



UNIVERSITE
JEAN LOROUGNON GUEDE
UFR ENVIRONNEMENT

REPUBLIQUE DE CÔTE D'IVOIRE

Union-Discipline-Travail

Ministère de l'Enseignement Supérieur et
de la Recherche Scientifique

ANNEE ACADEMIQUE :
2020-2021

N° D'ORDRE : 0452 /2021

N° CARTE D'ETUDIANT :

CI0414002824

MASTER

PHYSIQUE-CHIMIE

Option : Statistique

THEME :

**APPLICATION DES METHODES STATISTIQUES
MULTIVARIEES A L'ANALYSE DES PARAMETRES
PHYSICO-CHIMIQUES ET MICROBIOLOGIQUES DES
EAUX : CAS DES EAUX DE Puits DU VILLAGE M'PODY
(ABIDJAN, CÔTE D'IVOIRE)**

LABORATOIRE DES
SCIENCES ET
TECHNOLOGIES DE
L'ENVIRONNEMENT
(LSTE)

Présenté par :

N'GOUANDI Jean – Yves

JURY

- **PRESIDENT :** M. KOUASSI Kouakou Lazare, Professeur Titulaire,
Université Jean LOROUGNON GUEDE
- **DIRECTEUR :** M. YODE Armel Fabrice Evrard, Maître de Conférences,
Université Félix Houphouët-Boigny
- **ENCADREUR :** M. N'DRI Yao Aubin, Maître-Assistant, Université Jean
LOROUGNON GUEDE
- **EXAMINATEUR :** M. OKOU Gueï Cyrille, Maître-Assistant, Université
Jean LOROUGNON GUEDE

SOUTENU
PUBLIQUEMENT LE :

13 Octobre 2021



**UNIVERSITE
JEAN LOROUGNON GUEDE
UFR ENVIRONNEMENT**

REPUBLIQUE DE CÔTE D'IVOIRE

Union-Discipline-Travail

Ministère de l'Enseignement Supérieur et
de la Recherche Scientifique

ANNEE ACADEMIQUE :
2020-2021

N° D'ORDRE : 0452 /2021

N° CARTE D'ETUDIANT :

CI0414002824

MASTER

PHYSIQUE-CHIMIE

Option : Statistique

THEME :

**APPLICATION DES METHODES STATISTIQUES
MULTIVARIEES A L'ANALYSE DES PARAMETRES
PHYSICO-CHIMIQUES ET MICROBIOLOGIQUES DES
EAUX : CAS DES EAUX DE Puits DU VILLAGE M'PODY
(ABIDJAN, CÔTE D'IVOIRE)**

LABORATOIRE DES
SCIENCES ET
TECHNOLOGIES DE
L'ENVIRONNEMENT
(LSTE)

Présenté par :

N'GOUANDI Jean – Yves

JURY

- **PRESIDENT :** M. KOUASSI Kouakou Lazare, Professeur Titulaire,
Université Jean LOROUGNON GUEDE
- **DIRECTEUR :** M. YODE Arnel Fabrice Evrard, Maître de Conférences,
Université Félix Houphouët-Boigny
- **ENCADREUR :** M. N'DRI Yao Aubin, Maître-Assistant, Université Jean
LOROUGNON GUEDE
- **EXAMINATEUR :** M. OKOU Guei Cyrille, Maître-Assistant, Université
Jean LOROUGNON GUEDE

**SOUTENU
PUBLIQUEMENT LE :**

13 Octobre 2021

TABLE DES MATIERES

DEDICACES	III
REMERCIEMENTS.....	IV
LISTE DES ABREVIATIONS, ACRONYMES ET SIGLES	VI
LISTE DES TABLEAUX	VII
LISTE DES FIGURES	VIII
LISTE DES ANNEXES	IX
INTRODUCTION.....	1
PARTIE 1 : GENERALITES	3
1.1 PRESENTATION DE LA ZONE D'ETUDE.....	3
1.2 PARAMETRES DE L'EAU	3
<i>1.2.1 Paramètres physico-chimiques</i>	<i>4</i>
1.2.1.1 Ammonium (NH ₄ ⁺).....	4
1.2.1.2 Dureté Totale (DHT).....	4
1.2.1.3 Nitrates et nitrites	4
1.2.1.4 Sulfates	5
1.2.1.5 Température (T)	5
1.2.1.6 Titre alcalimétrique Complet (TAC).....	5
1.2.1.7 Turbidité (Turb).....	6
<i>1.2.2 Paramètres microbiologiques</i>	<i>6</i>
1.2.2.1 Coliformes totales	6
1.2.2.2 Entérocoques	6
1.2.2.3 Escherichia Coli	6
1.3 ANALYSE EXPLORATOIRE.....	7
<i>1.3.1 Analyse en composantes principales (ACP) (Lebart et al., 2006).....</i>	<i>7</i>
1.3.1.1 Objectif de l'ACP	7
1.3.1.2 Principe de l'ACP.....	8
1.3.1.3 Détermination des axes principaux	9
1.3.1.4 Caractéristiques des composantes principales.....	10
1.3.1.5 Procédé d'analyse	10
<i>1.3.2 Analyse Factorielle des Correspondances (AFC) (Lebart et al., 2006).....</i>	<i>12</i>
1.3.2.1 Jeu de données initiales	12
1.3.2.2 Transformation initiale des données	13
1.3.2.3 Détermination des composantes principales dans Rm	14
<i>1.3.3 Analyse de variance (ANOVA).....</i>	<i>16</i>
1.3.3.1 ANOVA à un facteur.....	16
1.3.3.2 Modèle.....	16
1.3.3.3 Tableau d'ANOVA	17
1.3.3.4 Test d'hypothèses	17
1.3.3.5 Comparaisons multiples	17
1.3.3.6 ANOVA non paramétrique	18

PARTIE 2: MATERIEL ET METHODES	19
2.1 MATERIEL	19
2.1.1 <i>Logiciels et langages de programmation</i>	19
2.1.2 <i>Présentation de données</i>	20
2.2 METHODE.....	21
2.2.1 <i>Analyses numériques</i>	21
2.2.2 <i>Méthode de l'analyse statistique descriptive</i>	22
2.2.3 <i>Méthode de l'ACP normée</i>	22
2.2.4 <i>Méthode de l'AFC</i>	24
2.2.5 <i>Méthode de l'ANOVA</i>	26
PARTIE 3: RESULTATS ET DISCUSSIONS.....	29
3.1 RESULTATS	29
3.1.1 <i>Résultats de l'analyse statistique descriptive</i>	29
3.1.2 <i>Résultats de l'ACP : analyse des puits</i>	30
3.1.2.1 <i>Distribution de l'inertie</i>	30
3.1.2.2 <i>Description des plans étudiés</i>	31
3.1.3 <i>Résultat obtenu avec l'AFC</i>	32
3.1.3.1 <i>Distribution de l'inertie</i>	33
3.1.3.2 <i>Description des plans étudiés</i>	33
3.1.4 <i>Résultats de l'ACP : paramètres physico-chimiques et microbiologiques</i>	35
3.1.5 <i>Résultat obtenu avec l'ANOVA</i>	36
3.2 DISCUSSION	38
CONCLUSION.....	40
REFERENCES	41
ANNEXES.....	44

Dédicaces

Ce mémoire est dédié à :

A l'Éternel des armées,

Dieu tu as pourvu à tout point de vue à la concrétisation de ce travail. Tu es toujours fidèle dans tes promesses, mon âme te bénis et te rend infiniment grâce. A toi seul la gloire aux siècles des siècles, amen.

A mes chers parents N'GOUANDI Kouassi et Amanzou Akoua Appoline,

En reconnaissance de tous les sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études. Que Dieu vous bénisse abondamment.

A tous mes frères et sœurs Sandrine, Christiane, Ehoussou et Evrard,

Pour vos encouragements et vos soutiens indéfectibles dans mes études. Que Dieu vous comble également.

A ceux qui n'ont jamais cessé de m'encourager, et de me conseiller.

A ceux qui n'ont jamais été avares ni de leur temps ni de leurs connaissances pour satisfaire mes interrogations.

Remerciements

Qu'il nous soit permis, au terme de ce travail, d'adresser notre reconnaissance à tous les enseignants de l'UFR environnement et de l'université Jean LOROUGNON GUEDE (UJLoG) de Daloa. Au cours de nos études, vous nous avez donné sans relâche avec clarté et précision, des connaissances et des conseils qui resteront gravés dans notre mémoire.

Nos profondes gratitude vont précisément :

A Madame TIDOU Abiba Sanogo Epse KONE,

Professeur Titulaire, Présidente de l'Université Jean LOROUGNON GUEDE, pour les moyens mis à notre disposition pour notre formation.

A Monsieur KOUASSI Kouakou Lazare,

Professeur Titulaire à l'Université Jean LOROUGNON GUEDE de Daloa, Directeur de l'Unité de Formation et de Recherche (UFR) Environnement, pour tous les efforts consentis en vue d'une meilleure réalisation de cette formation. Aussi, merci d'avoir accepté de présider le jury de la soutenance de ce mémoire.

A Monsieur YODE Armel Fabrice Evrard,

Maître de Conférences à l'Université Félix Houphouët-Boigny d'Abidjan, Directeur Scientifique de notre travail, pour sa disponibilité malgré ses multiples responsabilités, sa rigueur, et les moyens mis à notre disposition pour la réalisation de ce travail.

A Monsieur N'DRI Aubin,

Maître-Assistant à l'Université Jean LOROUGNON GUEDE de Daloa, encadreur de ce mémoire, pour ses judicieux conseils qui ont contribué à alimenter ma réflexion, ses orientations, son soutien, sa rigueur scientifique, sa patience et sa grande disponibilité, en un mot pour son encadrement. Encadrement sans lequel ce mémoire n'aurait été produit. Merci Docteur et que Dieu vous comble de sa grâce de jour en jour.

A Monsieur OKOU Gueï Cyrille,

Maître-Assistant à l'Université Jean LOROUGNON GUEDE de Daloa, pour avoir accepté d'être l'examineur de ce mémoire, grand merci pour sa rigueur et ses critiques objectives. Aussi, merci pour la pertinence de ses conseils méthodologiques et ses orientations dans le cadre de ce mémoire. Trouvez ici, monsieur, ma profonde reconnaissance pour votre contribution, votre soutien, la formation et l'encadrement apportés tout au long de ce travail et pour votre grande disponibilité. Grand merci. Que Dieu vous le rende au centuple.

A Monsieur DIBI Brou,

Maître de Conférences à l'Université JEAN LOROUGNON GUEDE de Daloa, Directeur du Laboratoire des Sciences et Technologies de l'Environnement (LSTE), qui nous a suivi depuis le début, dans notre travail. Il a suscité en nous une passion pour les questions relatives à la structure de notre document, l'adéquation entre les différentes parties. Nous le remercions infiniment pour sa disponibilité et ses critiques pertinentes. A travers lui nous remercions également tous les membres du LSTE pour leur accueil chaleureux au sein du laboratoire.

A Messieurs KEITA Kolé & ASSOHOON Stanislas,

Respectivement Maître-Assistant et Assistant, pour leur contribution à ce travail en termes de formation et conseils. Trouvez ici, Messieurs, ma profonde reconnaissance. Grand merci Messieurs et que Dieu vous le rende au centuple.

A tous nos confrères de la filière Physique Chimie Appliquée à l'Environnement,

Merci pour la fraternité, la cohésion et l'ambiance qui a régné entre nous pendant toute l'année mais aussi durant tout notre parcours universitaire.

Enfin, mes remerciements vont à l'endroit de tous ceux qui directement ou indirectement ont contribué par leurs conseils et compétence à la réalisation de ce mémoire et dont les noms ne sont pas personnellement cités. Ce travail, n'est en somme que la résultante des efforts des uns et des autres.

Liste des abréviations, acronymes et sigles

ACP : Analyse en Composantes Principales

ANOVA : Analyse de la Variance

AFC : Analyse Factorielle des Correspondances

Cl⁻ : Chlorures

CTH : Coliformes Thermotolérants

Cond : Conductivité

DHT : Dureté Totale

Efae: Enterocoques faecalis

Ecol: Escherichia coli

NH₄⁺ : ion ammonium

HCO₃⁻ : ion bicarbonate

NO₃⁻ : ion nitrates

NO₂⁻ : ion nitrites

M.D.S : Multidimensional Scaling

mg/L : milligramme par litre

NMF : Non Matrix Factorisation

OMS : Organisation Mondiale de la Santé

PM : Paramètres Microbiologiques

ppm : Partie par million

pH : Potentiel d'Hydrogène

PO₄³⁻ : Phosphates

SO₄²⁻ : Sulfates

T : Température

TAC : Titre Alcalimétrique Complet

Turb : Turbidité

Liste des Tableaux

Tableau I : Dureté de l'eau	4
Tableau II : Jeu de donnée initiale en AFC ((Tableau de contingence).....	12
Tableau III : Tableau de l'ANOVA.....	17
Tableau IV : Informations de la base de données	21
Tableau V : Fréquence de dépassement des normes	29
Tableau VI : Classement des puits selon leurs profils	32
Tableau VII : Matrice de corrélation entre les paramètres étudiés	35
Tableau VIII : Corrélations (Corr) et contributions (Contrib) des variables sur les différentes dimensions.....	36

Liste des Figures

Figure 1: Situation géographique de M'Pody	3
Figure 2: Jeu de données initiale en ACP	7
Figure 3: Passage des variables Xp aux nouvelles variables Fk	8
Figure 4: Projection orthogonale du nuage sur un sous-espace	9
Figure 5: Représentation plane des individus dans le premier plan principal	11
Figure 6: Cercle des corrélations.....	11
Figure 7: Projection simultanée en AFC	15
Figure 8: Icones des logiciels utilisés.....	20
Figure 9: Etapes de l'ACP.....	23
Figure 10: Courbe de l'inertie en fonction des valeurs propres	23
Figure 11: Etapes de l'AFC.....	25
Figure 12: Etapes de l'ANOVA	26
Figure 13: Décomposition de l'inertie totale (ACP)	31
Figure 14: Graphes des puits avec les différents 04 profils sur le plan (1,2) de l' ACP	32
Figure 15: Décomposition de l'inertie (AFC)	33
Figure 16: Graphe superposé en AFC sur les dimensions étudiées.....	34
Figure 17: Résultat des tests de Duncan.....	38

Liste des Annexes

Annexe 1 : Normes des paramètres de pollution physico-chimique et biochimiques de l'eau de boisson (OMS, 2009).....	41
Annexe 2 : Valeurs propres et variances en ACP	41
Annexe 3 : Projection des individus et variables sur les plans étudiés	42
Annexe 4 : Extraits des qualités de représentation et contributions des puits sur les cinq premières dimensions.....	43
Annexe 5 : Tableau des valeurs propres et variances en AFC.....	43
Annexe 6 : Représentation graphique du calcul de corrélations.....	43
Annexe 7 : Graphes des variables en ACP sur les dimensions étudiées.....	44
Annexe 8 : Résultat des tests de normalité	45
Annexe 9 : Résultat des tests de Kruskal-walis	46
Annexe 10 : Classement des puits par niveaux de concentration	46

INTRODUCTION

INTRODUCTION

L'eau est un bien indispensable aux hommes, animaux et végétaux. Tous les hommes doivent disposer d'une source d'approvisionnement en eau accessible, suffisante et sûre. Ainsi, un meilleur accès à une eau de boisson saine peut être bénéfique pour la santé. En effet, l'eau peut être source de maladie du fait de sa pollution par des rejets industriels, des eaux usées, des déchets ménagers ou agricoles, des excréta et divers déchets organiques (Scalon *et al.*, 2005 ; Eblin *et al.*, 2014). Selon l'Organisation Mondiale de la Santé (OMS), environ 1,1 milliard de personnes, n'ont pas accès à l'eau potable et 2,4 milliards n'ont pas accès à un système d'assainissement adéquat (OMS, 2004). La mise à disposition de l'eau potable au robinet à toutes les populations nécessite le captage, le contrôle ou le traitement et la distribution. Toutes ces opérations exigent de grands moyens techniques et financiers dont ne disposent pas les pays en voies de développement. En Côte d'Ivoire, cette tâche est la charge de la Société de Distribution d'eau de Côte d'Ivoire (SODECI). Les populations ayant accès à cette eau reste majoritairement réticentes quant à son utilisation comme eau de boisson du fait de sa couleur très souvent jaunâtre, des dépôts, de l'odeur de javel et de son mauvais goût. Les populations n'hésitent pas à recourir à d'autres sources d'eau à savoir les eaux de pluies, les cours d'eau et les puits. En Côte d'Ivoire, les eaux de puits sont couramment utilisées. En effet, certaines populations considèrent l'eau de puits comme étant plus saine en se basant uniquement sur son apparence incolore, sa limpidité et cela sans chercher à évaluer les paramètres physico-chimiques et microbiologiques. Or, les études réalisées sur la qualité de l'eau des puits dans le sud et à l'ouest de la Côte d'Ivoire ont montré qu'elle est impropre à la consommation (Yapo *et al.*, 2010 ; Ahoussi *et al.*, 2010). A titre d'exemple, en janvier 2020, une épidémie de diarrhée a été détectée dans le village M'Pody de la circonscription d'Anyama dans le sud de la Côte d'Ivoire suite à la consommation d'eau de puits (Gbagbo *et al.*, 2020). Selon l'OMS, plus de 2 millions d'enfants de moins de cinq ans, surtout des pays en développement où les mesures d'hygiène et d'assainissement sont insuffisantes, meurent chaque année des maladies diarrhéiques (Umesh *et al.*, 2003). Ces maladies diarrhéiques sont souvent liées à la mauvaise qualité des eaux de consommation et à un assainissement insuffisant des eaux usées. C'est pourquoi une étude a été menée sur les eaux de ces puits. Elle a montré que les eaux des puits de M'Pody sont à l'origine de cette épidémie car ces eaux ont une mauvaise qualité microbiologique (Gbagbo *et al.*, 2020). Ces auteurs ont mené une étude sur la qualité de ces eaux au moyen des méthodes électrochimiques et spectrophotométriques pour l'étude des paramètres physico-chimiques et d'une analyse microbiologique par filtration sur membrane. Ils ont fourni des mesures à adopter.

INTRODUCTION

Notre travail propose l'utilisation des techniques d'analyses statistiques multivariées pour analyser les paramètres de la qualité des eaux des puits du village M'Pody.

Pour atteindre l'objectif principal, il s'agit de :

- évaluer la qualité des paramètres physico-chimiques et microbiologiques l'eau des différents puits au vue des normes de l'OMS ;
- classier les puits selon leurs profils ;
- analyser l'impact des activités autour des puits sur les paramètres physico-chimiques et microbiologiques.

Pour mener à bien notre étude, nous présentons dans la partie I de ce mémoire, les connaissances sur la zone d'étude, les paramètres de l'eau et leurs normes, l'analyse exploratoire, la modélisation mathématique et le logiciel R. Ensuite, la partie II présente la démarche scientifique, c'est-à-dire le matériel utilisé et la méthodologie de travail. Enfin, la partie III présente les résultats obtenus et la discussion qui en découle.

PARTIE 1 :

GENERALITES

GENERALITES

1.1 PRESENTATION DE LA ZONE D'ETUDE

M'Pody est un village de la circonscription d'Anyama situé à 10 Km d'Abidjan en Côte d'Ivoire. Ses coordonnées géographiques sont approximativement 5°34'29" de latitude Nord et 4°14'8" de longitude Ouest (Figure 1).

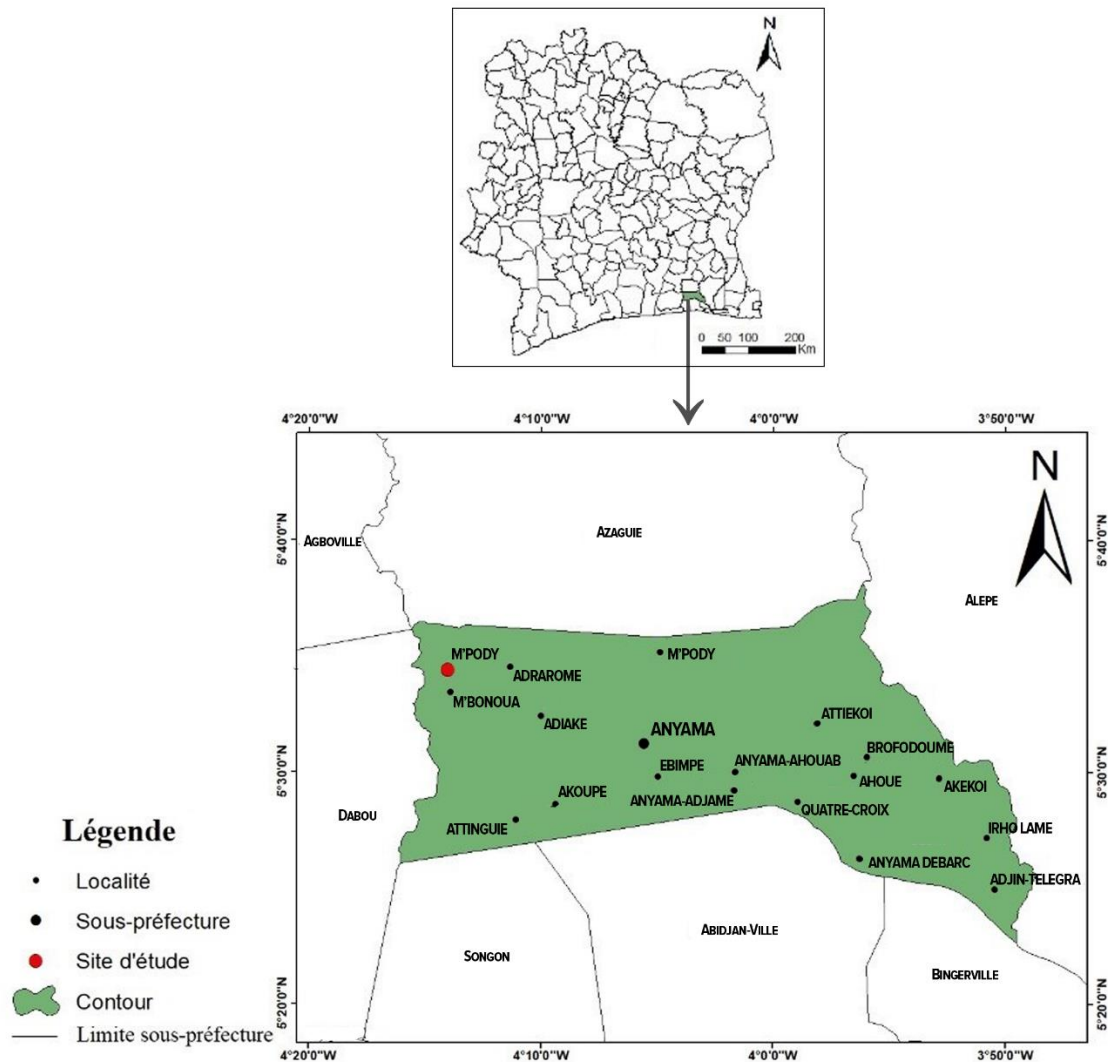


Figure 1: Situation géographique de M'Pody

1.2 PARAMETRES DE L'EAU

Pour savoir quelle technique utiliser s'il y a nécessité de traitement, on doit identifier les contaminants présents dans l'eau. C'est-à-dire faire une étude de pollution de cette eau. Ainsi, l'on définit trois (03) types de paramètres de pollution à savoir :

- les paramètres physico-chimiques : il s'agit entre autre de la température, l'oxygène dissout, le dioxyde de carbone, le pH, l'azote, la turbidité, la conductivité ;

GENERALITES

- les paramètres biochimiques : il s'agit de la Demande Biochimique en Oxygène (DBO) et de la Demande Chimique en Oxygène (DOC) ;
- les paramètres biologiques : il s'agit des espèces indicatrices, des bactéries.

Concernant notre étude, nous scindons les paramètres de l'eau des puits étudiés en trois (03) groupes : les paramètres physiques, les paramètres chimiques et les paramètres microbiologiques.

1.2.1 Paramètres physico-chimiques

1.2.1.1 Ammonium (NH_4^+)

Dans le traitement des eaux, il est important d'éliminer l'ammonium de l'eau parce qu'il réagit avec le chlore pour produire des chloramines, qui sont des désinfectants moins efficaces et peuvent provoquer des goûts désagréables. Certaines bactéries se développent aussi en transformant l'ammonium en nitrites puis en nitrates. L'ammonium n'a pas d'effet appréciable sur la santé du consommateur et sa présence dans les eaux est un indicateur de pollution. Il doit être éliminé dans les eaux de consommation car c'est un élément qui peut permettre à certaines bactéries de proliférer (Celerier *et al.*, 2002).

1.2.1.2 Dureté Totale (DHT)

La dureté totale ou titre hydrotimétrique d'une eau est une grandeur reliée à la somme des concentrations des cations métalliques (calcium, magnésium, aluminium, fer et strontium) présents dans l'eau. Les deux premiers cations (calcium Ca^{2+} et magnésium Mg^{2+}) sont généralement les plus abondants. La présence de ces deux cations dans l'eau tend souvent à réduire la toxicité des métaux (De Villers *et al.*, 2005). Une eau à titre hydrotimétrie élevée est dite « dure » (Tableau I). Dans le cas contraire, il s'agit d'une eau douce (Rodier *et al.*, 2009).

Tableau I : Dureté de l'eau

1°f = 10mg/L	0 à 8 °f	8 à 15 °f	15 à 30 °f	+ de 30 °f
	Eau très douce	Eau douce	Eau moyennement dure	Eau très dure

1.2.1.3 Nitrates et nitrites

L'azote peut se présenter sous un certain nombre de formes chimiques importantes telles que : l'azote organique, l'azote ammoniacal (NH_4^+), les nitrates (NO_3^-) et les nitrites (NO_2^-).

La présence des nitrates peut atteindre des taux anormalement élevés dans les zones ayant recours à l'utilisation des fertilisants agricoles ou contaminées par des excréta. Bien qu'ils soient moins toxiques, le principal danger des nitrates résulte de leur transformation en nitrites

GENERALITES

(dans l'appareil digestif) qui sont à l'origine de la méthémoglobinémie chez les nourrissons et les femmes enceintes.

Le nitrite ou azote nitreux est un élément toxique pour l'homme. La consommation d'une eau chargée de nitrite perturbe la fixation de l'oxygène par l'hémoglobine du sang (d'où la méthémoglobine). Les nitrites, une fois dans l'estomac peuvent se transformer en nitrosamines toxiques et cancérigènes (Chébékoué, 2008).

1.2.1.4 Sulfates

L'ion sulfate est l'un des anions les moins toxiques. Toutefois des concentrations élevées peuvent entraîner une déshydratation et une irritation gastro-intestinale. La présence de sulfates dans l'eau de boisson peut aussi lui communiquer un goût perceptible et contribuer à la corrosion du réseau de distribution. L'organisme humain s'adapte à des concentrations plus élevées en sulfates de l'eau de boisson. Le seuil de détection gustative est de 200-250 mg/L pour le sulfate de sodium (Na_2SO_4), 250900 mg/L pour le sulfate de calcium (CaSO_4) et 400-600 mg/l pour le sulfate de magnésium (MgSO_4) (OMS, 2017).

1.2.1.5 Température (T)

La température est un paramètre de l'eau. En effet, le réchauffement de l'eau favorise aussi le développement de certains microorganismes pathogènes pour l'homme ainsi que la diminution du taux de l'oxygène dissous (Pedoya, 1990). Elle influence la solubilité des gaz dans l'eau et la dissociation des sels dissous. 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 (De Villers *et al.*, 2005).

1.2.1.6 Titre alcalimétrique Complet (TAC)

Dans les eaux naturelles, le Titre alcalimétrique complet (Alcalinité) résulte le plus généralement de la présence d'hydrogénocarbonates, de carbonates et d'hydroxydes (Rodier *et al.*, 2009). En clair, le TAC mesure la somme des hydroxydes (OH^+), des carbonates (CO_3^-) et des bicarbonates (HCO_3^+). Ces derniers neutralisent les acides et les bases et limitent donc les variations du pH de l'eau. On dit aussi qu'ils ont un effet « tampon ». Ainsi, les variations du titre alcalimétrique complet ont un effet sur le pH (Qaouiyid *et al.*, 2016). L'alcalinité se mesure par la neutralisation d'un certain volume d'eau par une solution diluée d'acide sulfurique ou chlorhydrique. Le point d'équivalence étant déterminé par des indicateurs colorés. Il se mesure en °f (degrés français), ppm ou mg/L. Aussi, l'eau alcaline du fait de son pH élevé

GENERALITES

peut neutraliser l'acide gastrique dans l'estomac, elle peut donc soulager les troubles du tractus gastro-intestinal (Ignacio *et al.*, 2012).

1.2.1.7 Turbidité (Turb)

Généralement exprimée en unités de turbidité néphélométriques (UTN), la turbidité décrit la capacité de l'eau à se laisser traverser par la lumière. De plus, la turbidité peut être due à la mauvaise qualité de l'eau ou à un mauvais traitement de l'eau (OMS, 2017).

1.2.2 Paramètres microbiologiques

La qualité microbiologique de l'eau est évaluée en mesurant la présence de bactéries indicatrices de contamination fécale (Entérocoque faecalis, Escherichia Coli, Coliformes, etc...).

1.2.2.1 Coliformes totales

Les bactéries coliformes totales comprennent une grande diversité de bacilles aérobies et anaérobies facultatifs, non sporulant, capables de croître en présence de concentrations relativement élevées de sels biliaires, de fermenter le lactose et de produire des acides ou des aldéhydes dans les 24 heures à 35–37°C. Escherichia coli et les coliformes thermotolérants forment un sous-ensemble du groupe des coliformes totaux (OMS, 2017).

1.2.2.2 Entérocoques

Les entérocoques intestinaux sont un sous-groupe d'un groupe d'organismes plus grand définis comme streptocoques fécaux. Ce sont des bactéries relativement tolérantes au chlorure de sodium et à des pH alcalins. Elles sont anaérobies facultatives et se présentent isolées, par paires ou en courtes chaînes. Les entérocoques intestinaux sont présents dans beaucoup d'eaux usées et dans les environnements pollués par des eaux usées et des déchets humains et animaux.

1.2.2.3 Escherichia Coli

Il s'agit d'une bactérie que l'on trouve couramment dans le tube digestif de l'être humain et des organismes à sang chaud. La plupart des souches sont inoffensives. Certaines en revanche peuvent provoquer une intoxication. Escherichia coli se différencie des autres coliformes thermotolérants par sa capacité à produire de l'indole à partir du tryptophane ou par la production de β -glucuronidase.

1.3 ANALYSE EXPLORATOIRE

L'analyse exploratoire des données est une famille de méthodes statistiques dont les principales caractéristiques sont multidimensionnelles. Elle est utilisée pour traiter un nombre très important de données et de faire ressortir les aspects les plus importants de celles-ci. Les méthodes d'analyse exploratoire multidimensionnelle se classifient selon leur objectif (réduction de dimension ou classification) et le type des données à analyser (quantitatives et/ou qualitatives) :

- **Description et réduction de dimension (méthodes factorielles)**

- Analyse en Composantes Principales (p variables quantitatives) ;
- Analyse Factorielle Discriminante (p variables quantitatives, 1 variable qualitative) ;
- Analyse Factorielle des Correspondances (variables qualitatives) ;
- Analyse Canonique (p et q variables quantitatives) ;
- Multidimensional Scaling (M.D.S.) ;
- Non negative Matrix Factorisation ou NMF.

- **Méthodes de classification**

- Classification ascendante hiérarchique ;
- Algorithmes de réallocation dynamique ;
- Cartes de Kohonen (réseaux de neurones).

Pour cette étude, nous n'utiliserons que l'analyse en composantes principales, l'analyse factorielle des correspondances multiple.

1.3.1 Analyse en composantes principales (ACP) (Lebart *et al.*, 2006)

1.3.1.1 Objectif de l'ACP

On observe n individus de poids p_i sur p variables (caractères) quantitatives. L'ACP permet d'étudier les liaisons entre variables à l'aide de la notion de corrélation et les ressemblances entre individus à l'aide de la notion de distance. A partir des données, nous avons la matrice des p variables quantitatives observées sur les n individus (Figure 2).

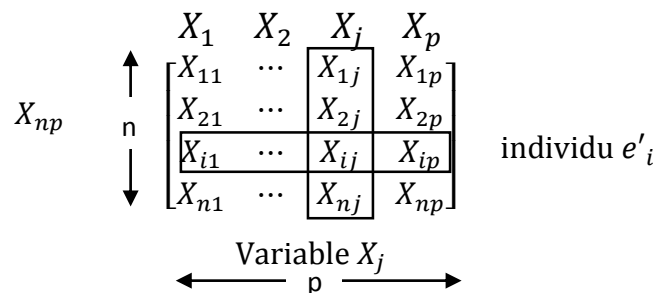


Figure 2: Jeu de données initiale en ACP

GENERALITES

On étudie chaque individu dans un espace de dimension p et chaque variable dans un espace de dimension n . On cherche donc à représenter le nuage des individus. A chaque individu noté e_i , on peut associer un point dans un espace \mathbb{R}^p . Cependant, lorsqu'on se retrouve dans un espace de dimension $p > 3$, la visualisation est impossible.

1.3.1.2 Principe de l'ACP

On cherche une représentation des n individus, dans un sous-espace F de \mathbb{R}^p de dimension k (petit). Autrement dit, la somme des carrés des distances des individus à F doit être minimale. Pour cela, on cherche à définir k nouvelles combinaisons linéaires des p variables initiales qui feront perdre le moins d'information possible. Ces variables seront appelées "composantes principales", les axes qu'elles déterminent : « axes principaux » et les formes linéaires associées : « facteurs principaux ».

En effet, on construit un ensemble de composantes ($F_1, F_2, \dots, F_k, \dots$), combinaisons linéaires des variables X_1, X_2, \dots, X_p (centrées et réduites), dont on peut apprécier la qualité de restitution de l'information (inerties ou dispersion maximales).

$$\begin{cases} F_1 = a_{11} z_1 + a_{21} z_2 + \dots + a_{p1} z_p \\ \vdots \\ F_k = a_{1k} z_1 + a_{2k} z_2 + \dots + a_{pk} z_p \\ \vdots \\ F_p = a_{1p} z_1 + a_{2p} z_2 + \dots + a_{pp} z_p \end{cases} \quad (1)$$

Où z_p est la variable p après centrage et réduction.

Elles permettent de calculer les coordonnées des individus dans le repère factoriel, et de juger de leur proximité dans les différents plans factoriels; on a :

$$F_{ik} = a_{1k} z_{i1} + a_{2k} z_{i2} + \dots + a_{pk} z_{ip} .$$

En outre, l'ACP revient à remplacer les variables X_1, X_2, \dots, X_p qui sont à priori corrélées en des nouvelles variables (F_1, F_2, \dots, F_k) non corrélées (Figure 3).

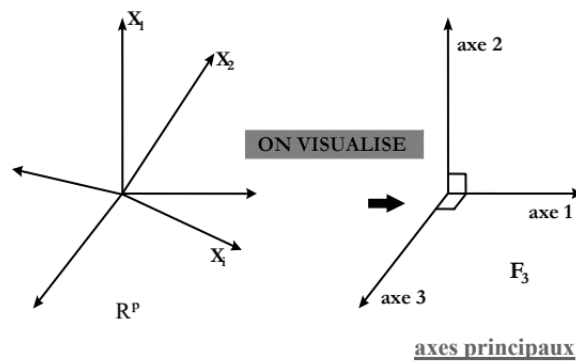


Figure 3: Passage des variables X_p aux nouvelles variables F_k

1.3.1.3 Détermination des axes principaux

Le principe sera de projeter les individus sur un espace de petite dimension k . Il faut donc calculer cette projection orthogonale pour la norme euclidienne sur un espace F .

- Inertie totale

L'inertie totale I_G est la somme pondérée des carrés des distances des individus au centre de gravité G . Elle mesure la dispersion totale du nuage de points. On a de façon générale

$$I_G = \sum_{i=1}^n p_i d^2(e_i, G) \quad \text{avec} \quad \sum_{i=1}^n p_i = 1. \quad (2)$$

En notant M la matrice de variances-covariances, l'inertie est donc aussi égale à :

$$M = \begin{bmatrix} s_1^2 & \cdots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{p1} & \cdots & s_p^2 \end{bmatrix}, \quad I_G = \sum_{i=1}^p s_i^2 = \text{Tr}(M). \quad (3)$$

Remarquons que, pour des variables sont centrées réduites, la variance de chaque variable vaut 1. L'inertie totale est alors égale à p (nombre de variables).

Notons que dans le cas d'une ACP normée, c'est la matrice de corrélation qui est utilisée.

- Comment trouver le sous – espace F ?

Sur la Figure 4, on a F étant un sous ensemble de l'espace \mathbb{R}^p , f_i la projection orthogonale de l'individu e_i sur F , on a d'après le théorème de Pythagore :

$$\|e_i - G\|^2 = \|e_i - f_i\|^2 + \|f_i - G\|^2 \quad \forall i = 1, \dots, n$$

$$\sum_{i=1}^n p_i \|e_i - G\|^2 - \sum_{i=1}^n p_i \|e_i - f_i\|^2 = \sum_{i=1}^n p_i \|f_i - G\|^2$$

On va chercher F tel que $\sum_{i=1}^n p_i \|e_i - f_i\|^2$ soit minimal.

Ce qui revient d'après le théorème de Pythagore à maximiser : $\sum_{i=1}^n p_i \|f_i - G\|^2$ (4)

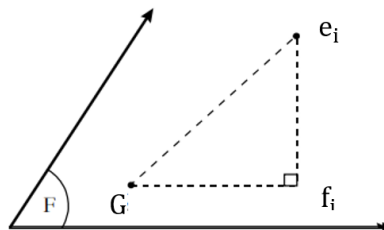


Figure 4: Projection orthogonale du nuage sur un sous-espace

1.3.1.4 Caractéristiques des composantes principales

Rechercher les axes portants le maximum d'inertie revient à construire de nouvelles variables (auxquelles sont associés ces axes) qui ont une variance maximale. En d'autres termes, on effectue un changement de repère dans \mathbb{R}^p de façon à se placer dans un nouveau système de représentation où le premier axe apporte le plus possible de l'inertie totale du nuage, le deuxième axe le plus possible de l'inertie non prise en compte par le premier axe, et ainsi de suite. Cette réorganisation s'appuie sur la diagonalisation de la matrice de variances-covariances.

Les axes principaux d'inertie sont les axes de direction des vecteurs propres de M normés à 1. A chacun de ces axes est associée une variable appelée composante principale. Sa variance est égale à l'inertie portée par l'axe principal qui lui est associé. En outre, les composantes principales sont non corrélées deux à deux car les axes associés sont orthogonaux.

1.3.1.5 Procédé d'analyse

- Qualité de représentation

Pour chaque individu, la qualité de sa représentation est définie par le carré du cosinus de l'angle entre l'axe de projection et le vecteur (5). Plus la valeur est proche de 1, meilleure est la qualité de représentation. En général, les qualités de représentation sont données axe par axe. Pour avoir la qualité de représentation dans un plan, on additionne les critères correspondant aux axes étudiés. Ce critère n'a pas de signification pour les individus proches de l'origine. Quand on détecte un individu pour lequel le cosinus carré est faible, on doit tenir compte de sa distance à l'origine avant d'indiquer qu'il est mal représenté.

$$\cos^2\theta = \cos^2\theta_1 + \cos^2\theta_2 \quad (5)$$

- Représentation des individus

La $j^{\text{ème}}$ composante principale C_j fournit les coordonnées des n individus sur le $j^{\text{ème}}$ axe principal. On désire une représentation plane des individus, la meilleure sera celle réalisée grâce aux deux premières composantes principales. Dans ce plan (Figure 5), nous allons nous intéresser à la qualité de représentation et la contribution de chaque individu.

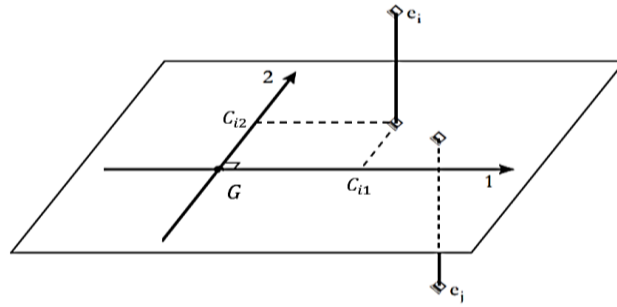


Figure 5: Représentation plane des individus dans le premier plan principal

- **Représentation des variables**

Les variables sont représentées dans un cercle des corrélations (Figure 6). Le cercle des corrélations est la projection du nuage des variables sur le plan des composantes principales.

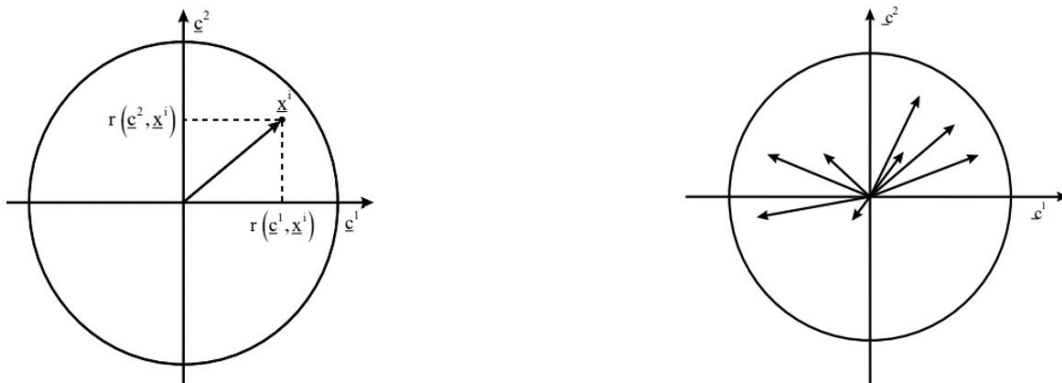


Figure 6: Cercle des corrélations

avec $r(C_j, X_i)$ est le coefficient de corrélation linéaire entre C_j et X_i .

En effet, les « proximités » entre les composantes principales et les variables initiales sont mesurées par les covariances, et surtout les corrélations $r(X_i, X_j)$. Les variables étant centrées

On a :

$$\langle X_i, X_j \rangle = \frac{1}{n} \sum_{k=1}^n X_{ik} X_{jk} = \text{Cov}(X_i, X_j) \quad (6)$$

$$\|X_i\|^2 = \langle X_i, X_i \rangle = \frac{1}{n} \sum_{k=1}^n X_{ik}^2 \quad (7)$$

$$\|X_i\|^2 = S_i^2 \quad : \text{Variance de } X_i \qquad \|X_i\| = S_i \quad \text{Ecart type de } X_i$$

$$\cos(\widehat{X_i, X_j}) = \frac{\langle X_i, X_j \rangle}{\|X_i\| \|X_j\|} = \frac{\text{Cov}(X_i, X_j)}{S_i S_j} = r(X_i, X_j) \quad (8)$$

Le cosinus de l'angle formé par les variables X_i et X_j est le coefficient de corrélation linéaire de ces deux variables. Ainsi, les variables bien représentées sont celles qui sont proches du cercle, celles qui sont proches de l'origine sont mal représentées.

1.3.2 Analyse Factorielle des Correspondances (AFC) (Lebart *et al.*, 2006)

L'analyse factorielle des correspondances (notée AFC), est une analyse destinée au traitement des tableaux de données où les valeurs sont positives, homogènes et qualitatives comme les tableaux de contingence. Elle a été introduite dans les années 60 par JP Benzecri. Son but principal est donc de décrire l'information contenue dans un espace multidimensionnel par une réduction de la dimension de cet espace tout en conservant un maximum d'information contenu dans l'espace de départ.

1.3.2.1 Jeu de données initiales

L'AFC s'applique essentiellement à des tableaux de contingence. C'est un tableau d'effectifs qui contient à l'intersection de la ligne i et de la colonne j des z_{ij} individus (Tableau II). Ce tableau donne des informations sur une population de taille N selon deux caractères qualitatifs nominaux et/ou ordinaux X en ligne et Y en colonne. L'étude d'un tel tableau se concentre le plus souvent sur la dépendance ou l'indépendance entre les deux caractères. Elle s'effectue généralement en utilisant le test du χ^2 et plus particulièrement par l'analyse de la variance (le rapport de corrélation) et la régression lorsque les deux caractères sont qualitatifs ordinaux en classes.

Le tableau des données $Z_{(n,m)}$ se présente alors de la façon suivante :

Tableau II : Jeu de donnée initiale en AFC ((Tableau de contingence)

$Z_{(n,m)} =$

Modalité de Y \ Modalité de X	1	...	j	...	m	Total
1	z_{11}	...	z_{1j}	...	z_{1m}	$z_{1.}$
...
I	z_{i1}	...	z_{ij}	...	z_{im}	$z_{i.}$
...
N	z_{n1}	...	z_{nj}	...	z_{nm}	$z_{n.}$
Total	$z_{.1}$...	$z_{.j}$...	$z_{.m}$	N

Avec $z_{i.} = \sum_{j=1}^m z_{ij}$, $z_{.j} = \sum_{i=1}^n z_{ij}$ $M = \sum_{i=1}^m \sum_{j=1}^n z_{ij} = \sum_{j=1}^n z_{.j}$

L'AFC s'intéresse plus particulièrement aux effectifs marginaux des tableaux que l'on appelle profils. Le tableau ci-dessus peut être alors transformé selon deux autres tableaux appelés tableaux de profils. Ainsi, de $Z_{(n,m)}$ on peut déduire deux matrices $X_{(n,m)}$ et $Y_{(n,m)}$.

$$X_{(n,m)}(I) = \left(\frac{z_{ij}}{z_{i.}} \right)_{1 \leq i \leq n, 1 \leq j \leq m} = \left(\frac{p_{ij}}{p_{i.}} \right)_{1 \leq i \leq n, 1 \leq j \leq m} \tag{a}$$

$$Y_{(n,m)}(J) = \left(\frac{z_{ij}}{z_{.j}} \right)_{1 \leq i \leq n, 1 \leq j \leq m} = \left(\frac{p_{ij}}{p_{.j}} \right)_{1 \leq i \leq n, 1 \leq j \leq m}. \quad (b)$$

Avec $p_{ij} = \frac{z_{ij}}{N}$; $p_{i.} = \frac{z_{i.}}{N}$; $p_{.j} = \frac{z_{.j}}{N}$ et p_{ij} : les fréquences relatives du tableau (les pourcentages).

La transformation du tableau $Z_{(n,m)}$ en un tableau $P_{(n,m)}$ est la première étape de l'AFC. On peut alors disposer de deux autres tableaux de sens différents : le tableau des profils lignes (a) et le tableau des profils colonnes (b).

1.3.2.2 Transformation initiale des données

La signification des mots individus et variables dans un tableau de contingence n'est pas la même que dans le tableau de l'ACP. En effet, dans le tableau de contingence les lignes et les colonnes représentent les modalités de deux caractères. Pour conserver une homogénéité dans la présentation des deux analyses, on pose par convention que les n modalités du caractère X en lignes portent le nom d'individu et que les m modalités du caractère Y en colonnes portent le nom de variable. Toutefois, le tableau de départ $Z_{(n,m)}$ peut-être écrit indifféremment avec la modalité X en ligne ou en colonne (pareillement pour Y) sans que la nature du tableau soit modifiée. Par contre, dans ce cas, les tableaux de profil ligne et colonne n'ont plus le même sens. Par analogie, l'AFC apparait donc comme une ACP. En effet, à partir de la matrice Z ou de ses transformées en matrices de profils, on peut considérer que l'information contenue dans le tableau peut être analysée à partir de deux espaces :

- l'espace R^m des « variables » (modalités colonnes) dans lequel on peut représenter le nuage des n points « individus » (modalité ligne). Chaque individu a pour coordonnée $X_{ij} = \frac{p_{ij}}{p_{i.}}$ et dans cet espace on utilise le tableau des profils lignes. Ainsi, on s'intéresse aux proximités relatives des points individus dans R^m .
- l'espace R^n des « individus » dans lequel on peut représenter les m points variables. Chaque variable a pour coordonnées $Y_{ij} = \frac{p_{ij}}{p_{.j}}$ et dans cet espace on utilise le tableau des profils colonne. Ainsi, on s'intéresse aux proximités relatives des variables dans R^n .

Dans le cas de l'ACP, l'information est donnée par la distance Euclidienne entre les points des nuages des deux espaces R^m et R^n .

1.3.2.3 Détermination des composantes principales dans R^m

- **Caractéristiques des « variables » et construction de la matrice d'information**

Comme pour l'ACP on se place dans l'espace des variables, on utilise donc la matrice des profils lignes. On vient de voir que dans cet espace les N points du nuage ont pour coordonnées :

$$x_{ij} = \frac{1}{\sqrt{P_j}} \frac{P_{ij}}{P_i} \tag{9}$$

On peut calculer les caractéristiques de ces variables (notées x_j pour $j=1$ à m) c'est-à-dire la moyenne et la covariance.

- La moyenne

$$\begin{aligned} \bar{x}_j &= \sum_i P_i \cdot X_{ij} \quad (\text{moyenne arithmétique pondérée}) \\ \bar{x}_j &= \sum_i P_i \cdot \frac{1}{\sqrt{P_j}} \frac{P_{ij}}{P_i} = \sum_i \frac{P_{ij}}{\sqrt{P_j}} = \frac{1}{\sqrt{P_j}} \sum_i P_{ij} = \frac{1}{\sqrt{P_j}} P_j \\ \bar{x}_j &= \sqrt{P_j} \end{aligned} \tag{10}$$

- La correspondance (entre deux variables x_j et $x_{j'}$) :

$$L(X_j, X_{j'}) = V_{jj'} = \sum_i P_i \cdot \left[\left(\frac{1}{\sqrt{P_j}} \frac{P_{ij}}{P_i} - \sqrt{P_j} \right) \left(\frac{1}{\sqrt{P_{j'}}} \frac{P_{ij'}}{P_i} - \sqrt{P_{j'}} \right) \right]$$

D'où
$$V_{jj'} = \sum_i \frac{P_{ij} P_{ij'}}{\sqrt{P_j} \sqrt{P_{j'}} P_i} - \sqrt{P_j} \sqrt{P_{j'}} \tag{11}$$

Remarquons qu'en faisant varier j et j' de 1 à m on construit alors la matrice $V(m, m)$ des variances/covariances des variables. C'est la matrice d'information des variables et par analogie avec l'ACP, l'étape suivante de l'AFC sera la diagonalisation de cette matrice.

- **Diagonalisation de la matrice des variances – covariances ou de la matrice d'inertie**

Le premier vecteur propre associé à cette première valeur propre définit un axe principal pour lequel les projections des individus et des variables possèdent une variance (dispersion) nulle. Ce qui signifie que toutes les projections possèdent les mêmes coordonnées. L'axe factoriel correspondant à cette valeur propre est donc exclu de l'analyse.

- **Les coordonnées des projections des individus et des variables sur les axes principaux**

Contrairement à l'ACP, en AFC, les projections sur les axes principaux du nuage des individus et des variables s'effectuent sur un même graphique (Figure 7).

$$F_{pi} = \sum_j F_{pj} \frac{P_{ij}}{P_i \sqrt{P_j}} = \sum_j \frac{P_{ij}}{P_i} a_{pj} \quad (12)$$

Cette relation permet de vérifier que les coordonnées de tous les individus sur l'axe principal qui a pour vecteur unitaire $u_{0j} = \sqrt{P_j}$ sont égales à 1. En effet, pour $p = 0$ on a :

$$F_{0i} = \sum_j u_{0j} \times \frac{P_{ij}}{P_i \sqrt{P_j}} = \sum_j \sqrt{P_j} \frac{P_{ij}}{P_i \sqrt{P_j}} = \frac{1}{P_i} \sum_j P_{ij} = \frac{P_i}{P_i} = 1. \quad (13)$$

Donc $F_{0i} = 1$ quel que soit i .

On pourrait vérifier avec les formules de transition que cette quantité a_{pj} correspond à la projection orthogonale de la variable j sur l'axe principal p noté dans cet espace a_p .

De ce fait, la formule : $F_{pi} = \sum_j \frac{P_{ij}}{P_i} a_{pj}$ n'est autre que le calcul du centre de gravité (de la moyenne pondérée) des coordonnées des projections des j variables.

D'où la propriété barycentrique de l'AFC : les coordonnées des projections orthogonales de chaque point i sont le barycentre (moyenne pondérée) des coordonnées des projections des points j . Et réciproquement, les coordonnées des projections de chaque point j sont le barycentre des projections des points i .

Au-delà de la formule, ce qui est important de retenir dans l'AFC, c'est la propriété barycentrique qu'elle révèle. Il s'agit d'une spécificité qui n'existe pas en ACP.

Ainsi on pourra dire par exemple que sur un plan factoriel choisi, un point i est d'autant plus proche d'un point j que le point j (variable) contribue le plus fortement possible au profil de l'individu i .

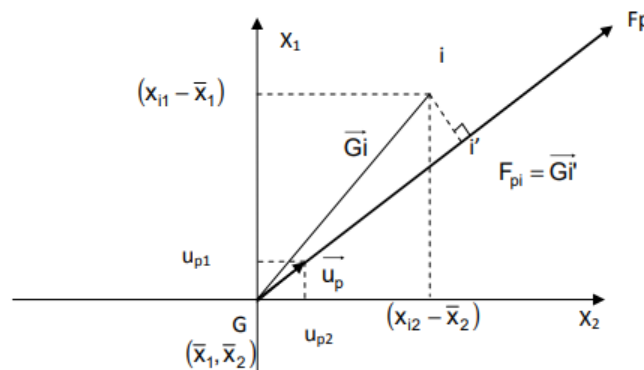


Figure 7: Projection simultanée en AFC

1.3.3 Analyse de variance (ANOVA)

L'ANOVA correspond à un modèle linéaire gaussien dans lequel toutes les variables explicatives (X_i) sont qualitatives. Dans ce contexte, elles sont appelées facteurs (d'où le terme de plans factoriels) et leurs modalités sont appelées niveaux. Ces niveaux sont supposés choisis, fixés, par l'utilisateur, de sorte que l'on parle souvent de facteurs contrôlés. De son côté, la variable aléatoire réponse Y est toujours quantitative et supposée gaussienne. L'analyse de la variance à k facteurs permet d'étudier le lien entre une variable quantitative continue et k variables qualitatives appelées facteurs. L'analyse de la variance est aussi considérée comme une généralisation du test de Student : elle permet de voir si la moyenne d'une variable quantitative est la même dans différents groupes (Azais & Bardet, 2005).

1.3.3.1 ANOVA à un facteur

On considère un facteur contrôle X présentant I modalités, chacune d'entre elles étant notée X_i . Pour chacune des modalités, nous effectuons $J \geq 2$ mesures d'une réponse Y . Ainsi, Le nombre total d'observations est $n = I \times J$.

1.3.3.2 Modèle

Notons

$$\bar{Y}_i = \frac{1}{J} \sum_{j=1}^J Y_{ij} : \text{la moyenne de la modalité } i \text{ du caractère } X$$

$$\bar{Y} = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J Y_{ij} : \text{la moyenne sur les } n \text{ observations}$$

Le modèle d'analyse de la variance s'écrit :

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ijk} \quad i = 1, \dots, I, \quad j = 1, \dots, J. \quad (14)$$

Avec : Y_{ij} l'observation j de la modalité i du facteur X

ε_{ij} le terme d'erreur sur l'observation j de la modalité i du facteur X

α_i l'effet propre de la modalité i du facteur X

GENERALITES

1.3.3.3 Tableau d'ANOVA

Les calculs effectués en ANOVA sont présentés dans le Tableau III suivant :

Tableau III : Tableau de l'ANOVA

Source	Variation	Degrés de liberté	Carré moyen	F
Facteur	SC_F	$I - 1$	$s_F^2 = \frac{SC_F}{I - 1}$	$F = \frac{s_F^2}{s_R^2}$
Résiduelle	SC_R	$n - I = I(J - 1)$	$s_R^2 = \frac{SC_R}{n - 1}$	
Totale	SC_{TOT}	$n - 1 = I \times J - 1$		

Nous avons trois sources de variation :

$$- \text{Variation due au facteur} \quad SC_F = J \sum_{i=1}^I (\bar{Y}_i - \bar{Y})^2. \quad (15)$$

$$- \text{Variation résiduelle} \quad SC_R = \sum_{i=1}^I \sum_{j=1}^J (Y_{ij} - \bar{Y}_i)^2. \quad (16)$$

$$- \text{Variation totale} \quad SC_{TOT} = \sum_{i=1}^I \sum_{j=1}^J (Y_{ij} - \bar{Y})^2. \quad (17)$$

Avec : $\bar{Y}_i = \frac{1}{J} \sum_{j=1}^J Y_{ij}$ la moyenne de la modalité i du caractère X

$\bar{Y} = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J Y_{ij}$ la moyenne sur les n observations

La relation fondamentale de l'ANOVA appelée équation de l'analyse de la variance est :

$$SC_{TOT} = SC_F + SC_R. \quad (18)$$

1.3.3.4 Test d'hypothèses

Nous souhaitons tester

$$\begin{cases} \mathbf{H}_0 : a_1 = a_2 = \dots = a_I = 0 & \text{(Absence d'effet du facteur } X) \\ \mathbf{H}_1 : a_1 = a_2 = \dots = a_I \neq 0 & \text{(Présence d'effet du facteur } X) \end{cases}$$

Nous pouvons aussi utiliser la p-valeur α^* pour conclure :

$$\alpha^* < \alpha \Rightarrow \text{on accepte } \mathbf{H}_1 \quad ; \quad \alpha^* > \alpha \Rightarrow \text{on accepte } \mathbf{H}_0$$

1.3.3.5 Comparaisons multiples

Lorsque l'hypothèse de Présence d'effet (\mathbf{H}_1) est acceptée, on procède à des comparaisons multiples pour détecter les groupes qui sont significativement différents.

1.3.3.6 ANOVA non paramétrique

Dans le cas où les hypothèses de validité, parfois assemblées sous l'acronyme "iid", qui veut dire "indépendant et identiquement distribué", ne sont pas vérifiées, l'ANOVA à un facteur n'est pas adaptée.

- Cas de non indépendance

Deux types de non-indépendance peuvent être rencontrés. Le premier est une non-indépendance des données à l'intérieur d'une modalité du facteur étudié. Ce genre de non-indépendance peut être mis en évidence par le test de Durbin Watson. Un deuxième cas de figure est celui de la non-indépendance des données entre les modalités. En cas de non-indépendance, l'ANOVA à un facteur n'est pas adaptée pour comparer les moyennes de k groupes. Dans cette situation, d'autres méthodes, comme des modèles mixtes doivent être employées (Zuur *et al.*, 2009).

- Cas de non homogénéité et/ou normalité

Si l'hypothèse de normalité et/ou l'hypothèse d'homogénéité des résidus est rejetée, deux solutions simples peuvent être appliquées. La première est l'utilisation d'une ANOVA non paramétrique, aussi appelée test de Kruskal-Wallis. Les comparaisons multiples (comparer toutes les moyennes deux à deux : procédure de Tukey ou comparer toutes les moyennes versus la moyenne du groupe contrôle : procédure de Dunnett) sont alors réalisées avec des tests de Wilcoxon (qui sont des tests non paramétriques). L'autre solution envisageable est une ANOVA par un test de permutation. Les comparaisons multiples étant réalisées comme précédemment, à l'aide de tests de Wilcoxon (Ruxton & Beauchamp, 2008).

PARTIE 2 :
MATERIEL ET
METHODES

MATERIEL ET METHODES

L'objectif de ce travail est d'utiliser des techniques d'analyses statistiques multivariées pour étudier la qualité des eaux des puits du village M'Pody. Pour cela, il nous faut impérativement des données de terrain sur ces puits, le matériel et la méthode adéquats. Les données terrain, le matériel et la méthodologie utilisés sont décrits dans cette partie.

2.1 MATERIEL

Pour la réalisation du présent travail, nous avons utilisé un matériel composé de :

- ✓ données d'analyses des mesures (paramètres physiques, chimiques et paramètres microbiologiques) au niveau des différents puits effectuées en 2020 ;
- ✓ outils informatiques (logiciels);
- ✓ langage de programmation.

2.1.1 Logiciels et langages de programmation

Les logiciels systèmes et applicatifs utilisés pour la réalisation du présent travail ainsi que les langages de programmation utilisés sont les suivants (Figure 8) :

- **Logiciels applicatifs et système**

- ✓ **Edition Windows** : Windows 10 Professionnel (Système d'exploitation installé sur le micro-ordinateur) ;
- ✓ **Type de système** : système d'exploitation 64 bits, processeur *64 ;
- ✓ **Microsoft Excel 2016** : jeu de données.

- **Langage de programmation R**

Le logiciel R est un environnement logiciel libre (open-source) pour le calcul statistique et graphique. Il a été créé par Ross Ihaka et Robert Gentleman (Ihaka & Gentleman, 1996). Il est à la fois un langage informatique et un environnement de travail. Ce logiciel sert à manipuler des données, à tracer des graphiques et à faire des analyses statistiques sur ces données (Lafaye *et al.*, 2011). Il fonctionne sous Linux, Windows et Macintosh. C'est aussi un outil très puissant et très complet, particulièrement bien adaptés pour la mise en œuvre informatique de méthodes statistiques. Il est plus difficile d'accès que certains autres logiciels du marché (comme SPSS ou Minitab par exemple), car il n'est pas conçu pour être utilisé à l'aide de « clics » de souris dans des menus. R est un logiciel dans lequel de nombreuses techniques statistiques modernes et classiques ont été implémentées sous forme de codes appelés "fonctions". Les méthodes les plus courantes permettant de réaliser une analyse statistique telles que la statistique descriptive;

MATERIEL ET METHODES

les tests d'hypothèses ; l'analyse de la variance ; les méthodes de régression linéaire simple et multiple ; etc... sont enchâssées directement dans le cœur du système. Notez également que la plupart des méthodes avancées de statistique sont aussi disponibles au travers de modules externes appelés "packages". Ceux-ci sont faciles à installer directement à partir d'un menu du logiciel. Ils sont tous regroupés sur le site internet du *Comprehensive R Archive Network* (CRAN) (<http://cran.r-project.org>) sur lequel vous pouvez les consulter. Ce site fournit aussi, pour certains grands domaines d'étude, une liste commentée des packages associés à ces thèmes (appelée *Task View*), ce qui facilite ainsi la recherche d'une méthode statistique particulière. Par ailleurs, une documentation détaillée en anglais de chaque package est disponible sur le CRAN. Il est par ailleurs utile de noter que les méthodes statistiques les plus récentes y sont régulièrement ajoutées par la communauté statistique elle-même.

Pour les analyses statistiques, nous avons utilisé le logiciel « R 4.0.4 » associés à « RStudio 1.3.1093 » et des packages. Nous avons entre autres :

- dplyr : pour renommer nos données ;
- FactoMineR : pour l'analyse exploratoire multidimensionnelle ;
- factoextra : pour extraire et visualiser les résultats des analyses exploratoires de données Multivariées ;
- ggplot2 : pour la visualisation des données ;
- pastecs : pour la statistique descriptive ;
- tidyverse, ggpubr, rstatix : pour la réalisation de l'ANOVA ;
- Summarytools : fournit des outils pour résumer les données de manière nette et rapide.



Figure 8: Icones des logiciels utilisés

2.1.2 Présentation de données

La base de données utilisées dans ce mémoire est constituée des résultats des analyses des paramètres physiques, chimiques et microbiologique réalisées sur des échantillons d'eau de soixante-douze (72) puits du village M'Pody en Côte d'Ivoire. En effet, elle se compose de la moyenne sur chacun des paramètres étudiés au cours de l'année 2020. En effet, les analyses terrains ont été réalisées par des méthodes électrochimiques et spectrophotométriques. Ces analyses ont été effectuées sur seize paramètres dont treize (13) paramètres physico-chimiques

MATERIEL ET METHODES

et trois (03) paramètres microbiologiques. Ainsi nous avons une matrice de [72, 16] données quantitatives.

Le tableau IV présente un résumé des informations de cette base de données.

Tableau IV : Informations de la base de données

Désignations des paramètres	Notations utilisées	Nombre de données	Valeurs nulles	Données manquantes	Minimum	Maximum	Range = max-min	Somme des valeurs	Médiane	Moyenne	Erreur standard de la moyenne	Variance	Ecart type	Unités de mesure
Turbidité	Tur	72	0	0	2,71	162,43	159,73	1613,32	16,04	22,41	2,78	554,62	23,55	NTU
Conductivité	Cond	72	0	0	24,73	594,00	569,28	11359,88	133,29	157,78	14,27	14669,47	121,12	µS/cm
Potentiel d'hydrogène	pH	72	0	0	4,22	6,34	7,30	375,44	5,08	5,21	0,10	0,78	0,88	-
Température	T	72	0	0	25,88	29,08	3,20	2001,38	28,03	27,80	0,10	0,68	0,83	°C
Nitrates	NO3-	72	0	0	1,11	44,16	43,05	1015,19	12,38	14,10	0,90	57,87	7,61	mg/L
Nitrites	NO2-	72	0	0	0,01	0,54	0,53	5,79	0,04	0,08	0,01	0,01	0,11	mg/L
Ammonium	NH4+	72	0	0	0,01	3,82	3,81	25,64	0,25	0,36	0,06	0,24	0,48	mg/L
Phosphates	PO43-	72	0	0	0,03	0,75	0,73	7,96	0,06	0,11	0,02	0,02	0,13	mg/L
Chlorures	Cl-	72	0	0	1,75	44,70	42,95	1308,70	17,16	18,18	1,16	96,09	9,80	mg/L
Titre Alcalimétrique Complet (Alcaliné)	TAC	72	0	0	26,25	216,25	190,00	5334,08	65,00	74,08	4,74	1618,24	40,23	mg/L
Durété Totale	DHT	72	0	0	5	106,25	101,25	1856,25	19,25	25,78	2,45	432,52	20,80	mg/L
Sulfates	SO42-	72	0	0	2,25	40,50	38,25	769,00	7,88	10,68	0,92	60,33	7,77	mg/L
Bicarbonates	HCO3-	72	0	0	30	216,25	186,25	6531,25	73,75	90,71	5,06	1840,00	42,90	mg/L
Coliformes Thermotolérants	CTH	72	0	0	4,25	7400,00	7395,75	87824,00	818,88	1219,78	149,65	1612423,04	1269,81	UFC/100ml
Escherichia coli	Ecol	72	0	0	2	5650,00	5648,00	68245,50	658,75	947,85	123,75	1102693,73	1050,09	UFC/100ml
Escherichia faecalis	Efae	72	0	0	3,5	3913,00	3909,50	67397,50	692,50	936,08	122,95	1088379,87	1043,25	UFC/100ml

2.2 METHODE

2.2.1 Analyses numériques

Les données des analyses des paramètres physiques, chimiques et microbiologiques ont été enregistrées dans un classeur Excel. Une fois sur R, nous cherchons à avoir un premier aperçu sur la situation des eaux de ces puits en évaluant les normes de leurs paramètres. Pour ce faire, nous procédons par une analyse statistique descriptive des paramètres physiques, chimiques et microbiologiques des eaux de puits de M'Pody. Ensuite, nous cherchons à classer ces puits selon différents profils. Pour ce faire, nous combinons l'analyse en composantes principales (ACP) et l'analyse factorielles des correspondances (AFC). En effet, l'ACP étudie les

MATERIEL ET METHODES

ressemblances entre les puits à partir du graphe des individus et l'AFC, à partir du graphe superposé des individus et variables afin de voir les liaisons entre les puits et les paramètres physico-chimiques et microbiologiques étudiés. Par la suite, nous cherchons à évaluer l'effet des puits sur les paramètres des eaux des puits. En d'autres termes, nous voulons savoir s'il existe des facteurs externes aux puits eux-mêmes qui influencent les paramètres des eaux de ces puits. Cette étude nous permettra de vérifier si les effets de la pollution causée par les bactéries se valent au niveau de tous les puits. Pour ce faire, nous utilisons l'analyse de variance (ANOVA). Les puits étant classés par groupes, deux individus ayant le même profil réagiront pareillement, c'est-à-dire que la différence de variance par rapport aux paramètres microbiologiques ne sera pas significative : elle sera nulle.

2.2.2 Méthode de l'analyse statistique descriptive

Les institutions Ivoirienne en charge de l'eau suivent les directives de l'Organisation Mondiale de la Santé (OMS) relatives à la qualité de l'eau de consommation. Par ailleurs, la réglementation de l'OMS sur les eaux de consommation publique impose la recherche d'indicateurs de contamination fécale (l'*Escherichia Coli*) en général. Selon cette réglementation, l'eau conditionnée ne doit comporter aucune de ces bactéries (Blé *et al.*, 2015). Ainsi pour l'analyse statistique descriptive menée dans cette étude, nous nous référons aux normes de l'OMS telles que données dans l'Annexe 1. Faire de l'analyse statistique descriptive revient à décrire, c'est-à-dire résumer ou représenter, par les statistiques, un ensemble relativement important de données.

Nous utilisons la fonction *stat.desc()* du package "pastecs" pour réaliser une statistique descriptive de nos données. Les normes de l'OMS servent de référence pour réaliser la détection des paramètres hors normes.

2.2.3 Méthode de l'ACP normée

La figure 9 présente les différentes étapes de la méthode de l'ACP sous R.

L'ACP est réalisé sous R à partir de la fonction *PCA()* du package "FactoMineR" et les résultats visualisés à l'aide de plusieurs autres fonctions du package "factoextra". En effet, les données sont importées dans RStudio sous forme de "data frame". Les paramètres n'ayant pas la même unité de mesure, il faut centrer-réduire via la formule suivante :

$$\frac{X_{ik} - \bar{X}_k}{S_k} \quad (19)$$

où \bar{X}_k et S_k sont respectivement la moyenne et l'écart-type de la variable X_k .

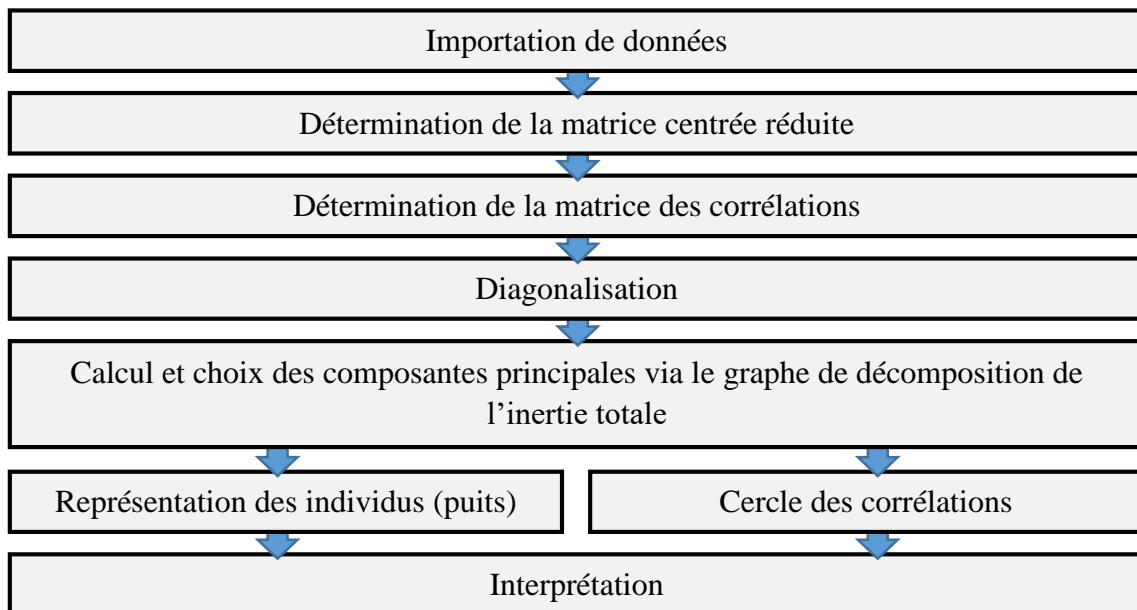


Figure 9: Etapes de l'ACP

Ensuite la matrice de corrélation est déterminée. La diagonalisation de cette matrice conduit à des valeurs propres qui nous donnent le nombre d'axes principaux. A chaque axe est associé une variable appelée composante principale. Cette dernière renferme les coordonnées des projections de l'individu sur l'axe considéré (voir formule 1). Puis nous nous intéressons à l'inertie portée par chaque axe principal pour faire le choix des plans principaux à analyser. Pour ce choix, il existe différentes procédures :

- Règle de la part d'inertie : on décide de garder q tel que la part d'inertie expliquée est supérieure à un seuil fixé a priori (généralement fixé à 70%);
- Règle de Kaiser : on ne conserve que les axes dont la valeur propre est supérieure à leur moyenne car ce sont les seules jugées informatives;
- Eboulis des valeurs propres (Figure 10) : on sélectionne q en localisant le « cassure » dans la courbe de l'inertie en fonction des valeurs propres.

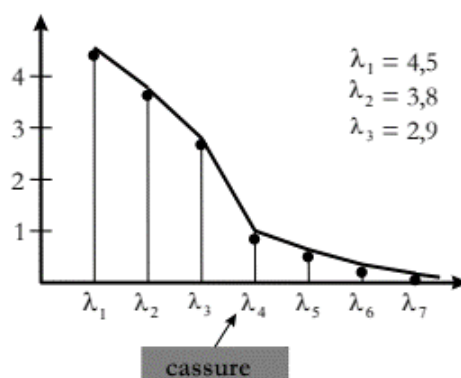


Figure 10: Courbe de l'inertie en fonction des valeurs propres

MATERIEL ET METHODES

Ici nous choisissons le nombre d'axe qui nous permet d'avoir une variance cumulée supérieure ou égale à 70%. Enfin, les graphes des puits et des paramètres sont tracés respectivement à partir des fonctions *fviz_pca_ind* et *fviz_pca_var* de "factoextra" suivant les différents plans factoriels sélectionnés. L'analyse de ces plans, en nous servant des notions de proximité entre individus, de contribution, de corrélation et qualité de représentation, nous conduit à des résultats qui seront interprétés.

- Notion de distance entre individu

Dans l'espace R^p , la distance euclidienne entre l'individu $e_i = (X_{i1}, X_{i2}, \dots, X_{ip})$ et $e_j = (X_{j1}, X_{j2}, \dots, X_{jp})$ s'écrit :

$$d^2(e_i, e_j) = \sum_{k=1}^p (X_{ik} - X_{jk})^2. \quad (20)$$

- Contribution

Il est très utile aussi de calculer pour chaque axe la contribution apportée par les divers individus à cet axe.

Considérons la $k^{ième}$ composante principale C_k , soit C_{ik} la valeur de la composante pour le $ième$ individu.

$$\sum_{i=1}^n \frac{1}{n} (C_{ik})^2 = \lambda_k \quad (21)$$

La contribution de l'individu e_i à la $k^{ième}$ composante est définie par :

$$\frac{\frac{1}{n} (C_{ik})^2}{\lambda_k} \quad (22)$$

2.2.4 Méthode de l'AFC

L'AFC est appliquée dans l'objectif d'analyser les relations liant les puits aux paramètres physiques, chimiques et microbiologiques. En d'autres termes, l'objectif est de déterminer les paramètres qui influencent les puits. Ainsi, deux puits sont considérés comme proches si elles s'associent de la même façon à l'ensemble des paramètres, c'est-à-dire si elles s'associent trop (ou trop peu) aux mêmes colonnes ; les termes « trop » et « trop peu » sont pris en référence à la situation d'indépendance (Escofier & Pagès, 2008). Symétriquement, deux paramètres sont proches si elles s'associent de la même façon à l'ensemble des puits. Schématiquement, l'étude de l'ensemble des puits revient à mettre en évidence une typologie dans laquelle on cherche les

puits dont la répartition s'écarte le plus de celle de l'ensemble de la population, celles qui se ressemblent entre elles et celles qui s'opposent. Le principe reste plus ou moins pareil à l'ACP.

- Mode opératoire

Le figure 11 présente les différentes étapes de la méthode de l'AFC sous R.

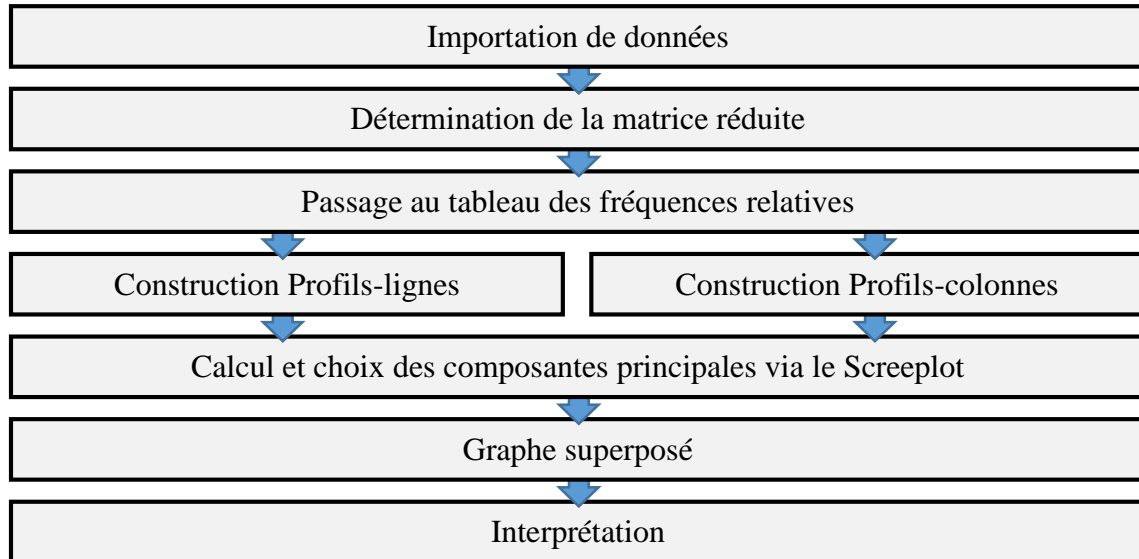


Figure 11: Etapes de l'AFC

L'AFC est réalisé sous R à partir de la fonction $CA()$ du package "FactoMineR". Les résultats visualisés à l'aide de plusieurs autres fonctions du package "factoextra". En effet, les données sont réduites dans RStudio sous forme de "data frame". Nous passons au tableau des fréquences relatives puis à la construction des Profils-lignes et Profils-colonnes (voir 1.3.2.2). Tout comme en ACP, les composantes principales choisies sont celles dont le pourcentage d'inertie cumulée atteint les 70%. La représentation graphique des puits et paramètres suivant le plan factoriel sélectionné est interprétée. Il s'agit d'un graphe représentant à la fois les liaisons entre individus et variables. Notons que la notion de proximité utilisé est la métrique du χ^2 que nous présentons brièvement.

► Passage de la distance euclidienne à la métrique du χ^2

Dans l'espace R^m , calculons la distance euclidienne entre deux points quelconques : x_i et $x_{i'}$:

$$d^2(x_i, x_{i'}) = \sum_{j=1}^m \left(\frac{P_{ij}}{P_{i.}} - \frac{P_{i'j}}{P_{i' .}} \right)^2 .$$

MATERIEL ET METHODES

Cette distance est utilisée en ACP. Par contre, on ne l'utilise pas exactement en AFC. Plus précisément on l'utilise après avoir effectué une transformation euclidienne préalable des coordonnées des points du nuage. En effet, cette transformation s'écrit :

$$x_{ij} = \frac{1}{\sqrt{P_j}} \frac{P_{ij}}{P_{i.}}$$

On obtient :

$$d^2(x_i, x_{i'}) = \sum_{j=1}^n \left(\frac{1}{\sqrt{P_j}} \frac{P_{ij}}{P_{i.}} - \frac{1}{\sqrt{P_j}} \frac{P_{i'j}}{P_{i'.}} \right)^2 = \sum_{j=1}^n \frac{1}{P_j} \left(\frac{P_{ij}}{P_{i.}} - \frac{P_{i'j}}{P_{i'.}} \right)^2 \quad (23)$$

On procède de la même façon dans l'espace R^n . Considérons deux points quelconques : y_j et $y_{j'}$ de cet espace. La transformation s'écrit :

$$y_{ij} = \frac{1}{\sqrt{P_{i.}}} \frac{P_{ij}}{P_j}$$

On obtient :

$$d^2(y_i, y_{i'}) = \sum_{j=1}^n \left(\frac{1}{\sqrt{P_{i.}}} \frac{P_{ij}}{P_j} - \frac{1}{\sqrt{P_{i'.}}} \frac{P_{i'j}}{P_{j'}} \right)^2 = \sum_{j=1}^n \frac{1}{P_{i.}} \left(\frac{P_{ij}}{P_j} - \frac{P_{i'j}}{P_{j'}} \right)^2 \quad (24)$$

On remarque que l'application de la distance euclidienne sur les données transformées est équivalente à l'application de la métrique du χ^2 sur les données non transformées. En outre, en AFC, la ressemblance entre deux lignes ou entre deux colonnes est définie par une distance entre leurs profils connus sous le nom de distance du χ^2 .

2.2.5 Méthode de l'ANOVA

Le figure 12 présente les différentes étapes de la méthode de l'ANOVA sous R.

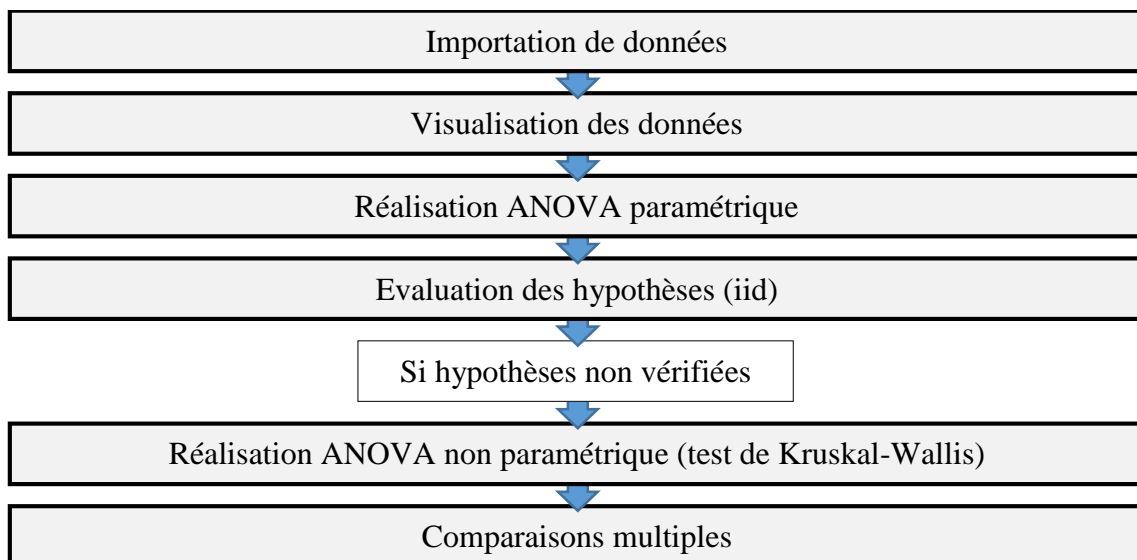


Figure 12: Etapes de l'ANOVA

L'ANOVA a été utilisée dans l'objectif d'étudier l'effet des paramètres microbiologiques (variables quantitatives) en fonction des différents puits, déterminés avec l'ACP. En clair, nous cherchons à savoir si les puits en eux-mêmes influencent les paramètres microbiologiques ou pas. Ce qui nous emmène à tester la significativité des différences entre les moyennes. Pour ce faire, nous devons comparer (c'est-à-dire analyser) les variances.

Nous avons sélectionné 03 paramètres pour la réalisation de l'ANOVA (Cond, TAC et Ecol). L'ANOVA non paramétrique (en ce qui nous concerne) ou le test de Kruskal-Wallis est réalisé sous R à partir de la fonction *kruskal.test()*. En effet, les données sont importées sous R puis visualisées via la fonction *ggplot()* du package "ggplot2" pour avoir une idée du comportement des résidus. Nous réalisons ensuite un test de normalité des résidus avec un test numérique (*Shapiro.test()*) et graphique (*hist()*). Si le test échoue c'est-à-dire l'ANOVA paramétrique est ensuite réalisée. Puis vient l'évaluation des hypothèses c'est-à-dire que nous vérifions que les résidus sont "iid". Ces hypothèses sont décrites dans la suite.

- Hypothèses de validité

Pour que les résultats de l'ANOVA à un facteur soient valides, nous postulons les trois hypothèses suivantes qui doivent être vérifiées :

- les résidus doivent être distribués de façon homogène. Autrement dit, les variances des résidus relatives aux différentes modalités de la variable explicative (ou indépendante) ne doivent pas être trop différentes.

$$\sum_{i=1}^I \alpha_i = 0. \quad (25)$$

- les réponses observées, pour une modalité donnée du facteur étudié, doivent être indépendantes les unes des autres, et indépendantes des réponses des autres modalités

$$\forall (i, j) \neq (k, l), \varepsilon_{ij} \text{ et } \varepsilon_{kl} \text{ sont indépendantes.} \quad (26)$$

- les résidus doivent être distribués selon une loi Normale, de moyenne 0.

$$\forall (i, j), 1 \leq i \leq I, 1 \leq j \leq J, \varepsilon_{ij} \text{ suit une loi normale } \mathcal{N}(0, \sigma^2). \quad (27)$$

Si ces hypothèses ne sont pas validées alors nous comparons les moyennes à l'aide d'une ANOVA non paramétrique. Ce qui nous conduit à une p-value. Enfin, si la p-value du test de Kruskal-Wallis est inférieure à 0,05, l'hypothèse de l'égalité des moyennes est rejetée. On conclut donc que les moyennes des groupes de puits sont globalement différentes. Afin

MATERIEL ET METHODES

d'évaluer quelles sont les moyennes qui diffèrent entre elles, nous réalisons des comparaisons multiples. Lorsque les hypothèses de normalité et d'homogénéité des résidus sont rejetées, ces comparaisons multiples sont réalisées avec des tests non-paramétriques.

PARTIE 3 :
RESULTATS ET
DISCUSSION

RESULTATS ET DISCUSSION

3.1 Résultats

Les résultats obtenus seront présentés dans cette partie du mémoire. Notons que les figures et tableaux de grandes tailles seront mis en annexe.

3.1.1 Résultats de l'analyse statistique descriptive

L'évaluation de la qualité des eaux des puits a donné les résultats du tableau V.

Tableau V : Fréquence de dépassement des normes

Désignations des paramètres	Notations utilisées	Moyennes Totales	Normes OMS	Fréquences des eaux de puits conformes (%)	Moyennes atteintes des eaux de puits conformes	Fréquences des eaux de puits non conformes (%)	Moyennes atteintes des eaux de puits non conformes	
Paramètres physico-chimiques	Turbidité	Tur	22,41 NTU	< 5 NTU	4,17	3,38 NTU	95,83	23,267 NTU
	Conductivité	Cond	157,7760 μ S/cm	180-1000 μ S/cm	27,78	312,6325 μ S/cm	72,22	98,2158 μ S/cm
	Potentiel d'hydrogène	pH	5,21	6,5 - 8,5	0	--	100	5,21
	Température	T	27,79°	\leq 25°C	0	--	100	27,79°
	Nitrates	NO3-	14,09 mg/l	\leq 50 mg/l	100	14,09 mg/l	0	--
	Nitrites	NO2-	0,080 mg/l	\leq 3 mg/l	100	0,080 mg/l	0	--
	Ammonium	NH4+	0,356 mg/l	\leq 0,5 mg/l	80,56	0,210 mg/l	19,44	0,958 mg/l
	Phosphates	PO43-	0,111 mg/l	0,5 mg/l	97,22	0,095 mg/l	2,78	0,6475 mg/l
	chlorures	Cl-	18,17 mg/l	\leq 250 mg/l	100	18,17 mg/l	0	--
	Titre Alcalimétrique Complet (Alcaliné)	TAC	74,08 mg/l	--	--	--	--	--
	Durété Totale	DHT	25,78 mg/l	--	--	--	--	--
	Sulfates	SO42-	10,68 mg/l	250 mg/l	100	10,68 mg/l	0	--
Bicarbonates	HCO3-	90,71	--	--	--	--	--	
Paramètres microbiologiques	Coliformes Thermotolérants	CTH	1219,77 UFC/250ml	0 UFC/250ml	0	--	100	1219,77 UFC/250ml
	Escherichia coli	Ecol	947,85 UFC/250ml	0 UFC/250ml	0	--	100	947,85 UFC/250ml
	Entérocoques faecalis	Efae	936,07 UFC/250ml	0 UFC/250ml	0	--	100	936,07 UFC/250ml

En effet, tous les puits de M'Pody présentent une pollution microbiologique. On distingue dans ces eaux, dans un premier temps, une présence élevée de plusieurs bactéries. La concentration en Coliformes Thermotolérants est de 1219,77 UFC/250ml en moyenne. La concentration moyenne est de 947,87 UFC/250ml pour les bactéries Escherichia coli et de 936,07 UFC/250ml pour les Entérocoques faecalis. Dans un second temps, ces résultats indiquent que 95,83% des puits présentent une eau aux propriétés optique inacceptable (en moyenne 23,267 NTU) et

RESULTATS ET DISCUSSION

72,22% présentent une conductivité électrique hors normes (en moyenne 98,2158 $\mu\text{S}/\text{cm}$). De plus, les eaux de ces puits sont globalement acides ; leur pH tourne autour de 5,2 avec une température de 27,79° et des concentrations en nitrates, nitrites et chlorures plutôt faibles (respectivement 14,09 mg/l, 0,080 mg/l et 18,17 mg/l). Enfin, 19,44% et 2,78% des puits du village de M’Pody ont des concentrations hors normes, respectivement, en Ammonium (0,958 mg/l) et en Phosphates (0,6475 mg/l).

3.1.2 Résultats de l’ACP : analyse des puits

L’analyse en composantes principales réalisée sur la moyenne des paramètres physiques, chimiques et microbiologiques des échantillons d’eau de chaque puits, nous a conduit aux résultats ci-dessous.

3.1.2.1 Distribution de l’inertie

La figure 13 présente la décomposition de l’inertie totale en ACP. Les valeurs propres (eigenvalues en anglais) mesurent l’inertie (la quantité de variance expliquée) de chaque composante principale (dimension). L’inertie de ces axes suggère le nombre judicieux de composantes principales à étudier. Ainsi, les premières composantes principales correspondent aux directions portant la quantité maximale de variation contenue dans le jeu de données. Dans notre cas, nous avons seize (16) dimensions ou axes principaux (Annexe 2). L’inertie des axes factoriels indique d’une part si les variables sont structurées et suggère d’autre part le nombre judicieux de composantes principales à étudier. Les 2 premiers axes principaux (Dim) de l’analyse expriment **48,37%** de l’inertie totale de nos données ; cela signifie que 48,37% de la variabilité totale du nuage des individus (ou des variables) est représentée dans ce plan. C’est un pourcentage relativement moyen, et le premier plan représente donc seulement une part de la variabilité contenue dans l’ensemble du jeu de données actif. En observant le graphe de décomposition de l’inertie, la cassure intervient au niveau des 4^e et 5^e composantes principales. Du fait de ces observations, il serait alors probablement nécessaire de considérer les dimensions inférieurs ou égales à la cinquième dans l’analyse.

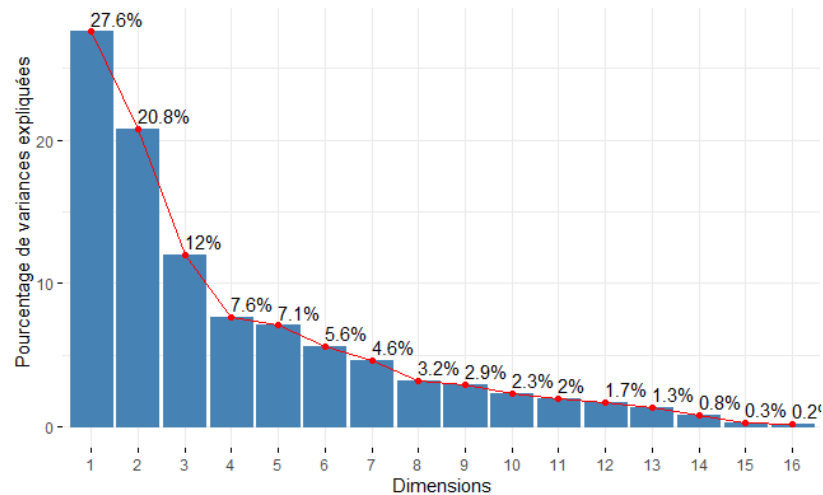


Figure 13: Décomposition de l'inertie totale (ACP)

En effet, l'inertie cumulée des cinq (05) premières composantes principales est de **75,07%** de l'inertie totale des données. Cependant, uniquement ces cinq premiers axes seront considérés dans notre analyse car elles renferment individuellement un grand pourcentage de l'information contenue dans notre tableau de données. Cette observation suggère que seuls ces axes sont porteurs d'une véritable information.

3.1.2.2 Description des plans étudiés

Sur l'Annexe 3, nous avons la projection des 72 individus (72 puits) et 16 variables (paramètres) physico-chimiques et microbiologique observés sur les cinq plans étudiés. Le nuage de points est observé sur ces cinq dimensions (axes principaux). Les puits sont positionnés en fonction de leurs coordonnées respectives. Leurs qualités de représentation et contributions aux plans étudiés servent pour une meilleure analyse. Un extrait de ces derniers peut être observé dans l'Annexe 4.

L'ACP nous a permis de faire ressortir différents profils (groupes) de puits en fonction de leur ressemblance. Ainsi comme illustré sur la figure 14, nous avons quatre types de puits dans le village M'Pody. Le tableau VI suivant résume les informations sur les différents profils de puits.

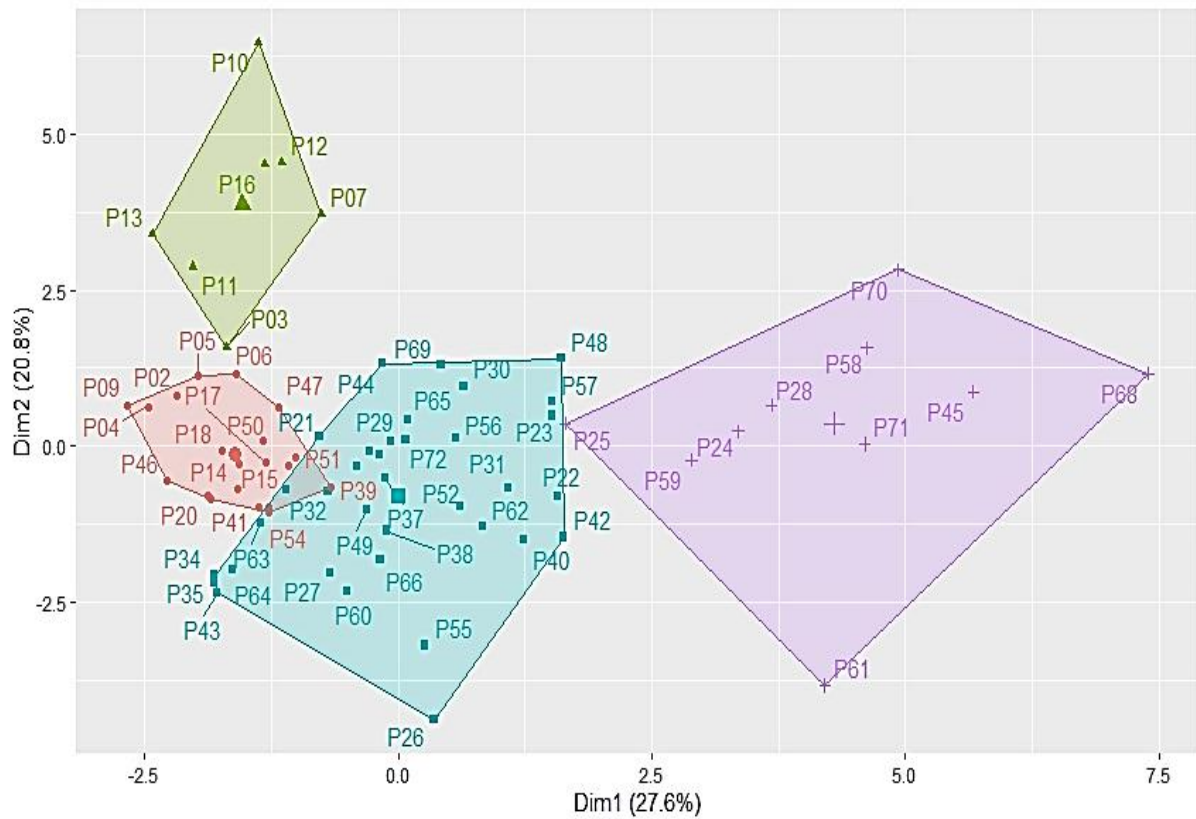


Figure 14: Graphes des puits avec les différents 04 profils sur le plan (1,2) de l' ACP

Tableau VI : Classement des puits selon leurs profils

PROFILS	PUITS
PROFIL 1 (Violet)	P24 ; P25 ; P28 ; P45 ; P58 ; P59 ; P61 ; P68 ; P70 ; P71.
PROFIL 2 (Vert)	P03 ; P07 ; P10 ; P11 ; P12 ; P13 ; P16.
PROFIL 3 (Rouge)	P01 ; P02 ; P04 ; P05 ; P06 ; P09 ; P14 ; P15 ; P17 ; P18 ; P19 ; P20 ; P39 ; P41 ; P46 ; P47 ; P50 ; P51 ; P53 ; P54.
PROFIL 4 (Bleu)	P08 ; P21 ; P22 ; P23 ; P26 ; P27 ; P29 ; P30 ; P31 ; P32 ; P33 ; P34 ; P35 ; P36 ; P37 ; P38 ; P40 ; P42 ; P43 ; P44 ; P48 ; P49 ; P52 ; P55 ; P56 ; P57 ; P60 ; P62 ; P63 ; P64 ; P65 ; P66 ; P67 ; P69 ; P72.

Ainsi, nous avons classé les puits selon différents profils. Les résultats de l'AFC permettent de connaître les caractéristiques de ses profils.

3.1.3 Résultat obtenus avec l'AFC

L'analyse factorielle des correspondances réalisée sur notre jeu de données nous a conduit aux résultats ci-dessous.

RESULTATS ET DISCUSSION

3.1.3.1 Distribution de l'inertie

L'AFC, nous donne quinze (15) dimensions ou axes principaux (voir Annexe 5). La décomposition de l'inertie en AFC (Figure 15) montre que les deux (02) premiers axes de l'analyse expriment **79,79%** de l'inertie totale des données ; cela signifie que 79,79% de la variabilité totale du nuage des lignes (ou des colonnes) est représentée dans ce plan. C'est un pourcentage assez important, et le premier plan représente donc convenablement la variabilité contenue dans le jeu de données. Ainsi, ce plan à lui seul résume la quasi-totalité de l'information. En conséquence, la description de l'analyse sera restreinte à ces seuls axes.

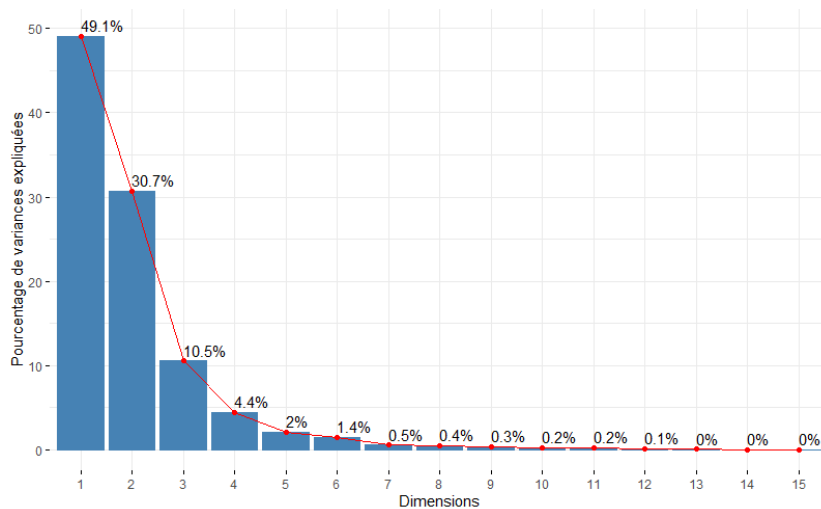


Figure 15: Décomposition de l'inertie (AFC)

3.1.3.2 Description des plans étudiés

L'AFC a donné le graphe superposé des puits et paramètres suivant sur le plan principal (1,2) (Figure 16).

La dimension 1 montre que les puits P64, P34, P35 et P43 (caractérisés par une coordonnée positive sur l'axe) partagent :

- une faible fréquence des paramètres pH, HCO_3^- et TAC (du plus rare au plus commun).

Les puits P60, P55, P26 et P54 (caractérisés par une coordonnée positive sur l'axe) partagent :

- une forte fréquence des paramètres PO_4^{3-} , NO_3^- , Cond et Cl^- (du plus commun au plus rare).
- une faible fréquence des paramètres Efae et CTH (du plus rare au plus commun).

Les puits P10, P13, P12 et P16 (caractérisés par une coordonnée négative sur l'axe) partagent :

- une forte fréquence des paramètres Ecol et CTH (du plus commun au plus rare).

Les puits P70, P07, P03 et P11 (caractérisés par une coordonnée négative sur l'axe) partagent :

- une forte fréquence des paramètres pH, Turb, Efae, DHT, SO_4^{2-} , TAC et HCO_3^- (du plus commun au plus rare).

RESULTATS ET DISCUSSION

- une faible fréquence du paramètre NO_3^- .

La dimension 2 montre que les puits P70, P07, P03 et P11 (caractérisés par une coordonnée positive sur l'axe) partage :

- une forte fréquence des paramètres pH, Turb, Efae, DHT, SO_4^{2-} , TAC et HCO_3^- (du plus commun au plus rare).

- une faible fréquence du paramètre NO_3^- .

Les puits P10, P13, P12 et P16 (caractérisés par une coordonnée négative sur l'axe) partage :

- une forte fréquence des paramètres Ecol et CTH (du plus commun au plus rare).

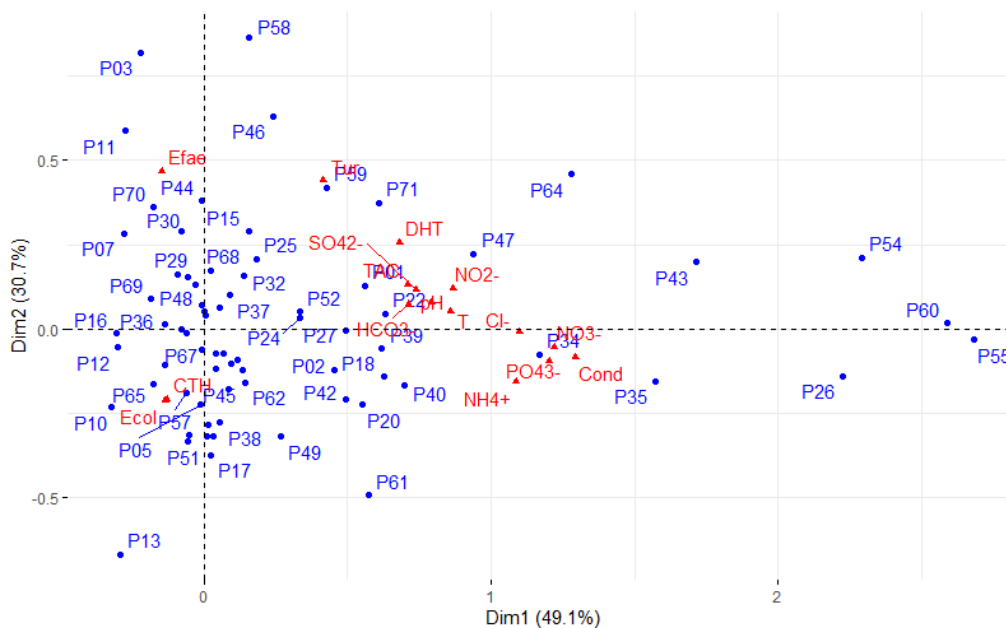


Figure 16: Graphe superposé en AFC sur les dimensions étudiées

Ainsi, ces résultats nous permettent de déduire les caractéristiques des profils de puits.

Le profil 1 regroupe les puits dont les eaux sont douces, très troubles, peu acides ($\text{pH} > 5,2$ la moyenne) et fortement minéralisées.

Le profil 2 concerne les puits ayant des eaux peu minéralisées, de pH autour de la moyenne. Ces puits contiennent une forte concentration de bactéries et très peu de chlorure donc sont fortement atteints par la pollution microbiologique.

Les eaux très douces ($\text{DHT} < 80\text{mg/L}$), de pH moyen, peu minéralisées, de TAC et chlorure faibles ainsi que faiblement atteint par la pollution microbiologique font partie du profil 3.

Le profil 4 est celui des eaux les plus chaudes, très acides, fortement minéralisées et contenant les plus grandes concentrations de chlorures et nitrates.

RESULTATS ET DISCUSSION

3.1.4 Résultats de l'ACP : paramètres physico-chimiques et microbiologiques

La matrice de corrélation donne une idée des associations entre les différents paramètres (Tableau VII). Une représentation graphique de cette matrice est en Annexe 6.

Tableau VII : Matrice de corrélation entre les paramètres étudiés

PARAMETRES	Tur	Cond	pH	T	NO ₃ ⁻	NO ₂ ⁻	NH ₄ ⁺	PO ₄ ³⁻	Cl ⁻	TAC	DHT	SO ₄ ²⁻	HCO ₃ ⁻	CTH	Ecol	Efae
Tur	1															
Cond	-0,167	1														
pH	0,230	-0,039	1													
T	0,000	0,434	-0,159	1												
NO ₃ ⁻	-0,099	0,687	-0,217	0,467	1											
NO ₂ ⁻	0,307	0,275	0,347	0,148	0,227	1										
NH ₄ ⁺	0,052	0,477	0,027	0,226	0,284	0,469	1									
PO ₄ ³⁻	-0,016	0,154	-0,005	0,123	0,163	0,296	0,197	1								
Cl ⁻	-0,109	0,525	0,175	0,422	0,410	0,080	0,153	0,044	1							
TAC	0,515	0,249	0,294	0,197	0,031	0,459	0,215	0,225	0,014	1						
DHT	0,632	0,270	0,275	0,204	0,115	0,451	0,132	0,208	0,023	0,891	1					
SO ₄ ²⁻	0,312	0,372	0,412	0,166	0,098	0,352	0,210	-0,039	0,234	0,672	0,612	1				
HCO ₃ ⁻	0,360	0,234	0,280	0,149	0,059	0,379	0,076	0,207	0,062	0,875	0,719	0,617	1			
CTH	0,059	-0,099	-0,017	-0,339	-0,069	0,011	-0,065	-0,035	-0,211	0,099	0,081	-0,030	0,120	1		
Ecol	0,049	-0,084	0,028	-0,326	-0,066	0,007	-0,060	0,017	-0,205	0,113	0,119	-0,004	0,121	0,960	1	
Efae	0,268	-0,141	0,102	-0,289	-0,148	0,092	-0,126	-0,099	-0,219	0,204	0,264	0,073	0,170	0,665	0,655	1

En effet, certains paramètres sont corrélés négativement (de -0,219 à -0,004) et d'autres positivement (de 0 à 0,960).

L'ACP se sert aussi des relations entre les paramètres à partir du calcul des corrélations existantes entre elle. Ces relations sont visibles dans les cercles de corrélations (Annexe 7). Ce cercle utilise la corrélation entre les paramètres et les axes principaux (dimensions) comme coordonnées des paramètres dans les différents plans principaux. Le tableau VIII suivant présente ces corrélations $r(C_j, X_i)$ sur les cinq dimensions étudiées ainsi que la contribution en pourcentage.

Ainsi, de l'analyse il apparaît clairement que la :

- Dim1 est une variable synthétique qui résume les informations des paramètres physico-chimiques suivants Turb, NO₂⁻, TAC, DHT, SO₄²⁻, HCO₃⁻ ;
- Dim2 est une variable synthétique qui résume bien les informations des paramètres physico-chimiques Cond, T, Cl⁻ et des paramètres microbiologiques CTH, Eol, Efae ;
- Dim3 est une variable synthétique qui résume bien les informations des paramètres physico-chimiques NO₃⁻ et des paramètres microbiologiques CTH, Eol ;
- Dim4 est une variable synthétique qui résume bien les informations du paramètre physico-chimique PO₄³⁻ ;
- Dim5 est une variable synthétique qui résume bien les informations du paramètre physico-chimiques pH.

Tableau VIII : Corrélations (Corr) et contributions (Contrib) des variables sur les différentes dimensions

		DIMENSIONS									
		Dim.1		Dim.2		Dim.3		Dim.4		Dim.5	
		Corr	Contrib	Corr	Contrib	Corr	Contrib	Corr	Contrib	Corr	Contrib
PARAMETRES	Tur	0,519	6,102	0,338	3,427	-0,352	6,445	-0,091	0,681	-0,179	2,829
	Cond	0,489	5,411	-0,557	9,328	0,494	12,688	0,170	2,362	0,044	0,170
	pH	0,390	3,447	0,188	1,065	-0,356	6,580	0,262	5,618	0,658	38,234
	T	0,326	2,402	-0,627	11,828	0,106	0,581	0,061	0,303	-0,365	11,800
	NO3-	0,285	1,838	-0,549	9,051	0,565	16,618	0,043	0,150	-0,159	2,223
	NO2-	0,652	9,642	-0,052	0,080	0,046	0,109	-0,361	10,684	0,403	14,352
	NH4+	0,387	3,402	-0,336	3,394	0,280	4,080	-0,355	10,335	0,395	13,776
	PO43-	0,275	1,711	-0,141	0,597	0,150	1,169	-0,678	37,671	0,020	0,035
	Cl-	0,246	1,374	-0,557	9,321	0,201	2,100	0,486	19,318	0,187	3,095
	TAC	0,902	18,422	0,194	1,132	-0,161	1,343	-0,048	0,189	-0,194	3,328
	DHT	0,875	17,363	0,196	1,158	-0,138	0,991	-0,032	0,081	-0,251	5,583
	SO42-	0,761	13,126	-0,002	0,000	-0,162	1,364	0,359	10,566	0,091	0,729
	HCO3-	0,809	14,826	0,181	0,985	-0,123	0,791	0,047	0,178	-0,201	3,573
	CTH	0,058	0,075	0,737	16,307	0,617	19,824	0,068	0,379	0,025	0,053
	Ecol	0,082	0,151	0,730	16,008	0,613	19,578	0,059	0,288	0,042	0,156
Efae	0,177	0,709	0,737	16,318	0,332	5,739	0,121	1,196	-0,027	0,065	

Notons que ces résultats pourraient être utiles pour la construction de modèle et pour le suivi de ses variables. En effet, la pollution microbiologique peut se résumer dans l'étude unique du paramètre *Escherichia coli* (Ecol). L'analyse de TAC pourrait nous renseigner sur la dureté (DHT), la concentration en bicarbonate (HCO_3^-), en sulfate (SO_4^{2-}). L'analyse de la conductivité (Cond) uniquement pourrait nous renseigner sur la température (T), la quantité de chlorure (Cl^-) et de nitrate (NO_3^-).

3.1.5 Résultat obtenus avec l'ANOVA

L'ANOVA a été réalisée afin de savoir si les puits ont un effet sur les paramètres. Trois paramètres ont été sélectionnés. En effet, nous mettons un lien entre l'ACP et l'ANOVA en ce sens que les paramètres corrélés auront des réponses similaires au niveau de l'ANOVA. En effet, les résultats obtenus avec :

- le paramètre « conductivité » seront semblables à ceux qu'on obtiendrait avec la température, les nitrates et les chlorures.
- le paramètre « titre alcalimétrique complet » seront semblables à ceux qu'on obtiendrait avec la dureté totale, le sulfate et le bicarbonate.
- le paramètre « *Escherichia coli* » seront semblables à ceux qu'on obtiendrait avec les coliformes thermotolérants et *Enterocoques faecalis*.

RESULTATS ET DISCUSSION

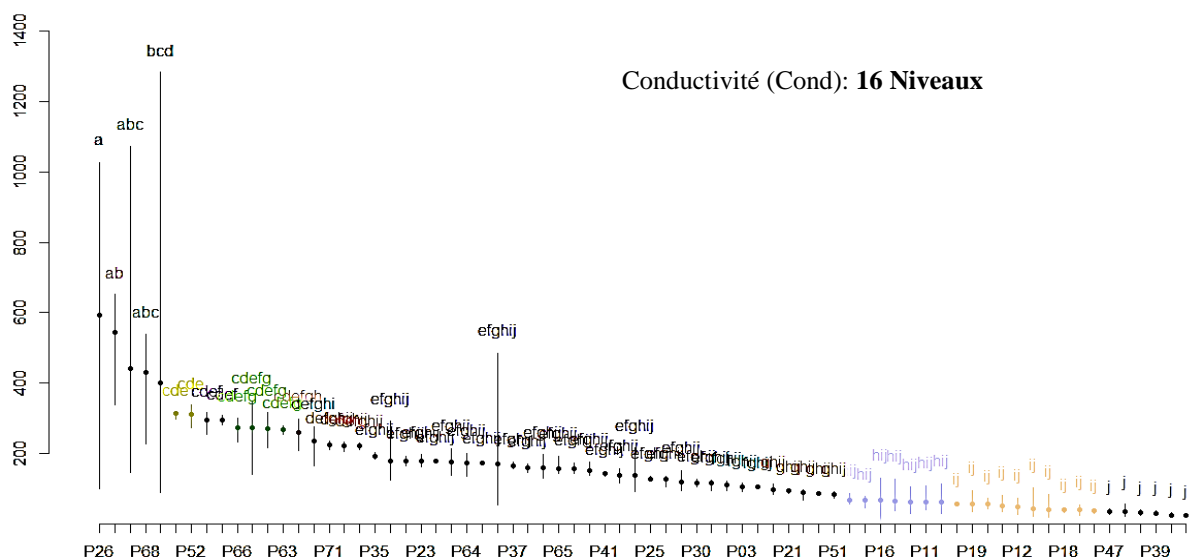
Les tests des conditions de validité l'ANOVA ont échoué. En effet, les tests de normalité (Annexe 8) de Shapiro-Wilk (test numérique) et l'histogramme (test graphique) réalisés sur les variables dépendantes y (paramètre) montrent que ces paramètres ne suivent pas la loi normale. Dans, ces conditions nous réalisons l'ANOVA non paramétrique (**kruskal.test**). Ces tests ont donné des p -values significatives (Figure 18), cela nous indique que les puits ont bien un effet sur ces paramètres. Ce qui signifie aussi que les facteurs influençant les paramètres des puits sont différents.

- Comparaisons multiples

Les tests de comparaisons multiples réalisés respectivement sur les paramètres Cond, TAC et Ecol ont donné les résultats suivants (Figure 17).

Ces résultats montrent les différences significatives entre les puits par paramètres étudiés. Ces résultats montrent aussi une classification des puits par niveau de concentration (du puits le plus concentré au puits le moins concentré) (voir Annexe 8). Ainsi, nous avons seize (16) niveaux significatifs avec la conductivité, 19 niveaux significatifs avec le titre alcalimétrique complet et 8 niveaux significatifs avec de l'Escherichia coli. En effet, ces résultats suggèrent que si trois puits appartiennent respectivement à différents niveaux significatifs à savoir a , ab et c par exemple, alors :

- le puits du niveau a est proche de celui du niveau ab mais différent du puits du niveau b ;
- le puits du niveau b est proche de celui du niveau ab mais différent du puits du niveau a .



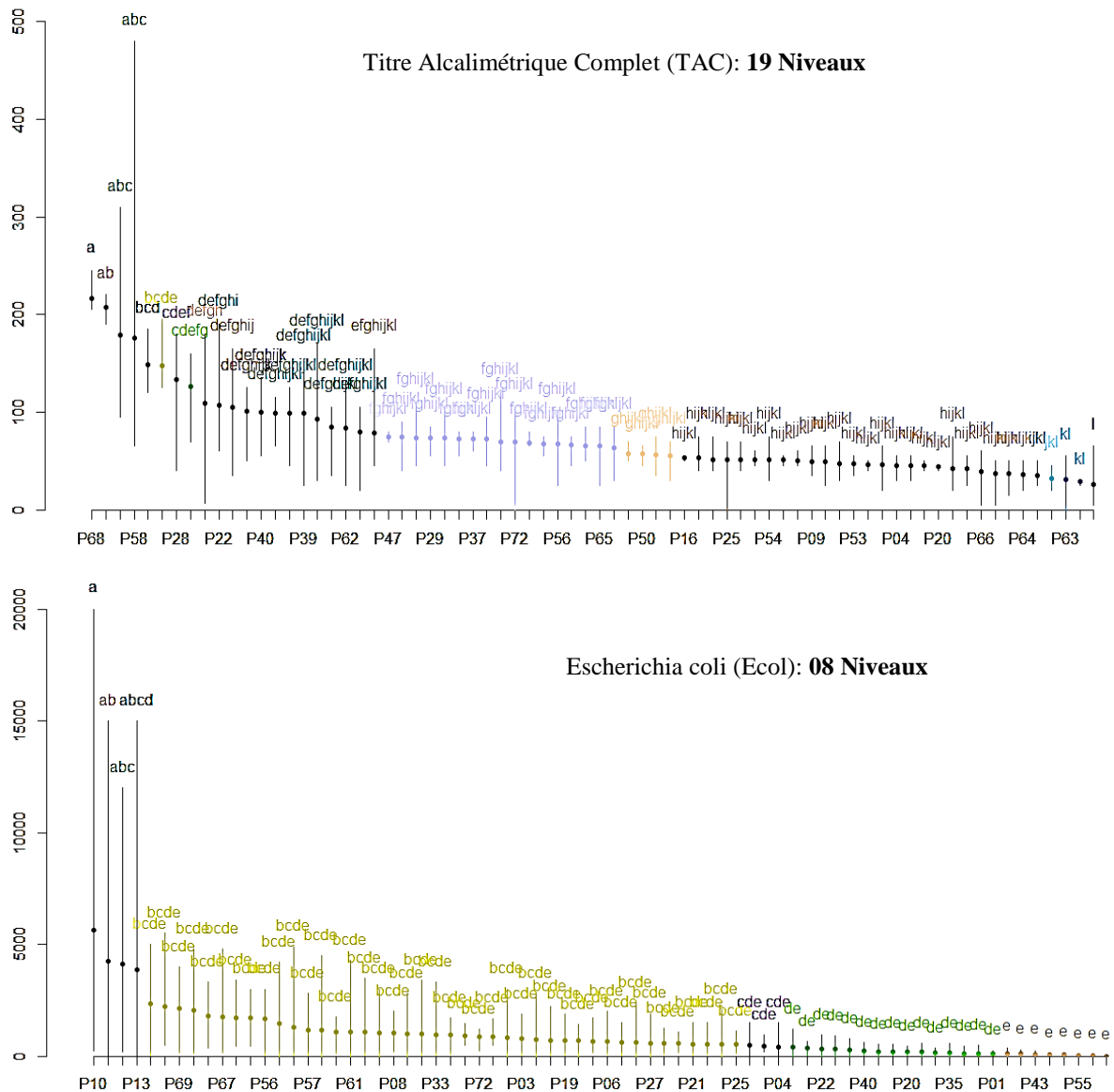


Figure 17: Résultat des tests de Duncan

3.2 Discussion

La statistique descriptive révèle que les eaux des puits du village M’Pody sont hors normes en ce qui concerne les paramètres de potentiel d’hydrogène, turbidité, coliformes thermotolérants, escherichia coli et entérocoques faecalis. Ainsi, ces eaux sont globalement acides, troubles et tous affectées par une pollution microbiologique. Ces résultats sont conformes à ceux trouvés par Gbagbo *et al.* (2020).

L’application de l’ACP et l’AFC a révélé quatre (04) types (profils) de puits ainsi que leurs caractéristiques. Tout comme l’étude effectuée par Lagnika *et al.* (2014), l’analyse faite au niveau des paramètres physico-chimiques et microbiologiques a permis de justifier la possibilité

RESULTATS ET DISCUSSION

de réduire les dimensions de l'étude. En effet, dans un groupe de variables fortement corrélées entre elles, l'on peut conserver la variable, la mieux corrélées avec les dimensions (variables synthétiques). En clair, la possibilité d'obtenir la quasi-totalité des informations des seize paramètres étudiés en réalisant une analyse sur huit paramètres uniquement. Ainsi, un suivi des paramètres des eaux des puits de M'Pody peut s'effectuer avec uniquement les paramètres suivants : Turbidité (Turb), Conductivité (Cond), Potentiel d'hydrogène (pH), Nitrite (NO_2^-), Ammonium (NH_4^+), Phosphate (PO_4^{3-}), Titre alcalimétrique complet (TAC) et Escherichia coli (Ecol).

L'application de l'ANOVA a montré que même étant sur la même zone d'étude, les paramètres des eaux des puits évoluent différemment. Il y a effectivement un impact significatif des puits sur la conductivité, la turbidité et la concentration de bactéries dans les puits ($p\text{-value} < 0.05$). Les bactéries (paramètres microbiologiques) sont d'origine fécale. Ainsi, un impact significatif des puits sur la présence de bactéries pourrait impliquer la présence de fosses septiques à proximité qui contaminent les eaux ou un mauvais entretien de certains puits par les populations. Les comparaisons multiples (test de Duncan) ont permis de classer les puits par niveau de concentrations significatives aux vues de trois paramètres que sont la conductivité, le titre alcalimétrique complet et l'escherichia coli. L'ACP a montré que ces paramètres sont corrélés à d'autres. Ainsi, une différence significative entre les puits pour l'un de ces paramètres implique celle des paramètres corrélés. En effet, une différence significative entre les puits par rapport au paramètre conductivité implique la même différence significative par rapport aux paramètres température, nitrates et chlorures. Il en est de même pour les paramètres « titre alcalimétrique complet » et « escherichia coli » et les paramètres qui leur sont corrélés respectivement. De plus, les classifications des puits par niveaux de concentration effectués peuvent s'avérer être utiles pour la prise de décision. Par exemple, les eaux des puits les plus concentrés significativement en bactéries (paramètre Ecol) identifiées par l'ANOVA peuvent être interdites à la consommation.

CONCLUSION

CONCLUSION

A la fin de cette étude, nous pouvons retenir que les techniques d'analyses multivariées nous ont effectivement permis d'étudier la qualité des eaux des puits du village M'Pody. L'analyse statistique descriptive a montré que les eaux des puits de M'Pody constituent un risque sanitaire pour les populations dont la survie en dépend. Sur treize paramètres physico-chimiques étudiées, la turbidité (Turb), la conductivité (Cond), le potentiel d'hydrogène (pH) et la température (T) se sont avérés être hors normes pour la majorité des eaux des puits étudiés. Il en est de même pour les trois paramètres microbiologiques (*Escherichia coli*, Coliformes thermotolérants et Entérocoques Faecalis). En effet, la totalité des 72 puits est concernée par cette pollution bactériologique. L'analyse exploratoire a permis dans un premier temps d'identifier quatre (04) profils de puits avec l'ACP. Dans un second temps, l'application de l'AFC a révélé les caractéristiques de ces profils. Les résultats de l'ACP ont permis de réduire la taille des puits (individus) et paramètres (variables) sans grande perte d'information. D'une part, les informations données par les trois paramètres microbiologiques se résument dans l'analyse unique des bactéries *Escherichia coli* (Ecol). D'autre part, la dureté, la teneur en sulfates (SO_4^{2-}) et bicarbonates (HCO_3^-) se déduisent de l'analyse unique du Titre Alcalimétrique Complet (TAC). Ces résultats peuvent servir à la construction de modèle. Enfin, les effets des puits sur les paramètres physico-chimiques et microbiologiques ont été mis en évidence avec l'application de l'ANOVA. Effets qui prouvent une mauvaise gestion des puits par la population. D'où la nécessité de prendre des mesures en matière de gestion de ces eaux. Cette étude a montré également l'importance et l'utilité ces techniques d'analyses multivariées dans l'étude et le suivi de la qualité des eaux de puits et la prévention contre toutes sortes de pollutions. Cette étude pourrait être approfondie par d'autres techniques d'analyse d'une part afin d'obtenir plus d'informations sur les bases de données et d'autre part par l'utilisation d'autres techniques d'analyse. L'on pourrait, par exemple, construire un modèle de régression linéaire du paramètre « *Escherichia coli* » en fonction des paramètres « turbidité », « conductivité », « potentiel d'hydrogène », « nitrites », « ammonium », « phosphates », « Titre Alcalimétrique Complet » et « coliformes thermotolérants ». Ou encore l'on peut construire un modèle non linéaire en utilisant les réseaux de neurones sur ACP et AFC.

REFERENCES

REFERENCES

- Ahoussi E.K., Soro N., Koffi B.Y., Soro G. & Biemi J. (2010). Origine de la minéralisation des eaux des aquifères discontinus sous couvert forestier de la zone Sud de la Côte d'Ivoire : cas de la région d'Abidjan-Agboville. *International Journal of Biological and Chemical Sciences*, 4(3) : 782-797.
- Azaïs J.M. & Bardet J. (2005). Le modèle linéaire par l'exemple : Régression, Analyse de la Variance et Plans d'Expériences. Illustrations numériques avec les logiciels R, SAS et Splus. Dunod, Paris (France), 366 p.
- Blé L.O., Soro T.D., Dje D.K., Degny G.S. & Biemi J. (2015). Eaux conditionnées en sachets : Quels risques d'exposition des populations du district d'Abidjan ?. *Larhyss Journal*, 24 : 85-107.
- Celerier J.L., Faby J.A., Loiseau G. & Juery C. (2002). La dégradation de la qualité de l'eau potable dans les réseaux. Rapport d'étude, Fonds National pour le Développement des Adductions d'Eau, 98 p.
- Chébékoué S.F. (2008). Évaluation du risque cancérigène associé à la contamination de l'eau potable de puits municipaux par les nitrates/nitrites dans certaines régions rurales du Québec. Mémoire de maîtrise, Université de Montréal (Montréal, Canada), 157 p.
- De Villers J., Squilbin M. & Yourassowsky C. (2005). Qualité physico-chimique et chimique des eaux de surface : cadre général. Rapport d'étude, Institut Bruxellois pour la Gestion de l'Environnement / Observatoire des Données de l'Environnement, 16 p.
- Eblin S.C., Sombo A.P., Soro G., Aka N., Kambiré O. & Soro N. (2014). Hydrochimie des eaux de surface de la région d'Adiaké (sud-est côtier de la Côte d'Ivoire). *Journal of Applied Biosciences*, 75 : 6259-6271.
- Escofier B. & Pagès J. (2008). Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation. Dunod, Paris (France), 328 p.
- Gbagbo T.A.C., Kpaibe S.A., Gbokpeya K.M., Able N., Seki T.O., Bakayoko A., Meless D.F.R. & Amin N. (2020). Caractérisation physicochimique et bactériologique des eaux de consommation de la nappe phréatique du village M'pody (Côte d'Ivoire). *African Journals Online*, 22(3) : 775-792.
- Ignacio R.M.C., Joo K.B. & Lee K.J. (2012). Clinical Effect and Mechanism of Alkaline Reduced Water. *Journal of Food and Drug Analysis*, 20 : 394-397.
- Ihaka R. & Gentleman R. (1996). A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3) : 299-314.
- Lafaye M.P., Drouihet R. & Liquet B. (2011). Le logiciel R Maîtriser le langage. Effectuer des analyses statistiques. Springer-Verlag, Paris (France), 527 p.

REFERENCES

- Lagnika M., Ibikounle M., Montcho J.C., Wotto V.D. & Sakiti N.G. (2014). Caractéristiques physico-chimiques de l'eau des puits dans la commune de Pobè (Bénin, Afrique de l'ouest). *Journal of Applied Biosciences*, 79 : 6887-6897.
- Lebart L., Piron M. & Morineau A. (2006). Statistique exploratoire multidimensionnelle. Visualisation et inférence en fouille de données - 4ème édition. Dunod, Paris (France), 480 p.
- OMS (2004). Directives de qualité pour l'eau de boisson : troisième édition. Services de production de documents de l'OMS, Genève, 110 p.
- OMS (2009). Potassium in drinking-water. Background document for preparation of WHO Guidelines for drinking-water quality. WHO Document Production Services, Geneva (Switzerland), 12 p.
- OMS (2017). Directives de qualité pour l'eau de boisson : 4ème édition intégrant le premier additif. Services de production de documents de l'OMS, Genève, 564 p.
- Qaouiyid A., Hmima H., Hourri K., Saad I., Saad E.H. & Mohamed F. (2016). Les teneurs métalliques et paramètres physico-chimiques de l'eau et du sédiment de Oued Beht, au niveau de Sidi Kacem et de Oued R'dom au niveau de Sidi Slimane. *European Scientific Journal*, 24(12) : 186-211.
- Pedoya C. (1990). La guerre de l'eau : Genèse, mouvement et échanges, pollution et pénuries. Frison Roche, Paris (France), 155 p.
- Rodier J., Legube B. & Merlet N. (2009). L'Analyse de l'Eau : Eaux Naturelles, Eaux résiduaires, Eau de Mer (9ème édition). Dunod, Paris (France), 1579 p.
- Ruxton G.D. & Beauchamp G. (2008). Some suggestions about appropriate use of the Kruskal-Wallis test. *Animal Behaviour*, 76 : 1083-1087.
- Scalon B.R., Reedy R.C., Stonestrom D.A., Prudic D.E. & Dennehy K.F. (2005). Impact of land use and land cover change on groundwater recharge and quality in southwestern US. *Global Change Biology*, 11 : 1577-1593.
- Umesh D., Bresee J.S., Glass R.I., (2003). The global burden of diarrhoeal disease in children. *Bulletin of the World Health Organisation*, 81(4): 236-7.
- Yapo O.B., Mambo V., Seka A., Ohou M.J.A., Konan F., Grouzilz V., Tidou A.S., Kouamé K.V. & Houenou P. (2010). Evaluation de la qualité des eaux de puits à usage domestique dans les quartiers défavorisés de quatre communes d'Abidjan (Côte

REFERENCES

- d'Ivoire) : Koumassi, Marcory, Port-Bouët et Treichville. *International Journal of Biological and Chemical Sciences*, 4(2) : 289-307.
- Zuur A.F., Ieno E.N., Walker N.J., Saveliev A.A. & Smith G.M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. Springer, New York, 600 p.

ANNEXES

ANNEXES

Annexe 1 : Normes des paramètres de pollution physico-chimique et biochimiques de l'eau de boisson (OMS, 2009)

N°	TYPES DE PARAMETRES	ELEMENTS	VALEURS INDICATIVES
1	Paramètres physico-chimiques	Conductivité	180-1000 μ S/cm
2		Température	$\leq 25^{\circ}\text{C}$
3		Turbidité	< 5 NTU
4		Ammonium	$\leq 0,5$ mg/l
5		Chlore	≤ 250 mg/l*
6		Nitrates	≤ 50 mg/l
7		Nitrites	≤ 3 mg/l
8		pH	6,5 – 8,5
9		Phosphate	0,5 mg/l
10		Sulfates	250 mg/l
11	Paramètres microbiologiques	Coliformes	0 UFC**/250ml
12		Entérocoque	0 UFC/250ml
13		Escherichia coli	0 UFC/250ml

*Pas de valeur en vrai mais on note un goût à partir de cette valeur

**UFC : Unité Formatrice de Colonie

Annexe 2 : Valeurs propres et variances en ACP

Dimensions	Valeur propre	Pourcentage de variance	Pourcentage de variance cumulée
Dim.1	4,411713	27,57320686	27,57320686
Dim.2	3,327518	20,79698886	48,37019571
Dim.3	1,922343	12,01464417	60,38483988
Dim.4	1,220239	7,626493393	68,01133328
Dim.5	1,130856	7,067851936	75,07918521
Dim.6	0,893195	5,582465703	80,66165092
Dim.7	0,735107	4,594419726	85,25607064
Dim.8	0,509452	3,184072125	88,44014277
Dim.9	0,466899	2,918120295	91,35826306
Dim.10	0,370246	2,314035674	93,67229874
Dim.11	0,32012	2,000747911	95,67304665
Dim.12	0,276809	1,73005427	97,40310092
Dim.13	0,214158	1,338487426	98,74158834
Dim.14	0,127571	0,797321633	99,53890998
Dim.15	0,04452	0,278250802	99,81716078
Dim.16	0,029254	0,182839222	100

ANNEXES

Annexe 3 : Projection des individus et variables sur les plans étudiés

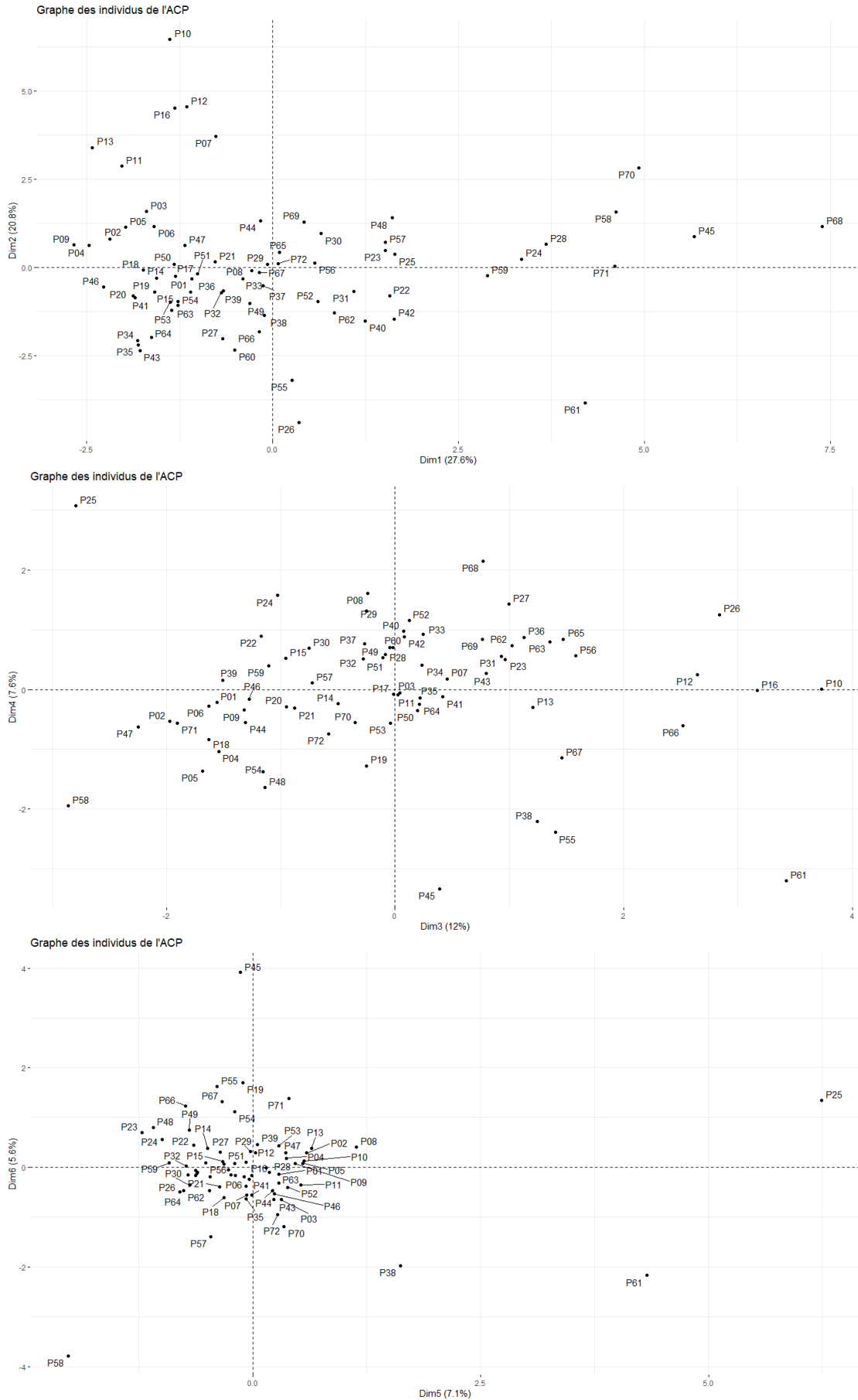


Figure : Graphes des individus en ACP sur les dimensions étudiées

ANNEXES

Annexe 4 : Extraits des qualités de représentation et contributions des puits sur les cinq premières dimensions

	Dim.1		Dim.2		Dim.3		Dim.4		Dim.5	
	Corrélation	Contribution	Corrélation	Contribution	Corrélation	Contribution	Corrélation	Contribution	Corrélation	Contribution
P01	0,2378298	0,3727804	0,0203931	0,0423795	0,4890164	1,7590841	0,0093261	0,0528505	0,0161983	0,0990505
P02	0,3631442	1,5055979	0,0490423	0,2695798	0,2956961	2,8135316	0,0212857	0,3190664	0,0268115	0,4336616
P03	0,2480491	0,9048333	0,2159822	1,0445658	0,0001657	0,0013874	0,0003263	0,004303	0,0084653	0,1204685
P04	0,5503608	1,9191694	0,0359426	0,1661735	0,2150566	1,7210555	0,0973964	1,2279229	0,0122869	0,1671505
P05	0,3183605	1,2314378	0,1053464	0,5402561	0,2307779	2,0486322	0,151676	2,1211564	0,0244586	0,3690831
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
P68	0,8341883	17,196903	0,020636	0,5640258	0,0089896	0,4253075	0,0705095	5,2552859	0,00771	0,620069
P69	0,0307613	0,0563975	0,2869572	0,6975238	0,1002148	0,4216608	0,1220567	0,8090567	0,0002329	0,0016658
P70	0,5522782	7,6451146	0,1815298	3,331662	0,0027446	0,087193	0,0070855	0,3546165	0,002663	0,1438115
P71	0,7232085	6,6566984	5,062E-05	0,0006177	0,1241622	2,6227762	0,0109996	0,3660453	0,0054112	0,1943084
P72	0,0006778	0,001703	0,0014427	0,0048063	0,0425367	0,2452973	0,0686811	0,6239539	0,0094695	0,0928282

Annexe 5 : Tableau des valeurs propres et variances en AFC

Axes Principaux (Dimensions)	Valeurs propres	Pourcentage de variance	Pourcentage de variance cumulé
Dim1	0,13931555	49,07701518	49,07701518
Dim2	0,08719737	30,71722187	79,79423705
Dim3	0,02988084	10,52619476	90,32043181
Dim4	0,01244657	4,384583811	94,70501562
Dim5	0,00578824	2,039037283	96,7440529
Dim6	0,00404987	1,42665867	98,17071157
Dim7	0,00149249	0,525762882	98,69647446
Dim8	0,00125301	0,441401756	99,13787621
Dim9	0,00092662	0,326422951	99,46429916
Dim10	0,00068742	0,242158606	99,70645777
Dim11	0,00054987	0,193703999	99,90016177
Dim12	0,00014555	0,051271989	99,95143376
Dim13	7,2037E-05	0,025376533	99,97681029
Dim14	4,8153E-05	0,016962947	99,99377324
Dim15	1,7676E-05	0,006226763	100

Annexe 6 : Représentation graphique du calcul de corrélations

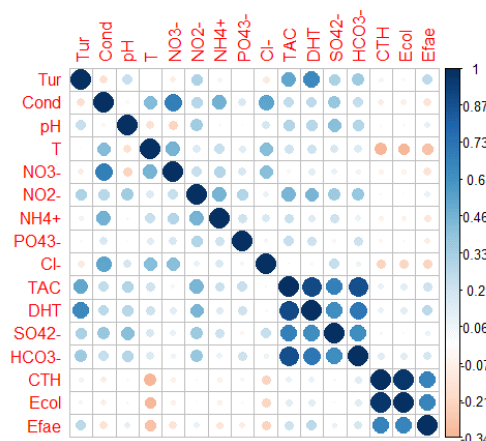
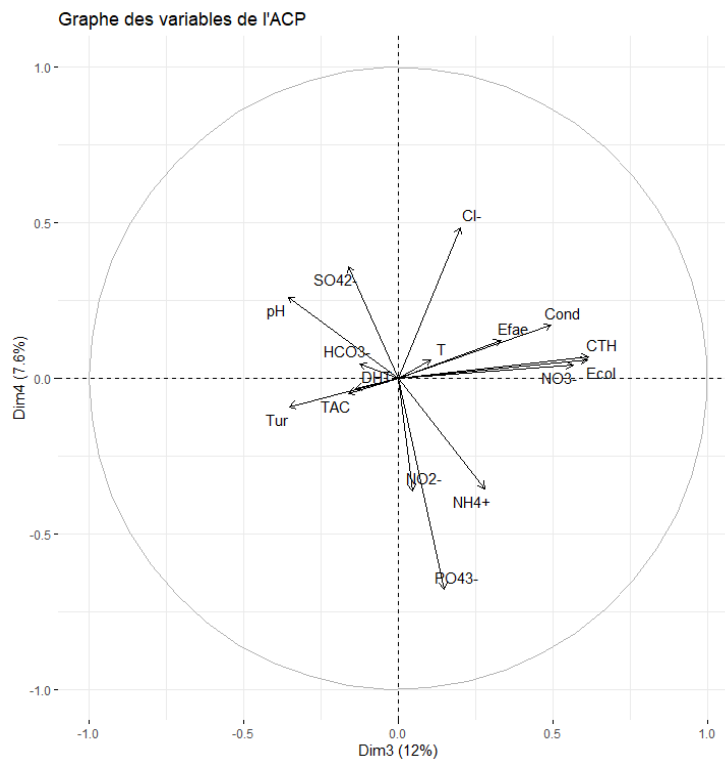
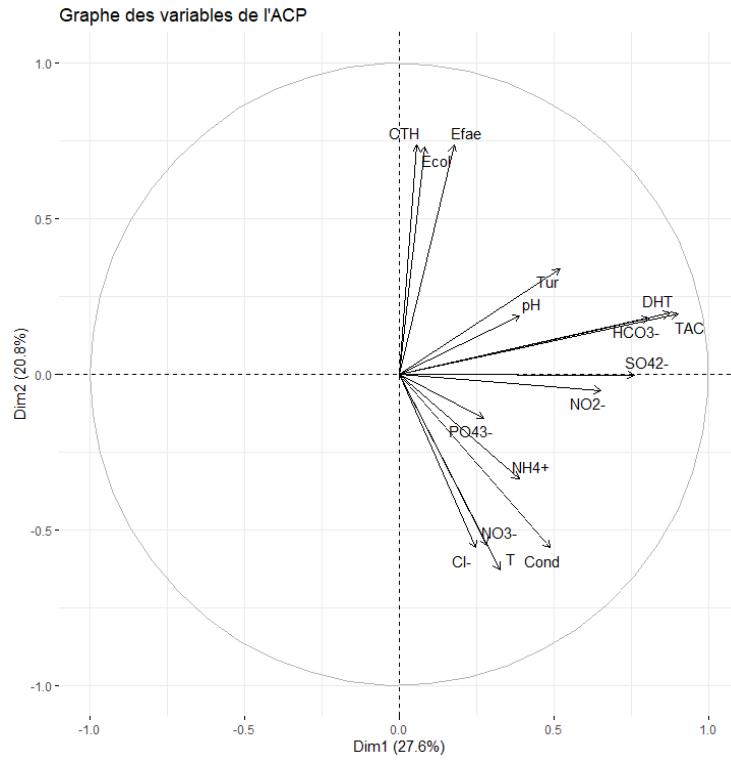


Figure : Représentation graphique des corrélations entre paramètres

Annexe 7 : Graphes des variables en ACP sur les dimensions étudiées



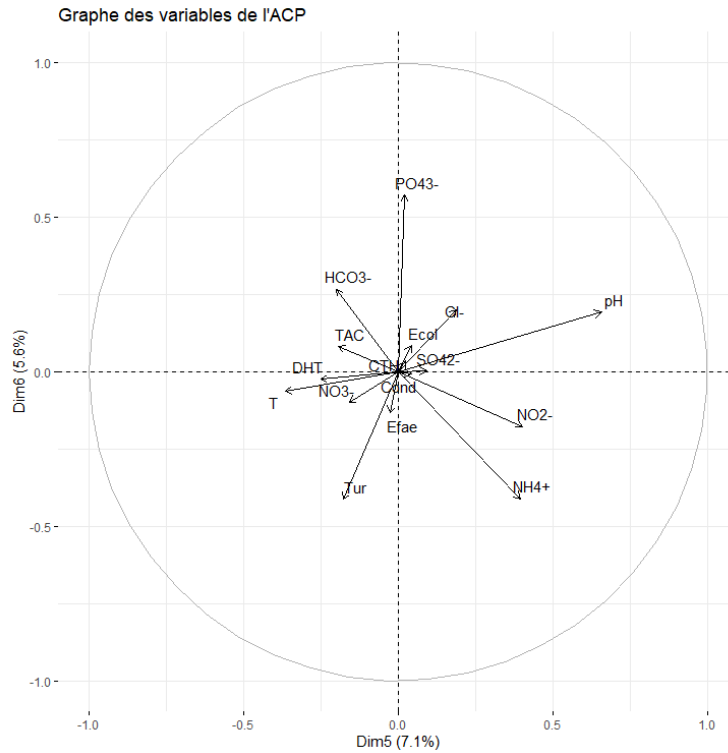
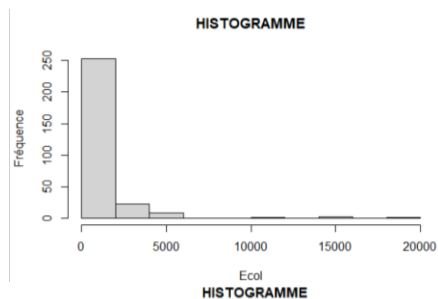
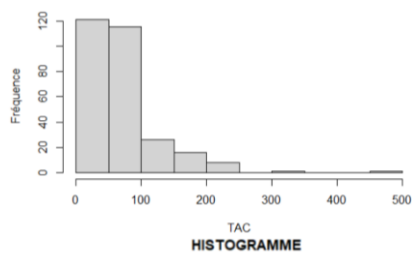


Figure : Cercle de corrélation en ACP sur les dimensions étudié

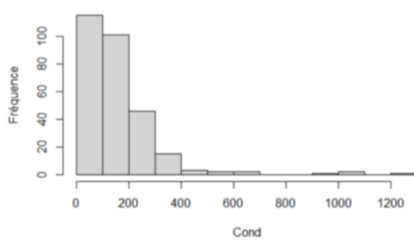
Annexe 8 : Résultat des tests de normalité



Shapiro-wilk normality test
 data: DATANO\$Ecol
 W = 0.44814, p-value < 2.2e-16



Shapiro-wilk normality test
 data: DATANO\$TAC
 W = 0.78174, p-value < 2.2e-16



Shapiro-wilk normality test
 data: DATANO\$Cond
 W = 0.68104, p-value < 2.2e-16

Annexe 9 : Résultat des tests de Kruskal-walis

Kruskal-wallis rank sum test
 data: DATANO\$Ecol by DATANO\$Puits
 Kruskal-wallis chi-squared = 103.52, df = 71, p-value = 0.007115

Kruskal-wallis rank sum test
 data: DATANO\$TAC by DATANO\$Puits
 Kruskal-wallis chi-squared = 149.66, df = 71, p-value = 1.504e-07

Kruskal-wallis rank sum test
 data: DATANO\$Cond by DATANO\$Puits
 Kruskal-wallis chi-squared = 257.22, df = 71, p-value < 2.2e-16

Annexe 10 : Classement des puits par niveaux de concentration

PUITS	Cond	NIVEAUX	PUITS	TAC	NIVEAUX	PUITS	Ecol	NIVEAUX	
P26	594	a	P68	216,25	a	P10	5650	a	
P61	545,25	ab	P45	207,5	ab	P16	4242,75	ab	
P27	440,95	abc	P71	178,75	abc	P12	4125	abc	
P68	430,75		P58	176,25		P13	3867,5	abcd	
P56	400,9	bcd	P24	148,75	bcd	P65	2329,5	bcde	
P31	313,75	cde	P70	147,5	bcde	P07	2205		
P52	311,75		P28	133,75	cdef	P69	2125		
P24	295,75	cdef	P31	126,25	cdefg	P68	2065		
P55	295,25		P59	109,25	defgh	P45	1817,5		
P66	273,75	cdefg	P22	107,5	defghi	P67	1750,25		
P45	272,9		P48	105	defghij	P23	1710		
P40	271,575		cdefgh	P42	101,25	defghijk	P70		1702,5
P63	267,75			P40	100		P56		1666
P62	259,35	P23		98,75	P28		1470		
P28	236,575	P30		98,75	P66		1286,5		
P71	224,825	defghij	P39	98,75	P57		1175		
P43	222,7		P61	92,5	P62		1148		
P60	221,2		efghij	P57	85		P36		1101,25
P35	193,65	P62		83,75	P61		1096,5		
P59	180,025	efghij		P44	80	P29	1095		
P08	179,275			P49	78,75	P50	1057,5		
P23	178,8		P47	75	P08	1025			
P34	178,725		efghijkl	P69	75	P30	1003,5		
P42	177,325	efghijkl		P01	73,75	P63	1000,25		
P64	173,075		P29	73,75	P33	957,25			

ANNEXES

P38	172,875		P52	73,75		P48	950,75	
P22	170,7		P33	72,5		P17	895	
P37	166,1		P37	72,5		P72	880	
P67	159,45		P51	72,5		P24	867,5	
P49	158,975		P07	70		P31	839	
P65	158,1		P72	70		P03	785	
P33	156,775		P21	68,75		P49	751,5	
P69	151,175		P12	67,5		P37	717,5	
P41	143,45		P56	67,5		P19	692,5	
P32	138,575		P55	66,25		P11	690,5	
P50	138		P10	65		P41	660	
P25	128,575		P65	65		P06	657,5	
P36	127,075		P32	63,75		P14	627	
P53	120,8		P15	57,5		P52	612,75	
P30	117,55		P50	57,5		P27	572,5	
P17	116,875		P06	56,25		P44	567,5	
P72	112,05		P67	55		P05	565	
P03	106,675	fg hij	P16	53,75	ghijkl	P21	532,75	
P29	104,925		P18	53,75		P38	532,5	
P70	96,95		P02	51,5		P32	532,25	
P21	94		P25	51,251		P25	527,5	
P44	90,4	ghij	P08	51,25		P53	488,5	cde
P57	87,175		P14	51,25		P42	435	
P51	84,275		P54	51,25		P04	406,25	
P54	67,475		P60	51,25		P59	385	
P48	67,35		P17	50		P09	352,5	
P16	67,025		P09	48,75		P22	329	
P14	64,55	hij	P13	48,75		P71	324,5	
P58	62,6		P27	47,5		P51	293,25	
P11	62,425		P53	47,5	hijkl	P40	252,5	
P01	62,35		P03	46,25		P39	210	
P46	58,7		P26	46,25		P02	200,75	de
P19	57,55		P04	45		P20	188	
P20	57,5		P11	45		P58	180	
P07	52,65		P19	45		P15	155	
P12	48,05		P20	43,75		P35	146,25	
P02	44,625	ij	P05	42,5		P34	130	
P09	42,0225		P36	42,5		P46	126,75	
P18	40,7		P66	38,75		P01	117,75	
P06	40,15		P38	37,5		P26	112,5	
P05	37,925		P41	37,5		P18	97,75	
P47	35,65		P64	36,25		P43	63,75	
P15	34,55		P34	35	ijkl	P47	51,5	
P10	32,5		P35	32,5	jkl	P64	45,75	e
P39	31,6	j	P63	31,25	kl	P55	12	
P04	25,8525		P43	28,75		P60	11	
P13	24,725		P46	26,25	l	P54	2	

RESUME

L'objet de ce travail est d'étudier la qualité des eaux de puits du village M'Pody en utilisant des méthodes d'analyses statistiques multivariées. Dans cette optique, seize (16) paramètres physico-chimiques et microbiologiques provenant de l'analyse des échantillons d'eaux de puits collectés au cours de l'année 2020 sont analysés sous R. L'analyse statistique descriptive est utilisée pour étudier la non potabilité de ces eaux. Ensuite, l'utilisation couplée de l'analyse en composantes principales (ACP) et l'analyse factorielle de correspondances (AFC) permet de faire ressortir les différents profils de puits de la zone d'étude. Enfin, pour prouver la contribution de ces puits sur les paramètres microbiologiques, l'analyse de la variance (ANOVA) est utilisée.

Mots clés : analyse en composantes principale, analyse factorielle de correspondances, modélisation, analyse des variances, eau de puits.

ABSTRACT

The purpose of this paper is to study the quality of well water in the village M'Pody using multivariate statistical analysis methods. To this end, sixteen (16) physico-chemical and microbiological parameters from the analysis of well water samples collected during the year 2020 are analyzed under R. Descriptive statistical analysis is used to study the non-potability of these waters. Then, the coupled use of principal component analysis (PCA) and factorial correspondence analysis (FCA) is used to highlight the different well profiles of the study area. Finally, to prove the contribution of these wells on the microbiological parameters, the analysis of variance (ANOVA) is used.

Key words: principal component analysis, factorial correspondence analysis, modeling, analysis of variance, well water.