

الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLICUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي و البحث العلمي
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
جامعة عمّار ثليجي بالأغواط
UNIVERSITE AMAR TELIDJI LAGHOUAT
كلية العلوم
FACULTE DES SCIENCES
قسم الإعلام الآلي
DEPARTEMENT D'INFORMATIQUE

Mémoire de MASTER

Domaine : Mathématiques et Informatique
Filière : Informatique
Option : Systèmes d'Information et de Décision

Par:

Mennad Oussama Taibi Mustapha

THEME

**Analyse descriptive sur les données climatiques :
cas d'étude la région de Laghouat**

Soutenu publiquement le 24-06-2019 Devant le jury composé de :

MrL.chellama
Mr M.H Maicha
Mr Y. Gellouma

Professeur
Professeur
M.C.(A)

Président
Examineur
Encadreur

N° : 07 / Année Universitaire 2018/2019

Table des matières

1	INTRODUCTION GENERALE	1
2	ETAT DE L'ART.....	3
2.1	INTRODUCTION.....	3
2.2	NOTIONS UTILISEES	3
2.3	REGIONALISATION ET VARIABILITE PLUVIOMETRIQUE DANS LE NORD CENTRE-OUEST ALGERIEN	7
2.4	ETUDE DE LA VARIABILITE ET DE LA STRUCTURE DE PLUIE ANNUELLE DANS UNE REGION SEMI-ARIDE.....	8
2.5	REGIONALISATION DES PLUIES ANNUELLES DANS L'ALGERIE NORD-OCCIDENTALE.....	9
2.6	VARIATIONS ET TENDANCES DES TEMPERATURES ET DES PRECIPITATIONS JOURNALIERES EN ALGERIE	10
2.7	CARACTERISATION CLIMATIQUE	10
2.8	EVOLUTION ET REGIONALISATION DES PRECIPITATIONS AU NORD DE L'ALGERIE	11
	CONCLUSION	12
3	ANALYSE FACTORIELLE	13
3.1	INTRODUCTION:.....	13
3.2	. ACP (ANALYSE EN COMPOSANTES PRINCIPALES) :.....	13
3.2.1	<i>Définition</i> :.....	13
3.2.2	<i>Historique</i> :.....	14
3.2.3	<i>Champs d'application</i> :	14
3.2.4	<i>Données d'une ACP</i> :.....	15
3.2.5	<i>Objectif de L'ACP</i>	15
3.2.5.1	L'espace des individus	16
3.2.5.2	L'espace des variables	18
3.2.5.3	Principe de construction de l'espace factoriel à partir du nuage des individus .	20
3.2.5.4	Principe de construction de l'espace factoriel à partir du nuage des variables .	29
3.2.5.5	Analyse en composantes principales normée.....	29
3.2.5.6	Matrice des données centrées réduites.....	30
3.2.5.7	Matrice de corrélation.....	31
	CONCLUSION	32
4	IMPLEMENTATION	33
4.1	INTRODUCTION.....	33
4.2	COLLECTED'INFORMATION	ERROR! BOOKMARK NOT DEFINED.
4.3	ANALYSE FACTORIELLE.....	34
4.3.1	<i>Analyse 2^K</i>	34
4.3.1.1	Résultat mensuelle	35
4.3.1.2	Résultat Annuel	36
4.4	INTERPRETATION DES RESULTANTS ACP	37
4.4.1.1	Interprétation des Résultats mensuelle	37
4.4.1.2	Interprétation des Résultats Annuelle	39

5 CONCLUSION GENERALE	41
ANNEXE.....	43
• <i>Présentation des données mensuelle</i>	<i>45</i>
• <i>Présentation des données annuelle</i>	<i>49</i>

Table des Figures

Figure 3-1 : Espace des individus.....	16
Figure 3-2 : Espace des variables.....	16
Figure 3-3 : Nuage des individus.....	17
Figure 3-4 : Nuage des variables.....	19
Figure 4-1 : Cercle de corrélation des variables.....	38
Figure 4-2 : présentation des variables mensuelle	38
Figure 4-3 : Espaces des variables des plans factoriels	40
Figure 4-4: présentation des variables Annuelle	40

Liste des tableaux

Tableau 3.1 Représentation type d'un tableau A.C.P.	15
Tableau 4.9 : Matrices de corrélation.....	35
Tableau 4.10 : Signifiante de Principal Components	35
Tableau 4.11 : Factor Loadings [Communality Estimates]	35
Tableau 4.12 : Matrices de corrélation (annuel).....	36
Tableau 4.13 : Signifiante de Principal Components	36
Tableau 4.14 : Factor Loadings [Communality Estimates]	36
Tableau 4.1 : donnée Précipitation montielle (mm).....	45
Tableau 4.2: donnée température mensuelle (°C)	46
Tableau 4.3 : donnée température min mensuelle (°C).....	46
Tableau 4.4: donnée température moyenne mensuelle (mm)	47
Tableau 4.5: donnée humidité mensuelle en (%)	47
Tableau 4.6 : données isolation mensuelle.....	48
Tableau 4.7: données de vent mensuelle	48
Tableau 4.8 : Présentation des données mensuelle	49

Remerciements

En premier lieu, nous remercions Dieu le tout puissant dont la sagesse et son savoir sont infinis. Il nous a donné santé, connaissance, force, courage et patience pour mener à bien notre travail et nous a guidés pour l'accomplissement de ce mémoire.

Nous tenons à exprimer notre profonde gratitude et nos sincères remerciements à Mr. GUELLOMA Founes

Pour ses dévouements, ses conseils qui nous ont éclairé dans l'élaboration de ce travail. Nous tenons également à remercier tous les enseignants qui nous ont encadrés tout au long de notre cursus universitaire. Tous les membres de nos familles et nos ami(e)s.

Nous tenons à remercier aussi les membres du jury d'avoir accepté de juger notre travail, et tout l'encadrement du Département d'informatique.

Chapitre 1

1 Introduction générale

Le changement climatique est un phénomène naturel qui influe positivement et négativement sur la vie humaine et la terre en générale directement et indirectement. Dans ce mémoire, nous faisons une étude descriptive de ces changements climatiques durant les 10 dernières années. Cette analyse nous aide pour faire une estimation du changement climatique futur.

Les statistiques descriptives comme le nom l'indique servent à analyser et décrire des données pour obtenir un rendu final. Ce sont de simples calculs Mathématiques qui permettent de dégager des données une réelle tendance positive ou négative des résultats. A partir de ces chiffres, des graphiques viennent en complément pour appuyer l'analyse statistique.

La variabilité climatique, en tant que phénomène a longtemps été étudiée et caractérisée. La question la plus importante, tant pour Laghouat que dans les autres régions de l'Algérie, est la recherche de la corrélation entre les paramètres climatiques (humidité, vent, température, précipitations,etc.)

Pour analyser ces paramètres et mettre en valeur la sensibilité de certains par rapport aux autres et dans le but de rendre l'information en connaissance, nous utilisons dans le cadre de ce mémoire quelques méthodes ; la méthode statistique dite ACP « Analyse en Composantes Principales », et la méthode de 2K factorial design.

Ce mémoire est organisé comme suite :

Le chapitre 1 est réservé à l'état de l'art. Nous présentons des études utilisées localement et mondialement.

Le chapitre 2 détaille les techniques ciblées pour notre étude à savoir ACP, 2K.

Le chapitre 3 contient les résultats de notre étude ainsi que l'analyse et l'interprétation.

Chapitre 2

2 Etat de l'art

2.1 Introduction

Le climat est un domaine de recherche vaste et riche et les chercheurs donnent des recherches très importantes qui nous aident dans la vie quotidienne parmi ces recherches l'analyse des données climatiques qui donne des résultats presque exacts, efficace et utiles dans d'autres domaines comme la navigation aérienne. Dans ce chapitre nous présentons une panoplie des études locales et mondiales sur l'analyse des données climatiques.

2.2 Notions utilisées

Nous expliquons quelques notions nécessaires pour la compréhension du reste de ce mémoire.

- **L'insolation solaire :**

L'évolution de la nébulosité d'une part, l'interposition fugace d'autres objets d'autre part, ont pour conséquence que la durée de l'insolation réellement subie par l'objet ne peut être calculée ni évaluée a priori, mais seulement mesurée chaque jour sur le terrain après le coucher du Soleil : cette durée effective d'insolation est inférieure ou, très rarement, égale à la durée géographique d'insolation.

Dans un site disposé de telle façon que la durée effective d'insolation n'y soit pas soumise à des influences aléatoires autres que celles des nuages, la mesure de cette durée représente par définition l'ensoleillement du site au lieu et au jour considérés ; il est ainsi extrêmement fréquent que l'on confonde les expressions "ensoleillement" et "durée d'insolation", voire "insolation". Il faut noter aussi que le

terme d'insolation est parfois employé comme simple synonyme de l'éclairement d'une surface horizontale par le rayonnement solaire. (Météo France, 2008)

▪ **Température :**

La température n'est pas une grandeur directement mesurable. Sa perception varie d'un individu à l'autre et selon les conditions atmosphériques (vent, pluie, ensoleillement...). Pour déterminer la grandeur physique de la température, plusieurs échelles ont été définies. En France et dans la majorité des pays on utilise le degré Celsius. Le 0°C de cette échelle correspond au point de congélation de l'eau, et le 100°C au point d'ébullition de l'eau. (Météo France, 2008)

▪ **Température moyenne :**

- température moyenne quotidienne : elle est calculée à partir de deux mesures de la température (demi-sommeil de la température maximale du jour et de la température minimale) ou plus selon les mesures disponibles.
- La température moyenne mensuelle : moyenne calculées à partir des 28, 29, 30 ou 31 températures moyennes quotidiennes.
- La température moyenne annuelle : moyenne des 365 ou 366 moyennes quotidiennes. (Météo France, 2008)

▪ **Vent :**

En météorologie, le vent désigne le mouvement horizontal de l'air. Sa mesure comprend deux paramètres : sa direction et sa vitesse ou force. La vitesse est exprimée communément en km/h ou m/s. Marins et pilotes utilisent les nœuds (1 nœud = 1,852 km/ h). La mesure du vent est toujours une moyenne sur une période donnée. L'anémomètre permet de mesurer la vitesse du vent. (Météo France, 2008)

- **L'humidité**

Le paramètre humidité désigne en météorologie la quantité de vapeur d'eau contenue dans l'air. La vapeur d'eau est l'une des formes que l'eau – indispensable à la vie – peut prendre sur terre. Les autres sont la forme liquide (océans, lacs, pluie...) et la forme solide (glaciers, banquise, neige...). La vapeur d'eau est absolument transparente et invisible. Rappelons que l'atmosphère est en moyenne essentiellement composée d'azote (78 %) et d'oxygène (21%) laissant aux autres gaz 1% en moyenne. Ainsi la proportion de vapeur d'eau varie de 0,1 % à tout au plus 5% de l'air qui nous entoure ! Pourtant c'est sa présence dans l'atmosphère qui conditionne la présence ou non de nuages, de précipitations, de brouillard. Ainsi, l'air ambiant est humide même lorsqu'il ne « mouille » pas ! Sortez un verre de thé glacé du réfrigérateur, rapidement, les parois vont ruisseler de gouttelettes d'eau. Cette eau provient de la condensation de la vapeur d'eau présente dans l'air. L'air contient, selon sa température, une quantité maximale variable de vapeur d'eau. En effet, le gaz « vapeur d'eau » se condense lorsque la température baisse et passe alors de la forme gazeuse à la forme liquide : c'est la saturation. Au-delà de la saturation, les molécules d'eau, jusque-là indépendantes, se déposent sur tout support (la tasse de thé, la végétation, ou des poussières en altitude) et forment des gouttelettes visibles à l'œil nu. C'est la condensation. En altitude, ces gouttelettes formeront les nuages ; au sol, on observera de la rosée (ou de la gelée blanche en cas de température négative) ou, s'il y a un peu de vent, de la brume voire du brouillard. (Météo France, 2008)

- **Réchauffement climatique :**

Le réchauffement climatique, réchauffement planétaire, réchauffement global¹,

N1 ou dérèglement climatique est le phénomène d'augmentation des températures moyennes océaniques et de l'air, induit par la quantité de chaleur piégée à la surface terrestre, mesurée depuis plusieurs décennies, du fait des émissions de gaz à effet de serre (CO₂, etc.). Ce terme désigne communément le

réchauffement mondial observé depuis le début du siècle. On rencontre fréquemment l'expression « changement climatique » utilisée pour désigner le réchauffement climatique, alors qu'en principe le changement climatique désigne les épisodes de réchauffement ou refroidissement d'origine naturelle qui se sont produits avant l'ère industrielle. (wikipedia, 2019)

- **Station météorologique :**

Une station météorologique, parfois désignée par synecdoque abri météorologique, est un ensemble de capteurs qui enregistrent et fournissent des mesures physiques et des paramètres météorologiques liés aux variations du climat, ces capteurs étant placés dans un boîtier, abri météorologique qui réalise l'équilibre thermique du thermomètre avec l'air et le protège du rayonnement solaire¹. Les variables à mesurer sont la température, la pression, la vitesse et direction du vent, l'hygrométrie, le point de rosée, la pluviométrie, la hauteur et le type des nuages, le type et l'intensité des précipitations ainsi que la visibilité. Les stations peuvent comporter des capteurs pour toutes ou une partie seulement de ces informations, selon leur type : agro-météorologique, d'aéroport, météo routière, climatologique, etc.

Les stations météorologiques sont le plus souvent à des coordonnées fixes sur les îles et les continents mais elles peuvent être aussi être à bord d'avions avec (AMDAR) ou de navires (en particulier les navires météorologiques). Elles peuvent également être déplacées pour des usages particuliers comme des expériences scientifiques ou pour déterminer les conditions météorologiques lors d'un désastre tel un feu de forêt. Les données qu'on en obtient peuvent être envoyées directement comme rapport météorologique, dans le cas d'une station automatique, ou faire partie des observations METAR émises par un observateur humain. (wikipedia, 2019)

- **Les pluviométries**

Les précipitations désignent tous les météores qui tombent dans une atmosphère et il peut s'agir de solides ou de liquides selon la composition et la température de cette dernière. Ce terme météorologique est le plus souvent au

pluriel et désigne sur la Terre les hydrométéores (cristaux de glace ou gouttelettes d'eau) qui, ayant été soumis à des processus de condensation et d'agrégation à l'intérieur des nuages, sont devenus trop lourds pour demeurer en suspension dans l'atmosphère et tombent au sol ou s'évaporent en virage avant de l'atteindre. Par extension, le terme peut également être utilisé pour des phénomènes similaires sur d'autres planètes ou lunes ayant une atmosphère. (France, Météo, 2008)

2.3 Régionalisation et variabilité pluviométrique dans le nord Centre-ouest algérien

L'objectif principal de cette étude c'est l'évolution de la variabilité de pluie ainsi que la détermination des régions ayant les mêmes conditions pluviométriques dans le Centre et Ouest Nord d'Algérie. Les données sont récoltées sur un total de 127 postes pluviométriques durant la période d'observation entre 1930-2012. Elle utilise deux techniques statistiques complémentaires :

Les tests de statistiques de détection de rupture et l'Analyse en Composantes Principales (ACP).

D'après les tests statistiques une variabilité intra annuelle des précipitations est plus importante pour l'ensemble des stations surtout à l'Ouest à partir de la moitié des années 70 (17%).

La régionalisation obtenue par l'analyse en composantes principale a permis d'identifier cinq régions principales :

La région Nord Centre : caractérisée par une diminution des précipitations annuelles à partir des années 70 surtout pour les saisons d'automne et d'hiver.

La région Nord-Ouest : caractérise par un déficit pluviométrique annuelle important à partir de 1976. Cette région a connu une baisse des précipitations en hivers et printemps.

La région Centre-Est : aucune variation pluviométrique enregistrée pendant la période d'observation.

Le Centre et Centre-Ouest : caractérisée par les déficits pluviométriques les plus importants de la zone d'étude surtout en hiver et au printemps à partir des années 70. (mounia, 2017)

2.4 Etude de la variabilité et de la structure de pluie annuelle dans une région semi-aride

L'objectif de cet article est de caractériser la variabilité de la pluie dans le bassin versant de la Macta. D'une superficie de 14390 Km², le bassin versant de la Macta est située au nord-ouest de l'Algérie. Les données sont récoltées sur un réseau de 48 pluviomètres. Cette caractérisation permettra de mieux gérer l'incertitude, liée au caractère aléatoire, des précipitations et de mieux prévoir les risques qui peuvent y être engendrés. Les fluctuations dans la répartition annuelle des hauteurs de pluie sont considérables et préoccupantes pour le développement et la mise en valeur des terres agricoles. L'évidence des premiers caractères du régime de la distribution géographique des précipitations en Algérie a été mise, uniquement, par l'Analyse en Composantes Principales (ACP). Ce travail pose la même problématique, mais en utilisant deux techniques statistiques complémentaires : l'Analyse en Composantes Principales (ACP) et l'Analyse Factorielle Discriminante (AFD). Dans cette étude, la régionalisation de la pluie annuelle est approchée par l'analyse des proximités sur le premier facteur principal (ACP). Les régions obtenues seront validées par la méthode de l'AFD ; plus appropriée. L'étude de la structure de la pluie est effectuée à l'aide du variogramme moyen. L'isotropie du champ pluviométrique est analysée par des variogramme directionnels. L'optimisation de la carte pluviométrique est dressée par Krieger en voisinage glissant, elle sera accompagnée de la carte d'estimation des erreurs d'interpolation. (CHIBANE, BENTCHAKAL, MEDJERAB, & BENFARES, 2015)

2.5 Régionalisation des pluies annuelles dans l'Algérie nord-occidentale

La pluviométrie présente un intérêt majeur comme signature au-delà de son apport pour la définition du climat. Sa régionalisation s'avère très utile dans de nombreuses applications où il est nécessaire de minimiser la variabilité spatiale du paramètre analysé : détermination des « stations étalons », contrôle et reconstitution de données manquantes (Champeaux et al., 1996).

L'objectif d'une régionalisation (ou zonage) climatique est d'obtenir un découpage d'un territoire en zones homogènes, à l'intérieur desquelles le comportement climatique est similaire au(x) paramètre(s) étudié(s). À cet effet, nous pouvons distinguer, dans la zone d'étude, plusieurs échelles spatiales emboîtées :

Une échelle microclimatique, de distance inférieure au kilomètre, qui représente en fait le parcellaire souvent confondue, par l'expression, avec l'échelle locale ;

Une échelle régionale, de distance caractéristique de l'ordre de 100 à 200 kilomètres, pour laquelle interviennent essentiellement la latitude, la distance à la mer, et le positionnement par rapport aux massifs montagneux, une échelle, appelée topoclimatique de l'ordre de 10 à 20 kilomètres ;

Les recherches entreprises sur le climat de l'Algérie par l'Agence Nationale des Ressources Hydrauliques et l'Office National de la Météorologie, ont mis en évidence les premiers caractères du régime de la distribution géographique des précipitations par ACP seulement. Ce travail pose la même problématique, mais en utilisant deux techniques mathématiques complémentaires : l'Analyse en Composantes Principales et l'Analyse Factorielle Discriminante, en essayant toutefois d'interpréter le régime de la distribution géographique des pluies pour les régions homogènes mises en évidence. (Medjerab & Henia, 2005)

2.6 Variations et tendances des températures et des précipitations journalières en Algérie

Les variations et les tendances des températures et des précipitations journalières en Algérie ont été établies sur une période de trente-quatre ans (1982-2016), en utilisant la méthode graphique, les tests statistiques, l'indice pluviométrique standardisé, l'indice d'aridité et l'analyse en composantes principales. Une baisse des précipitations est observée dans la région d'étude surtout à l'ouest de l'Algérie avec des années déficitaires.

L'analyse en composantes principales a mis en évidence quatre grandes régions. La région centre littorale est caractérisée par une légère baisse des précipitations annuelles. La région Est littorale et les hauts plateaux Est connaissent une stabilité du régime pluviométrique. La région Ouest littorale subit un déficit pluviométrique annuel important à partir de (1982-2004). La région des hauts plateaux Ouest et Centre sont caractérisées quant à elles par des déficits pluviométriques. La région Sud où le climat est sec. Le Sahara est une région très ventée et aride.

Les températures ont connu une augmentation de plus de 0.5°C à partir des années 1980. La tendance à la baisse des précipitations et l'augmentation des températures s'explique principalement par le renforcement de l'anticyclone des Açores (Il empêche l'arrivée de l'air froid polaire). L'année 2013 pourrait être le début d'une période prolongée de sécheresse et de faible pluviométrie.

(BENYETTOU & BOUKLIKHA, 2016)

2.7 Caractérisation Climatique

Cette étude présente l'intérêt de certaines méthodes statistiques, pour interpréter la variabilité spatiotemporelle des précipitations des seize stations pluviométriques situées dans le bassin versant de Medjerda (Nord- Est Algérien) sur des périodes allant de 19 à 38 ans. Les séries analysées présentent toutes des ruptures dont la majorité se trouve dans les décennies 80 et 90. La cartographie des indices pluviométriques interannuels calculés par décennie sur la période 1971-2007

montre que les décennies 80 et 90 apparaissent comme déficitaires sur tout le territoire du bassin. Ainsi, une tendance à un retour des précipitations est remarquée pour la période (2001-2007). À l'échelle mensuelle, les précipitations ont augmenté significativement après les années de rupture pour les mois d'hivers et du printemps, à l'exception du mois de mars. On observe aussi, une concentration exceptionnelle de pluie dans le mois d'Août. À l'échelle journalière, l'analyse des fractions pluviométriques journalières met en évidence une tendance à l'augmentation des pluies fortes supérieures à 20 mm après les années de ruptures. Le bassin de la Medjerda connaît aussi une réduction de la pluviométrie et une augmentation de la température sur une période allant de (1913-1938) à (1969-2007). (Hammar, Djebbar, & Khoualdia, 2014)

2.8 Evolution et régionalisation des précipitations au nord de l'Algérie

Une analyse du régime pluviométrique du Nord de l'Algérie a été établie sur une période de sept décennies (1936–2009). Pour cela 102 séries pluviométriques ont été sélectionnées et soumises à cinq tests statistiques pour la détection de tendance ou de rupture. Une baisse significative des précipitations est observée à partir de la moitié des années 70 particulièrement dans la région Ouest (déficit pluviométrique entre 16 et 43%). L'indice pluviométrique standardisé a montré que les décennies 80 et 90 étaient les plus déficitaires. Ce résultat coïncident dans l'ensemble avec les dates de rupture trouvées au Maroc. La régionalisation des précipitations par l'analyse en composantes principales a mis en évidence six régions homogènes. Ce découpage spatial montre aussi la vulnérabilité de chaque région au changement climatique. (MEDDI, SOUAG, TAIBI, & GIL, 2013)

Conclusion

L'analyse des données est un domaine de recherche très vaste et n'est pas facile. Basé sur le partage de connaissance, ce chapitre présente quelques résumés des études du même but pour nous aider à réaliser notre analyse.

Pour une analyse de paramètres, nous proposons des techniques (méthodes) statistique comme ACP pour explorer la corrélation entre les paramètres climatique. C'est ainsi que nous consacrerons le chapitre qui suit.

Chapitre 3

3 Analyse Factorielle

3.1 Introduction:

Dans le cadre de cette étude, nous nous basons sur les méthodes ACP et 2K pour analyser les paramètres définis, afin de définir les corrélations existantes entre ces derniers pour éviter la redondance.

C'est ainsi que nous consacrons ce deuxième chapitre à la présentation des méthodes et l'utilité de ces derniers.

3.2 . ACP (Analyse en Composantes Principales) :

3.2.1 Définition :

L'Analyse en Composantes Principales (ACP) fait partie du groupe des méthodes descriptives multidimensionnelles appelées méthodes factorielles, qui consiste à transformer des variables liées entre elles (dites "corrélées" en statistique) en nouvelles variables dé-corrélées les unes des autres. Ces nouvelles variables sont nommées "composantes, principales", ou axes principaux. Elle permet au praticien de réduire le nombre de variables et de rendre l'information moins redondante.

Il s'agit d'une approche à la fois géométrique (les variables étant représentées dans un nouvel espace, selon des directions d'inertie maximale) et statistique (la recherche portant sur des axes indépendants expliquant au mieux la variabilité (la variance) des données).

Lorsqu'on veut compresser un ensemble de p variables aléatoires, les q premiers axes de l'analyse en composantes principales sont un meilleur choix, du point de vue de l'inertie ou de la variance.

L'A.C.P. permet d'explorer les liaisons entre variables et les ressemblances entre individus.

3.2.2 Historique :

L'analyse en composantes principales est la plus utilisée des analyses multi variées. C'est aussi la plus ancienne. Ses fondements datent de 1901 quand Karl PEARSON en a conçu les idées essentielles à deux dimensions (Benzédrine, 1973). Il prolonge ses travaux dans le domaine de la régression et des corrélations entre plusieurs variables. Pearson utilise ces corrélations non plus pour expliquer une variable à partir des autres (comme en régression), mais pour décrire et résumer l'information contenue dans ces variables.

Sa généralisation a été faite en France en 1933 par Harold HOTELLING qui en a fait une première publication sous forme actuelle à plusieurs dimensions. Elle a attendu l'ordinateur vers les années 1960 pour être mise en pratique par J-Penzé.

3.2.3 Champs d'application:

Les champs d'application sont aujourd'hui multiples, allant de la biologie à la recherche économique et sociale. L'ACP est majoritairement utilisée pour :

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

3.2.4 Données d'une ACP :

L'ACP s'intéresse à des tableaux de données rectangulaires avec des individus en lignes et des variables quantitatives en colonnes (tableau 1.1). L'ACP vise à décrire graphiquement un tableau de données d'individus n avec leurs variables quantitatives p de grande taille :

Individus/ Variables	Var ₁ ...var _j ...Var _p
Ind ₁	
.	
.	
Ind _i	X _{ij}
.	
.	
ind _n	

Tableau 3.1 Représentation type d'un tableau A.C.P.

Un tel tableau de données sera tout simplement une matricée réelle à n lignes (les individus) et à p colonnes (les variables) (formule 1.1) :

$$X = (X_{ij})_{i = 1 \dots n; j = 1 \dots p} \quad (1.1)$$

$$X = \begin{bmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{bmatrix} \quad (1.2)$$

3.2.5 Objectif de L'ACP

L'objectif de L'ACP est de condenser l'information contenu dans le tableau par un Analyse de corrélation linéaire entre les variables et une visualisation graphique des distance entre des individus.

On dispose d'un tableau de donnée X_{ij} , On peut interpréter géométriquement les lignes et les colonnes du tableau de donnée par des points dans deux espaces différents : l'espace des individus (figure 3.1) et/ou l'espace des variables (figure 3.2)

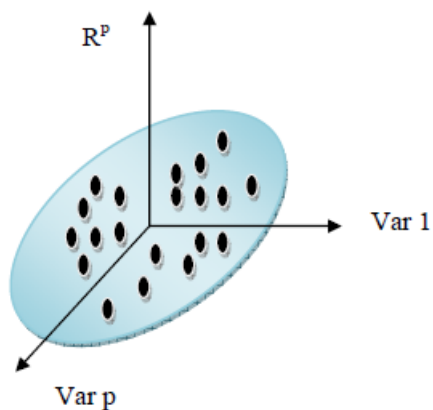
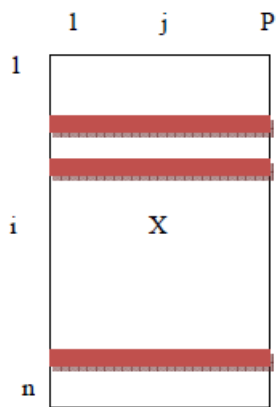


Figure 3-1 : Espace des individus.

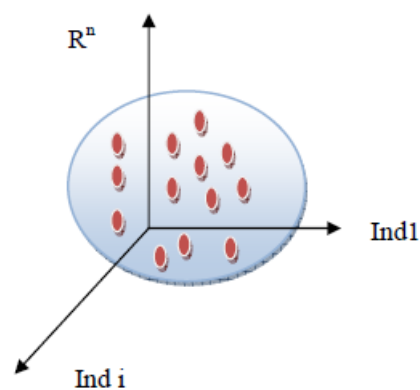
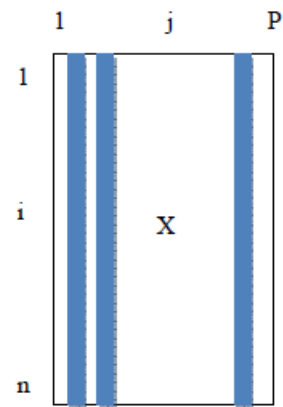


Figure 3-2 : Espace des variables.

3.2.5.1 L'espace des individus

Dans l'espace des individus (figure 3-1), les axes représentent les variables et les individus sont représentés comme un nuage de points d'un ensemble d'individus. Etudier les individus revient à considérer ce nuage de points dans un espace à R^p , avec p le nombre de variables ; cela revient à étudier les lignes du tableau 3.1.

Dans l'espace des individus, chacun de ces derniers est représenté par un point et l'ensemble forment ainsi un nuage de points. Le but de l'ACP est de visualiser ce nuage dans un espace de faible dimension le plus fidèlement possible.

Nous identifierons l'individu i avec le point ligne $x_i = (x_{i1}, \dots, x_{ip})$ qui sera considéré comme un point dans un espace affine (l'espace des individus) à p dimension.

L'analyse du nuage des individus utilise la notion fondamentale de distance.

Pour faire une représentation géométrique, il faut choisir une distance entre deux points de l'espace.

1 individus = une ligne du tableau \Rightarrow un point dans un espace à p dimensions

Si $j = 1$: on a une représentation axiale ;

Si $j = 2$: on peut construire un nuage de point avec la régression simple ;

Si $j = 3$: on peut avoir une représentation plus difficile en 3 dimensions ;

Si $j = 4$: impossible à représenter mais le concept est simple ; pour cela on doit définir la notion de ressemblance.

On peut dire que deux individus se ressemblent s'ils sont proches, et la ressemblance entre deux individus est mesurée par la distance euclidienne.

La distance utilisée par l'ACP dans l'espace R^p à p dimensions ou sont représentés les unités, est la distance euclidienne classique. La distance euclidienne entre deux individus e_i et e_j (figure 1-3) s'écrit :

$$d^2(e_i, e_{i'}) = \sum_{j=1}^p (x_i^j - x_{i'}^j)^2 \quad (3.3)$$

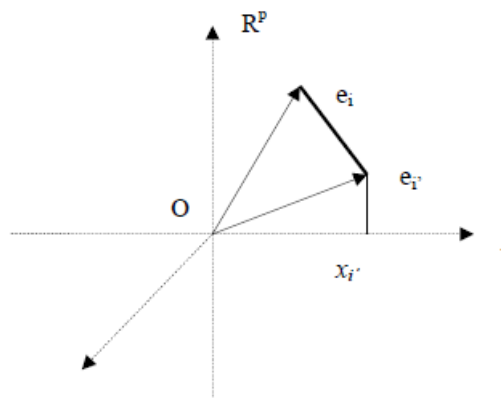


Figure 3-3 : Nuage des individus

Deux points sont très proche si les p coordonnées de ces deux points sont très proche (mêmes valeurs pour ces différentes variables).

❖ **Remarque :**

Pour permettre de donner à chaque individu la même importance dans le calcul de la distance on suppose que tous les individus ont le même poids et il est égales a $\frac{1}{n}$; Chaque individus i est muni d'un poids $p_i > 0$ et $\sum_{i=1}^n p_i = 1$ (3.4)

Dans le cas le plus usuel de poids égaux, la matrice diagonale des poids

$$D = \frac{1}{n}I \quad (3.5)$$

Où I est la matrice identité.

$$D = \text{diag}(p_i) = \begin{pmatrix} p_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & p_n \end{pmatrix} \quad (3.6)$$

3.2.5.2L'espace des variables

Dans l'espace des variables (figure 3-2), les axes représentent les individus et les variables sont représentés comme un nuage de points d'un ensemble de variables. Etudier les variables revient à considérer ce nuage de point dans un espace à R^n , avec n le nombre d'individu ; cela revient à étudier les colonnes du (tableau 3-1).

Les p colonnes peuvent être considérés comme p points dans un espace vectoriel à n dimensions dans R^n . Cet espace est appelé l'espace des variables.

Si les valeurs prises par deux variables sont très voisines pour l'ensemble des individus, ces variables seront présentées par des variables très proches (ce qui peut signifier que les variables mesurent la même chose ou encore qu'elles sont liées par une relation particulière). Pour chaque variable aléatoire j on définit sa moyenne par :

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij} \quad (3.7)$$

Et son écart type par :

$$\sigma_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2} \quad (3.8)$$

On peut alors définir la corrélation entre les variables, cela revient à étudier la forme du nuage.

Le cosinus de l'angle formé par 2 variables (figure 3.4) est égal au coefficient de corrélation linéaire entre ces variables :

$$\cos(\alpha) = \frac{\langle \overline{ol}, \overline{ok} \rangle}{\|\overline{ol}\| \|\overline{ok}\|} = \frac{\sum_{j=1}^p x_{il}^j x_{ik}^j}{\sqrt{\sum_{j=1}^p (x_{il}^j)^2} \sqrt{\sum_{j=1}^p (x_{ik}^j)^2}} = r(x_{i'}^j, x_{i''}^j) \quad (3.9)$$

Tout en sachant qu'un coefficient de corrélation entre les variables de x et y est égale à :

$$r_{xy} = \frac{\cos(x,y)}{\sqrt{\text{var}(x)}\sqrt{\text{var}(y)}} = \cos(x,y) \quad (3.10)$$

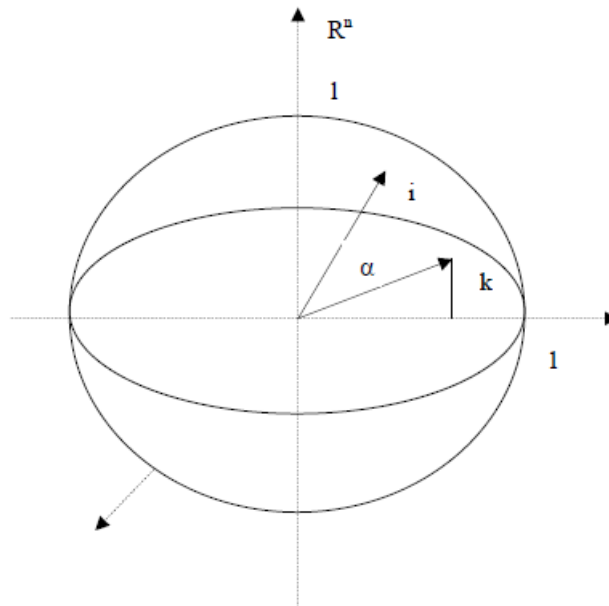


Figure 3-4 : Nuage des variables.

❖ Problématique

Difficulté à mettre en évidence les relations globales existantes entre les variables dès que $p > 3$, et les ressemblances entre individus dès que $n > 3$, car impossibles de les visualiser.

Lorsqu'il n'y a que deux dimensions (largeur et longueur par exemple), il est facile de représenter les données sur un plan :

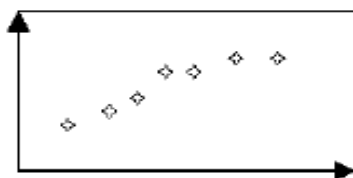


Figure 3-5 : Représentation du nuage à deux dimensions

Avec trois dimensions (largeur, hauteur et profondeur par ex.), c'est déjà plus difficile :

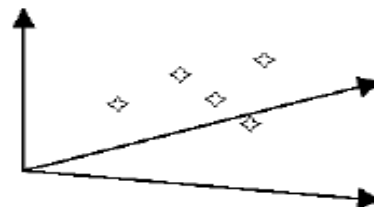


Figure 3-6 : Représentation du nuage à trois dimensions

❖ Solution

Condenser l'information du tableau de manière à retirer les relations vraiment caractéristiques (proximités entre variables et individus), ceci en limitant la perte d'information. C-à-d déterminer un sous-espace de dimension $q < p$ (q nouveaux axes) ($ou\ q < n$), sur lequel projeter les nuages de points relatifs au tableau de données qui soit :

- « compréhensible » par l'œil : q faible, de préférence $q=1,2$ ou 3 ;
- le moins déformant possible (projection la plus fidèle possible)

Ce sous-espace est appelé *espace factoriel* du nuage.

3.2.5.3 Principe de construction de l'espace factoriel à partir du nuage des individus

On effectue un changement de repère, passant du repère défini par les p variables à un repère de dimension p le moins déformant possible pour le nuage. Il sera défini par p nouveaux axes, appelés axes factoriels ou axes principaux.

On retient ensuite les q premiers axes du nouveau repère, ce qui nous donnera l'espace factoriel de dimension q . Il permet de récupérer les liens les plus significatifs contenus dans le tableau.

Les p axes factoriels sont définis séquentiellement :

- On détermine l'axe (premier axe factoriel) sur lequel le nuage se déforme le moins possible en projection, c à d le premier axe factoriel est donc l'axe selon lequel est préservé, par projection, le maximum de la dispersion initiale des points du nuage.

- On cherche un second axe, sur lequel le nuage se déforme le moins en projection, après le premier axe, tout en étant orthogonal au premier,

- On réitère jusqu'à l'obtention de p axes.

Un facteur est une variable composite fabriquée à partir des variables d'origine ; il s'exprime comme combinaison linéaire des anciennes variables. Le 1er axe factoriel correspond à la variable composite qui différencie le mieux les individus.

- **Comment obtenir une déformation minimale ?**

Chacune des n lignes du tableau X peut être considérée comme un vecteur, ou encore un point de R^p . L'ensemble du tableau constitue alors un nuage de n points.

L'ACP vise à fournir une image simplifiée du nuage de point qui soit la plus fidèle possible. Elle permet d'avoir une idée sur les distances entre individus, cela revient à trouver un sous espace qui résume le mieux les données, c.-à-d. une image est bonne si elle restitue le plus fidèlement possible le nuage originale. Pour avoir une bonne qualité d'une image il faut :

- restituer le plus fidèlement la forme générale du nuage. Les points doivent être bien séparés pour permettre de visualiser les distances entre les individus, donc cela revient à augmenter la dispersion, i.e. la variabilité (l'inertie), du nuage ;
- ne pas perturber les distances entre individus.
- Choix de l'origine du nuage d'individus (vecteur moyen ou centre de gravité) :

Le point o correspondant au vecteur de coordonnées toutes nulles n'est pas forcément une origine satisfaisante, car si les coordonnées des points du nuage des individus sont grandes, le nuage est éloigné de cette origine. Il apparaît plus judicieux de choisir une origine liée au nuage lui-même : le centre de gravité du nuage (figure 3.5), cela va permettre au nuage d'être plus dispersé.

Le centre de gravité G du nuage des individus est alors le point dont les coordonnées sont les valeurs moyennes des variables :

$$G = \left(\frac{1}{n}\right) \sum_{i=1}^n X_i = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} \\ \cdot \\ \cdot \\ \frac{1}{n} \sum_{i=1}^n x_{ij} \\ \cdot \\ \cdot \\ \frac{1}{n} \sum_{i=1}^n x_{ip} \end{pmatrix} = \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ x_j \\ \cdot \\ \cdot \\ x_p \end{pmatrix} = \begin{pmatrix} \bar{X}_1 \\ \cdot \\ \cdot \\ \bar{X}_2 \\ \cdot \\ \cdot \\ \bar{X}_3 \end{pmatrix} \quad (3.11)$$

Comme nous considérons chaque variable, comme chaque individu, ayant le même poids, G a alors pour coordonnées dans le repère : $(0, \vec{e}_1, \vec{e}_2, \vec{e}_3)$

$$OG = \begin{pmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_j \\ \vdots \\ \bar{X}_p \end{pmatrix}$$

(3.12)

Sous forme graphique :

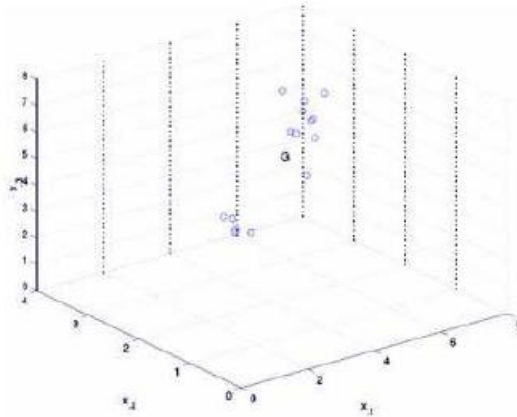


Figure 3-7 : Représentation de G dans le repère $(0, \vec{e}_1, \vec{e}_2, \vec{e}_3)$.

Prendre G comme origine, conformément à la figure précédente, revient alors à travailler sur le tableau des données centrées :

$$X_c = \begin{bmatrix} x_{11} - x_{.1} & \cdots & x_{1j} - x_{.j} & \cdots & x_{1p} - x_{.p} \\ \vdots & & \vdots & & \vdots \\ x_{i1} - x_{.1} & \cdots & x_{ij} - x_{.j} & \cdots & x_{ip} - x_{.p} \\ \vdots & & \vdots & & \vdots \\ x_{n1} - x_{.1} & \cdots & x_{nj} - x_{.j} & \cdots & x_{np} - x_{.p} \end{bmatrix} \quad (3.13)$$

Centrer les données ne modifie pas la forme du nuage, ça revient juste à translater le nuage vers une autre origine qui est le centre de gravité du nuage.

❖ **Remarque**

La matrice des données centrées contient les coordonnées centrées (que nous noterons X_{eij}) des individus dans le repère $(G, \vec{e}_1, \vec{e}_2, \vec{e}_3)$. Nous nous placerons dans la suite toujours dans ce repère pour le nuage de points des individus et nous prendrons $O = G$ (figure 3.6)

Et sous forme graphique :

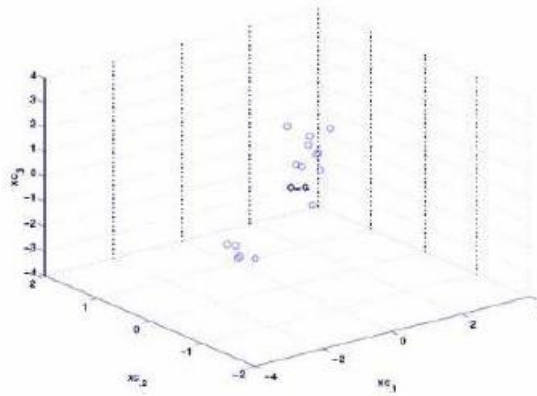


Figure 3-8 : Représentation des variables centrées dans le repère $(G, \vec{e}_1, \vec{e}_2, \vec{e}_3)$.

Quantification de la qualité de l'image

On quantifie la qualité d'une image avec la notion de dispersion ou de variabilité appelé inertie (formules (3.14), (3.15)) qui n'est autre que la mesure des dispersions, qui est l'équivalent d'une variance (formule (3.11)) mais généralisé à plusieurs dimensions. Si cette inertie est grande, cela signifie que le nuage est très dispersé, tandis que si elle est petite, alors le nuage est très concentré sur son centre de gravité.

- Détermination de la matrice de variance –covariance

Avant de déterminer le plan factoriel, nous allons chercher à détecter les liens possibles entre les variables, pour cela nous devons déterminer la matrice variance-covariance afin de connaître la variabilité du nuage.

Nous rappelons que la variance et la covariance entre deux variables et est donnée par :

$$Var_X = \sigma_X^2 \quad (3.14)$$

Et

$$COV_{xy} = \frac{1}{n} \sum_{i=1}^n X_i Y_i - \bar{X}\bar{Y} \quad (3.15)$$

Soit X_c la matrice des données centrées. Pour avoir la matrice des variances covariances nous allons pré-multiplier la matrice des données centrées par sa transposée. La somme des éléments de la diagonale principale (des variances), la trace par définition, représente l'inertie du nuage de points (formule 3.16).

La matrice des variances - covariances est égale à :

$$V = \frac{1}{n} X_c^t X_c = \frac{1}{n} \begin{bmatrix} varx_1 & cov(x_1, x_2) & \cdots & cov(x_1, x_p) \\ & \vdots & & \\ cov(x_2, x_1) & varx_2 & \cdots & cov(x_2, x_p) \\ & & \vdots & \\ cov(x_n, x_1) & cov(x_n, x_2) & \cdots & varx_p \end{bmatrix} \quad (3.16)$$

Nous voyons facilement que la matrices des covariances et au coefficient 1/n près de la matrice des produit scalaires canoniques des vecteurs de la matrice des données centrées (en d'autres termes, chaque composante de la matrice des covariances est égale au produit scalaire des variables centrées).

- **Calcul de l'inertie totale du nuage des individus**

On note I_G le moment d'inertie du nuage des individus par rapport au centre de gravité G :

$$I_G = \frac{1}{n} \sum_{i=1}^n d^2(G, u_i) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - x_{.j})^2 \quad (3.17)$$

Ce moment d'inertie totale est intéressant car c'est une mesure de la dispersion du nuage des individus par rapport à son centre de gravité.

Remarque :

On peut voir, en inversant l'ordre des signes sommes, que I_G peut aussi s'écrire sous la forme suivante :

$$I_G = \sum_{j=1}^p \left[\frac{1}{n} \sum_{i=1}^n (x_{ij} - x_{.j})^2 \right] = \sum_{j=1}^p var(v_j) \quad (3.18)$$

Ou $Var(v_j)$ est la variance empirique de la variable v_j . Sous cette forme, on constate que l'inertie totale est égale à la trace de la matrice de variance-covariance des p variables v_j :

$$I_G = trace(V) = \sum_{i=1}^p v_i = \sum_{i=1}^p \lambda_i \quad (3.19)$$

Avec λ_i les valeurs propres de la matrice des variances-covariances :

$$det(V - \lambda I) = 0$$

Une valeur propre représente la variance des individus sur l'axe correspondant.

- Interprétation de l'espace factoriel du nuage de points individus

Chaque axe factoriel k , de vecteur directeur u_k , représente une nouvelle variable C_k de dimension p (figure 3.7), construite comme combinaison linéaire des variables (axes) de départ, appelée composante principale (formule (3.21)). La coordonnée C_{ik} d'un individu i donné sur cet axe correspond à la valeur de la composante principale prise par cet individu.

Les composantes principales sont construites de manière à restituer la majeure partie de l'information du tableau, (elles déforment le moins possible l'information). La première composante principale sera une combinaison linéaire des variables de départ de dispersion maximale (c'est-à-dire celle qui ont la plus grande variance). Les axes qu'elles déterminent sont des axes principaux (formules (3.18)) et les formes linéaires associées sont des facteurs principaux.

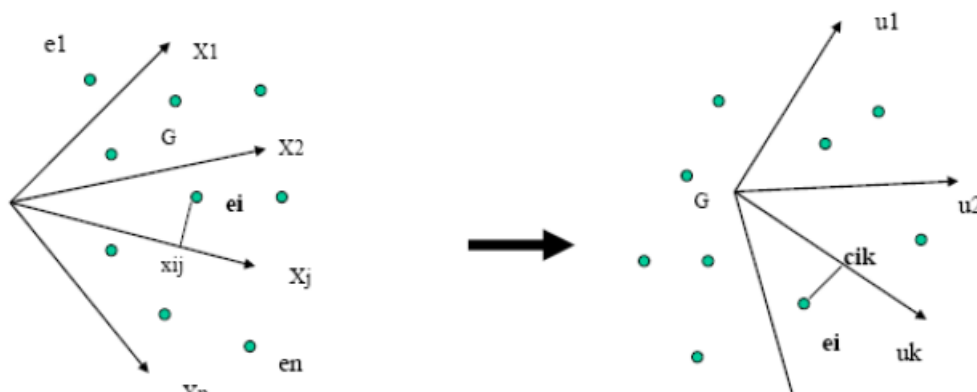


Figure 3-9 : Construction d'un espace factoriel (Duby et al, 2008).

Figure 3-9 : Construction d'un espace factoriel .

Pour pouvoir trouver la meilleure image approchée du nuage, il faut que l'axe sur lequel on projette le nuage permette une dispersion maximale, pour cela on doit déterminer les axes factoriels (axes principaux) par le principe d'ajustement ou par calcul numérique.

- Détermination des axes principaux

- Ajustement du nuage des individus

Le principe d'ajustement est de perdre le moins d'information possible, c.-à-d. chercher un sous espace qui permet de voir au mieux le nuage des individus.

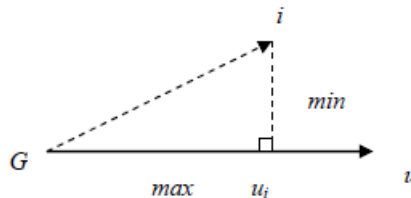


Figure 3-10 : Ajustement d'un individu i.

Pour trouver l'axe factoriel (facteur) qui déforme le moins possible le nuage, il faut que la distance Gu_i soit grand ;

Donc :

$(Gu_i)^2$ Grand \leftrightarrow $(iu_i)^2$ Petit avec $u_i <$ axe (Pythagore) \Rightarrow On doit maximiser la somme $\sum_i (Gu_i)^2$

- Maximiser $\sum_i (Gu_i)^2$ revient à trouver le meilleur plan, avec $u_i <$ plan (donc la dispersion soit la plus grande possible).

- Le meilleur plan contient le meilleur axe factoriel : pour trouver le meilleur plan on doit trouver le meilleur axe et ensuite trouver un axe orthogonal au premier axe ($u_1 \perp u_2$) et qui maximise l'inertie projetée $\sum_i (Gu_i)^2$;

- Puis on cherche un 3^{ème} axe, orthogonale au deux premiers et d'inertie maximum.

- Calcul numérique des axes principaux

On appelle axes principaux d'inertie les axes de direction des vecteurs propres.

Un vecteur dit « vecteur propre » par une application linéaire si l'application ne fait que modifier sa taille sans changer sa direction.

Un vecteur propre x de composante (x', x'') , Associé à la valeur propre λ doit vérifier la relation :

$$(V - \lambda I_n) \begin{pmatrix} x' \\ x'' \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \quad (3.21)$$

Tel que V est la matrice des variances-covariances.

Le premier axe est celui associé à la plus grande valeur propre λ_1 ;

Le deuxième axe est celui associé à la deuxième valeur propre λ_2 ;

Et ainsi de suite...

- Sélection du nombre q axes factoriels à retenir

La sélection du nombre d'axes à retenir est une étape importante d'une ACP. Les critères utilisés sont en général empiriques : la méthode du coude qui consiste à détecter un coude sur le diagramme des valeurs propres, ou le critère de Kaiser. Ce dernier consiste à ne retenir que les valeurs propres supérieures à la moyenne ; et lorsqu'on est dans une ACP normée la somme des valeurs propres est égale au nombre p de ces valeurs propres, et la moyenne des valeurs propres est donc égale à 1. On ne retiendra donc par le critère de Kaiser que les valeurs propres supérieures à 1.

- Pourcentage d'inertie expliquée

Le pourcentage de variance expliquée par les axes factoriels retenus est obtenu par la somme de leurs valeurs propres divisée par la trace. Cette valeur permet d'évaluer en quelque sorte la quantité d'information initiale recueillie par chaque axe.

Le % d'inertie (ou "variance" du nuage ou "dispersion") expliquée par un axe factoriel

$$\frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \times 100 \quad (3.22)$$

Elle mesure la part d'inertie expliquée par l'axe u_i

- Détermination des composantes principales

La matrice de variance-covariance est diagonalisée. Les lignes et colonnes de cette matrice correspondent à de nouvelles variables appelées composantes principales ;

A chaque axe est associée une composante principale.

La composante c^1 est le vecteur renfermant les coordonnées de projections des individus sur l'axe 1

La composante c^2 est le vecteur renfermant les coordonnées de projections des individus sur l'axe 2 ...

Pour obtenir ces coordonnées, on écrit que chaque composante principale est une combinaison linéaire des variables initiales de la forme :

Pour $i=1$

$$c^1 = x_1^1 u_1^1 + x_2^1 u_1^2 + \dots + x_p^1 u_1^n \quad (3.23)$$

Les composantes principales forment le vecteur :

$$c = \begin{pmatrix} c^1 \\ c^2 \\ \vdots \\ c^n \end{pmatrix} = \begin{pmatrix} x_1^1 x_2^1 \dots x_p^1 \\ x_1^2 x_2^2 \dots x_p^2 \\ \vdots \\ x_1^n x_2^n \dots x_p^n \end{pmatrix} \begin{pmatrix} u_1^1 u_2^1 \dots u_p^1 \\ u_1^2 u_2^2 \dots u_p^2 \\ \vdots \\ u_1^n u_2^n \dots u_p^n \end{pmatrix} \quad (3.24)$$

Tel que :

$$X_c = \begin{pmatrix} x_1^1 x_2^1 \dots x_p^1 \\ x_1^2 x_2^2 \dots x_p^2 \\ \vdots \\ x_1^n x_2^n \dots x_p^n \end{pmatrix} \text{ Est la matrice des données centrées}$$

ET

$$\Delta = \begin{pmatrix} u_1^1 u_2^1 \dots u_p^1 \\ u_1^2 u_2^2 \dots u_p^2 \\ \vdots \\ u_1^p u_2^p \dots u_p^p \end{pmatrix} \text{ Est la matrice des axes principaux}$$

On a:

$$c^1 = \begin{pmatrix} c_{11} \\ c_{12} \\ \cdot \\ \cdot \\ c_{1p} \end{pmatrix} \text{ et } c^2 = \begin{pmatrix} c_{21} \\ c_{22} \\ \cdot \\ \cdot \\ c_{2p} \end{pmatrix} \text{ et } \dots \dots \dots c^n = \begin{pmatrix} c_{n1} \\ c_{n2} \\ \cdot \\ \cdot \\ c_{np} \end{pmatrix} \quad (3.25)$$

Donc les variables des composantes principales forment une matrice $c = X \Delta$

La variance d'une composante principale est égale à l'inertie apportée par l'axe principale qui lui est associé

1^{er} composante c^1 variance : λ_1

2^{eme} composante c^2 variance : λ_2

3^{eme} composante c^3 variance : λ_3

Ces p composantes principales C^i s'expriment comme combinaison linéaire des anciennes variables. Elles sont telles que la somme des carrés de leurs corrélations avec les anciennes variables soit maximum.

Les composantes principales sont non corrélées deux à deux

En effet, les axes associés sont orthogonaux. Des graphes sont alors dessinés croisant les composantes principales, tant pour les variables que pour les individus, séparément car l'interprétation ne suit pas les mêmes principes. Les graphes des variables comprennent le cercle des corrélations.

3.2.5.4 Principe de construction de l'espace factoriel à partir du nuage des variables

Même règle que pour les individus, cela consiste à rechercher des axes orthogonaux qui permettent de représenter au mieux les variables, et à la fin on obtiendra des espaces factoriels identique à ceux obtenu par l'ajustement de l'ensemble des individus.

3.2.5.5 Analyse en composantes principales normée

Dans les paragraphes précédents, nous avons étudié une ACP simple, pour laquelle, non seulement tous les individus ont le même poids dans l'analyse, mais aussi, toutes les variables sont traitées de façon symétrique (on leur fait jouer le même

rôle) et les nouveaux axes sont issus de la matrice de variance-covariance empirique des variables (formule (3.13)). Cela pose parfois des problèmes. Le premier reproche fait par des praticiens est que, si les variables initiales sont hétérogènes, comme par exemple des poids, des tailles et des âges. Quel sens peut-on donner aux composantes principales qui sont alors des combinaisons linéaires de variables hétéroclites ? Le deuxième reproche, est que si on change d'unités sur ces variables, on peut changer complètement les résultats de l'ACP. Le dernier reproche vient du fait qu'une variable contribuera d'autant plus à la confection des premiers axes, que sa variance est forte.

Pour échapper à tous ces problèmes, on cherche à *normaliser* les variables et à travailler sur des variables *sans dimension*. Il y a plusieurs façons de normaliser les variables, mais la plus couramment utilisée est celle qui consiste à diviser les valeurs des variables par leur écart-type, c'est-à-dire que l'on travaille sur des variables *centrées et réduites*.

Centrer les données revient à travailler sur des variables sans unité, et la réduction permet d'accorder la même importance à chaque variable, ne pas réduire permet de donner plus d'importance aux variables qui ont une variabilité plus grande (une variance plus grande). C'est-à-dire l'opération de réduction diminue l'importance des variables les plus dispersées, elle peut inversement, augmenter l'importance des variables peu dispersées ! A la limite, une variable qui aurait à peu près la même valeur pour tous les individus (donc avec un écart-type faible) se verrait attribuer le même poids que les autres variables alors que le bon sens nous obligerait à l'éliminer.

3.2.5.6 Matrice des données centrées réduites

La matrice des données centrées réduites (sans dimensions) est alors :

$$Y = \begin{bmatrix} (x_{11} - x_{.1})/\sigma_1 & \cdots & (x_{1j} - x_{.j})/\sigma_j & \cdots & (x_{1p} - x_{.p})/\sigma_p \\ \vdots & & \vdots & & \vdots \\ (x_{i1} - x_{.1})/\sigma_1 & \cdots & (x_{ij} - x_{.j})/\sigma_j & \cdots & (x_{ip} - x_{.p})/\sigma_p \\ \vdots & & \vdots & & \vdots \\ (x_{n1} - x_{.1})/\sigma_1 & \cdots & (x_{nj} - x_{.j})/\sigma_j & \cdots & (x_{np} - x_{.p})/\sigma_p \end{bmatrix} \quad (3.26)$$

Si nous notons $D_{1/0}$ la matrice diagonale suivante :

$$D_{1/\sigma} = \begin{bmatrix} 1/\sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_p \end{bmatrix} \quad (3.27)$$

Nous avons alors :

$$Y = X_c D_{1/\sigma} \quad (3.28)$$

❖ Remarque :

Dans le cas où les variables sont centrées réduites, la moyenne de la variable y_i est nulle et donc sa variance est alors 1 (ce qui revient à dire que la norme de la variable centrée réduite est de norme unitaire).

3.2.5.7 Matrice de corrélation

La matrice des corrélations des variables initiales permet de voir si l'analyse peut être globalement possible. En effet, une des conditions de mise en œuvre est que les variables initiales soient corrélées, sinon, le nombre de dimensions ne peut être réduit. On ne peut résumer que ce qui est partiellement redondant et lié. Il faut donc qu'un maximum de valeurs de la matrice des corrélations soit proche de 1 en valeur absolue.

Pour avoir la matrice de corrélation nous allons pré-multiplier la matrice des données centrées réduite par sa transposée ; on aura alors :

$$R = \frac{1}{n} Y_c^t Y_c = \frac{1}{n} \begin{bmatrix} 1 & cor(x_1, x_2) \cdots cor(x_1, x_2) \\ \vdots & \vdots \\ cor(x_2, x_1) & 1 \cdots cor(x_1, x_2) \\ \vdots & \vdots \\ cor(x_n, x_1) & cor(x_n, x_2) \cdots 1 \end{bmatrix} \quad (3.29)$$

Avec :

$$cor(x_i, x_j) = cor(x_j, x_i) = \frac{cov(x_i, x_j)}{\sigma_i \sigma_j} \quad (3.30)$$

Puisque les variables sont centrées, la variance de chaque variable vaut 1. Donc l'inertie totale est alors égale à p le nombre de variables.

Rappelons que les matrices des covariances et de corrélations sont toutes deux des matrices carrées symétriques.

Représenter le nuage de point des données centrées réduites ne modifie rien à la forme de celui-ci. En effet, la différence entre les deux n'est qu'un changement d'échelle.

L'information intéressante pour les individus est la distance entre les points. En effet, plus cette distance sera grande entre deux individus X_i et X_{iF} . Plus les deux individus ne seront différents et mieux on pourra les caractériser. Mais, il faut d'abord choisir une distance. Nous prenons une autre distance euclidienne que la distance euclidienne classique, la distance choisie est alors ;

$$d^2(e_i, e_{i'}) = \sum_{j=1}^p \frac{1}{\sigma_j^2} (x_{ij} - x_{i'j})^2 \quad (3.31)$$

Cette nouvelle distance ne traite plus les variables de façon symétrique, mais elle permet de faire jouer un rôle équitable à chacune d'entre elles. (MILLOUDI, 2014)

Conclusion

Après avoir exposé le principe de l'ACP, nous allons passer au prochain chapitre. Nous appliquons l'ACP sur des paramètres climatiques de la station météo de la région de Laghouat.

Pour pouvoir traiter les données, nous faisons appel au logiciel R.

Chapitre 4

4 Implémentation

4.1 Introduction

Nous présentons dans cette partie, l'analyse des données climatiques dans la région de Laghouat qui est le sujet de notre étude.

Premièrement, nous présentons les données et nous commentons les résultats de L'ACP et leurs intégrité et compatibilité.

En deuxième lieu, nous expliquons la phase analyse descriptive des donnée et après, nous terminons par la phase d'interprétation.

4.2 Collecte d'information

En effet, le premier palier est de récupérer un échantillon constant sur une période assez significative. Pour atteindre cet objectif nous nous somme déplacé à la station météo de Laghouat.

Les données récupérées s'étendent sur une période de 10 ans. Nous avons pu récupérer d'autres données depuis le département de biologie et celui d'agronomie (à l'université de Laghouat).

Néanmoins, nous avons contacté qu'il existe des années sur lesquels les données sont manquante sur des parties assez conséquentes. Pour surmonter ce problème, nous avons remplacé ce vide par les données statistiques en sorte que l'analyse ne soit pas altérée.

Les données sont présentées en annexe et sont divisées en données mensuelles et annuelles.

4.3 Analyse Factorielle

L'analyse factorielle est un terme qui désigne plusieurs méthodes d'analyse de grands tableaux de données visant à déterminer des facteurs corrélés.

4.3.1 Analyse 2^k

Nous voulions à travers l'analyse 2^k étudier les variables les plus significative pour une variable d'étude fixée à l'avance en se basant sur l'analyse des extrémités. Néanmoins, les donnée collectées ne sont pas complètes afin d'être adaptées à nous besoin .en effet, nous avons fixé en premier lieu le but d'analyse le comportement de sol (qualité) para port au changement climatiques. Cette données n'est pas encours disponible sur les sources d'information consultées. Pour cette raison qu'on s'est contentes de l'analyse ACP et K-means.

4.3.1.1 Résultat mensuelle

	TEMPERATURE MOYENNE	TEMPERATURE MIN	TEMPERATURE MAX	HUMIDITE_%	ISOLATION TOTAL	PRECIPITATIONS	VENT
TEMPERATURE MOYENNE	1	0,95502	0,96292	-0,75255	0,55677	0,03285	-0,10378
TEMPERATURE MIN	0,95502	1	0,90569	-0,72935	0,56084	0,05237	-0,09648
TEMPERATURE MAX	0,96292	0,90569	1	-0,70754	0,54075	0,05903	-0,12674
HUMIDITE_%	-0,75255	-0,72935	-0,70754	1	-0,57983	0,24253	0,08
ISOLATION TOTAL	0,55677	0,56084	0,54075	-0,57983	1	-0,24382	0,00938
PRECIPITATIONS	0,03285	0,05237	0,05903	0,24253	-0,24382	1	-0,08691
VENT	-0,10378	-0,09648	-0,12674	0,08	0,00938	-0,08691	1

Tableau 4.1 : Matrices de corrélation

Axis	Eigen value	Proportion (%)	Cumulative (%)
1	3,95768	56,54%	56,54%
2	1,231335	17,59%	74,13%
3	0,935409	13,36%	87,49%
4	0,484498	6,92%	94,41%
5	0,276458	3,95%	98,36%
6	0,092995	1,33%	99,69%
7	0,021625	0,31%	100,00%

Tableau 4.2 : Significance de Principal Components

Attribut	Axis_1		Axis_2		Axis_3	
	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)
Température MOYENNE	0,96501	93 % (93 %)	0,13835	2 % (95 %)	0,08964	1 % (96 %)
Température MIN	0,94631	90 % (90 %)	0,14403	2 % (92 %)	0,10268	1 % (93 %)
Température MAX	0,94063	88 % (88 %)	0,17326	3 % (91 %)	0,07942	1 % (92 %)
Humidité_%	-0,8521	73 % (73 %)	0,20298	4 % (77 %)	0,10274	1 % (78 %)
Isolation total	0,70618	50 % (50 %)	-0,35705	13 % (63 %)	-0,04754	0 % (63 %)
Précipitations	-0,07821	1 % (1 %)	0,86882	75 % (76 %)	0,39325	15 % (92 %)
Vent	-0,12355	2 % (2 %)	-0,48775	24 % (25 %)	0,86201	74 % (100 %)
Var. Expl.	3,95768	57 % (57 %)	1,23134	18 % (74 %)	0,93541	13 % (87 %)

Tableau 4.3 : Factor Loadings [Communality Estimates]

4.3.1.2 Résultat Annuel

	ISOLATION_TOT OTAL	PRECIPITATIONS	TEMPERATURE MIN	TEMPERATURE MAX	HUMIDITE _%	VENT	TEMPERATURE MOYENNE
ISOLATION_TOT AL	1	-0,8746	0,64734	-0,56261	-0,1619	0,06578	-0,05061
PRECIPITATIONS	-0,8746	1	-0,64238	0,28653	0,30044	0,12934	-0,17795
TEMPERATURE_ MIN	0,64734	-0,64238	1	-0,5221	0,06426	0,10505	0,48394
TEMPERATURE_ MAX	-0,56261	0,28653	-0,5221	1	-0,13906	-0,17162	0,35836
HUMIDITE_%	-0,1619	0,30044	0,06426	-0,13906	1	0,31884	0,05024
VENT	0,06578	0,12934	0,10505	-0,17162	0,31884	1	0,02638
TEMPERATURE_ MOYENNE	-0,05061	-0,17795	0,48394	0,35836	0,05024	0,02638	1

Tableau 4.4 : Matrices de corrélation (annuelle)

Axis	Eigen value	Proportion (%)	Cumulative (%)
1	2.828475	40,41 %	40,41%
2	1,521952	21,74 %	62,15%
3	1,382410	19,83 %	81,98%
4	0,704724	10,07 %	92,01%
5	0,440969	6,30 %	98,35%
6	0,062670	0,90 %	99,25%
7	0,052800	0,75%	100,00%

Tableau 4.5 : Significance de Principal Components

Attribute	Axis_1		Axis_2		Axis_2	
	Corr.	% (Tot. %)	Corr.	% (Tot. %)	Corr.	% (Tot. %)
-						
Isolation total	-0,93004	86 % (86 %)	0,00428	0 % (86 %)	-0,19587	4 % (90 %)
Précipitations	0,87991	77 % (77 %)	-0,32797	11 % (88 %)	0,02122	0 % (88 %)
Température MIN	-0,86054	74 % (74 %)	-0,07658	1 % (75 %)	0,38796	15 % (90 %)
Température MAX	0,63366	40 % (40 %)	0,52249	27 % (67 %)	0,39453	16 % (83 %)
Humidité %	0,13377	2 % (2 %)	-0,74227	55 % (57 %)	0,34935	12 % (69 %)
Vent	-0,05912	0 % (0 %)	-0,70661	50 % (50 %)	0,25414	6 % (57 %)
Température MOYENNE	-0,16064	3 % (3 %)	0,29197	9 % (11 %)	0,92563	86 % (97 %)
Var. Expl.	2,82847	40 % (40 %)	1,52195	22 % (62 %)	1,38841	20 % (82 %)

Tableau 4.6 : Factor Loadings [Communality Estimates]

4.4 Interpretation des resultants ACP

4.4.1.1 Interprétation des Résultats mensuelle

Il ressort de (Tableau 4.9) que l'ensemble des valeurs de corrélation est relativement moyenne. En effet, plus de 55 % des coefficients de corrélation sont supérieurs à 0,50. Il s'en suit que les variables analysées sont corrélées entre elles. (Température min –température max- température moy) constatées comme Les fortes corrélations, (Température min –température max- température moy -humidité) comme moyenne corrélation, Température min –température max- température moy- isolation total) comme faible corrélation.

Les tableaux (Tableau 4.10) et (Tableau 4.11) présentent respectivement les valeurs propres de chaque facteur aux stations de Laghouat. Seules les composantes principales dont les valeurs propres sont supérieures ou égales à 1 ont été considérées comme significatives. Seulement les deux (2) premières composantes remplissent cette condition. Les résultats montrent que les variances exprimées par les deux premiers facteurs sont de 74.13 %. En considérant les deux premières composantes principales, les variances exprimées sont de 74,13 % (près de 75 %) soit plus de la moitié des variances exprimées. A la lumière de ces résultats, seul ce cas va être présenté à partir des deux premières composantes principales. En effet, 75 % représente la valeur de référence acceptable dans le cadre d'une ACPN. Les deux premiers facteurs contiennent donc l'essentiel des informations relatives aux matrices des données.

L'analyse dans l'espace des variables a été réalisée dans les plans factoriels C1-C2 (Figure 4.1) et qui rendent compte de manière satisfaisante de la structure des données.. Un premier regroupement se situant dans les valeurs positives comprend la température moyenne mensuelle (T_{moy}), la température maximale mensuelle (T_{max}), la température minimale mensuelle (T_{min}), l'humidité (AH) avec les mois juin et juillet. Un deuxième regroupement se situant dans les valeurs négatives comprend l'isolation totale avec le mois de mai. Et mois d'avril avec précipitation. Température min /max/moy est en inversement proportionnelle avec l'humidité et l'isolation total avec précipitation presque en inversement proportionnelle.

4.4.1.2 Interprétation des Résultats Annuelle

Il ressort de ce tableau (Tableau 4.12) que l'ensemble des valeurs de corrélation est relativement faible. En effet, plus de 77 % des coefficients de corrélation sont inférieurs à 0,50. Il s'en suit que les variables analysées ne sont pas fortement corrélées entre elles. Ce qui réduit significativement la redondance de l'information et justifie du même coup la pertinence dans le choix de ces variables pour mener l'étude.

Le (Tableau 4.13) présente respectivement les valeurs propres de chaque facteur aux stations de Laghouat. Seules les composantes principales dont les valeurs propres sont supérieures ou égales à 1 ont été considérées comme significatives. Seulement les trois (3) premières composantes remplissent cette condition. Les résultats montrent que les variances exprimées par les trois premiers facteurs sont de 81.98 %. En considérant les deux premières composantes principales, les variances exprimées sont de 62.15 % (loin de 75 %). A la lumière de ces résultats, Nous remarquons qu'il y a une relation entre les températures et précipitations.

L'analyse dans l'espace des variables a été réalisée dans les plans factoriels C1-C2 (Figure 4.4) et qui rendent compte de manière satisfaisante de la structure des données. Un premier regroupement se situant dans les valeurs positives les années 2009 -2012-2013-2014-2015-2016-2017-2018. Un deuxième regroupement se situant dans les valeurs négatives comprend les années 2010 2011.

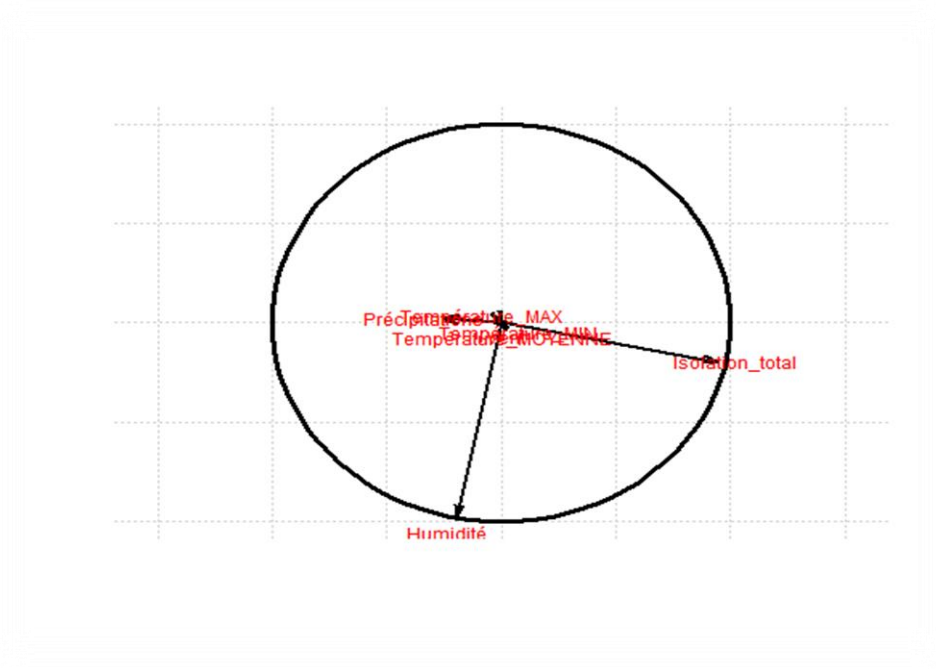


Figure 4-3 : Espaces des variables des plans factoriels

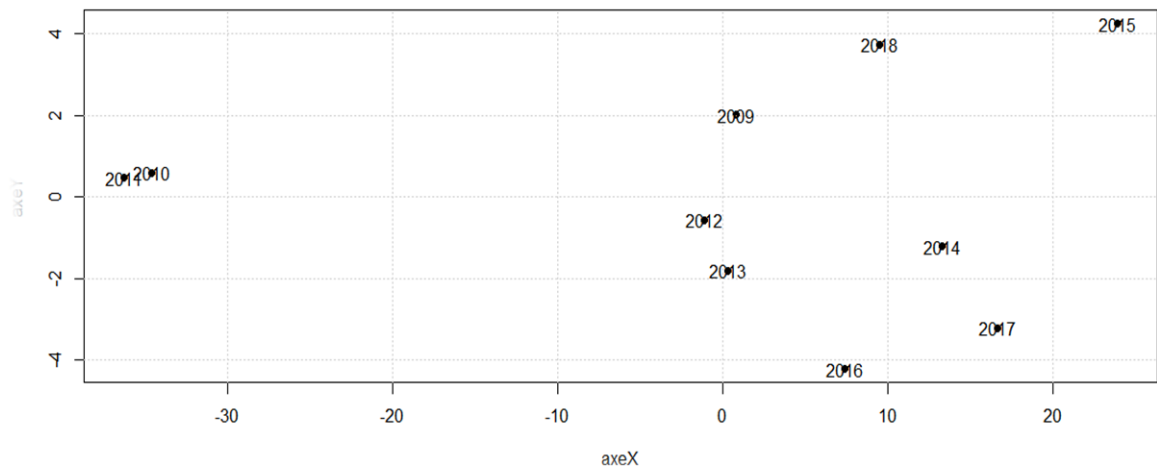


Figure 4-4: présentation des variables Annuelle

5 Conclusion générale

Cette étude a permis de caractériser les principales de la variabilité climatique dans la région de Laghouat. L'humidité relative mensuelle varie entre 49 % (janvier 2009) et 30 % (août 2009) et entre 56 % (janvier 2018) et 20 % (août 2018) avec une moyenne de 5 % en 2009 et 42% en 2018. Les températures moyennes mensuelles varient entre 31, ° C (août 2009) et 09° C (février 2009) et entre 31.6, ° C (août 2018) et 06° C (février 2018) avec une moyenne de 18.9°C en 2009 et 18.6°C en 2018. Les variations interannuelles de la température de l'air montrent que la température de l'air connaît une hausse régulière. La variabilité climatique se manifeste également par une dynamique spatio-temporelle des pluies annuelles, les régimes pluviométriques mensuels varient entre 10, mm (août 2009) et 24mm (février 2009) et entre 00.8, mm (août 2018) et 10mm (février 2018) avec une moyenne de 14.1mm en 2009 et 14.4mm en 2018. Nous avons proposé sept variables pour l'analyse en composantes principales normées qui a permis d'analyser les facteurs de variabilité temporelle des régimes climatiques saisonniers dans Laghouat. Ces variables se rapportent aux paramètres atmosphériques (température de l'air et humidité de l'air) et aux précipitations (fréquence et hauteurs de précipitations). L'analyse en composantes principales normées a permis d'extraire quatre composantes principales dont les trois premières ont été utilisées pour l'interprétation des données. Les séries chronologiques constituées à partir de ces sept variables s'étendent sur la période 2009-2018. L'analyse en composantes principales normées basées sur ces différentes variables, a permis de mettre en évidence l'impact de la variabilité climatique. La première composante principale (C1) est associée aux facteurs définissant les températures, humidité, précipitations. La deuxième et les troisièmes composantes principales (C2) sont associées aux variables liées à l'isolation totale. Quant aux facteurs qui influencent la variabilité temporelle des régimes pluviométriques saisonniers, il est apparu que les variables pluviométriques sont influencées par la température de l'air ainsi que l'humidité de l'air. On conclut donc que la température de l'air et l'humidité de l'air sont des facteurs de la variabilité temporelle des régimes pluviométriques saisonniers dans Laghouat.

Bibliographie

- (BENYETTOU & BOUKLIKHA, 2016) BENYETTOU, M. A., & BOUKLIKHA, A. (2016). *Variations et tendances des températures et des précipitations journalières en Algérie*. tlemcene: Univ ABOU BEKR BELKAID.
- (Hammar, Djebbar, & Khoualdia, 2014) 2014 *Characterization of climate variability* Annaba Université Badji Mokhtar Annaba
- (CHIBANE, BENTCHAKAL, MEDJERAB, & BENFARES, 2015) CHIBANE, B., BENTCHAKAL, M., MEDJERAB, A., & BENFARES, B. (2015). ETUDE DE LA VARIABILITE ET DE LA STRUCTURE DE PLUIE ANNUELLE DANS UNE REGION SEMI-ARIDE. *Larhyss Journal* .
- (MEDDI, SOUAG, TAIBI, & GIL, 2013) MEDDI, M., SOUAG, D., TAIBI, S., & GIL, M. (2013). *Evolution et régionalisation des précipitations au nord de l'Algérie (1936–2009)*. alger: Université des Sciences et Technologie Houari Boumediene.
- (CHIBANE, BENTCHAKAL, MEDJERAB, & BENFARES, 2015) Medjerab, A., & Henia, L. (2005). Régionalisation des pluies annuelles dans l'Algérie nord-occidentale. *journals openedition* .
- (Météo France, 2008) Météo France. (2008). *Glossaire*. Consulté le 06 04, 2019, sur Météo France: <http://www.meteofrance.fr>
- (MILOUDI, 2014) MILOUDI, N. (2014). *CONTRIBUTION A L'ETUDE DE LA VULNERABILITE*. TIZI OUZOU: UNIVERSITE MOULOUD MAMMERI.
- (mounia, 2017) 2017 *régionalisation et variabilité pluviométrique dans le nord Centre-ouest algérien* khemise miliana univ Djilali bounaama
- (wikipedia, 2019) wikipedia. (2019, 05 14). *Réchauffement climatique*. Consulté le 06 05, 2019, sur wikipedia: <https://fr.wikipedia.org>

Annexe

- **Code source ACP en R :**

```
data = read.table("data.txt",sep="\t",header=T)
colnames(data)
rownames(data) = data[,1] # On donne à chaque ligne le nom de son cancer (var 1, var 2)
x = data[,-1] ;x # On exclue la colonne 1 qui contenait les noms des Facteur.
library(ade4)
acp2 <- dudi.pca(x ,scannf= F,scale=FALSE,,nf=3)
acp2
# Fonction partto
partto = function(x, bar.col="cyan", line.col="red", pch=16, h=80,
h.lty=3,main="",xlab="Défauts",ylab="Fréquence (%)", names.arg=c(), ylab2="Cumul",mar=c(5,4,3,4)) {
  if (length(names.arg)>0) {names.arg=names.arg[order(x, decreasing = TRUE)]}
  x = sort(x,decreasing=T); x = x*100/sum(x);
  cumul = (cumsum(x)/sum(x))*100
  simulation = barplot(x,col=bar.col) ; simulation
  plot.new()
  par(mar=mar)
  barplot(x,col=bar.col,axes=F,ylim=c(0,100),main=main,xlab=xlab,ylab="",names.arg=names.arg)
  #par(new=TRUE)
  points(simulation,cumul,pch=pch,col=line.col,xlab="",ylab="",type="o")
  abline(h=h,lty=h.lty) ; box()
  axis(2) ; axis(4,c(0,20,40,60,80,100),col.axis=line.col,col=line.col)
  mtext(ylab,side=2,line=2,cex=1.2) ; mtext(ylab2,side=4,col="red",line=2,cex=1.2)
  result = c(x , cumul) ; result = matrix(result,nc=length(x), byrow=T)
  if (length(names.arg)>0) {colnames(result) = names.arg }
  rownames(result) = c("frequency","cumul")
  return(result)}
```

```

partto(acp2$eig, h=80) # h = 95 : afficher un seuil de cumul de la variance à 95%
axeX<- acp2$li[,1] ; axeY<- acp2$li[,2]
plot(axeX,axeY,pch=16);grid()
text(axeX,axeY,rownames(x))

# Récupérer la description donnant la corrélation de chaque variable par composante
poids_var_x<- acp2$c1[,1]
poids_var_y<- acp2$c1[,2]
vecnames<- rownames(acp2$c1)
# Tracer un fond blanc orthonormé
plot(c(-1,1),c(-1,1),type="n",xlab="",ylab="",asp=1,axes=F) ; grid()
# Tracer le cercle
library("plotrix") ;draw.circle(0,0,1,lwd=3)
# Tracer les flèches
arrows(0,0,poids_var_x,poids_var_y,length=0.1,angle=17,lwd=2)
# Calculer la position du nom de chaque vecteur // aux pointes des flèches
poids_x_bin<- poids_var_x ; poids_x_bin[poids_x_bin<0]<- -1 ; poids_x_bin[poids_x_bin>0]<- 1
poids_y_bin<- poids_var_y ; poids_y_bin[poids_y_bin<0]<- -1 ; poids_y_bin[poids_y_bin>0]<- 1
poids_z<- poids_x + poids_y
poids_x<- poids_x / poids_z * poids_x_bin * 0.08 ; poids_y<- poids_y / poids_z * poids_y_bin * 0.08
# Tracer les noms des vecteurs
text(poids_var_x+poids_x,poids_var_y+poids_y,vecnames,cex=0.75,col="red")

```

• Présentation des données mensuelle

Les tableaux suivants contiennent les données météo collectées de façon mensuelle (précipitation, température max, température min, température moyenne, humidité, insolation, vent).

• Précipitation

date	Précipitations	date	Précipitations	date	Précipitations	date	Précipitations	date	Précipitations
janv-09	24.5	janv-11	06.3	janv-13	08.2	janv-15	04.6	janv-17	10.4
févr-09	13.7	févr-11	00.4	févr-13	12.6	févr-15	18.0	févr-17	0.8
mars-09	14.1	mars-11	50.6	mars-13	02.6	mars-15	05.3	mars-17	00.0
avr_-09	1.8	avr_-11	42.5	avr_-13	07.6	avr_-15	02.1	avr_-17	08.0
mai_-09	4.7	mai_-11	22.9	mai_-13	09.2	mai_-15	0.3	mai_-17	05.8
juin-09	2.7	juin-11	29.9	juin-13	00.0	juin-15	04.8	juin-17	13.4
juil-09	3.1	juil-11	08.1	juil-13	05.6	juil-15	00.0	juil-17	02.6
août-09	10.1	août-11	06.6	août-13	08.8	août-15	15.4	août-17	00.8
sept-09	79.0	sept-11	24.5	sept-13	12.2	sept-15	36.8	sept-17	20.2
oct_-09	00.2	oct_-11	55.1	oct_-13	27.4	oct_-15	38.4	oct_-17	03.4
nov_-09	13.3	nov_-11	34.0	nov_-13	02.6	nov_-15	4.6	nov_-17	01.4
déc_-09	05.8	déc_-11	04.6	déc_-13	41.6	déc_-15	0	déc_-17	00.0
janv-10	06.3	janv-12	4.0	janv-14	03.7	janv-16	0.2	janv-18	07.8
févr-10	00.4	févr-12	1.4	févr-14	02.2	févr-16	0.6	févr-18	08.5
mars-10	50.6	mars-12	25.3	mars-14	00.4	mars-16	1.6	mars-18	02.5
avr_-10	42.5	avr_-12	5.3	avr_-14	00.2	avr_-16	5.8	avr_-18	07.9
mai_-10	22.9	mai_-12	13.3	mai_-14	24.2	mai_-16	1.8	mai_-18	22.6
juin-10	29.6	juin-12	13.3	juin-14	14.6	juin-16	0.8	juin-18	01.0
juil-10	08.1	juil-12	5.9	juil-14	03.8	juil-16	5.8	juil-18	02.3
août-10	06.6	août-12	0.4	août-14	03.2	août-16	13.8	août-18	72.1
sept-10	24.5	sept-12	18.2	sept-14	20.7	sept-16	28.2	sept-18	10.2
oct_-10	55.1	oct_-12	40.8	oct_-14	03.0	oct_-16	0.8	oct_-18	24.4
nov_-10	34.0	nov_-12	32.6	nov_-14	23.5	nov_-16	10.4	nov_-18	09.1

Tableau 0.1 :donnee Précipitation montielle (mm)

- Température MAX

date	Temp_MAX	date	Temp_MAX	date	Temp_MAX	date	Temp_MAX	date	Temp_MAX
janv-09	13.0	janv-11	16.4	janv-13	14.5	janv-15	13.0	janv-17	12.4
févr-09	15.8	févr-11	15.8	févr-13	14.2	févr-15	11.9	févr-17	17.6
mars-09	20.6	mars-11	18.1	mars-13	20.2	mars-15	19.0	mars-17	21.5
avr_-09	21.2	avr_-11	26.0	avr_-13	23.4	avr_-15	26.5	avr_-17	24.5
mai_-09	30.4	mai_-11	28.2	mai_-13	27.6	mai_-15	21.2	mai_-17	31.7
juin-09	36.5	juin-11	32.5	juin-13	33.3	juin-15	33.3	juin-17	34.6
juil-09	40.4	juil-11	39.1	juil-13	37.9	juil-15	37.7	juil-17	38.1
août-09	38.8	août-11	38.7	août-13	35.7	août-15	36.5	août-17	38.6
sept-09	29.0	sept-11	34.8	sept-13	31.6	sept-15	31.7	sept-17	31.4
oct_-09	27.1	oct_-11	24.1	oct_-13	29.1	oct_-15	28.0	oct_-17	24.8
nov_-09	21.7	nov_-11	18.5	nov_-13	17.1	nov_-15	18.2	nov_-17	12.1
déc_-09	19.1	déc_-11	14.5	déc_-13	12.3	déc_-15	15.7	déc_-17	13.4
janv-10	16.5	janv-12	14.5	janv-14	14.2	janv-16	17.1	janv-18	15.1
févr-10	18.4	févr-12	12.5	févr-14	16.6	févr-16	17.1	févr-18	13.8
mars-10	21.8	mars-12	19.5	mars-14	17.2	mars-16	19.1	mars-18	19.3
avr_-10	26.0	avr_-12	23.3	avr_-14	25.7	avr_-16	25.9	avr_-18	22.7
mai_-10	27.7	mai_-12	31.2	mai_-14	29.5	mai_-16	29.7	mai_-18	25.8
juin-10	35.9	juin-12	38.7	juin-14	32.8	juin-16	35.1	juin-18	33.5
juil-10	36.9	juil-12	40.6	juil-14	38.5	juil-16	38.1	juil-18	41.0
août-10	39.0	août-12	36.9	août-14	38.5	août-16	36.2	août-18	33.2
sept-10	31.9	sept-12	32.7	sept-14	33.1	sept-16	31.2	sept-18	11.1
oct_-10	29.0	oct_-12	27.0	oct_-14	27.2	oct_-16	27.2	oct_-18	22.7
nov_-10	19.5	nov_-12	19.2	nov_-14	18.6	nov_-16	18.5	nov_-18	17.8
déc_-10	17.1	déc_-12	16.4	déc_-14	12.7	déc_-16	13.2	déc_-18	17.3

Tableau 0.2: donnée température mensuelle (°C)

- Température MIN

date	Temp_MIN	date	Temp_MIN	date	Temp_MIN	date	Temp_MIN	date	Temp_MIN
janv-09	03.6	janv-11	01.8	janv-13	03.9	janv-15	02.2	janv-17	01.7
févr-09	02.1	févr-11	02.3	févr-13	02.8	févr-15	03.6	févr-17	05.8
mars-09	06.3	mars-11	06.0	mars-13	08.6	mars-15	07.0	mars-17	08.4
avr_-09	04.2	avr_-11	12.0	avr_-13	11.3	avr_-15	13.4	avr_-17	11.8
mai_-09	14.6	mai_-11	15.0	mai_-13	14.2	mai_-15	17.4	mai_-17	19.5
juin-09	20.9	juin-11	19.7	juin-13	19.4	juin-15	19.8	juin-17	21.5
juil-09	24.2	juil-11	23.9	juil-13	24.2	juil-15	24.3	juil-17	24.7
août-09	23.5	août-11	23.7	août-13	21.9	août-15	23.4	août-17	25.4
sept-09	16.8	sept-11	02.2	sept-13	26.4	sept-15	19.9	sept-17	18.5
oct_-09	11.6	oct_-11	11.5	oct_-13	25.8	oct_-15	14.1	oct_-17	12.2
nov_-09	06.2	nov_-11	06.7	nov_-13	07.4	nov_-15	08.0	nov_-17	05.7
déc_-09	04.1	déc_-11	02.7	déc_-13	02.7	déc_-15	03.3	déc_-17	02.8
janv-10	04.0	janv-12	00.7	janv-14	04.2	janv-16	04.4	janv-18	03.9
févr-10	06.1	févr-12	-0.1	févr-14	04.8	févr-16	05.4	févr-18	03.4
mars-10	06.9	mars-12	05.8	mars-14	06.4	mars-16	06.5	mars-18	08.8
avr_-10	11.1	avr_-12	09.6	avr_-14	13.3	avr_-16	12.8	avr_-18	11.0
mai_-10	13.5	mai_-12	15.3	mai_-14	16.2	mai_-16	16.1	mai_-18	14.4
juin-10	19.5	juin-12	22.7	juin-14	19.4	juin-16	21.4	juin-18	20.5
juil-10	23.5	juil-12	25.2	juil-14	24.6	juil-16	24.3	juil-18	27.2
août-10	23.6	août-12	24.1	août-14	24.9	août-16	23.5	août-18	21.6
sept-10	18.3	sept-12	18.1	sept-14	21.2	sept-16	19.1	sept-18	20.3
oct_-10	12.1	oct_-12	13.6	oct_-14	13.8	oct_-16	15.1	oct_-18	22.7
nov_-10	07.4	nov_-12	07.9	nov_-14	09.1	nov_-16	07.8	nov_-18	07.6
déc_-10	04.0	déc_-12	02.8	déc_-14	03.3	déc_-16	05.1	déc_-18	03.7

Tableau 0.3 : donnée température min mensuelle (°C)

- Température MOYENNE

date	Temp_MOY	date	Temp_MOY	date	Temp_MOY	date	Temp_MOY	date	Temp_MOY
janv-09	08.1	janv-11	08.7	janv-13	10.3	janv-15	07.5	janv-17	06.6
févr-09	09.0	févr-11	09.1	févr-13	09.6	févr-15	07.8	févr-17	11.2
mars-09	13.6	mars-11	12.1	mars-13	15.5	mars-15	13.0	mars-17	14.2
avr_-09	14.2	avr_-11	19.0	avr_-13	18.5	avr_-15	19.9	avr_-17	18.0
mai_-09	22.7	mai_-11	21.6	mai_-13	22.5	mai_-15	24.7	mai_-17	25.4
juin-09	29.3	juin-11	25.9	juin-13	28.3	juin-15	26.7	juin-17	28.1
juil-09	32.7	juil-11	31.6	juil-13	32.6	juil-15	31.6	juil-17	31.6
août-09	31.3	août-11	31.3	août-13	30.4	août-15	29.9	août-17	31.6
sept-09	22.6	sept-11	27.1	sept-13	26.8	sept-15	25.8	sept-17	24.9
oct_-09	19.6	oct_-11	17.8	oct_-13	23.7	oct_-15	19.5	oct_-17	18.5
nov_-09	13.4	nov_-11	12.6	nov_-13	13.2	nov_-15	13.1	nov_-17	08.9
déc_-09	10.8	déc_-11	08.3	déc_-13	08.3	déc_-15	09.5	déc_-17	08.1
janv-10	10.0	janv-12	07.3	janv-14	09.1	janv-16	10.7	janv-18	09.6
févr-10	12.2	févr-12	06.0	févr-14	10.7	févr-16	11.2	févr-18	08.6
mars-10	14.6	mars-12	12.6	mars-14	11.9	mars-16	12.8	mars-18	14.0
avr_-10	18.6	avr_-12	16.4	avr_-14	19.3	avr_-16	19.3	avr_-18	16.7
mai_-10	20.4	mai_-12	23.7	mai_-14	23.1	mai_-16	22.9	mai_-18	20.1
juin-10	27.9	juin-12	30.7	juin-14	26.5	juin-16	28.2	juin-18	27.3
juil-10	32.8	juil-12	33.2	juil-14	31.8	juil-16	31.2	juil-18	34.3
août-10	31.1	août-12	32.4	août-14	31.8	août-16	29.8	août-18	27.3
sept-10	24.9	sept-12	25.4	sept-14	27.1	sept-16	25.2	sept-18	25.8
oct_-10	18.7	oct_-12	20.1	oct_-14	20.6	oct_-16	21.2	oct_-18	17.4
nov_-10	13.2	nov_-12	13.2	nov_-14	13.6	nov_-16	13.2	nov_-18	12.4
déc_-10	10.2	déc_-12	09.2	déc_-14	07.9	déc_-16	09.3	déc_-18	10.3

Tableau 0.4: donnée température moyenne mensuelle (mm)

- Humidité

date	Humidité	date	Humidité	date	Humidité	date	Humidité	date	Humidité
janv-09	71	janv-11	59	janv-13	50	janv-15	53	janv-17	56
févr-09	59	févr-11	52	févr-13	44	févr-15	52	févr-17	44
mars-09	56	mars-11	55	mars-13	40	mars-15	35	mars-17	32
avr_-09	53	avr_-11	46	avr_-13	33	avr_-15	27	avr_-17	33
mai_-09	42	mai_-11	43	mai_-13	31	mai_-15	21	mai_-17	26
juin-09	37	juin-11	43	juin-13	21	juin-15	27	juin-17	23
juil-09	29	juil-11	28	juil-13	21	juil-15	16	juil-17	17
août-09	30	août-11	29	août-13	23	août-15	30	août-17	20
sept-09	59	sept-11	40	sept-13	33	sept-15	32	sept-17	29
oct_-09	51	oct_-11	61	oct_-13	37	oct_-15	50	oct_-17	41
nov_-09	52	nov_-11	68	nov_-13	51	nov_-15	59	nov_-17	39
déc_-09	61	déc_-11	74	déc_-13	65	déc_-15	56	déc_-17	53
janv-10	60	janv-12	65	janv-14	69	janv-16	46	janv-18	49
févr-10	61	févr-12	58	févr-14	59	févr-16	36	févr-18	48
mars-10	50	mars-12	54	mars-14	52	mars-16	36	mars-18	42
avr_-10	48	avr_-12	40	avr_-14	36	avr_-16	32	avr_-18	40
mai_-10	41	mai_-12	38	mai_-14	36	mai_-16	24	mai_-18	37
juin-10	33	juin-12	49	juin-14	35	juin-16	19	juin-18	24
juil-10	32	juil-12	25	juil-14	26	juil-16	17	juil-18	18
août-10	35	août-12	25	août-14	29	août-16	24	août-18	37
sept-10	45	sept-12	40	sept-14	41	sept-16	26	sept-18	43
oct_-10	49	oct_-12	55	oct_-14	45	oct_-16	44	oct_-18	52
nov_-10	57	nov_-12	73	nov_-14	63	nov_-16	16	nov_-18	56
déc_-10	49	déc_-12	67	déc_-14	69	déc_-16	27	déc_-18	61

Tableau 0.5: donnée humidité mensuelle en (%)

- Isolation Total

date	Isolation	date	Isolation	date	Isolation	date	Isolation	date	Isolation
janv-09	164	janv-11	197	janv-13	255	janv-15	255	janv-17	229
févr-09	217	févr-11	205	févr-13	255	févr-15	204	févr-17	222
mars-09	244	mars-11	195	mars-13	255	mars-15	293	mars-17	297
avr_-09	279	avr_-11	221	avr_-13	255	avr_-15	299	avr_-17	287
mai_-09	330	mai_-11	241	mai_-13	255	mai_-15	327	mai_-17	317
juin-09	296	juin-11	272	juin-13	255	juin-15	323	juin-17	333
juil-09	272	juil-11	284	juil-13	255	juil-15	379	juil-17	363
août-09	268	août-11	242	août-13	255	août-15	256	août-17	291
sept-09	205	sept-11	175	sept-13	255	sept-15	258	sept-17	301
oct_-09	279	oct_-11	226	oct_-13	255	oct_-15	270	oct_-17	131
nov_-09	225	nov_-11	196	nov_-13	255	nov_-15	259	nov_-17	245
déc_-09	217	déc_-11	215	déc_-13	255	déc_-15	239	déc_-17	225
janv-10	200	janv-12	219	janv-14	217	janv-16	229	janv-18	232
févr-10	168	févr-12	223	févr-14	234	févr-16	217	févr-18	212
mars-10	188	mars-12	188	mars-14	269	mars-16	126	mars-18	310
avr_-10	221	avr_-12	289	avr_-14	307	avr_-16	259	avr_-18	234
mai_-10	283	mai_-12	327	mai_-14	324	mai_-16	299	mai_-18	210
juin-10	312	juin-12	282	juin-14	271	juin-16	327	juin-18	353
juil-10	283	juil-12	304	juil-14	307	juil-16	381	juil-18	277
août-10	229	août-12	293	août-14	406	août-16	312	août-18	313
sept-10	211	sept-12	258	sept-14	212	sept-16	252	sept-18	234
oct_-10	222	oct_-12	245	oct_-14	288	oct_-16	275	oct_-18	245
nov_-10	190	nov_-12	202	nov_-14	194	nov_-16	249	nov_-18	225
déc_-10	175	déc_-12	250	déc_-14	211	déc_-16	188	déc_-18	354

Tableau 0.6 : données isolation mensuelle

- Vent

date	Vent	date	Vent	date	Vent	date	Vent	date	Vent
janv-09	02.7	janv-11	02.7	janv-13	02.7	janv-15	03.3	janv-17	03.1
févr-09	02.7	févr-11	02.7	févr-13	02.7	févr-15	04.5	févr-17	02.8
mars-09	02.7	mars-11	02.7	mars-13	02.7	mars-15	03.5	mars-17	02.7
avr_-09	02.7	avr_-11	02.7	avr_-13	02.7	avr_-15	02.6	avr_-17	02.6
mai_-09	02.7	mai_-11	02.7	mai_-13	02.7	mai_-15	03.3	mai_-17	02.9
juin-09	02.7	juin-11	02.7	juin-13	02.7	juin-15	02.9	juin-17	02.7
juil-09	02.7	juil-11	02.7	juil-13	02.7	juil-15	01.9	juil-17	02.0
août-09	02.7	août-11	02.7	août-13	02.7	août-15	02.5	août-17	02.5
sept-09	02.7	sept-11	02.7	sept-13	02.7	sept-15	02.1	sept-17	02.0
oct_-09	02.7	oct_-11	02.7	oct_-13	02.7	oct_-15	02.1	oct_-17	01.0
nov_-09	02.7	nov_-11	02.7	nov_-13	02.7	nov_-15	01.7	nov_-17	02.2
déc_-09	02.7	déc_-11	02.7	déc_-13	02.7	déc_-15	00.8	déc_-17	02.5
janv-10	02.7	janv-12	02.7	janv-14	03.4	janv-16	02.5	janv-18	02.7
févr-10	02.7	févr-12	02.7	févr-14	04.3	févr-16	03.7	févr-18	03.2
mars-10	02.7	mars-12	02.7	mars-14	04.5	mars-16	02.0	mars-18	04.6
avr_-10	02.7	avr_-12	02.7	avr_-14	03.7	avr_-16	03.2	avr_-18	03.4
mai_-10	02.7	mai_-12	02.7	mai_-14	03.0	mai_-16	03.9	mai_-18	03.3
juin-10	02.7	juin-12	02.7	juin-14	03.4	juin-16	02.9	juin-18	03.3
juil-10	02.7	juil-12	02.7	juil-14	03.0	juil-16	02.7	juil-18	02.9
août-10	02.7	août-12	02.7	août-14	02.5	août-16	02.0	août-18	02.1
sept-10	02.7	sept-12	02.7	sept-14	03.0	sept-16	01.8	sept-18	02.9
oct_-10	02.7	oct_-12	02.7	oct_-14	02.4	oct_-16	01.6	oct_-18	02.7
nov_-10	02.7	nov_-12	02.7	nov_-14	03.1	nov_-16	01.8	nov_-18	02.6
déc_-10	02.7	déc_-12	02.7	déc_-14	03.3	déc_-16	01.8	déc_-18	01.3

Tableau 0.7: données de vent mensuelle

- **Présentation des données annuelle**

Les tableaux suivant présentent les données annuelles des paramètres suivant (précipitation, température max, température min, température moyenne, humidité, insolation, vent).

<i>date</i>	<i>Précipitations</i>	<i>Température MAX</i>	<i>Température MIN</i>	<i>Température MOYENNE</i>	<i>Humidité_%</i>	<i>Isolation total</i>	<i>Vent</i>
2009	14.4	26.1	11.5	18.9	05.0	249.6	02.7
2010	23.7	26.6	12.5	19.5	46.6	223.5	02.7
2011	23.7	25.5	10.6	18.7	49.8	222.4	02.7
2012	13.3	26.0	12.1	19.1	49.0	256.6	02.7
2013	11.5	24.7	14.0	19.9	37.4	255	02.7
2014	08.7	25.3	13.4	19.4	46.6	270	03.3
2015	10.8	24.3	13.0	19.0	38.1	280.1	02.6
2016	07.1	25.7	13.4	19.5	28.9	259.5	02.4
2017	05.5	25.0	13.1	18.9	34.4	270.0	02.4
2018	14.1	22.7	13.7	18.6	42.2	266.5	02.9

Tableau 0.8 : Présentation des données mensuelle