

**MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE**

ECOLE NATIONALE POLYTECHNIQUE



المدرسة الوطنية المتعددة التخصصات
Ecole Nationale Polytechnique

Mémoire présenté par :

Melle MADI Maïssa

DEPARTEMENT HYDRAULIQUE

Mémoire de fin d'études

En vue de l'obtention du diplôme d'Ingénieur d'Etat en Hydraulique

Thème

**Caractérisation de la qualité des eaux de
surface par hybridation des techniques de
Data Mining**

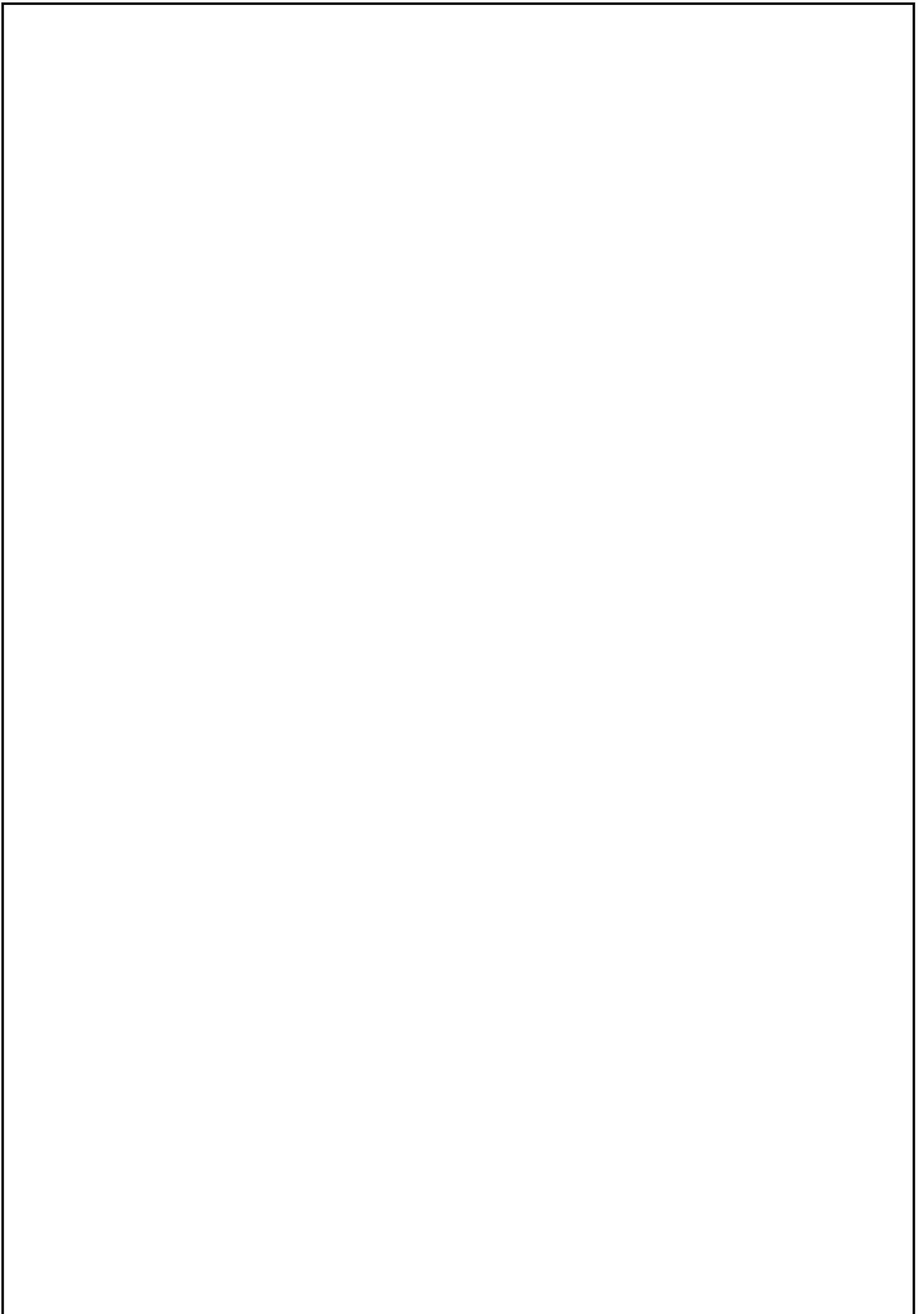
Composition du jury :

Président : Abdelkader NAMANE MCA ENP

Promoteur : Abdelmalek BERMAD Professeur ENP

Examineurs: Abdelouahab LEFKIR MCA ENSTP

Rafik OULEBSIR Doctorant USTHB



**MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE**

ECOLE NATIONALE POLYTECHNIQUE



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

Mémoire présenté par :

Melle MADI Maïssa

DEPARTEMENT HYDRAULIQUE

Mémoire de fin d'études

En vue de l'obtention du diplôme d'Ingénieur d'Etat en Hydraulique

Thème

**Caractérisation de la qualité des eaux de
surface par hybridation des techniques de
Data Mining**

Composition du jury :

Président : Abdelkader NAMANE MCA ENP

Promoteur : Abdelmalek BERMAD Professeur ENP

Examineurs: Abdelouahab LEFKIR MCA ENSTP

Rafik OULEBSIR Doctorant USTHB

ملخص

تعد المياه السطحية و مياه السدود المصدر الاكثر شيوعا للتزود بالماء العذب وذلك لسهولة الوصول إليها. هذا ما جعلها معرضة لخطر تلوث المياه الذي جاء بسبب تطور الصناعة والنمو السكاني الذي عرفه بلدنا في السنوات الأخيرة. يعد الحفاظ على جودة المياه أمرًا أساسيًا للتنمية المستدامة والحفاظ على هذه الثروة للأجيال القادمة.

الهدف من عملنا هو تصنيف المعلمات المسؤولة عن تدهور جودة المياه السطحية في منطقة السد "بني هارون" وباستخدام الطرق المختلفة لتعدين البيانات مثل "تحليل المكونات الرئيسية", "طريقة الك المتوسط" و "التصنيف التصاعدي الهرمي".
الكلمات المفتاحية : تحليل المكونات الرئيسية, طريقة الك المتوسط, التصنيف التصاعدي الهرمي. تلوث المياه, جودة المياه.

Abstract :

Surface water is the most common and accessible source of water. Because of that and due to the development of industry and the population growth that know our country the last decencies, it became threatened by water pollution. The maintain of the water quality is fundamental for sustainable development and to conserve the source for the next generations. The object of our work is to classify the parameters responsible for the degradation of the quality of surface water in the area of the dam "Beni Haroun" and by using the different methods of Data Mining such as "Principal Component Analysis (PCA)" and "The K-Means Method" combined with "Hierarchical Ascending Classification (HAC)".

Key words: Principal Component Analysis (PCA), K-means, Hierarchical Ascending Classification (HAC), Water pollution, Water Quality.

Résumé:

Les eaux de surfaces ont toujours été la source la plus utilisable et la plus accessible pour l'approvisionnement d'eau. Mais à cause du développement de l'industrie et de la croissance démographique que connaît notre pays les dernières décences, elles sont devenues menacé par la pollution de l'eau. Le maintien de la qualité de l'eau est fondamental pour le développement durable et pour conserver la ressource pour les prochaines générations.

L'objet de nos travaux est de classer les paramètres responsables de la dégradation de la qualité des eaux de surface dans la zone du barrage «Beni Haroun» et en utilisant les différentes méthodes de Data Mining telles que «Analyse en Composantes Principales (ACP)» et "La méthode K-Means" combinée à la "Classification hiérarchique ascendante (CAH)".

Mots clés : Analyse en composantes principales (ACP), K-means, Classification ascendante hiérarchique (CAH), pollution des eaux, qualité des eaux.

Dédicace

A celle qui m'a élevée et m'as vu grandir, pour tout l'amour qu'elle m'a donnée toutes ses années, pour ses prières et son soutien, que dieu la protège et la préserve en bonne santé : ma grand-mère

A ma chère mère, pour tous ses sacrifices et ses encouragements, pour son soutien continue et son amour inconditionnelle et surtout pour toute la force qu'elle m'a transmise pour poursuivre mes objectifs, pour sa patience et ses prières qui j'espère vont m'accompagner toute ma vie

A mon pilier dans la vie, mon père, celui qui a sacrifié tant d'années pour notre confort, qui a toujours essayer de donner le mieux de lui-même pour qu'on manque de rien, mais surtout je le remercierai jamais assez pour tout l'amour qu'il m'a apporté, pour ses encouragements, son soutien et sa confiance en moi, cette confiance qui m'a toujours poussée vers l'avant pour donner le meilleur de moi-même et de le rendre fier.

A ma douce tante, celle qui a toujours cru en moi, qui a toujours été derrière moi toujours fière de ce que j'ai pu accomplir, je ne te remercierai jamais assez pour tout l'amour que tu m'as apportée toutes ses années, merci pour tout ma deuxième mère

A mon frère et ma sœur, qui ont toujours eu confiance en moi et m'ont toujours vu comme un bon exemple et qu'ont toujours été présents pour moi

A mes chers amis : Arslane, Amina, Mouh, Fethia et Latif, ces amis avec lesquels j'ai passé les meilleurs moments de ma vie et qui ont toujours été là pour moi à m'écouter et m'encourager,

A ma chère grande famille, tous mes oncles, mes cousins et cousines, et une pensée spéciale à mon cher grand-père et grand-mère « qu'Allah les accueillent en son vaste paradis »

A toute la promotion hydraulique 2019-2020

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 infailible,

Remerciements

Je tiens 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 notre encadreur Mr BERMAD Abdelmalek pour ses précieux conseils et son aide durant toutes mes années au sien du département et surtout durant ce travail

Mes remerciements s'étendent également à Mr A. LEFKIR et R. OULEBSIR pour leurs bonnes explications qui nous ont éclairé le chemin de la recherche et leur collaboration avec nous dans l'accomplissement de ce travail. Mes vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner notre travail et de l'enrichir par leurs propositions.

Je tiens à exprimer mes sincères remerciements à tous les professeurs qui m'ont enseigné et qui par leurs compétences m'ont soutenu dans la poursuite de mes études. Enfin, je tiens également à remercier toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail.

TABLE DES MATIERES

LISTE DES TABLEAUX.....	
LISTE DES FIGURES	
INTRODUCTION GENERALE	9
CHAPITRE 1 : LA POLLUTION DES EAUX DE SURFACE ET SES CARACTERISTIQUES.....	12
INTRODUCTION :.....	13
1-1- TYPES DE POLLUTION.....	13
1-1-1- ORIGINES DE POLLUTION.....	13
<i>1-1-1-1- Pollution d'origine organique</i>	<i>13</i>
<i>1-1-1-2- Pollution d'origine chimique.....</i>	<i>14</i>
1-1-2- SOURCE DE POLLUTION	14
<i>1-1-2-1- Pollution industrielle.....</i>	<i>14</i>
<i>1-1-2-2- Pollution agricole.....</i>	<i>15</i>
<i>1-1-2-3- Pollution domestique.....</i>	<i>15</i>
<i>1-1-2-4- Pollution accidentelle.....</i>	<i>15</i>
1-2- LES PARAMETRES CARACTERISTIQUES DE L'EAU :.....	16
1-2-2- LE PH :	17
1-2-3- LA CONDUCTIVITE :	17
1-2-5- ELEMENTS EN SOLUTION :	17
1-2-6- LA DURETE :	17
1-2-7- SODIUM ET POTASSIUM	18
1-2-8- SUBSTANCES EUTROPHISANTES : DIFFERENTES FORMES D'AZOTE ET DE PHOSPHORE	18
1-2-9- OXYGENE DISSOUS (OD) ET % DE SATURATION EN OXYGENE :.....	18
1-2-10 - CHARGE EN MATIERES ORGANIQUES : DEMANDE BIOCHIMIQUE EN OXYGENE (DBO) ET DEMANDE CHIMIQUE EN OXYGENE (DCO) :.....	19
1-2-11- METAUX LOURDS :	19
1-3- NORMES ALGERIENNES POUR LES EAUX DE SURFACES :	20
CHAPITRE 2 : THEORIES D'ACP ET DE CLUSTERISATION.....	22
INTRODUCTION :.....	23
2-1- ANALYSE EN COMPOSANTES PRINCIPALES.....	23
<i>2-1-1- Définition.....</i>	<i>23</i>
<i>2-1-2- Notions d'individu et de caractère</i>	<i>23</i>
<i>2-1-4- Objectifs</i>	<i>24</i>
<i>2-1-5- Application de l'ACP.....</i>	<i>24</i>
<i>2-1-6- Procédé d'application de l'ACP.....</i>	<i>25</i>
<i>2-1-7- Analyse des résultats</i>	<i>25</i>
2-2- LA CLASSIFICATION ASCENDANTE HIERARCHIQUE	27
<i>2-2-1- Définition.....</i>	<i>27</i>
<i>2-2-2- Avantages de la CAH :</i>	<i>27</i>
2-3- K-MEANS :.....	28
<i>2-3-1- Définition.....</i>	<i>28</i>
<i>2-3-2- Algorithme</i>	<i>29</i>
<i>2-3-3- Avantages de K-means</i>	<i>30</i>
<i>2-3-4- Inconvénients</i>	<i>30</i>
CHAPITRE 3 : PRESENTATION DE LA ZONE D'ETUDE.....	31
INTRODUCTION.....	32
3-1- BARRAGE DE BENI HAROUN	32
3-2- CARACTERISTIQUES DU BASSIN VERSANT	33
<i>3-2-1- Le relief.....</i>	<i>34</i>
<i>3-2-2- les altitudes</i>	<i>35</i>
<i>3-2-3- La géologie :</i>	<i>35</i>

3-2-4- Le climat:	36
3-3- CARACTERISTIQUE DES STATIONS :	36
3-3-1- <i>Ain Smara</i> :	36
3-3-2- <i>Menia</i> :	37
3-3-3- <i>Le barrage</i> :.....	37
3-3-4- <i>Activité industrielle de la région</i>	37
3-3-5- <i>Densité urbaine</i> :	40
3-3-6 <i>L'activité agricole</i> :	40
3-4- IMPACT DES REJETS DES EAUX USEES SUR LA QUALITE DES EAUX.....	41
LA MAJORITÉ DES REJETS DES TRENTE DEUX (32) AGGLOMÉRATIONS ET DE L'INDUSTRIE DE LA RÉGION, DÉVERSENT LEURS EAUX USÉES DANS LES COURS D'EAU DU BASSIN. CES REJETS DOMESTIQUES ET INDUSTRIELS DANS LES COURS D'EAU DU BASSIN ET DANS LA RETENUE DU BARRAGE FAVORISENT L'ENRICHISSEMENT DE L'EAU EN ELEMENTS NUTRITIFS, PHOSPHORE ET AZOTE ET NOTAMMENT L'AUGMENTATION CORRELATIVE DE LA PRODUCTION VEGETALE AQUATIQUE, PRINCIPALEMENT DES ALGUES PLANCTONIQUES OU FIXEES, CE QUI ALTERE LES PROPRIETES PHYSIQUES ET CHIMIQUES DE L'EAU. (DECAMPS,1996).....	41
3-4-1- <i>Données et prélèvements des échantillons</i> :	41
CHAPITRE 4: COLLECTE ET CRITIQUE DES DONNES	44
INTRODUCTION.....	45
4-1- CRITIQUE DE LA DONNEE :	45
- <i>Ain Smara</i>	45
- <i>Menia</i>	46
- <i>Beni Haroun</i>	48
4-2- ELIMINATION DES VALEURS ABERRANTES ET EXTREMES :	50
- <i>Ain Smara</i>	51
- <i>Menia</i>	54
- <i>Beni Haroun</i>	55
CHAPITRE 5 : RESULTATS ET DISCUSSIONS	57
INTRODUCTION :	58
5-1- ANALYSE EN COMPOSANTES PRINCIPALES	58
5-1-1- <i>Application de l'ACP</i>	58
- <i>Ain Smara</i>	58
- <i>Menia</i>	60
- <i>Beni Haroun</i>	62
V-2- K-MEANS	64
5-3- LA CLASSIFICATION ASCENDANTE HIERARCHIQUE	70
5-4- RESUME DES RESULTATS DE CLASSIFICATION SELON L'OMS :	74
- <i>Ain Smara</i>	75
- <i>Menia</i>	76
- <i>Beni Haroun</i>	78
CONCLUSION GENERALE	80
REFERENCES BIBLIOGRAPHIQUES	82

Liste des tableaux

Tableau 1-1-Normes algériennes des eaux de surface	19
Tableau 2-1: Classification des résultats en fonction du KMO	25
Tableau 2-2: Classification des résultats par le test de Bartlett	26
Tableau 3-1 : Les sous bassins et leurs superficies	33
Tableau 3-2 : Les classes des altitudes dans le bassin	35
Tableau 3-3- : Caractéristiques statistiques des données mensuelles climatiques, apports liquides et volume de la cuvette du barrage Béni Haroun	36
Tableau 3-4: Liste des activités principales de la région et leurs principaux rejets	37
Tableau 3-5: Besoins, consommations et déficits (Agglomération supérieur à 5000 hab)	39
Tableau 3-6: Coordonnées des stations	40
Tableau 4-1: Caractéristiques statistiques Ain Smara	43
Tableau 4-2: Série de données (1994-2010) Ain Smara	44
Tableau 4-3 : Caractéristiques statistiques Menia	45
Tableau 4-4 : Série de données (1994-2010) Menia	46
Tableau 4-5 : Paramètres statistiques Beni Haroun	47
Tableau 4-6 : Série de données (1994-2010) Beni Haroun	48
Tableau 4-7: Nombre de valeurs extrêmes par paramètre « Ain Smara »	51
Tableau 4-8: Nombre de valeurs extrêmes par paramètre « Menia »	52
Tableau 4-9: Nombre de valeurs extrêmes par paramètre « Beni Haroun »	54
Tableau 5-1 : Tableau des valeurs propres ACP, Ain Smara	56
Tableau 5-2 : Tableau des cosinus ACP, Ain Smara	57
Tableau 5-3 : Tableau des valeurs propres ACP, Menia	58
Tableau 5-4 : Tableau des cosinus ACP, Menia	59
Tableau 5-5 : Tableau des valeurs propres ACP, Beni Haroun	60
Tableau 5-6 : Tableau des cosinus ACP, Beni Haroun	61
Tableau 5-7 : Résultats K-means Ain Smara	63
Tableau 5-8:Résultats K-mean Menia	65
Tableau 5-9:Résultats K-mean Beni Haroun	66
Tableau 5-10 : Classification par CAH avec coefficient de corrélation de Pearson Ain Smara	68
Tableau 5-11 : Classification par CAH avec coefficient de corrélation de Pearson Menia	69
Tableau 5-12 : Classification par CAH avec coefficient de corrélation de Pearson Beni Haroun	70
Tableau 5-13 : Classification selon les normes de l’OMS	72
Tableau 5-14 : Classification selon les normes de l’OMS de la station d’Ain Smara avant réception du barrage	72
Tableau 5-15: Classification selon les normes de l’OMS de la station d’Ain Smara après réception du barrage	73
Tableau 5-16: Récapitulatif de classification selon les paramètres station Ain Smara	73
Tableau 5-17: Récapitulatif de classification selon les matières station Ain Smara	73
Tableau 5-18 : Classification selon les normes de l’OMS de la station de Menia avant réception du barrage	74
Tableau 5-19 : Classification selon les normes de l’OMS de la station de Menia après réception du barrage	74
Tableau 5-20: Récapitulatif de classification selon les paramètres station Menia	75
Tableau 5-21: Récapitulatif de classification selon les matières station Menia	75
Tableau 5-22 : Classification selon les normes de l’OMS de la station de Beni Haroun avant réception du barrage	77
Tableau 5-23 : Classification selon les normes de l’OMS de la station de Beni Haroun après réception du barrage	77
Tableau 5-24 : Récapitulatif de classification selon les paramètres station Beni Haroun	77
Tableau 5-25: Récapitulatif de classification selon les matières station Beni Haroun	78

Liste des figures

Figure 1-1: Répartition des produits polluants selon leurs source de rejets	15
Figure 2-1: Un dendrogramme	28
Figure 2-2: Principe de la méthode de clustering K-means	29
Figure 2-3: La progression de l'algorithme pour k=3	29
Figure 3-1: Situation géographique et carte du bassin versant Béni Haroun	34
Figure 3-2: Carte de l'emplacement des trois stations	41
Figure 3-3 : Situation géographique des stations de mesure dans le bassin Kebir-Rhumel	41
Figure 4-1: Test de Grubbs	49
Figure 4-2 : Résultat test de Grubbs (paramètre Cu, station Ain Smara)	50
Figure 4-3 : Résultat test de Grubbs (paramètre Na, station Ain Smara)	50
Figure 4-4 : Résultat test de Grubbs (paramètre K, station Ain Smara)	51
Figure 4-5 : Résultat test de Grubbs (paramètre Cl, station Menia)	52
Figure 4-6 : Résultat test de Grubbs (paramètre NH ₄ , station Menia)	52
Figure 4-7 : Résultat test de Grubbs (paramètre T, station Beni Haroun)	53
Figure 4-8: Résultat test de Grubbs (paramètre SO ₄ , station Beni Haroun)	53
Figure 5-1: Cercle des Corrélations ACP Ain Smara	59
Figure 5-2: Cercle des Corrélations ACP Menia	61
Figure 5-3: Cercle des Corrélations ACP Beni Haroun	63
Figure 5-4: Courbe de l'évolution des variances K-means Ain Smara	64
Figure 5-5 : Courbe de l'évolution des variances K-means Menia	66
Figure 5-6: Courbe de l'évolution des variances K-means Beni Haroun	67
Figure 5-7: Le dendrogramme avec coefficient de corrélation de Pearson Ain Smara	68
Figure 5-8: Le dendrogramme avec coefficient de corrélation de Pearson Menia	69
Figure 5-9: Le dendrogramme avec coefficient de corrélation de Pearson Beni Haroun	70
Figure5-10: Regroupement des paramètres CAH-Pearson	70

Introduction générale

L'eau est une ressource indispensable pour la continuité de toute forme de vie. Pour cette raison, l'homme a toujours privilégié de bâtir ses communautés dans des zones humides près des rivières et des lacs.

L'eau, élément vital, nécessaire à tous les aspects de la vie, constitue un élément essentiel de tous les écosystèmes terrestres et indispensable à l'homme dans sa vie quotidienne.

L'agriculture d'une autre part, a été la première activité que l'homme avait connue. L'irrigation de ses terrains est devenue sa priorité, c'est pour cela qu'il a toujours cherché de nouvelles ressources d'eau.

Avec l'évolution de l'être humain, son secteur d'activité s'est développé et donc ses besoins en eau aussi. L'arrivée de l'industrie et la croissance démographique dans le monde ont créé une forte demande en eau, ce qui a poussé les états à réfléchir à de nouvelles sources d'alimentation qui sont les barrages.

Les barrages sont des ouvrages utilisés par les anciennes civilisations mais qui ont réapparus dans le début du 20ème siècle.

Ces ouvrages ont pu satisfaire la demande croissante en eau, mais d'autres problèmes sont apparus.

La diversité de l'industrie et l'utilisation des engrais en agriculture ont été la cause de la création d'une nouvelle forme de pollution : la pollution de l'eau.

Cette pollution qui est essentiellement causée par les rejets industriels, les débits de ruissellement de surface à partir des terres agricoles, sans oublier les rejets domestiques qui eux aussi avec le temps ont évolués.

Plusieurs substances nouvelles, biodégradables ou pas, se retrouvent dans les cours d'eaux qui ne sont pas familiers avec elles surtout pas dans ces quantités. De plus, leurs pouvoirs auto-épurateur n'est pas prêt à en tenir compte, d'où la création de la pollution.

La ressource en eau se fait rare à cause de sa surconsommation, et l'effet du changement climatique a déséquilibré l'écosystème et le cycle de l'eau naturel plus précisément.

Les précipitations ne sont plus aussi régulières et les zones semi arides deviennent petit à petit arides, ce qui a causé aussi la réduction des nappes et l'augmentation des fréquences des crues. La protection et la surveillance de cette ressource est désormais la responsabilité de tous les états du monde, et de garantir sa disponibilité aux générations futures. C'est une

nouvelle politique de la gestion des ressources en eau qui a été adoptée par 110 états en 1992 lors du sommet de Rio de Janeiro qui cherchaient une stratégie mondiale de développement durable.

Après ce sommet, les pays participants sont devenus plus vigilants concernant la gestion et l'exploitation de la ressource en eau, ainsi que leur traitement et leur réutilisation après usage. En Algérie, l'évaluation des ressources en eau potable a mis en évidence une situation tendue en matière de ressources en eau, en particulier en période estivale, avec des besoins en eau plus importants, et des ressources à l'étiage. Cette évaluation a suscité une mobilisation importante de tout le pays sur le thème de l'eau, tant d'un point de vue quantitatif que qualitatif.

Globalement l'Algérie, reçoit 100 milliards de m³ de pluie avec des pertes en évapotranspiration de 80% . Le tell, comprenant les bassins tributaires de la Méditerranée (Chélif, Seybouse, Kebir Rhumel, Sebaou, Isser, Tafna, Macta et Medjerda-Mellegue) reçoit près de 70% des précipitations, sur une superficie de 130.000 km² avec une pluviométrie moyenne qui varie de 500 à 600 mm.

Ce potentiel déjà faible, est aggravé par un grand nombre de problèmes, d'une part, l'envasement qui atteint ces barrages et retenues collinaires est de plus de 60%, et d'autre part, les réseaux d'adduction d'eau potable aux populations urbaines, suburbaines et rurales non entretenues avec des pertes qui atteignaient 40%, et enfin 600 millions de m³ d'eau, soit l'équivalent de 10 barrages sont inutilisables pour cause de pollutions urbaine, industrielle et agricole.

Le barrage de Béni Haroun ayant la plus grande réserve d'eau de surface en Algérie, fait l'objet d'un suivi de qualité. Pour cela, l'ANRH (Agence Nationale des Ressources Hydrauliques) a mis en place quatre (4) stations de mesure pour contrôler et caractériser la qualité de l'eau du barrage et de ses différents affluents.

Ce travail est consacré à la recherche des substances les plus nuisibles à la qualité d'eau. En utilisant les outils du Data Mining tels que l'Analyse en Composantes Principales, les méthodes du clustering K-Means et de la Classification Ascendante Hiérarchique.

Ce mémoire est scindé en cinq (5) chapitres, le **premier** est dédié à la caractérisation de la pollution de l'eau, ses causes et les paramètres qui y contribuent. Dans le **second chapitre**, on a abordé les différents traitements appliqués aux séries de données recueillies à l'amont du barrage, à leur analyse, et à leur classification. Les théories des différentes méthodes d'analyse, et de clustering que vont subir les données observées ont fait l'objet du **Chpître III**. Alors que le chapitre IV a été consacré à la présentation de la zone d'étude. Enfin, dans le

5ème chapitre, on a détaillé l'ensemble des résultats obtenus, issus de l'application des techniques du Data Mining, et leurs interprétations.

Chapitre 1 : La pollution des eaux de surface et ses caractéristiques

Introduction :

Les eaux de surface sont la source primaire pour l'alimentation en eau potable, l'irrigation et l'industrie. Elles sont très affectées par les diverses sources de pollution et de contamination ce qui va dégrader leur qualité et donc mettre des restrictions pour leurs utilisations. Dans ce chapitre nous allons détailler les types de pollutions existantes et tous les paramètres qui peuvent altérer la qualité des eaux et les normes utilisées pour classer les eaux de surfaces étudiées.

La pollution des eaux de surface se partage en 2 catégories principales :

1- Pollution d'origine organique

2- Pollution d'origine chimique

Et peut provenir à l'origine de ces 4 sources :

1- Pollution industrielle

2- Pollution agricole

3- Pollution domestique

4- Pollution accidentelle

1-1- Types de pollution

1-1-1- Origines de pollution

1-1-1-1- Pollution d'origine organique

On peut décrire cette pollution comme étant la pollution naturelle que les eaux de surface subissent. Elle se caractérise par la décomposition de la matière organique, constituée d'éléments vivants du système (feuilles d'arbres, ..). Cette matière étant plus au ou moins biodégradable consomme de l'Oxygène Dissous pour se décomposer, et peut donc altérer la qualité de l'eau de surface.

On peut en conclure que les bas niveaux de l'Oxygène Dissous dans les eaux de surfaces est un indice de l'existence de la pollution organique. D'autres indices peuvent être mentionnés dans ce cas, comme l'augmentation des niveaux de nitrites (NO₂) et de l'ammoniaque (NH₄) qui sont un indice de la décomposition de la matière

organique riche en azote. Ces deux éléments en cas des valeurs extrêmes peuvent nuire à la vie aquatique et altérer la qualité des eaux de surface.

1-1-1-2- Pollution d'origine chimique

C'est une pollution d'origine humaine. Elle se manifeste par le biais des produits chimiques de synthèse. Le problème avec cette pollution contrairement à la pollution organique est que les substances chimiques sont allogènes par rapport au cours d'eau. Donc son pouvoir auto-épurateur ne tient pas compte de ces substances nocives. La plupart de ces composants chimiques ne sont pas biodégradables (métaux lourds, pesticides, carburants, acides) et donc restent en suspension ou bien en sédiments plusieurs centaines d'années. Cette pollution est l'origine d'une menace pour la vie aquatique et la dégradation totale des eaux de surfaces qui ne peuvent plus être utilisées sans un traitement préliminaire.

1-1-2- Source de pollution

1-1-2-1- Pollution industrielle

L'origine primaire de la pollution industrielle est les rejets d'eaux polluées des usines de production dans les cours d'eaux. Ces rejets sont appelés effluents industriels. Le traitement de ces eaux doit être effectué impérativement à l'amont des cours d'eaux pour diminuer l'effet des toxines existantes dans ces eaux sur les organismes vivants des cours d'eaux. Par conséquent, épurer les effluents industriels jusqu'à atteindre le niveau des paramètres du cours d'eau pour laisser ce dernier s'occuper du reste de la pollution avec son pouvoir auto-épurateur. Ce processus élimine les substances telles que les métaux lourds, les éléments radioactives et toxique que le cours d'eau ne pourra jamais éliminer et donc éviter la création d'une pollution chimique au niveau de ce dernier.

Cette pollution peut provenir de 4 secteurs d'activités principaux :

- a- Industries agro-alimentaires
Principaux rejets : matière organique ou azoté
- b- Industries chimiques et transformation des métaux
Principaux rejets : métaux lourds
- c- Industries de tannage des peaux

Principaux rejets : acides et chromes

d- Carrières et centrales thermiques

Principaux rejets : matière en suspension des mines et augmentation de la température de l'eau.

1-1-2-2- Pollution agricole

Cette pollution est à l'origine de l'utilisation intensive des phytosanitaires et des pesticides (herbicides, insecticides, fongicides). Pour augmenter le rendement des terres agricoles, les agriculteurs ont eu recours à l'utilisation des engrais. Ces derniers et sous l'effet du ruissellement de surface rejoignent les cours d'eau et créent une pollution bactériologique. Les deux composants principaux des engrais sont les nitrates et les phosphates. Ces éléments sont à l'origine de l'eutrophisation qui est une forme de pollution bactériologique et qui produit des algues dans les fonds des cours d'eau provoquant l'augmentation de la turbidité de ces derniers. Cette pollution dégrade non seulement l'aspect visuel de l'eau avec l'apparition des algues, mais aussi influe sur la qualité et l'exploitation de cette eau et son utilisation pour couvrir la demande en eau potable et en irrigation.

1-1-2-3- Pollution domestique

Dans un habitat, on peut distinguer deux types de rejets :

a- Les eaux de lavage : regroupent toutes les eaux provenant des salles de bains et des cuisines et qui sont généralement chargées de graisses, de détergents et de débris organiques.

b- Les eaux vannes : regroupent les eaux provenant des sanitaires et qui sont généralement chargées des matières organiques azotées et des germes fécaux.

Dans un milieu urbain, il faut en tenir compte des rejets des habitations, des hôtels, des écoles et des hôpitaux qui contiennent d'autres substances polluantes tel que les produits désinfectants et antiseptiques mais aussi des produits contenant des métaux lourds.

1-1-2-4- Pollution accidentelle

Cette pollution contrairement à celles citées précédemment, est une pollution ponctuelle qui vient perturber l'écosystème. Elle peut provenir d'une fuite des réseaux d'assainissements urbains, déversements de produits polluants ou bien

d'eaux polluées dues à des accidents (camion-citerne, bacs endommagés, fuites sur canalisations, fuites au cours du chargement ou du déchargement des produits...). Elle peut aussi être d'origine agricole dans le cas des fuites d'effluents, issus directement des bâtiments d'élevage ou fuites de produits de traitement des cultures liées à leur manipulation. L'avantage de ce type de pollution est qu'une fois localisé, elle peut être réglé mais peut causer d'énormes dégâts à la qualité de l'eau vue qu'elle ne peut être estimée d'une façon précise. (MATEF, 2002)

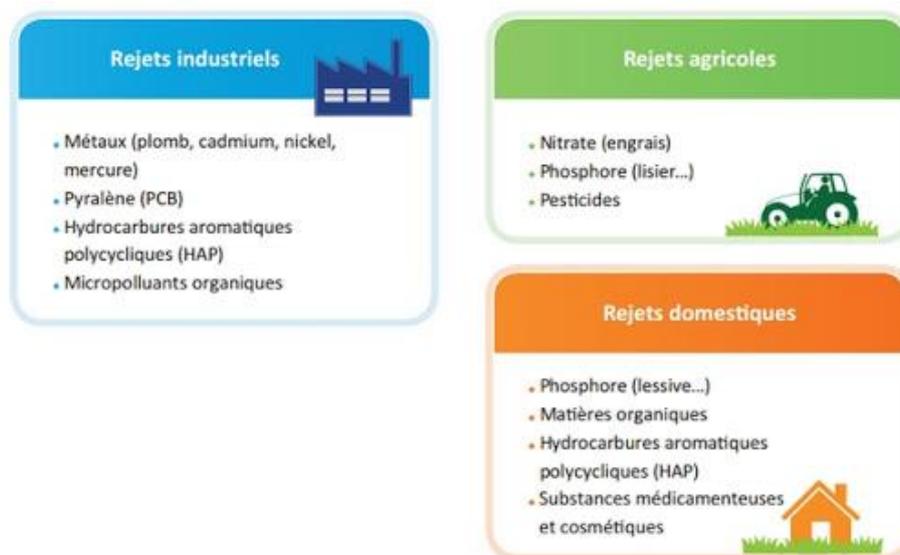


Figure 1-1 : Répartition des produits polluants selon le type de rejets (SMMAR, 2016)

1-2- Les paramètres caractéristiques de l'eau :

Pour pouvoir caractériser la qualité d'une eau il faut se référer à des paramètres chimiques et physico-chimiques qui évaluent le degré de pollution des eaux de surfaces.

1-2-1- Température :

La température varie selon un cycle saisonnier naturel. Ces variations affectent les organismes aquatiques qui eux ont une limite supérieure et inférieure de tolérance. Ces limites varient selon chaque espèce.

La température de l'eau joue un rôle important par exemple en ce qui concerne la solubilité des sels et des gaz dont, entre autres, l'oxygène nécessaire à l'équilibre de la vie aquatique. Par ailleurs, la température accroît les vitesses des réactions chimiques et biochimiques d'un facteur 2 à 3 pour une augmentation de température de 10 degrés Celsius (°C). L'activité

métabolique des organismes aquatiques est donc également accélérée lorsque la température de l'eau s'accroît. (IBGE 2005)

1-2-2- Le pH :

Le pH indique la concentration des ions H^+ présents dans l'eau. Pour une eau de surface le pH peut varier de 4 à 10. Une eau trop acide, peut indiquer la présence des métaux lourds qui dégradent la qualité de l'eau, et un pH trop élevé indique la présence de l'ammoniac qui est très dangereux pour la vie aquatique.

1-2-3- La conductivité :

C'est la capacité d'une solution à conduire le courant électrique. Elle est mesurée à 20°C. Elle provient de la présence des sels minéraux dans l'eau qui sont de bons conducteurs. L'estimation de la quantité de matière minérale dissoute est obtenue par une formulation empirique en fonction de la conductivité.

Une variation trop importante dans la conductivité peut nuire à la vie aquatique et elle est généralement causée par un déversement d'eaux usées.

1-2-4- La turbidité :

On dit qu'une eau est turbide quand on perd la visibilité à travers cette dernière. Cette perte est due à la présence des matières en suspension (MES) qui sont généralement constituées d'argiles et de particules organiques très fines résultants de la fragmentation des roches, l'érosion ou bien la décomposition des matières organiques.

Une forte turbidité réduit l'activité photosynthétique de la végétation aquatique submergée et enracinée et des algues, ce qui ralentit leur croissance et peut nuire, par conséquent, à la croissance et à la reproduction des poissons. Elle n'alterne pas la qualité d'eau potable mais réduit l'aspect esthétique par le changement de couleur de cette dernière.

1-2-5- Eléments en solution :

Les ions Cl^- et SO_4^{2-} doivent être surveillés régulièrement car ils indiquent deux types de pollution bien particuliers. Les eaux chargées en Cl^- proviennent des ménages et indiquent une pollution ménagère et les eaux chargées en sulfates sont plutôt des eaux usées industrielles. Les grandes variations de ces deux paramètres sont à surveiller car ça dégrade la qualité des eaux de surface.

1-2-6- La dureté :

La dureté est liée aux concentrations des ions de calcium et magnésium. La présence de ces deux cations dans l'eau tend souvent à réduire la toxicité des métaux. La dureté se mesure en mg de CaCO₃ par litre.

On distingue :

- La dureté totale = [Ca²⁺] + [Mg²⁺]
- La dureté totale = dureté calcique + dureté magnésienne
- La dureté calcique TCa= [Ca²⁺]

1-2-7- Sodium et potassium

Le sodium et le potassium, le sixième et septième élément les plus abondants à l'état naturel sont en proportions très variables.

Le potassium, beaucoup moins abondant que le sodium, et rarement présent dans l'eau à des teneurs supérieures à 20 mg/l. Il ne représente aucun inconvénient particulier bien que le K soit une des sources possibles de radioactivité de l'eau.

1-2-8- Substances eutrophisantes : différentes formes d'azote et de phosphore

L'azote et le phosphore constituent la base nutritionnelle de la quasi-totalité de la végétation. Les concentrations en nitrites (NO₂-), nitrates (NO₃-), ammoniac (NH₃) et ammonium (NH₄+), phosphates (PO₄-), azote (N) et phosphore (P) sont dès lors des paramètres importants pour le suivi de la qualité des eaux de surface. Des concentrations de nitrates et de phosphates trop importantes induisent le phénomène d'eutrophisation (étouffement de la vie aquatique). Ce qui va créer des algues et réduit la pénétration de la lumière dans les couches profondes du cours d'eau ou bien du barrage. Les algues vont consommer l'oxygène dissous dans l'eau sans en créer et l'eau va être mal oxygénée et va perdre en qualité. En absence d'oxygène, des bactéries anaérobies transforment les nitrates et produisent de l'ammoniac (NH₃) ou de l'ammonium (NH₄+). L'ammonium en lui-même n'est pas nuisible. Lorsque le pH augmente, on retrouve de l'ammoniac, un gaz soluble dans l'eau et toxique pour la vie aquatique. (l'IBGE, 2005)

1-2-9- Oxygène dissous (OD) et % de saturation en oxygène :

L'oxygène existe sous sa forme dissoute dans l'eau. La quantité d'O₂ qui est dissolvable varie et dépend fortement de la température et la salinité de l'eau. L'oxygène dissous dans l'eau provient soit de l'atmosphère, soit de la photosynthèse par les végétaux aquatiques, entre autres le phytoplancton. Les besoins en oxygène dissous des différents organismes aquatiques

sont fonction de la température et varient considérablement selon les organismes ; voilà pourquoi il ne serait pas très utile de recommander une seule concentration d'oxygène pour tous les organismes présents dans toutes les eaux. Une très faible concentration d'oxygène dissous a des effets particulièrement nocifs sur les organismes aquatiques. Il n'est pas approprié d'établir une concentration minimale acceptable pour l'oxygène dissous, mais il a été prouvé que des concentrations inférieures à 4 mg/l nuisent à la plupart des organismes aquatiques.

1-2-10 - Charge en matières organiques : demande biochimique en oxygène (DBO) et demande chimique en oxygène (DCO) :

La demande biochimique en oxygène (DBO) représente la quantité d'oxygène utilisée par les bactéries pour décomposer partiellement ou pour oxyder totalement les matières biochimiques oxydables présentes dans l'eau et qui constituent leur source de carbone. L'indicateur utilisé est généralement la DBO5 qui correspond à la quantité d'oxygène (exprimée en mg/l) nécessaire aux microorganismes décomposeurs pour dégrader et minéraliser en 5 jours la matière organique présente dans un litre d'eau polluée. Plus la DBO5 est élevée, plus la quantité de matières organiques présentes dans l'échantillon, est élevée. La demande chimique en oxygène (DCO), correspond à la quantité d'oxygène nécessaire pour la dégradation par voie chimique, effectuée à l'aide d'un oxydant puissant, des composés organiques présents dans l'eau. Elle permet de mesurer la teneur en matières organiques totales (excepté quelques composés qui ne sont pas dégradés), y compris celles qui ne sont pas dégradables par les bactéries. Il s'agit donc d'un paramètre important permettant de caractériser la pollution globale d'une eau par des composés organiques. (Khaymastv, 2018)

1-2-11- Métaux lourds :

Le suivi des concentrations en éléments traces métalliques (ETM) (densité $>$ à 5 g/cm³) est particulièrement important vu leur toxicité et leur capacité de bioaccumulation le long des chaînes alimentaires. Contrairement aux polluants organiques, les métaux ne peuvent pas être dégradés biologiquement ou chimiquement. Les concentrations en cuivre, nickel, chrome, plomb, zinc, cadmium, arsenic sont régulièrement mesurées. Les métaux lourds caractérisent certains types de pollution, comme par exemple :

- la présence de cuivre et de nickel est un signe des rejets provenant d'industries de traitement de surface des métaux ;

- le plomb est lié à des pollutions diffuses (apports dus aux transports routiers et à l'existence de sites industriels désaffectés) ;
- le zinc est évacué par des industries qui pratiquent la galvanisation ou la préparation d'alliages tels que le laiton et le bronze, il est également libéré lors du contact entre les eaux de ruissellement et les matériaux galvanisés (toitures métalliques, gouttières) ;
- Le fer ferreux est soluble dans l'eau alors que le fer ferrique l'est très peu et tend à précipiter sous forme d'hydroxyde. Sous l'action de l'oxygène, le fer ferreux s'oxyde en fer ferrique et il est difficile de les doser séparément. Il est tout aussi difficile de distinguer nettement entre fer en solution, en suspension, à l'état colloïdal, sous forme de complexes minéraux ou organiques. (DUPONT, 1981)
- Le cuivre peut se rencontrer dans certaines eaux naturelles généralement à des teneurs inférieures à 1 mg/l sous forme ionique ou de complexes (cyanures, ammoniacale, produits organiques, etc.). En dehors de pollutions industrielles ou de traitements agricoles, ce métal provient habituellement de la corrosion des tuyauteries de distribution (0,5 à 1 mg/l). (RODIER, 1996)
- Le manganèse est généralement associé au fer quoiqu'en quantité plus faible. C'est le Mn^{2+} qui est soluble dans les eaux naturelles bien que le manganèse présente de nombreux autres états d'oxydation.

1-3- Normes algériennes pour les eaux de surfaces :

Afin de gérer et caractériser la pollution des eaux, l'état met en place des normes à vérifier pour les différents paramètres cités préalablement. Ceci va aider dans la gestion et la distribution des eaux et pour préserver les cours d'eaux et la vie aquatique.

Tableau 1-1 : Normes algériennes des eaux de surfaces (JORA n°18 2011)

Paramètre	Norme
pH	6,5-9
O ₂ (mg/l)	<5
Cond (µS/cm)	<2800
Mes (mg/l)	<2000
Ca ²⁺ (mg/l)	<200
Mg ²⁺ (mg/l)	<10
Na ⁺ (mg/l)	<200
K ⁺ (mg/l)	<12

Cl⁻ (mg/l)	<500
SO₄²⁻ (mg/l)	<400
HCO₃⁻ (mg/l)	90-500 ^a
NO₃⁻ (mg/l)	<50
NO₂⁻ (mg/l)	<0.2
NH₄⁺ (mg/l)	<0.5
PO₄³⁻ (mg/l)	<5
DCO (mg/l)	<90
DBO₅ (mg/l)	<30
Cu (mg/l)	<2
Zn (mg/l)	<10
Fe (mg/l)	<0.3
Mn (mg/l)	<0.05
RS (mg/l)	<1500

Chapitre 2 : Théories d'ACP et de clustérisation

Introduction :

L'analyse des données pour une série qui contient plusieurs variables a toujours posé un problème. Etant donné que l'espace graphique ne peut contenir que 3 axes, et tenant en compte que la représentation graphique reste toujours la meilleure façon pour la visualisation. L'analyse en composantes principales est venue avec une solution. Elle permet de regrouper les paramètres les plus pertinents dans le même espace graphique en effectuant des corrélations entre les différents paramètres de la série de données.

2-1- Analyse en composantes principales

2-1-1- Définition

L'analyse en composantes principales est une méthode factorielle d'analyse de données. Elle consiste à représenter graphiquement des valeurs d'un tableau rectangulaire $m \times p$ tel que m représente les variables quantitatives et p les individus. Les données peuvent être extraites d'une population ou bien d'un échantillon.

Les représentations des unités permettent de voir s'il existe une structure non connue a priori, sur cet ensemble d'unités de façon analogue, les représentations des variables permettent d'étudier les structures de liaisons linéaires sur l'ensemble des variables considérées. Ainsi on cherchera si l'on peut distinguer des groupes dans l'ensemble des unités en regardant quelles sont les unités qui se ressemblent, celles qui se distinguent des autres. Pour les variables, on cherchera quelles sont celles qui sont très corrélées entre elles, celles qui au contraire ne sont pas corrélées aux autres. (THARRAULT, 2008)

2-1-2- Notions d'individu et de caractère

2-1-2-1- Individu

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

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

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

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

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

2-1-2-2- Caractère

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

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

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

2-1-3- Définition géométrique

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

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

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

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

2-1-4- Objectifs

L'Analyse en Composantes Principales a pour objectifs :

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

2-1-5- 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. (XLstat ACP)

2-1-6- Procédé d'application de l'ACP

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

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

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

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

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

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. (BENSALEM, 2018)

2-1-7- Analyse des résultats

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

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

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

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

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

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

Tableau 2-1 : Classification des résultats en fonction du KMO (BENSALEM, 2018)

Inacceptable	< 0.5
Médiocre	$0.5 < \text{KMO} < 0.6$
Moyen	$0.6 < \text{KMO} < 0.7$
Bien	$0.7 < \text{KMO} < 0.8$
Très Bien	$0.8 < \text{KMO} < 0.9$

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

Tableau 2-2: Classification des résultats par le test de Bartlett (BENSALEM, 2018)

Très significatif	$\rightarrow 0$
Significatif	< 0.05
Acceptable	$0.05 < I < 0.1$
Rejeté	> 0.1

• Combien de facteurs retenir ?

Trois règles sont applicables :

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

2ème règle : On choisit le nombre d'axes en fonction du cumulé de la variabilité, on s'arrête dans l'axe qui comprend 80% de la variabilité.

3ème règle : Le « Scree-test » ou test du coude. On observe le graphique des valeurs Propres et on ne retient que les valeurs qui se trouvent à gauche du point d'inflexion. Graphiquement, on part des composants qui apportent le moins d'information (qui se trouvent à droite), on relie par une droite les points presque alignés et on ne retient que les axes qui sont Au-dessus de cette ligne. (BENSALEM, 2018)

2-2- La Classification Ascendante Hiérarchique

2-2-1- Définition

La classification ascendante hiérarchique est une méthode statistique qui permet de partitionner une population en différentes classes ou sous-groupes.

Elle fonctionne avec les deux principes d'homogénéité intra-classe qui consiste à regrouper les individus les plus semblables dans la même classe pour obtenir les classes les plus dissemblables d'où le principe de l'hétérogénéité inter-classes.

Ce regroupement d'individus semblables se fait en calculant la distance entre chaque paire d'individus et qui vont par la suite s'exprimer en matrice de distance. Plus cette distance est petite plus sera la chance que les deux individus se retrouvent dans la même classe et vice versa.

2-2-2- Avantages de la CAH :

La classification ascendante hiérarchique (CAH) est une méthode de classification qui présente les avantages suivants :

- On travaille à partir des similarités/dissimilarités entre les objets que l'on veut regrouper. On peut donc choisir un type de similarités /dissimilarité adapté au sujet étudié et à la nature des données.
- L'un des résultats est le dendrogramme, qui permet de visualiser le regroupement progressif des données. On peut alors se faire une idée d'un nombre adéquat de classes dans lesquelles les données peuvent être regroupées. (CAH Xlstat)

2-2-3- Résultats de la CAH :

- **Le dendrogramme :**

Le dendrogramme est une arborescence qui affiche les groupes formés par le regroupement des observations à chaque étape et leurs niveaux de similarité. Le niveau de similarité est mesuré le long de l'axe vertical (alternativement, vous pouvez afficher le niveau de distance) et les différentes observations sont répertoriées le long de l'axe horizontal. (Mini tab)

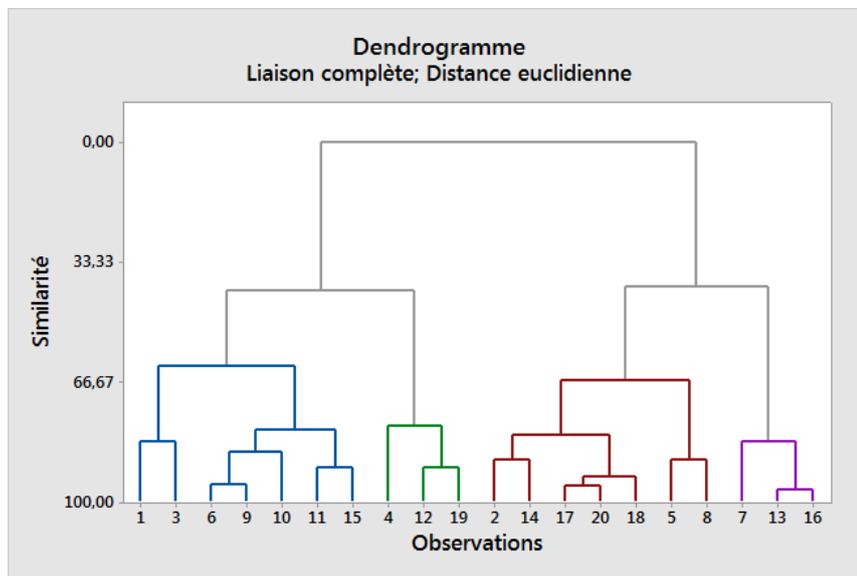


Figure 2-1 : Un dendrogramme (Mini tab)

- **Le tableau**

Un autre résultat de la CAH effectuée avec le logiciel XLstat est le tableau qui regroupe les paramètres de chaque classe pour un niveau de similarité/dissimilarité choisi à l'amont de la classification.

2-3- K-Means :

2-3-1- Définition

C'est un algorithme de clustering qui prend un jeu de données pour former des clusters. Ces clusters sont des données similaires dans leurs caractéristiques.

Cette similarité est mesurée par la distance qui sépare deux données. Plus la distance est petite plus la similarité est grande. Pour chaque cluster, on a un centroïde. Cette donnée autour de

laquelle se regroupent les données similaires. Le choix du centroïde se fait en se basant sur la distance qui le sépare des données qu'ils l'entourent, mais aussi de la moyenne descriptive du groupe. Donc l'algorithme répète les deux opérations jusqu'à obtenir un résultat optimal et stable.

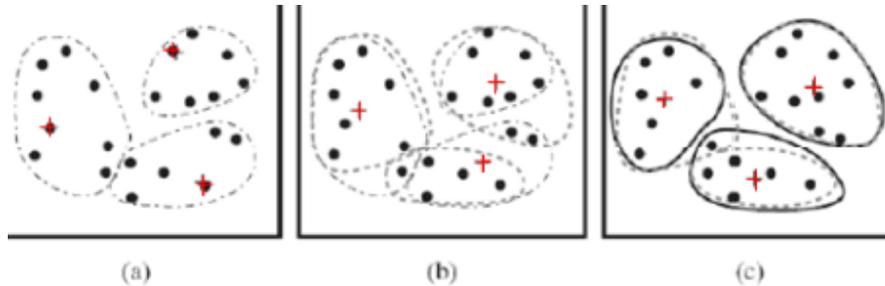


Figure 2-2 : Principe de la méthode de clustering par K-Means (BENSALEM, 2018)

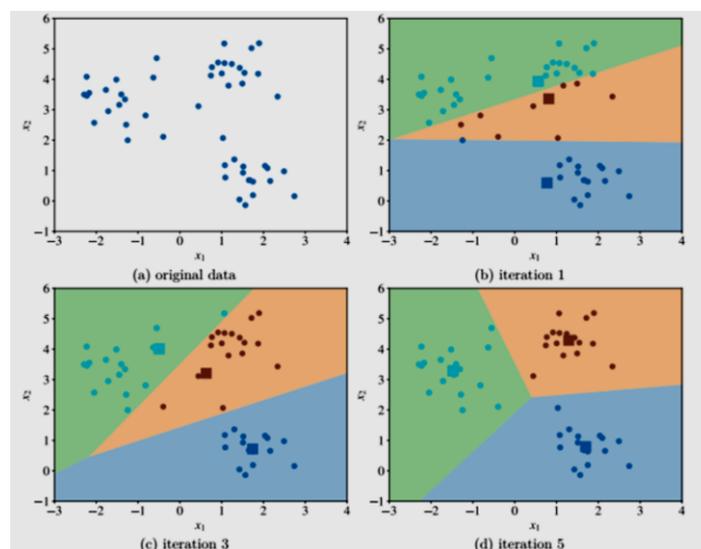


Figure 2-3 : La progression de l'algorithme pour K=3 (BENSALEM, 2018)

2-3-2- Algorithme

Entrée

Ensemble de N données, noté par x

Nombre de groupes souhaité, noté par k

Sortie

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

Début

1) Initialisation aléatoire des centres C_k ;

Répéter

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

$$x_i \in c_k \text{ si } \forall j |x_i - \mu_k| = \min |x_i - \mu_j|$$

Avec μ_k le centre de la classe K ;

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

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

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

Fin.

2-3-3- Avantages de K-means

On peut citer les avantages suivants :

- Un algorithme simple à utiliser et facile à comprendre
- Applicable à un grand nombre de données
- Réduit l'erreur quadratique

2-3-4- Inconvénients

- Le nombre de classe doit être fixé au départ.
- Ne détecte pas les données bruitées.
- Le résultat dépend de tirage initial des centres des classes.
- Les clusters sont construits par rapports à des objets inexistantes (les milieux)
- N'est pas applicable en présence d'attributs qui ne sont pas du type intervalle (Bradley, 1998).

Chapitre 3 : Présentation de la zone d'étude

Introduction

Le but de ce travail est la caractérisation de la pollution des eaux qui arrive au barrage de Beni Haroun, et pour mieux interpréter les résultats que l'on va obtenir, il est judicieux de bien connaître la région et d'avoir une vision globale sur les sources de pollution qui peuvent y exister et les rejets principaux. Cela nous facilitera, sans nul doute, la caractérisation de la qualité des eaux de ce barrage.

3-1- Barrage de Béni Haroun

Pour faire face à une demande en eau de plus en plus croissante de quelques wilayas de l'Est du pays, l'état algérien a construit le plus grand projet de transfert des eaux de surface, le barrage de Beni Haroun. Avec une capacité de stockage de 960 millions de m^3 , il est considéré comme la plus grande réserve d'eau en Algérie.

Ce barrage est situé dans la Wilaya de Mila, au Nord-Est de l'Algérie sur l'Oued El Kebir, à une quarantaine de kilomètres de son embouchure dans la mer méditerranée. La retenue créée par le barrage se trouve au Sud de ce dernier. Le barrage est situé à 4 km du confluent de l'Oued Rhumel et de l'Oued Enndja. Le bassin total de l'Oued Kebir au site du barrage couvre une superficie voisine de 7.725 km^2 .

Il est destiné à régulariser un volume annuel de 504 Hm^3 (dont 435 de son volume propre et 69 du barrage de Boussiaba) pour alimenter en eau potable, et en irrigation les régions des hautes plaines du CONSTANTINOIS et des AURES à travers les Wilayas de Jijel, Constantine, Mila, Oum El Bouaghi, Batna et Khenchela. Il constitue l'ouvrage clé du projet de transfert de Béni Haroun dont l'objectif est de transférer les ressources mobilisées par le barrage de Beni Haroun vers ces six Wilayas.

Le barrage de Beni Haroun est un barrage poids en BCR (Béton Compacté au Rouleau) avec un volume utile de 732 hm^3 , entre les cotes 172 et 200 m.

L'objectif de cet aménagement est :

1. L'alimentation en eau potable de cinq wilayas : Mila, Constantine, Batna, Khenchela et Oum el Bouaghi à raison de 204 millions de mètre cubes.
2. L'irrigation d'environ 40.000 hectares dans les plaines de Teleghma, Chemoura et Toufana à raison de 262 millions de mètres cube.
3. Le laminage des crues de l'Oued Kébir pour la protection des infrastructures à l'aval.

Cette distribution est assurée par 5 grands couloirs d'adduction, 3 stations de traitement et plusieurs de station de pompage. Il a été mis en service en 2003.

3-2-Caractéristiques du bassin versant

Le bassin versant du barrage de Béni Haroun couvre une superficie de 5320 km², soit plus de 60% de la superficie totale du grand bassin Kebir Rhumel noté 10 par l'Agence Nationale des Ressources Hydrauliques (ANRH), qui lui s'étend sur une superficie totale de 8800 km², et dont il fait partie. Ce dernier se divise en deux parties distinctes, l'une occidentale, constituée du bassin de l'Oued Ennja qui se caractérise par des précipitations relativement élevées et d'une topographie montagneuse, et l'autre orientale, constituée par le bassin de l'Oued Rhumel qui se caractérise, par une topographie plus douce et des précipitations modérées. Les deux cours d'eaux confluent près de l'extrémité amont de la gorge Beni Haroun pour former l'Oued El Kebir. (MAROUF, 2012)

Cette entité hydrographique est composée de quatre (04) sous bassins versant, (Tableau III-1), (Figure III-1).

Tableau 3-1 : Les sous bassins et leurs superficies (MAROUF 2012)

Sous bassin	Superficie km ²)
Rhumel-Amont	1130
Rhumel-Seguen	1146
Boumerzoug	1860
Rhumel-Smendou	1184
Total	5320

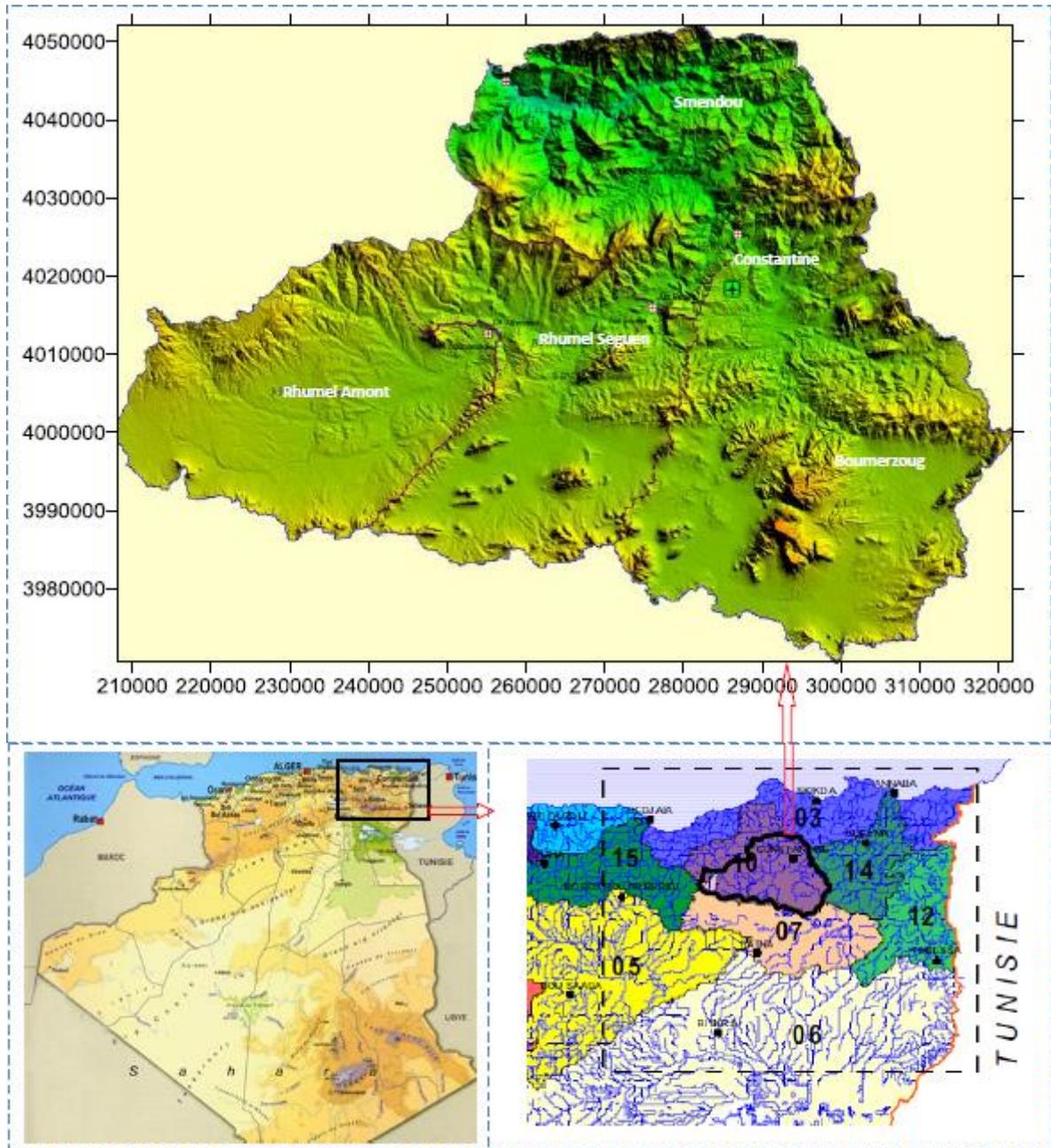


Figure 3-1: Situation géographique et carte du bassin versant Béni Haroun. (BRAHIMI, 2017)

3-2-1- Le relief

Le bassin versant de Beni Haroun est constitué d'un relief compartimenté, caractérisé par des altitudes allant de 15 m à 1729 m, et il se range dans la classe de relief modéré. La majorité des formations du bassin ont une perméabilité assez faible, caractérisé par une couverture végétale discontinue. (HAFSI, 2016)

3-2-2- les altitudes

Les altitudes dans le bassin de notre étude varient de 200 à 1700 m. Le tableau ci-dessous, montre les tranches d'altitudes dans le bassin.

. La classe maximum de la courbe des fréquences altimétriques, se situe entre 800 et 900 m, avec une superficie de 1461,2 km² de l'aire réceptrice du bassin.

Tableau 3-2- : Les classes des altitudes dans le bassin (Kerdoud, 2006)

Classes d'altitudes	Superficie (km²)
<200	20,81
200-300	69,19
300-400	178,07
400-500	263,23
500-600	253,55
600-700	446,38
700-800	1337,93
800-900	1461,32
900-1000	1145,22
1000-1100	88,55
1100-1200	42,34
1200-1300	18,51
1300-1400	2,90
>1400	5328

3-2-3- La géologie :

La géologie du bassin versant de Béni Haroun par sa complexité montre une opposition tant topographique que lithologique avec un Atlas tellien au nord à limite bien nette, contre les hautes plaines au sud à limite floue. La superposition de roches dures et perméables telles que calcaires et grés, fissurés et tectonisés, sur des roches tendres et imperméables comme les argiles et les marnes, (ou inversement), joue un rôle non négligeable dans le déclenchement des processus érosifs.

3-2-4- Le climat:

La situation du bassin Beni Haroun lui confère un caractère original, avec au Nord un climat humide et pluvieux où les précipitations moyennes annuelles atteignent 1000 mm, et au Sud un climat de type semi-aride où les précipitations moyennes annuelles tournent autour de 300 mm. Ce contraste trouve son originalité dans de multiples facteurs comme l'influence Méditerranéenne au Nord et Continentale au Sud ainsi que la répartition des ensembles montagneux.

Les précipitations constituent évidemment, le paramètre climatique essentiel. Leur intensité, leur continuité et leur périodicité sont l'origine même de l'écoulement, de sa localisation et de sa violence

Dans certaines régions, où le régime des écoulements est fortement lié aux précipitations.

La pollution mesurée par milligrammes et par litres est modifiée lors des fortes pluies. En effet, les concentrations de certains polluants diminuent par le phénomène de dilution, d'autres augmentent (comme la pollution par les matières en suspension d'origine érosive).

Tableau 3-3- : Caractéristiques statistiques des données mensuelles climatiques, apports liquides et volume de la cuvette du barrage Béni Haroun (BRAHIMI, 2017)

Paramètre	Max	Min	Moy	Ecart-type	Cv (%)
Précipitations (mm)	253	0	55.66	4.48	99.5
Température (°C)	29.52	4.78	16.55	0.61	42.4
Evaporation (Hm³)	10.48	0.09	3.03	0.23	87.1
Apport liquide (Hm³)	501.55	2.53	77.91	8.99	132.7
Volume de la cuvette (Hm³)	1018.4	40.5	604.9	27.67	47.8

3-3- Caractéristique des stations :

3-3-1- Ain Smara :

C'est une commune dans la daïra d'El Khroub, wilaya de Constantine. Elle se caractérise par une forte densité de population et une activité industrielle naissante. Plusieurs industries de différents domaines sont implantées près d'oued Rhumel. Par exemple les industries de construction mécanique d'Ain Smara et d'Oued Hamimime, les unités de textile d'Ibn Badis,

le dépôt d'hydrocarbures d'El Khroub, les unités de matériaux de construction et autres activités qui polluent les cours d'eau et donc le barrage avec leurs rejets nocifs .Malgré l'existence d'une station d'épuration dans la région (la station d'épuration Ibn Ziad), les rejets y sont toujours. Sans oublier l'activité agricole dans la région.

3-3-2- Menia :

C'est une station implantée sur un autre affluent du barrage. Elle est plutôt caractérisée l'amont par une forte activité agricole, vu l'étendu des terrains à l'amont. Mais aussi on peut repérer la cimenterie de Hammam Bouziane, qui se situe juste avant la station. Donc on s'attend à retrouver des traces de la pollution causée par cette industrie.

3-3-3- Le barrage :

Contrairement aux deux autres stations qui se retrouvent sur un cours d'eau en mouvement, le barrage est un stockage donc il regroupe tous les types et les traces de pollution qui sont collecté par les cours d'eau. Donc on peut s'attendre à ce que la pollution qui se trouve dans le barrage soit globale et plus apparente.

3-3-4- Activité industrielle de la région

Tableau 3-4 : Liste des activités industrielles principales de la région et leurs principaux rejets (KERDOUD, 2006)

Secteur	Définition du secteur	Etablissement concernés	Principaux rejets
Agroalimentaire	Toutes les activités agroalimentaires telles que : conserveries, produits de boissons, Sucrieries, distilleries, levureries, laiteries et activités dérivées, abattoirs, biscuiteries, huileries, confiseries, traitement et transformation de tabacs, etc	ERIAS (Grarem Gouga, Hamma Bouziane, El Khroub), Laiterie Numédia, SNTA, SO.AL.CO, Limonaderie PULPA, E.B.G.S (ex- EMIB), Abattoir Municipal	Pollution organique biodégradable, DCO, DBO avec un pH acide

		d'El Khroub	
Industries Minérales	Fabrication de ciment ou de chaux, fabrication de céramique, fabrication d'amiante-ciment, verreries, briqueteries, centrales à béton, etc.	SOFEM, ENATUB, S.C.O.A, SO.M.I.C, S.C.C Ibn Ziad, E.R.C.E Hamma Bouziane	MES
Tanneries et Industries de Cuir	Tanneries et Mégisseries	Tannerie Hadj Arab.	SO4, MES
Bois- Papier -Carton	Transformation du Bois, Fabrication de Pâte, Fabrication de Papier, ou Carton à partir de fibres primaires ou recyclées, etc.	Néant	DCO, DBO, Ca, Mg, NH4, Na, PO4
Industries Textiles	Transformation liés à la Fabrication ou à la Transformation de fibres textiles, naturelles ou synthétiques telles que, ennoblissement, blanchiment, Apprêt,	COTEST (ex - COTITEX)	DCO, DBO, MES pH basique

	Teintureries, Filature, Blanchisseries, etc.		
Chimie-para chimie- Pétrole	Toutes activités de fabrication ou de transformation de substances chimiques ou organiques ou minérales, y compris le raffinage et le stockage d'hydrocarbures, la fabrication d'engrais, la fabrication de produits pharmaceutiques, la fabrication de peintures, la fabrication de vernis ou de résines, la fabrication de produits phytosanitaires, la production ou la transformation de matières plastiques, etc.	E.N.A.D-SO.DE.R, E.N.G.I, NAFTAL.	SO4, M.O, MES, Cl, NH4
Sidérurgie et Métallurgie	Activités liées à la fabrication ou la transformation des métaux ferreux ou non ferreux.	P.M.A, P.M.O, CPG, CCA, GERMAN (ex-CCE), CAB.AM, SNVI	NH4, Zn, Fe, Cu

3-3-5- Densité urbaine :

Le bassin de Beni Haroun compte quelque trente-deux (32) agglomérations de plus de 5000 habitants. Elles abritent un total de 1,06 millions d'habitants selon le recensement de la population et de l'habitat de 1998. Les plus fortes concentrations urbaines se limitent à Constantine (462.187 habitants), El Khroub (65.344 habitants), Mila (54.557 habitants), Ain M'lila (50.744habitants), Chelghoum Laid (41.417 habitants) et Ain Fakroun (40.701 habitants).

Généralement, ces centres urbains sont branchés au réseau d'alimentation en eau potable avec un taux de raccordement variant de 70 à 98%.

Les principaux rejets d'une agglomération urbaine sont comme mentionné dans le premier chapitre essentiellement de la pollution organique et la pollution azotée ainsi que les résidus des détergents.

Le traitement des eaux rejetées dans le bassin se fait dans deux stations de traitement. La première, située à Chelghoum Laid, est destinée à traiter les eaux rejetées par cette agglomération. La seconde, située à Ibn Ziad, est destinée à épurer les eaux rejetées par la ville de Constantine. Ces dernières fonctionnent partiellement et le débit traité par les stations en question est évalué alors à 23918 m³/j soit un taux de dépollution de 32%. (KERDOUD, 2006)

Tableau3-5: Besoins, consommations et déficits (Agglomération supérieur à 5000 hab) (KERDOUD, 2006)

Année	Besoins Hm ³ /an	Consommations Hm ³ /an	Déficits Hm ³ /an
1999	49.37	34.51	14.86
2000	50.93	36.69	14.24
2001	52.47	35.32	17.14
2002	53.39	32.39	21.00
2003	54.82	32.39	22.43

3-3-6 L'activité agricole :

Autour du barrage de Beni Haroun, on retrouve quatre (4) périmètres d'irrigation, seulement deux (2) sont opérationnels : le périmètre de Fourchi à Ain M'lila, et celui de Hamma Bouziane. D'après les données collectées au niveau des différentes directions et services agricoles (D.S.A), la superficie irrigable du périmètre de Fourchi est estimée à 616 hectares, dont 1801 hectares irrigués. Ce périmètre satisfait ses besoins grâce aux ressources

souterraines: deux (02) forages d'un débit d'ensemble de 300 l/s. Malgré la disponibilité dans ce périmètre en eau, ou les besoins sont pratiquement égaux à la demande 1200 hm³/an, il n'en demeure qu'il rencontre quelques difficultés dues à la détérioration du réseau d'alimentation à presque 70%. L'agriculture et l'élevage sont responsables du rejet de nombreux polluants organiques et inorganiques dans les eaux de surface et souterraines provenant de l'érosion des terres agricoles, de composés azotés issus des déchets des animaux et engrais chimiques ou naturels utilisés pour la production agricole. (KERDOUD, 2006)

3-4- Impact des rejets des eaux usées sur la qualité des eaux

La majorité des rejets des trente deux (32) agglomérations et de l'industrie de la région, déversent leurs eaux usées dans les cours d'eau du bassin. Ces rejets domestiques et industriels dans les cours d'eau du bassin et dans la retenue du barrage favorisent l'enrichissement de l'eau en éléments nutritifs, phosphore et azote et notamment l'augmentation corrélative de la production végétale aquatique, principalement des algues planctoniques ou fixées, ce qui altère les propriétés physiques et chimiques de l'eau. (Decamps,1996)

Les uniques stations de dépollution (Chelghoum El Aid et Constantine) fonctionnent partiellement, la dépollution ne dépasse pas les 32% (Kerdoud, 2006), et en dehors de toutes mesures de protection de l'environnement, la ressource en eau de surface se trouve exposée à des risques potentiels d'altérations chroniques ou épisodiques. Ce qui a contribué à la dégradation de la qualité des eaux du barrage de Beni Haroun. Pour évaluer l'impact de ces affluents sur le qualité de ses eaux, une étude des différentes pollutions (minérale, azotés, ...) en comparaison avec la grille de qualité de l'OMS a été menée afin de classer les eaux du barrage.

3-4-1- Données et prélèvements des échantillons :

Vu l'importance du volume stocké dans le barrage de Beni Haroun, le contrôle de la qualité de ses eaux est devenu primordial. Quatre stations d'analyse des paramètres physico-chimiques et bactériologiques sont installées par l'ANRH, au bord de la retenue du barrage (Ain Smara, Hammam Grouz, Grarem, Menia), et des analyses journalières sont effectuées par l'ANRH de Constantine.

Tableau 3-6 : Coordonnées des stations (Kerdoud, 2006)

Code station	Nom station	Coordonnées Lambert	
		X(m)	Y(m)
100403	Ain Smara	841,94	335,80
1006 24	El Menia	849,300	348,40
1006 01	Amont Barrage Beni Haroun (Grarem)	821,550	363,35
1003 15	Hammam Grouz	822,000	322,100

Malheureusement, nous n'avons pu acquérir que les données des trois premières stations du tableau III-6 sur une période d'observation s'étalant de 2000 à 2010 (contenant 22 paramètres mesurés), et une autre série de valeurs moyennes annuelles s'étalant de 1994 à 2003 (contenant 7 paramètres mesurés).

Ces trois stations sont représentées sur les cartes ci-dessous :

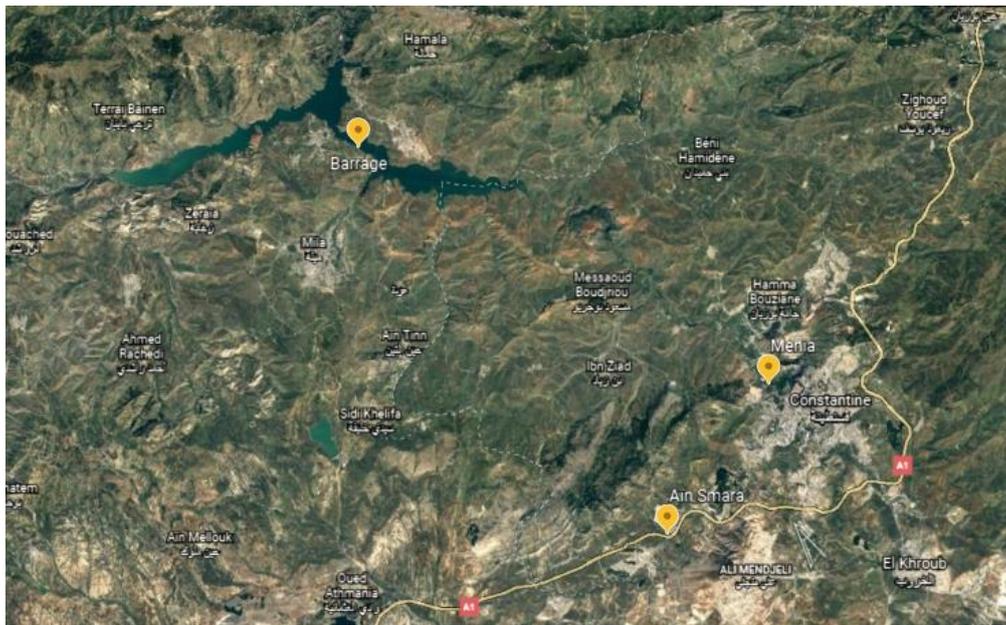


Figure 3-2 : Carte de l'emplacement des trois stations

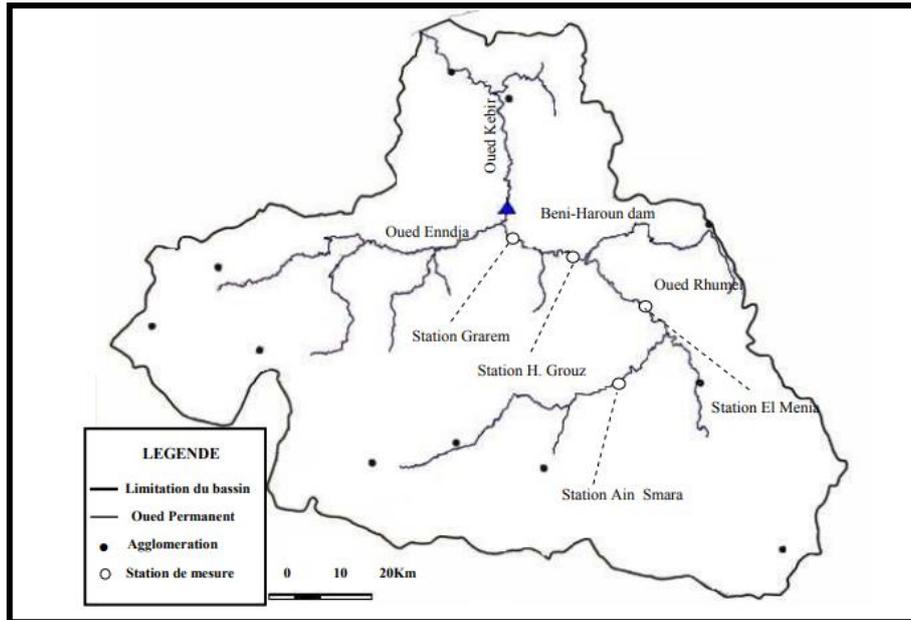


Figure 3-3 : Situation géographique des stations de mesure dans le bassin Kebir-Rhumel (MAROUF, 2012)

Chapitre 4: Collecte et critique des données

Introduction

Pour avoir de bons résultats issus d'une méthode de classification, il faut tout d'abord effectué un prétraitement de données. Ce traitement va permettre d'avoir une série plus homogène avec moins d'incohérences et d'aberrations, pour aboutir à des résultats plus significatifs à l'aval des méthodes de clustring.

4-1- Critique de la donnée :

Les données dont on dispose sont des mesures des paramètres caractéristiques de la qualité de l'eau dans 3 stations :

- Ain Smara

La série recueillie s'étale du 16/01/2000 au 14/06/2010. Les mesures sont des mesures journalières avec une fréquence d'une mesure par mois, donc en théorie on doit avoir 126 valeurs mais on ne dispose que de 98, ce qui fait un taux de dates sans mesure de 22.22%.

Le tableau suivant regroupe les caractéristiques statistiques de chaque paramètre.

Tableau 4-1: Caractéristiques statistiques Ain Smara

Paramètre	Unité de mesure	Min	Max	Moyenne	Ecart type	Cv(%)	Taux de valeurs manquantes (%)
T	°C	5,00	30,00	15,97	6,62	41,46	0
pH	/	6,90	8,81	7,76	0,38	4,94	0
Cond	S.m ⁻¹	219,00	2800,00	1874,71	479,22	25,56	0
MES	mg/l	18,00	1970,00	194,36	264,48	136,07	2,04
Rs	mg/	161,20	1942,00	1256,03	349,70	27,84	4,081
Ca	mg/l	38,00	255,10	165,26	49,01	29,66	0
Mg	mg/l	6,00	166,00	59,42	27,06	45,54	0
Na	mg/l	19,40	550,00	147,94	66,84	45,18	0
K	mg/l	0,00	174,00	8,05	17,44	216,54	1,02
Cl	mg/l	34,00	610,00	231,11	88,66	38,36	0
SO4	mg/l	100,00	760,00	376,79	109,63	29,10	0
HCO3	mg/l	113,90	1870,80	312,07	181,40	58,13	0
NO3	mg/l	0,00	98,00	24,66	16,58	67,25	1,02
NO2	mg/l	0,00	4,00	0,99	0,82	83,09	3,06
NH4	mg/l	0,00	90,00	4,14	11,16	269,75	1,02
PO4	mg/l	0,00	20,00	2,05	2,71	132,19	1,02
DCO	mg/l	6,80	607,20	76,30	80,55	105,56	2,04
DBO5	mg/l	0,10	150,00	11,83	18,38	155,41	1,02

O2-dis	mg/l	0,00	14,50	7,42	2,26	30,48	31,63
Cu	mg/l	0,00	0,10	0,01	0,02	191,91	22,448
Pb	mg/l	0,00	0,25	0,01	0,05	519,62	72,448
Zn	mg/l	0,00	0,79	0,03	0,10	391,05	21,428
Fe	mg/l	0,00	4,59	0,66	0,95	143,49	21,428
Mn	mg/l	0,00	2,56	0,12	0,30	240,88	22,448

/

En plus de la série de donnée citée préalablement, on dispose d'une deuxième série de donnée avec un pas temps annuel (moyenne annuelle par paramètre) pour les années de 1994 à 2010 et un nombre de paramètres restreint. La série est présentée dans le tableau suivant.

Tableau 4-2: Série de données (1994-2010) Ain Smara

	Matière organique et oxydable		Matière azotées		Matière phosphorée	Particules en suspension	Minéralisation
	O2 Dissous	DCO	NO2	NO3	PO4	MES	Conductivité
1994	2.489	204	0,4	50	4,25	12,1	2200
1995	1.72	340	0,15	6	18	10	2200
1996	1.339	220	0,13	50	15	17,15	2300
1997	3.929	516	0,17	52,2	15	12,19	2200
1998	3.959	236	0,69	6,85	19,2	10,6	2700
1999	0.29	133	3,51	6,4	7,59	8,39	2600
2000	6,0	75,0	0,4	12,3	3,6	126,7	1733,3
2001	7,9	91,0	1,0	28,3	2,9	135,0	2300,0
2002	31,5	336,0	0,4	24,0	8,0	32,0	2800,0
2003	6,9	83,0	0,7	20,3	3,3	130,8	2016,7
2004	6,9	83,0	0,7	20,3	3,3	130,8	2016,7
2005	7,9	66,8	1,0	29,9	1,1	187,7	1829,1
2006	7,9	53,4	0,8	26,7	0,9	133,0	1706,6
2007	8,5	46,0	1,4	36,4	2,6	352,3	2050,0
2008	9,2	56,4	1,3	35,4	2,1	117,4	2003,0
2009	7,5	63,7	0,9	20,9	1,1	272,2	1659,3
2010	6,7	52,7	1,0	34,5	1,5	84,7	1935,0

- Menia

On a 107 mesures brutes s'étalant du 16/01/2000 au 03/06/2010. En théorie on doit avoir 126 valeurs mais on ne dispose que de 107 ce qui fait un taux de dates sans mesure de 15%.

Le tableau suivant regroupe les caractéristiques statistiques de chaque paramètre.

Tableau 4-3 : Caractéristiques statistiques Menia

Paramètres	Min	Max	Moyenne	Ecart type	CV (%)	Taux de valeurs manquantes (%)
T	6	35	18,77	6,60	35,16	0
pH	7	8,77	7,75	0,35	4,52	0
Cond	1,3	13000	1955,99	1135,31	58,04	0
MES	7	2702	253,45	401,08	158,25	0
Rs	436	1942	1194,75	224,50	18,79	0,93
Ca	37	531	149,13	53,32	35,75	0
Mg	21	142	58,71	25,33	43,14	0
Na	49,8	355	152,51	47,44	31,11	0
K	0	24	9,56	5,59	58,47	0
Cl	20	2055	249,08	190,64	76,54	0
SO4	146	592	296,26	76,68	25,88	0
HCO3	146,4	689	374,14	111,50	29,80	0
NO3	0	64,8	13,14	12,91	98,25	0
NO2	0	8,03	1,41	1,41	100,00	0,93
NH4	0	90	9,88	11,89	120,34	0,93
PO4	0	16,8	5,34	3,69	69,10	0
DCO	12,4	670	92,85	79,53	85,65	0
DBO5	0,5	145	15,80	19,03	120,44	1,86
O2-dis	0,85	40,9	6,51	4,41	67,74	26,16
Cu	0	0,45	0,02	0,05	250,00	21,49
Pb	0	0,38	0,04	0,08	200,00	69,15
Zn	0	0,726	0,04	0,09	225,00	21,49
Fe	0	7,389	0,72	1,15	159,72	19,62
Mn	0	2,004	0,15	0,25	166,67	22,429

Pour cette station aussi on dispose d'une deuxième série de donnée qui s'étale de 1994 à 2010 et un nombre de paramètres restreint. La série est présentée dans le tableau suivant :

Tableau 4-6 : Série de données (1994-2003) Menia

	Matière organique et oxydable		Matière azotées		Matière phosphorée	Particules en suspension	Minéralisation
	O2 dissous	DCO	NO2	NO3	PO4	MES	Conductivité
1994	16.6	270	0.2	44	10.8	16.1	1900
1995	12.39	296	0.28	50	15	14.19	1800

1996	20,29	99	0,13	20	8,39	19	1900
1997	10,1	154	0,25	54	14,3	14,19	1800
1998	21,1	97	0,2	17	13,19	25	2100
1999	30,7	96	1,89	9,6	8,39	12,5	2100
2000	4,77	88,80	1,11	11,00	8,14	11,00	1985,71
2001	5,62	142,11	1,17	10,00	7,94	10,00	1911,11
2002	5,81	276,91	0,80	6,67	7,43	6,67	1700,00
2003	6,79	87,57	1,65	13,14	2,68	13,14	1742,86
2004	6,27	63,77	1,74	19,00	2,95	19,00	1862,50
2005	6,36	64,70	1,80	16,22	3,39	16,22	2031,11
2006	6,36	69,94	2,47	10,33	5,74	10,33	1944,44
2007	6,02	53,64	1,57	5,38	5,98	5,38	2004,44
2008	7,88	65,18	1,34	13,99	4,05	13,99	1742,73
2009	6,89	69,75	0,66	15,30	2,71	15,30	1623,75
2010	6,18	51,60	1,04	33,07	3,50	33,07	1706,67

- Beni Haroun

Pour la station du barrage on dispose de 105 mesures brutes pour la période du 16/01/2000 au 02/05/2010. En théorie on doit avoir 125 valeurs mais on ne dispose que de 105 ce qui fait un taux de dates sans mesure de 16%.

Le tableau suivant regroupe les caractéristiques statistiques de chaque paramètre. Il est à noter que contrairement aux deux autres stations qui mesurent 23 paramètres, la station de Beni Haroun ne mesure que 21 (Les 2 paramètres manquants sont le plomb (Pb) et les résidus secs (Rs)).

Tableau 4-5 : Paramètres statistiques Beni Haroun

Paramètre	Min	Max	Moyenne	Ecart type	Cv(%)	Taux de valeurs manquantes(%)
T	6,5	30	18,86	6,33	33,56	0
pH	7	8,69	7,87	0,30	3,81	0
Cond	800	2700	1508,99	448,30	29,71	0
MES	17	320	141,72	78,77	55,58	0,952
Ca	40	210	117,21	34,87	29,75	0
Mg	17	168	50,34	29,77	59,14	0
Na	45	280	123,61	45,61	36,90	0
K	0	18	6,62	4,58	69,18	0

Cl	45	480	180,27	67,54	37,47	0
SO4	174	680	286,59	83,60	29,17	0
HCO3	109,8	561	256,16	119,80	46,77	0
NO3	0	62	14,18	11,05	77,93	0
NO2	0	2,913	0,52	0,60	115,38	0,952
NH4	0	1,833	0,37	0,49	132,43	0
PO4	0	14	1,75	2,64	150,86	0
DCO	17,5	175,5	83,03	29,81	35,90	0
DBO5	0	27,5	10,79	6,33	58,67	0
O2-dis	1,92	15	7,37	2,26	30,66	29,523
Cu	0	0,3	0,02	0,04	200,00	20
Zn	0	0,86	0,03	0,10	333,33	20,952
Fe	0	3,166	0,55	0,82	149,09	20
Mn	0	1,799	0,11	0,23	209,09	22,857

La série de mesure au pas annuel du barrage s'étalant de 1994 à 2010 est présentée dans le tableau suivant :

Tableau 4-6 : Série de données (1994-2010) Beni Haroun

	Matière organique et oxydable		Matière azotées		Matière phosphorée	Particules en suspension	Minéralisation
	O2 dissous	DCO	NO2	NO3	PO4	MES	Conductivité
1994	55	101	0.2	8	4	6.09	2000
1995	60.7	152	0.44	3.2	3.7	11.8	2100
1996	40.7	91	0.31	6	6.19	12.3	1800
1997	43.59	111	0.62	8.69	8.8	13.1	2900
1998	63.9	91	0.3	8.8	8	14.6	2300
1999	22.89	166	6.09	4.03	6.4	9.1	2300
2000	6,3	72,8	1,1	28,4	5,2	127,2	1980,0
2001	6,5	87,7	0,7	21,6	3,1	120,0	1952,3
2002	6,3	86,4	0,5	12,2	4,5	138,7	1950,0
2003	5,4	58,0	0,4	19,0	1,3	146,7	1733,3
2004	7,1	67,3	0,7	18,5	1,8	166,8	1883,3
2005	7,5	71,0	0,3	14,7	0,3	153,0	1406,7
2006	7,5	87,6	0,6	9,7	0,3	73,6	1250,0
2007	7,0	89,1	0,2	7,0	0,4	181,0	1241,8
2008	8,9	84,4	0,3	7,7	0,3	148,4	1171,7
2009	8,3	92,3	0,2	6,5	0,2	167,9	1118,2

2010	9,4	111,5	0,7	8,0	0,2	198,4	1064,0
------	-----	-------	-----	-----	-----	-------	--------

-Traitement des valeurs manquantes :

Pour des raisons de réduction de la taille d'échantillon qui est à la base pas très grande, l'option adoptée pour combler les vides est de remplacer le manque de la mesure « i » par la moyenne des deux mesures « i-1 » et « i+ 1 », et ce pour toutes les stations.

4-2- Elimination des valeurs aberrantes et extrêmes :

On peut toujours réaliser de longues études statistiques avec les méthodes les plus sophistiquées. Si les données comportent des anomalies, les résultats sont significativement impactés et les conclusions de l'étude risquent de ne pas valoir grand-chose. D'où l'importance de traiter les données en amont. Parmi les anomalies, que peuvent comporter les données, on peut citer les valeurs aberrantes.

En statistique, une donnée aberrante est une valeur ou une observation qui est « distante » des autres observations effectuées sur le même phénomène, c'est-à-dire qu'elle contraste grandement avec les valeurs « normalement » mesurées. ... Ces dernières sont parfois exclues de la série de données.

La technique la plus simple pour leur détection, est l'identification sur graphique (courbe ou nuage de points, en général), ou par utilisation de certains logiciels qui intègrent des algorithmes d'identification.

Nous avons utilisé pour notre part le test de Grubbs, qui est un test statistique qui permet de déterminer si, la valeur la plus grande ou la plus petite, ou dans le cas du test de Grubbs double, si les deux valeurs les plus grandes, ou si les deux plus petites peuvent être considérées comme extrêmes (ou aberrantes).

Pour utiliser le test de Grubb, on calcule dans un premier temps :

- \bar{X} : La moyenne de l'ensemble des mesures
- S : L'écart-type de l'ensemble des mesures
- G : Distance entre la valeur suspectée d'être un point aberrant et la moyenne

$$G = \frac{|\bar{x} - \text{Valeur}|}{S}$$

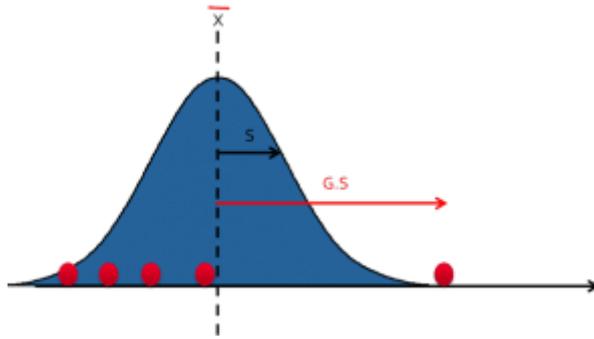


Figure 4-1 : Test de Grubbs (XLstat 2018)

On compare ensuite la valeur de G obtenue à un G limite :

$$G_{limite} = \frac{N - 1}{\sqrt{N}} \cdot \sqrt{\frac{t^2_{\alpha, N-2}}{N - 2 + t^2_{\alpha, N-2}}}$$

Si $G > G_{limite}$ la valeur est considérée comme aberrante et inversement.

Après avoir appliqué le test de Grubbs unilatéral avec un $\alpha = 10\%$, nous avons obtenus plusieurs valeurs extrêmes signalées par le test et déjà remarquées à l'œil nu après avoir représenté sous forme graphique l'évolution temporelle des différents paramètres.

On a utilisé un test unilatéral car les valeurs nulles peuvent faire partie des données observées. Le test de Grubbs a été effectué à l'aide du logiciel Xlstat et voici quelques résultats obtenus pour les différentes stations :

- **Ain Smara**

Pour le cuivre (Cu) : pas de valeurs extrêmes

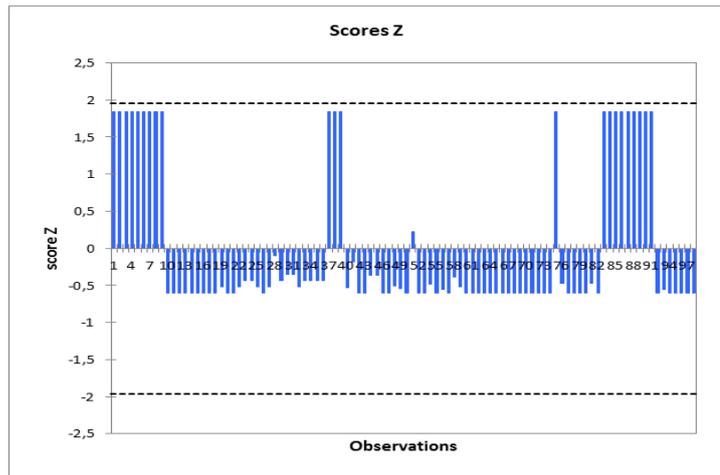


Figure 4-2 : Résultat test de Grubbs paramètre Cu, station Ain Smara)

Pour le sodium (Na) : une seule valeur extrême

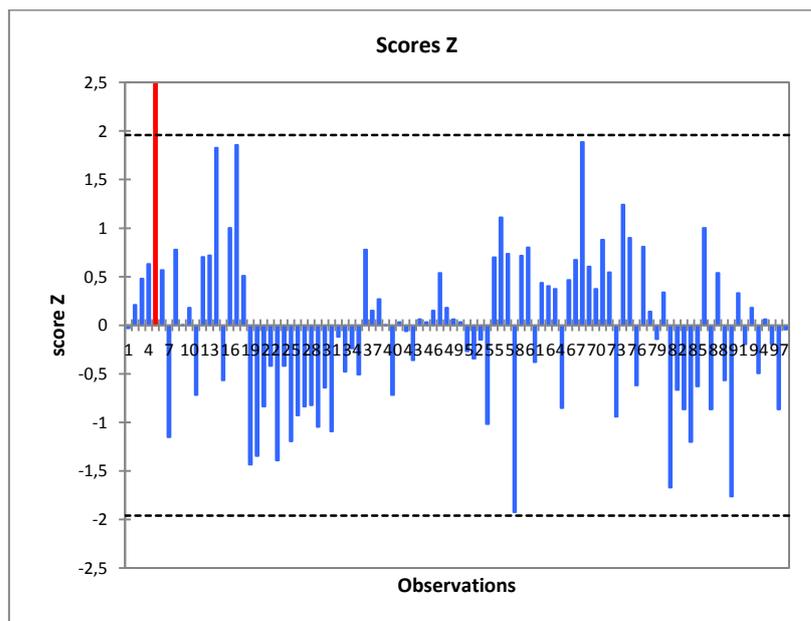


Figure4-3 : Résultat test de Grubbs (paramètre Na, station Ain Smara)

Pour le potassium (K) : par exemple cette valeur est associé probablement à une erreur de saisie (K= 17.4 mg/l)

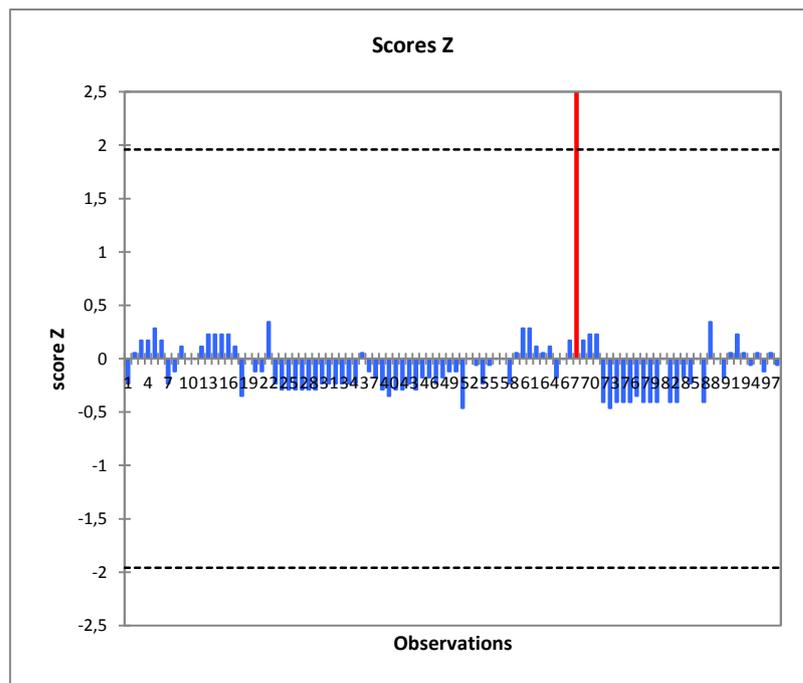


Figure 4-4 : Résultat test de Grubbs (paramètre K, station Ain Smara)

Tableau 4-7: Nombre de valeurs extrêmes par paramètre « Ain Smara »

Nombre de valeurs extrêmes	[0,2[[2,4[[4,6[>=6
Paramètres	T, pH, K, HCO ₃ , Cu, Zn, Mn, O ₂ , Pb	No ₃ , No ₂ , Po ₄ , Dbo, Cond, Rs, Mes, Ca, Cl,	Mg, NH ₄	Fe, SO ₄

Après avoir effectué le test de Grubbs et repérer les valeurs extrêmes, on a procédé à l'élimination des lignes qui les contiennent. Il faut noter qu'on peut avoir plusieurs valeurs extrêmes de différents paramètres pour la même date donc les lignes à éliminer ne sont pas forcément la somme des nombres de valeurs extrêmes par paramètre. La série est passée de 98 lignes à 73 après ce test.

- **Menia**

Pour le chlore (Cl) : Cl=2055 mg/l

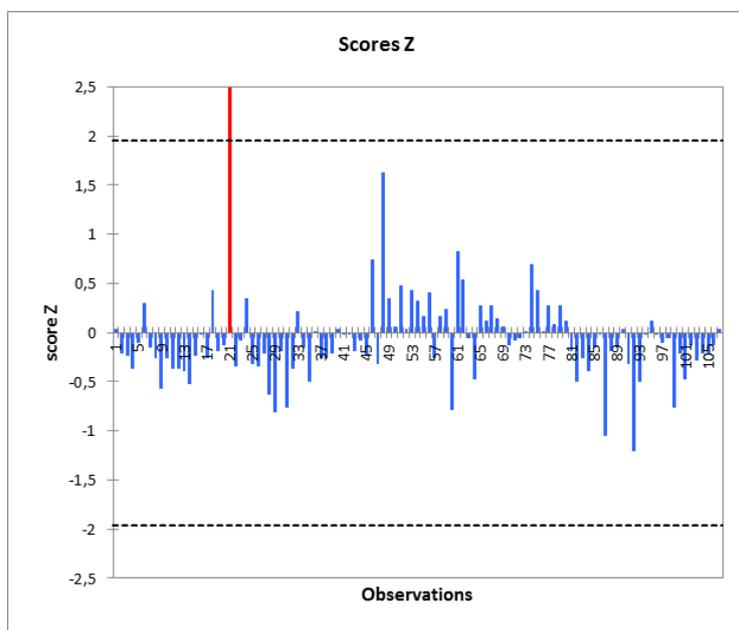


Figure 4-5 : Résultat test de Grubbs (paramètre Cl, station Menia)

Pour l'ammonium (NH₄) : 6 valeurs extrêmes

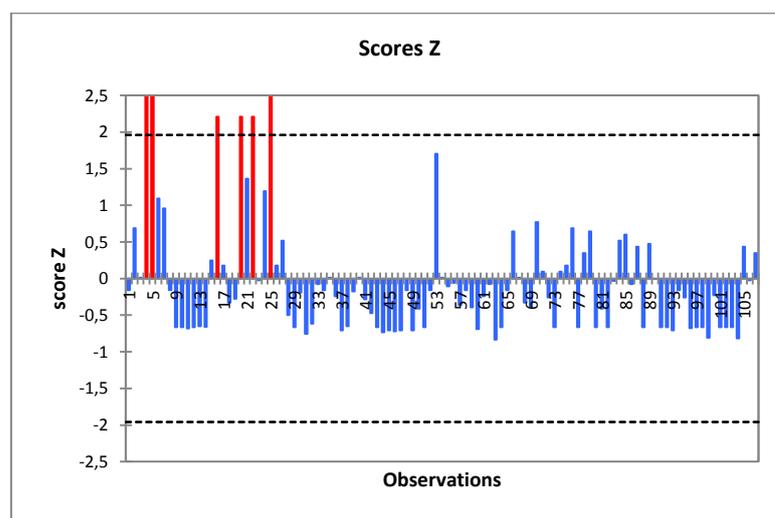


Figure4-6 : Résultat test de Grubbs (paramètre NH₄, station Menia)

Tableau 4-8: Nombre de valeurs extrêmes par paramètre « Menia »

Nombre de valeurs extrêmes	[0,2[[2,4[[4,6[>=6

Paramètres	T, Ca, Cl, K, Pb, O2, Cu, Zn	Cond, Rs, Na, PO4, Mn, Fe, Dco	Mes, SO4, HCO3, NH4	Mg, NO3, NO2
------------	---------------------------------	-----------------------------------	------------------------	--------------

La série est passée de 107 lignes à 86 après application du test de Grubbs.

- **Beni Haroun**

Pour la température : pas de valeurs extrêmes

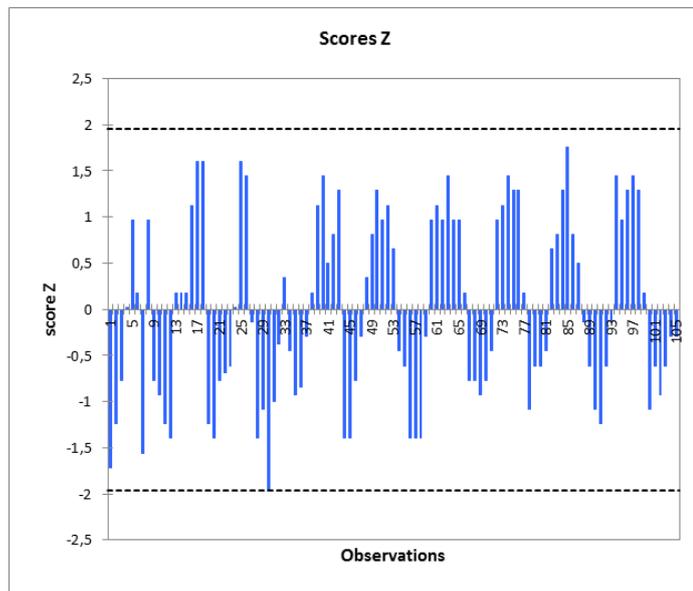


Figure4-7 : Résultat test de Grubbs (paramètre T, station Beni Haroun)

Pour les sulfates (SO4) : 4 valeurs extrêmes

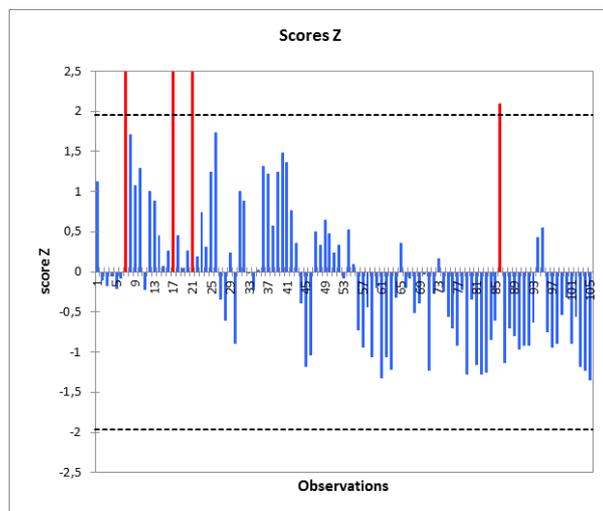


Figure 4-8: Résultat test de Grubbs (paramètre SO4, station Beni Haroun)

Tableau 4-9: Nombre de valeurs extrêmes par paramètre « Beni Haroun »

Nombre de valeurs extrêmes	[0,2[[2,4[[4,6[>=6
Paramètres	T, Cu, Zn, Mn	Cond.,Cl,, Dco, O2	pH, Mes, K, So4, Hco3, No3, Nh4, Po4, Dbo	Ca, Mg, No2, Fe

La série de données est passée de 105 lignes à 89 après application de ce test.

Chapitre 5 : Résultats et discussions

Introduction :

Dans ce chapitre, on va présenter et discuter les différents résultats obtenus. Dans un premier temps, on va décrire toutes les étapes de l'analyse en composantes principales, afin de définir les axes principaux et les paramètres prépondérants sur chaque axe. Ensuite, on va utiliser deux méthodes de clustérisation afin de regrouper les paramètres dans leurs classes correspondantes. Enfin, on va classer les eaux des 3 stations recueillies selon la classification de l'OMS.

5-1- Analyse en composantes principales

Il existe plusieurs méthodes statistiques qui permettent de résumer l'information contenue dans une matrice de données. Parmi celles-ci, l'A.C.P qui permet de détecter une structure dans les relations entre variables, ce qui revient à ordonner les variables. Par conséquent, cette technique est utilisée comme une méthode de synthèse des données (on dit encore réduction)

5-1-1- Application de l'ACP

L'A.C.P appliquée est centrée réduite ceci dans le but d'identifier et de quantifier la liaison entre les différents paramètres, bien que l'A.C.P est une méthode exploratoire et descriptive, elle est largement appliquée pour enquêter sur les impacts environnementaux et les processus hydrogéochimiques à travers le monde. (BRAHIMI, 2017)

Cette analyse a été effectuée à l'aide du logiciel XLSTAT 2017

- Ain Smara

Cette analyse a été réalisée sur un tableau de 23 variables (T, pH, Cond, MES, Rs, Ca, Mg, Na, K, Cl, SO₄²⁻, HCO₃⁻, NO₃⁻, NO₂⁻, NH₄⁺, PO₄³⁻, DCO, DBO₅, O₂-dissous, Pb, Zn, Mn, Cu) et de 73 individus.

- Tableau des valeurs propres

Tableau 5-1 : Tableau des valeurs propres ACP, Ain Smara

	F1	F2	F3	F4	F5	F6	F7
Valeur propre	5,666	3,311	1,980	1,457	1,268	1,204	1,062
Variabilité (%)	24,634	14,395	8,610	6,333	5,512	5,235	4,616
% cumulé	24,634	39,029	47,639	53,972	59,484	64,720	69,335

D'après ce tableau, sept composantes principales ont été retenues en utilisant la règle de Kaiser mentionnée dans le chapitre III, et qui consiste à s'arrêter à l'axe qui contient la valeur propre supérieure à 1.

• Cercle des corrélations

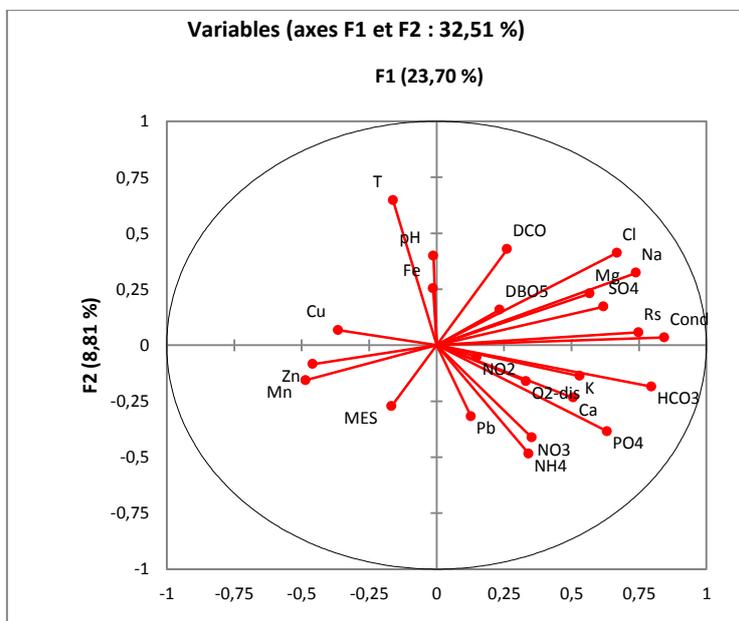


Figure 5-1: Cercle des Corrélations ACP Ain Smara

• Tableau des cosinus

Tableau 5-2 : Tableau des cosinus ACP, Ain Smara

	F1	F2	F3	F4	F5	F6	F7
T	0,026	0,419	0,051	0,000	0,027	0,139	0,013
pH	0,000	0,159	0,015	0,112	0,086	0,212	0,101
Cond	0,713	0,001	0,004	0,005	0,016	0,000	0,001
MES	0,028	0,074	0,019	0,048	0,534	0,015	0,001
Rs	0,560	0,003	0,000	0,000	0,000	0,026	0,008
Ca	0,257	0,055	0,107	0,316	0,034	0,001	0,009
Mg	0,322	0,053	0,132	0,159	0,001	0,010	0,107
Na	0,546	0,105	0,087	0,022	0,005	0,012	0,094
K	0,281	0,019	0,118	0,065	0,006	0,000	0,017
Cl	0,447	0,170	0,084	0,047	0,006	0,015	0,081
SO4	0,382	0,030	0,004	0,194	0,040	0,012	0,094
HCO3	0,635	0,034	0,003	0,036	0,015	0,013	0,001
NO3	0,125	0,171	0,064	0,005	0,014	0,042	0,065
NO2	0,022	0,003	0,009	0,003	0,002	0,411	0,005
NH4	0,116	0,235	0,090	0,044	0,019	0,117	0,001

PO4	0,399	0,148	0,012	0,017	0,060	0,000	0,011
DCO	0,068	0,184	0,249	0,061	0,014	0,004	0,024
DBO5	0,055	0,025	0,317	0,267	0,079	0,008	0,060
O2-dis	0,110	0,026	0,023	0,000	0,010	0,162	0,006
Cu	0,133	0,005	0,019	0,018	0,033	0,087	0,080
Pb	0,016	0,101	0,002	0,079	0,003	0,004	0,148
Zn	0,211	0,007	0,095	0,000	0,201	0,006	0,198
Fe	0,000	0,064	0,287	0,026	0,007	0,000	0,020
Mn	0,236	0,025	0,106	0,015	0,211	0,018	0,120

Dans le premier axe, on retrouve les paramètres de la minéralisation et 2 métaux lourds. Le 2^{ème} axe contient des matières azotées. Le 3^{ème} axe contient les paramètres organiques DBO et DCO avec le Fe. Le 4^{ème} et le 5^{ème} axe contiennent le calcium et les matières en suspensions respectivement puis sur le 6^{ème} axe on retrouve le pH avec les nitrites.

Les résultats obtenus sont assez bons, notamment les minéraux qui sont bien corrélés avec la conductivité, la DBO et la DCO qui se regroupent avec l'axe 2. Même si on remarque quelques aberrations, comme les métaux lourds, qui ne sont pas tous sur le même axe, mis à part le Zn, et le Mn, mais aussi les nitrites qu'on prévoyait les voir bien corrélées avec les autres matières azotées.

- **Menia**

Cette analyse a été effectuée aux mêmes variables que pour la station d'Ain Smara, mais pour 86 individus.

- Tableau des valeurs propres

Tableau 5-3 : Tableau des valeurs propres ACP, Menia

	F1	F2	F3	F4	F5	F6	F7
Valeur propre	5,666	3,311	1,980	1,457	1,268	1,204	1,062
Variabilité (%)	24,634	14,395	8,610	6,333	5,512	5,235	4,616
% cumulé	24,634	39,029	47,639	53,972	59,484	64,720	69,335

D'après ce tableau, sept composantes principales sont retenues selon la règle de Kaiser.

- Cercle de corrélations

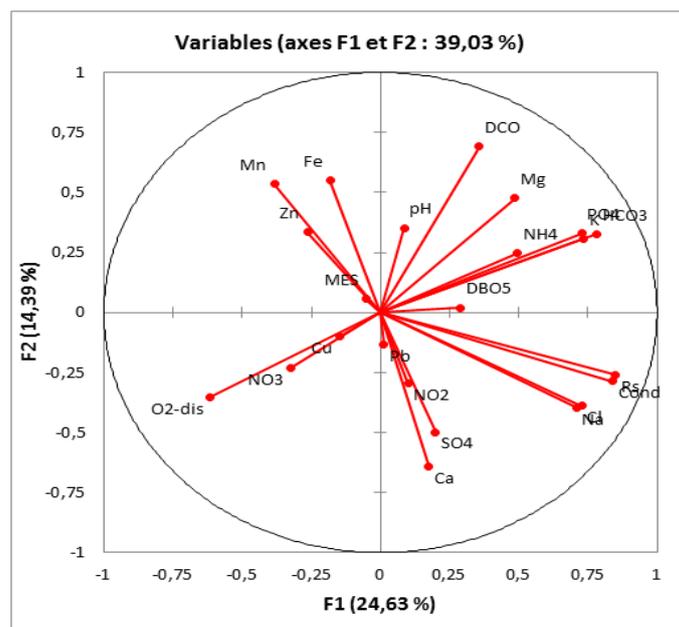


Figure 5-2: Cercle des Corrélations ACP Menia

- Tableau des cosinus

Tableau 5-4 : Tableau des cosinus ACP, Menia

	F1	F2	F3	F4	F5	F6	F7
pH	0,007	0,124	0,271	0,096	0,080	0,000	0,071
Cond	0,704	0,082	0,049	0,002	0,001	0,001	0,004
MES	0,003	0,003	0,021	0,174	0,432	0,000	0,155
Rs	0,717	0,067	0,097	0,006	0,001	0,000	0,003
Ca	0,030	0,412	0,003	0,130	0,000	0,146	0,003
Mg	0,235	0,226	0,199	0,134	0,004	0,024	0,007
Na	0,503	0,157	0,005	0,051	0,012	0,022	0,010
K	0,540	0,094	0,009	0,000	0,006	0,003	0,005
Cl	0,528	0,150	0,000	0,015	0,006	0,035	0,005
SO4	0,040	0,247	0,481	0,002	0,013	0,000	0,001
HCO3	0,612	0,107	0,001	0,002	0,001	0,064	0,005
NO3	0,104	0,053	0,148	0,001	0,158	0,090	0,017
NO2	0,010	0,087	0,001	0,008	0,249	0,164	0,145
NH4	0,243	0,061	0,137	0,065	0,013	0,001	0,019
PO4	0,528	0,108	0,048	0,000	0,002	0,007	0,002
DCO	0,127	0,478	0,057	0,008	0,054	0,001	0,000
DBO5	0,082	0,000	0,096	0,004	0,000	0,181	0,152
O2-dis	0,379	0,124	0,032	0,002	0,027	0,008	0,045
Cu	0,022	0,010	0,115	0,080	0,086	0,003	0,179
Pb	0,000	0,018	0,020	0,021	0,015	0,391	0,205
Zn	0,070	0,113	0,024	0,326	0,074	0,010	0,020

Fe	0,033	0,302	0,152	0,021	0,003	0,052	0,002
Mn	0,148	0,287	0,016	0,309	0,032	0,001	0,006

Le premier axe contient les matières minérales, les phosphates et l'oxygène dissous. Le 2^{ème} axe contient le fer la DCO et le calcium. Le 3^{ème} axe contient le pH et les sulfates. Le 4^{ème} axe regroupe 2 métaux lourds qui sont le zinc et le manganèse, alors que le 5^{ème} axe regroupe 2 matières azotées (les nitrates et les nitrites) et le 6^{ème} axe contient le Pb.

Les résultats sont assez cohérents en ce qui concerne les minéraux qui sont bien corrélés avec la conductivité, et les nitrates qui se retrouvent sur le même axe avec les nitrites. Mais, on remarque quelques incohérences notamment les métaux lourds qui ne se regroupent pas sur le même axe et aussi la DBO et la DCO.

- **Beni Haroun**

Cette analyse a été effectuée aux mêmes variables que pour la station d'Ain Smara, mais pour 89 individus.

- Tableau des valeurs propres

Tableau 5-5 : Tableau des valeurs propres ACP, Beni Haroun

	F1	F2	F3	F4	F5	F6
Valeur propre	7,228	2,256	2,166	1,250	1,145	1,095
Variabilité (%)	34,418	10,741	10,314	5,952	5,451	5,215
% cumulé	34,418	45,159	55,473	61,425	66,875	72,090

Six composantes principales ont été retenues en utilisant la règle de Kaiser.

- Cercle de corrélations

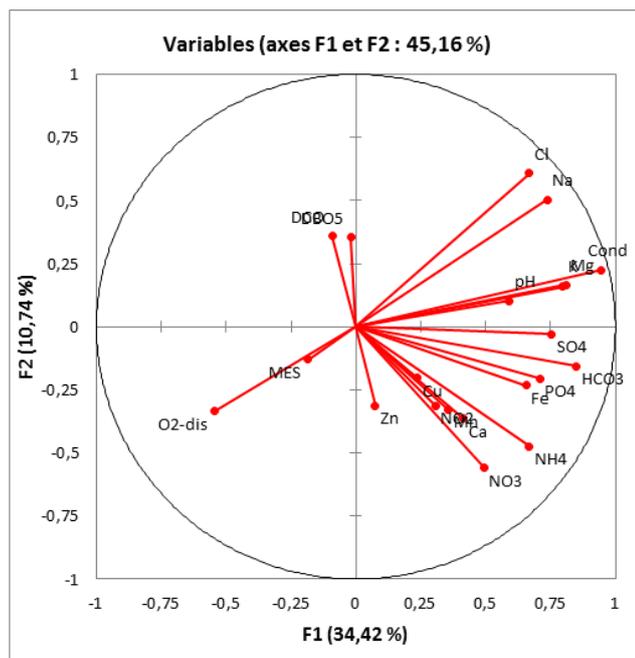


Figure 5-3: Cercle des Corrélations ACP Beni Haroun

- Tableau des cosinus

Tableau 5-6 : Tableau des cosinus ACP, Beni Haroun

	F1	F2	F3	F4	F5	F6
pH	0,348	0,010	0,159	0,058	0,015	0,095
Cond	0,895	0,050	0,000	0,000	0,000	0,005
MES	0,034	0,017	0,118	0,092	0,073	0,381
Ca	0,171	0,133	0,378	0,022	0,013	0,054
Mg	0,653	0,028	0,150	0,011	0,003	0,031
Na	0,542	0,254	0,015	0,050	0,001	0,005
K	0,636	0,025	0,003	0,003	0,028	0,040
Cl	0,447	0,370	0,000	0,057	0,000	0,003
SO4	0,565	0,001	0,008	0,079	0,006	0,020
HCO3	0,723	0,024	0,052	0,002	0,006	0,040
NO3	0,245	0,314	0,073	0,016	0,039	0,000
NO2	0,094	0,098	0,244	0,044	0,001	0,026
NH4	0,446	0,223	0,000	0,109	0,005	0,007
PO4	0,503	0,042	0,000	0,089	0,000	0,000
DCO	0,008	0,129	0,041	0,140	0,004	0,304
DBO5	0,000	0,125	0,169	0,002	0,273	0,025
O2-dis	0,299	0,112	0,012	0,063	0,000	0,017
Cu	0,056	0,041	0,133	0,144	0,457	0,001
Zn	0,005	0,098	0,276	0,246	0,080	0,030

Fe	0,431	0,054	0,009	0,002	0,037	0,010
Mn	0,125	0,107	0,325	0,021	0,103	0,001

Le 1^{er} axe contient divers paramètres tel quel les minéraux, l'ammonium, le phosphate, l'oxygène dissous et le fer. Le 2^{ème} axe contient les nitrates seulement, puis le 3^{ème} axe contient le zinc et le manganèse comme métaux lourds avec le calcium et les nitrites. Le 5^{ème} axe contient la DBO et le cuivre et en dernier le 6^{ème} axe les MES avec la DCO.

Les résultats sont assez bons en ce qui concerne les minéraux qui sont bien corrélés avec la conductivité, mais on remarque des incohérences comme la DBO qui est corrélée avec le cuivre ou encore les nitrates et les nitrites qui ne se retrouvent pas sur le même axe.

Discussion des résultats :

L'application de l'ACP aux 3 stations étudiées, montent d'assez bons résultats, et ce malgré quelques incohérences, dues probablement aux différentes relations liant les paramètres mesurés. Ainsi, on retrouve par exemple, dans les 3 stations une bonne corrélation des minéraux avec la conductivité, alors que les nitrites et les métaux lourds posent problème.

L'explication que l'on peut donner au sujet de l'incohérence des résultats obtenus, à partir de l'application de l'ACP, à ce stade de l'étude, c'est qu'on soupçonne, la non linéarité des relations, liants les différents paramètres étudiés, alors que celle-ci est l'essence même de l'ACP. De plus, on a constaté un effet d'échelle, trop important au niveau des données (il existe des paramètres dont les valeurs varient entre 0 et 1 et d'autres entre 500 et 2000).

Pour mieux comprendre les relations qui peuvent exister entre les différents paramètres mesurés et tenter d'avoir des résultats plus pertinents, on se propose de passer à d'autres techniques de clustérisation, qui tiennent justement compte, de la non linéarité des relations entre les paramètres étudiés.

V-2- K-means

Le K-Means a été effectué sur les séries de données recueillies en utilisant le logiciel Xlstat.

Le but de l'utilisation de cet algorithme de classification non supervisé 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.

-Ain Smara

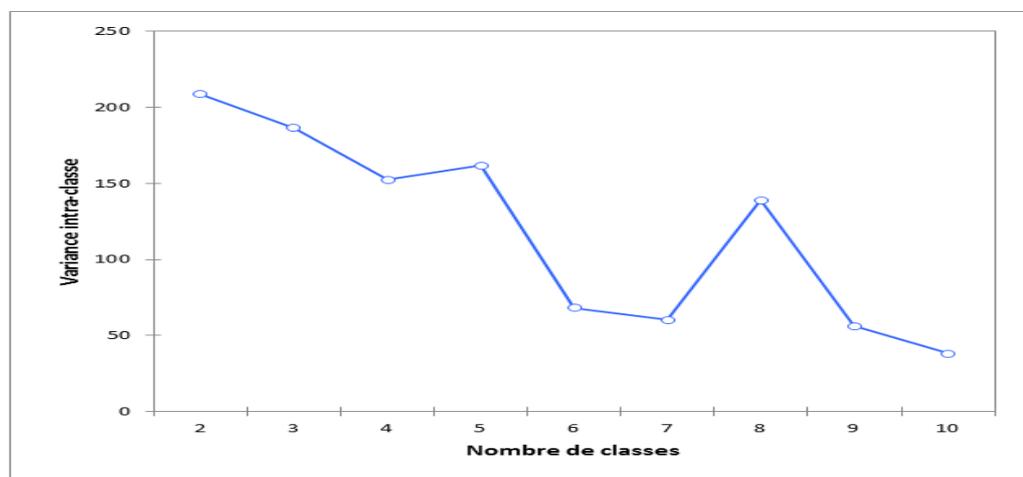


Figure 5-4 : Courbe de l'évolution des variances K-means Ain Smara

Selon le graphe ci-dessus, le nombre de classes optimales en appliquant la méthode du coude est de 4.

Pour K = 4, on obtient les résultats suivants :

Tableau 5-7 : Résultats K-Means Ain Smara

Classe	1	2	3	4
Objets	13	6	3	1
Somme des poids	13	6	3	1
Variance intra-classe	30,468	204,820	754,213	0,000
Distance minimale au barycentre	2,859	5,903	15,810	0,000
Distance moyenne au barycentre	4,934	12,159	21,144	0,000
Distance maximale au barycentre	9,333	21,267	31,704	0,000

barycentre				
	pH	Cond	MES	NO2
	Rs	Na	Ca	
	Cl	K	Mg	
	HCO3	SO4		
	NH4	NO3		
	PO4	O2-dis		
	DCO			
	DBO5			
	Cu			
	Pb			
	Zn			
	Fe			
	Mn			

D'après ce tableau, on retrouve dans la première classe, le pH, les résidus secs, le chlore, les bicarbonates, l'ammonium, le phosphore, la DCO et DBO et les métaux lourds. La deuxième classe contient quelques paramètres de minéralisation avec les sulfates, l'oxygène dissous et les nitrates. Alors que la troisième classe regroupe les MES avec le calcium et le magnésium, et on retrouve dans la dernière, les nitrites.

On peut dire là aussi, que cette deuxième technique de clustérisation, permet d'avoir des regroupements assez intéressants de paramètres, comme par exemple, les métaux lourds qui se retrouvent tous dans la même classe, ou bien la DCO et la DBO qui se regroupent aussi dans la même classe. En revanche, on remarque quelques incohérences notamment les nitrites et les nitrates qui sont dans des classes différentes.

- **Menia**

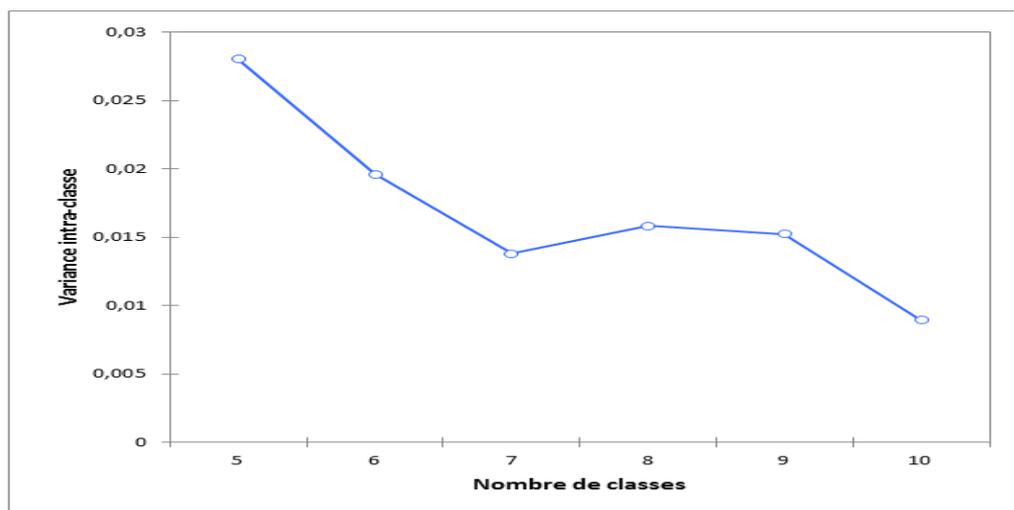


Figure 5-5 : Courbe de l'évolution des variances K-means Menia

Selon le graphe ci-dessus, le nombre de classes optimales en appliquant la méthode du coude est 7.

Pour K = 7, on obtient les résultats suivants :

Tableau 5-8 : Résultats K-Means Menia

Classe	1	2	3	4	5	6	7
Objets	10	5	2	1	3	1	1
Somme des poids	10	5	2	1	3	1	1
Variance intra-classe	0,004	0,024	0,067	0,000	0,012	0,000	0,000
Distance minimale au barycentre	0,029	0,107	0,183	0,000	0,057	0,000	0,000
Distance moyenne au barycentre	0,054	0,136	0,183	0,000	0,086	0,000	0,000
Distance maximale au barycentre	0,088	0,198	0,183	0,000	0,127	0,000	0,000
	pH	Cond	Mg	Na	HCO3	NO2	NH4
	MES	Rs	K		PO4		
	SO4	Ca			O2-dis		
	NO3	Cl					
	DBO5	DCO					
	Cu						
	Pb						
	Zn						
	Fe						
	Mn						

Dans la première classe on retrouve les métaux lourds avec le pH, les MES, les sulfates, la DBO et les nitrates. Puis la 2^{ème} classe contient la conductivité, la DCO, les résidus secs, le calcium et le chlore. La 3^{ème} classe contient le magnésium et le potassium. Les dernières classes contiennent respectivement, le sodium, les bicarbonates le phosphore et l’oxygène dissous, les nitrates et l’ammonium.

On remarque que les métaux lourds, se trouvent là aussi regroupés dans la même classe, ce qui constitue un très bon résultat, alors que plusieurs incohérences subsistent, notamment la DCO et la DBO qui ne sont pas dans la même classe ou encore les nitrates et les nitrites.

- **Beni Haroun**

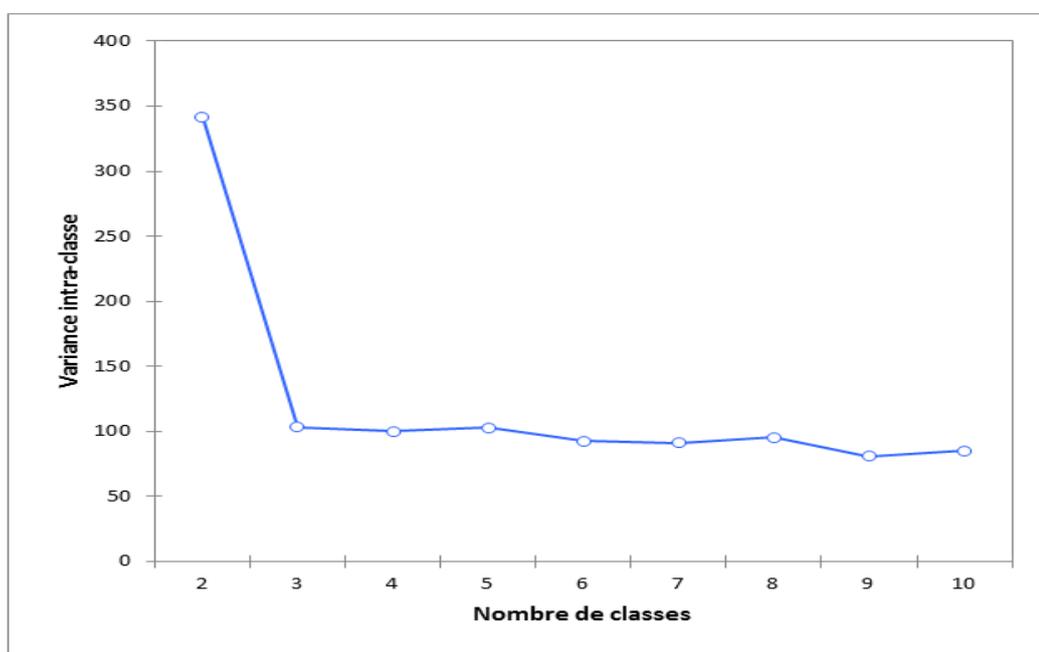


Figure 5-6 : Courbe de l'évolution des variances K-Means Beni Haroun

Selon le graphe ci-dessus, le nombre de classes optimales en appliquant la méthode du coude est 3.

Pour K = 3, on obtient les résultats suivants :

Tableau 5-9: Résultats K-Means Beni Haroun

Classe	1	2	3
Objets	6	2	13
Somme des poids	6	2	13
Variance intra-classe	174,926	635,166	29,549
Distance minimale au barycentre	6,951	17,821	2,588
Distance moyenne au barycentre	11,634	17,821	4,668

Distance maximale au barycentre	16,410	17,821	9,988
	pH	Cond	MES
	Na	Mg	Ca
	K		Cl
	SO4		HCO3
	NO3		NO2
	O2-dis		NH4
			PO4
			DCO
			DBO5
			Cu
			Zn
			Fe
			Mn

La première classe contient le pH, le sodium, le potassium, les sulfates, les nitrates et l'oxygène dissous. La 2^{ème} classe contient la conductivité et le magnésium et la 3^{ème} classe regroupe les MES, le calcium, le chlore, les bicarbonates, les nitrites, l'ammonium, le phosphore, la DCO, la DBO et les métaux lourds.

Les résultats sont nettement meilleurs pour cette station, car on retrouve les métaux lourds, la DBO et la DCO, mais aussi les nitrites et l'ammonium, dans la même classe, alors que les nitrates qui devaient être dans la même classe ne le sont pas.

Discussion des résultats :

Les résultats de la clustérisation par K-Means sont plus pertinents que ceux obtenus à partir de l'application de l'ACP. Ceci confirme nos soupçons que les liaisons entre paramètres étudiés ne sont pas nécessairement linéaires et que l'ACP ne pouvait pas prendre en charge ce type de liaisons. En effet, on remarque, par exemple, qu'en appliquant le K-Means, on obtient un regroupement des métaux lourds dans une même classe, pour les 3 stations. Néanmoins, quelques incohérences subsistent toujours, comme par exemple, pour les nitrites et les nitrates, qui ne se retrouvent jamais regroupés dans la même classe, et ce pour les 3 stations. Le problème avec cette méthode, c'est qu'on remarque qu'elle contient plusieurs regroupements dans la même classe, donc des sous-classes, mais qu'elle n'arrive pas à faire ressortir. C'est pour cela qu'on se propose d'appliquer une troisième et dernière méthode de clustérisation, qui est la classification ascendante hiérarchique et qui à l'aide de ses résultats peut apporter un plus pour ce travail.

5-3- La Classification Ascendante Hiérarchique

Le logiciel utilisé est XLstat. La classification a été faite selon la similarité du coefficient de corrélation de Pearson.

-Ain Smara

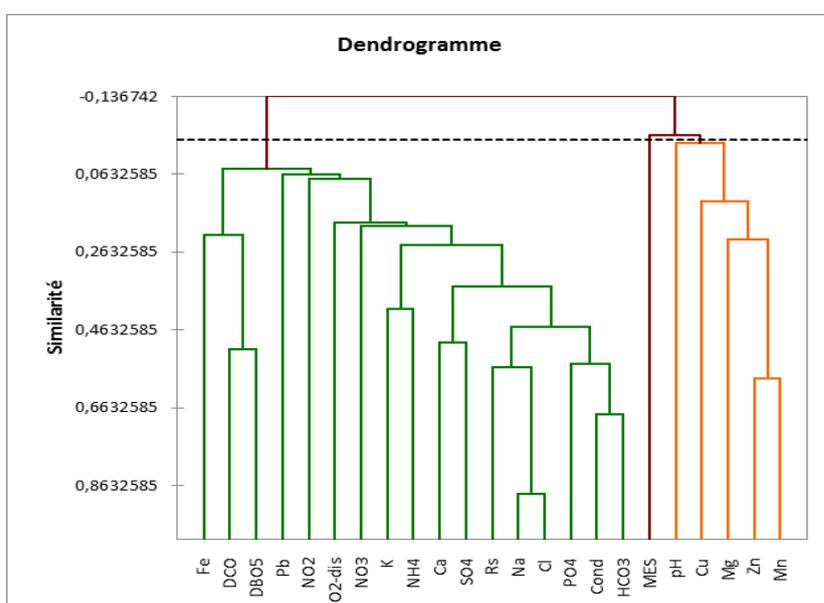


Figure 5-7: Le dendrogramme avec coefficient de corrélation de Pearson Ain Smara

Tableau 5-10 : Classification par CAH avec coefficient de corrélation de Pearson Ain Smara

Classe	1	2	3
Objets	5	17	1
Somme des poids	5	17	1
Variance intra-classe	526,969	209,371	0,000
Distance minimale au barycentre	7,394	5,576	0,000
Distance moyenne au barycentre	16,189	12,432	0,000
Distance maximale au barycentre	39,825	33,376	0,000
	pH, Mg, Cu, Zn, Mn	Cond, Rs, Ca, Na, K, Cl, SO4, HCO3, NO3, NO2, NH4, PO4, DCO, DBO, O2-dis, Pb, Fe	MES

On retrouve dans la première classe 3 métaux lourds avec le pH et le magnésium. Puis dans la 2^{ème} classe on retrouve les minéraux les matières azotées, les matières organiques et les 2 métaux restant et en dernier la 3^{ème} classe contient les matières en suspension.

- **Menia**

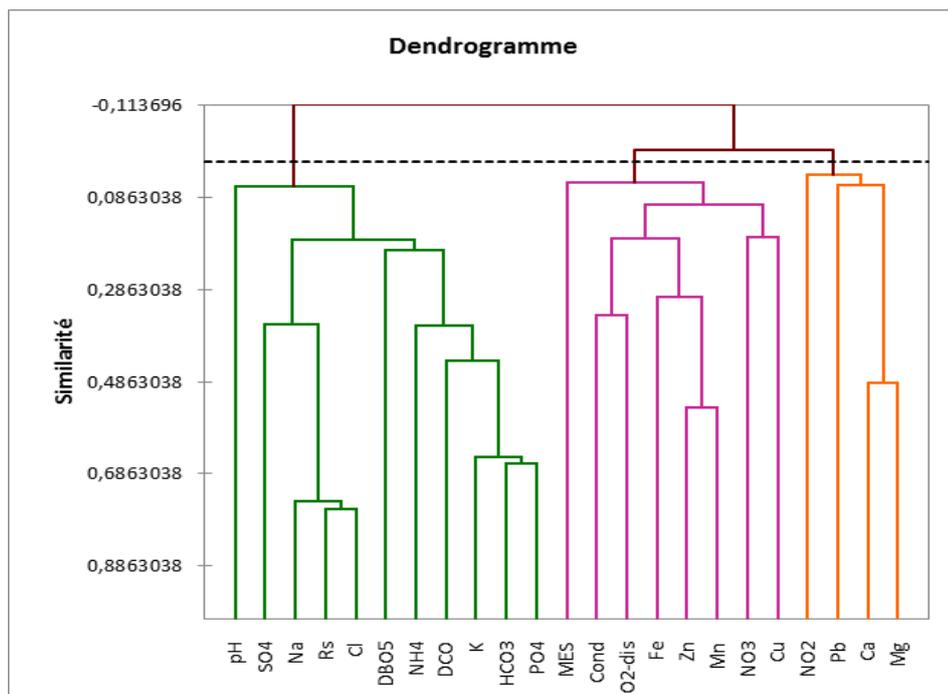


Figure 5-9: Le dendrogramme avec coefficient de corrélation de Pearson Menia

Tableau 5-11: Classification par CAH avec coefficient de corrélation de Pearson Ain Smara

Classe	1	2	3
Objets	11	8	4
Somme des poids	11	8	4
Variance intra-classe	914,409	74,940	805,371
Distance minimale au barycentre	10,730	1,201	8,149
Distance moyenne au barycentre	23,299	5,207	20,969
Distance maximale au barycentre	73,824	18,289	34,105
	pH, Rs, Na, K, Cl, SO4, HCO3, NH4, PO4, DCO, DBO	Cond, MES, NO3, O2-dis, Cu, Zn, Fe, Mn	Ca, Mg, NO2, Pb

Dans la première classe on trouve quelques minéraux avec le pH, les phosphates, l'ammonium, la DBO et la DCO. Puis, dans la 2^{ème} classe on trouve les métaux lourds avec la conductivité, les MES, les nitrates et l'oxygène dissoute. En dernier, on retrouve des minéraux avec les nitrites et le plomb.

- Beni Haroun

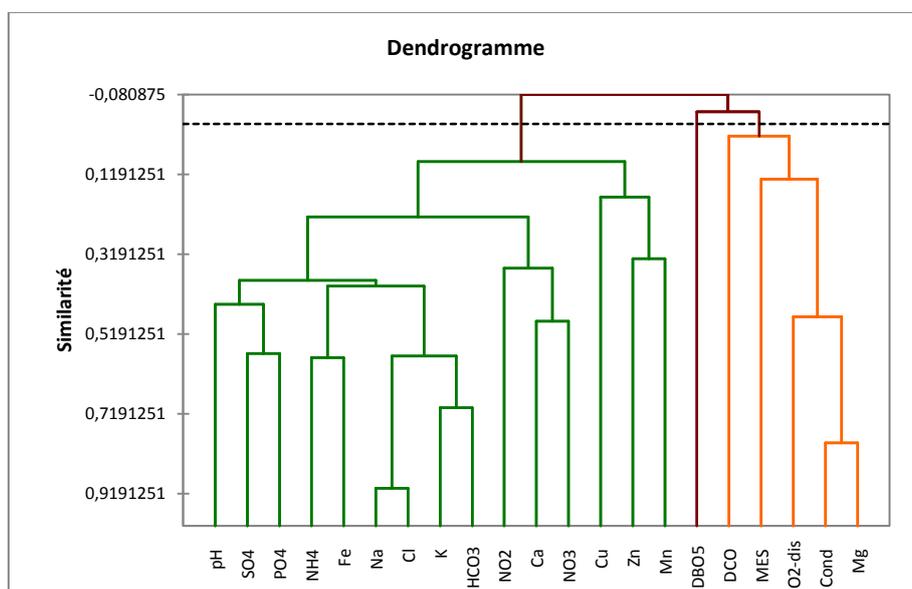


Figure 5-8: Le dendrogramme avec coefficient de corrélation de Pearson Beni Haroun

Tableau 5-12: Classification par CAH avec coefficient de corrélation de Pearson

Classe	1	2	3
Objets	15	5	1
Somme des poids	15	5	1
Variance intra-classe	151,292	1511,412	0,000
Distance minimale au barycentre	3,626	20,626	0,000
Distance moyenne au barycentre	10,290	32,877	0,000
Distance maximale au barycentre	25,334	53,972	0,000
	pH, Ca, Na, K, Cl, SO4,	Cond, MES, Mg,	DBO5

	HCO ₃ , NO ₃ , NO ₂ , NH ₄ , PO ₄ , Cu, Zn, Fe, Mn	DCO, O ₂ -dis	
--	--	--------------------------	--

Dans la première classe on trouve plusieurs minéraux avec les métaux lourds et les matières azotées. Dans la 2^{ème} classe on retrouve la conductivité, les matières en suspension, le magnésium, la DCO et l'oxygène dissoute et en dernier la DBO dans la 3^{ème} classe.

En analysant les trois dendrogrammes obtenus pour chacune des stations, on juge que celui de la station de Beni Haroun présente le meilleur regroupement d'éléments. Ces sous classes ont été obtenues par l'analyse graphique du dendrogramme en modifiant le niveau de similarité on a repéré les regroupements suivants :

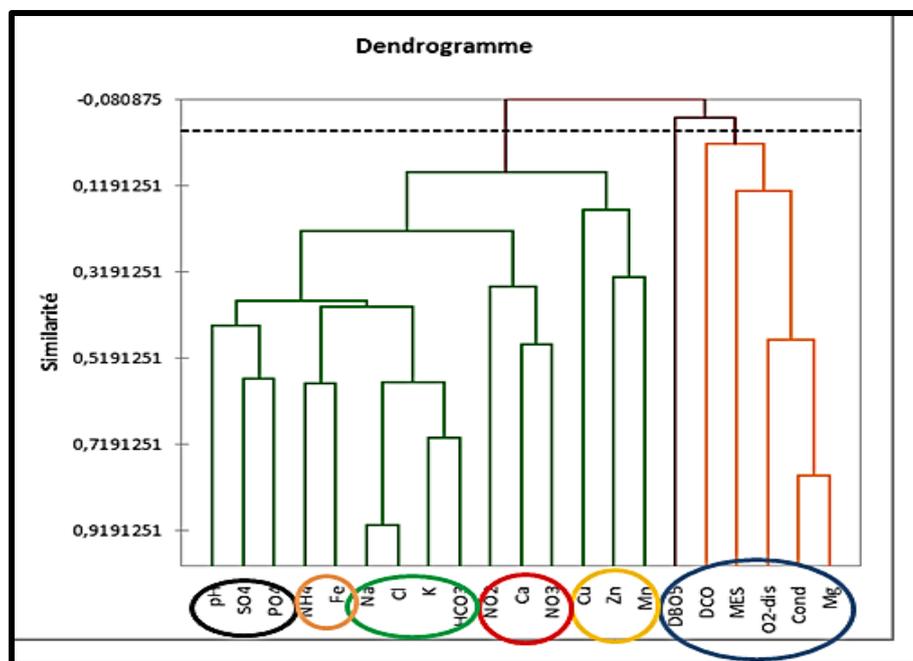


Figure5-9: Regroupement des paramètres CAH-Pearson

On retrouve 6 classes principales :

- Le cercle bleu contient à deux éléments près les paramètres responsables de la pollution organique
- Le cercle jaune, regroupe les métaux lourds
- Le cercle rouge, regroupe les éléments de la pollution agricole
- Le cercle vert, regroupe les minéraux
- Le cercle noir, c'est consacré à la pollution industrielle
- En ce qui concerne les éléments du cercle orange, l'ammonium est en excès par rapport à la norme, et les valeurs du fer sont très élevées, comparées aux

autres métaux lourds qui ont tous la même norme, donc c'est pour cela qu'on considère cette classe comme étant la classe des éléments en excès. De plus les résultats de l'ACP ont montrés qu'ils sont bien corrélés entre eux vu qu'ils se retrouvent sur le même axe. (Voir tableau V-6)

Discussion des résultats:

Les résultats obtenus par la CAH sont assez satisfaisants. La majorité des classes se rapprochent de la théorie malgré quelques différences qui sont peut-être dû à la taille de la série qui n'est pas très grande et donc ne décrit pas le phénomène comme il se doit. Les meilleurs résultats sont obtenus à la station de Beni Haroun, cela peut s'expliquer par le fait que c'est les eaux du barrage et donc la qualité des données est meilleure car c'est un stockage contrairement aux deux autres stations qui sont implantées sur des cours d'eau en mouvement.

5-4- Résumé des résultats de classification selon l'OMS :

Après avoir comparé les résultats de la CAH avec la classification de l'OMS on a constaté une ressemblance entre les deux, même si, la classification de l'OMS est restreinte et comprend moins de paramètres, on l'a utilisé pour avoir une vision globale sur la qualité de l'eau et avoir des conclusions sur nos résultats.

Tableau 5-13 : Classification de L'OMS

Classe Paramètre	Excellente	Bonne	Passable	Médiocre	Pollution extrême
O2-dissous	>9	7-9	5-7	3-5	<3
DCO	<20	20-30	30-40	40-80	>80
MES	<5	5-25	25-38	38-50	>50
NO2	<0.03	0.03-0.1	0.1-0.5	0.5-1	>1
NO3	<2	2-10	10-25	25-50	>50
PO4	<0.1	0.1-0.5	0.5-1	1-2	>2
Conductivité	<2500	2500-3000	3000-3500	3500-4000	<4000

L'OMS a caractérisé les matières en 5 classes :

- Matière minérale : Conductivité
- Matière organique et oxydables : DCO, O2-dis
- Matière azotées : NO2, NO3
- Matière en suspension : MES
- Matière phosphatée : PO4

- **Ain Smara**

- Avant réception

Tableau 5-14 : Classification selon les normes de l'OMS de la station d'Ain Smara avant réception du barrage

Date	1994	1995	1996	1997	1998	1999	2000	2001	2002
Cond	Blue								
MES	Green						Red		
NO3	Yellow	Green	Yellow	Red	Green	Red	Yellow	Yellow	Green
NO2	Yellow					Red	Yellow	Red	Yellow
DCO	Red						Yellow	Red	Yellow
O2-dis	Red						Yellow	Green	
PO4	Red								

- Après réception

Tableau 5-15: Classification selon les normes de l'OMS de la station d'Ain Smara après réception du barrage

Date	2003	2004	2005	2006	2007	2008	2009	2010
Cond	Blue							
MES	Red							
NO3	Yellow							
NO2	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow
DCO	Red	Yellow						
O2-dis	Green							Yellow
PO4	Red	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow

Nous avons adopté la même démarche que pour la station d'El Menia, les résultats obtenus sont comme suit :

Tableau 5-16: Récapitulatif de classification selon les paramètres station Ain Smara

Paramètre	classe moyenne de toute la période	classe moyenne avant la réception	classe moyenne après la réception
Cond			
MES			
NO3			
NO2			
DCO			
O2-dis			
PO4			

Enfin, nous avons regroupé les matières selon la catégorisation de l'OMS mentionnée précédemment.

Tableau 5-17: Récapitulatif de classification selon les matières station Ain Smara

	94-10	94-02	03_10
Matière minérales			
Matière en suspension			
Matière azotée			
Matière organique et oxydable			
Matière phosphatée			

Tout d'abord on constate une stabilité dans la classe des matières minérales et des matières azotées qui sont respectivement excellente et médiocre pendant toute la période d'observation. Les matières en suspension, et les matières organiques et oxydables sont toutes les deux d'une qualité médiocre. En ce qui concerne les matières organiques, on remarque une amélioration après la mise en eau du barrage (passée de pollution excessive à une qualité passable), en revanche les matières en suspension ont connu une dégradation de qualité (de passable à une pollution excessive après 2003). Enfin, la matière phosphorée cause une pollution excessive même si on constate une légère amélioration après 2003.

- Menia

- Avant la réception du barrage

Tableau 5-18 : Classification selon les normes de l’OMS de la station de Menia avant réception du barrage

Date	1994	1995	1996	1997	1998	1999	2000	2001	2002	
Cond	Blue									
MES	Green					Red				
NO3	Yellow					Red		Yellow		Green
NO2	Yellow					Red				Yellow
DCO	Red									
O2-dis	Red					Yellow		Yellow		
PO4	Red									

- Après la réception

Tableau 5-19 : Classification selon les normes de l’OMS de la station de Menia après réception du barrage

Date	2003	2004	2005	2006	2007	2008	2009	2010
Cond	Blue							
MES	Red							
NO3	Yellow							
NO2	Red							
DCO	Red	Yellow						Yellow
O2-dis	Yellow					Green	Yellow	
PO4	Red							

Nous avons calculé la classe moyenne avant, et après la réception du barrage, ainsi que la totalité de la période, les résultats obtenus sont présentés ci-dessous :

Tableau 5-20: Récapitulatif de classification selon les paramètres station Menia

Paramètre	classe moyenne de toute la période	classe moyenne avant la réceptic	classe moyenne après la réception
Cond	Blue		
MES	Yellow	Yellow	Red
NO3	Yellow	Yellow	Yellow
NO2	Red		
DCO	Red		Yellow
O2-dis	Yellow	Red	Yellow
PO4	Red		

Enfin, nous avons regroupé les matières selon la catégorisation de l’OMS mentionnée précédemment.

Tableau 5-21: Récapitulatif de classification selon les matières station Menia

	94-10	94-02	03_10
Matière minérales			
Matière en suspension			
Matière azotée			
Matière organique et oxydable			
Matière phosphatée			

On constate que les eaux analysées dans la station de mesure de Menia ont une excellente qualité matières minérales. Les matières en suspension sont médiocres et ont connu une augmentation après la réception du barrage (la qualité est passée de passable à une pollution excessive). Les matières azotées sont stables tout au long de la période d’observation avec une qualité médiocre. En ce qui concerne les matières organiques et oxydables qu’ont connus une amélioration après 2003 et la qualité est passée d’une pollution excessive à médiocre. Enfin, les eaux connaissent une pollution excessive en matières phosphorées.

- **Beni Haroun**

- Avant réception

Tableau 5-22: Classification selon les normes de l’OMS de la station de Beni Haroun avant réception du barrage

Date	1994	1995	1996	1997	1998	1999	2000	2001	2002	
Cond										
MES										
NO3										
NO2										
DCO										
O2-dis										
PO4										

- Après réception

Tableau 5-23: Classification selon les normes de l’OMS de la station de Beni Haroun après réception du barrage

Date	2003	2004	2005	2006	2007	2008	2009	2010
Cond	Blue							
MES	Red							
NO3	Yellow			Green				
NO2	Yellow	Orange	Yellow	Orange	Yellow	Yellow	Orange	Orange
DCO	Orange			Red				
O2-dis	Yellow	Green					Blue	Blue
PO4	Orange		Yellow					Green

La même démarche a été adoptée que pour les deux autres stations, les résultats obtenus dans ce cas sont :

Tableau 5-24 : Récapitulatif de classification selon les paramètres station Beni Haroun

Paramètre	classe moyenne de toute la période	classe moyenne avant la réceptio	classe moyenne après la réception
Cond	Blue		
MES	Orange	Yellow	Red
NO3	Green		
NO2	Orange		
DCO	Red		
O2-dis	Yellow	Orange	Green
PO4	Orange	Red	Yellow

Enfin, nous avons regroupé les matières selon la catégorisation de l’OMS

Tableau 5-25: Récapitulatif de classification selon les matières station Beni Haroun

	94-10	94-02	03_10
Matière minérales	Blue		
Matière en suspension	Orange	Yellow	Red
Matière azotée	Yellow		
Matière organique et oxydable	Orange		Yellow
Matière phosphatée	Orange	Red	Yellow

En premier lieu, on est sur une stabilité sur toute la période au niveau des matières minérales et des matières azotées. La première étant d’une qualité excellente et la deuxième est passable. En second lieu, on constate une amélioration dans les matières organiques, et les matières phosphorées, qui sont passées de médiocre à passable, et de pollution excessive à passable respectivement après la réception du barrage. Les deux sont en qualité médiocre sur l’ensemble des années d’observation. Enfin, Les matières en suspension ont connu une augmentation, les faisant passer de la qualité passable à une pollution excessive.

Conclusion générale

Le but de ce travail est de trouver les paramètres qui influent le plus, et qui altèrent la qualité des eaux de surface du barrage de Beni Haroun, et des cours d'eau dont il est tributaire.

Cette caractérisation de la qualité est essentielle pour la gestion, et la préservation de la ressource en eau qui se fait rare dans la région en question.

Pour ce faire, on s'est basé sur les techniques du Data Mining, dont on a utilisé plusieurs méthodes de clustérisation afin de regrouper les paramètres polluants dans des classes bien définis.

Au préalable de ce travail, on a prédéfini les types de pollutions et leurs origines. Ensuite, on a présenté la région d'étude, et repéré les sources de pollutions possibles, dans différents domaines d'activité. Puis, on a enchainé avec la présentation des stations étudiées, et critiqué les données recueillies avec l'élimination des valeurs aberrantes, afin de fiabiliser les résultats, étant donné que les techniques utilisées ne tolèrent pas de valeurs aberrantes.

En ce qui concerne ces techniques, en premier lieu, on a effectué une ACP pour chacune des stations, et on a obtenu des résultats assez satisfaisants, avec quelques incohérences. Mais en raison de la non linéarité des relations inter-paramètres, on est passé à une deuxième technique qui est celle du K-Means.

Cette technique, nous a permis d'obtenir des résultats acceptables pour les 3 stations, mais le problème résidait dans le fait qu'on retrouvait des sous groupements dans la même classe. Pour y remédier, on a utilisé une autre technique de clustérisation qui est la classification ascendante hiérarchique. Même si les résultats pour les stations de Menia et de Ain Smara se rapprochaient nettement des résultats obtenus par le K-Means, cette technique appliquée sur la série de données de Beni Haroun a donnée, plutôt de bons résultats et a regroupé les paramètres de pollutions en 6 classes différentes, qui sont : la pollution organique, les métaux lourds, la pollution agricole, la pollution minérale et la pollution industrielle. A l'exception des deux paramètres, qui sont le fer et l'ammonium, qui se sont déjà avérés, erronés lors de la critique des données et qui sont bien corrélés entre eux selon l'ACP.

Enfin, on a projeté les résultats obtenus par la CAH, à la classification de l'OMS déjà existante, pour classer les eaux des trois stations. Suite à cette classification, on a eu deux types d'informations. En ce qui concerne les eaux des cours d'eau, elles ont une excellente

minéralisation, une qualité médiocre en matière azotées, en matières organiques, et une pollution excessive en matière phosphorée. Pour les matières en suspension on a remarqué une chute de la qualité après la mise en eau du barrage. Pour les eaux de barrage, on constate aussi une excellente minéralisation, une chute de la qualité dans les matières en suspension et une amélioration considérable de qualité en matières phosphorée, organique et azotée.

En conclusion, on peut dire que ce travail nous a permis de franchir un pas dans le domaine de la caractérisation des eaux de surface, en utilisant les différentes techniques du Data Mining, un domaine qui reste nouveau et prêt à être exploré .

Comme perspectives pour la suite de ce travail on peut citer :

- Créer une classification algérienne semblable à celle de l’OMS mais en considérant plus de variables et de types de pollution.
- Utiliser la méthode sur une série de données plus grande pour fiabiliser les résultats obtenus.

REFERENCES BIBLIOGRAPHIQUES

- BENSALÉM, Ouail., FERRAOUN, Rafik., *Modélisation de l'énergie électrique des STEP Boumerdes et Tipaza*, Mémoire de projet de fin d'études, ENP Alger, 2018, 102 pages.
- Bradley, Ussama Fayad. *Scaling clustering algorithms to large databases In knowledge Discovery and Data Mining*, 1998, pp.9-15.
- BRAHIMI, Serhane., *Comportement des éléments chimiques dans les eaux de surface du barrage de Beni Haroun (Est Algérien)*, Mémoire de Magister, Université de Tebessa, 2017, pp.40-60.
- Decamps.H. *Les forêts riveraines des cours d'eau : écologie, fonction et gestion*. Verth Internat. Verien. Limnol, 1996, pp.423-426.
- DUPONT André. *Hydraulique urbaine : Hydrologie-captage et traitement des eaux* Ed. Eyrolles, tome I, 1981, pp.146-150.
- HAFSI. Bilel. *Le Transport solide et la Qualité des Eaux du Barrage Beni-Haroun et leur Impacts Sur l'environnement*, Mémoire de Projet de fin d'études, Université Larbi Ben M'hidi– Oum El Bouaghi, 2016, pp.54-66.
- HAMRICHE, Amel. *Contribution à l'étude et à la simulation des paramètres hydrométriques par l'analyse en composantes principales (ACP)*, Mémoire de Projet de Fin d'Etude, ENP Alger, 1993, pp.34-48.
- Institut bruxellois pour la gestion de l'environnement, *Qualité physico-chimique et chimique des eaux de surface*, Cadre général Fiche 2. Institut Bruxellois pour la gestion de l'environnement/observatoire des données de l'environnement, 2005.
- Journal officiel de la république algérienne N°18 18 Rabie Ethani 1432 ,23 mars 2011.
- KERDOUD, Souhila. *Le bassin versant de Beni Haroun eau et pollution*, Mémoire de Magister, Université Mentouri- Constantine, 2006, pp.55-70.
- Khaymastv : *Contrôle de la qualité et de la salubrité des milieux naturels*. [En ligne]. [Consulté le 20/03/2020]. Disponible à l'adresse :

<http://www.khayma.com/svtfat/bacinterpollu6.htm>

- MAROUF, Nadir.: *Etude de la qualité des eaux et de transport solide dans le barrage de Beni Haroun (Mlia), son impact sur l'environnement de la région*. Thèse de doctorat, Université de Biskra, 2012, pp.62-64.
- Ministère de l'aménagement du territoire et de l'environnement français, *Pollutions d'origine agricoles: le PMPOA* 2002. [En ligne]. [Consulté le 20/05/2020]. Disponible à l'adresse : http://www.waternunc.com/fr/medd65_2002.htm
- Minitab. *Le dendrogramme*. [En ligne]. [Consulté le 20/05/2020]. Disponible à l'adresse :
<https://support.minitab.com/fr-fr/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/cluster-observations/interpret-the-results/all-statistics-and-graphs/dendrogram/>
- RODIER, Jean André. *L'analyse de l'eau: Eaux naturelles, eaux résiduaires, eaux de mer*, Ed. Dunod, 8ème Edition, Paris. 1996, pp.1527-1536.
- Syndicat mixte des milieux aquatiques et des rivières. *Pollution des eaux* récupéré sur : [En ligne]. [Consulté le 25/03/2020]. Disponible à l'adresse : <http://www.smmar.org/article/tout-savoir-sur-l-eau/inf-eau/pollutions-des-eaux>,
- THARRAULT, Yvon. *Diagnostic de fonctionnement par analyse en composantes principales : Application à une station de traitement des eaux usées*, Thèse de Doctorat Institut National Polytechnique de Lorraine, 2008, pp.65-74.
- XLstat. *Analyse en composantes principales*. [En ligne]. [Consulté le 20/04/2020]. Disponible à l'adresse :
<https://help.xlstat.com/s/article/analyse-en-composantes-principales-acp-avec-excel?language=fr>,
- XLstat. *Classification ascendante hiérarchique*. [En ligne]. [Consulté le 05/05/2020]. Disponible à l'adresse :
<https://help.xlstat.com/s/article/classification-ascendante-hierarchique-cah-dans-excel?language=fr>