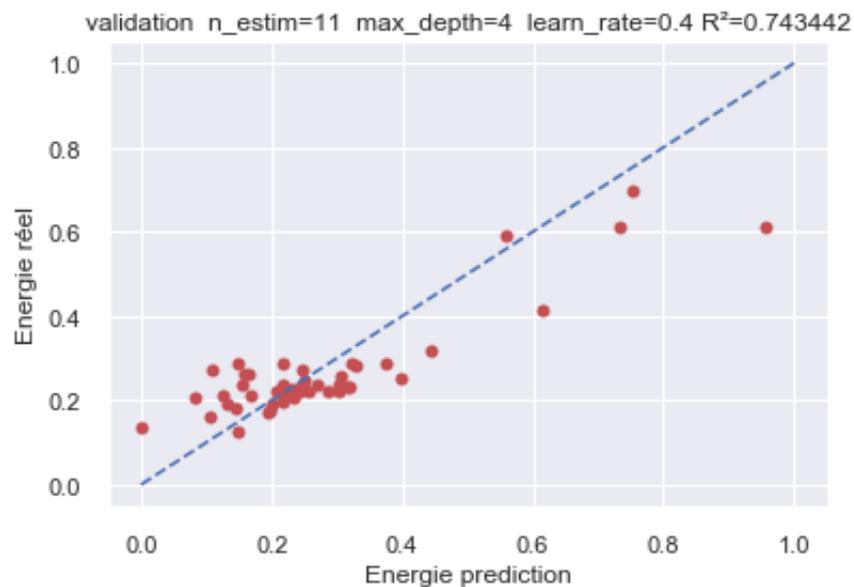


FIGURE 4.20 – validation cas 1

Le critère d'évaluation numérique retenu donne un très bon résultat pour la calibration ($R^2 = 0.94$), et un bon résultat pour la validation ($R^2=0.74$). On a remarqué que même si on élimine les valeurs aberrantes, le coefficient de détermination R^2 reste influencé par les valeurs max.

Cas 2 : DCO

Après plusieurs essais on a abouti aux résultats optimums suivants :

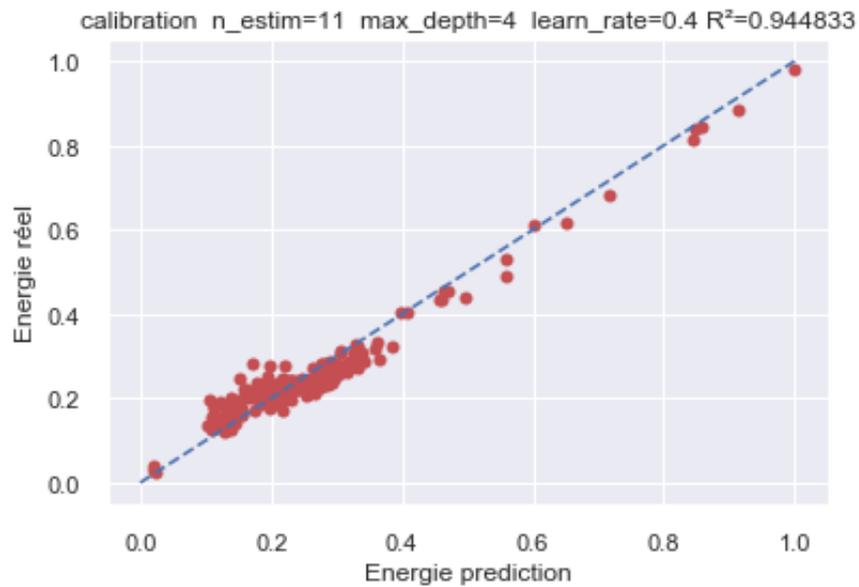


FIGURE 4.21 – Calibration cas 2

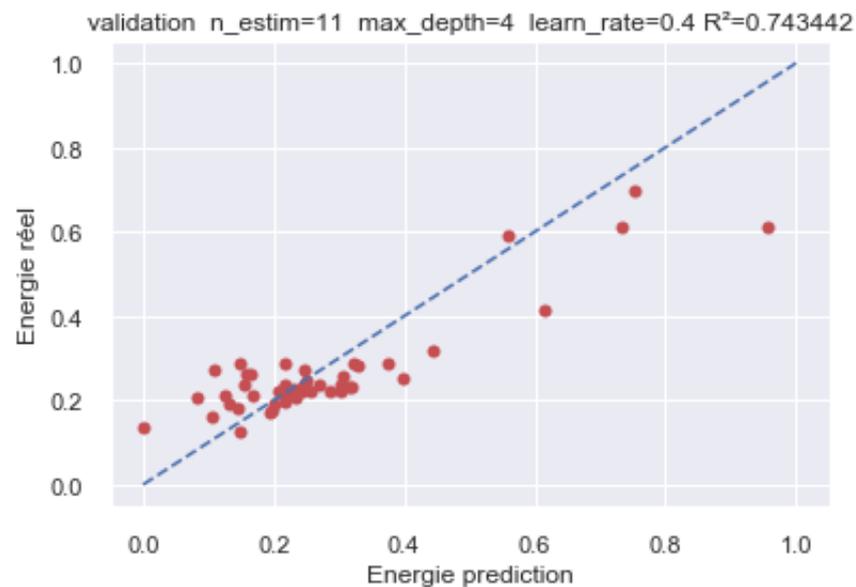


FIGURE 4.22 – Validation cas 2

Le critère d'évaluation numérique retenu donne un bon résultat que ce soit pour la calibration ($R^2 = 0.78$), ou pour la validation ($R^2=0.60$). La même remarque a été faite concernant l'élimination des valeurs aberrantes.

Cas 3 : MES

Après plusieurs essais on a abouti aux résultats optimums suivants :

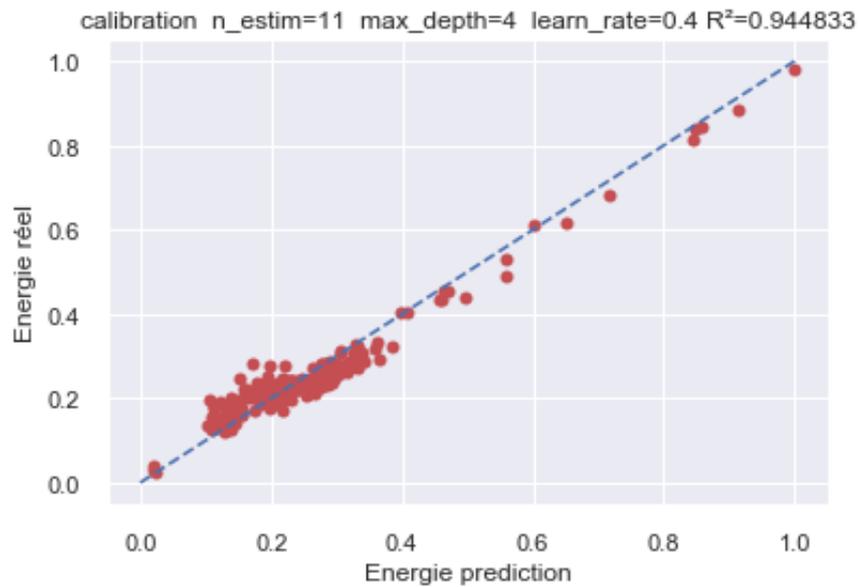


FIGURE 4.23 – Calibration cas 3

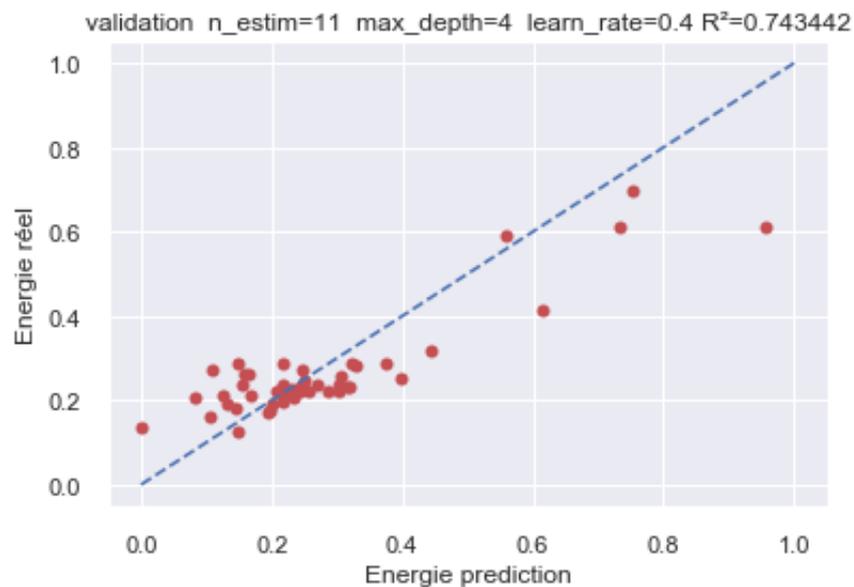


FIGURE 4.24 – Validation cas 3

Le critère d'évaluation numérique retenu donne un très bon résultat pour la calibration ($R^2 = 0.93$), et un bon résultat pour la validation ($R^2=0.70$).

La même remarque a été faite concernant l'élimination des valeurs aberrantes.

Cas 4 : WQI

Après plusieurs essais on a abouti aux résultats optimums suivants :

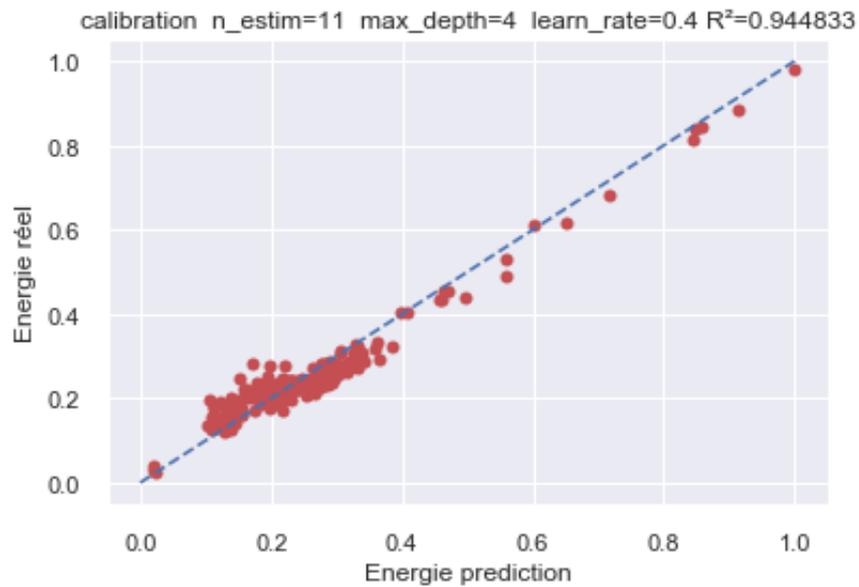


FIGURE 4.25 – Calibration cas 4

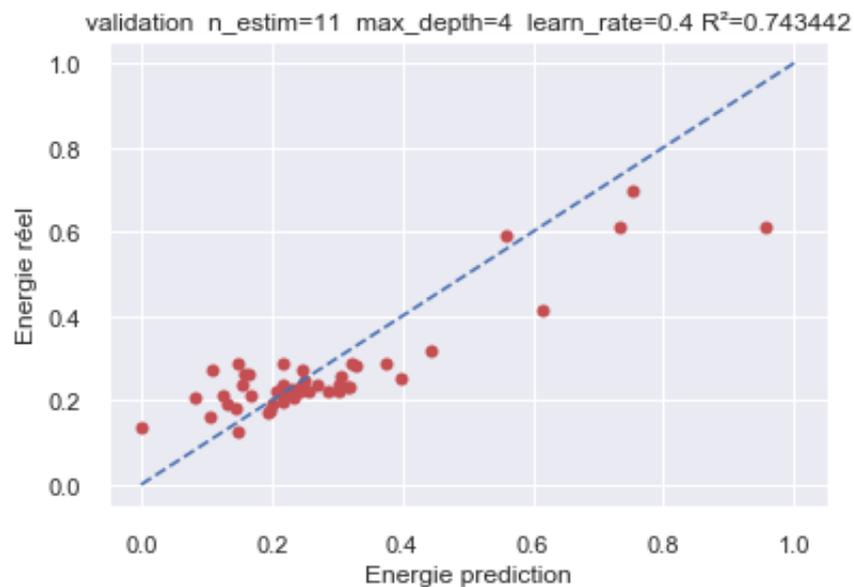


FIGURE 4.26 – Validation cas 4

Le critère d'évaluation numérique retenu donne un très bon résultat que ce soit pour la calibration ($R^2 = 0.90$), ou pour la validation ($R^2=0.84$).

La même remarque a été faite concernant l'élimination des valeurs aberrantes.

Les résultats obtenus dans ce cas sont très pertinents, en effet une variation minimale a été observée entre les valeurs de R^2 en validation, et en calibration.

TABLE 4.3 – Résultats de modèle pour les différents cas étudiés

| | <i>R2 calibration</i> | <i>R2 validation</i> |
|-------|-----------------------|----------------------|
| Cas 1 | 0.94 | 0.74 |
| Cas 2 | 0.77 | 0.66 |
| Cas 3 | 0.931 | 0.695 |
| Cas 4 | 0.89 | 0.838 |

On remarque de très bonnes performances du modèle développé pour le quatrième cas par rapport aux autres cas.

4.4.2 Application du modèle sur les données à optimiser

Afin d'exploiter les résultats pertinents obtenus par le modèle développé, on a déterminé le gain en matière d'énergie, durant les années « 2008-2016 » pour la STEP de Beni Messous.

Les résultats de la simulation de l'ensemble des journées non sélectionnées qui correspond au deuxième cluster pour le 4eme cas, sont illustrés dans la figure suivante :

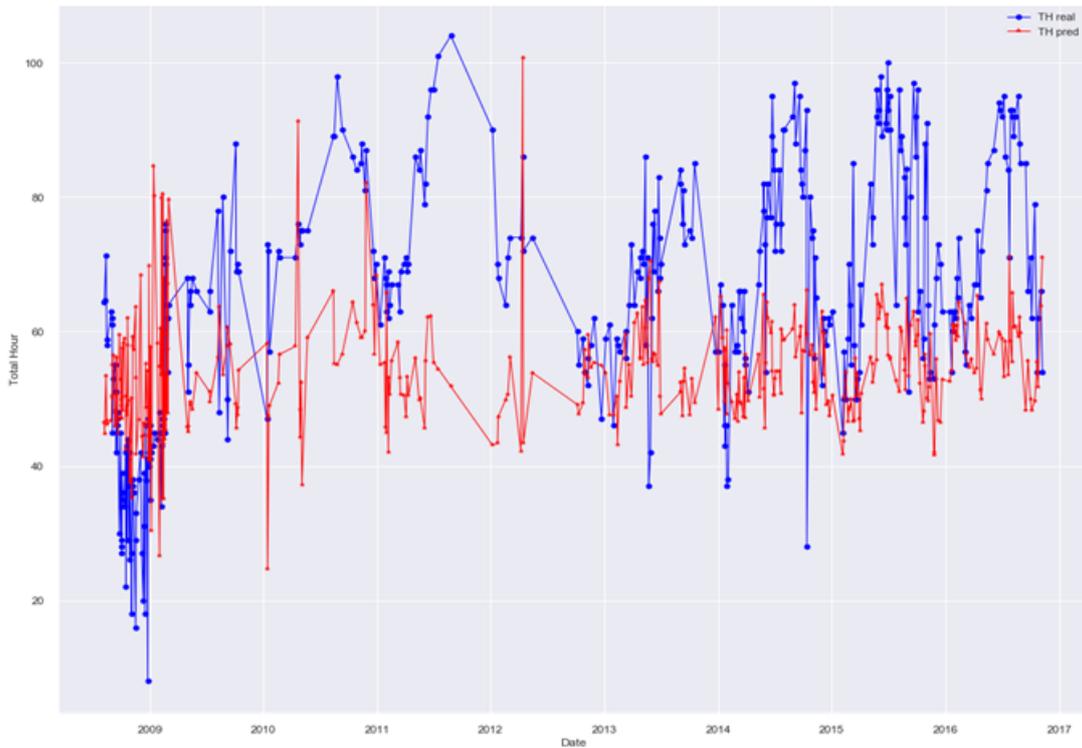


FIGURE 4.27 – Comparaison entre les heures de fonctionnement réel et simulée

Cette figure représente les valeurs de la Somme des heures de fonctionnement réelle (en bleu), et simulée (en rouge) des aérateurs par le modèle développé.

On remarque que ce dernier a surestimé seulement, l'énergie consommée entre 2008 et 2009, ceci s'explique par le fait que la STEP était en période d'essai, ce qui a pu occasionné plusieurs dysfonctionnements.

Les économies d'énergies estimées (soustraction des sommes d'heures de fonctionnement réelle et simulée par le modèle développé) sont représentées dans la figure suivante.

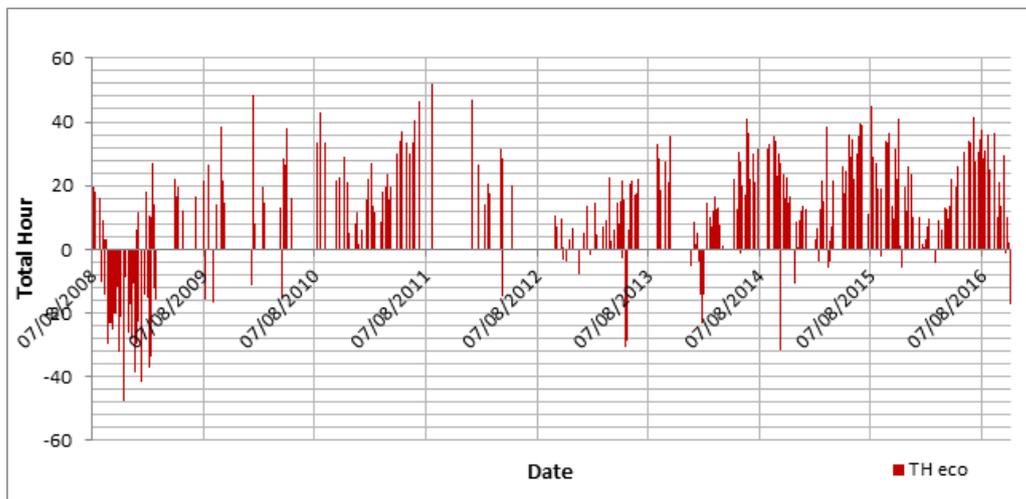


FIGURE 4.28 – Economies d'énergie réalisée pendant sur la période 2008-2016

En triant les observations des heures de fonctionnement par ordre croissant avec leur économie respective, on observe qu'il y'a une tranche de valeur ou le gain est nulle (entre 50 et 60 heures), et cela pourrait représenter la tranche des heures de fonctionnement optimal pour cette STEP.

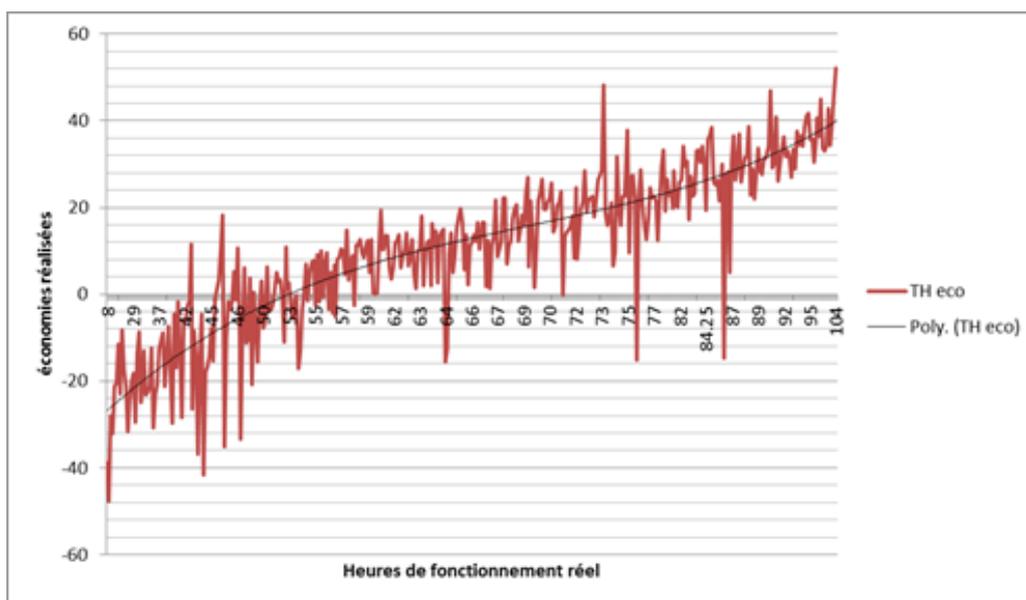


FIGURE 4.29 – Economies d'énergie réalisée par rapport aux valeur réel

L'économie d'énergie journalière moyenne pour la STEP de Béni Messous est représentée par l'histogramme suivant :

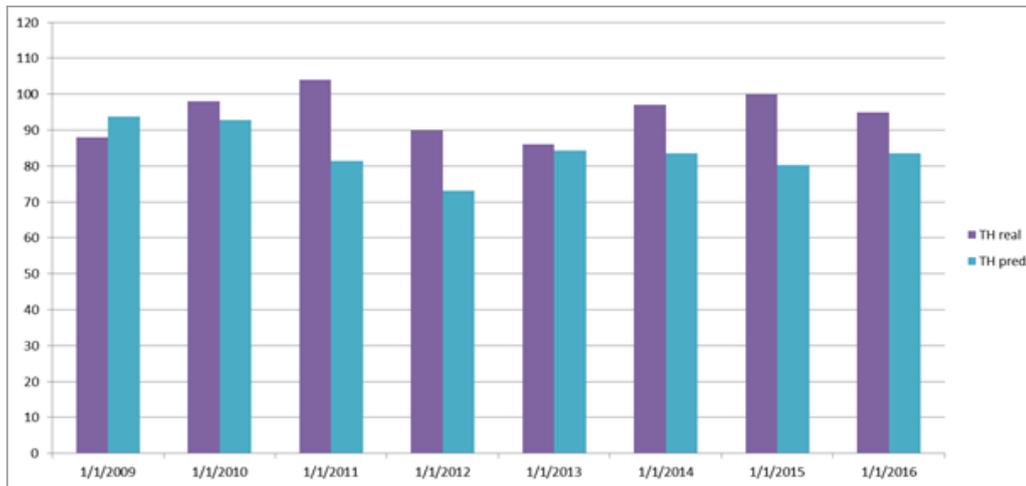


FIGURE 4.30 – Economie d'heure de fonctionnement et heure de fonctionnement réel par an

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

Sur la période 2008-2016 le modèle a pu économiser 4225 heures de fonctionnement, ce qui représente une moyenne de 10,75 h/j, qui correspond à une consommation énergétique de 806 kWh /j.

En supprimant les valeurs où l'économie estimée est négative (valeur concentrée principalement sur l'année 2008), le modèle économise 5608 heures de fonctionnement, ce qui représente une moyenne de 19 h/j, et une consommation énergétique de 1435.5 kWh /j.

4.5 Conclusion

Dans ce chapitre, nous avons montré comment nous avons utilisé deux approches statistiques qui sont l'algorithme K-Means et la méthode ACP pour préparer les données nécessaires à l'élaboration de notre modèle. Les résultats obtenus nous ont permis d'affirmer que l'analyse en utilisant l'algorithme K-means en ajoutant une troisième variable pour les différents cas est viable. Ce qui nous a permis d'obtenir des regroupements de données justifiées par des classes bien structurées pour l'élaboration du modèle de régression avec le Gradient Boosting.

Les résultats obtenus dans les différents cas en appliquant la technique du Gradient Boosting sur les données permettent de dire que notre modèle est performant et reflètent des résultats pertinents plus particulièrement dans le cas où on a utilisé les variables WQI et RG avec TH/QE, avec un résultat très bon de $R^2 > 0.8$ pour la validation, tout en vérifiant que ces résultats n'ont pas été obtenus à cause d'un sur apprentissage (overfitting).

Finalement le modèle obtenu a permis de réduire la consommation énergétique d'une manière significative, tout en assurant le même rendement épuratoire. Ce qui conduirait à abaisser grandement la facture énergétique et aussi la pollution engendrée par ce type de traitement.

Conclusion Générale

Le but de ce travail est de porter un jugement sur la consommation de l'énergie dans la station d'épuration de Béni Messous. 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.

La modélisation et l'optimisation des processus de traitement des eaux usées font face à trois Défis : Le premier est lié aux manques de données. Comme le traitement des eaux usées comprend les processus chimiques et biologiques, ainsi que les instruments collectant de gros volumes de données. De nombreuses variables de l'ensemble de données sont fortement couplées. Les données sont bruyantes, incertaines et incomplètes. Par conséquent, plusieurs algorithmes de prétraitement doivent être utilisés pour prétraiter les données, réduire sa dimensionnalité (en utilisant l'ACP) et déterminer les variables importants. Le deuxième défi est dans la nature temporelle du processus. Différents algorithmes d'exploration de données sont utilisés pour obtenir des modèles précis. En utilisant l'analyse par K means pour identifier un pattern au sein des données, et de regrouper les individus ayant des caractéristiques similaires.

Le dernier défi est l'optimisation des modèles de processus. Comme les phénomènes sont généralement hautement non linéaires et dynamiques, de nouveaux algorithmes évolutifs sont utilisés tel que le Gradient Boosting.

Pour ce faire la conception de modèles de régressions basés sur la nouvelle approche du Gradient Boosting utilisant les arbres de décision a été établie et expliquée en détail.

A l'amont de modèle, on a représenté l'évolution de chaque effluent au fil du temps afin de pouvoir faire une critique des données d'une part. D'autre part on a testé et critiqué à travers deux techniques statistiques, qui sont le K-Means, et l'Analyse en Composantes Principales, dans le but de justifier l'utilisation des différents paramètres de la STEP pour l'élaboration de modèle.

Le modèle développé, est le fruit de l'application de la méthode K means, avec trois variables sur la série de mesures des différents paramètres, caractérisant le fonctionnement de la STEP.

Différents logiciels et outils de programmation, ont été utilisés lors de ce travail, tel que le logiciel XLSTAT, pour l'application de la méthode ACP, et le langage de programmation PYTHON, avec ses bibliothèques (Scikit learn, Matplotlib, Pandas, Numpy) pour le K-Means et le Gradient Boosting.

Les résultats obtenus montrent une meilleure économie d'énergie pour la STEP de Béni Messous. Il nous est donc permis de dire que cette étude, nous a permis de franchir un pas très important dans la recherche de modèles d'optimisation de la consommation en énergie électrique des stations d'épuration des eaux usées, qui reste un domaine vierge qui n'a pas encore livré tous ses secrets.

Les futurs travaux devront se concentrer sur la distribution des économies réalisées avec le modèle développé sur les différents aérateurs, et leur distribution sur la journée. On pourra aussi comparée les résultats obtenues avec d'autres types de modèle utilisant l'apprentissage automatique comme par exemple les réseaux de neurones.

Bibliographie

- [1] *Analyse de l'eau : Aspects et règlementaire et technique.*
- [2] *Analyse en Composantes Principales.*
- [3] *mémento technique de l'eau.*
- [4] *L'épuration biologique des eaux résiduaires.* Paris, 1979.
- [5] *Gestion des eaux usées urbaines et industrielles.* Paris, 1982.
- [6] *épuration biologique des eaux usées urbaines.* 1984.
- [7] *Hydrauliques urbaine.* 1986.
- [8] *Activated sludge models ASM1, ASM2, ASM2d and ASM3.* IWA Publishing, 2000.
- [9] *l'analyse de l'eau : eaux naturelle, eau résiduaires et l'eau de mer.* Paris, 2005.
- [10] *Water Quality Indices.* Elsevier, 2012.
- [11] *The Hundred-page Machine Learning Boo.* 2019.
- [12] Amel.Hamriche, K.Tachet, and A.Bermad. Contribution à l'étude et la simulation des paramètres hydrométéorologiques par l'analyse en composantes principales (a.c.p.). Master's thesis, Ecole Nationale Polytechnique d'Alger, 1993.
- [13] JEAN CLAUDE BOEGLIN. Traitement biologique des eaux résiduaires. Technical report, Revue techniques de l'ingénieur.
- [14] R.M Brown, N.I. McClelland, R.A. Deininger, and R.G. Tozer. A water quality index do we dare? *Water Sewage Works*, 1970.
- [15] CNAM. Cours - arbres de décision, 2018.
- [16] BELAHMADI.M et SEDDIK.O. Etude de la biodégradation du 2,4- dichlorophénol par le macrobiote des effluents d'entrée et de sortie de la station d'épuration des eaux usées d'ibn ziad. Master's thesis, •, 2011.
- [17] Sylvie Gillot and Jean-marc Choubert. Biodegradable organic matter in domestic wastewaters : comparison of selected fractionation techniques. *Water Science and Technology*, 2010.
- [18] Goldstein.R and W.Smith. Water and sustainability :u.s electricity consumption for water supply and treatment—the next half century. Technical report, EPRI, 2002.
- [19] Goldstein.R and W.Smith. Water and sustainability :u.s electricity consumption for water supply and treatment—the next half century. Technical report, EPRI, 2002.
- [20] Rainier Hreiz, Abderrazak M. Latifi, and Nicolas Roche. Optimal design and operation of activated sludge processes : State-of-the-art. *The Chemical Engineering Journal*, 2015.
- [21] Stefano Longo, Benedetto Mirko d'Antoni, Michael Bongards, Antonio Chaparro, Andreas Cronrath, Francesco Fatone, Juan M. Lema, Miguel Mauricio-Iglesias, Ana Soares, and Almudena Hospido. Monitoring and diagnosis of energy consumption in wastewater treatment plants. a state of the art and proposals for improvement. *Applied Energy*, 2016.
- [22] N.Beck. Application de méthodes de clustering traditionnelles et extension au cadre multi-critère. Master's thesis, Université Libre de Bruxelles, 2006.

- [23] PNUE / OMS. Recommandation pour la surveillance sanitaire des zones côtières à usage récréatif et des zones conchylicoles. Technical report, Bureau régional de l’OMS pour l’Europe, 1979.
- [24] Erika Pittoors, Yaping Guo, and Stijn W. H. Van Hulle. Modeling dissolved oxygen concentration for optimizing aeration systems and reducing oxygen consumption in activated sludge processes : A review. *Chemical Engineering Communications*, 2014.
- [25] R.Maachou, A. Lefkir, A. Bermad, T. Djaoui, and A. Khouider. Statistical analysis of pollution parameters in activated sludge process. *Desalination and Water Treatment*, 2016.
- [26] R.Maachou, A. Lefkir, A. Bermad, and A. Khouider. Factorization of physicochemical parameters of activated sludge process using the principal component analysis. *Desalination and Water Treatment*, 2015.
- [27] R.Maachou, A. Lefkir, A. Bermad, and A. Khouider. Modeling of activated sludge process using artificial neuro-fuzzy-inference system (anfis). *Desalination and Water Treatment*, 2015.
- [28] DARIO TORREGROSSA. *A decision support system for energy saving in WasteWater Treatment Plants*. PhD thesis, UNIVERSITÉ DU LUXEMBOURG EN SCIENCE DE L’INGÉNIEUR, 2018.
- [29] Torregrossa.D, Schutz.G, Cornelissen.A, Hernández-Sancho.F, and Hansen.J. Energy saving in wwtp : Daily benchmarking under uncertainty and data availability limitations. *Environmental Research*, 2016.
- [30] Le Anh Tuan and Ifi Hanoi. Réduction de base de données par la classification automatique. Master’s thesis, Institut de la francophonie pour l’informatique (IFI), 2004.