



- PFE
- SI
- SIP

## Rapport de Stage

### Sujet:

**Caractérisation spatiotemporelle des  
précipitations provinciales annuelles en  
utilisant les données de CGMS**

### résumé :

L'objectif de ce stage est l'étude de la variabilité spatiotemporelle des précipitations au Maroc et sur la période 1981-2019, en se basant sur des séries chronologiques de précipitations annuelles fournies par le modèle CGMS-Maroc.

Pour réaliser cet objectif, Nous avons exploité des méthodes et techniques d'analyse de données et de statistiques multidimensionnelles : dans un premier temps, nous avons procédé avec une analyse exploratoire, où on a utilisé les caractéristiques statistiques des données pour construire une idée générale sur les précipitations au Maroc. Ensuite, nous avons employé l'ACP pour réduire la dimension des tableaux de données et compresser l'information, distribuée au début sur 75 variables, dans un nombre réduit de composantes principales (8 CPs). Ensuite, nous avons exploité les méthodes de classification automatique pour classer les années et les provinces selon leurs cumuls annuels de pluies, ce qui nous a permis de tirer les résultats suivants : 74% des années de la période de référence sont sèches (les années 80 et 90) ; selon les cumuls de pluies, les provinces Marocaines peuvent se classer en 8 classes ( 8 climats).

#### Réalisé par :

CHAQDID Abdelaziz  
*2ème année Météorologie*

#### Encadré par :

Mr. EL KHATRI Said  
*Le chef du Centre National du Climat*

**Année universitaire : 2019- 2020**



La Direction Générale de la Météorologie (**DGM**)  
Maroc Météo - Boulevard Mohamed Taieb Naciri –  
Hay Hassani , B.P 8106 Casa-Oasis - Casablanca -  
Maroc



## Rapport de stage d'ingénieur

(Stage à distance)

---

# Caractérisation spatiotemporelle des précipitations provinciales annuelles en utilisant les données de CGMS

---

Réalisé par :

**CHAQDID Abdelaziz**

3ème année génie Météorologie

Ecole Hassania des Travaux Publics (EHTP)

Encadré par :

**M.EL KHATRI Said**

Chef du Centre National du Climat

Direction Générale de la Météorologie (DGM)

Du 12 Juin au 7 Août 2020

## **Remerciement**

Avant tous, je remercie Dieu le tout-puissant et miséricordieux, qui m'a donné le courage, la force et la patience pour accomplir ce travail.

Je tiens également à exprimer ma gratitude et mon respect à mon prof et encadrant M. El khatri Said, pour son enseignement de qualité, son encadrement et ses précieux conseils professionnels ainsi que personnels.

À cette occasion, je tiens à remercier, les professeurs et le staff administratif de l'école Hassania des Travaux Publics pour tout l'effort qu'ils ont fourni pour que tout se passe dans l'ordre et dans les bonnes conditions malgré la situation sanitaire critique que connaît notre pays et le monde entier cette année.

Finalement, je voudrais exprimer ma reconnaissance à toute personne ayant contribué de près ou de loin, même avec un simple sourire, à la réussite de ce stage.

## **Avant-propos**

Dans le cadre de la formation des ingénieurs, l'Ecole Hassania des Travaux Publics exige à ses élèves ingénieurs d'effectuer un stage d'ingénieur (SI) à la fin de leur deuxième année.

Ce stage d'ingénieur permet aux élèves ingénieurs d'agrémenter leurs notions théoriques par la pratique, de s'accoutumer et se familiariser avec le monde professionnel et d'enrichir leur savoir-faire et l'esprit technique. Durant ce stage, l'élève ingénieur est appelé à mettre en œuvre les capacités acquises pendant sa formation, à combler ses lacunes et à approfondir ses connaissances. Tout en restant conscient de la responsabilité que le poste d'ingénieur exige en termes de prise de décision, de créativité et de développement.

La situation épidémique au Maroc et dans le monde entier imposa, exceptionnellement cette année, à l'EHTP comme à d'autres écoles d'ingénieurs de prendre la décision de réaliser tous les stages à distance. Une décision qui aura sans doute son impact sur la réalisation des objectifs du stage SI. Mais, grâce à la coordination entre le staff administratif, les encadrant et les élèves ingénieurs les stages se sont déroulés dans les bonnes conditions et sans problèmes, et ce rapport en est la preuve.

# Table des matières

Remerciement . . . . .	2
Avant-propos . . . . .	3
Table des figures . . . . .	7
Liste des tableaux . . . . .	8
<b>1 Introduction</b>	<b>11</b>
1.1 Présentation du sujet . . . . .	11
1.2 Le système CGMS . . . . .	12
1.3 Le rôle de la DGM dans CGMS-Maroc . . . . .	12
1.4 Description des données . . . . .	13
<b>2 Analyse exploratoire</b>	<b>14</b>
2.1 Les paramètres statistiques utilisés . . . . .	14
2.2 Analyse des résultats . . . . .	15
2.2.1 La moyenne des pluies annuelles . . . . .	15
2.2.2 Indice de dispersion absolu des pluies annuelles . . . . .	18
2.2.3 Indice de dispersion relatif des pluies annuelles . . . . .	18
<b>3 Analyse en composantes principales (ACP)</b>	<b>20</b>
3.1 Présentation de la méthode (ACP) . . . . .	20
3.1.1 Les données d'entrée . . . . .	21
3.1.2 Le principe . . . . .	21
3.1.3 Les données de sortie . . . . .	21
3.2 Application de la méthode . . . . .	21
3.2.1 Le nombre de composantes principales à conserver . . . . .	22
3.2.2 Etude des composantes principales et des vecteurs propres . . . . .	23
3.2.3 Etude des années . . . . .	29
<b>4 Classification automatique</b>	<b>34</b>
4.1 Classification hiérarchique ascendante . . . . .	34
4.1.1 Présentation de la méthode . . . . .	34
4.1.2 Classification des années . . . . .	35
4.1.3 Classification des provinces . . . . .	39
4.2 Méthode des centres mobiles (k-means) . . . . .	41
4.2.1 Présentation de la méthode . . . . .	41
4.2.2 Classification des années . . . . .	42

4.2.3	Classification des provinces . . . . .	44
4.3	Classification d'une nouvelle observation . . . . .	44
4.3.1	Positionnement du problème . . . . .	44
4.3.2	Solution . . . . .	44
4.3.3	Résultats . . . . .	45
<b>Conclusion</b>		<b>47</b>
Annexe A	. . . . .	49
Annexe B	. . . . .	55
Annexe C	. . . . .	59
<b>Bibliographie</b>		<b>61</b>

# Table des figures

2.1	La moyenne des précipitations annuelles au Maroc pour la période 1981-2019	16
2.2	carte des climats mondiaux selon la classification de Köppen [1] . . . . .	17
2.3	L'écart-type des précipitations annuelles au Maroc pour la période 1981-2019	18
2.4	Le coefficient de variation des précipitations annuelles au Maroc pour la période 1981-2019 . . . . .	19
3.1	Graphe des valeurs propres . . . . .	22
3.2	Evolution temporelle des quatres premières composantes principales . . . . .	24
3.3	Représentation graphique du vecteur propre numéro 1 . . . . .	25
3.4	Comparaison entre la moyenne des précipitations annuelles au Maroc et la première composante principales . . . . .	26
3.5	Représentation graphique du vecteur propre numéro 2 . . . . .	26
3.6	Représentation graphique du vecteur propre numéro 3 . . . . .	27
3.7	Représentation graphique du vecteur propre numéro 5 . . . . .	28
3.8	Représentation graphique du vecteur propre numéro 7 . . . . .	29
3.9	Projection des individus sur le plan (CP1, CP2) . . . . .	29
3.10	Les 20 premiers individus, classés par le cosinus carré, projetés dans le plan (CP1, CP2) . . . . .	30
3.11	L'anomalie réduite de l'année 1987 reconstituée par CP1, CP1 et CP2 et observation complète. . . . .	31
3.12	L'anomalie réduite de l'année 2014 reconstituée par CP1, CP1 et CP2 et observation complète. . . . .	31
3.13	L'anomalie réduite de l'année 2000 reconstituée par CP1, CP1 et CP2 et observation complète. . . . .	32
3.14	L'anomalie réduite de l'année 2009 reconstituée par CP1, CP1 et CP2 et observation complète. . . . .	32
4.1	Dendrogramme des années . . . . .	35
4.2	Les écarts des précipitations annuelles par rapport à la moyenne sur la période 1981-2019, au Maroc . . . . .	37
4.3	Représentations graphiques de l'anomalie par rapport à la normale des étalons réels et fictifs des classes fournies par la CHA des années . . . . .	38
4.4	Dendrogramme des provinces . . . . .	39
4.5	Evolution de la distance d'agrégation en fonction des étapes . . . . .	40
4.6	Evolution du saut d'agrégation entre deux classes en fonction des étapes .	40

4.7	Représentation graphique des résultats de la classification hiérarchique des provinces . . . . .	41
4.8	Projection des classes fournies par la méthode Kmeans sur le plan (CP1, CP2) . . . . .	42
4.9	Représentations graphiques de l'anomalie par rapport à la normale des étalons réels et fictifs des classes fournies par la méthode K-means . . . . .	43
4.10	Représentation graphique des résultats de la méthode Kmeans appliquée aux provinces . . . . .	44
4.11	Contributions des individus à la formation des CPs . . . . .	55
4.12	Contributions des variables à la formation des CPs . . . . .	56
4.13	Représentations graphiques des vecteurs propres (non présentées dans les chapitres) . . . . .	57
4.14	Projection des 10 premières variables sur les plans formés par les CPs deux à deux . . . . .	58

# Liste des tableaux

2.1	Les valeurs extrêmes des moyennes annuelles provinciales des précipitations sur la période 1981-2019 . . . . .	16
3.1	La variance expliquée par les dix premières composantes principales . . . . .	23
4.1	Les classes et leur effectifs, obtenues par CHA . . . . .	36
4.2	Comparaison des classes fournies par les deux méthodes : k-means et CHA	42
4.3	Les cumuls de précipitations pour l'années 2020 estimée par la moyenne mobile . . . . .	45
4.4	Classification des années selon la quantité des précipitations annuelles sur la période 1981-2019 . . . . .	47
4.5	Statistiques descriptives des provinces : . . . . .	59
4.6	Statistiques descriptives des années : . . . . .	60

# **Acronymes**

**ACP** Analyse en Composantes Principales.

**CA** Classification Automatique.

**CGMS** Crop Growth Monitoring System.

**CHA** Classification Hiérarchique Ascendante.

**DGM** Direction Générale de la Météorologie.

Cette page a été laissé en blanc intentionnellement

# Chapitre 1

## Introduction

### Sommaire

---

1.1	Présentation du sujet . . . . .	11
1.2	Le système CGMS . . . . .	12
1.3	Le rôle de la DGM dans CGMS-Maroc . . . . .	12
1.4	Description des données . . . . .	13

---

L'intérêt porté à l'étude de la variabilité des précipitations au Maroc devient de plus en plus important, vu qu'elle est reliée à un sujet d'actualité, qui est les changements climatiques. Mais, certainement cet intérêt n'est pas récent. En effet, en 1921 Augustin, du laboratoire de géologie physique et cartographie de l'institut scientifique de Rabat publia les valeurs moyennes de précipitation au Maroc, et dressa la première carte de précipitations au Maroc. Et en 1956 Debrach, Gaußen et Joly publièrent une base de données pluviométriques du Maroc (1925/1949) et une notice sur les procédés de traitement et de calculs. En 1957, Gaußen et Bagnouls rétablirent la carte des précipitations au Maroc sur la période 1925/1950. Et jusqu'à date, les publications et les travaux dans ce sens n'ont jamais cessé. Au contraire, ce sujet commence à avoir plus d'intérêt, vu sa relation avec d'autres sujets d'actualité national comme le risque de la sécheresse et le stress hydrique, et des sujets globaux comme les changements climatiques.

A partir de ce point, nous avons déclenché notre étude qui sera d'une part une approche pour comprendre la variabilité des précipitations au Maroc, et d'autre part, une occasion pour la confrontation des connaissances théoriques acquises à l'école avec les problèmes de la vie réelle.

### 1.1 Présentation du sujet

Le cerveau humain, en générale, est très familiarisé avec l'interprétation des couleurs et des formes géométriques. Or il trouve une grande difficulté à interpréter des tableaux de nombres. Pire, notre cerveau perd le fil dès que le tableau dépasse cent lignes et dix colonnes, par exemple.

L'objectif de notre projet, c'est l'extraction et la synthèse de l'information cachée dans des séries chronologiques de précipitations fournies par le modèle CGMSMaroc. La synthèse sera sous forme de cartes colorées, graphes et tables synthétiques.

Pour réaliser cet objectif, Nous exploiterons des méthodes et techniques d'analyse de données et de statistiques multidimensionnelles : dans un premier temps, nous procéderons avec une analyse exploratoire, où nous utiliserons les caractéristiques statistiques des données pour construire une idée générale sur les précipitations au Maroc. Ensuite, nous emploierons l'ACP pour réduire la dimension des tableaux de données et compresser l'information, distribuée au début sur 75 variables, dans un nombre réduit de composantes principales, et qui nous permettra de conclure facilement sur la variabilité des précipitations. Souvent, l'ACP est suivie de la CA cette méthode permet la subdivision de la donnée en des classes, ce qui rend la visualisation des individus qui se ressemblent et ceux qui se diffèrent plus claire et facile. Nous exploiterons cette méthode dans notre cas pour classer les années selon leurs cumuls de précipitations d'une part, classer les provinces dans le but d'une régionalisation d'autre part. Et finalement, nous faisons usage de l'outil ArcGis pour représenter nos résultats sous forme de cartes synthétiques.

## 1.2 Le système CGMS

Un système national de suivi de la campagne agricole et de prédiction agrométéorologique des récoltes céréalières, appelé « CGMS-MAROC » (Crop Growth Monitoring System – Maroc), a été initié par l'Institut National de la Recherche Agronomique (INRA), dans le cadre du projet E-AGRI. Le CGMSMAROC est piloté par l'INRA et géré en consortium formel avec la Direction de la Météorologie Nationale (DGM) et la Direction de la Stratégie et des Statistiques (DSS). Le CGMS-MAROC est constitué de trois niveaux :

**Niveau 1** : la collecte des données météorologiques et leur interpolation sur une grille carrée, de 9x9 km de superficie, disponible sur tout le territoire national ;

**Niveau 2** : la simulation de la croissance des cultures, par plusieurs modèles de simulations agrométéorologiques, grâce à la plate-forme BioMA ;

**Niveau 3** : la prédiction des récoltes à partir d'une approche combinée, mettant à contribution des analyses statistiques paramétriques et non paramétriques des données météorologiques, des données de simulation et des données satellitaires. [1] [2]

## 1.3 Le rôle de la DGM dans CGMS-Maroc

La DGM est l'un des organismes marocains et africains les plus armés en termes d'équipements informatiques de modélisation et de prévision. L'exploitation et la maintenance de ce matériel informatique, pour une longue période, lui ont également permis d'accumuler une expertise dans ce domaine. En outre, la DGM dispose d'une large base de données des paramètres météorologiques, dont le traitement constitue le cœur et l'étape initiale de CGMS. Cette base de données contient l'historique des paramètres météorologique sur une période d'années et pour la plupart du territoire marocain et est actualisée chaque

jour. Ainsi, l'apport de la DGM à CGMS-Maroc se résume en :

- L'hébergement et la maintenance informatique de CGMS-MAROC au profit des trois institutions ;
- L'interpolation des données climatiques du réseau des stations météorologiques et l'utilisation de ces données interpolées au niveau 1 de CGMS-MAROC.

## 1.4 Description des données

Les données brutes fournies par CGMS sont sous format de shapefile (points de grille), les points de grille dans ce fichier sont espacés de 9 km et répartis sur l'ensemble du territoire marocain, en chaque point de grille on peut récupérer, via le tableau attributaire, le cumul de précipitations annuelles pour un certain nombre d'années.

Dans cette étude, nous travaillerons sur des données de l'année agricole (du mois Septembre de l'année grégorienne à Août), tout en acceptant que l'année 2019-2020 est complète, même si on n'a pas encore reçu les données de juillet et Août, qui sont peu influentes sur le cumul annuel. Les données utilisées sont sous forme de table (75 lignes contenant les provinces et 39 colonnes pour les années de 1981 à 2019). Cette table est dérivée du shapefile en calculant la moyenne spatiale des précipitations annuelles par provinces et pour chaque année. La table des données peut être représentée comme suit :

$$P = \begin{array}{c} Tanger \\ \vdots \\ Aousserd \end{array} \quad \left( \begin{array}{ccc} p_{ij} & \dots & p_{i,39} \\ \vdots & \ddots & \vdots \\ p_{75,j} & \dots & p_{39,75} \end{array} \right)$$

## Bibliographie

- [1] T.El Hairech D.Qinghan A.de Wit, Riad Ballaghi. Cgms piloting report morocco. 26/02/2014.
- [2] M.TAHRI M.LAHLOU R.BALAGHI, T.EL HAIRECH. Cgms-maroc manuel d'utilisation. Avril 2013.

# Chapitre 2

## Analyse exploratoire

### Sommaire

---

<b>2.1</b>	<b>Les paramètres statistiques utilisés . . . . .</b>	<b>14</b>
<b>2.2</b>	<b>Analyse des résultats . . . . .</b>	<b>15</b>
2.2.1	La moyenne des pluies annuelles . . . . .	15
2.2.2	Indice de dispersion absolu des pluies annuelles . . . . .	18
2.2.3	Indice de dispersion relatif des pluies annuelles . . . . .	18

---

Comme le titre le suggère, on essayera à travers ce chapitre de faire une analyse exploratoire, dont le but sera ; premièrement, de découvrir la donnée en situant les valeurs des séries de pluies à l'aide des paramètres statistiques appelés ‘indices de position’. Deuxièmes, par le biais des indices de dispersion (absolu et relatif), on tentera d'étudier l'irrégularité des précipitations annuelles sur toutes les provinces du Royaume.

### 2.1 Les paramètres statistiques utilisés

Il existe de nombreux paramètres ou statistiques qui permettent d'avoir une idée générale de la donnée. Ces paramètres sont de deux types :

Les indices de position : ils permettent de situer les valeurs d'une série de nombres réels. Parmi ces indices, on peut citer : la moyenne, la médiane, les quartiles, le mode etc.

Les indices de dispersions : quant à eux, ils expriment la variabilité ou la dispersions des valeurs de la série autour d'un indicateur de position. Par exemple : l'écart-type, la variation ou l'écart interquartile. Dans notre cas nous avons opté pour les indices suivants :

**Indices de position** : la moyenne arithmétique, qui est définie comme suit : Pour une série de cumuls annuels  $P = \{p_1, p_2, \dots, p_n\}$ , on définit sa moyenne par :

$$\bar{P} = \frac{1}{n} \sum_{i=1}^n p_i$$

**Indices de dispersion** : ce paramètre nous permettra de mieux comprendre la variabilité spatiotemporelle des précipitations en quantifiant l'ampleur de l'irrégularité plu-

viométrique, pour cela nous choisissons deux indices :

*L'indice de dispersions absolue* : l'écart-type, qui se définit par :

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - \bar{P})^2}$$

En utilisera cet indice pour quantifier la variabilité temporelle des précipitations par rapport à la moyenne au sein de chaque province.

*L'indice de dispersions relatif* : le coefficient de variation, qui se définit par :

$$C_v = \frac{\sigma}{\bar{P}} \times 100$$

Ce paramètre nous permettra de comparer les provinces selon la variabilité des précipitations.

## 2.2 Analyse des résultats

### 2.2.1 La moyenne des pluies annuelles

Afin de situer les valeurs de la série de précipitations dont nous disposons, nous avons employé la moyenne arithmétique. Les résultats obtenus sont représentés graphiquement sur la carte (figure 2.1) et sur la table 2.1 qui représente les valeurs extrêmes des moyennes annuelle et provinciale des précipitations sur la période 1981-2019 .

De prime abord, la carte des moyennes montre un contraste puissant entre les provinces de Sud et celles du Nord. En effet, la moyenne des précipitations varie selon deux axes : un premier axe Sud-Nord, où elle croît du sud vers le nord en passant de 33,5 mm à Aousserd à 811,2 mm à Chefchaouen. Le deuxième axe Est-Ouest, concerne plutôt la moitié Nord du pays, où la moyenne varie entre 163 mm à Figuig et 519 à Salé.

Sur le premier axe, L'opposition entre les provinces de Sud et celles du Nord est dû, en général, à l'effet de la latitude. En effet, plus on descend du Nord vers le Sud du pays, plus on quitte les régions relativement humides et ayant un climat méditerranéen caractérisé par les étés secs et chauds et les hivers humides et doux (selon la classification de Köppen (figure 2.2), et on pénètre les régions désertiques ayant un climat Aride caractérisé par la sécheresse permanente et une rareté des précipitations pendant toute l'année, selon la même classification.

Sur le deuxième axe, l'opposition entre les provinces est causée par deux facteurs : l'altitude et la distance par rapport aux côtes. Les provinces de l'Ouest, allant de Tanger au Nord à Safi au Sud, bénéficient d'un climat méditerranéen influencé par les perturbations océaniques venant de l'Atlantique pendant les périodes de pluie. Plus on s'avance vers l'Est, le climat méditerranéen se dégrade et le climat montagnard prend sa place. Par conséquent les provinces côtières ainsi que les provinces montagneuses qui sont face à l'océan Atlantique reçoivent des pluies plus importantes, au contraire des provinces internes dites 'sous le vent'. En fait, les chaînes de l'Atlas forment une barrière qui empêche les masses d'air humides chargées de pluies d'atteindre ces provinces.

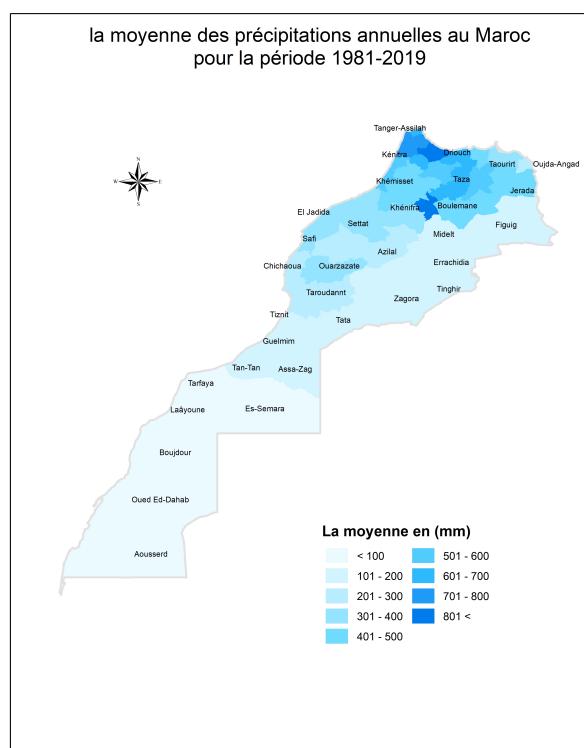
**Table 2.1 – Les valeurs extrêmes des moyennes annuelles provinciales des prrécipitations sur la période 1981-2019**

3 provinces les plus sèches	Moyenne (en mm)	3 provinces les plus humides	Moyenne (en mm)
<b>Oued Ed-Dahab</b>	<b>32.7</b>	<b>Chefchaouen</b>	<b>811.2</b>
Aousserd	33.5	Ifra	806.9
Boujdour	42.2	Tétouan	722.3

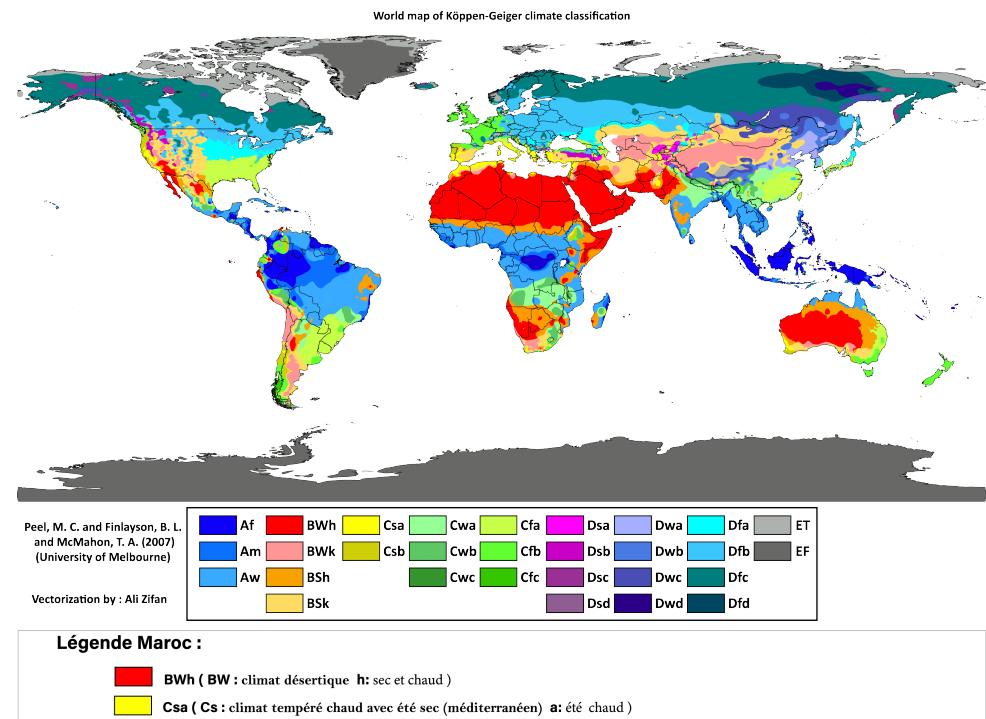
  

3 Années les plus sèches	Moyenne (en mm)	3 Années les plus humides	Moyenne (en mm)
<b>1994</b>	<b>213.3</b>	<b>2009</b>	<b>638.1</b>
1982	226.4	1995	615.2
1992	226.8	2008	590.2

**Figure 2.1 – La moyenne des précipitations annuelles au Maroc pour la période 1981-2019**



**Figure 2.2 – carte des climats mondiaux selon la classification de Köppen [1]**



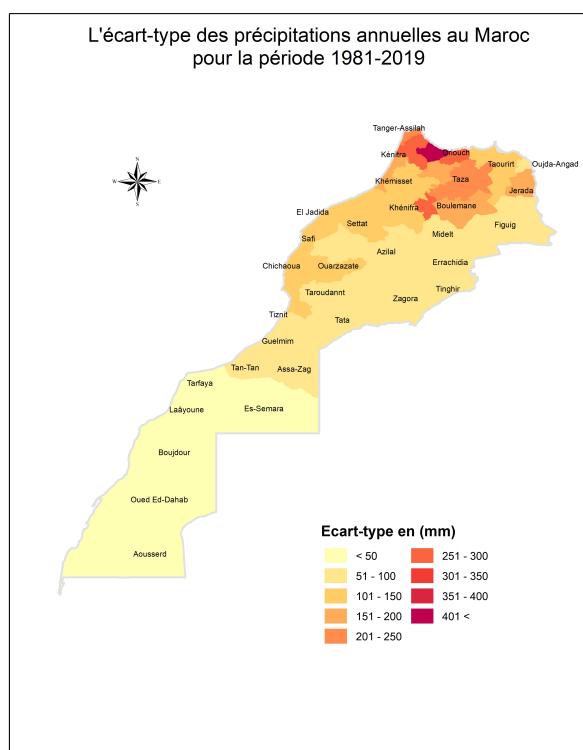
## 2.2.2 Indice de dispersion absolu des pluies annuelles

Pour calculer la dispersion des valeurs des séries de précipitation par rapport à la moyenne, en chaque province, nous avons calculé l'écart-type de ces séries. Les résultats sont représentés sous forme de carte (figure 2.3).

Tout d'abord, on remarque que l'ampleur de la dispersion des précipitations annuelles n'est pas uniforme sur tout le territoire marocain. En effet, les écarts des pluies annuelles par rapport à la moyenne, croît du Sud au Nord : les pluies annuelles au Sud, sont faiblement dispersées par rapport à la moyenne (27 mm à Boujdour), tandis que celles du Nord sont fortement dispersées (447 mm à Chefchaouen).

En outre, on constate une grande ressemblance entre la carte de la moyenne et celle de l'écart type. Ainsi, on conclut que plus la moyenne des précipitations annuelles est grande plus l'écart-type est important. Par conséquent, l'écart-type à lui seul ne permet pas la comparaison de l'ampleur de l'irrégularité des précipitations annuelles entre les différentes provinces du pays. Pour remédier à ce problème, nous utiliserons le coefficient de variation qui normalise les écarts-type des provinces par leurs moyennes.

**Figure 2.3 – L'écart-type des précipitations annuelles au Maroc pour la période 1981-2019**



## 2.2.3 Indice de dispersion relatif des pluies annuelles

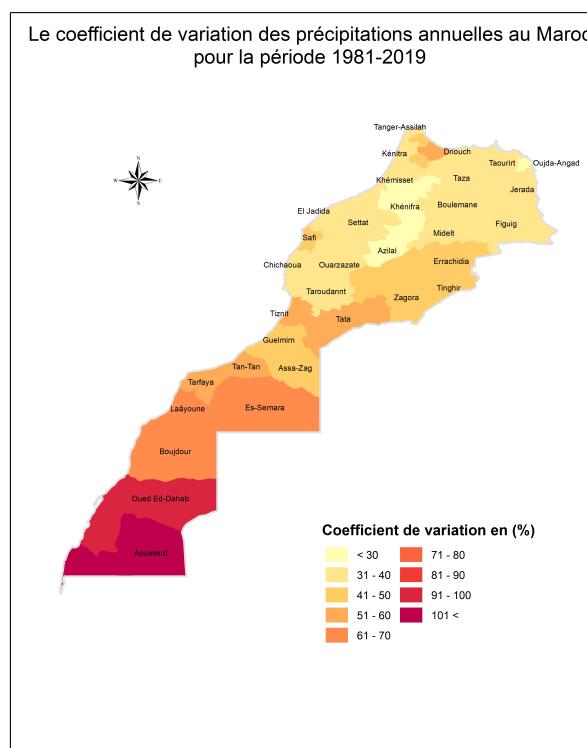
L'analyse de la figure 2.4 montre les deux informations suivantes :

**L'irrégularité des pluies annuelles au Maroc décroît avec la latitude :** Sachant que Cv exprime la dispersion des précipitations après élimination de l'effet de la moyenne,

on conclut que les provinces du Sud révèlent une forte irrégularité des précipitations. En effet, les valeurs de Cv les plus grandes sont prises par les provinces du Sud : Aousserd 103.3%, Oued Ed-Dahab 93.2%, Laâyoune 67.4% etc. Alors que, les valeurs les plus faibles sont prises par les provinces du nord et du centre : Béni Mellal 28.1%, Azilal 29.5%, Tanger-Assilah 33.8% etc. ce qui prouve que les pluies annuelles au Nord du pays sont relativement plus régulières.

**L'effet de l'altitude combiné avec l'effet de l'exposition à l'océan donnent lieu à des valeurs très faibles de Cv :** L'effet de l'altitude combiné avec l'effet de l'exposition à l'océan donnent lieu à des valeurs très faibles de Cv : Béni Mellal 28.1% (située au pied du haut Atlas (500 m d'altitude), parfaitement exposée aux masses d'air humides provenant de l'Atlantique), Ouarzazate 30% (située au Sud-Est du Haut Atlas (1151 m), mal exposée à l'océan Atlantique). Afin d'avoir une idée globale sur l'irrégularité des pluies annuelles au Maroc, nous avons calculé, en plus, le coefficient de variation des moyennes annuelles sur tout le Maroc, ce qui nous a permis de conclure que les pluies annuelles au Maroc sont irrégulières et très dispersées par rapport à la moyenne, avec un coefficient de variation Cv=56 ( sachant que la moyenne provinciale est 356.7 mm et l'écart-type est 201.7 mm).

**Figure 2.4 – Le coefficient de variation des précipitations annuelles au Maroc pour la période 1981-2019**



## Bibliographie

- [1] [www.wikipedia.org](http://www.wikipedia.org) (classification de köppen).

# Chapitre 3

## Analyse en composantes principales (ACP)

### Sommaire

---

<b>3.1 Présentation de la méthode (ACP) . . . . .</b>	<b>20</b>
3.1.1 Les données d'entrée . . . . .	21
3.1.2 Le principe . . . . .	21
3.1.3 Les données de sortie . . . . .	21
<b>3.2 Application de la méthode . . . . .</b>	<b>21</b>
3.2.1 Le nombre de composantes principales à conserver . . . . .	22
3.2.2 Etude des composantes principales et des vecteurs propres . . . . .	23
3.2.3 Etude des années . . . . .	29

---

Ce deuxième chapitre traitera l'application de l'analyse en composantes principales (ACP) sur les séries des cumuls de précipitations annuels. Mais, avant de se lancer dans l'application directe, il serait judicieux de présenter brièvement la méthode d'analyse en composantes principales :

### 3.1 Présentation de la méthode (ACP)

L'Analyse en Composantes principales (ACP) fait partie du groupe des méthodes descriptives multidimensionnelles appelées méthodes factorielles. L'ACP prend sa source dans un article de Karl Pearson publié en 1901, le père du test du  $\xi$ . Elle est une méthode descriptive, qui ne s'appuie pas sur un modèle probabiliste, mais elle dépend d'un modèle géométrique. L'ACP propose, à partir d'un tableau rectangulaire de données comportant les valeurs de  $p$  variables quantitatives pour  $n$  unités (appelées aussi individus), des représentations géométriques de ces unités et de ces variables.

### 3.1.1 Les données d'entrée

Les données sont les mesures de précipitations effectuées sur la période allant de 1981 à 2019 et dans l'ensemble des provinces marocaines. Ainsi, les 39 ans 1981, 1982, ..., 2000, ... 2019 représenteront les observations (ou les individus) enregistrées dans les 75 provinces Marrakech, Tanger, ..., Casablanca, ...Aousserd (les variables). Le tableau des données brutes à partir duquel on fera l'analyse est noté P et a la forme suivante :

$$P = \begin{array}{c} 1981 \\ \vdots \\ 2019 \end{array} \left( \begin{array}{ccc} p_{ij} & \dots & p_{i,39} \\ \vdots & \ddots & \vdots \\ p_{75,j} & \dots & p_{39,75} \end{array} \right)$$

La matrice à diagonalisée est, selon le besoin, la matrice de variances-covariances ou la matrice de corrélation équivalente à P.

### 3.1.2 Le principe

L'ACP est une technique permettant de réduire la dimensionnalité des données, d'augmenter l'interprétabilité tout en minimisant la perte d'informations. Ceci est rendu possible en créant de nouvelles variables non corrélées qui maximisent successivement la variance. Trouver ces nouvelles variables, les composantes principales (CPs), revient à résoudre un problème de valeur propre / vecteur propre (problème d'optimisation), et les nouvelles variables sont classées de façon à ce que la première CP exprime le maximum de variance. Les références [1] et [2] fournissent plus de détails sur le principe de l'ACP.

### 3.1.3 Les données de sortie

Les données de sortie, principales, sont les solutions du problème d'optimisation : le vecteurs propres et leurs valeurs propres. La matrice des vecteurs propres aura la forme suivante :

$$A = \begin{array}{c} Tanger \\ \vdots \\ Aousserd \end{array} \left( \begin{array}{ccc} a_{ij} & \dots & a_{i,39} \\ \vdots & \ddots & \vdots \\ a_{75,j} & \dots & a_{39,75} \end{array} \right)$$

## 3.2 Application de la méthode

Les méthodes statistiques employées dans ce projet sont toutes implémentées à l'aide du langage de programmation R (tous les scripts utilisés sont bien commentés et présentés en annexe A). Ainsi, les résultats discutés dans ce chapitre sont obtenus à l'aide de la fonction, prédéfinie en R, *prcomp* (librairie : *factoextra*). En arguments de *prcomp*, nous

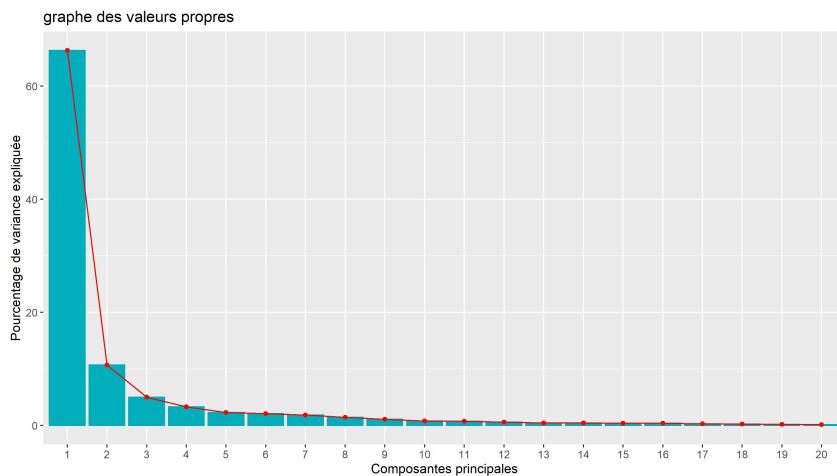
avons choisi de centrer et de réduire les données, ce qui revient à faire l'analyse avec la matrice de corrélation. Toutefois, une comparaison entre les résultats obtenus avec matrice de corrélation et ceux obtenus avec matrice de variancescovariances sera fait de temps en temps afin de mieux comprendre et interpréter les résultats.

### 3.2.1 Le nombre de composantes principales à conserver

Le choix du nombre de composantes à garder se fait selon plusieurs critères, dont on peut citer : le critère de Cattell, le critère de Kaisser, le critère de Nicolaev etc. Mais, les deux critère Kaisser et Cattell restent les plus utilisés parmi les autres.

**Critère de Cattell (coude) :**

**Figure 3.1 – Graphe des valeurs propres**



Selon le graphe, le nombre de composantes à conserver, d'après Cattell, est 4, avec un total de variance expliquée égale à 85.4%.

### Critère de Kaisser :

**Table 3.1 – La variance expliquée par les dix premières composantes principales**

composante principale	Valeurs propres	variance(en%)	variance cumulée (en %)
CP 1	49.7	66.3	66.3
CP 2	8.0	10.7	77.0
CP 3	3.7	5.0	82.0
<b>CP 4</b>	<b>2.5</b>	<b>3.3</b>	<b>85.4</b>
CP 5	1.7	2.3	87.7
CP 6	1.6	2.1	89.8
CP 7	1.4	1.8	91.6
<b>CP 8</b>	<b>1.1</b>	<b>1.5</b>	<b>93.1</b>
CP 9	0.8	1.1	94.2
CP 10	0.6	0.8	95.0

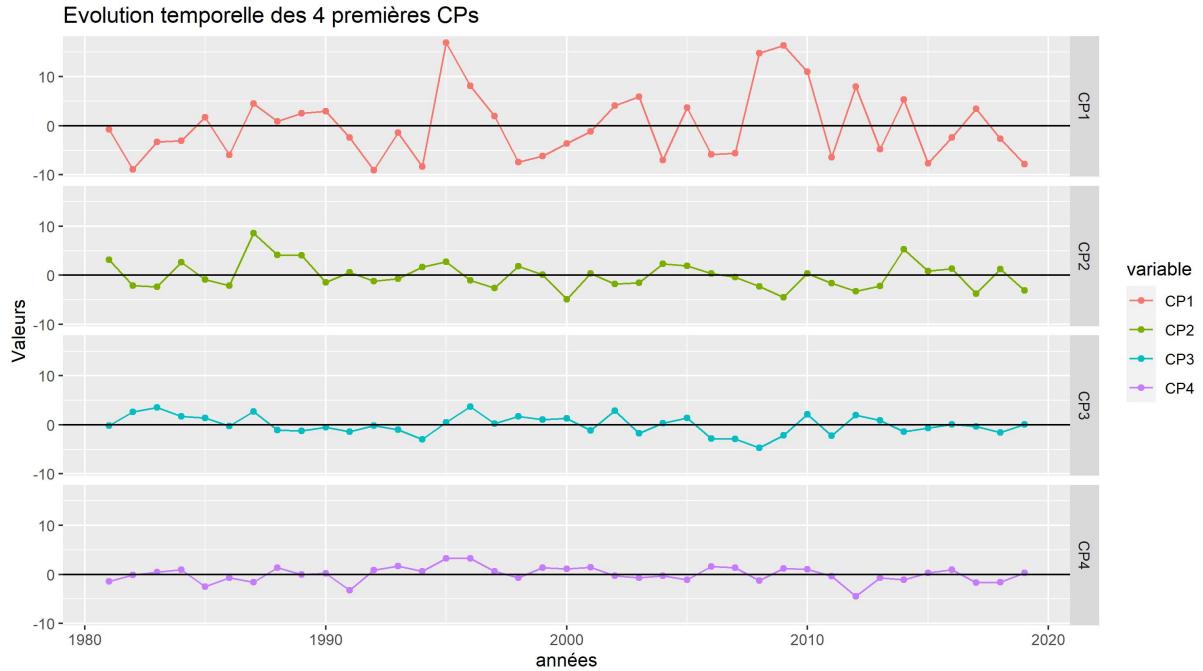
Selon le critère de Kaiser, on conserve seulement les premières composantes ayant une valeur propre supérieure ou égale à 1. Alors on doit conserver les 8 premières composantes principales, exprimant 93.1% de la variance totale.

### 3.2.2 Etude des composantes principales et des vecteurs propres

Dans ce qui suit, nous adoptons le critère de Kaisser, ainsi le nombre de CPs à gardées est 8, permettant une perte de 6.9% de l'information totale. En outre, l'analyse dans cette section portera seulement sur les CPs qui nous semblent utiles et qui fournissent une information claire et facilement interprétable, les représentations graphiques du reste des composantes et vecteurs propres seront présentées en annexe B.

## Evolution des composantes principales :

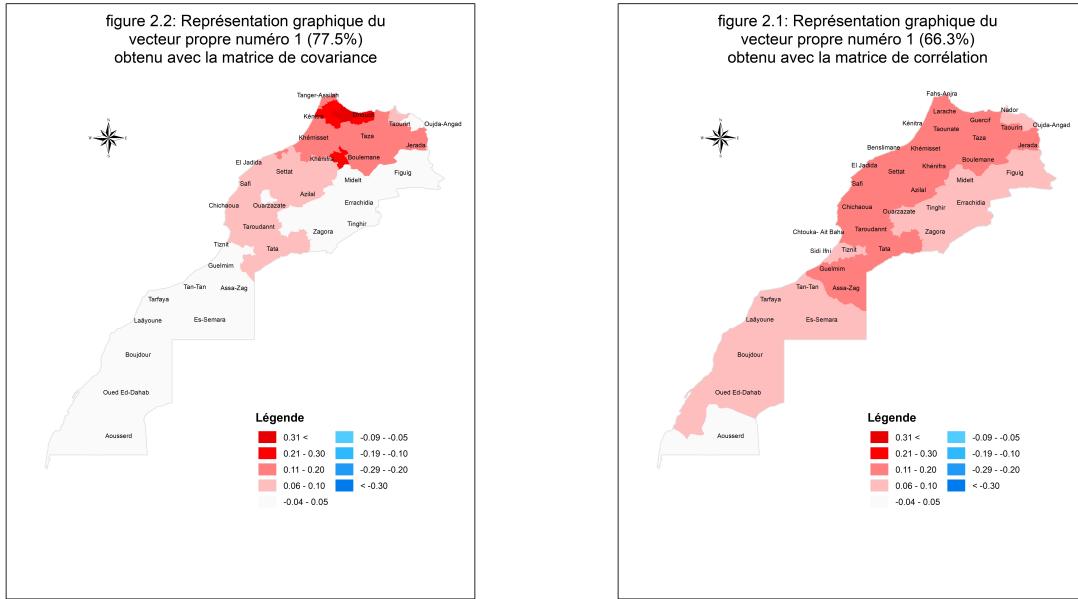
**Figure 3.2 – Evolution temporelle des quatres premières composantes principales**



De prime abord, on constate que l'évolution des quatre premières composantes ne présente aucune périodicité et que l'amplitude des composantes décroît rapidement en passant de la première à la quatrième composante.

La courbe d'évolution de cette composante marque plusieurs maximums relatifs intéressants par exemple 1995, 2008 et 2009, ainsi que des minimums relatifs tels que 1994, 1992 et 1982. En fait, ces piques correspondent exactement aux valeurs extrêmes présentées en table 2.1 (chapitre 2), qui représentent les valeurs extrêmes de la moyenne spatiale sur toutes les provinces du Royaume ayant le même poids dans notre étude.

**Composante principale numéro 1 (CP1) :** En plus, la représentation graphique du vecteur propre numéro 1 (figure 3.3(a)) est très proche de celle de la moyenne (figure 2.1 , chapitre 1), cette ressemblance devient de plus en plus claire sur la représentation graphique du vecteur propre numéro 1 (figure 3.3(b)), obtenu sans réduire les données.



(a) Obtenu avec la matrice de corrélation

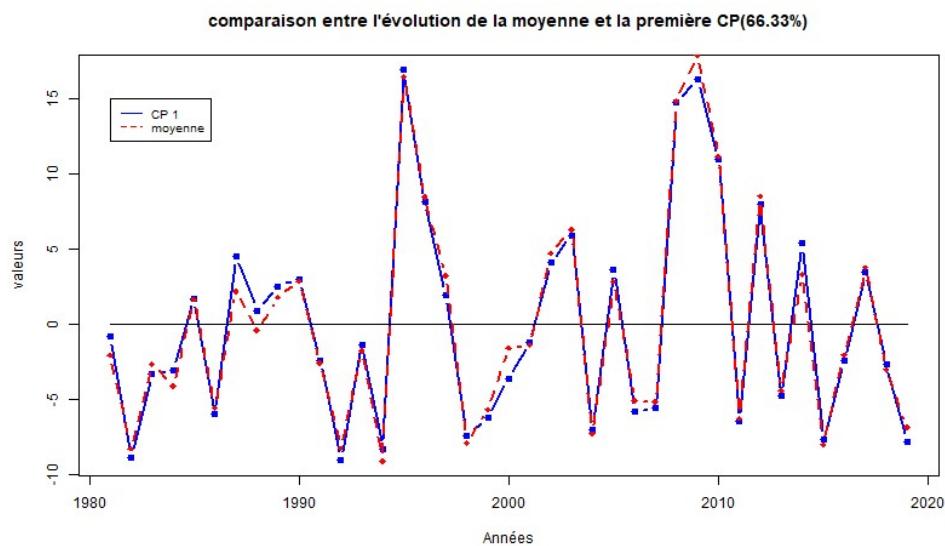
(b) Obtenu avec la matrice de covariance

**Figure 3.3 – Représentation graphique du vecteur propre numéro 1**

Le tracé de l'évolution de CP1 et la moyenne des précipitations annuelles au Maroc (figure 3.4), complète nos remarques précédentes et montre que CP1 suit une évolution pareille à celle de la moyenne. Par conséquent, **la première composante principale décrit la moyenne des précipitations annuelles au Maroc.** la première composante principale décrit la moyenne des précipitations annuelles au Maroc.

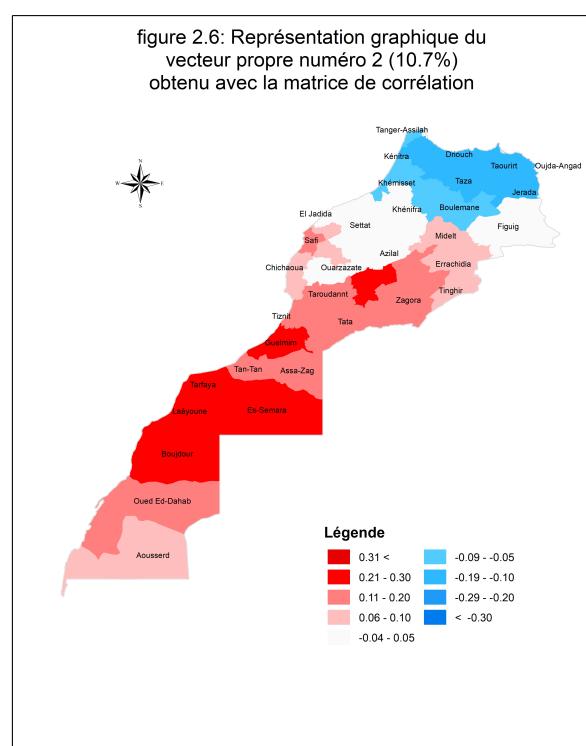
**Remarque :** avant de tracer CP1 et la moyenne (figure 3.4), nous avons centré et réduit les deux signaux pour pouvoir comparer leurs évolutions, par conséquent l'axe des ordonnées n'a pas d'unité, il représente les valeurs centrées réduites de la moyenne et de CP1. Cette procédure est obligatoire pour pouvoir faire une comparaison visuelle, vu que les écarts types et les moyennes des deux signaux sont très différentes.

**Figure 3.4 – Comparaison entre la moyenne des précipitations annuelles au Maroc et la première composante principales**



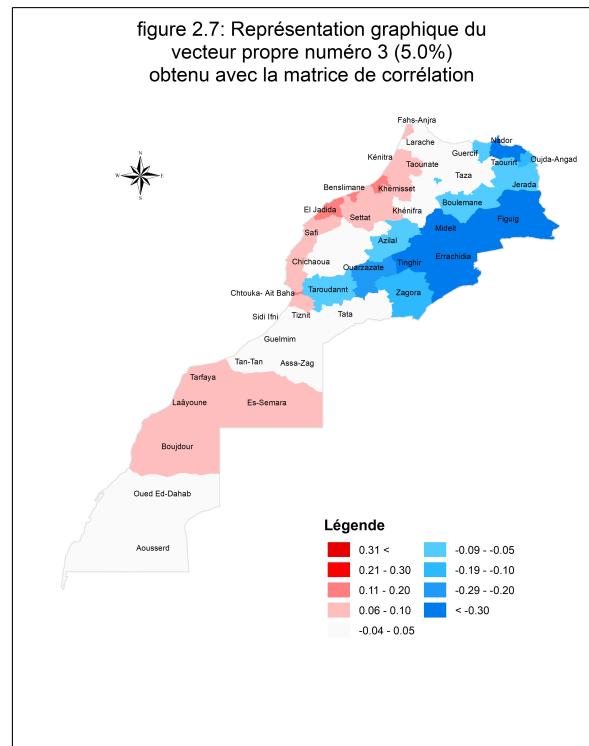
**La composante principale numéro 2 (CP2) :** La représentation graphique du vecteur propre numéro 2, figure ci-dessous, montre une opposition entre les provinces méditerranéennes et les provinces du Sud. Par conséquent, **la deuxième composante principale exprime la variabilité des précipitations sur l'axe Nord-Sud sous l'effet de la latitude.**

**Figure 3.5 – Représentation graphique du vecteur propre numéro 2**



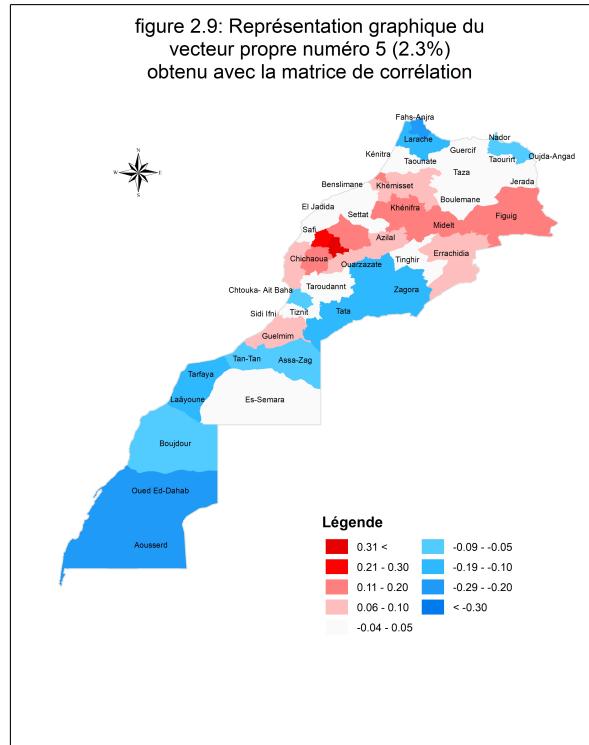
**La composante principale numéro 3 (CP3) :** La représentation graphique du vecteur propre numéro 3 (figure 3.7), permet de distinguer facilement entre les provinces de la côte atlantique nord en rouge et les provinces à l'Est de la chaîne de l'Atlas en bleu. Par conséquent, **la troisième composante principale décrit l'effet de l'exposition à l'océan et la distance par rapport à la côte.**

**Figure 3.6 – Représentation graphique du vecteur propre numéro 3**



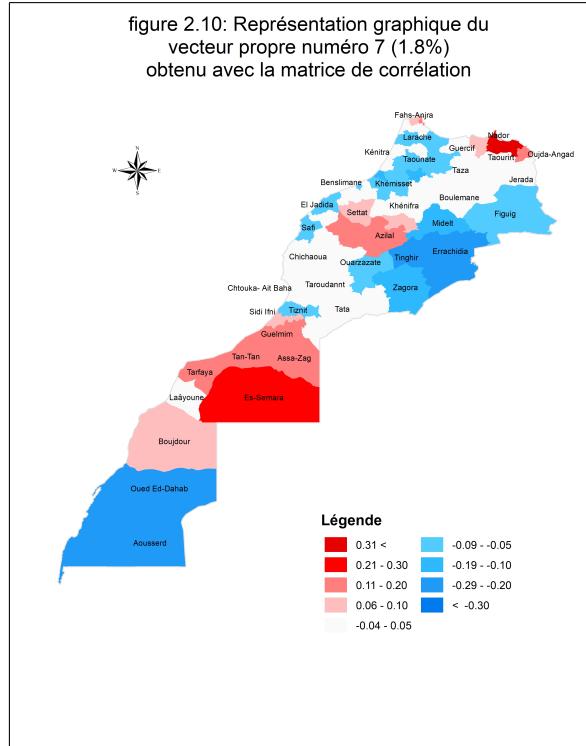
**La composante principale numéro 5 (CP5) :** Idem pour la 5ième CP, l'analyse de la représentation graphique du vecteur propre numéro 5 permet d'identifier une opposition entre les provinces montagneuses et le reste du pays. Donc **la cinquième composante principale décrit l'effet de l'altitude.**

**Figure 3.7 – Représentation graphique du vecteur propre numéro 5**



En conclusion, la première composante principale exprime l'aspect général des précipitations annuelles au Maroc, ce dernier coïncide avec la moyenne des précipitations annuelles. Tandis que la deuxième, la troisième et la cinquième composante principale décrivent les effets qui causent la variabilité des pluies sur les deux axes Nord-Sud et Ouest-Est, à savoir l'effet de la latitude, l'altitude et la distance par rapport à la côte. Donc, plus l'ordre de la composante augmente plus elle traite des phénomènes de plus en plus locaux, à titre d'exemple : le vecteur propre numéro 7 (figure 3.8), montre une opposition entre les provinces de Es-Semara, Berkane et Nador et les autres.

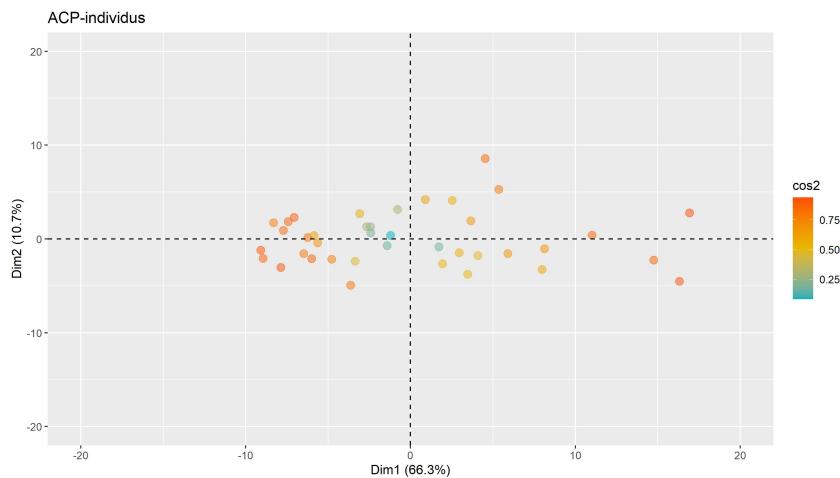
**Figure 3.8 – Représentation graphique du vecteur propre numéro 7**



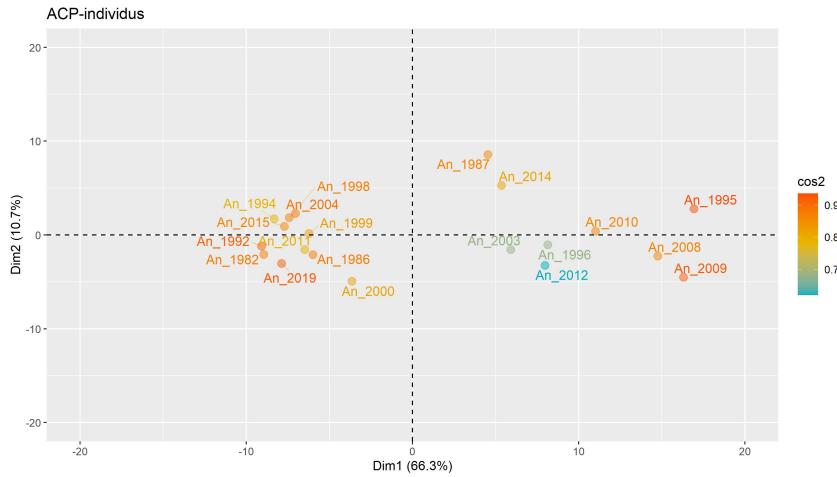
### 3.2.3 Etude des années

L'analyse des composantes principales et des vecteurs propres nous permettra dans cette section de lire facilement le graphe des individus projetés dans le plan formé par les deux premières composantes principales (figure 3.9 et 3.10). En effet, l'axe horizontale (CP1) exprime la moyenne des précipitations annuelles, alors que l'axe verticale (CP2) exprime la variabilité des pluies annuelles sur l'axe Nord-Sud du pays.

**Figure 3.9 – Projection des individus sur le plan (CP1, CP2)**



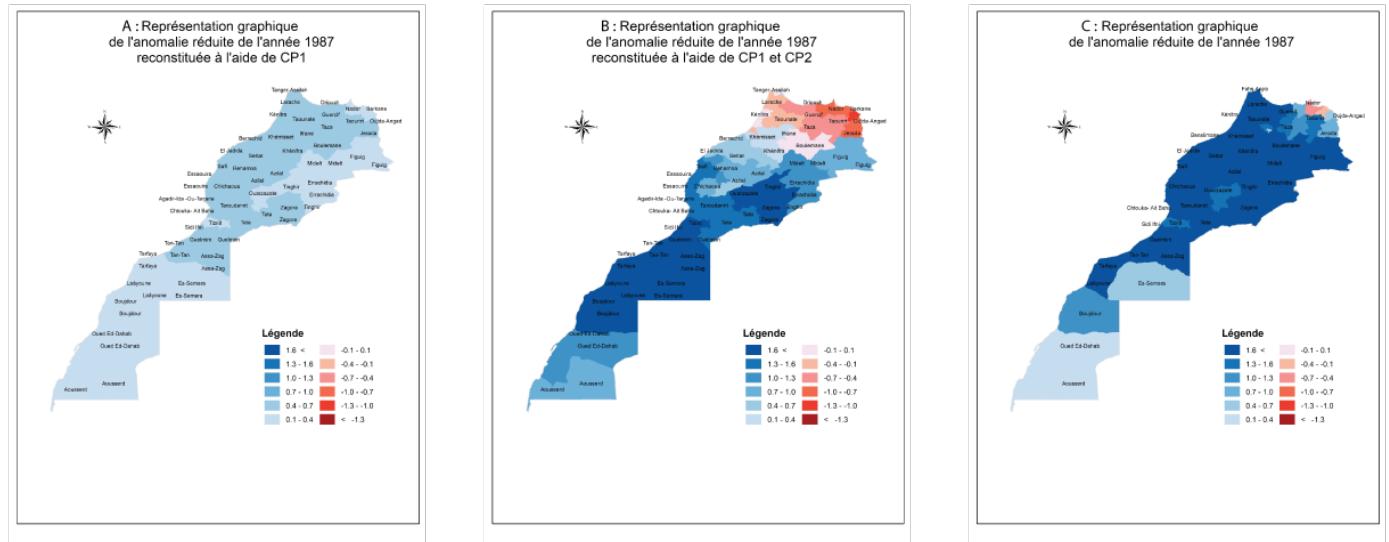
**Figure 3.10 – Les 20 premiers individus, classés par le cosinus carré, projetés dans le plan (CP1, CP2)**



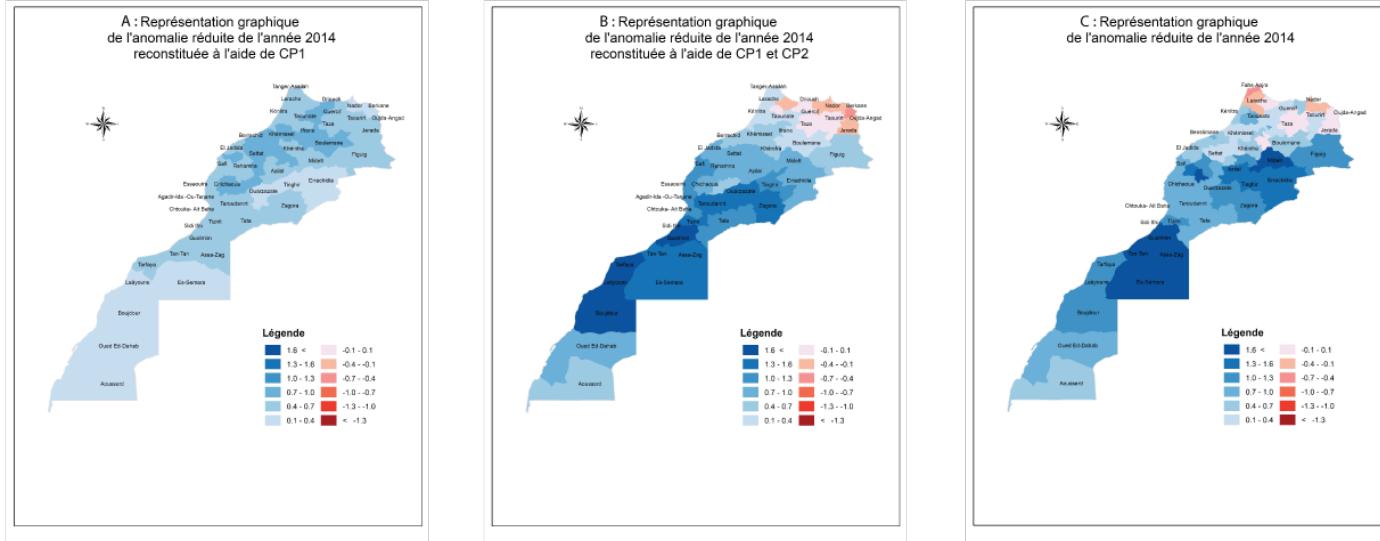
Le graphe ci-dessus permet de repérer facilement les années remarquables qui contribuent le plus à la formation des deux axes. En effet, les années 1995, 2009 et 2008 en opposition avec 1992, 1982 et 1994 sont les années qui contribuent le plus à la formation de CP1, ces six années sont exactement les mêmes présentées dans les table 2.1 (chapitre 1), comme étant les trois années les plus sèches et les trois les plus humides de la période 1981-2019 au Maroc. Sur l'axe CP2, les années 1987 et 2014 en opposition avec 2000 et 2009 sont les observations les plus remarquables qui méritent d'être analysées séparément :

**Les années 1987 et 2014 :** tout d'abord, ces deux observations appartiennent au demi-plan droit du repère donc leurs pluies provinciales sont audessus de leurs normales, ceci est bien illustré par les figures 3.11(A) et 3.12(A) où nous avons représenté l'anomalie réduite des précipitations provinciales des années 1987 et 2014 reconstituées à l'aide de CP1 seule. Ensuite, nous avons recalculer et tracer l'anomalie réduite de ces mêmes années, mais cette fois ci reconstituées à l'aide de CP1 et CP2, figures 3.11(B) et 3.12(B) Par conséquent, les effets apportés par l'ajout de CP2 sont : la création des écarts négatifs par rapport à la normale aux provinces du Nord et l'amplification des écarts positifs au provinces du Sud.

**Figure 3.11 – L'anomalie réduite de l'année 1987 reconstituée par CP1, CP1 et CP2 et observation complète.**



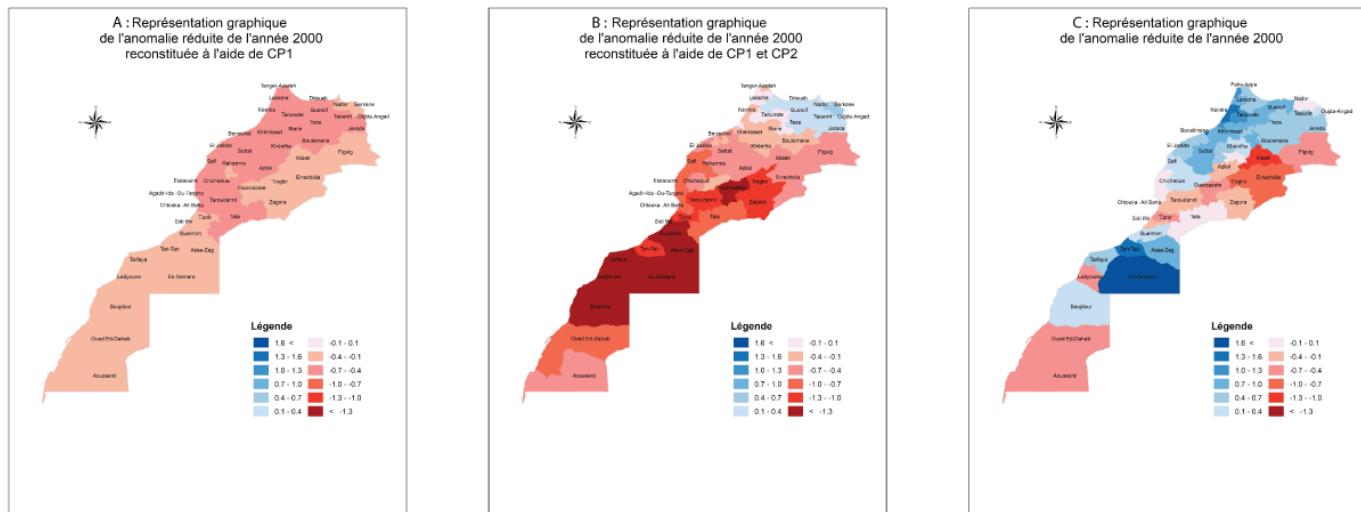
**Figure 3.12 – L'anomalie réduite de l'année 2014 reconstituée par CP1, CP1 et CP2 et observation complète.**



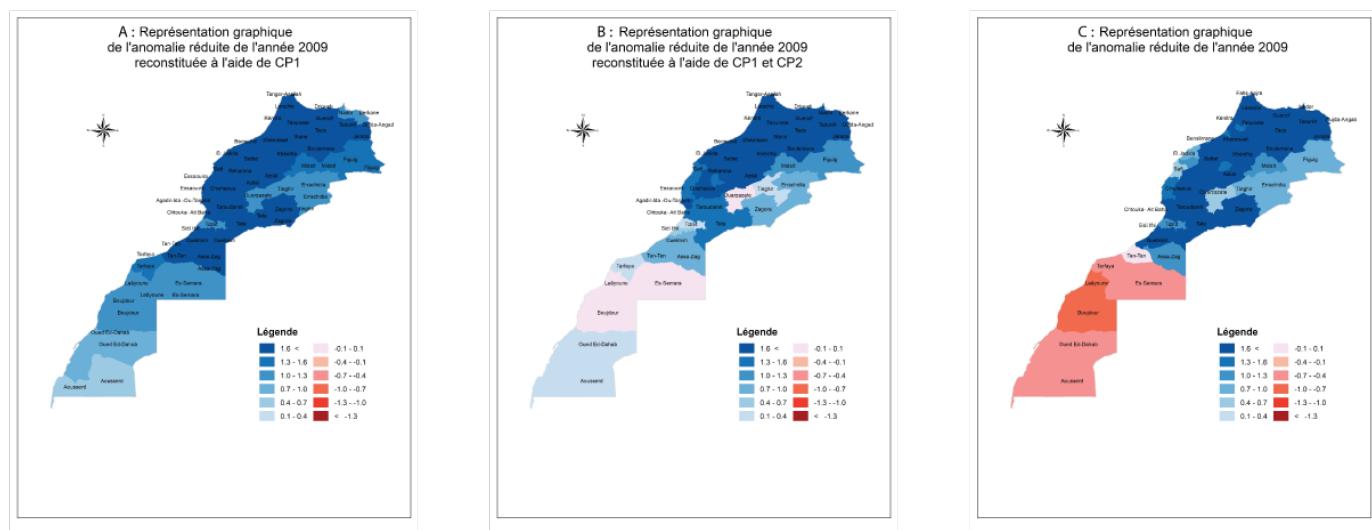
**Les années 2000 et 2009 :** idem, pour ces deux années, Nous avons reconstitué leurs précipitations premièrement à l'aide de CP1 ensuite à l'aide de CP1 et CP2, les résultats sont présentés en figures 3.13 et 3.14. Cette fois ci, l'observation 2009 appartient au demi-plan droit où les précipitations des observations sont au-dessus de leurs normales alors que l'observation 2000 appartient au demi-plan gauche où les précipitations des années sont au-dessous de leurs normales, ce qui explique la différence des couleurs entre les figures 3.13(A) et 3.14(A). Les effets apportés par l'ajout de CP2 cette fois sont l'inverse du premier cas. En effet, l'ajout de CP2 pour l'année 2000 a créé des écarts positifs par

rapport à la normale aux provinces du Nord et a amplifié les écarts négatifs des provinces du Sud. Tandis que, pour l'année 2009, l'ajout de CP2 a réduit les écarts positifs au Sud et a amplifié les écarts positifs au Nord.

**Figure 3.13 – L'anomalie réduite de l'année 2000 reconstituée par CP1, CP1 et CP2 et observation complète.**



**Figure 3.14 – L'anomalie réduite de l'année 2009 reconstituée par CP1, CP1 et CP2 et observation complète.**



En conclusion, l'axe horizontale (CP1) exprime une opposition entre les années humides (exemple : 1995 et 2009) et les années sèches (exemple : 1982 et 1994). Tandis que, l'axe verticale (CP2) exprime la variabilité des précipitations provinciales annuelles sur l'axe Nord-Sud, de telle sorte que le demi-plan supérieur contient les années où les

provinces du Nord sont au-dessous de leurs normales et celles du Sud sont au-dessus de leurs normales (exemple : 1987 et 2014), et l'inverse pour le demiplan inférieur (exemple : 2008 et 2009).

## Bibliographie

- [1] S.El khatri. Cours statistiques multidimensionnelles. 2010-2020.
- [2] S.Robin C.Duby. Cours analyse en composantes principales.

# Chapitre 4

## Classification automatique

### Sommaire

---

<b>4.1</b>	<b>Classification hiérarchique ascendante</b>	<b>34</b>
4.1.1	Présentation de la méthode	34
4.1.2	Classification des années	35
4.1.3	Classification des provinces	39
<b>4.2</b>	<b>Méthode des centres mobiles (k-means)</b>	<b>41</b>
4.2.1	Présentation de la méthode	41
4.2.2	Classification des années	42
4.2.3	Classification des provinces	44
<b>4.3</b>	<b>Classification d'une nouvelle observation</b>	<b>44</b>
4.3.1	Positionnement du problème	44
4.3.2	Solution	44
4.3.3	Résultats	45

---

La classification (clustering) est une méthode mathématique d'analyse de données : pour faciliter l'étude d'une population d'effectif important, on les regroupe en plusieurs classes de telle sorte que les individus d'une même classe soient le plus semblables possible et que les classes soient le plus distinctes possibles. Afin de réaliser une classification, il y a plusieurs méthodes (dont les résultats peuvent ne pas être les mêmes). Dans ce projet, nous utiliserons les deux méthodes les plus connues : la classification hiérarchique ascendante et la méthode des centres mobiles (K-means).

### 4.1 Classification hiérarchique ascendante

#### 4.1.1 Présentation de la méthode

Pour classifier une population d'effectif  $n$  dont les individus sont numérotés 1, 2, ...,  $n$ , on considère cette population comme la réunion de  $n$  classes à un seul élément et on regroupe progressivement les classes deux à deux selon l'algorithme suivant :

**Etape 1 :** Calculer la matrice des distances.

**Etape 2 :** Remplacer les deux individus de distance minimale par une classe (à 2 éléments) numéroté  $n+1$ . La population compte alors  $n - 1$  classes  $n - 2$  classes à un élément et une à 2 éléments).

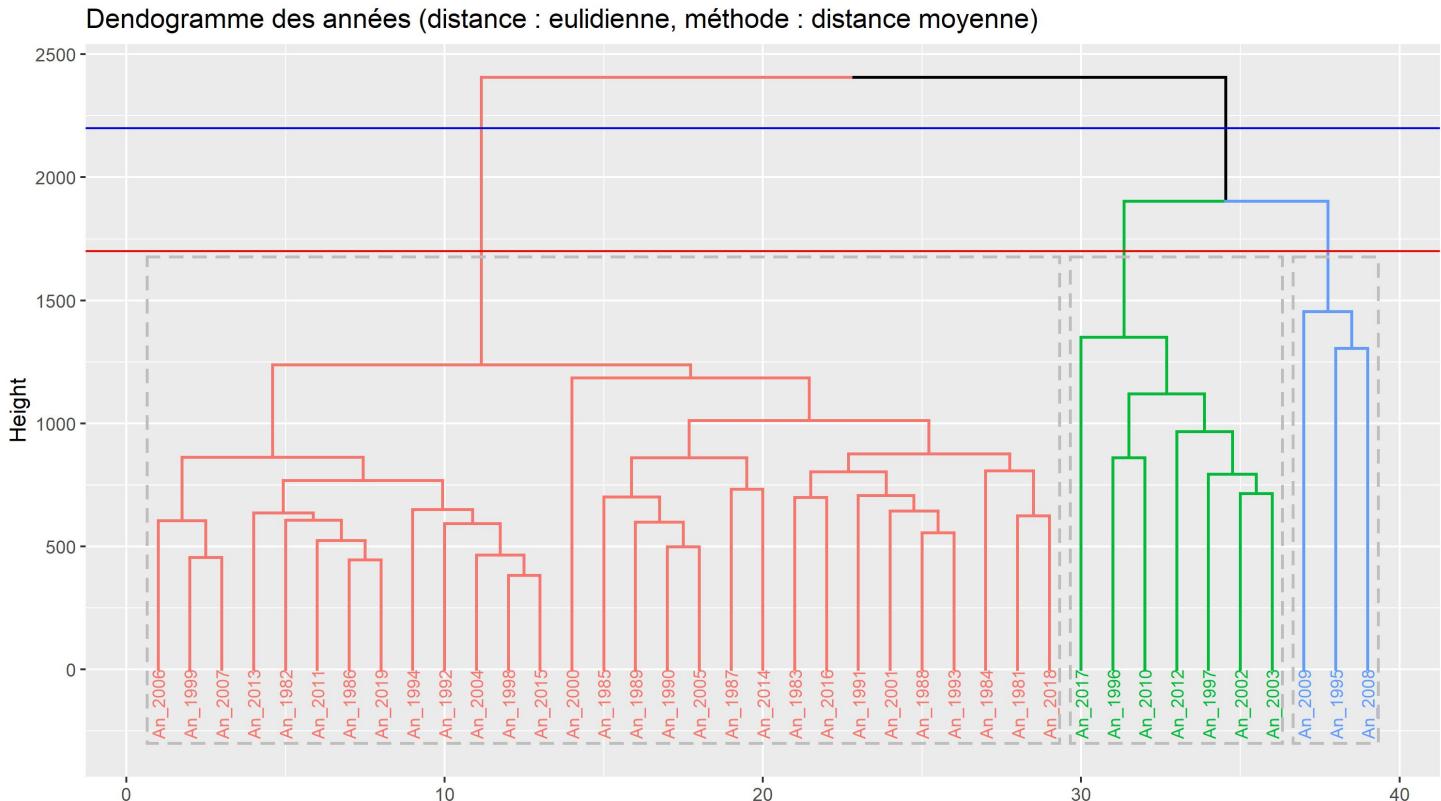
On peut donc recommencer à l'étape 1, recalculer la matrice des distances entre ces  $n - 1$  classes, en remplaçant si nécessaire "distance entre individus" par "écart entre classes", puis regrouper (étape 2) les deux classes de distance minimale en une seule. On appelle cela l'agglomération au plus proche voisin. Après  $n - 1$  itérations, tous les individus seront regroupés en une classe unique.

En se basant sur cet algorithme on construit un arbre (appelé dendrogramme) permettant de visualiser toutes les étapes de l'algorithme, ainsi que les individus regroupés en premier et ceux regroupés en dernier. Le choix de nombre de classes à garder (où couper l'arbre) peut-être soit : un choix objectif, alors on coupe l'arbre où on a le plus long saut des troncs, ou un choix subjectif, qui suppose une connaissance profonde du jeu de données et de nombre de classes possibles. [1] [2]

Afin d'implémenter la classification hiérarchique, nous avons utilisé la fonction, prédéfinie en R, hclust. En argument de hclust, nous avons choisi d'utiliser la distance euclidienne comme distance de comparaison et la distance moyenne comme critère d'agglomération.

#### 4.1.2 Classification des années

**Figure 4.1 – Dendrogramme des années**



D'après le dendrogramme, le choix objectif de nombre de classes peut être soit 2 classes soit 3, vu que les deux troncs correspondants sont les plus long et sont presque égaux. Si on veut choisir subjectivement le nombre de classes, il serait judicieux de prendre 3 classes : classe des années humides, classe des années sèches et classe des années normales.

Le choix de 3 classes nous donne une distribution assez intéressante : En effet, la classe 1 englobe plus de 75% de l'effectif et la classe 2 moins de 8 % (table 4.1), cela permet de conclure que les années au Maroc sont majoritairement des années de la classe 1 et le climat du Maroc est déterminé par le type de ces années. Tandis que, l'occurrence des observations comme celle de la classe 2 est rare (nombre des années classe 2/ nombre des années classe1 =0.1).

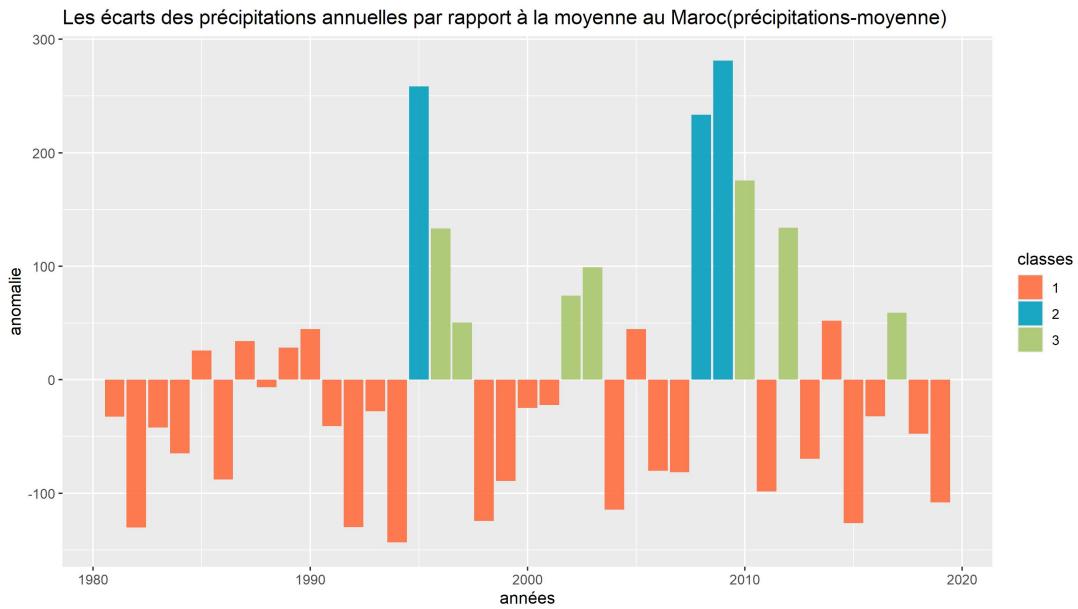
**Table 4.1 – Les classes et leur effectifs, obtenues par CHA**

groupe	N-individus	pourcentage
1	29	74.35
2	3	7.69
3	7	17.94

L'analyse de l'histogramme des écarts de précipitations par rapport à la moyenne (figure 4.2) ainsi que les représentations graphiques des étalons réels (figure 4.3) permet d'identifier le type des années dans chacune des classes.

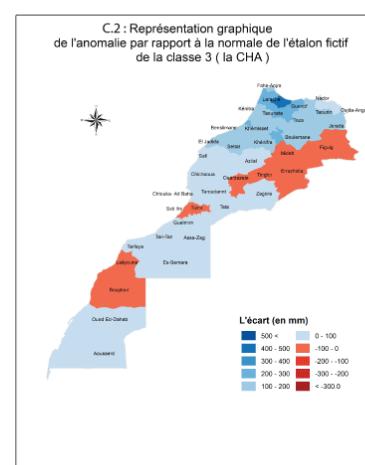
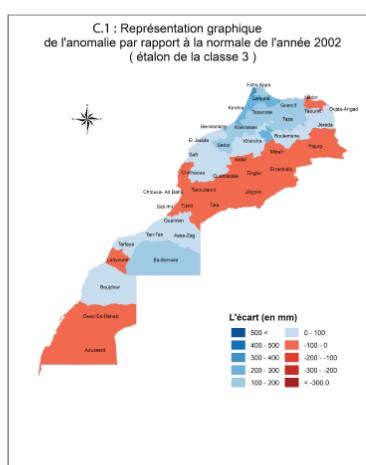
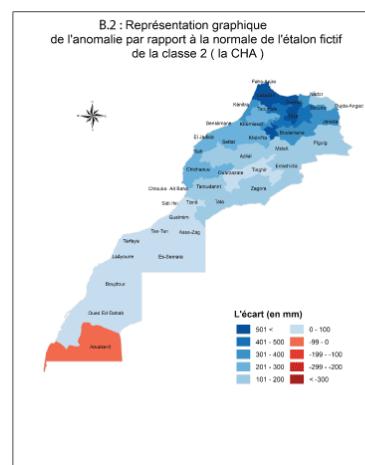
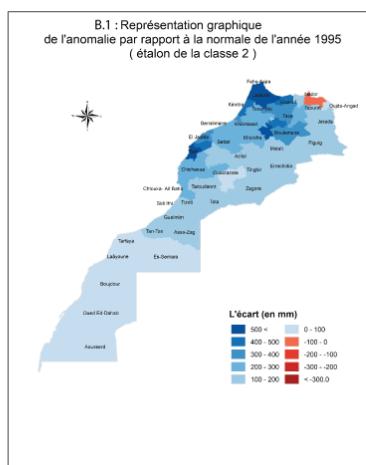
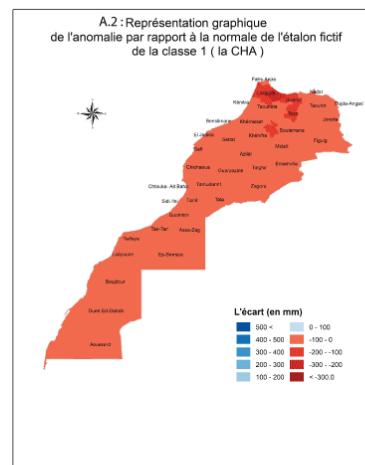
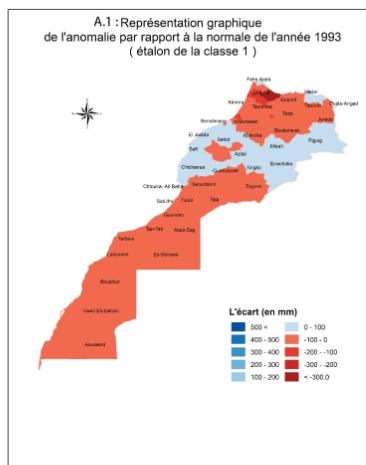
En effet, on constate sur la figure (4.2) que pour la majorité des années les précipitations sont au-dessous de la moyenne, et si on considère la moyenne comme étant seuil de sécheresse, on peut qualifier ces années d'années sèches. En fait, ces années sont toutes incluses dans la classe1. D'autre part, les 3 années ayant les écarts positifs par rapport à la moyenne les plus grands correspondent aux années de la classe 2.

**Figure 4.2 – Les écarts des précipitations annuelles par rapport à la moyenne sur la période 1981-2019, au Maroc**



On conclusion, la classe 1 correspond aux années sèches, la classe 2 contient les années humides et la classe 3 contient les années normales. Cette conclusion est tout à fait en accord avec les représentations graphiques de l'anomalie par rapport à la normale des étalons réels des classes, figures ci-dessous. En effet, la figure 4.3(A) montre que les pluies annuelles dans la majorité des provinces sont au-dessous de la normale, tandis que dans la figure 4.3(C) toutes les provinces sont au-dessus de la normale.

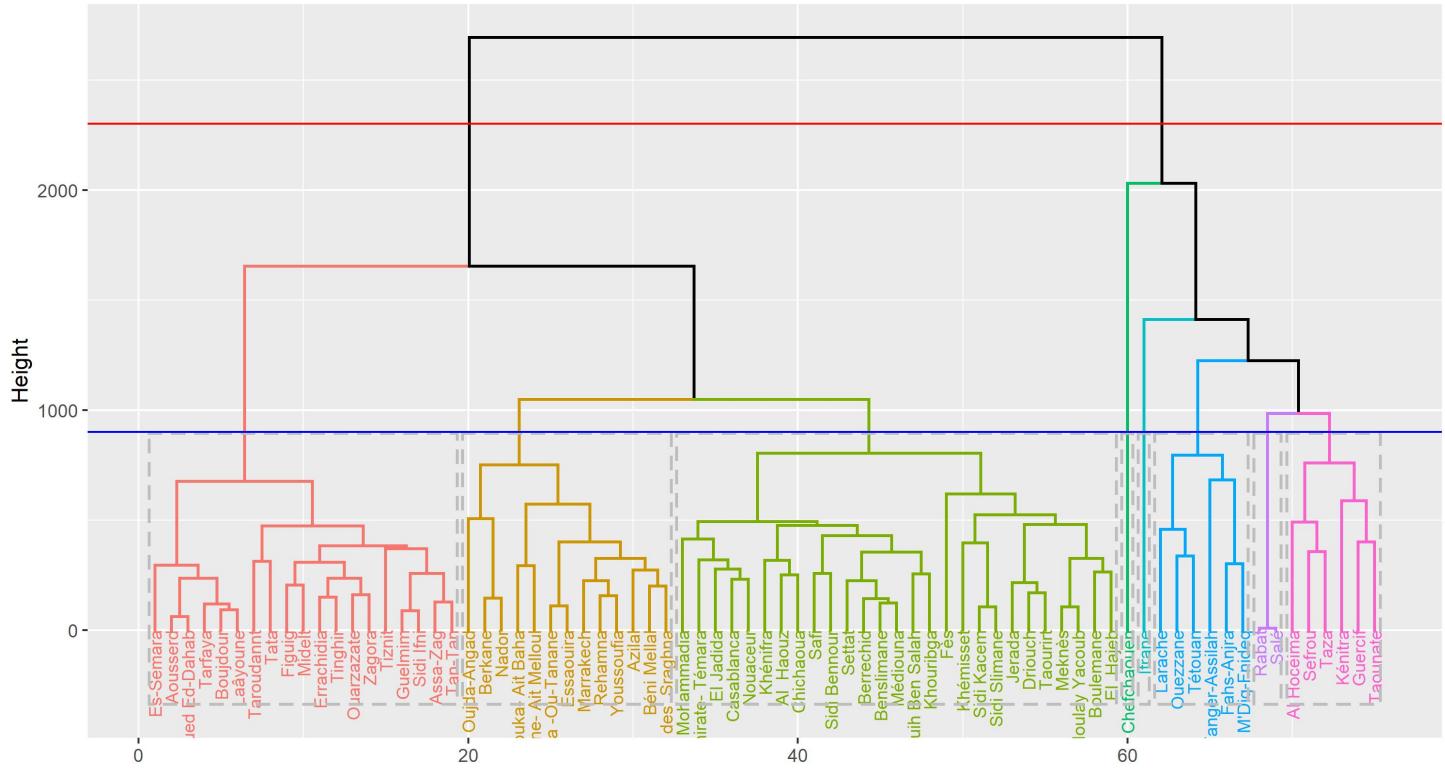
**Figure 4.3 – Représentations graphiques de l'anomalie par rapport à la normale des étalons réels et fictifs des classes fournies par la CHA des années**



### 4.1.3 Classification des provinces

**Figure 4.4 – Dendrogramme des provinces**

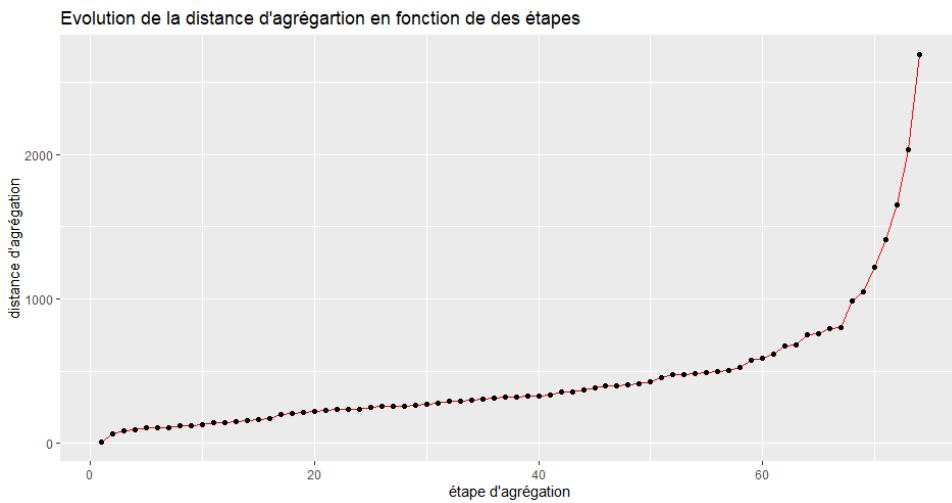
Dendrogramme des provinces (distance : eulidiennne, méthode : distance moyenne)



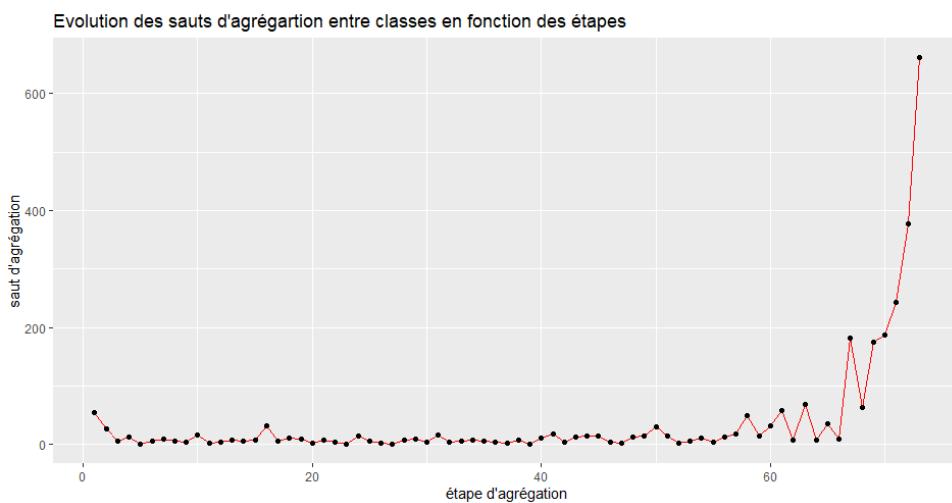
D'après le dendrogramme, le choix objectif de nombre de classes est 2. Si on veut choisir subjectivement le nombre de classes, on doit prendre en considération les différentes influences que subit le Maroc : les influences océaniques, méditerranéennes, montagnardes, continentales et désertiques. Ainsi, d'une part, le nombre de classes doit dépasser 2 pour mettre en évidence les différentes régions climatiques. D'autre part, la distance d'agrégation doit être suffisamment grande afin de distinguer entre les classes.

La méthode du « coude » appliquée aux courbes (figure 4.5 et 4.6), montre que le meilleur choix est 8 classes, tout en respectant les conditions précédemment mentionnées.

**Figure 4.5 – Evolution de la distance d'agrégation en fonction des étapes**

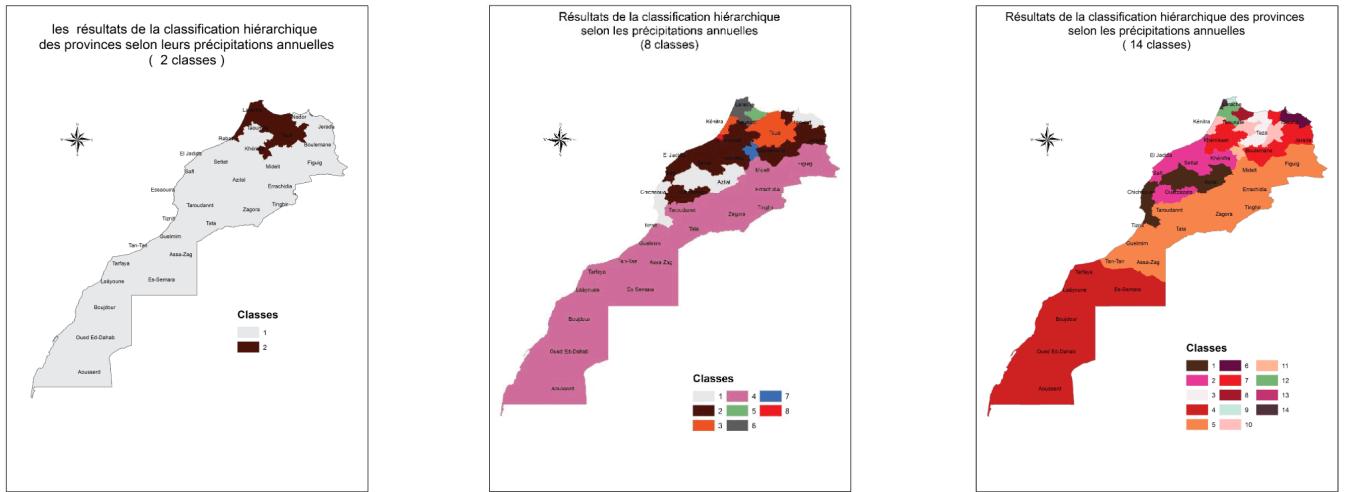


**Figure 4.6 – Evolution du saut d'agrégation entre deux classes en fonction des étapes**



La classification des provinces à l'aide de la CHA, nous donne une répartition très familière, c'est celle de la moyenne des précipitations annuelles au Maroc. Une répartition où les provinces de l'extrême Nord, caractérisées par des pluies annuelles très importantes, sont séparées des autres provinces du pays. Et plus en augmente le nombre de classes plus en s'approche de la répartition géographique de la moyenne des précipitations annuelles, cela est parfaitement illustré par la figure 4.7.

**Figure 4.7 – Représentation graphique des résultats de la classification hiérarchique des provinces**



## 4.2 Méthode des centres mobiles (k-means)

### 4.2.1 Présentation de la méthode

La méthode des centres mobiles s'applique lorsque l'on sait à l'avance combien de classes on veut obtenir. Appelons  $k$  ce nombre de classes. L'algorithme est le suivant :

Etape 0 : Pour initialiser l'algorithme, on tire au hasard  $k$  individus appartenant à la population,  $C_1^0, C_2^0, \dots, C_K^0$  : ce sont les  $k$  centres initiaux. On notera que l'indice numérote les différents centres et l'exposant indique qu'il s'agit des  $k$  centres initiaux. On choisit aussi une distance entre individus. On va ensuite répéter un grand nombre de fois les deux étapes suivantes :

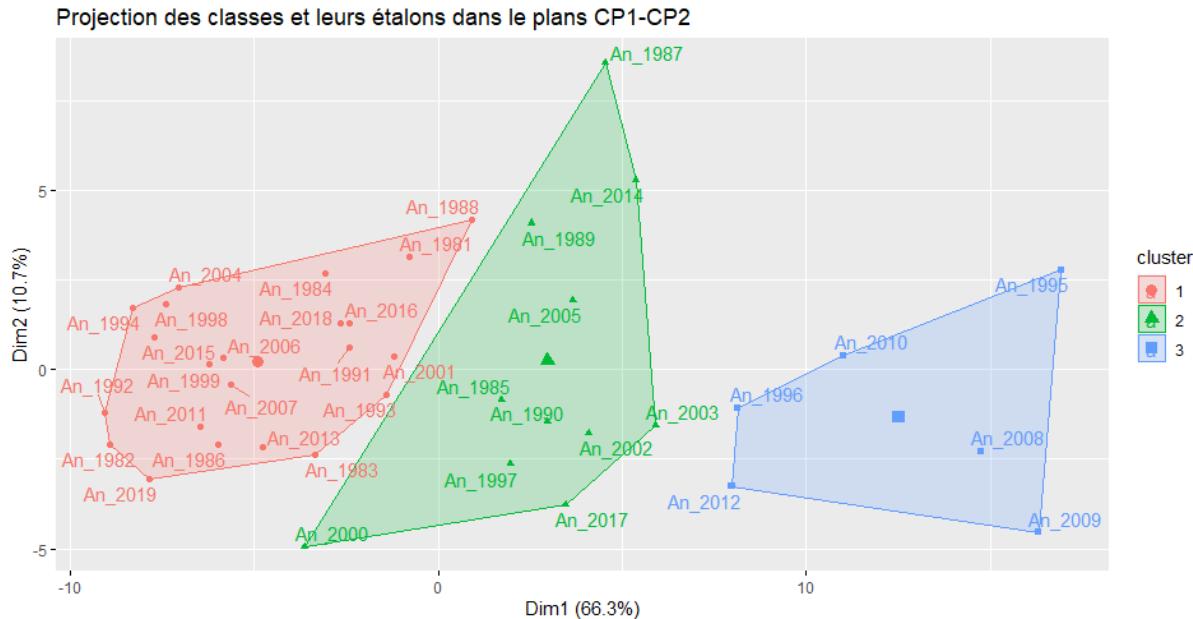
**Etape 1 :** Constitution de classes : On répartit l'ensemble des individus en  $k$  classes  $\Gamma_0^0, \Gamma_1^0, \dots, \Gamma_k^0$  en regroupant autour de chaque centre  $C_i^0$  pour  $i = 1, \dots, k$  l'ensemble des individus qui sont plus proches du centre  $C_i^0$  que des autres centres  $C_j^0$  pour  $j$  (au sens de la distance choisie).

**Etape 2 :** Calcul des nouveaux centres : On détermine les centres de gravité  $G_1, G_2, \dots, G_k$  des  $k$  classes ainsi obtenues et on désigne ces points comme les nouveaux centres  $C_1^1 = G_1, C_2^1 = G_2, C_k^1 = G_k$ .

Répétition des étapes 1 et 2 : on répète ces deux étapes jusqu'à la stabilisation de l'algorithme, c'est-à-dire jusqu'à ce que le découpage en classes obtenu ne soit (presque) plus modifié par une itération supplémentaire. [2] [3]

#### 4.2.2 Classification des années

**Figure 4.8 – Projection des classes fournies par la méthode Kmeans sur le plan (CP1, CP2)**



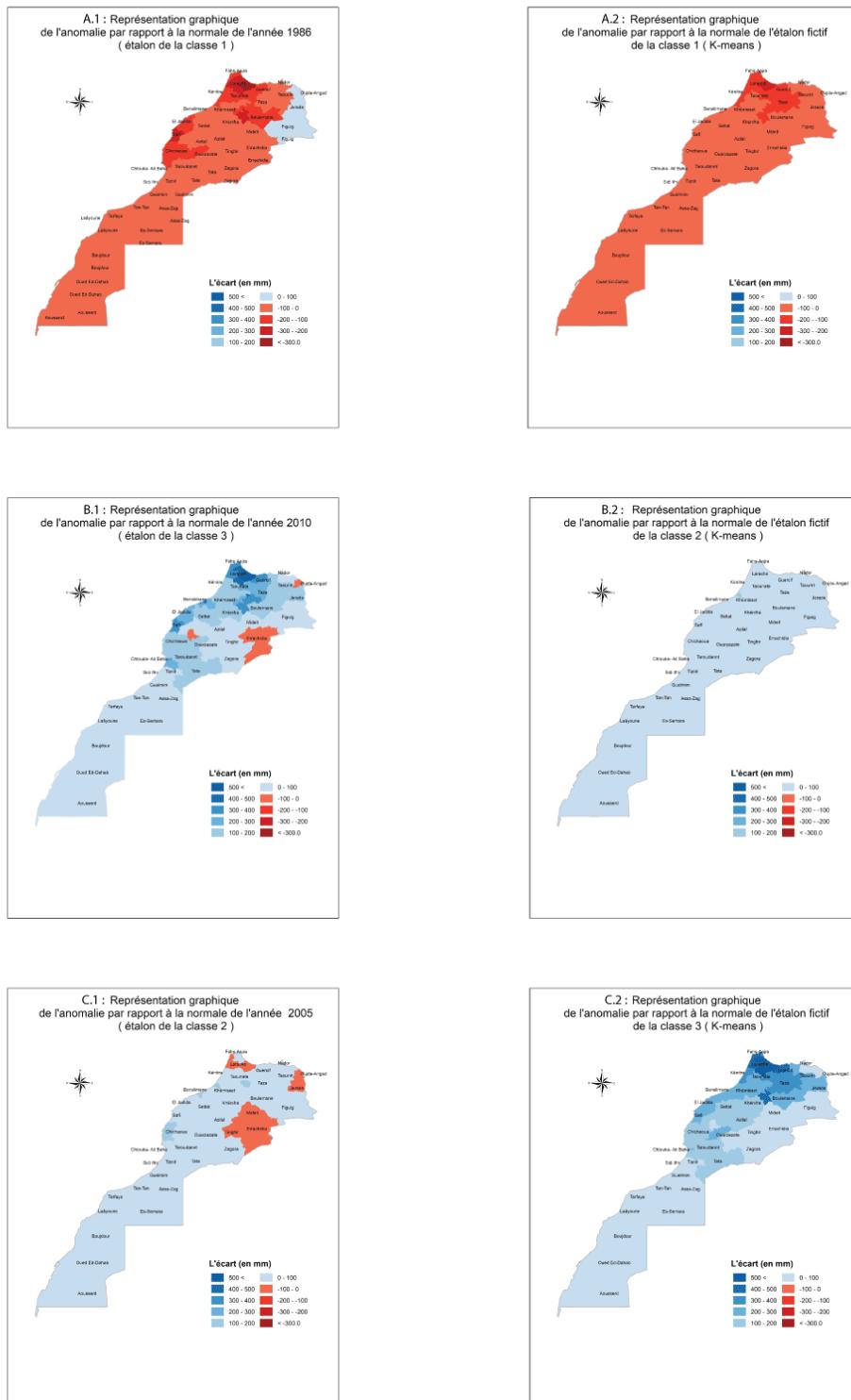
Comparés aux résultats de la CHA, les effectifs des classes ont peu changé : les effectifs des classes 2 et 3 en CHA (2 et 1 en K-means) ont augmenté au détriment de l'effectif de la classe 1 en CHA (classe 3 en K-means). La table 4.2 illustre, clairement, les différences entre les résultats des deux méthodes. Malgré cette différence en résultats, les aspects d'écrit par les classes restent les mêmes. En effet, la classe 3 correspond aux années sèches, la classe 2 contient les années humides et la classe 1 contient les années normales.

**Table 4.2 – Comparaison des classes fournies par les deux méthodes : k-means et CHA**

Groupe-CHA	N-individus	Pourcentage %	Groupe K-means	N-individus	pourcentage %
1	29	74.35	1	22	56.41
2	3	7.69	3	6	15.38
3	7	17.94	2	11	28.2

Comme conséquence du changement des effectifs des classes, les étalons ont aussi changé figure (4.9). Mais, ils restent représentatifs de la nature des années contenues par chacune des classes : l'an 1986 représente les années sèches de la classe 1, 2005 représente les années normales de la classe 2 en K-means (classe 3 en CHA) et 2010 est l'étalon des années humides de la classe 3 en K-means (classe 2 en CHA).

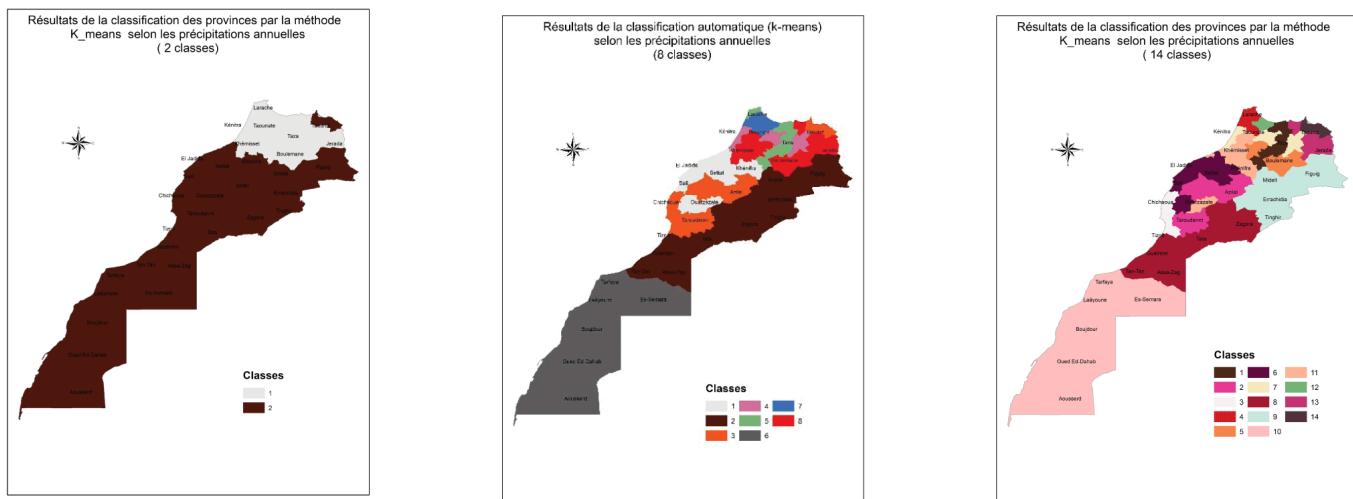
**Figure 4.9 – Représentations graphiques de l'anomalie par rapport à la normale des étalons réels et fictifs des classes fournies par la méthode K-means**



### 4.2.3 Classification des provinces

La classification des provinces par la méthode K-means fournit des résultats un peu différents par rapport à ceux déjà obtenus par la CHA, mais en global les deux classifications sont analogues et fournissent une classification qui tend vers le profil de la moyenne des précipitations annuelles, par conséquent les deux méthodes classent les provinces selon leurs moyennes des précipitations annuelles.

**Figure 4.10 – Représentation graphique des résultats de la méthode Kmeans appliquée aux provinces**



## 4.3 Classification d'une nouvelle observation

### 4.3.1 Positionnement du problème

Supposant qu'on dispose des cumuls de précipitations au Maroc pour l'années 2020 (observation 2020), et on souhaite savoir les années les plus proches de 2020 ainsi que la classe où elle appartient, en se basant sur les résultats des classifications précédentes.

### 4.3.2 Solution

```

1 #fonction pour déterminer les Nb vecteurs les plus proches
2 plus.proche<- function(vect , matrice, Nb){
3   #la matrice des distances entre les vecteurs de matrice et vect
4   combin<-cbind(matrice,vect)
5   distance<- data.frame(as.matrix(dist(t(combin),method="euclidean")))
6   #trier la matrice distance par vect (décroissant)
7   distance= distance[order(distance$vect),]
8   #éliminer vect de la matrice distance
9   distance<-(distance[ !(rownames(distance)%in% "vect") , !(names(distance)%in% "vect"))])
10  # renvoyer les Nb les plus proche de vect
11  rownames(head(distance,Nb))
12 }
```

### 4.3.3 Résultats

Les observations des cumuls de 2020 ne sont pas encore disponibles, vu que l'année n'est pas encore achevée. Par conséquent, on utilisera la technique de la moyenne mobile pour faire une estimation simple de l'observation 2020.

L'observation 2020 (estimée) :

**Table 4.3 – Les cumuls de précipitations pour l'années 2020 estimée par la moyenne mobile**

provinces	précipitation (mm)	provinces	précipitation (mm)	provinces	précipitation (mm)	provinces	précipitation (mm)
Agadir-Ida-Ou-Tanane	259.0	M'Diq-Fnideq	520.0	Fès	541.1	Sidi Kacem	436.2
Al Haouz	384.1	Médiouna	338.3	Figuig	165.2	Sidi Slimane	424.5
Al Hoceima	688.9	Meknès	480.6	Fquih Ben Salah	300.1	Skhirate- Témara	397.7
Aousserd	48.7	Midelt	170.1	Guelmim	116.5	Tanger-Assilah	630.3
Assa-Zag	106.5	Mohammadia	418.4	Guercif	594.9	Tan-Tan	90.3
Azilal	236.3	Moulay Yacoub	490.2	Ifra	819.5	Taounate	559.9
Béni Mellal	275.2	Nador	284.2	Inezgane- Ait Melloul	245.1	Taourirt	437.7
Benslimane	321.3	Nouaceur	370.6	Jerada	459.6	Tarfaya	40.7
Berkane	280.8	Ouarazate	114.1	Kénitra	562.4	Taroudannt	194.6
Berrechid	313.2	Oued Ed-Dahab	44.5	Khémisset	421.1	Tata	152.3
Boujdour	33.2	Ouezzane	649.6	Khénifra	392.6	Taza	657.8
Boulemane	477.1	Oujda-Angad	268.8	Khouribga	323.1	Tétouan	641.4
Casablanca	408.1	Rabat	547.8	Laâyoune	30.3	Tinghir	126.6
Chefchaouen	812.2	Rehamna	245.2	Larache	662.0	Tiznit	154.3
Chichaoua	349.3	Safi	349.5	Marrakech	221.6	Youssoufia	244.5
Chtouka- Aït Baha	211.1	Salé	546.8	El Kelâa des Sraghna	263.0	Zagora	127.8
Driouch	444.8	Sefrou	697.5	Errachidia	143.5	Fahs-Anjra	562.1
El Hajeb	501.8	Settat	293.4	Essaouira	269.4	Sidi Ifni	125.3
El Jadida	384.9	Sidi Bennour	331.5	Es-Semara	56.8		

Les 3 premières années les plus proches de 2020 sont :

> plus.proche(An\_2020,data,3)

[1] "An\_2005" "An\_1985" "An\_2016"

L'an 2020 appartient à la classe (CHA) :

> plus.proche(An\_2020, etalons\_CHA,1)

[1] "classe 1"

L'an 2020 appartient à la classe (K-means) :

> plus.proche(An\_2020,etalons\_kmeans,1)

[1] "classe 2"

Alors, l'année 2020 est sèche selon la CHA et normale selon la CA( k-means).

## Bibliographie

- [1] Classification automatique de données par la méthode hiérarchique ascendante. 2011-2012.
- [2] S.El khatri. Cours statistiques multidimensionnelles. 2010-2020.
- [3] Cours 8 : Classification automatique de données par la méthode des centres mobiles. 2011-2012.

# Conclusion

En guise de conclusion, les techniques employées dans ce projet nous ont permis d'acquérir une connaissance profonde des caractéristiques et des modalités (paterns) de notre jeu de données.

En effet, l'ACP a permis de réduire les 75 variables en 8 composantes principales qui résument plus de 93% de l'information totale. A l'aide de l'ACP, l'information sur la moyenne est décrite par une seule nouvelle variable qui est la première composante principale. Idem, pour la variabilité des précipitations : du Nord vers le sud sous l'effet de la latitude et de l'Ouest vers l'Est sous l'effet de l'altitude ou l'éloignement par rapport à la côte, ces caractéristiques sont résumées par les premiers vecteurs propres (paterns).

La classification automatique, quant à elle, a servi pour regrouper les provinces ainsi que les années dans des classes où les individus se ressemblent le maximum possible en termes de la caractéristique de leurs précipitations. La classification des années, soit par la classification hiérarchique ascendante ou par la méthode K-means, a permis l'identification et le partitionnement des individus en trois classes : classe des années sèches, classe des années humides et la classe des années moyenne, ce regroupement nous permet de tirer une conclusion assez intéressante sur le climat du Maroc : la classe des années sèches contient plus de 75% des années, alors que la classe des années humides ne représente que 7.7% des années . Donc le climat du Maroc est généralement un climat sec avec des périodes de sécheresse très longues : par exemple, la plus longue période c'est celle de 1981 à 1994, le tableau ci-dessous illustre cela :

**Table 4.4 – Classification des années selon la quantité des précipitations annuelles sur la période 1981-2019**

An_1981	sèche	An_1991	sèche	An_2001	sèche	An_2011	sèche
An_1982	sèche	An_1992	sèche	An_2002	normale	An_2012	normale
An_1983	sèche	An_1993	sèche	An_2003	normale	An_2013	sèche
An_1984	sèche	An_1994	sèche	An_2004	sèche	An_2014	sèche
An_1985	sèche	An_1995	humide	An_2005	sèche	An_2015	sèche
An_1986	sèche	An_1996	normale	An_2006	sèche	An_2016	sèche
An_1987	sèche	An_1997	normale	An_2007	sèche	An_2017	normale
An_1988	sèche	An_1998	sèche	An_2008	humide	An_2018	sèche
An_1989	sèche	An_1999	sèche	An_2009	humide	An_2019	sèche
An_1990	sèche	An_2000	sèche	An_2010	normale		

La classification des provinces par les deux méthodes, a fournie des résultats très

proches et qui partitionnent les provinces du royaume, selon la moyenne des précipitations annuelles, en des classes où les provinces ont des précipitations annuelles moyennes très faibles, voire rares, ces classes regroupent les provinces sahariennes et quelques provinces du Sud-Est du pays. Ainsi que des classes contenant des provinces où les pluies sont abondantes et les précipitations annuelles sont très importantes, ces classes regroupent les provinces du Nord et les provinces côtières.

# Annexe A : Les scripts R

## Chapitre 1 : Analyse exploratoire

```
:::::::::::::::::::  
# Libraries  
pack<-c("factoextra", "plotrix") # les packages nécessaires  
  
if (!require("easypackages")) installed.packages()  
packages(pack)  
:::::::::::::::::::  
#importation des données  
data<- read.csv("RR_data.csv", header=T, row.names = 1)  
  
#caractéristiques des précipitations par provinces  
#variabilité des précipitations annuelles par province  
M<-apply(data, 1, mean)      # vecteur moyen  
S<-apply(data, 1, sd)        # vecteur ecart-type  
Cv<-round((S/M)*100 , 3)    # vecteur coefficient de variation  
sommaire=cbind(M,S,Cv)  
sommaire<-`colnames<-`(sommaire,c("moyenne","écart-type","coef de variation %"  
sommaire  
#variabilité des moyennes provinciales au maroc  
MM<-mean(M)                  # la moyenne des moyennes provinciales  
SM<-sd(M)                    # ecart-type  
CvM<-round((SM/MM)*100 , 3)   # coefficient de variation  
Msommaire=cbind(MM,SM,CvM)  
Msommaire<-`colnames<-`(Msommaire,c("moyenne","écart-type","coef de variation %"  
Msommaire<-`rownames<-`(Msommaire,"moyenne")  
Msommaire  
sommaire<-rbind(sommaire,Msommaire)  
#caractéristiques des précipitations par année (moyenne sur le maroc)  
moy_maroc<- apply(data, 2, mean)  
# anomalie de la série des précepeitaion annuelles sur le maroc  
moy_moy<- replicate(length(moy_maroc),mean(moy_maroc))  
anomal_add=data.frame(anomalie=moy_maroc-moy_moy,années=1981:2019)  
ggplot(data=anomal_add, aes(x= années, y=anomalie), fill= anomalie) +  
  geom_bar(stat="identity", aes(fill = anomalie))+  
  scale_fill_gradient2(low = "red", high = "blue", mid = "#fee0d2", midpoint =  
  labs(title="Les écarts des précipitations annuelles par rapport à la moyenne  
précipitations-moyenne)",  
       x="années")
```

## Chapitre 2 : Analyse en composantes principales

```
#Application de l'ACP
data<- t(data)
View(data)
res.acp <- prcomp(data, center=TRUE, scale = TRUE)      #implémentation de l'ACP

# valeurs propres
#####
#la courbe des valeurs propres (graphe en coude)
fviz_eig(res.acp,linecolor="red",barfill = "#00AFBB", barcolor = "#00AFBB",
          title ="graphe des valeurs propres",
          x = "Composantes principales",
          y = "Pourcentage de variance expliquée",
          ncp=20)+theme_get()

# le pourcentage de variance expliquée
eig.val <- get_eigenvalue(res.acp)
View(eig.val)

# vecteurs propres
#####
vct_pr<-as.matrix(res.acp$rotation)

# individus
#####
# Resultats pour les individus
res.ind <- get_pca_ind(res.acp)
res.ind$coord           # Coordonnées
res.ind$contrib          # Contributions au CPs
res.ind$cos2             # Qualité de la représentation
# projection des individus dans les plans des CPs
N_CP=4
for (i in 1:N_CP)
{
  for (j in i:N_CP)
  {
    if (i!=j){
      fviz_pca_ind(res.acp,
                    axes = c(i,j),
                    geom.ind = c("point"),
                    select.ind = list(contrib = 20),
                    alpha.ind = 0.5,
                    pointsize=3,
                    col.ind = "cos2", # colorer selon la qualité de la représentation
```

```

        gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
        repel = TRUE      # éviter les text superposés
    )+ theme_grey()+
    labs(title ="ACP-individus")+
    ylim(-20,20)+ xlim(-20,20)
}
}

# Contributions des individus à CPs
for (i in 1:N_CP) {
  titre= paste("Contribution des individus à la formation de CP",i)
  fviz_contrib(res.acp, choice = "ind", axes = i,
               title =titre)
}

# variables
#####
# Resultats des variables
res.var <- get_pca_var(res.acp)
res.var$coord          # coordonnées
res.var$contrib         # Contributions au CPs en %
res.var$cos2            # Qualité de la représentation
# projection des variables dans le plan des deux premières comp. princip
N_CP=5
for (i in 1:N_CP)
{
  for (j in i:N_CP)
  {
    if (i!=j){
      fviz_pca_var(res.acp,
                    axes = c(i,j),
                    select.var = list(cos2 =30),
                    col.var = "cos2", # colorer selon la contribution
                    gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
                    geom.var = c('point','text'),
                    title ="Variables-ACP(10 premières) ",
                    repel = TRUE,
                    )+ theme_get()

    }
  }
}

```

```

#projection des variables remarquables
fviz_pca_var(res.acp,
              select.var=list(name=c("moy","An_2008","An_2005","An_1999")),
              col.var = "contrib", # colorer selon la contributions dans les CPs
              gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
              repel = TRUE , # éviter les text superposés
              geom.var = c('point','text'),
              pointsize=2,
              title ="Variables-ACP",

)+ theme_get()
# Contributions des toutes les variables à PCs
for (i in 1:N_CP) {
  titre= paste("Contribution des variables à la formation de CP",i)
  fviz_contrib(res.acp, choice = "var", axes = i,
               title =titre)+ylim(0,20)
}

#les CPs
#####
#tracé des 4 1ères CPs
N_CP=4
z= res.acp$x      # matrice des coordonnées sur les CPs
years=39
x = seq(as.Date("1981-1-1"), by = "year", length.out = years)
df_plt<-data.frame(x,z[,1:4])
colnames(df_plt)=c("années","CP1","CP2","CP3","CP4")
df_plt_fus <- reshape2::melt(df_plt, id.var='années')
ggplot(data = df_plt_fus,
        mapping = aes(x = années, y = value, col=variable)) +
  geom_line()+geom_point() +
  facet_grid(rows = vars(variable))+geom_hline(yintercept = 0)+
  labs(title = "Evolution temporelle des 4 premières CPs",
       x = "années",
       y = "Valeurs")

```

## Chapitre 3 : Classification automatique

```
#Application de l'ACP
data<- t(data)
View(data)
res.acp <- prcomp(data, center=TRUE, scale = TRUE)      #implémentation de l'ACP

# valeurs propres
#####
#la courbe des valeurs propres (graphe en coude)
fviz_eig(res.acp,linecolor="red",barfill = "#00AFBB", barcolor = "#00AFBB",
          title ="graphe des valeurs propres",
          x = "Composantes principales",
          y = "Pourcentage de variance expliquée",
          ncp=20)+theme_get()
# le pourcentage de variance expliquée
eig.val <- get_eigenvalue(res.acp)
View(eig.val)

# vecteurs propres
#####
vct_pr<-as.matrix(res.acp$rotation)

# individus
#####
# Resultats pour les individus
res.ind <- get_pca_ind(res.acp)
res.ind$coord           # Coordonnées
res.ind$contrib          # Contributions au CPs
res.ind$cos2             # Qualité de la représentation
# projection des individus dans les plans des CPs
N_CP=4
for (i in 1:N_CP)
{
  for (j in i:N_CP)
  {
    if (i!=j){
      fviz_pca_ind(res.acp,
                    axes = c(i,j),
                    geom.ind = c("point"),
                    select.ind = list(contrib = 20),
                    alpha.ind = 0.5,
                    pointsize=3,
                    col.ind = "cos2", # colorer selon la qualité de la représentation
```

```

        gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
        repel = TRUE      # éviter les text superposés
    )+ theme_grey()+
    labs(title ="ACP-individus")+
    ylim(-20,20)+ xlim(-20,20)
}
}
}

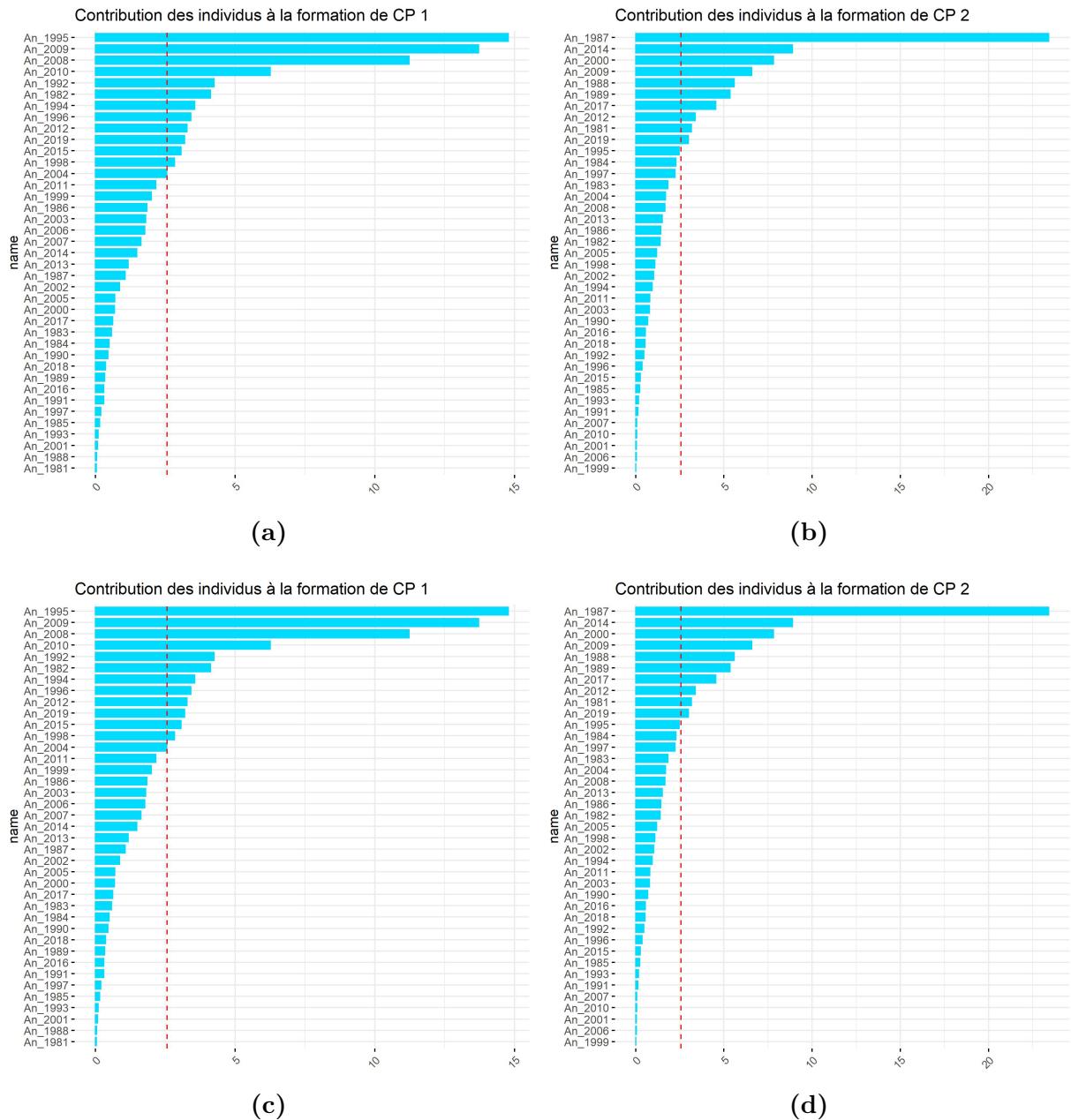
# Contributions des individus à CPs
for (i in 1:N_CP) {
  titre= paste("Contribution des individus à la formation de CP",i)
  fviz_contrib(res.acp, choice = "ind", axes = i,
               title =titre)
}

# variables
::::::::::
# Resultats des variables
res.var <- get_pca_var(res.acp)
res.var$coord          # coordonnées
res.var$contrib         # Contributions au CPs en %
res.var$cos2            # Qualité de la représentation
# projection des variables dans le plan des deux premières comps princip
N_CP=5
for (i in 1:N_CP)
{
  for (j in i:N_CP)
  {
    if (i!=j){
      fviz_pca_var(res.acp,
                    axes = c(i,j),
                    select.var = list(cos2 =30),
                    col.var = "cos2",  # colorer selon la contribution
                    gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"
                    geom.var = c('point','text'),
                    title ="Variables-ACP(10 premières) ",
                    repel = TRUE,
                    )+ theme_get()

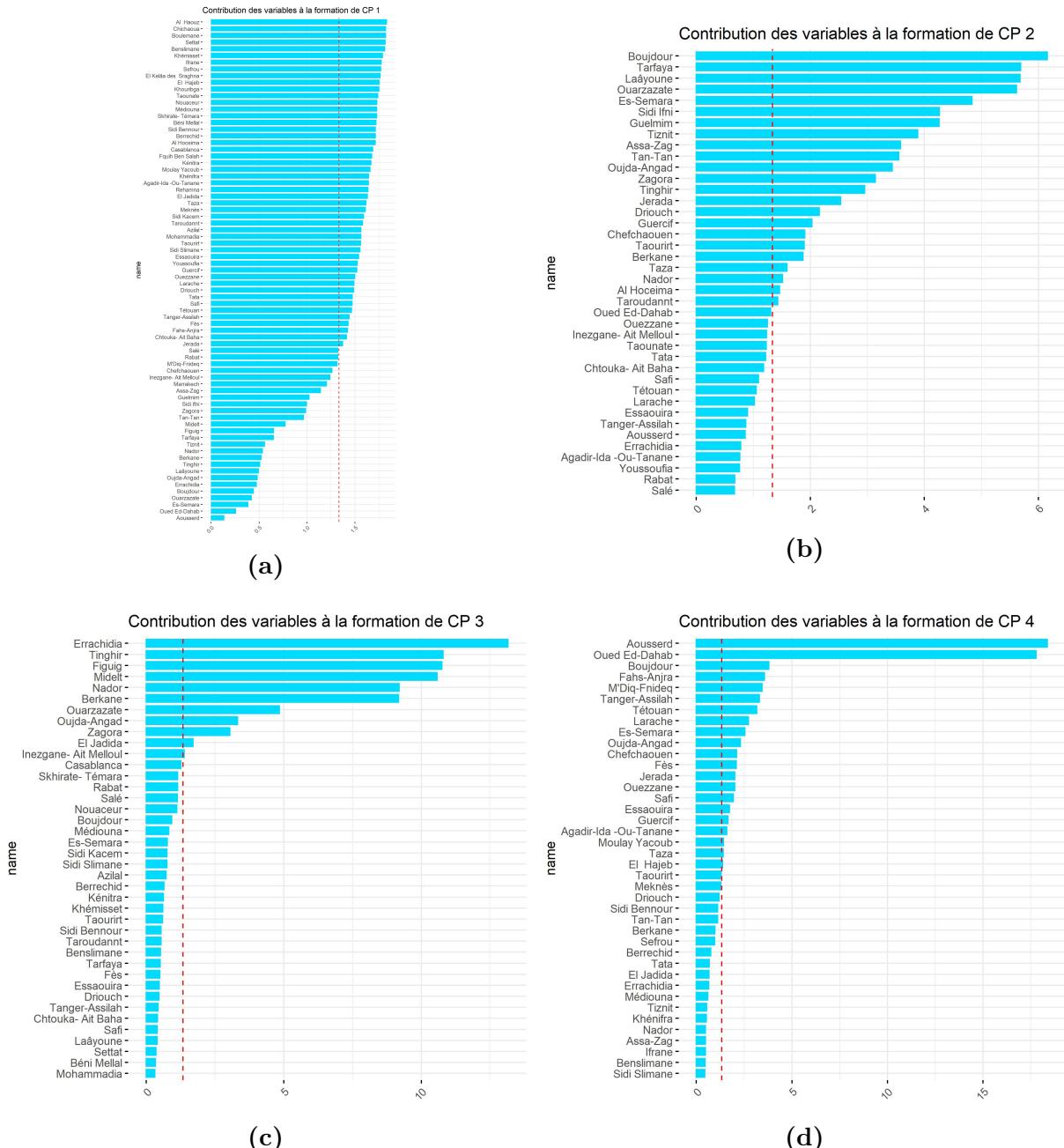
    }
  }
}

```

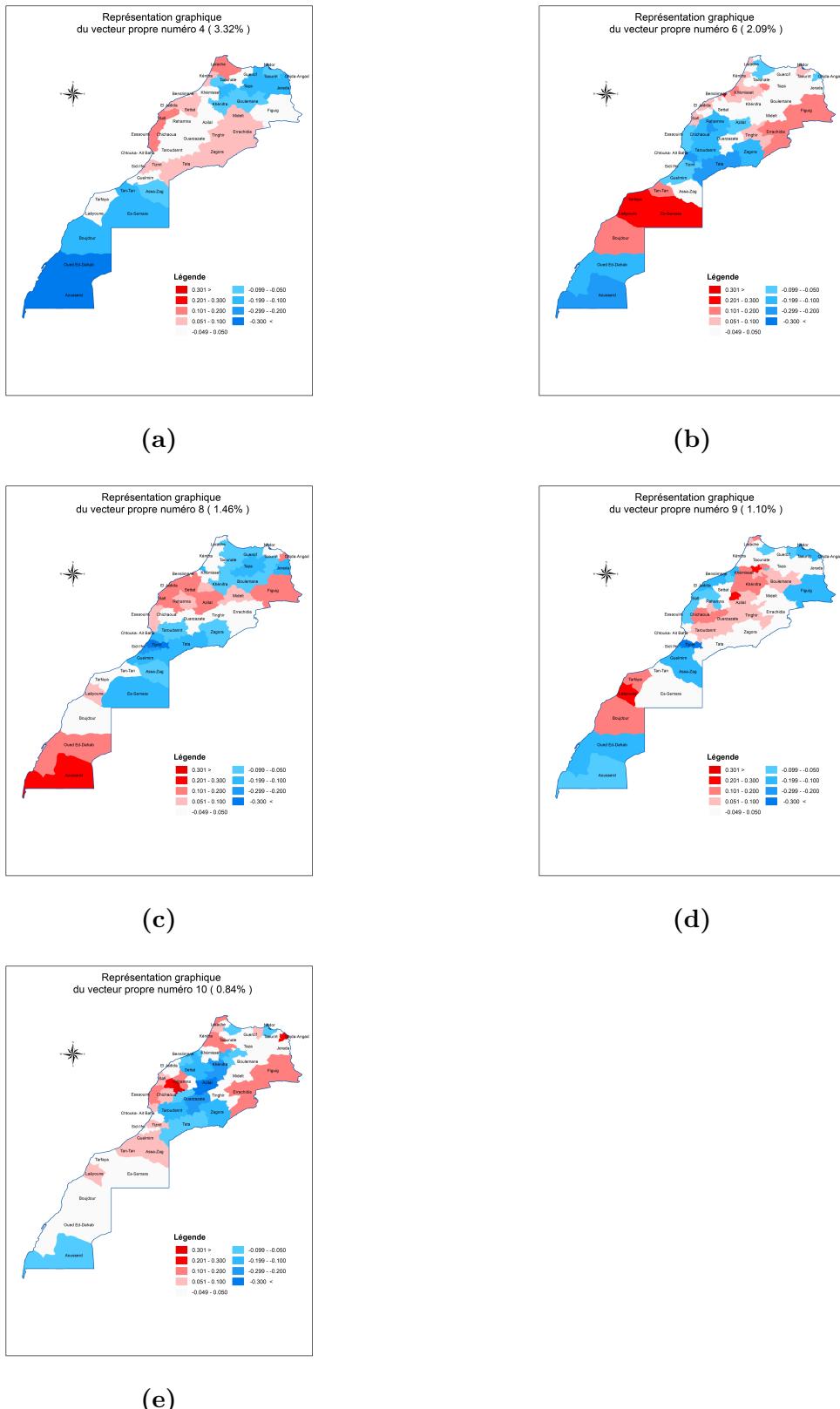
## Annexe B : Représentations graphiques



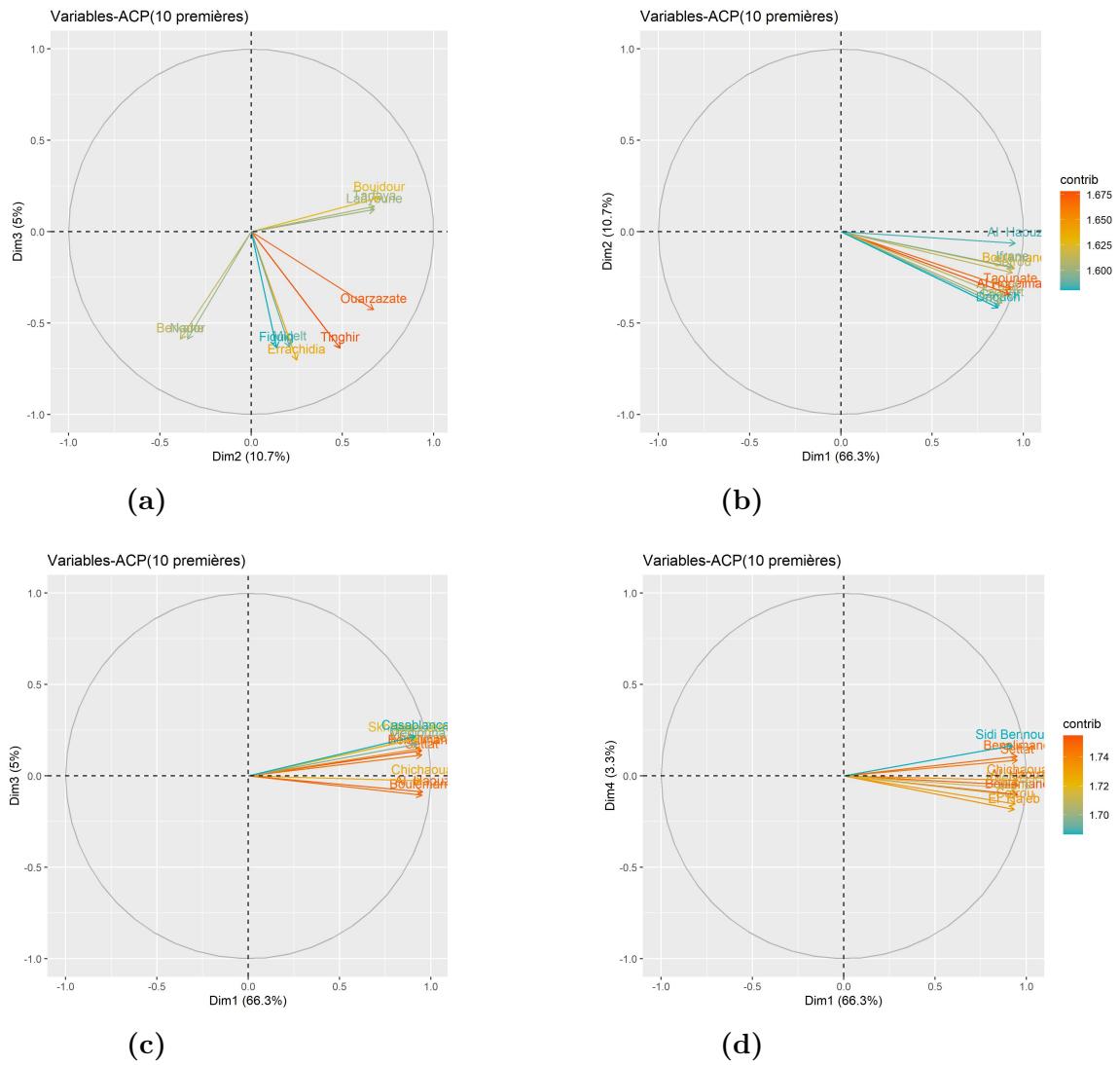
**Figure 4.11 – Contributions des individus à la formation des CPs**



**Figure 4.12 – Contributions des variables à la formation des CPs**



**Figure 4.13 – Représentations graphiques des vecteurs propres (non présentées dans les chapitres)**



**Figure 4.14 – Projection des 10 premières variables sur les plans formés par les CPs deux à deux**

## Annexe C : Tables des résultats statistiques

**Table 4.5 – Statistiques descriptives des provinces :**

Provinces	moyenne	écart-type	coef de variation %	Provinces	moyenne	écart-type	coef de variation %
Agadir-Ida -Ou-Tanane	271.3	104.97	38.70	M'Diq-Fnideq	645.2	238.06	36.90
Al Haouz	383.3	118.64	30.95	Médiouna	337.4	116.90	34.65
Al Hoceima	683.9	253.39	37.05	Meknès	476.0	136.47	28.67
Aousserd	33.5	34.66	103.40	Midelt	155.6	54.67	35.14
Assa-Zag	113.6	55.76	49.10	Mohammadia	391.4	149.74	38.26
Azilal	246.8	72.76	29.48	Moulay Yacoub	477.8	138.31	28.95
Béni Mellal	280.3	78.99	28.18	Nador	315.6	112.34	35.60
Benslimane	334.6	115.44	34.50	Nouaceur	366.8	129.49	35.30
Berkane	301.6	101.23	33.56	Ouarzazate	126.0	51.83	41.14
Berrechid	319.3	111.65	34.97	Oued Ed-Dahab	32.8	30.54	93.25
Boujdour	42.2	26.74	63.34	Ouezzane	690.9	299.33	43.33
Boulemane	467.7	150.04	32.08	Oujda-Angad	265.2	77.50	29.22
Casablanca	397.6	144.21	36.27	Rabat	519.2	193.12	37.20
Chefchaouen	811.3	447.29	55.14	Rehamna	250.1	78.40	31.35
Chichaoua	351.0	110.48	31.47	Safi	358.6	148.35	41.37
Chtouka- Ait Baha	210.9	114.59	54.33	Salé	519.0	192.69	37.13
Driouch	440.0	142.24	32.33	Sefrou	664.8	217.03	32.65
El Hajeb	481.6	141.31	29.34	Settat	321.3	113.81	35.42
El Jadida	376.3	126.96	33.74	Sidi Bennour	340.3	128.80	37.85
El Kelâa des Sraghna	277.6	85.20	30.69	Sidi Ifni	135.3	64.75	47.87
Errachidia	130.5	62.74	48.08	Sidi Kacem	421.6	141.64	33.60
Essaouira	280.3	110.40	39.39	Sidi Slimane	409.0	142.18	34.77
Es-Semara	71.3	46.48	65.15	Skhirate- Témara	390.4	128.74	32.98
Fahs-Anjra	673.3	234.58	34.84	Tanger-Assilah	720.6	243.56	33.80
Fès	489.7	169.92	34.70	Tan-Tan	103.1	53.78	52.14
Figuig	163.0	57.12	35.05	Taounate	551.1	188.80	34.26
Fquih Ben Salah	340.5	107.72	31.64	Taourirt	429.5	136.47	31.78
Guelmim	127.2	62.46	49.09	Tarfaya	58.3	34.27	58.79
Guercif	575.8	200.27	34.78	Taroudannt	203.5	74.41	36.56
Ifrane	806.9	273.19	33.86	Tata	158.0	79.65	50.41
Inezgane- Ait Melloul	239.1	136.76	57.21	Taza	637.0	224.26	35.20
Jerada	442.9	152.75	34.49	Tétouan	722.3	298.60	41.34
Kénitra	592.7	187.99	31.72	Tinghir	126.0	52.99	42.05
Khémisset	431.4	126.02	29.21	Tiznit	145.7	79.27	54.41
Khénifra	385.9	115.57	29.95	Youssoufia	248.6	79.25	31.88
Khouribga	348.1	116.00	33.33	Zagora	134.4	59.83	44.51
Laâyoune	45.2	30.45	67.40				
Larache	716.2	273.37	38.17				
Marrakech	223.6	72.66	32.50				

**Table 4.6 – Statistiques descriptives des années :**

années	moyenne	écart-type	coef de variation %	années	moyenne	écart-type	coef de variation %
1981	324.0	165.86	51.19	2000	331.7	264.17	79.63
1982	226.4	154.89	68.41	2001	334.4	193.50	57.87
1983	314.5	209.75	66.68	2002	430.8	278.59	64.67
1984	292.0	138.30	47.37	2003	455.6	277.33	60.87
1985	382.3	227.99	59.63	2004	242.1	123.81	51.14
1986	268.7	162.12	60.33	2005	401.4	208.96	52.06
1987	390.8	159.18	40.73	2006	276.5	191.04	69.10
1988	350.1	165.78	47.35	2007	275.2	171.42	62.30
1989	385.0	201.80	52.41	2008	590.3	349.81	59.26
1990	401.5	231.16	57.58	2009	638.1	450.79	70.65
1991	315.7	182.37	57.76	2010	532.2	323.24	60.74
1992	226.8	154.06	67.92	2011	258.1	151.90	58.85
1993	328.9	181.92	55.32	2012	490.5	308.62	62.92
1994	213.3	100.75	47.24	2013	287.1	175.40	61.09
1995	615.2	370.59	60.24	2014	408.6	172.77	42.28
1996	489.9	322.30	65.79	2015	230.5	117.47	50.97
1997	407.2	288.51	70.85	2016	324.6	192.48	59.30
1998	232.2	114.24	49.20	2017	415.8	269.69	64.87
1999	267.5	174.08	65.08	2018	309.2	180.14	58.26
				2019	248.7	170.67	68.63

## Bibliographie

- [1] www.wikipedia.org (classification de köppen).
- [2] T.El Hairech D.Qinghan A.de Wit, Riad Ballaghi. Cgms piloting report morocco. 26/02/2014.
- [3] M.TAHRI M.LAHLOU R.BALAGHI, T.EL HAIRECH. Cgms-maroc manuel d'utilisation. Avril 2013.
- [4] S.El khatri. Cours statistiques multidimensionnelles. 2010-2020.
- [5] S.Robin C.Duby. Cours analyse en composantes principales.
- [6] Classification automatique de données par la méthode hiérarchique ascendante. 2011-2012.
- [7] Cours 8 : Classification automatique de donn'ees par la méthode des centres mobiles. 2011-2012.