

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



Université de Ghardaïa

Faculté des Sciences et Technologies

Département des Sciences et Technologie

N° d'ordre :

N° de série :

Mémoire présenté en vue de l'obtention du diplôme de

**MASTER**

**Domaine :** *Sciences et Technologies.*

**Filière :** *Hydraulique.*

**Spécialité :** *Hydraulique urbaine.*

**Par :** BOUREZG Elarbi

OULEDOULI Mohamed

**Thème**

**Étude comparative de la variabilité de données pluviométriques des stations : - Ghardaïa- Hassi R'mel - Touggourt sur la période 1975-2014**

**Soutenu publiquement le : 11 /06/2018**

**Devant le jury :**

<b>Mr. MECHERI Bachir</b>	MAA	Univ. Ghardaïa	<b>Président</b>
<b>Mme. AMIER Rekaia</b>	MAA	Univ. Ghardaïa	<b>Examineur</b>
<b>Mr. CHEBIHI Lakhdar</b>	MAB	Univ. Ghardaïa	<b>Examineur</b>
<b>Mr. OULED BELKHIR Cheikh</b>	MAA	Univ. Ghardaïa	<b>Encadreur</b>

**Année universitaire : 2017/2018.**

---

## Dédicaces

*A mon support dans ma vie, mon exemple éternel, mon soutien moral et source De joie et de bonheur, celui qui s'est toujours sacrifié pour me voir réussir, Que dieu te garde dans son vaste paradis, à toi mon père «Abdelkader»*

*Je tien à te dévoiler mes sincères sentiment d'amour et de respect pour celle Qui m'a appris le courage, qu'elle a fait tant de sacrifices durant mes études du primaire à ce jour ma mère «Saida»*

*A mes beaux-frères «Rekia» «Mokhtar» «Fatima» «Fatiha» «Youcef» «Aicha» «Zineb» «Meriem» «Khadîdja» «Ahmed» «Ibrahim» «Oumkaltom» «Moustapha» «Lamin» et toute la famille «Bourezg» .*

*A mes amies de la promo : «Mohamed» «Abdel Hadi» «Yacine» «Ahmed» «Adelrrazak» «Abdelfattah» «Djalal» «Messaoud» «Aissaoui» «Meriem» «Kenza» «Djihad».*

*A mes amies : «Younes» «Faredj» «Abdellah» «Bachir» «Imad» «Arafa»*

*A la promo de la 2 éme année Master «Hydraulique Urbain» et «Ressource Hydraulique »*

*Aux personnes qui m'ont toujours aidé et encouragé, qui étaient toujours à mes côtés et qui m'ont accompagnaient durant mon chemin d'études supérieures, mes aimables amis, collègues d'études.*

**BOUREZG ELARBI**

## Dédicace

*A mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études.*

*A mes frère (AbdelRazak, Abdel Rahim, Abdelkrim, Siham, widade et Les fils de mes oncles) pour leurs encouragements permanents, et leur soutien moral.*

*A mes chers cousins (Abderrahmane, Abdelkader et Hamo) leur soutien moral.*

*A toute ma famille pour leur soutien tout au long de mon parcours universitaire.*

*Aux amis de la famille (Mr : Lahouadji Bachir).*

*Merci à mon collègue pour ce travail (Elarbi)*

*A mes chers frères dans la spécialisation (Elarbi, Abdel Hadi, Abdel Fattah, AbdelRazak, Yacine, Djalal, Ahmed, Massoud, Meriem, Kenza) pour leur appui et leur encouragement.*

*A mes aimes donne la spécialité hydraulique urbaine et ressource hydraulique.*

*Merci d'être toujours là pour moi.*

*Ouledouali Mohamed*

### **Remercîments**

*Merci à Allah ; notre guide, notre force, notre bonheur et la raison de notre existante. C'est lui qui nous a fait comprendre le but de cette vie, et qui nous a donné le pouvoir d'aimer les gens et d'apprécier les choses. Merci d'être là dans les moments les plus difficiles.*

*Je remercie fortement mon promoteur : Mr. OULED BELKHIR CHEIKH de m'avoir orienté par ses conseils judicieux dans le but de mener à bien ce travail. Mon respect aux membres de jury qui me feront l'honneur de juger ce travail, je les remercie vivement*

*Mes vifs et sincères remerciements vont à Mr.MECHERI Bachir. Pour avoir accepté de présider ce jury.*

*A Mme. AMIER Rekaia et Mr.CHEBIHI Lakhdar. D'avoir accepté d'examiner ce travail ;*

*A la fin je remerciement A toute la Promotion de 2eme année Mastère En hydraulique urbain, mais amis et tous ceux qui ont contribué de près ou de loin à la réalisation de ce mémoire.*

*Ouledouali Mohammed*

*Bourezg Elarbi*

## Tableau de matière

<i>Dédicaces</i> .....	<b>I</b>
<i>Remercîments</i> .....	<b>III</b>
<b>Liste des tableaux</b> .....	<b>IX</b>
<b>Liste des figures</b> .....	<b>X</b>
<b>Liste des abréviations</b> .....	<b>XII</b>
<b>Introduction générale</b> .....	<b>1</b>
<b>Chapitre 01 : Mesures des précipitations et correction de données pluviométriques</b> .....	<b>3</b>
1.1 Introduction .....	4
1.2 Etude des précipitations .....	4
1.2.1 Définition des précipitations.....	4
1.2.2 Mesures de la hauteur d'eau précipitée .....	4
1.2.3 Instruments de mesure .....	4
1.2.3.1 Le pluviomètre .....	5
1.2.3.2 Les pluviographes.....	5
1.2.3.3 L'implantation des appareils de mesures .....	6
1.2.3.4 Le dépouillement d'un pluviogramme .....	6
1.3 Correction des données .....	7
1.3.1 Les étapes d'organisation.....	7
1.3.1.1 Acquisition des données .....	7
1.3.1.2 Traitement primaire des données.....	7
1.3.1.3 Contrôle des données .....	8
1.3.1.4 Organisation des données.....	8
1.3.1.5 Diffusion et publication .....	8
1.3.2 Le contrôle des données .....	8
1.3.3 Recherche des erreurs et corrections des mesures.....	9
1.3.3.1 Les erreurs dans les mesures .....	9
1.3.3.1.1 Les erreurs d'observation.....	9
1.3.3.1.3 Les erreurs systématiques.....	9
1.3.3.2 La correction des erreurs .....	10
1.4 Conclusion .....	12
<b>Chapitre 02 : Analyse des données pluviométriques et le choix du type de loi d'ajustement</b> .....	<b>14</b>
2.1 Introduction .....	15
2.1 Analyse des données statistiques .....	15

2.1.1 Analyse descriptive .....	15
2.1.2. Choix de la loi d'ajustement.....	16
2.1.2.1 Loi de Gauss ou loi normale .....	16
2.1.2.1.1 Fonction de répartition .....	16
2.1.2.1.2 Estimation des paramètres .....	18
2.1.2.1.3. Intervalles de confiance .....	19
2.1.2.2 Loi de Galton (loi log-normale) .....	20
2.1.2.2.1 Fonction de répartition .....	20
2.1.2.2.2 Estimation des paramètres .....	21
2.1.2.2.3 Intervalles de confiance .....	23
2.1.2.3. Loi de Gumbel .....	24
2.1.2.3.1. Fonction de distribution.....	24
2.1.2.3.2. Estimation des paramètres .....	24
2.1.2.3.3 Intervalle de confiance.....	26
2.2 Teste d'ajustement .....	26
2.2.1 Teste de Khie 2 .....	26
2.2.2 Test Wn2 d'Anderson .....	27
2.3 Conclusion.....	28
<b>Chapitre 03 : Analyse multi variation (Corrélation multiple et ACP) .....</b>	<b>29</b>
3.1 Introduction .....	30
3.2 Corrélation multiple .....	30
3.2.1 Généralités .....	30
3.2.2 Objectif.....	30
3.2.3 Propriétés.....	31
3.2.4 Domaine de définition.....	31
3.2.5 Statistique de corrélation.....	32
3.2.6 Corrélation simple.....	33
3.2.7 Corrélation multiple .....	35
3.3 Analyse en composantes principales ACP .....	38
3.3.1 Généralités .....	38
3.3.2 Objectif.....	38
3.3.3 Méthodologie de l'ACP .....	38
3.3.3.1 Les étapes d'un ACP .....	40
3.3.3.2 Interprétation géométrique de l'ACP.....	41
3.3.3.3 Analyse dans l'espace des échantillons.....	41

3.3.4 Exemple numérique .....	41
3.3.5 ACP de la matrice des covariances.....	42
3.3.6 ACP de la matrice des corrélations .....	42
3.3.7 Exemples d'application .....	45
3.4 Conclusion.....	47
<b>Chapitre 04: présentation dès les régions d'étude .....</b>	<b>49</b>
4.1 Introduction .....	50
4.2Présentation de la Wilaya de Ghardaïa.....	50
4.2.1 Situation géographique.....	50
4.2.2 Le climat .....	51
4.2.2.1Températures.....	51
4.2.2.2 Pluviométrie .....	51
4.2.2.3 Evaporation .....	52
4.2.2.4 Les vents.....	52
4.3 Présentation la région de Hassi R'mel.....	52
4.3.1 Situation géographique .....	52
4.3.2 Climat .....	53
4.3.2.1 Température .....	53
4.3.2.2 Pluviométrie.....	53
4.3.2.3 Humidité de l'air.....	54
4.3.2.4 Vents .....	54
4.3.2.5 Evaporation .....	54
4.4 Présentation la région de Touggourt .....	54
4.4.1 Situation géographique .....	54
4.4.2 Climat .....	55
4.4.2.1Température .....	55
4.4.2.3 L'humidité relative de l'air .....	55
4.4.2.4 Le vent.....	55
4.5 Conclusion.....	56
<b>Chapitre 05 : Statistique d'ordre applique sur les 03 stations .....</b>	<b>57</b>
5.1 Introduction .....	58
5.2 Analyse de la précipitation.....	58
5.3 Station de Ghardaïa .....	58
5.3.1 Précipitations maximales journalières .....	58
5.3.2 Analyse des données statistiques .....	59

5.3.3 Choix de la loi d'ajustement.....	60
5.3.3.1 Ajustement de la série pluviométrique à la loi de Galton (Log Normal).....	60
5.3.3.2 Ajustement de la série pluviométrique à la loi de Gumbel.....	60
5.3.4 Interprétation des graphes .....	61
5.3.5 Test de Khi <sup>2</sup> .....	61
5.3.5.1 Interprétation des tableaux .....	62
5.4 Station de Hassi R'mel.....	62
5.4.1 Précipitations maximales journalières .....	62
5.4.2 Analyse des données statistiques .....	62
5.4.3 Choix de la loi d'ajustement.....	63
5.4.3.1 Ajustement de la série pluviométrique à la loi de Galton (Log Normal).....	63
5.4.3.2 Ajustement de la série pluviométrique à la loi Gumbel.....	64
5.4.4 Interprétation de graphe .....	65
5.4.5 Teste khi2 .....	65
5.4.5.1 Interprétation des tableaux .....	65
5.5 Station de Touggourt .....	66
5.5.1 Précipitations maximales journalières .....	66
5.5.2 Analyse des données statistiques .....	66
5.5.3 Choix de la loi d'ajustement.....	67
5.5.3.1 Ajustement de la série pluviométrique à la loi de Galton (Log Normal).....	67
5.5.3.2 Ajustement de la série pluviométrique à la loi Gumbel.....	67
5.5.4 Interprétation de graphe .....	68
5.5.5 Teste khi2 .....	68
5.5.5.1 Interprétation des tableaux .....	69
5.6 Conclusion.....	69
<b>Chapitre 06 : Analyse composantes principales.....</b>	<b>71</b>
6.1 Introduction .....	72
6.2 Généralités .....	72
6.3 Modelé de L'A.C.P .....	72
6.3.1 Détermination analytique .....	73
6.4 Application aux pluies maximales journalières.....	76
6.4.1 Interprétation dans l'espace des variables .....	77
6.4.2 Interprétation dans l'espace des individus .....	80
6.5 Conclusion.....	81
<b>Conclusion générale.....</b>	<b>82</b>



**Annexes ..... IX**  
**Références bibliographiques.....XVII**  
**Résumé**

## Liste des tableaux

### Chapitre 02 :

Tableau 2.1 Table de la probabilité au non-dépassement est le complément à 1 de $F( u )$ .....	17
Tableau 2.2 Les bornes de l'intervalle de confiance.....	24

### Chapitre 03 :

Tableau 3.1 Table d'inter corrélation.....	35
Tableau 3.2 données sont exemple numérique.....	41
Tableau 3.3 Matrice des coefficients de corrélation.....	41
Tableau 3.4 Les données sont des analyses géochimiques d'éléments majeurs .....	46

### Chapitre 05 :

Tableau 5.1 Pluies journalières maximales observées à Ghardaïa (1975-2014).....	59
Tableau 5.2 Test de Khi 2 appliqué sur la série des pluies (Galton).....	61
Tableau 5.3 Test de Khi 2 appliqué sur la série des pluies (Gumbel).....	62
Tableau 5.4 Pluies journalières maximales observées à Hassi R'mel (1975-2014).....	63
Tableau 5.5 Test de Khi 2 appliqué sur la série des pluies (Galton).....	65
Tableau 5.6 Test de Khi 2 appliqué sur la série des pluies (Gumbel).....	65
Tableau 5.7 Pluies journalières maximales observées à Touggourt (1975-2014).....	66
Tableau 5.8 Test de Khi 2 appliqué sur la série des pluies (Galton).....	68
Tableau 5.9 Test de Khi 2 appliqué sur la série des pluies (Gumbel).....	69

### Chapitre 06 :

Tableau 6.1 Données formant la matrice initial .....	73
Tableau 6.2 Statistiques descriptives.....	74
Tableau 6.3 Moyenne et l'écart types des variables .....	74
Tableau 6.4 Valeurs propres et pourcentage de variance exprimé.....	76
Tableau 6.5 Matrice des coefficients de corrélation.....	76
Tableau 6.6 Corrélations entre les variables et les composantes principales .....	77
Tableau 6.7 Corrélations entre les individus et les composantes principales .....	79

## Liste des figures

### Chapitre 01:

Figure 1.1 pluviomètre .....	5
Figure 1.2 pluviographe a augets basculeurs et pluviographe a siphon .....	5
Figure 1.3 Enregistrement pluviographique.....	6
Figure 1.4 Précipitation annuelle cumulative -moyenne de 6 stations .....	11
Figure 1.5 la méthode du double cumul .....	12

### Chapitre 02:

Figure 2.1 Ajustement graphique à une loi de Gauss. ....	19
Figure 2.2 Variations de $\beta_1$ et $\beta_2$ avec $\sigma$ . ....	21
Figure 2.3 Ajustement graphique Loi de Galton .....	23
Figure 2.4 Ajustement graphique à la loi de Gumbel .....	26

### Chapitre 03:

Figure 3.1 Quelques valeurs possibles de coefficient de corrélation .....	32
Figure 3.2 Représentation d'une corrélation pour trois variables .....	35
Figure 3.3 Ellipse définie par l'ACP .....	42
Figure 3.4 Régressions et ACP .....	45
Figure 3.5 Graphe de valeurs propres (en%) .....	47
Figure 3.6 ACP des données des montérégiennes (variables).....	47

### Chapitre 04:

Figure 4.1 Situation géographique de Ghardaïa. ....	51
Figure 4.2 Situation géographique de la ville de Hassi R'mel. ....	53
Figure 4.3 Situation géographique de Touggourt .....	55

### Chapitre 05:

Figure 5.1 Ajustement graphique d'une loi de Galton (Ghardaïa) .....	60
Figure 5.2 Ajustement graphique d'une loi de Gumble (Ghardaïa) .....	61
Figure 5.3 Ajustement graphique d'une loi de Galton (Hassi R'mel) .....	64
Figure 5.4 Ajustement graphique d'une loi de Gumble (Hassi R'mel) .....	64
Figure 5.5 Ajustement graphique d'une loi de Galton (Touggourt) .....	67
Figure 5.6 Ajustement graphique d'une loi de Gumble (Touggourt) .....	68

### Chapitre 06:

Figure 6.1 Contributions des variables.....	77
Figure 6.2 Projection de les stations sur le plan factoriel .....	78

Figure 6.3 Projection des mois sur les composantes principales ..... 80

---

## Liste des abréviations

- ANRH : Agence nationale des ressources hydriques.
- ONM : Office national de météorologie.
- FDP : Fonction de distribution de probabilité.
- SDC : Série de durées complètes.
- IDF : Intensité-durée-fréquence.
- ACP : Analyse composantes principales
- FND : fréquence non dépasses
- FDP : Fonction de distribution de probabilité.
- MNT : Modèle numérique de terrain.
- SDC : Série de durées complètes.
- SMA : Série max annuels.
- Cv : Coefficient de variation
- Cs : Coefficient d'asymétrie
- Ck : Coefficient d'aplatissement
- P : Précipitation ou pluie brute en (mm).
- g : gradex
- $\sigma$  : l'écart-type
- $\bar{x}$  : moyenne
- X mod : le mode
- r : le coefficient de corrélation
- u : Variable centrée réduite de Gauss.
- y : la variable de Gumbel
- P10 : La pluie journalière décimale
- P j max : pluies maximales journalière (mm)

- Nb : Nombre d'année d'observation.
- NS : Nombre de valeurs inférieures et supérieures à la médiane.
- Ts : Taille de la plus grande suite de valeurs inférieures ou supérieures à la médiane dans la série chronologique.
- N1 : Taille de la série des pluies moyennes annuelles de la station à étudier.
- N2 : Taille de la série des pluies moyennes annuelles de la station de référence.
- Q f% : débit pluvial de fréquence
- P t, T : pluie de courte durée de fréquence donnée (mm)
- P j max % : pluie journalière de fréquence donnée (mm)
- b : exposant climatique
- T : Période de retour (années)
- FD : Fréquence au dépassement
- FND : Fréquence au non dépassement
- $\Delta t$  : Durée de l'intervalle (min)
- I **min** : Intensité de la pluie minimale (mm/min)
- I **moy** : Intensité de la pluie moyenne (mm/min)
- I max : Intensité de la pluie maximale (mm/min)
- t : temps de pluie (min)
- A t : paramètre géographique
- Nt : L'exposant climatique
- P<sub>obsrvé</sub>*** : La précipitation mesurée
- P<sub>corrigé</sub>*** : La précipitation corrigée
- S<sub>ajesté</sub>*** : La pente de la portion du graphe fiable
- S<sub>obsrvé</sub>*** : La pente de la portion du graphe à corriger

# **Introduction générale**

Les changements climatiques et la variabilité sont connus depuis longtemps, un intérêt évident des hydrologues. Les conséquences de ces phénomènes sur l'activité humaine en général et sur les ressources en eau en particulier sont très fortes et affectent de nombreux secteurs

En hydrologie, concentrez-vous d'abord sur la distribution elle-même, l'étendue des précipitations, quand et où elle se produit. Ainsi, l'étude des précipitations accent hydrologique sur la quantité et le rythme, dans le temps et l'espace, cette étude est également intéressé par les problèmes de mesure, ainsi que les espèces et l'utilisation appropriée des données de la pluie est l'information de la pluie est d'une importance capitale pour la vie quotidienne précisément, les ingénieurs ont besoin de connaître les paramètres de pluie Pour déterminer la taille de leur travail ou de leur maintenance

Il est très important de considérer les précipitations journalières maximales dans les zones arides et semi-arides pour estimer l'eau qui sont susceptibles de couler les montants, les projets de loi climat hydrologique que nous sommes intéressés à ce principalement dans cette partie du désert du nord, se rapportent à « étudier l'analyse détaillée des paramètres. »

La zone qui fait l'objet de l'étude est située au Sahara Septentrional, elle correspond aux trois stations le plus en voisinage dans la région à savoirs : Ghardaïa, Hassi R'mel et Touggourt

Parmi les problèmes rencontrés par l'enregistrement de la quantité de précipitations, l'absence de données sur les précipitations dans les stations du désert en raison de l'absence de certains enregistrements et de la perte due à un défaut dans le processus d'enregistrement. L'estimation des valeurs manquantes a toujours été un problème majeur en hydrologie en raison de la variation spatiale et temporelle des précipitations et de la complexité de la physique.

Ajouter un autre au problème réside dans le problème de la couverture et la distance entre les stations où le vaste réseau déjà dans les pays du désert, et deviennent de plus en plus vulnérables à rendre la région plus Sahara, il serait nécessaire de savoir que la distance entre deux stations peut dépasser 150 km.

L'objectif principal de cette étude est le recensement des précipitations sur la zone de Sahara septentrional et de chercher s'il ya une possibilité de corrélation quel quant entre les



trois stations; traduisant ainsi l'étendue des perturbations touchants le Sahara septentrional. La disponibilité des données pluviométriques sur les trois stations notamment les précipitations maximales journalières s'étale la période: 1975-2014.

Nous avons accentuées sur l'utilisation d'un ensemble de techniques statistiques tel que les méthodes des ajustements statistique ainsi que l'analyse dans les composantes principales ACP

Nous avons divisé cette étude en six chapitres où nous parlons dans le premier chapitre sur la façon de mesurer les précipitations et comment remédier au manque de données dans les enregistrements des retombées radioactives, Le deuxième chapitre parlera des types de méthodes statistiques et comment les appliquer, Le troisième chapitre traitera des types de corrélation et de la manière de déduire et d'introduire l'application de l'ACP et comment l'appliquer, le quatrième chapitre parle des détails dès les zones d'étude et des caractéristiques.

Comme pour le cinquième chapitre, nous appliquons les méthodes adaptatives aux trois régions afin de connaître la meilleure formule, formulons les résultats auxquels nous sommes parvenus traitement des chroniques, précipitations maximales journalières. Enfin, dans un dernier chapitre en effectue le traitement de ces données par l'analyse en composantes principales (ACP)

# **Chapitre 01 : Mesures des précipitations et correction de données pluviométriques**

## **1.1 Introduction**

L'étude se limite ici à l'examen des variations des quantités précipitées depuis le début des observations. Nous présenterons ensuite les résultats obtenus après le traitement statistique des observations et notamment les aspects liés à la stationnarité de la série, à la distribution des précipitations et aux différentes échelles de mesure envisagées.

## **1.2 Etude des précipitations**

### **1.2.1 Définition des précipitations**

Sont dénommées précipitations, toutes les eaux météoriques qui tombent sur la surface de la terre, tant sous forme liquide (bruine, pluie, averse) que sous forme solide (neige, grêle) et les précipitations déposées ou occultes (rosée, gelée blanche, givre...) [23]

Elles sont provoquées par un changement de température ou de pression. Les précipitations constituent l'unique entrée des principaux systèmes hydrologiques continentaux que sont les bassins versants. [12]

### **1.2.2 Mesures de la hauteur d'eau précipitée**

Comme les précipitations varient selon différents facteurs (déplacement de la perturbation, lieu de l'averse, influence de la topographie, etc.), leur mesure est relativement compliquée.

Quelle que soit la forme de la précipitation, liquide ou solide, on mesure la quantité d'eau tombée durant un certain laps de temps. On l'exprime généralement en hauteur de précipitation ou lame d'eau précipitée par unité de surface horizontale (mm). On définit aussi son intensité (mm/h) comme la hauteur d'eau précipitée par unité de temps. La précision de la mesure est au mieux de l'ordre de 0,1 mm. En Algérie toute précipitation supérieure à 0,5 mm est considérée comme pluie effective. [12]

### **1.2.3 Instruments de mesure**

Les différents instruments permettant la mesure des précipitations sont :

### 1.2.3.1 Le pluviomètre

Instrument de base de la mesure des précipitations liquides ou solides. Il indique la quantité d'eau totale précipitée et recueillie à l'intérieur d'une surface calibrée dans un intervalle de temps séparant deux relevés. Il se compose d'un seau tronconique de contenance variable suivant l'abondance des pluies journalières et d'un entonnoir surmonté d'une bague à arête chanfreinée matérialisant la surface de réception. [12]

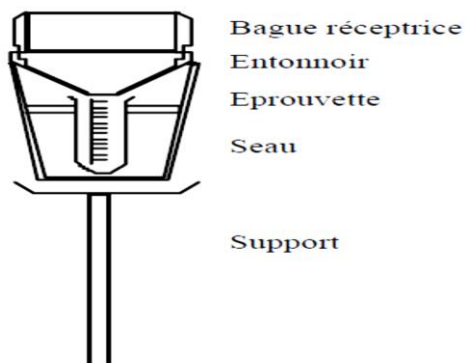


Figure 1.1 pluviomètre

### 1.2.3.2 Les pluviographes

Ces appareils sont destinés à l'enregistrement de la hauteur de pluie cumulée en fonction du temps. Deux types principaux ont eu un certain développement : les pluviographes à augets basculeurs et ceux à siphons. Actuellement, ces derniers tendent à être abandonnés.

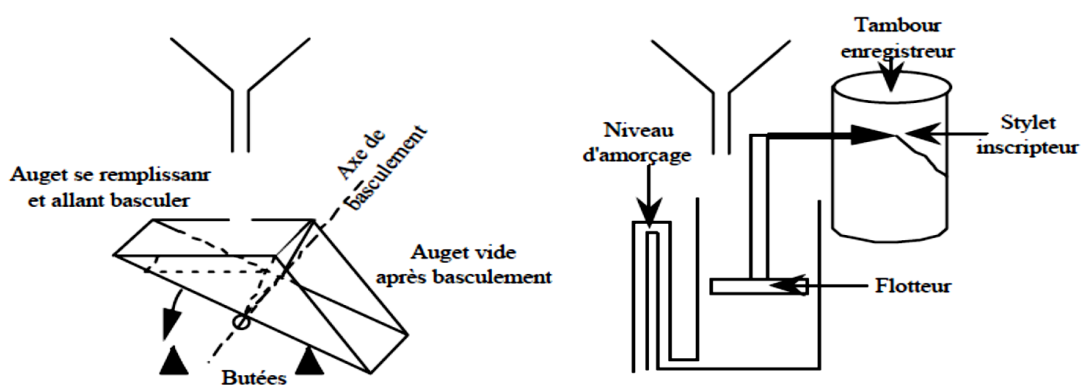


Figure 1.2 pluviographe a augets basculeurs et pluviographe a siphon

### 1.2.3.3 L'implantation des appareils de mesures

- être représentatif du secteur étudié en étant exposé normalement aux vents,
- être éloigné de tout obstacle, en générale a une distance minimum égale à 4 fois la hauteur de l'obstacle (arbre, bâtiment, etc.,).
- être à proximité de la résidence de l'observateur ;
- avoir une surface réceptrice rigoureusement horizontal ; on admet qu'un écart de 1 peut provoquer des erreurs de l'ordre de 1%. [24]

### 1.2.3.4 Le dépouillement d'un pluviogramme

On a pris l'habitude d'appeler pluviogramme la courbe des pluies cumulées fournie par le pluviographe, et l'hectogramme le graphique (ou histogramme) des intensités (mm/h)

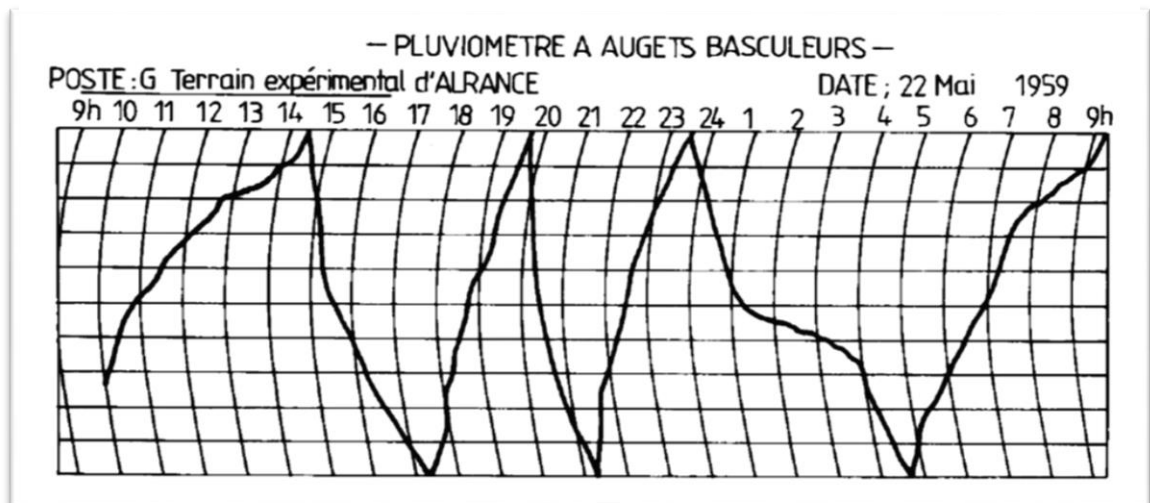


Figure 1.3 Enregistrement pluviographique

Le papier enregistreur a les caractéristiques suivantes

- Verticalement :
  - l'espace entre deux traits fins représente :
    - \* 1 mm de pluie pour une bague de 400 cm<sup>2</sup>
    - \* 1/5 mm de pluie pour une bague de 2000 cm<sup>2</sup>
  - un basculement représente :
    - \*0.5 mm pour une bague de 400 cm<sup>2</sup>

\*0.1 mm pour une bague de 2000 cm<sup>2</sup>

- Horizontalement :
  - l'espace entre deux traites épaisses représente 1 heure
  - l'espace entre deux traits fins représente 15 minutes.

Il existe plusieurs méthodes de dépouillement d'un pluviogramme :

- le dépouillement à pas de temps constant
- le dépouillement à intensités constantes
- le dépouillement à automatique sur digitaliseur [24]

### **1.3 Correction des données**

De façon générale, pour permettre le passage de l'acquisition des données à leur utilisation effective dans le cadre d'une analyse hydrologique on distingue les étapes suivantes : acquisition, traitement, contrôle et validation, organisation, diffusion et publication [23]

#### **1.3.1 Les étapes d'organisation**

##### **1.3.1.1 Acquisition des données**

L'acquisition de données consiste à procéder, par le biais d'un instrument de mesure, à acquérir de l'information [21]

##### **1.3.1.2 Traitement primaire des données**

La donnée acquise précédemment nécessite souvent un traitement préalable ou traitement primaire - afin de la rendre pertinente et exploitable. Il s'agit pour l'essentiel de la conversion de la mesure effectuée en une grandeur qui soit hydrologiquement significative

Le traitement des données inclut aussi le contrôle primaire des données qui comprend les contrôles de cohérence à l'exclusion de tous traitements statistiques [23]

### **1.3.1.3 Contrôle des données**

Avant de pouvoir exploiter les données et bien qu'elles soient dans un format adéquat, il importe de contrôler la fiabilité et la précision de ces dernières. Le contrôle permet de valider les données avant leur organisation au sein d'une banque de données pour leur mise à disposition à des fins opérationnelles. Lors de cette opération, on introduit des indices de qualité de la donnée ainsi que des indices indiquant que celle-ci est reconstituée, calculée voire manquante [21]

### **1.3.1.4 Organisation des données**

Au vu de l'importance quantitative et qualitative des données, il importe de les organiser avec soin. Ceci se fait à partir d'un corpus de documents originels (formulaires de terrain, diagrammes, unité de stockage électronique) constituant les archives qui sont en règle générale accessibles uniquement à un personnel spécifique (responsable du centre de collecte, archiviste...). La traduction des archives sous la forme de fichiers de base génère les "fichiers en l'état" et fournit une indication sur la provenance de la donnée (mesure, calcul, copie etc.) ainsi que sur sa qualité (complète ou non) et sa précision. Enfin, on constitue un fichier de travail provisoire permettant une visualisation des données et permettant de procéder aux différents tests de qualité et de précision des données [23]

### **1.3.1.5 Diffusion et publication**

La diffusion des données en vue de leur utilisation opérationnelle peut être faite selon deux modes :

- Publication papier
- Diffusion informatique [12]

## **1.3.2 Le contrôle des données**

La constitution d'une série de valeurs, constituant un échantillon au sens statistique du terme, est un processus long, et au cours duquel de nombreuses erreurs, de nature fort différente, sont susceptibles d'être commises.

Une erreur de mesure est définie comme étant la différence entre la vraie valeur (qui est l'idéal recherché, mais qui n'est en principe) et la valeur mesurée. Il est commode, tant pour les présenter que pour différencier la façon de les aborder, de considérer deux types d'erreur : les **erreurs aléatoires** et les **erreurs systématiques** [12]

### **1.3.3 Recherche des erreurs et corrections des mesures**

Selon la nature des erreurs constatées ou supposées la recherche de ces dernières fait appel à différentes techniques et méthodes

- « in situ » qui consiste à vérifier sur place la manière dont les données ont été organisées, traitées et/ou transformées.
- Investigation de bureau qui consiste à vérifier la chaîne de traitement de la mesure/donnée à chaque étape de son élaboration, tout comme la manière dont on a constitué les séries de données soumises à contrôle et/ou publication.
- Investigation statistique qui, à l'aide d'outils spécifiques, permet de mettre en évidence certaines erreurs ou inconsistance. Ces techniques efficaces ont largement été utilisées dans la pratique professionnelle et se basent sur des hypothèses spécifiques qu'il convient de bien connaître [21]

#### **1.3.3.1 Les erreurs dans les mesures**

##### **1.3.3.1.1 Les erreurs d'observation**

- ❖ lecteur peu consciencieux
- ❖ Erreurs fortuites de lecteur de l'éprouvette
- ❖ Erreurs dues à l'évaporation
- ❖ Débordement du pluviomètre quand la pluie est très intense
- ❖ Pluviomètre percé
- ❖ Pertes d'eau pendant le transvasement de l'éprouvette dans le sceau
- ❖ Pluviomètre sous un arbre [24]

##### **1.3.3.1.3 Les erreurs systématiques**

- ❖ la graduation de l'éprouvette ne correspondant pas à l'ouverture du pluviomètre
- ❖ un changement dans l'exploitation dû à :
  - un déplacement du pluviomètre



- une modification de l'environnement du pluviomètre
- un changement d'observateur
- une éprouvette cassée remplacée par une autre qui ne convient pas [12]

### 1.3.3.2 La correction des erreurs

- **La méthode des doubles cumuls**

L'homogénéisation par cette technique graphique nécessite la connaissance d'une série de données annuelles et homogènes et observées dans une station de référence dite station témoin ou station de base, voisine et régionale avec la station à corriger.

La méthode du double cumul est une méthode bidimensionnelle. Elle est employée pour évaluer, s'il y a lieu, la présence d'une anomalie dans la série étudiée pour la corriger [9]

- **Principe de la méthode**

Il s'agit de comparer la tendance de la station étudiée par rapport à celle de la station témoin, en traçant le graphe des données cumulées à la station étudiée par rapport aux données cumulées de la station témoin. [9]

-La méthode est fondée sur le principe suivant :

En l'absence d'anomalie, deux stations A, B, voisines mesurent chaque année une pluviométrie annuelle dans un rapport sensiblement constant d'une année à l'autre

En conséquence les points  $M(i)$  de coordonnées les pluies cumulées calculées à chaque station A et B jusqu'à l'année  $i$  sont pratiquement alignés. En revanche si une erreur systématique à la station étudiée s'est produite alors la droite des doubles cumuls présenterait une cassure de sa pente à l'année de l'introduction de l'erreur [9]

### Remarque

La méthode du double cumul a l'avantage d'être simple, bien connue de tous et très rapidement réalisable. Elle permet de détecter l'erreur mais pas de la corriger, du moins pas directement.

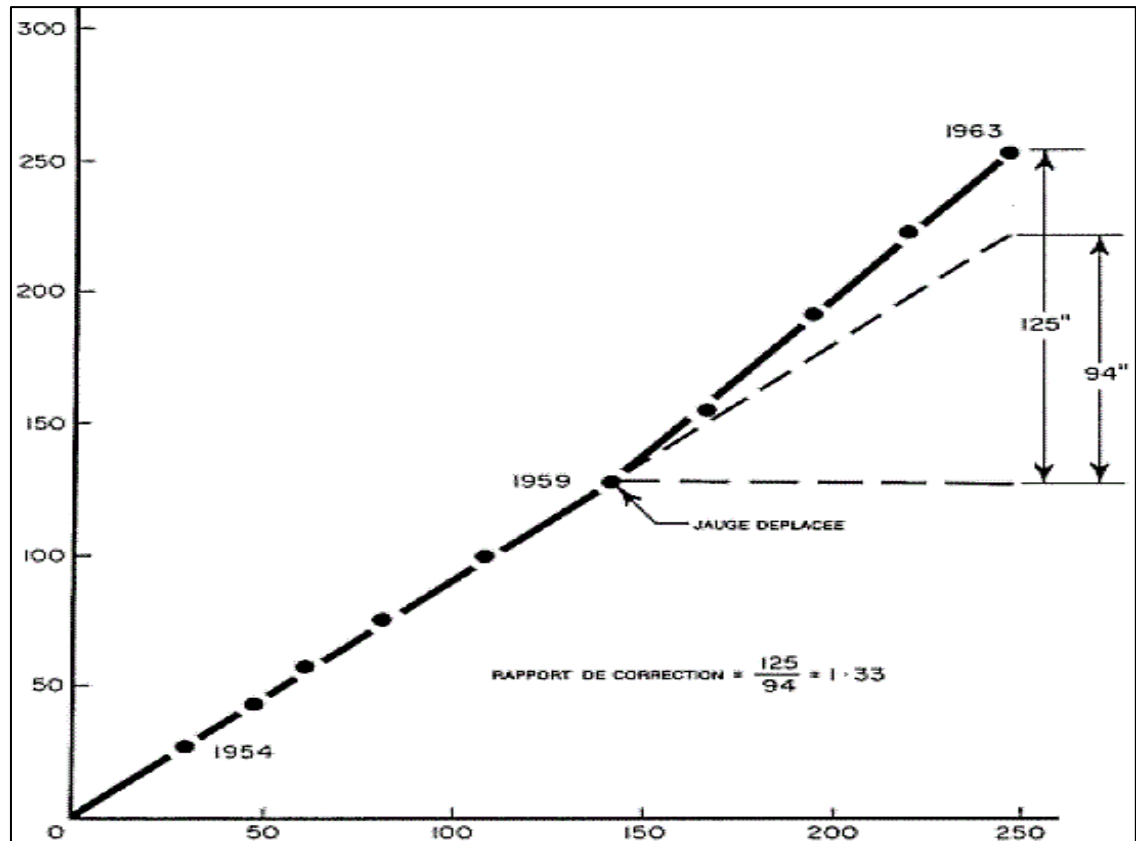


Figure 1.4 Exemple de Précipitation annuelle cumulative -moyenne de 6 stations

#### - Procédure de l'homogénéisation

On regarde l'historique de la station en cherchant des éléments indicatifs pouvant faciliter la prise de décision sur la période à partir de laquelle on soupçonne l'introduction d'erreurs dans les mesures. Dans le cas où aucun élément indicatif n'est disponible, on considérera quelles données les plus récentes sont les plus fiables [9]

La procédure de correction des données de la portion du graphe non fiable se fait en prolongeant la pente plus fiable selon la formule :

$$P_{\text{corrigé}} = \frac{S_{\text{ajesté}}}{S_{\text{obsrvé}}} P_{\text{obsrvé}}$$

$P_{obsrvé}$  Est la précipitation mesurée

$P_{corrigé}$  Est la précipitation corrigée

$S_{ajesté}$  Est la pente de la portion du graphe fiable

$S_{obsrvé}$  Est la pente de la portion du graphe à corriger

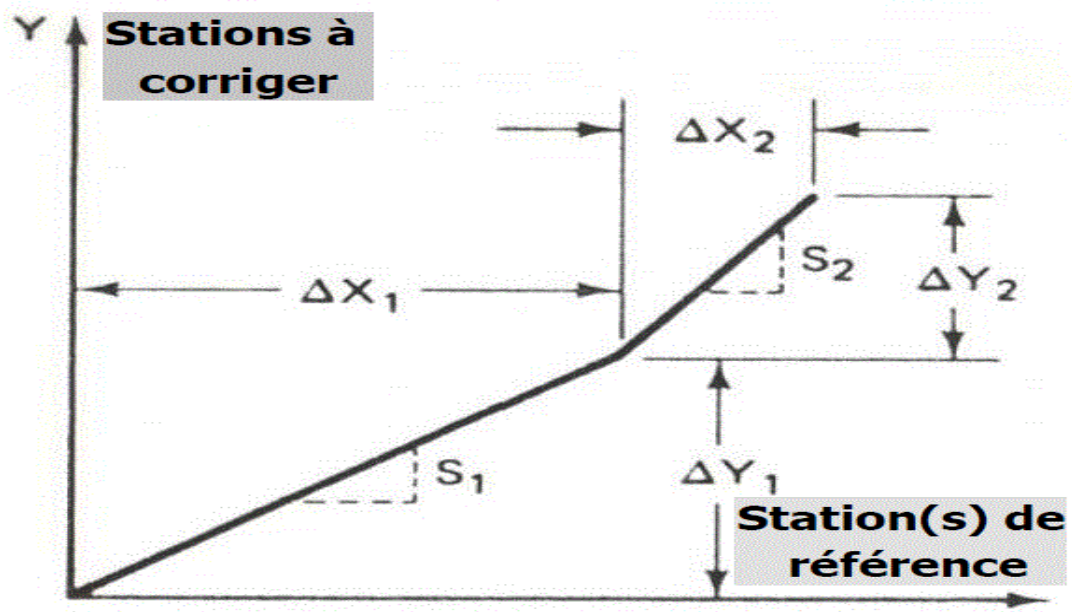


Figure 1.5 la méthode du double cumul

## 1.4 Conclusion

La recherche de modèles optimaux qui contrôlent le phénomène de la pluie a toujours été l'une des préoccupations de nombreux chercheurs dans ce domaine afin de corriger le manque de données et de les corriger pour résoudre les problèmes rencontrés par l'étude.

**Chapitre 02 : Analyse des  
données pluviométriques et  
le choix du type de loi  
d'ajustement**

## 2.1 Introduction

L'analyse statistique des données pluviométrique vise à définir des régimes à travers certaines valeurs caractéristiques et représentatives (moyennes, quartiles, valeurs extrêmes

Pour l'étude statistique des précipitations en hydrologie, on dispose d'une série pluviométrique qui comporte les précipitations maximales journalières pour la période la plus longue possible.

## 2.1 Analyse des données statistiques

### 2.1.1 Analyse descriptive

Pour l'étude des précipitations, nous avons besoin d'une série pluviométrique qui comporte les précipitations maximales journalières pour la période la plus longue possible.

Nous prenons comme base de calcul la série pluviométrique de station météorologique).

L'analyse statistique des données pluviométriques consiste à déterminer les caractéristiques empiriques d'un échantillon d'une série d'observations de précipitations maximales journalières, de 30 années. [13]

Les caractéristiques de cette série sont :

- Moyenne des précipitations maximales journalières ( $\bar{X}$ ) :

$$\bar{X} = \frac{\sum X_i}{N} \dots\dots\dots (1)$$

Où :

$X_i$  : précipitations journalières maximales afférentes à l'année considérée (**mm/j**)

$N$  : taille de l'échantillon (**N=40**)

$\bar{X}$  : Pluie journalière maximale annuelle moyenne (**mm/j**)

- Coefficient de variation ( $C_v$ ) :

$$C_v = \frac{\delta_x}{\bar{X}} \dots\dots\dots (2)$$

- Ecart type :

$$\delta_x = \sqrt{\frac{\sum(X_i - \bar{X})^2}{N}} \dots\dots\dots (3)$$

- **Médiane.** La médiane d'une série de nombres disposés en ordre croissant est la valeur centrale ou la moyenne arithmétique des deux valeurs centrales. Dans un polygone cumulatif de fréquences, la médiane est la valeur correspondant à 0,5.
- **Mode.** Le mode est la valeur de fréquence maximale.
- L'exposant climatique :  
**b** : Selon les études régionales. [13]

L'ensemble de ces valeurs ponctuelles, condensées sous forme statistique, est utilisé pour déterminer la fréquence et les caractéristiques d'un événement pluvieux isolé ou encore la variabilité de la pluviométrie dans l'espace:

- La hauteur totale de pluie.
- la durée.
- l'intensité moyenne.
- les intensités maximales sur des intervalles de temps quelconques.

La distribution d'intensité instantanée  $i(t)$  [13]

## 2.1.2. Choix de la loi d'ajustement

### 2.1.2.1 Loi de Gauss ou loi normale

#### 2.1.2.1.1 Fonction de répartition

$$F(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{u^2}{2}} \quad \text{Du avec} \quad u = \frac{x - \bar{X}}{\sigma_x} \dots\dots\dots (4)$$

**u** est appelée variable réduite de Gauss. L'intervalle de définition est donc :  $x \in ] - \infty, + \infty$  [Cette loi est symétrique ( $\beta_1 = 0$ ) et présente, par définition, l'aplatissement moyen ( $\beta_2 = 0$ ). [20]

Le tableau ci-dessous donne la fréquence au non-dépassement de **u** :

$$F(u) = \text{Prob}(U \leq u)$$

**Tableau 2.1 Table de la probabilité au non-dépassement est le complément à 1 de F (u).**

u	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7290	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8483	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9013
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9662	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2	0,9772	0,9779	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

Le paramètre de tendance centrale  $\bar{X}$  et à la fois la moyenne, le mode et la médiane. Le paramètre de dispersion  $\sigma_x$  est l'écart-type (racine carrée de la variance).i

Le paramètre de tendance centrale  $\bar{X}$  et à la fois la moyenne, le mode et la médiane. Le paramètre de dispersion  $\sigma_x$  est l'écart-type (racine carrée de la variance). [20]

La densité de probabilité a pour expression :

$$f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \dots\dots\dots(5)$$

La dérivée première s'annule pour  $u = 0$  (valeur modale) et la dérivée seconde pour  $u \pm 1$ , La fonction de densité est représentée par la courbe "en cloche" symétrique par rapport au mode  $u = 0$  et avec des points d'inflexion pour  $u \pm 1$ . [20]

**2.1.2.1.2 Estimation des paramètres****a) Méthode des moments :**

On obtient immédiatement les estimations suivantes :

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \text{ Et } \sigma_x = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}} \dots\dots(6)$$

**b) Méthode du maximum de vraisemblance :**

Cette méthode conduit à une autre estimation de l'écart-type :

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}} \dots\dots(7)$$

Elle est consistante mais moins correcte que la précédente que l'on préférera donc [12]

**c) Ajustement graphique :**

A chaque élément de l'échantillon, on associe une fréquence expérimentale au non-dépassement :  $F = \frac{i-0.5}{n}$

A chaque valeur de  $F(x_i)$  correspond une valeur de  $u$ . Pour chaque élément de l'échantillon, on peut porter dans un graphique gradué en abscisse selon les valeurs de  $u$  et en ordonnée selon la variable  $x$ , un point de coordonnées :  $(\frac{i-0.5}{n}, X_i)$

Si les valeurs de  $x_i$  sont distribuées normalement, les points s'alignent sur une droite d'équation :  $x_i = \bar{x} + u i \sigma_x$ , appelée droite de Henry.

D'après cette droite, on détermine la moyenne  $\bar{x}$  puisque pour  $u = 0$ , on a  $x = \bar{x}$ , et l'écart-type en faisant par exemple  $u = 2$ , donc  $x = \bar{x} + 2\sigma_x$ . [20]



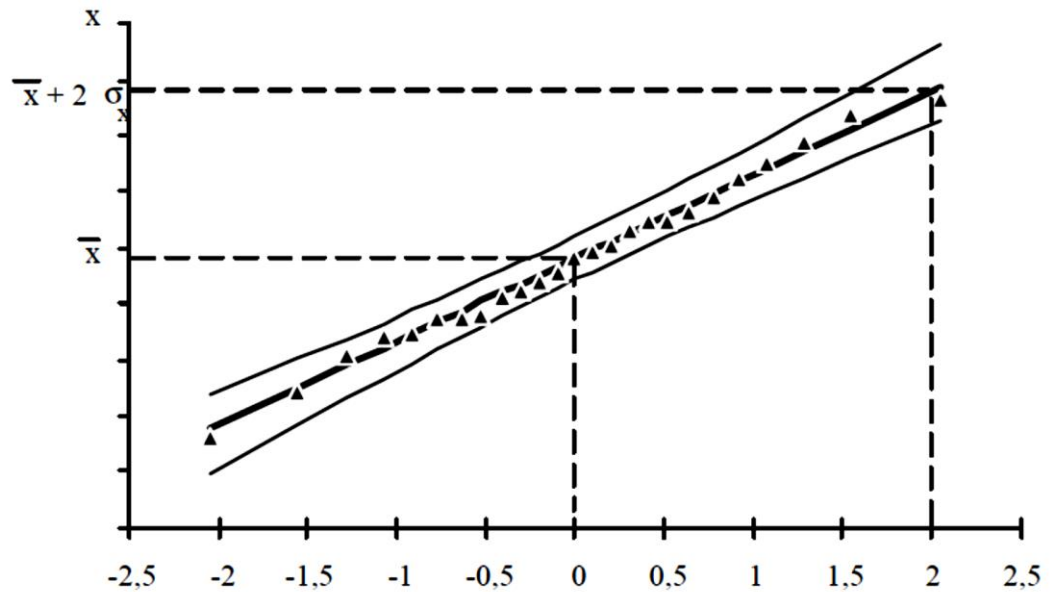


Figure 2.1 Ajustement graphique à une loi de Gauss.

### 2.1.2.1.3. Intervalles de confiance

On appelle intervalle de confiance à  $\alpha$  % sur l'estimation  $\hat{p}$  d'un paramètre  $p$ , l'intervalle limité par deux bornes  $p_1$  et  $p_2$  telles que l'on ait  $\alpha/2$  % de chance d'avoir  $p < p_1$  ou  $p > p_2$  du seul fait des erreurs d'échantillonnage. [21]

On a donc  $\alpha$  % de chance pour que la véritable valeur  $p$  (du paramètre que l'on estime par  $\hat{p}$ ) soit comprise entre  $p_1$  et  $p_2$ . Compte tenu des réflexions sur les distributions des estimations, on obtient :

\* Intervalle de confiance à  $\alpha$  % sur la moyenne :

Les bornes de l'intervalle sont :

$$\bar{X} \pm t_{1-\frac{1-\alpha}{2}} \frac{\sigma}{\sqrt{n}} \dots\dots(8)$$

- si  $n < 30$ ,  $t_{1-\frac{1-\alpha}{2}}$  est la valeur du  $t$  de Student de fréquence au non-dépassement  $F=1-\frac{1-\alpha}{2}$  pour :  $v=n-1$

- si  $n > 30$ ,  $t_{1-\frac{1-\alpha}{2}}$  la variable réduite de Gauss de fréquence au non-dépassement  $F=1-\frac{1-\alpha}{2}$  pour  $v=n-1$

\* Intervalle de confiance à  $\alpha$  % sur l'écart-type :

- si  $n < 50$ , les bornes de l'intervalle de confiance à  $\alpha$  % sont :

$$\frac{n\sigma_X^2}{x_1^2} < V < \frac{n\sigma_X^2}{x_2^2} \dots\dots(9)$$

$x_1^2$  : Valeur ayant pour fréquence au non-dépassement  $F = \frac{1-\alpha}{2}$

$x_2^2$  : Valeur ayant pour fréquence au non-dépassement  $F = 1 - \frac{1-\alpha}{2}$

- si  $n > 50$ , les bornes sont :

$$\sigma_X \pm t_{1-\frac{1-\alpha}{2}} \frac{\sigma_X}{\sqrt{2n}} \dots\dots(10)$$

Avec  $t_{1-\frac{1-\alpha}{2}}$ : variable réduite de Gauss de fréquence au non-dépassement :  $F = 1 - \frac{1-\alpha}{2}$

\* Intervalle de confiance à  $\alpha$  % sur le quantile  $x_F$  :

Les bornes de l'intervalle de confiance sont alors :

$$XF \pm \frac{t\alpha}{\sqrt{n}} \sqrt{1 + \frac{tF^2}{2} \pm \frac{tFt\alpha^2}{2n}} \alpha \dots\dots(11)$$

(2 fois plus pour la borne supérieure et 2 fois moins pour la borne inférieure)

$tF$  : variable réduite de Gauss ayant la fréquence au non-dépassement  $F$

$t\alpha$  : variable réduite de Gauss ayant la fréquence au non-dépassement  $1 - \frac{1-\alpha}{2}$ . [13]

### 2.1.2.2 Loi de Galton (loi log-normale)

#### 2.1.2.2.1 Fonction de répartition

Parmi les nombreuses formulations possibles, nous retiendrons la suivante, qui fait le mieux le parallèle avec la loi de Gauss :

$$F(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{u^2}{2}} \dots\dots(12)$$

Du :  $u = a \log(x - x_0) + b$

Intervalle de définition :  $x \in ]x_0, +\infty [$

On remarque que l'on passe de la loi de Gauss à la loi de Galton en faisant le changement de variable de  $x$  en  $\log(X - X_0)$ . Cette loi fait intervenir trois paramètres dont  $X_0$  est le paramètre de position [11]

Une autre formulation qui fait mieux ressentir le rôle des paramètres est :

$$F(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^u \frac{1}{u} e^{-\frac{\text{Ln}^2(u)}{2\sigma^2}} \dots\dots(13)$$

Du :

Xo : paramètre de position (identique au précédent)

σ : paramètre de forme positif différent de zéro

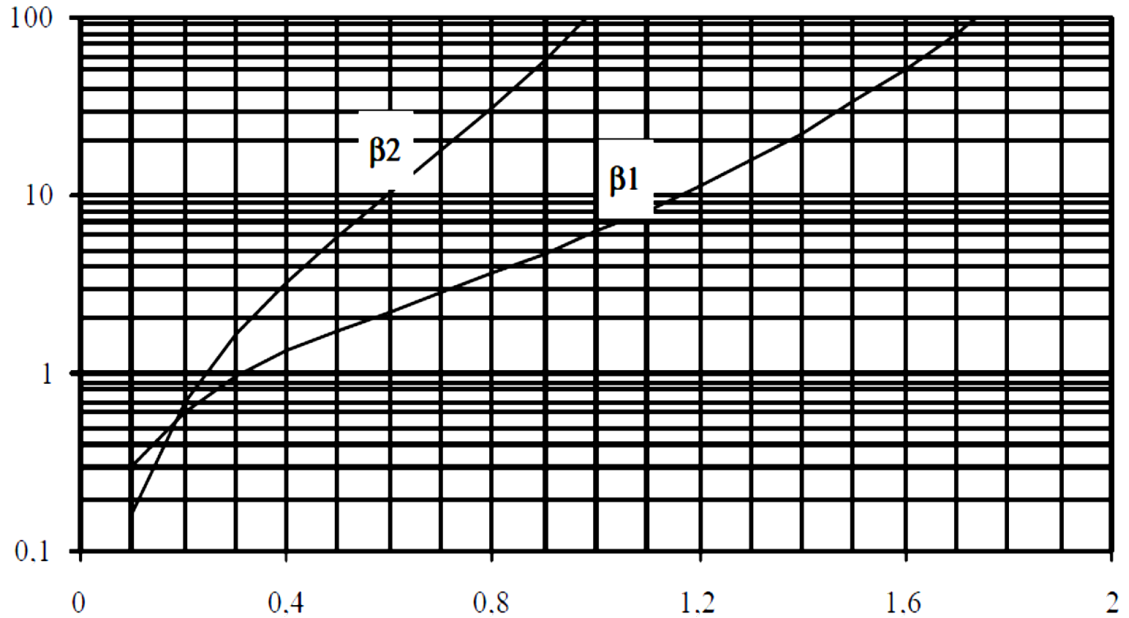


Figure 2.2 Variations de β1 et β2 avec σ.

Le coefficient d'aplatissement β2 est toujours positif ; la loi de Galton est plus "pointue" que la loi de Gauss.

Le coefficient d'asymétrie β1 est toujours positif ; la loi de Galton est dissymétrique, étendue vers la droite [11].

### 2.1.2.2.2 Estimation des paramètres

#### a) Méthode des moments :

En utilisant la première formulation, les trois paramètres à déterminer sont xo, a et b. En Posant  $\bar{x}$ : moyenne des x

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \quad : \text{Écart-type des } x,$$

$$\mu_3 = \frac{1}{(n-1)(n-2)} \sum_{i=1}^n (X_i - \bar{X})^3 \text{ Moment centré d'ordre 3 on obtient trois équations à}$$

trois inconnues :

$$\frac{\sigma^4}{\mu_3} = \frac{(\bar{X}-X_0)^3}{\sigma^2+3(\bar{X}-X_0)^3} \dots\dots(14)$$

$$a = \frac{1.517}{\sqrt{\log\left[1+\frac{\sigma^2}{(\bar{X}-X_0)^2}\right]}} \dots\dots(15)$$

$$b = \frac{1.1513}{a} - a \log(\bar{X} - X_0) \dots\dots(16)$$

L'équation (1) ne dépend que de  $X_0$  et des valeurs prises dans l'échantillon.

Malheureusement, il n'est pas possible d'expliciter  $X_0$  et on ne peut résoudre l'équation (1) par approximations successives. Connaissant  $X_0$  on résout l'équation (2) qui donne  $a$  ; puis connaissant  $a$  et  $x_0$  l'équation (3) qui donne  $b$ .

La méthode des moments présente le double inconvénient de nécessiter un calcul itératif et de donner parfois des valeurs de  $X_0$  supérieures à la plus petite des valeurs rencontrées dans l'échantillon [11].

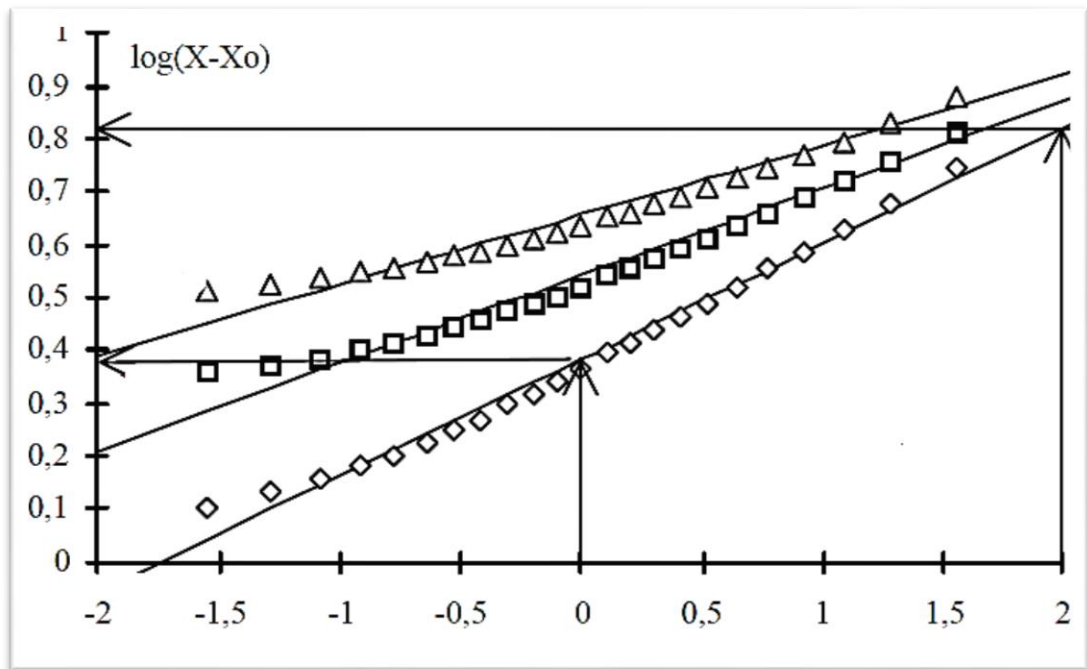
**b) Méthode du maximum de vraisemblance :**

On obtient trois équations à trois inconnues  $X_0$ ,  $a$  et  $b$  :

$$2.3026 \left[ \sum_{i=1}^n \frac{1}{X_i - X_0} \right] \left[ \frac{1}{n} \sum_{i=0}^n \log^2(X_i - X_0) - \frac{1}{n^2} \left( \sum_{i=0}^n \log(X_i - X_0) \right)^2 \right] = \frac{1}{n} \left[ \sum_{i=0}^n \log(X_i - X_0) \right] \left[ \sum_{i=0}^n \frac{1}{X_i - X_0} \right] - \sum_{i=0}^n \frac{\log(X_i - X_0)}{X_i - X_0} \dots\dots(17)$$

$$a^2 = \frac{1}{\frac{1}{n} \sum \log^2(X-X_0) - \frac{1}{n^2} [\sum \log(X-X_0)]^2} \dots\dots(18)$$

$$b = - \frac{a \sum \log(X-X_0)}{n} \dots\dots(19) \quad [20]$$

**c) Méthode graphique :**

**Figure 2.3 Ajustement graphique Loi de Galton**

On passe de la loi de Gauss à la loi de Galton par le changement de variable  $x$  en  $\log(x - x_0)$ . Par conséquent, si on connaît  $X_0$ , on pourra tracer l'équivalent de la droite de Henry en portant  $\log(x_i - x_0)$  en fonction de  $F = \frac{i-0.5}{n}$  sur un papier à probabilité graduée suivant les valeurs de la variable réduite de Gauss. Pour déterminer  $X_0$ , on procédera par tâtonnements en se fixant différentes valeurs de  $x_0$  jusqu'à obtenir l'alignement des points expérimentaux comme le montre la figure ci-dessous. [13]

Pour déterminer  $a$  et  $b$ , on prendra deux points, par exemple :

$$u = 0 \Rightarrow a \log(X - X_0) + b = 0$$

$$u = -2 \Rightarrow a \log(X - X_0) + b = -2 \quad [13]$$

**2.1.2.2.3 Intervalles de confiance**

L'étude théorique de la distribution des erreurs sur l'estimation des trois paramètres n'a pas donné de résultats facilement exploitables.

La seule méthode utilisable pour déterminer un intervalle de confiance sur un quantile  $X$  est de faire une simulation. On suppose que les trois paramètres calculés sont exacts et on tire au hasard dans la population-mère ainsi définie  $m$  échantillons de taille  $n$ . On obtient ainsi  $m$  estimations du quantile  $X_p$ . On admet, après avoir rangé ces  $m$  valeurs, que celles de rang  $i =$

10 et de rang  $m-9$  représentent, avec une bonne approximation, les bornes de l'intervalle de confiance à  $\frac{m-19}{m}$  [11]

**Tableau 2.2 les bornes de l'intervalle de confiance**

Nombre de tirages à effectuer	38	76	190	1900
Intervalles de Confiance à	50%	75%	90%	99%

### 2.1.2.3. Loi de Gumbel

#### 2.1.2.3.1. Fonction de distribution

Sous sa forme simple et généralement utilisée on a :

$$F(x) = e^{-e^{-u}} \dots\dots(20)$$

$$\text{Avec : } \frac{x - x_0}{s}$$

Le terme  $u$  est la variable réduite de Gumbel. L'intervalle de Variation de  $x$  est donc :  $x \in ] - \infty, \infty [$

- $X_0$  est le paramètre de position (mode)
- $s$  est le paramètre d'échelle différent de zéro et positif.

Le coefficient d'asymétrie  $\beta_1$  est constant et égal à 1,139 ; ce qui correspond à une dissymétrie étendue vers la droite [20]

Le coefficient d'aplatissement  $\beta_2$  est constant et égal à 2,4 ; la loi de Gumel est donc plus pointue que la loi de Gauss.

La densité de probabilité a pour expression :

$$f(u) = e^{-u} e^{-e^{-u}} \dots\dots(21)$$

Elle a un maximum pour  $u = 0$  et deux points d'inflexion pour  $u = \text{Ln} [1/2 (3 \pm \sqrt{5})]$ .

On peut noter dès à présent le comportement asymptotique de la distribution de Gumbel : si  $F$  tend vers 1, en posant  $T = 1/ 1 - F$ , la variable réduite de Gumel  $u$  tend vers  $\text{Ln } T$  [13]

#### 2.1.2.3.2. Estimation des paramètres

**a) Méthode des moments :**

La méthode des moments conduit à des estimations de calcul particulièrement simple et d'emploi très fréquent :

$$s = 0,78\sigma_x$$

$$X_0 = \bar{x} - 0,577 s$$

$$\left[ \sigma_x = \sqrt{\frac{\sum(x-x)^2}{n-1}} \quad \text{et} \quad \bar{X} = \frac{\sum X}{n} \right] \dots\dots (22)$$

**b) Méthode du maximum de vraisemblance :**

On aboutit au système résolutif suivant :

$$S + \frac{\sum X e^{-X/s}}{\sum e^{-X/s}} = \bar{X} \dots\dots(23)$$

$$X_0 = S \text{ Ln} \left[ \frac{\sum e^{-X/s}}{n} \right]$$

La première équation est implicite en s ; elle ne peut donc se résoudre que par itération. Ayant déterminé s, la deuxième équation permet la détermination de la méthode du maximum de vraisemblance ne s'utilise que lorsque l'on dispose de moyens de calcul automatique. [20]

**c) Méthode graphique :**

Cette loi à deux paramètres se prête très facilement à des ajustements graphiques.

La variable réduite de Gumbel  $u = \frac{x-x_0}{s}$  s'exprime très facilement en fonction de la fréquence puisque l'on a  $F(x) = e^{-e^{-u}}$  d'où  $u = -\text{Ln} \left[ \text{Ln} \frac{1}{F} \right]$

Ainsi, en portant arithmétiquement les valeurs de x en fonction de leur fréquence expérimentale au non dépassement selon un axe gradué arithmétiquement en u, on doit obtenir une droite passant par exemple :

$$\text{Paru} = 0 \text{ pour } x = x_0$$

$$\text{Et } u = \pm 1 \text{ pour } x = x_0 \pm s \text{ d'où la détermination immédiate de } x_0 \text{ et.}$$

s représente la pente de cette droite ; aussi appelle-t-on souvent s le "gradex" (gradient exponentiel). [11]

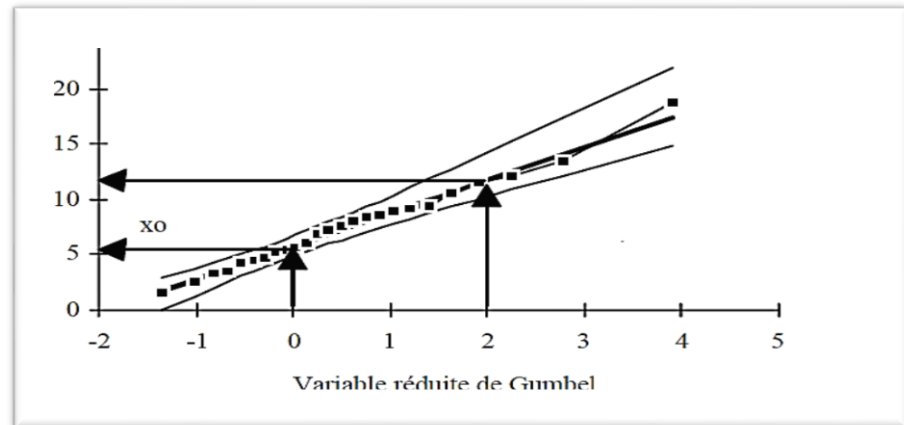


Figure 2.4 Ajustement graphique à la loi de Gumbel

### 2.1.2.3.3 Intervalle de confiance

BERNIER et VERON ont étudié les intervalles de confiance de la loi de Gumbel.

L'intervalle de confiance à  $\alpha$  % sur un quantile  $x_F$  s'exprime en fonction de l'écart-type  $\sigma_x$  par

$$\hat{x}_F - h_1 \sigma_x \leq x_F < \hat{x}_F + h_2 \sigma_x$$

Où  $h_1$  et  $h_2$  sont des paramètres dépendant de la taille  $n$  de l'échantillon de la fréquence  $F$  et de la valeur de  $\alpha$ .  $h_1$  et  $h_2$  seront évalués par la formule suivante (avec le signe + pour  $h_2$  et le signe - pour  $h_1$ ) :

$$h_{1,2} = \frac{\frac{t_\alpha}{n} \sqrt{1 + 1.13 t_F + 1.1 t_F^2} \pm \frac{t_\alpha^2 \sigma}{n} (1.1 t_F + 0.57)}{1 - 1.1 \frac{t_\alpha^2}{n}}$$

$t_\alpha$  est la variable réduite de Gauss correspondant à la fréquence au non-dépassement

$$1 - \frac{1-\alpha}{2}$$

$t_F$  est la variable réduite de Gumbel correspondant à la fréquence au non-dépassement

$F$ , ramenée à sa moyenne et à son écart-type :  $t_F = \frac{-\text{Ln}(-\text{Ln } F) - 0.577}{1.28}$  [20]

## 2.2 Teste d'ajustement

### 2.2.1 Teste de Khie 2

Parmi tous les tests existants, le test du  $\chi^2$  de PEARSON est le plus pratique et le plus utilisé bien qu'il ne soit pas très puissant [13]



Pour mettre en œuvre ce test, on procède de la façon suivante :

- On découpe l'échantillon de n éléments, en K classes (K>4) de façon à ce que chaque classe contienne théoriquement plus de 5 éléments ;
- Soit  $n_i$  le nombre d'éléments contenus dans la  $i^{\text{ème}}$  classe et  $n'_i$  (>5) le nombre théorique qui devrait s'y trouver ; on forme le terme suivant :

$$X^2 = \sum_{i=1}^k \frac{(n_i - n'_i)^2}{n'_i} \quad [20]$$

Cette variable  $X^2$  suit une loi de probabilité dite loi du  $X^2$ , à degrés de liberté avec :

$$V = K - P - 1$$

K = nombre de classes

p = nombre de paramètres estimés d'après l'échantillon.

Les tables données en annexes permettent de connaître la valeur numérique du  $X^2$  qui  $\alpha$  % de chance d'être dépassé par le seul fait du hasard, en fonction des valeurs de v [24]

### 2.2.2 Test Wn2 d'Anderson

Ce test, un peu plus fastidieux à mettre en œuvre (tout au moins manuellement), est plus efficace que le test du  $X^2$ . Analogue dans son esprit, il favorise les valeurs extrêmes et considère individuellement chaque élément de l'échantillon.

Soit un échantillon de taille n où  $x_i$  est la  $i^{\text{ème}}$  valeur dans l'échantillon classé par ordre croissant ; on peut estimer pour chaque  $x_i$ , sa fréquence théorique au non-dépassement F ( $x_i$ ) à partir de l'ajustement que l'on désire tester. On construit alors ainsi Wn2 :

$$W_n = n - \frac{1}{n} \sum_{i=1}^n ((2i - 1) \ln[F(x_i)] + (2n - 2i + 1) \ln[1 - F(x_i)]) \dots\dots (24)$$

La variable u définie comme suit à partir de Wn2 permet de caractériser la qualité de l'ajustement :

$$U = \frac{\ln[W_n^2 - 0.18/n^{1/4}] + 0.8\sqrt{n}}{0.65} \dots\dots (25)$$

## 2.3 Conclusion

L'analyse statistique permet la collecte des informations hydrologiques qu'elle représente la série mesurée sur plusieurs années dans quelques paramètres qui reflètent ce phénomène.

Les méthodes de modulation donnent des résultats très proches les uns des autres: la méthode des moments est beaucoup plus rapide, mais elle présente des inconvénients par rapport à la représentation graphique: l'ajustement graphique permet d'identifier et d'ignorer les éventuels points non alignés. Changer la pente et ainsi changer les normes du code statistique. En général, l'édition manuelle fournit souvent beaucoup d'informations sur la chaîne étudiée

# **Chapitre 03 : Analyse multi variation (Corrélation multiple et ACP)**

### 3.1 Introduction

Les régressions et corrélations multiples sont d'un grand intérêt en hydrologie pour étudier les relations entre les variables hydrologiques, climatologiques, morphologiques, etc.

La théorie complète des régressions et corrélations multiples est d'un abord assez difficile et il existe déjà à son sujet une abondante littérature. Nous nous bornerons ici à en donner les grandes lignes et insisterons plus particulièrement sur les points qui sont essentiels pour une application correcte de la théorie.

Les méthodes d'analyse multidimensionnelles traitent des tableaux des données pour lesquels les colonnes correspondent aux variables et les lignes aux unités statistiques (individus). L'analyse en composantes principales est un autre outil permettant une meilleure visualisation de nos données.

### 3.2 Corrélation multiple

#### 3.2.1 Généralités

Leur utilisation peut se classer sous trois rubriques :

a) extension dans le temps de séries d'observations hydrologiques qui sont de trop courte durée ou comportent des lacunes.

b) prévisions de données hydrologiques (apports mensuels, crues, étiages, etc.) en fonction des conditions hydrométéorologiques observées au moment de la prévision.

c) extension géographique à des bassins non observés des caractéristiques hydrologiques déterminées sur divers bassins versants de régime analogue.

La théorie complète des régressions et corrélations multiples est d'un abordasse difficile et il existe déjà à son sujet une abondante littérature. Nous nous bornerons ici à en donner les grandes lignes et insisterons plus particulièrement sur les points qui sont essentiels pour une application correcte de la théorie [18]

#### 3.2.2 Objectif

Soient  $X$  et  $Y$  deux grandeurs statistiques quantitatives observées.

On souhaite :

1. Déterminer s'il existe une relation entre  $X$  et  $Y$ .

2. Caractériser la forme de la liaison (la relation) entre  $X$  et  $Y$  (positive ou négative, linéaire ou non linéaire, monotone ou non monotone).
3. Tester si la liaison est statistiquement significative.
4. Quantifie l'intensité de la liaison.
5. Valider la liaison identifiée. Est-ce qu'elle n'est pas le fruit d'un simple artefact ou le produit d'autres informations sous-jacentes dans les données ?

Attention, la position des variables est symétrique dans ce cadre. On ne veut pas évaluer l'influence d'une des variables sur l'autre, à la différence de la régression [22]

### 3.2.3 Propriétés

1. Il est de même signe que la covariance, avec les mêmes interprétations.
2.  $X$  et  $Y$  sont indépendants, alors  $r = 0$ . La réciproque est fausse.
3. Lorsque le couple de variables ( $X; Y$ ) suit une loi normale bi variée, et uniquement dans ce cas, nous avons l'équivalence  $r = 0$ ,  $X$  et  $Y$  sont indépendants. Dans ce cas, le coefficient de corrélation caractérise parfaitement la liaison entre  $X$  et  $Y$ . Dans les autres cas, le coefficient de corrélation constitue une mesure parmi les autres de l'intensité de la corrélation.
4. Le coefficient de corrélation constitue une mesure de l'intensité de liaison linéaire entre 2 variables.  
Il peut être égal à zéro alors qu'il existe une liaison fonctionnelle entre les variables. C'est le cas lorsque la liaison est non monotone.
5. La corrélation d'une variable avec elle-même est  $r_{xx} = 1$  [22]

### 3.2.4 Domaine de définition

Le coefficient de corrélation est indépendant des unités de mesure des variables, ce qui autorise les comparaisons. La mesure est normalisée, elle est définie entre 2

$$-1 < r < 1$$

Lorsque :

- $r = +1$ , la liaison entre  $X$  et  $Y$  est linéaire, positive et parfaite c.-à-d. la connaissance de  $X$  nous fournit la valeur de  $Y$  (et inversement).
- $r = -1$ , la liaison est linéaire et négative [18]

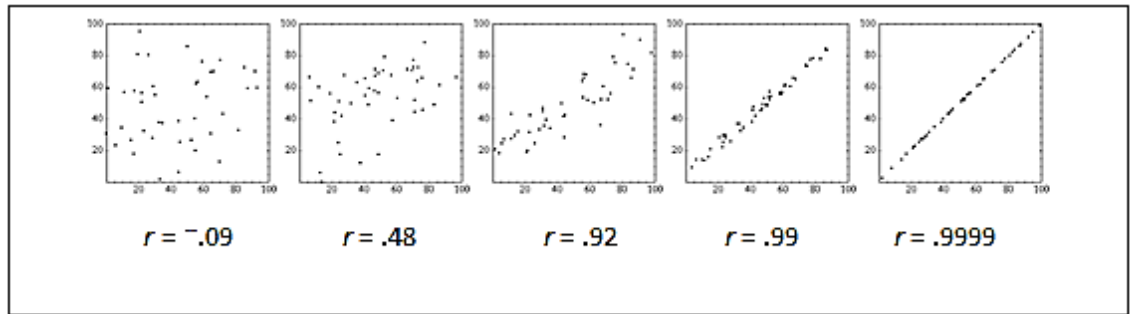
### 3.2.5 Statistique de corrélation

Qu'entend-t-on par corrélation ? Supposons deux échantillons sur un groupe, par exemple, quotient intellectuel et habilité en lecture. On s'attend à ce que ces deux mesures varient ensembles. C'est à dire que si une personne a un score élevé sur une mesure, l'autre mesure devrait aussi être élevée. Inversement, si une personne a un score faible sur une mesure, l'autre devrait aussi être faible. Dans ce cas, les mesures sont dites positivement corrélées [5]

Imaginons par opposition, deux autres mesures, l'habilité en lecture et le temps pour lire un passage donné. Dans ce cas-ci, on s'attend plutôt à ce qu'une personne avec un score élevé dans l'habilité en lecture montre un score petit (rapide) en lecture, et vice-versa. Dans ce second cas, les mesures sont dites négativement corrélées [5].

On peut concevoir le coefficient de corrélation comme un indice de la qualité de la droite idéale passant par les points (ou encore comme la pente quand les valeurs des deux variables ont été normalisée -transformée en cote  $z$ ). Les moyennes des deux variables sont alors zéro. Et la variance est 1. Les données autant de  $X$  que de  $Y$  s'étendent vraisemblablement entre  $-3$  et  $+3$  [17]

La corrélation est une statistique qui caractérise l'existence ou l'absence d'une relation entre deux échantillons de valeurs prise sur un même groupe de sujets. Le coefficient de corrélation permet de quantifier cette relation -1 par le signe de la corrélation (positive et négative), et par la force de cette corrélation. Le degré de corrélation, comme nous le verrons plus loin, se mesure sur une échelle de 0 à 1. Zéro signifie une totale absence de corrélation entre les deux mesures, alors que 1 signifie une corrélation parfaite, c'est à dire que connaître la valeur d'une mesure nous permet de connaître exactement la valeur de l'autre. Les illustrations de la **Figure 3.1** (appelées « squatteront » quand on illustre une mesure en fonction d'une autre mesure) donnent quelques valeurs possibles pour le coefficient de corrélation [8]



**Figure 3.1** Quelques valeurs possibles de coefficient de corrélation

On peut concevoir le coefficient de corrélation comme un indice de la qualité de la droite idéale passant par les points (ou encore comme la pente quand les valeurs des deux variables ont été normalisées -transformées en cote  $z$ ). Les moyennes des deux variables sont alors zéro. Et la variance est 1. Les données autant de  $\mathbf{X}$  que de  $\mathbf{Y}$  s'étendent vraisemblablement entre  $-3$  et  $+3$  [5]

On se rend compte que dans le cas où  $r = 0$ , les valeurs  $\mathbf{Y}$  élevées pourraient être autant associées à des valeurs  $\mathbf{X}$  élevées qu'à des valeurs  $\mathbf{X}$  basses. Et vice-versa. La meilleure

Prédiction possible de  $\mathbf{Y}$  ne dépend pas de la connaissance de  $\mathbf{X}$ . La connaissance de  $\mathbf{X}$  ne donne aucune information sur  $\mathbf{Y}$ . [8]

Comme on le voit, le nuage de point devient de plus en plus étroit au fur et à mesure que le coefficient devient élevé. Quand  $r$  est à son maximum (1), les données transformées de  $\mathbf{X}$  sont parfaitement prédites par les données transformées de  $\mathbf{Y}$ , c'est à dire

$$\frac{Y_i - \bar{Y}}{\bar{Y}} = \frac{X_i - \bar{X}}{\bar{X}} \dots\dots (1)$$

Dans le cas où  $r = -1$ , la relation est toujours vraie, sauf pour un signe moins :

$$\frac{Y_i - \bar{Y}}{\bar{Y}} = - \frac{X_i - \bar{X}}{\bar{X}} \dots\dots(2)$$

### 3.2.6 Corrélation simple

Nous noterons  $r_{\mathbf{XY}}$  le coefficient de corrélation entre deux échantillons  $\mathbf{X}$  et  $\mathbf{Y}$ . Il est aussi souvent appelé le coefficient de corrélation de Pearson, du nom de son inventeur, pour le distinguer d'autres indices de corrélations (tel le coefficient de Spearman) [17]

- **Calcul du coefficient de corrélation r**

Pour calculer le coefficient de corrélation, il faut premièrement pouvoir calculer la covariance entre deux échantillons. On se rappelle que la variance (non biaisée) se calcule [5]

Comme suit :

$$\overline{X^2} = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2 \dots\dots(3)$$

La covariance est une mesure de la variance présente dans deux échantillons simultanément. L'idée étant que si les deux échantillons covariant, la covariance devrait être grande, alors que s'ils ne covariant pas, la covariance devrait être modérément faible. Une façon d'atteindre cette mesure est d'utiliser le produit des différences, comme suit :

$$\overline{XY^2} = \frac{1}{n-1} \sum_i (X_i - \bar{X})(Y_i - \bar{Y}) \dots\dots(4)$$

Dans cette équation, si  $X_i$  tend à être très supérieur à sa moyenne en même temps que  $Y_i$ , la somme sera grande, indiquant une forte Co variation. La mesure de Co variation est exprimée en unité de  $X$  fois l'unité de  $Y$ . Dans le premier exemple ci-haut, la Co variation serait exprimée en point de Q.I par mots lus. Pour éliminer ces unités, on peut diviser par les écarts types des échantillons pris individuellement. Cette division a aussi pour résultat de normaliser la covariance entre  $-1$  et  $1$ , ce qui est donc l'indice de corrélation souhaité :

$$r_{XY} = \frac{\overline{XY^2}}{\overline{X} \times \overline{Y}} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2} \times \sqrt{\sum_i (Y_i - \bar{Y})^2}} \dots\dots(5)$$

Ce que nous avons en fait, c'est un ratio entre combien de variation les deux mesures ont en commun divisée par la quantité de variation qu'elles pourraient avoir au plus. Si on élève  $r$  au carré,  $r^2$  donne la quantité de variance en commun entre les deux échantillons. On parle aussi souvent de « pourcentage de la variance expliquée », car si on prend le point de vue que, disons,  $X$  explique les résultats obtenus en  $Y$ , une certaine quantité de variance en  $X$  explique la variance en  $Y$ , et cette quantité est donnée par  $r^2$ . Autrement dit, si nous connaissons la variable  $X$ , l'incertitude à propos de la variable  $Y$  est réduite de moitié. Une autre façon d'illustrer la variance expliquée est sous la forme d'un diagramme de Vienne. Dans ce cas, on peut voir  $r^2$  en termes de superposition de cercles [8]



### 3.2.7 Corrélation multiple

Jusqu'à présent, nous n'avons considéré que des corrélations entre deux variables. Il existe aussi des cas où trois variables ou plus sont impliquées. Par exemple, pour prédire le revenu familial d'un individu (**Z**), les indicateurs Degré de scolarité (**X**) et Revenu des parents (**Y**), pris individuellement n'expliquent peut-être qu'une partie de la variance, alors que lorsque les deux sont pris en considération simultanément, une bien meilleure prédiction peut être atteinte. Voir la Figure 3.2 qui illustre les contributions de la variance expliquée de chaque variable sur les autres :

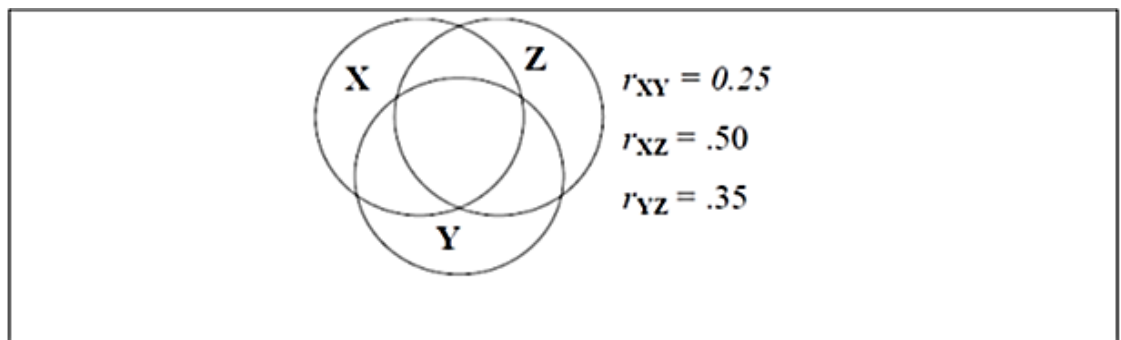


Figure 3.2 Représentation d'une corrélation pour trois variables

Une façon de considérer ce diagramme de Venn est de regarder la matrice des corrélations en prenant les variables deux par deux, ce qu'on appelle une table d'inter corrélation :

Tableau 3.1 une table d'inter corrélation

	Y	Z
X	0.25	0.50
Y		0.35

Cependant, cette table ne répond toujours pas à notre question puisqu'elle continue à prendre les prédictors un par un. Nous souhaitons prédire **Z** étant donné un couple **X** et **Y**

Connu simultanément. Pour y arriver, nous utilisons un indice de corrélation multiple  $R$ . [5]

- **Calcul du coefficient de corrélation  $R$**

L'indice de corrélation multiple (pour plus de deux variables) est représenté par un  $R$  majuscule pour le différencier du cas particulier où il n'y a qu'un total de deux variables.

Tout comme le  $r$ , l'indice  $R$  va de plus 1 à  $-1$ . Pour éviter les confusions, on utilise les indices tels :

$R_{Z, XY}$  pour indiquer que l'on cherche à prédire  $Z$  à partir des valeurs simultanées de  $X$  et  $Y$ . [17]

De la même façon,  $R_{Z, XY}$  indique le pourcentage de la variance de  $Z$  expliqué par  $X$  et  $Y$ . [5]

Prenons comme exemple une étude où l'on veut déterminer la relation entre la qualité des programmes qu'une personne écoute (selon son évaluation personnelle)  $X$  et le prix de son équipement  $Y$ , pour déterminer le nombre d'heure que cette personne va passer devant la télévision par semaine  $Z$ . Nous supposons que le chercheur a déjà obtenu les corrélations simples  $r_{XY}$ ,  $r_{XZ}$  et  $r_{YZ}$  pour chaque pair  $(X, Y)$ ,  $(X, Z)$ , et  $(Y, Z)$ . Le  $R$  multiple se calcule comme suit :

$$R^2_{Z,XY} = \frac{r^2_{XZ} + r^2_{YZ} - 2r_{XZ}r_{YZ}r_{XY}}{1 - r^2_{XY}} \dots\dots(6)$$

Un coefficient de corrélation multiple s'interprète de la même façon qu'un  $r$  régulier dans le cas d'un problème à deux variables. De plus, il est aussi possible de tester des hypothèses concernant  $R$ , tel  $H_0 : R = 0$ . Nous n'allons pas entrer dans les détails.

Cependant, un bon logiciel d'analyse statistique va rapporter si le  $R$  obtenu est significatif ou non. De la même façon, une droite de régression passant par les triplets de points est aussi possible. L'idée étant de prédire les  $Z$  suivant une équation de la forme

$$Z_i = b_{XZ,Y} + b_{YZ,X}Y_i + \alpha \dots\dots(7)$$

Une chose importante à voir dans cette équation est que l'effet de  $X$  et de  $Y$  sont additifs, c'est à dire que  $X$  affecte  $Z$  indépendamment de  $Y$ . Autrement dit, ce modèle stipule qu'il n'existe pas d'interaction entre les facteurs  $X$  et  $Y$  sur la variable dépendante  $Z$ . Cette hypothèse teste à l'aide d'une ANOVA. Si l'interaction  $A * B$  est significative, il faut simplement ne pas faire de régression multiple linéaire. [17]

Le facteur  $XZ, Y$   $b$  est le ratio entre combien d'unité  $Z$  change pour chaque unité de Changement dans  $X$  quand  $Y$  est tenu constant. Parce que chacun de ces coefficients représente seulement une portion de la prédiction de  $Z$ , ils sont appelés coefficients de corrélation partielle. Les équations pour calculer les pentes des effets de chaque variable individuelle sont donnés par :

$$b_{XZ,Y} = \frac{\bar{Z}}{\bar{X}} \times \frac{r_{XZ} - r_{YZ}r_{XY}}{1 - r^2_{XY}} \dots\dots(8)$$

$$b_{YZ,X} = \frac{\bar{Z}}{\bar{Y}} \times \frac{r_{YZ} - r_{XZ}r_{XY}}{1 - r^2_{XY}} \dots\dots (8')$$

### Exemple

Soit la recherche où l'on veut déterminer la relation entre la qualité des programmes qu'une personne écoute (selon son évaluation personnelle)  $X$  et le prix de son équipement  $Y$  (en \$), pour déterminer le nombre d'heure que cette personne va passer devant la télévision par semaine  $Z$  (en heures).

Le chercheur a obtenu ces trois mesures sur un échantillon de 50 personnes (non présentés). La moyenne des observations est  $X = 5$  (cote),  $Y = 205$  \$,  $Z = 18$  heures.

La variance (non biaisée) des observations est :  $X = 2.2$  (cote),  $Y = 88.0$ ,  $Z = 11.8$  heures. Il observe les corrélations simples  $r_{xy} = 0.894$ ,  $r_{xz} = 0.750$  et  $r_{yz} = 0.918$ . Il regarde en premier le coefficient de régression multiple  $R$  :

$$\begin{aligned} R^2_{Z,XY} &= \frac{r^2_{XZ} + r^2_{YZ} - 2r_{XZ}r_{YZ}r_{XY}}{1 - r^2_{XY}} \dots\dots(9) \\ &= \frac{0.894^2 + 0.918^2 - 2 \times 0.894 \times 0.918 \times 0.750}{1 - 0.750^2} = \frac{0.799 + 0.843 - 1.231}{0.438} \\ &= \frac{0.411}{0.438} = 0.938 \end{aligned}$$

Près de 94% de la variance est expliquée quand on considère  $X$  et  $Y$  simultanément, ce qui serait significatif si on regardait un listing d'ordinateur. Les coefficients de corrélation partielle sont :

$$\begin{aligned} b_{XZ,Y} &= \frac{\bar{Z}}{\bar{X}} \times \frac{r_{XZ} - r_{YZ}r_{XY}}{1 - r^2_{XY}} = \frac{11.8}{2.2} \times \frac{0.894 - 0.918 \times 0.750}{1 - 0.750^2} = 2.51 \\ b_{YZ,X} &= \frac{\bar{Z}}{\bar{Y}} \times \frac{r_{YZ} - r_{XZ}r_{XY}}{1 - r^2_{XY}} = \frac{11.8}{88.0} \times \frac{0.918 - 0.894 \times 0.750}{1 - 0.750^2} = 0.076 \end{aligned}$$

Avec les valeurs moyennes, on trouve l'ordonnée à l'origine,  $a$  :

$$\alpha = \bar{Z} - b_{XZ,Y}\bar{X} - b_{YZ,X}\bar{Y} = 18 - 2.51 \times 5 - 0.076 \times 205 = -10.1$$

Étant données ces différentes valeurs, nous pouvons prédire le temps passé devant la télévision si le prix de la télévision  $X$  et la cote des programmes écoutés  $Y$  sont connus. Par exemple, si un individu rapporte écouter des émissions qu'il cote 3 et que son équipement coûte 100\$, on s'attend à ce qu'il passe  $0.076 \times 100 + 2.51 \times 3 - 10.13 = 5$  heures par semaine de vante la télévision

### 3.3 Analyse en composantes principales ACP

#### 3.3.1 Généralités

L'analyse des données consiste essentiellement à établir quelles sont les relations existantes entre les observations, entre les variables, et entre les observations et les variables. Il s'agit donc de mettre un peu d'ordre dans le fichier de données, souvent de taille considérable (quelques centaines ou milliers d'analyses et quelques dizaines de variables. Ceci peut être fait à l'aide de diagrammes binaires, à l'aide des corrélations et corrélations partielles, à l'aide de diagrammes spécifiques propres au domaine étudié (ex. diagramme AFM et diagrammes de Harker en pétrologie). L'analyse en composantes principales est un autre outil permettant une meilleure visualisation de nos données [11]

#### 3.3.2 Objectif

La description des liaisons entre deux variables par des techniques statistiques bidimensionnelles conduisent à se poser la question des représentations simultanées de données en dimension plus grande que 2. Quel graphique permettrait de “généraliser” le nuage de points tracé dans le cas de deux variables permettant d’aborder la structure de corrélation présente entre plus de 2 variables. L’outil utilisé est alors l’analyse en composantes principales. [6]

Mathématiquement, l’analyse en composantes principales est un simple changement de base : passer d’une représentation dans la base canonique des variables initiales à une représentation dans la base des facteurs définis par les vecteurs propres de la matrice des corrélations [10]

#### 3.3.3 Méthodologie de l'ACP

Le tableau de données  $n \times p$  forme un nuage de  $n$  points dans un espace à  $p$  dimensions, ou un nuage de  $p$  points dans un espace à  $n$  dimensions. Un diagramme binaire consiste à projeter ces points sur deux des dimensions choisies plus ou moins arbitrairement. L'ACP consiste à projeter les points sur une droite, un plan... un sous-espace à  $s$  dimensions (avec  $s \leq p$ ) choisi de façon à optimiser un certain critère. Intuitivement, on cherchera le sous-espace donnant la meilleure visualisation possible de notre nuage de points. Un bon choix consiste à rechercher la plus grande dispersion (le plus grand étalement) possible des projections dans le

sous-espace choisi. On est amené ainsi à chercher une rotation de notre système d'axes initial (les variables) permettant de mieux voir notre nuage. Définissons  $u_1$  le vecteur unitaire (i.e. de norme 1;  $u_1'u_1=1$ ) recherché; c'est le vecteur présentant la plus grande dispersion des projections. [4]

Soit la matrice  $X_{n \times p}$ ; chaque ligne représente une observation. Chaque colonne représente une variable. On supposera chaque variable centrée, i.e. on a soustrait la moyenne de chaque variable au préalable. Ceci est fait de façon à faire coïncider le centre de gravité du nuage de points avec l'origine.

Les projections des  $n$  observations sur le vecteur  $u_1$  sont données par:

$$C = Xu_1 \text{ (produit scalaire)}$$

La somme des carrés de ces projections (inertie) est :  $C'C = u_1'X'Xu_1$

On choisira  $u_1$  de façon à maximiser cette dernière quantité. Le problème est donc : Maximiser  $u_1'X'Xu_1$  sujet à  $u_1'u_1=1$

Il s'agit d'un problème classique d'optimisation sous contrainte que l'on peut solutionner par la méthode de Lagrange.

On forme le Lagrangien =  $u_1'X'Xu_1 - \lambda (u_1'u_1 - 1)$

On dérive par rapport à chacune des  $p$  composantes du vecteur  $u_1$  ainsi que par rapport au multiplicateur de Lagrange ( $\lambda$ ) et on pose les dérivées partielles égales à zéro [4]

$$2[X'Xu_1 - \lambda u_1] = 0$$

$$u_1'u_1 = 1$$

Simplifiant, on trouve :

$$X'Xu_1 = \lambda u_1$$

$$u_1'u_1 = 1$$

On reconnaît là l'équation de vecteurs propres et de valeurs propres de la matrice  $X'X$ . On note que cette matrice est, à un facteur  $1/n$  ou  $1/(n-1)$  près, la matrice de variances (sur la diagonale) et de covariances (hors-diagonale) des  $p$  variables. Le vecteur donnant les

projections ayant la plus grande dispersion est donc le 1<sup>er</sup> vecteur propre de la matrice de variances-covariances de X (cf. remarque i. ici-bas). [6]

### 3.3.3.1 Les étapes d'un ACP

#### a) Choix du tableau X :

- **On travaille toujours sur le tableau centré** : On montre que tout axe factoriel passe par le centre de gravité : le nouveau repère est centré en G.

- **Travailler sur le tableau brut (centré par défaut) ou centré réduit?**

Si X n'est pas réduit, l'importance que prendront les variables dans le calcul des composantes principales est fonction de leur ordre de grandeur ; une variable d'écart-type important aura plus de poids qu'une variable d'écart-type faible. Des variables de fort écart-type construiront les premières composantes principales : les calculs ne sont pas faux, et conduisent aux mêmes interprétations mais la lecture des résultats risque d'être brouillée. [8]

⇒ On commence souvent par centrer et réduire X

#### b) Analyse directe

Construction de l'espace factoriel du nuage de points individus associé au tableau. On garde pour l'instant les p axes factoriels. [22]

#### c) Analyse duale

Construction de l'espace factoriel du nuage de points variables : elle est *déduite* de la première. [5]

#### d) Interprétation de ces analyses

Choix du nombre d'axes q à retenir, construction des nuages de points projetés sur ces axes, interprétation des axes principaux et étude des proximités entre points. [22]

#### e) Synthèse des résultats

Construction éventuelle du tableau C réduit (tableau des composantes principales) et visualisation des nuages de points associés. [16]

### 3.3.3.2 Interprétation géométrique de l'ACP

L'ACP revient à effectuer une rotation du système d'axes initial puisque les vecteurs propres sont orthogonaux entre eux et constituent donc un nouveau repère de coordonnées. Les cosinus entre les nouveaux axes et les anciens sont les composantes des vecteurs propres.

Une autre interprétation procède par analogie avec la régression. Le 1er vecteur propre est le vecteur qui explique le mieux, simultanément toutes les variables de la matrice X. Ce vecteur minimise la somme des carrés entre les projections sur le vecteur et la position des points dans l'espace original. Le deuxième vecteur propre est celui qui explique le mieux, simultanément, l'ensemble des résidus obtenus, et ainsi de suite... [4]

### 3.3.3.3 Analyse dans l'espace des échantillons

On s'est intéressé, jusqu'à maintenant, au nuage des n observations dans l'espace des p variables. On pourrait également considérer le nuage de p variables dans l'espace des n observations. On cherche le sous-espace de dimension s ( $s \leq p$ ) pour lequel la somme des carrés des projections est maximale [10]

On applique la même technique que précédemment et on trouve que la solution, pour le premier vecteur est donnée par le système suivant:

$$XX \cdot V_1 = \beta V_1 V_1' V_1$$

### 3.3.4 Exemple numérique

**Tableau 3.1 Les données sont exemple numérique**

<b>x</b>	3	4	6	6	6	7	7	8	9	9	9	10	11
<b>y</b>	2	10	5	8	10	2	13	9	5	8	14	7	12
<b>x</b>	11	12	12	13	13	13	13	14	15	17	17	18	20
<b>y</b>	12	10	11	6	14	15	17	7	13	13	17	19	20

On effectue l'ACP de la matrice des covariances :

**Tableau 3.3 Matrice des coefficients de corrélation**

	$u_1$	$u_2$
X	0.66314	0.7485
Y	0.7485	-0.66314

Les vecteurs  $\lambda_1 u_1$  et  $\lambda_2 u_2$  définissent les axes majeurs d'une ellipse passant par les points  $(\text{Var}(x), \text{Cov}(x, y))$  et  $(\text{Cov}(x, y), \text{Var}(y))$ . Les vecteurs  $21\lambda_1 u_1$  et  $22\lambda_2 u_2$  sont les axes majeurs d'une ellipse qui tracée dans le plan des coordonnées des observations devraient renfermer approximativement 86% des observations lorsque la distribution est (bi) normale. [10]

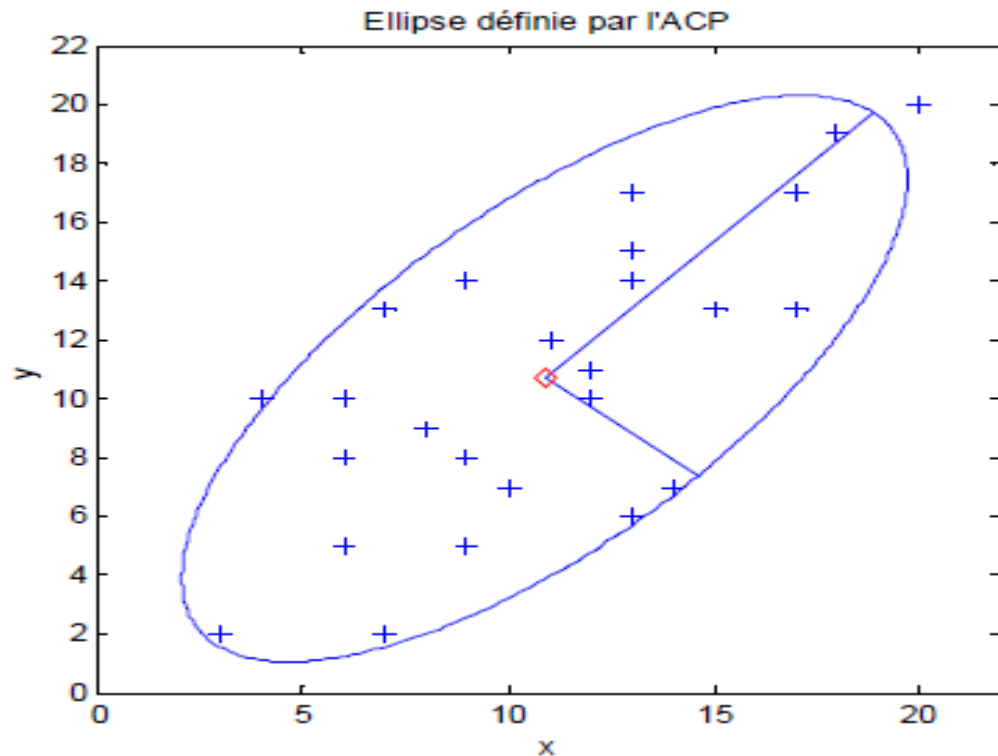


Figure 3. 3 Ellipse définie par l'ACP

### 3.3.5 ACP de la matrice des covariances

Dans cette forme d'ACP, on cherche les vecteurs propres et valeurs propres de la matrice  $X'X/n$  au lieu de  $X'X$  ( $X$  est centrée). Habituellement, on projette ensuite  $X$  sur les vecteurs propres. Les valeurs propres de cette matrice représentent des variances plutôt que des sommes de carré. Toutefois rien d'important ne change par rapport à l'analyse de  $X'X$ . Les vecteurs propres demeurent les mêmes, les projections également [14].

### 3.3.6 ACP de la matrice des corrélations

Si, en plus de centrer la matrice  $X$ , on avait divisé chaque valeur par l'écart-type de la variable correspondante, alors le produit  $X'X/n$  serait la matrice des corrélations. Il s'agit de



l'ACP la plus courante. L'avantage de cette approche est que les variables ne possédant plus d'unités, les résultats sont indépendants des unités originales choisies pour mesurer les variables. On peut ainsi considérer des variables dont l'ordre de grandeur des variations est très différent. Auparavant, une variable présentant une variation très grande aurait contribué beaucoup plus à la définition des vecteurs propres qu'une variable montrant peu de variations. Ainsi, le  $\text{TiO}_2$  exprimé en pourcentage, aura une importance beaucoup moins grande dans l'ACP de la matrice des variances-covariances que s'il est exprimé en ppm. Ce n'est pas le cas avec l'ACP de la matrice des corrélations et c'est ce qui fait que cette méthode est habituellement préférée. [6]

#### Particularités de l'ACP de la matrice des corrélations

- La trace de la matrice des corrélations (donc la somme des valeurs propres) est égale au nombre de variables. Comme pour l'ACP de la matrice des covariances, les valeurs propres représentent la variance des projections sur chaque vecteur. Dans les calculs des contributions des observations, il faudra en tenir compte (cf. ACP matrice des covariances).

Les coordonnées des variables sur les vecteurs propres sont égales aux corrélations entre ces variables et les composantes principales correspondantes. Par conséquent les coordonnées des variables sont toutes inférieures à 1 et supérieures à -1. Également, si on trace un cercle unité sur un plan factoriel quelconque, on est assuré que les projections des variables tomberont toutes à l'intérieur de ce cercle. [10]

En effet, les coordonnées des variables sont données par:  $X'V=U\Lambda^{1/2}$

La variance d'une composante principale est donnée par  $\lambda_i$  (note : ici  $\lambda_i$  est valeur propre de la Matrice des corrélations).

La variance d'une variable originale est 1.

Les covariances entre composantes principales et variables originales sont données par :

$$1/n X'C= 1/n X'XU = U\Lambda$$

La corrélation est donc :  $U\Lambda\Lambda^{-1/2} = U\Lambda^{1/2}$ , i.e. la corrélation est égale à la coordonnée des variables sur les vecteurs propres. Finalement, puisque la qualité de la représentation est ici égale à la coordonnée, au carré, de la variable et que la somme des qualités de représentation donne nécessairement 1, il suit que les coordonnées des variables sont sur la

surface d'une hyper sphère de rayon unité. Plusieurs programmes impriment automatiquement le cercle unité sur les divers plans factoriels. [5]

- **Cas particulier de l'ACP avec 2 variables**

Lorsqu'il n'y a que deux variables, des relations très simples existent pour les valeurs propres et vecteurs propres de l'ACP de la matrice des corrélations. Les composantes des vecteurs propres sont toutes égales à  $\pm (1/2)^{0.5}$ . Les valeurs propres sont égales à  $1 \pm r_{xy}$ . Dans une figure mettant en relation  $X/s_x$  et  $Y/s_y$ , la pente du 1er vecteur propre est donc 1 ou -1 selon que la corrélation entre x et y est positive ou négative. Si on construit la figure Y vs X, alors on aura une droite dont la pente est  $s_y/s_x$  que l'on peut utiliser pour effectuer des prédictions lorsque X et Y ont des rôles similaires (ex. longueur et largeur d'un fossile). Cette droite de prédiction est connue sous le nom de "redue major axis". On peut démontrer que c'est la droite qui minimise la somme des produits croisés des erreurs sur X et des erreurs sur Y. [16]

$$\text{Min}[\sum_{i=1}^n |(X_i - \hat{X}_i)(Y_i - \hat{Y}_i)|] \dots\dots(10)$$

Avec les valeurs prédites données par le modèle unique:

$$\hat{Y} = b_0 + b_1 \dots\dots(11)$$

$$\hat{X} = \frac{Y - b_0}{b_1} \dots\dots(12)$$

Géométriquement, ceci revient à minimiser la somme des surfaces des triangles rectangles formés par la droite et les points expérimentaux. [16]

Pour l'ACP de la matrice des covariances, on parle d'axe majeur, mais il n'y a pas les mêmes relations simples pour déterminer les valeurs propres et vecteurs propres.

On peut aussi généraliser la notion d'axe majeur ou d'axe majeur réduit à plus de deux dimensions. On obtient ainsi, en général, un hyperplan de dimension p-1 orienté dans un espace à p dimensions et qui permet de prédire n'importe quelle variable étant donné que l'on connaît les p-1 autres variables. Cette équation s'obtient à partir du vecteur propre associé à la plus petite valeur propre. En effet, la projection sur le plus petit vecteur propre est  $Xu_1$ . Si

cette projection est 0, alors l'observation se trouve dans l'espace complémentaire (i.e. hyperplan de dimension p-1). Donc, on doit avoir, dans le cas de l'ACP de la matrice des corrélations:

$$\mathbf{0} = \frac{(X_1 - \bar{X}_1)}{s_1} \mathbf{u}_{p1} + \frac{(X_1 - \bar{X}_1)}{s_2} \mathbf{u}_{p2} + \dots + \frac{(X_p - \bar{X}_p)}{s_p} \mathbf{u}_{pp} \dots\dots(13)$$

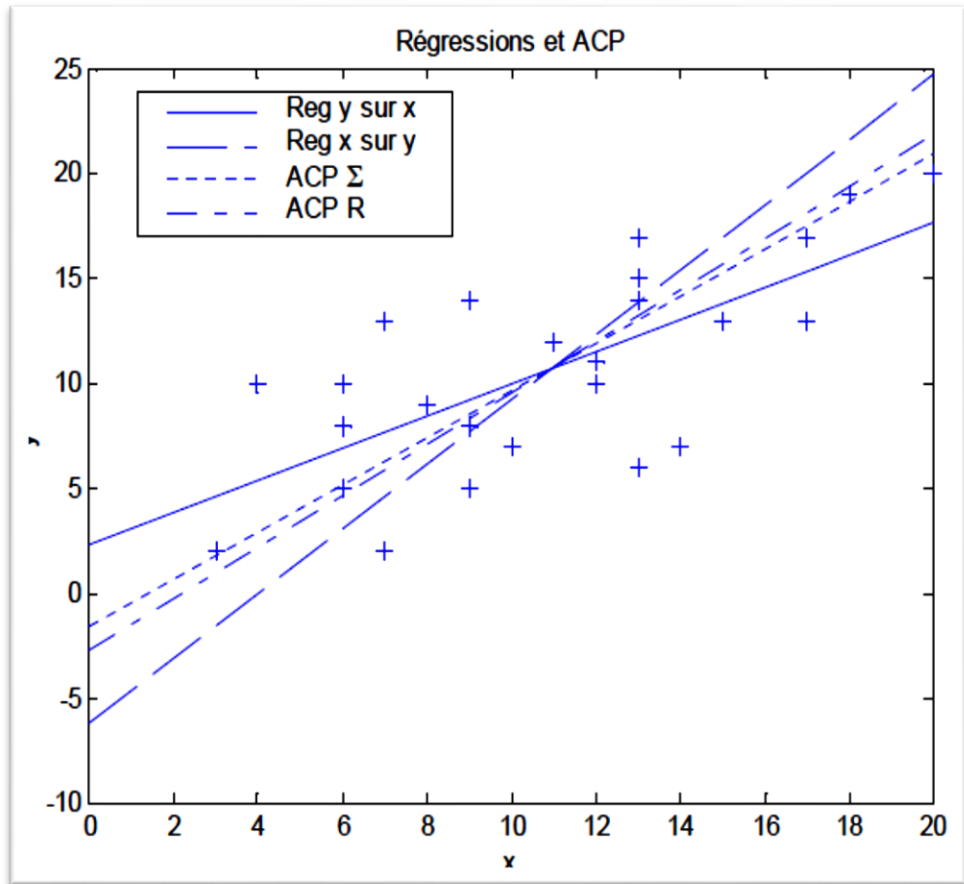


Figure 3.4 Régressions et ACP

### 3.3.7 Exemples d'application

La donnée numérique :

Les données sont des analyses géochimiques d'éléments majeurs (10 éléments) de roches prélevées dans les différentes montrougiennes de la région. On a 68 analyses de roches réparties en 4 groupes principaux de 17 observations et plusieurs sous-groupes. [22]

On remarque que les deux premiers vecteurs propres expliquent à eux seuls 77% de la variance totale. Le premier vecteur propre oppose  $\text{SiO}_2$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{Na}_2\text{O}$  et  $\text{K}_2\text{O}$  à  $\text{FeO}$ ,  $\text{MgO}$ ,

TiO<sub>2</sub>, CaO et P<sub>2</sub>O<sub>5</sub>. Les divers éléments contribuent à peu près également à la définition du vecteur (sauf MnO). Ce premier vecteur oppose donc les éléments féls qu'aux éléments mafiques. A ce titre, on peut l'interpréter comme un axe de maturité magmatique. La disposition des observations le long de cet axe confirme d'ailleurs cette interprétation puisqu'on retrouve les roches mafiques à droite sur le graphe, les felsiques à gauche, et les intermédiaires au centre. [22]

**Tableau 3.4 Les données sont des analyses géochimiques d'éléments majeurs (10 éléments)**

Group	Sous-groupe	Observations
Ultramafiques	Jacupirangite	1.2
	Alnoite	3 à 13
	Mélilité à biotite	14
	Yamaskite	15 à 17
Mafiques	Essexite	18 à 21
	Montréalite	22
	Gabbro	23 à 32
	Diabase	33.34
Roches intermédiaires	Dykes mésocratiques	35 à 51

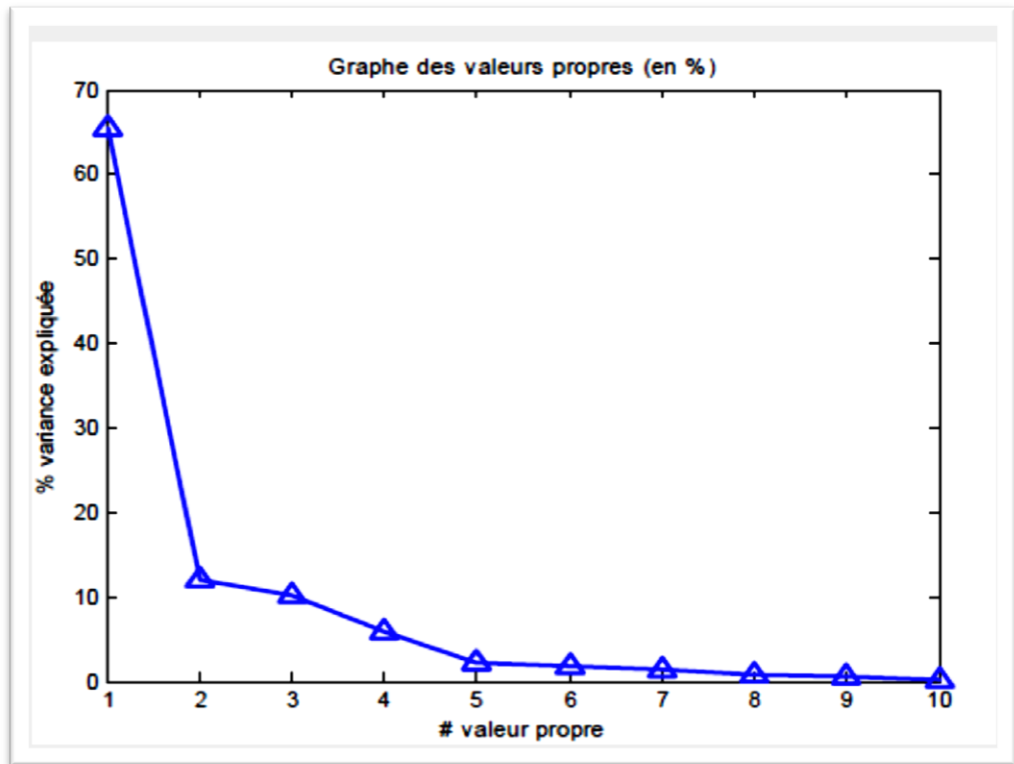


Figure 3.5 Graphe de valeurs propres (en%)

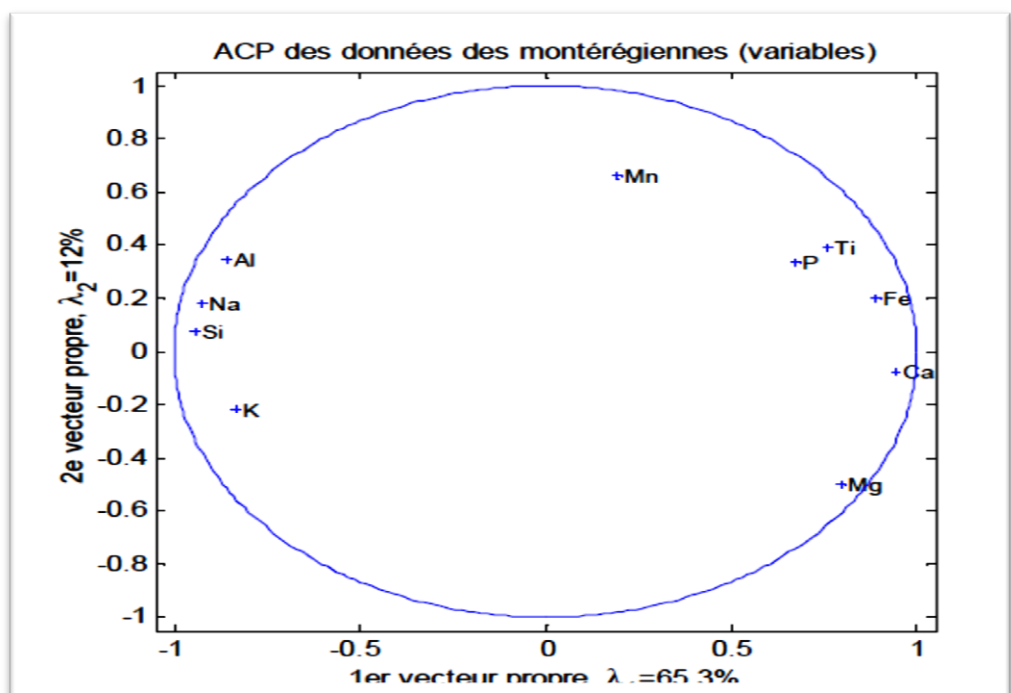


Figure 3.6 ACP des données des montérégiennes (variables)

### **3.4 Conclusion**

Les analyses multivariées sont largement utilisées dans diverses recherches, en particulier pour décrire la corrélation entre deux variables en contrôlant l'effet d'autres variables

L'objectif principal des analyses descriptives multivariées est d'étudier ou de décrire un ensemble de variables prises globalement, cette étude permet de synthétiser et de visualiser rapidement une grande quantité d'information ; Elles reposent sur un examen des interdépendances entre toutes les variables

L'ACP est méthode s'applique aux ensembles de données quantitatives d'au moins deux variables, Puisqu'il s'agit d'une méthode d'analyse de données multifactorielle, son but est de résumer cet ensemble de données

# **Chapitre 04: présentation**

## **dès les régions d'étude**

## 4.1 Introduction

Au cours de ce chapitre, nous nous proposons de présenter les principales données qui concernent le cadre physique de notre région d'étude situation géographique, climat.

La zone étudiée appartient au Bas Sahara. Cet environnement socio-économique se situe dans un contexte climatique de type désertique hyperaride, où Les précipitations sont très faibles et irrégulières ne jouent qu'un rôle limité dans la recharges des nappes. Ce chapitre comportera les données essentielles de la région de Ghardaïa et Hassi R'mel et Touggourt, afin d'assurer et le bon déroulement de cette étude avec les différentes caractéristiques reliées spécialement à cette région.

## 4.2Présentation de la Wilaya de Ghardaïa

### 4.2.1 Situation géographique

La région du M'zab, située à 600 km au sud de la capitale est considérée comme l'une des grandes oasis du Sahara algérien. Administrativement, elle fait partie de la wilaya de Ghardaïa, regroupant trois chefs lieu de communes ; Ghardaïa chef-lieu de la wilaya occupant l'amont et le centre de la vallée, Bounoura et El Atteuf occupant la partie aval de la vallée. [1]

Géographiquement la vallée du M'zab est comprise entre 32° 26' et 32° 31' de latitude Nord et 3° 37' et 3°46' de longitude Est. Elle est alignée approximativement NWSE, en partant de Bouchen, zone de confluence de l'oued El Haimeur (Laadira) avec l'oued Touzouz, jusqu'à El Atteuf sur une longueur de 23,50 km. [1]

La largeur de la vallée varie entre 2,35 et 1,50 km en amont et diminue au fur et à mesure vers l'aval pour atteindre 0,55 km tout au sud de la ville d'El Atteuf. [2]

Les altitudes quant à elles sont relativement faibles, elles diminuent progressivement de l'amont vers l'aval, 514 à 508 mètres à Bouchen, 505 à 485 b mètres au niveau de la ville de Ghardaïa, 455 à 450 mètres au niveau d'El Atteuf, et 435 mètres à Ahbas point le plus bas de l'oasis d'El Atteuf. [2]



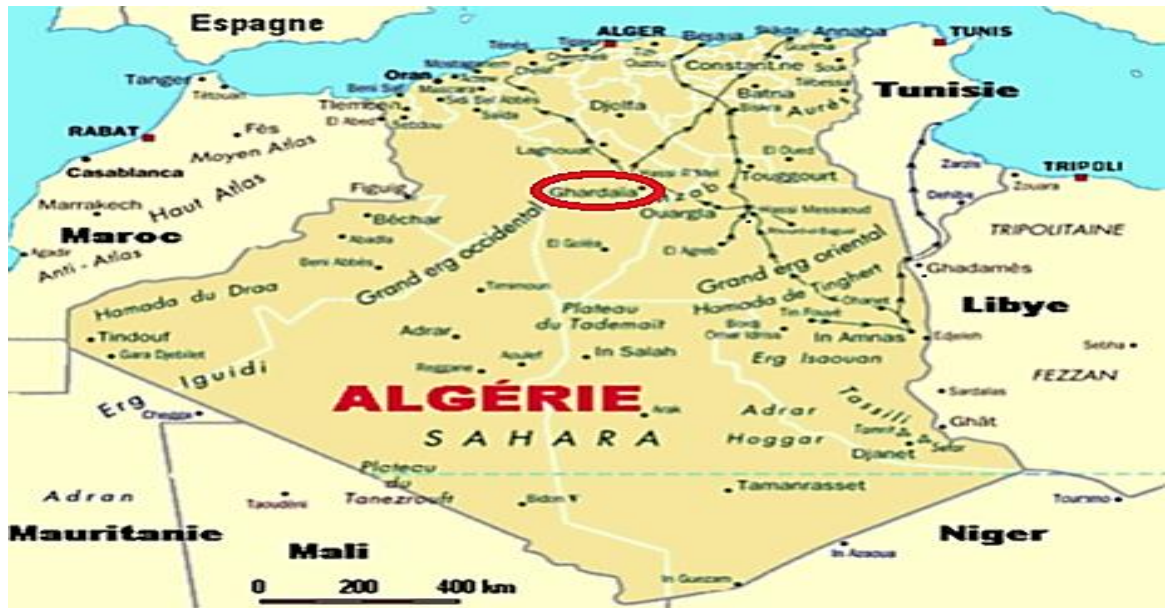


Figure 4.1 Situation géographique de Ghardaïa.

#### 4.2.2 Le climat

La région de Ghardaïa se caractérise, par un climat saharien, qui se distingue par une grande amplitude thermique entre le jour et la nuit, d'Été et d'hiver. La moyenne pluviométrique est de 65.52 mm/an . [1]

##### 4.2.2.1 Températures

Les températures enregistrées durant l'année 2005 sont:

- **Température minimale du mois le plus froid (m) :** Dans la région de Ghardaïa, le mois de janvier est le mois le plus froid, avec une température de 5,6 °C.
- **Température maximale du mois le plus chaud (M) :**  
Le mois d'Août est le plus chaud avec une température de 40,8 °C.
- Température minimale moyenne 16,5 °C.
- Température maximale moyenne 28 °C.

##### 4.2.2.2 Pluviométrie

D'une manière générale, les précipitations sont faibles et d'origine orageuse caractérisées par des écarts annuels et interannuels très important et également par leur intensité  $P = 101,9$  mm. [1]

#### **4.2.2.3 Evaporation**

Durant cette période elle a été évaluée à 2439 mm. [2]

#### **4.2.2.4 Les vents**

Ils sont de deux types :

- Les vents de sables en automne, printemps et hiver de direction nord -ouest.
- Les vents chauds (Sirocco) dominant en été, de direction ouest nord; sont très sec et entraînent une forte évapotranspiration, nécessitent des irrigations importantes.
- La vitesse moyenne mensuelle est de 20 m/s. [2]

### **4.3 Présentation la région de Hassi R'mel**

#### **4.3.1 Situation géographique**

La ville de Hassi R'mel, porte du désert, est située dans la wilaya de Laghouat à 550 Km au sud d'Alger, à environ 70 Km à l'ouest de Berriane et à 120 Km au nord-ouest de Ghardaïa aux points méridiens  $2^{\circ}55'$  et  $3^{\circ}50'$  Est et les parallèles  $33^{\circ}15'$  et  $33^{\circ}45'$  Nord. [19]

Géographiquement, cette région est limitée par le grand Erg occidental au Sud et à l'Ouest, l'Atlas Saharien au Nord et la région d'Ouargla à l'Est. L'altitude moyenne de la zone est de l'ordre de 800 m. Le champ de Hassi R'mel s'étend sur 80 km x 60 km. On y trouve le plus grand gisement de gaz naturel du continent africain. Il produit essentiellement du gaz à condensât (Figure4.2). [3]



Figure 4.2 Situation géographique de la ville de Hassi R'mel.

### 4.3.2 Climat

Le champ gazier de Hassi R'mel est située en zone saharienne dont le climat désertique à une aridité bien exprimée. Il est caractérisé par des précipitations moyennes annuelles très irrégulières caractérisées par des écarts types et des coefficients de variation très élevées. [3]

#### 4.3.2.1 Température

Dans cette région l'hiver est trop froid, et comportant d'importants risques de gelée et pendant l'été le climat est chaud. La région de Hassi R'mel est caractérisée par des températures irrégulier, la température moyenne annuelle est de 19,86 °C, avec 31,67 °C en juillet pour le mois le plus chaud et 8,96 °C en janvier pour le mois le plus froid, avec des extrêmes maximale est égale à 43,90 °C en juillet et minimum est égale -3,1 °C en janvier. [19]

#### 4.3.2.2 Pluviométrie

La pluviométrie est très faible dans cette région, la moyenne annuelle est légèrement supérieure à 82,4 mm pour un nombre de jour de pluie de 33 jours. [3]

### 4.3.2.3 Humidité de l'air

L'humidité relative pour cette région est moyenne, elle est généralement comprise entre 49 et 55% pendant l'été et s'élève à 70% et 85% en l'hiver, Elle varie entre 40% et 85%, sur la période d'observation. La moyenne de l'humidité la plus élevée est enregistrée au mois de décembre, avec 84,4% et le taux le plus faible est au mois de juillet, avec 42,5%, et une moyenne annuelle de 62,41%. [19]

### 4.3.2.4 Vents

Dans la région de Hassi R'mel les vents soufflent du Sud Ouest au Nord Est, les vitesses de vents assez importants sur toute la période estivale (Février – Octobre) avec un des valeurs maximales entre le mois du Septembre et Octobre (19 et 22 m/s). [19]

### 4.3.2.5 Evaporation

L'évaporation dans le région de Hassi R'mel est très élevée, elle est en moyenne de 400,40 mm au mois de juillet et de 70,36 mm au mois de février, soit 219,74 mm moyenne annuelle, elle est très variable suivant les années, les mois et aussi les semaines. [3]

## 4.4 Présentation la région de Touggourt

### 4.4.1 Situation géographique

La vallée de l'oued Righ est une entité économique bien précise qui regroupe près de 50oasis situées au Nord- Est du Sahara du grand erg oriental et au Sud du Massif des Aurès. [7]

Elle s'étend sur un axe Nord-Sud d'environ 150 Km entre les latitudes Nord 32°54' et 34°9' et couvre 15000ha environ des palmeraies; la vallée de l'oued Righ plus souvent nommée simplement oued Righ débute au Nord à Oum El Tiour à plus de 500 Km au Sud-est d'Alger et elle se termine à 150 Km plus au Sud de la palmeraie d'El Goug. [7]

La région de Touggourt qui occupe la moitié sud de la vallée le long d'un axe de 70 km qui débute à la commune de Blid et Amor au sud jusqu'à la commune de Sidi Slimane au



Figure 4. 3 Situation géographique de Touggourt

#### 4.4.2 Climat

En général, la vallée d'Oued Righ est caractérisée par un climat sec et aride, accusant un écart de températures important diurne et nocturne et entre saisons. [15]

##### 4.4.2.1 Température

La température est soumise à des variations mensuelles importantes, le mois d'Août avec 33.42°C est le plus chaud ; alors que le mois de Janvier, avec 11.07°C, est le plus froid de l'année. [15]

##### 4.4.2.3 L'humidité relative de l'air

L'humidité est en fonction des saisons, nous enregistrons un maximum de 65.28 % au mois de Décembre et un minimum de 32.45 % au mois de Juillet. [7]

##### 4.4.2.4 Le vent

Les vents dominants sont surtout ceux du printemps provenant d'Ouest au Nord-Ouest. Les vents les plus forts soufflent en fin d'hiver, début printemps. Les vents de sable arrivent pendant, le mois de Juin et parfois en Juillet. [7]

## **4.5 Conclusion**

Les données ci-dessus donnent un aperçu du domaine couvert par cette étude. Dans ce chapitre, nous définissons les différentes caractéristiques de Ghardaïa Hassi R'mel Touggourt.

La zone en question est située au nord-est du Sahara et couvre trois régions (Ghardaïa, Hassi R'mel et Touggourt), dominées par le climat aride, illustrées par les données climatiques de la station météorologique de chaque région.

**Chapitre 05 : Statistique  
d'ordre applique sur les 03  
stations**



## **5.1 Introduction**

Pour l'étude statistique des précipitations en hydrologie, on dispose d'une série pluviométrique qui comporte les précipitations maximales journalières pour la période la plus longue possible.

Nous prenons comme base de calcul la série pluviométrique de la station expérimentale de Ghardaïa et Hassi R'mel et Touggourt. .

La série a une période de fonctionnement de 1975 à 2014 qui a été fournie par l'Office Nationale de la Météorologie (ONM).

L'analyse statistique des données pluviométriques consiste à déterminer les caractéristiques empiriques d'un échantillon d'une série d'observations de précipitations maximales journalières, de 40 années.

## **5.2 Analyse de la précipitation**

Selon Dubief (1953), les précipitations ont pratiquement toujours lieu sous forme des pluies. Ces dernières sont caractérisées par leur faible importance quantitative et les pluies torrentielles sont rares. Elles sont liées aux perturbations soudano-sahariennes ou sahariennes.

Cette insuffisance de pluies sahariennes est accompagnée d'une irrégularité très marquée du régime pluviométrique et d'une variabilité interannuelle considérable, ce qui accentue la sécheresse. C'est le cas de Ghardaïa et Hassi R'mel et Touggourt.

## **5.3 Station de Ghardaïa**

### **5.3.1 Précipitations maximales journalières**

Les précipitations maximales journalières permettent de calculer les intensités des pluies les plus fréquentes. La série d'observation enregistrée entre 1975 jusqu'à 2014 nous a permis de déduire la pluviométrie maximale journalière qui est de 54.8 mm.



**Remarque :**

Pour l'étude statistique des pluies maximales journalières, les ajustements effectués, montrent que ces pluies s'ajustent plutôt sur les lois dissymétriques, log-normal (Galton) et Gumbel

**5.3.2 Analyse des données statistiques**

- La moyenne interannuelle des précipitations maximales journalières  $P_j$  max, Durant 40 ans d'observations est

$$\bar{x} = 19,18 \text{ mm}$$

$n$  : le nombre d'années d'observations ( $n= 40$  ans).

- L'écart type «  $\sigma$  Pmax, j » :  $\delta_x = 12,96$  mm
- Coefficient de variation : «  $C_v$  » :  $C_v = 0,68$
- Moyenne  $\log(P)$  : 2.74
- L'écart type  $\log(P)$  : 0.69

**Tableau 5.1 Pluies journalières maximales observées à Ghardaïa (1975-2014)**

Anne	Pluie (mm)	Anne	Pluie (mm)
1975	24.7	1995	30
1976	17.2	1996	12
1977	17.9	1997	23
1978	11.3	1998	3.5
1979	18.6	1999	26
1980	46.5	2000	18
1981	13	2001	18
1982	9.4	2002	24
1983	2.9	2003	33
1984	20.2	2004	46
1985	6.8	2005	7
1986	36.6	2006	16.2
1987	8.8	2007	10
1988	8.9	2008	23
1989	6.7	2009	19.5
1990	27.9	2010	8.6
1991	10.9	2011	50
1992	12.6	2012	7.11
1993	11.9	2013	10
1994	54.8	2014	14.99

### 5.3.3 Choix de la loi d'ajustement

#### 5.3.3.1 Ajustement de la série pluviométrique à la loi de Galton (Log Normal)

Le tableau (01) dans l'annexe indique tous les valeurs d'ajustement

On tracer le grave sur l'Excel :

a) l'axe des X : les variables réduites

b) l'axe des Y : les valeurs théoriques

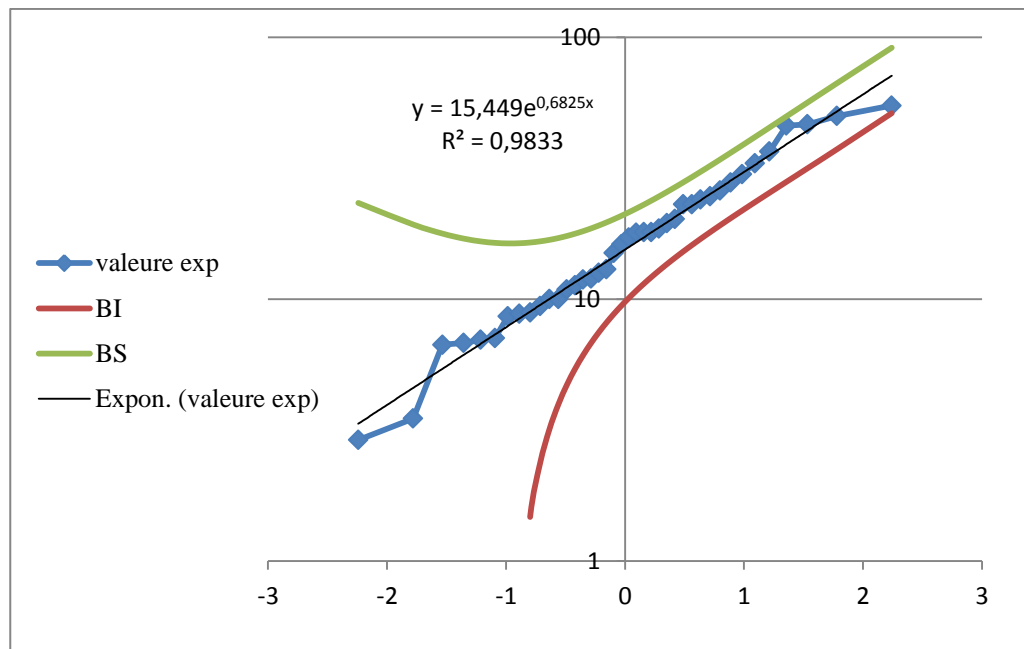


Figure 5.1 Ajustement graphique d'une loi de Galton (station Ghardaia)

#### 5.3.3.2 Ajustement de la série pluviométrique à la loi de Gumbel

Le tableau (02) dans l'annexe indique tous les valeurs d'ajustement

On tracer le grave sur l'Excel :

a) l'axe des X : les variables réduites

b) l'axe des Y : les valeurs théoriques

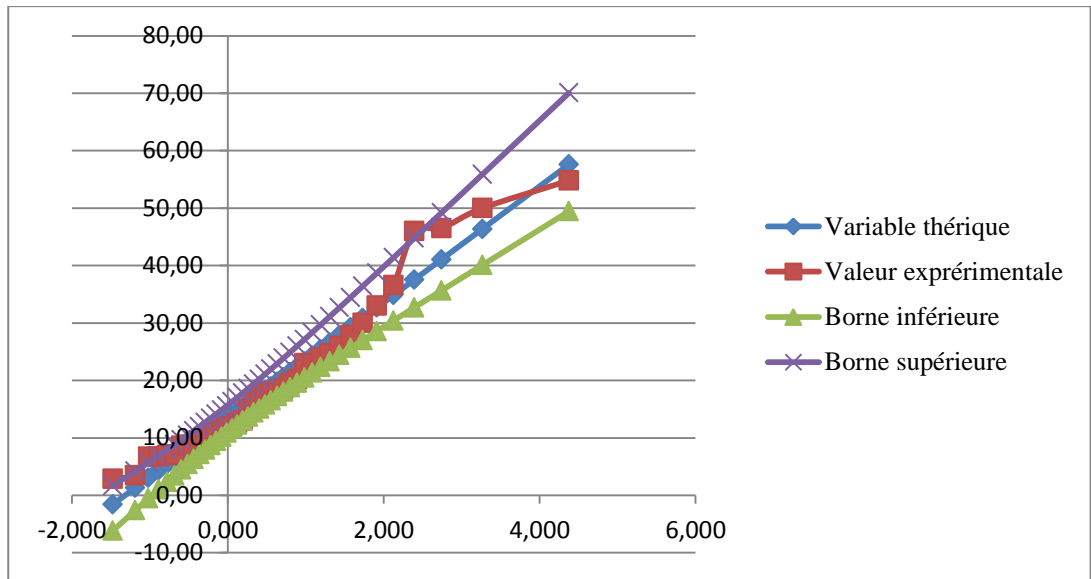


Figure 5. 2 Ajustement graphique d'une loi de Gumble (station Ghardaia)

### 5.3.4 Interprétation des graphes

La projection des données expérimentales sur les deux graphes a permis de voir que la loi de Galton est mieux ajustée par rapport à la loi de Gumbel vu que ces données sont plus proches de la droite théorique.

### 5.3.5 Test de $\text{Khi}^2$

#### a) Loi de Galton (log normal)

Tableau 5. 2 Test de  $\text{Khi}^2$  appliqué sur la série des pluies (Galton)

$\text{Khi}^2$ (Valeur calculé)	2.66
$\text{Khi}^2$ (Valeur théorique)	5.99
DDL	2
Alpha	0,05

**b) Loi de Gumbel****Tableau 5. 3 Test de Khi 2 appliqué sur la série des pluies (Gumbel)**

Khi <sup>2</sup> (Valeur calcule)	3.29
Khi <sup>2</sup> (Valeur théorique)	7.83
DDL	3
Alpha	0,05

**5.3.5.1 Interprétation des tableaux**

- D'après le test de Khi<sup>2</sup> applique sur la série des pluies Dans la loi de Galton (log normal)

, En remarque que :

Valeur théorique > valeur calcule

- D'après le test de Khi<sup>2</sup> applique sur la série des pluies Dans la loi de Gumbel, En remarque que :

Valeur théorique > valeur calcule

**5.4 Station de Hassi R'mel****5.4.1 Précipitations maximales journalières**

Les précipitations maximales journalières permettent de calculer les intensités des pluies les plus fréquentes. La série d'observation enregistrée entre 1975 jusqu'à 2014 nous a permis de déduire la pluviométrie maximale journalière qui est de 88 mm.

**5.4.2 Analyse des données statistiques**

- La moyenne interannuelle des précipitations maximales journalières P<sub>j</sub> max, Durant 40 ans d'observations est

$$\bar{x} = 45 \text{ mm}$$

n : le nombre d'années d'observations (n= 40 ans).

- L'écart type «  $\bar{\sigma}$  P<sub>max, j</sub> » :  $\bar{\sigma}=25$

- Coefficient de variation : «  $Cv$  » :  $cv = 0.56$
- Moyenne  $\log(P)$  : 3.62
- L'écart type  $\log(P)$  : 0.64

**Tableau 5. 4 Pluies journalières maximales observées à Hassi R'mel (1975-2014)**

Anne	Pluie (mm)	Anne	Pluie (mm)
1975	48	1995	32
1976	56	1996	88
1977	33	1997	76
1978	15	1998	41
1979	28	1999	28
1980	20	2000	29
1981	75	2001	27
1982	80	2002	87
1983	5	2003	79
1984	83	2004	80
1985	24	2005	50
1986	49	2006	50
1987	27	2007	26
1988	29	2008	58
1989	17	2009	65
1990	61	2010	21
1991	27	2011	77
1992	31	2012	20
1993	17	2013	31
1994	82	2014	18

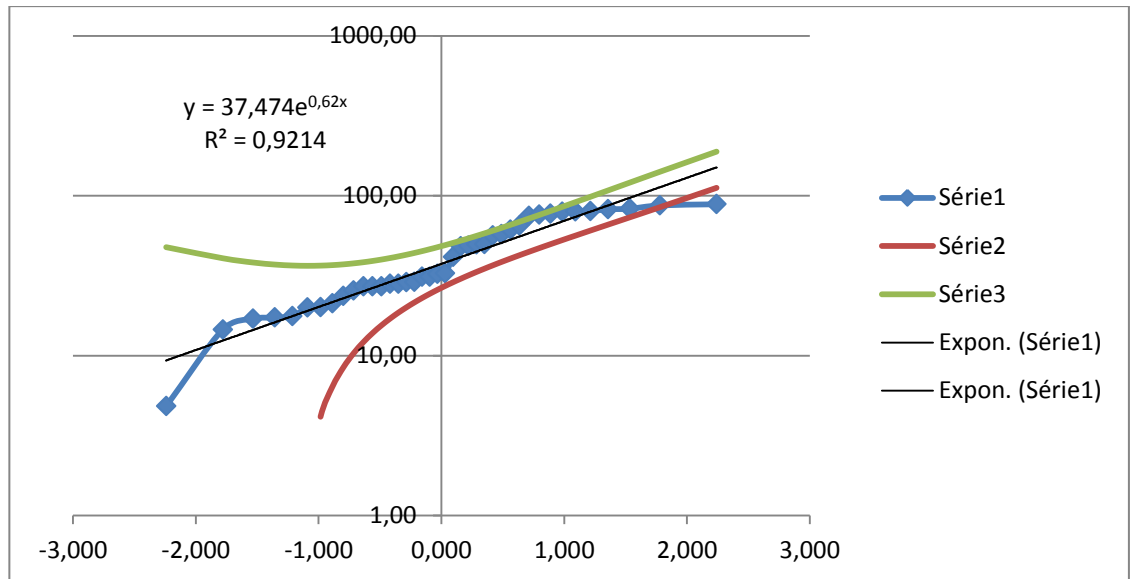
### 5.4.3 Choix de la loi d'ajustement

#### 5.4.3.1 Ajustement de la série pluviométrique à la loi de Galton (Log Normal)

Le tableau (03) dans l'annexe indique tous les valeurs d'ajustement

On tracer le grave sur l'Excel :

- l'axe des X : les variables réduites
- l'axe des Y : les valeurs théoriques



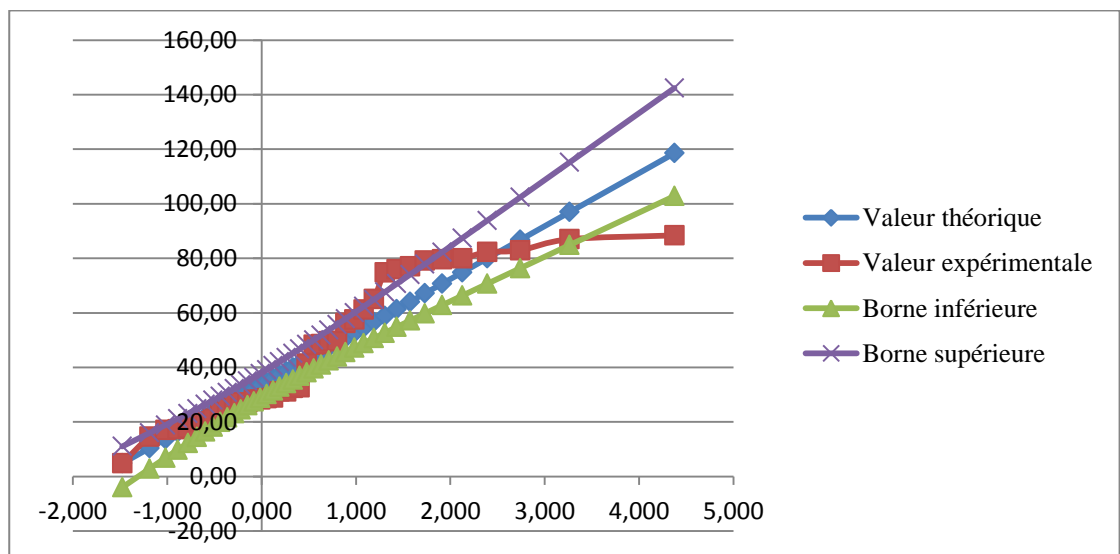
**Figure 5. 3 Ajustement graphique d'une loi de Galton (station Hassi R'mel)**

#### 5.4.3.2 Ajustement de la série pluviométrique à la loi Gumbel

Le tableau (04) dans l'annexe indique tous les valeurs d'ajustement

On tracer le grave sur l'Excel :

- l'axe des X : les variables réduites
- l'axe des Y : les valeurs théoriques



**Figure 5. 4 Ajustement graphique d'une loi de Gumble (station Hassi R'mel)**

#### 5.4.4 Interprétation de graphe

La projection des données expérimentales sur les deux graphes a permis de voir que la loi de Galton est mieux ajustée par rapport à la loi de Gumbel vu que ces données sont plus proches de la droite théorique.

#### 5.4.5 Teste khi2

##### a) Loi de Galton (log-normal)

**Tableau 5. 5 Test de Khi 2 appliqué sur la série des pluies (Galton)**

Khi <sup>2</sup> (Valeur calculé)	5.826
Khi <sup>2</sup> (Valeur théorique)	5.991
DDL	2
Alpha	0.05

##### b) Loi de Gumbel

**Tableau 5. 6 Test de Khi 2 appliqué sur la série des pluies (Gumbel)**

Khi <sup>2</sup> (Valeur calculé)	10.76
Khi <sup>2</sup> (Valeur théorique)	7.82
DDL	3
Alpha	0,05

##### 5.4.5.1 Interprétation des tableaux

- D'après le test de Khi<sup>2</sup> applique sur la série des pluies Dans la loi de Galton (log normal), En remarque que :  
Valeur théorique > valeur calculé.
- D'après le test de Khi<sup>2</sup> applique sur la série des pluies Dans la loi de Gumbel, En remarque que :  
Valeur théorique < valeur calculé.

## 5.5 Station de Touggourt

### 5.5.1 Précipitations maximales journalières

Les précipitations maximales journalières permettent de calculer les intensités des pluies les plus fréquentes. La série d'observation enregistrée entre 1975 jusqu'à 2014 nous a permis de déduire la pluviométrie maximale journalière qui est de 28.10 mm.

### 5.5.2 Analyse des données statistiques

- La moyenne interannuelle des précipitations maximales journalières  $P_j$  max, Durant 40 ans d'observations est  $\bar{x} = 11.49$  mm  
n : le nombre d'années d'observations (n= 40 ans).
- L'écart type «  $\bar{\sigma} P_{max, j}$  » :  $\bar{\sigma} = 5.78$
- Coefficient de variation : « Cv » :  $cv = 0.50$
- Moyenne  $\log(P)$  : 2.30
- L'écart type  $\log(P)$  : 0.54

**Tableau 5.7 Pluies journalières maximales observées à Touggourt (1975-2014)**

Anne	Pluie (mm)	Anne	Pluie (mm)
1975	15.10	1995	5.41
1976	14.10	1996	14.73
1977	20.10	1997	16.64
1978	7.40	1998	6.90
1979	13	1999	4.68
1980	23.30	2000	4.83
1981	11.70	2001	4.51
1982	14.60	2002	22.59
1983	15	2003	13.19
1984	12.30	2004	13.33
1985	6.30	2005	8.28
1986	11.80	2006	8.26
1987	15.90	2007	4.25
1988	11.80	2008	9.59
1989	11.40	2009	10.84
1990	12.40	2010	3.53
1991	15.80	2011	12.83
1992	28.10	2012	3.32
1993	11.80	2013	5.18
1994	11.90	2014	2.94



### 5.5.3 Choix de la loi d'ajustement

#### 5.5.3.1 Ajustement de la série pluviométrique à la loi de Galton (Log Normal)

Le tableau (05) dans l'annexe indique tous les valeurs d'ajustement

On tracer le grave sur l'Excel :

a) l'axe des X : les variables réduites

b) l'axe des Y : les valeurs théoriques

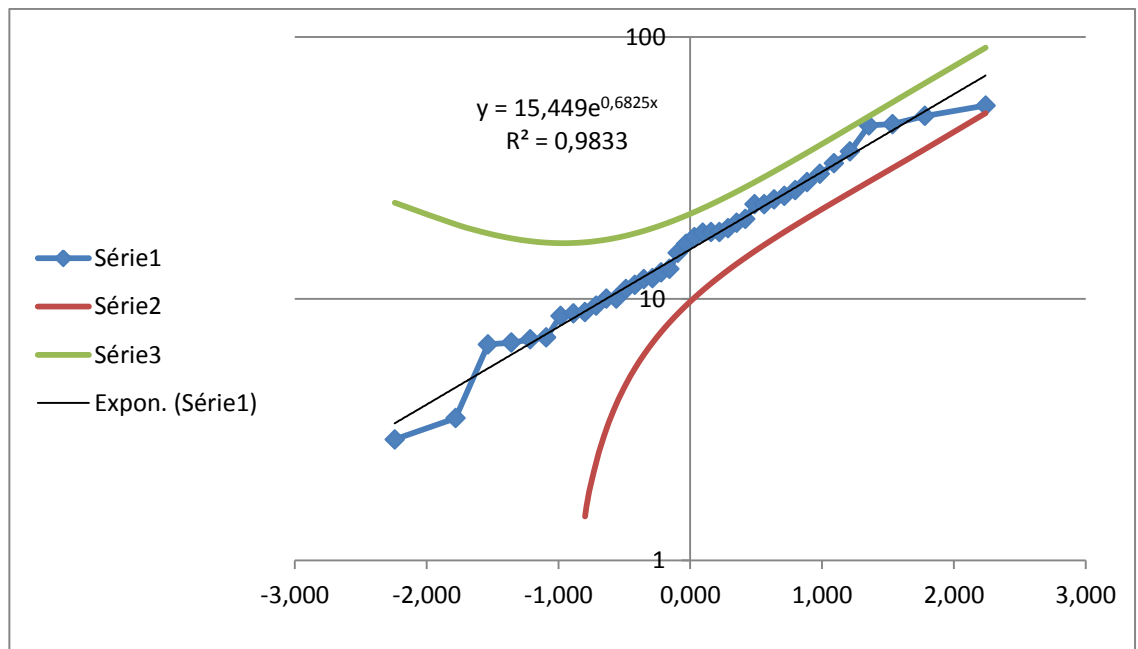


Figure 5. 5 Ajustement graphique d'une loi de Galton (station Touggourt)

#### 5.5.3.2 Ajustement de la série pluviométrique à la loi Gumbel

Le tableau (06) dans l'annexe indique tous les valeurs d'ajustement

On tracer le grave sur l'Excel :

a) l'axe des X : les variables réduites

b) l'axe des Y : les valeurs théoriques

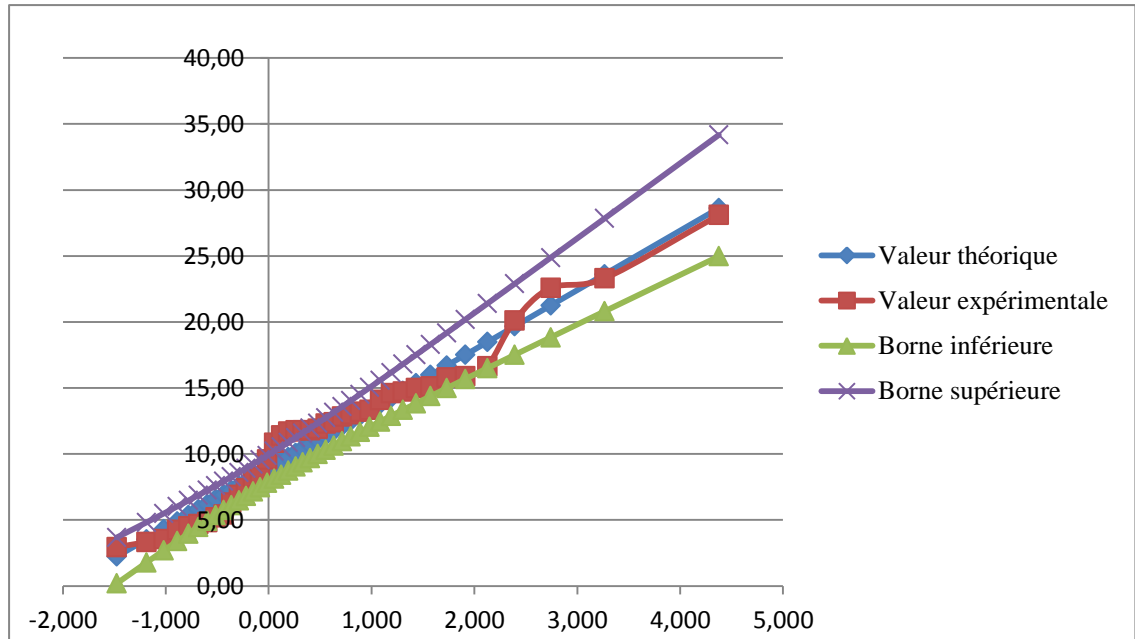


Figure 5. 6 Ajustement graphique d'une loi de Gumble (station Touggourt)

#### 5.5.4 Interprétation de graphe

La projection des données expérimentales sur les deux graphes a permis de voir que la loi de Galton est mieux ajustée par rapport à la loi de Gumbel vu que ces données sont plus proches de la droite théorique.

#### 5.5.5 Teste khi2

##### a) Loi de Galton (log-normal)

Tableau 5. 8 Test de Khi 2 appliqué sur la série des pluies (Galton)

Khi <sup>2</sup> (Valeur calculé)	12.07
Khi <sup>2</sup> (Valeur théorique)	5.991
DDL	2
Alpha	0,05

**a) Loi de Gumbel****Tableau 5. 9 Test de Khi 2 appliqué sur la série des pluies (Gumbel)**

Khi <sup>2</sup> (Valeur calculé)	21.16
Khi <sup>2</sup> (Valeur théorique)	7.82
DDL	3
Alpha	0,05

**5.5.5.1 Interprétation des tableaux**

- D'après le test de Khi<sup>2</sup> applique sur la série des pluies Dans la loi de Galton (log normal), En remarque que :

Valeur théorique < valeur calculé.

- D'après le test de Khi<sup>2</sup> applique sur la série des pluies Dans la loi dans loi de Gumbel , En remarque que :

Valeur théorique < valeur calculé.

**5.6 Conclusion**

Pour les stations étudiées, on remarque, qu'elles s'ajustent correctement avec la loi de GALTON, cela explique l'alignement des points expérimentaux.

Il apparaît, que très longues périodes d'observation sont nécessaires pour connaître la pluviosité avec une approximation suffisante.

D'après le test de Khi<sup>2</sup> applique sur la série des Pluies journalières maximales observées à Ghardaïa (1975-2014) dans la loi de Galton (log normal) et dans la loi de Gumbel, en remarque que :

Valeur théorique > valeur calculé. Ce qui montre que l'ajustement est accepté avec la loi de Galton et accepté avec la loi de Gumbel

D'après le test de Khi<sup>2</sup> applique sur la série des Pluies journalières maximales observées à Hassi R'mel (1975-2014) dans la loi de Galton (log normal), en remarque que :

Valeur théorique > valeur calculé. Ce qui montre que l'ajustement est accepté avec la loi de Galton.

D'après le test de Khi<sup>2</sup> applique sur la série des Pluies journalières maximales observées à Hassi R'mel (1975-2014) dans la loi de Gumbel, en remarque que :

Valeur théorique < calculé. Ce qui montre que l'ajustement est rejeté avec la loi de Gumbel.

D'après le test de  $\text{Khi}^2$  applique sur la série des Pluies journalières maximales observées à Touggourt (1975-2014) dans la loi de Galton (log normal) et dans la loi de Gumbel, en remarque que :

Valeur théorique < valeur calculé. Ce qui montre que l'ajustement est rejeté avec la loi de Galton et rejeté avec la loi de Gumbel.

# **Chapitre 06 : Analyse composantes principales**

## 6.1 Introduction

L'analyse en composantes principales (ACP), permet d'analyser et de visualiser un jeu de données contenant des individus décrits par plusieurs variables quantitatives.

C'est une méthode statistique qui permet d'explorer des données dites multivariées (données avec plusieurs variables).

Dans ce chapitre, nous décrivons les bases de l'ACP et démontrons comment calculer, visualiser et interpréter l'ACP. De plus, nous montrerons comment révéler les variables les plus importantes qui expliquent les variations dans un jeu de données.

## 6.2 Généralités

La méthode d'analyses multidimensionnelles traite des tableaux des données pour lesquels les colonnes correspondent aux variables et les lignes aux unités statistiques (individus).

Chaque variable peut être représentée par un vecteur à  $n$  dimensions (d'où le nom analyse multidimensionnelle) de même, chaque individu peut être représenté par un vecteur qui sera à  $n$  dimensions. Ainsi, le tableau de données est transformé à une représentation graphique du type vectoriel.

Toutefois, comme il s'agit d'un espace à plusieurs dimensions, cette représentation est difficilement interprétable.

L'idée consiste à ramener les vecteurs qui représentent les données dans un espace à une, ou deux dimensions, seule possibilité pour avoir une représentation simple et accessible.

Le but de toute l'analyse multidimensionnelle consiste à extraire l'information qui est contenu dans le tableau des données selon les principaux axes factoriel.

## 6.3 Modélé de L'A.C.P

Avec le modèle d'analyse en composant principale (ACP) les lignes du tableau de données correspondent aux variables : la métrique utilisée est la métrique usuelle (matrice unité).

### 6.3.1 Détermination analytique

On part de la matrice des variances - covariances

$$V = j \begin{bmatrix} \text{varx1} & & \\ & \text{cov}(x_j, x_i) & \\ & & \text{varxp} \end{bmatrix}$$

Le métrique utilisée étant la métrique usuelle, elle correspond à :

**Tableau 6.1 Les données formant la matrice initial Pluie( mm)**

Ghardaïa	Hassi R'mel	Touggourt
24.7	48.25	15.10
17.2	56.30	14.10
17.9	32.70	20.10
11.3	14.55	7.40
18.6	28.15	13.00
46.5	20.05	23.30
13	74.80	11.70
9.4	79.64	14.60
2.9	4.83	15.00
20.2	82.94	12.30
6.8	23.63	6.30
36.6	48.55	11.80
8.8	27.10	15.90
8.9	28.70	11.80
6.7	17.00	11.40
27.9	61.15	12.40
10.9	27.10	15.80
12.6	31.15	28.10
11.9	17.30	11.80
54.8	82.25	11.90
30	32.45	5.41
12	88.40	14.73
23	76.00	16.64
3.5	41.41	6.90
26	28.07	4.68
18	28.96	4.83
18	27.05	4.51
24	87.00	22.59
33	79.13	13.19
46	80.01	13.33
7	49.67	8.28
16.2	49.53	8.26
10	25.53	4.25
23	57.54	9.59
19.5	65.03	10.84

8.6	21.21	3.53
50	76.97	12.83
7.11	19.95	3.32
10	31.11	5.18
14.99	17.65	2.94

**Tableau 6.2 Statistiques descriptives**

Variable	Observations	Obs. avec données manquantes	Obs. sans données manquantes	Minimum	Maximum
Ghardaïa	40	0	40	2.900	54.800
Hassi R'mel	40	0	40	4.830	88.400
Touggourt	40	0	40	2.941	28.100

**Tableau 6.3 Moyenne et l'écart types des variables**

Variable	Moyenne (mm)	Ecart-type
Ghardaïa	19.18	12.96
Hassi R'mel	44.71	24.91
Touggourt	11.49	5.78

$$M = \begin{bmatrix} 1 & 0 \\ & 1 \\ 0 & 1 \end{bmatrix} \text{ qui est une matrice unité.}$$

Pour calculer V on peut employer deux procédés. En effet on a :

$$V = X' P X,$$

Ou X' est la matrice des variable x centrées, X est la transposée de cette matrice, et P est une matrice diagonale qui correspond au poids.

Donc on peut détermine V à partir de calcule au poids.

Le problème va consister à obtenir successivement les valeurs propres de la matrice VM et leurs vecteurs propres associées ; VM est identique à V, puisque M un matrice unité.

Pour chaque variété on peut ainsi définir le pourcentage de variance exprimée :

$$\% \text{ variance exprimée} = \frac{Y_k}{Y_1 + Y_2 + \dots + Y_p}$$



On somme  $Y_1 + Y_2 + \dots + Y_p$  par définition n'est autre que la trace de la matrice V (somme des éléments diagonaux).

Le vecteur propre  $t_1 (t_1^1, t_1^2, \dots, t_1^p)$  est appelé vecteur axial factoriel.

Dans le calcul il ne faut pas oublier de norme le vecteur propre. Ces vecteurs permettent de reconstituer la variable à partir des facteurs centrés.

La corrélation entre variable et facteur, compte tenu de ce que a été dit plus haut, concernant la visualisation, est donnée par :

$$\rho(x^j, F_k) = \frac{a_k^j}{\sqrt{\text{var}x^j}} \dots \dots \dots (1)$$

De la même façon, on peut chercher à exprimer la corrélation multiple entre la variable et les variétés factorielles. Avec  $F^1$ , cela correspond au coefficient de corrélation défini ci-dessus, on calcule ensuite avec  $(F^1, F^2) \dots$  en fait, on préfère pour le calcul, prendre le carré des corrélations multiples variables-variétés factorielles

$$R^2(x^j, \{F^1, \dots, F^k\}) = \frac{\sum_{i=1}^{i=p} (a_i^j)^2}{\text{var}x^j} \dots \dots \dots (2)$$

Pour ce qui des u. s, il est nécessaire, dans un premier temps, de déterminer la contribution des unités statistique à l'inertie de chaque u. s. et l'inertie du nuage, les inerties étant calculées par le rapport au centre de gravité. Soit :

$$\frac{\sum_{j=1}^{j=p} \theta_i (u_i^j - X^j)^2}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \dots \dots \dots (3)$$

Si on donne le même poids à chaque u. s. l'expression devient :

$$\frac{\frac{1}{n} \sum_{j=1}^{j=p} (u_i^j - X^j)^2}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \dots \dots \dots (4)$$

Lorsque l'on va projeter les nuages sur les hyperplans solution de l'espace des u. s., chaque u.s. aura par rapport aux axes factoriels qui auront été définies une nouvelle coordonnée appelée composant principale.

Lorsque l'on connaît les composantes principales, il est possible de considérer la contribution d'u.s. à la détermination des facteurs. Pour cela, on compare l'inertie qu'exprime la projection de l'u,s par rapport à l'u s par rapport à inertie exprimée par le facteur. Ainsi, on a :

$$\text{Contribution de la } j^{\text{ième}} \text{ u.s. à } \mathbf{f}^k = \frac{\theta_i}{k} (\mathbf{c}_i)^2 \dots \dots \dots (5)$$

De même que l'on établit les corrélations multiples est variable et variétés factorielles, de même on définit les cosinus entre .u. s et variétés factorielles.

Comme pour les variables, on préfère prendre le carré des coefficients de corrélations multiples que l'on nomme carré des M cosinus des (u.s.) avec les variétés factorielles

#### 6.4 Application aux pluies maximales journalières

L'interprétation des résultats va se faire à partir de la représentation graphique obtenue avec les saturations dans l'espace des variables et les composantes principale dans l'espace des unités statistiques, et à partir également des différents indices fournis par l'analyse ; ces indices sont à la fois une aide à l'interprétation et un contrôle dans l'extraction de l'information contenue dans les tableaux traités.

Compte tenu des variables, on réalisera une A.C.P centrée sur les pluies moyennes journalières de toutes nos stations

**Tableau 6.4 Valeurs propres et pourcentage de variance exprimé**

	F1	F2	F3
Valeur propre	1.676	0.797	0.527
Variabilité (%)	55.876	26.562	17.561
% cumulé	55.876	82.439	100.000

**Tableau 6.5 Matrice des coefficients de corrélation**

Variables	Ghardaïa	Hassi R'mel	Touggourt
Ghardaïa	<b>1</b>	<b>0.468</b>	0.236
Hassi R'mel	<b>0.468</b>	<b>1</b>	0.294
Touggourt	0.236	0.294	<b>1</b>

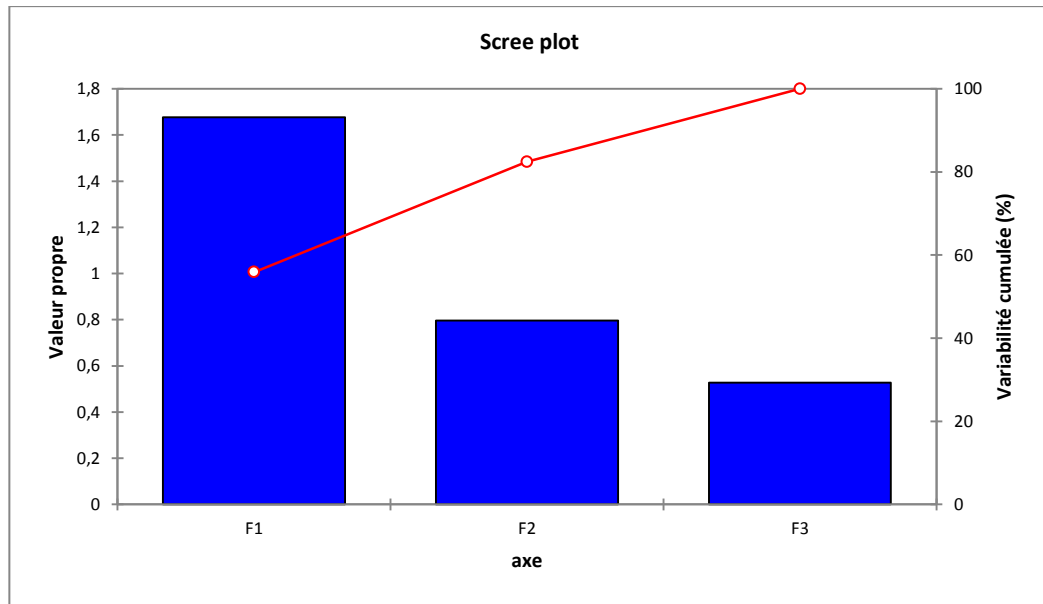


Figure 6.1 Contributions des variables

Tableau 6.6 Corrélations entre les variables et les composantes principales

Variable	Composantes principales					
	Axe 1		Axe 2		Axe 3	
Ghardaïa	0.784	0.615	-0.389	0.151	-0.483	0.233
Hassi R'mel	0.816	0.666	-0.221	0.049	0.534	0.285
Touggourt	0.629	0.395	0.772	0.597	-0.090	0.008

Ou :

- Colonne 1 : Corrélations entre les variables et les axes principaux
- Colonne 2 : Cosinus carrés des variables

#### 6.4.1 Interprétation dans l'espace des variables

Le tableau 6.4 des valeurs propres que expriment la variance de chacun des facteurs: le facteur 1 exprime 55.88 de la variance (fig.6.1), il donc plus important par rapport aux autres, puisque le plan factoriel (1-2) = 82.44 % de la variance, la représentation des données dans ce plan rend compte de façon satisfaisante de leur structure.

Les stations (Ghardaïa, Hassi R'mel) témoignée d'une moyenne corrélation, ce que donne un coefficient de corrélation =0.47 (fig.6.2).

Par contre, la station de Touggourt trouve très faible corrélation avec les autres stations, ce que donne un coefficient de corrélation = 0.29 par la corrélation (Hassi R'mel,

Touggourt)(fig.6.2), et donne un coefficient de corrélation = 0.24 par la corrélation (Ghardaia, Touggourt) (fig.6.2).

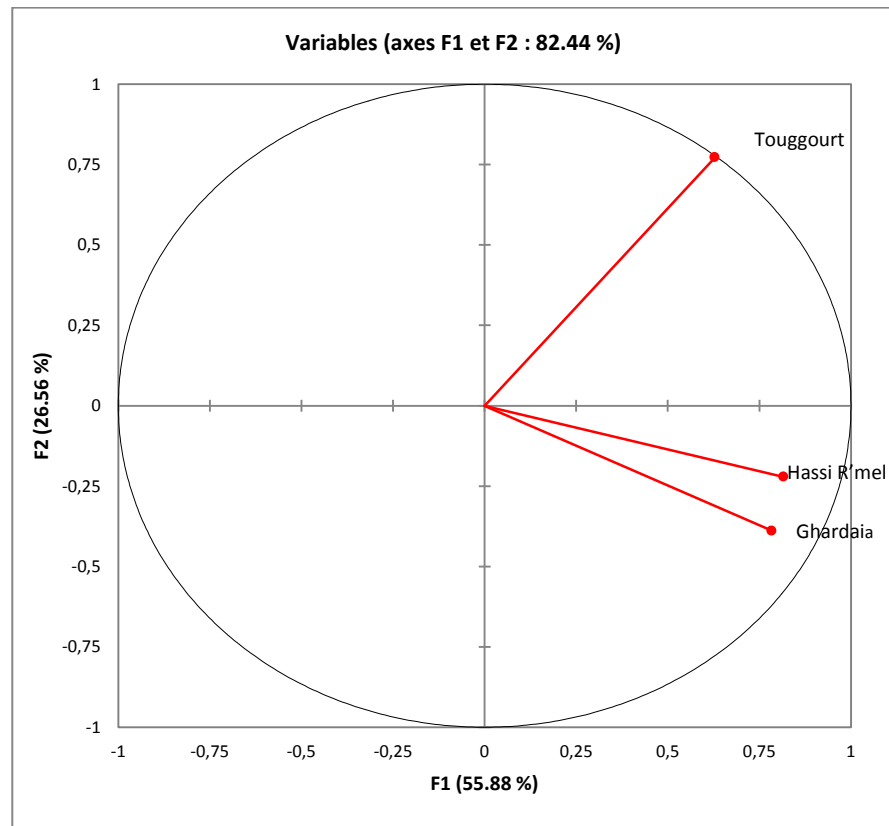


Figure 6.2 Projection de les stations sur le plan factoriel

**Tableau 6.7** Corrélation entre les individus et les composantes principales

Individus	Composantes principales					
	F1		F2		F3	
1975	0.658	0.715	0.324	0.173	-0.260	0.111
1976	0.425	0.397	0.347	0.264	0.393	0.339
1977	0.363	0.052	1.470	0.856	-0.480	0.091
1978	-1.494	0.931	-0.048	0.001	-0.403	0.068
1979	-0.324	0.200	0.416	0.329	-0.498	0.472
1980	1.665	0.282	1.110	0.125	-2.415	0.593
1981	0.496	0.142	-0.061	0.002	1.217	0.856
1982	0.696	0.167	0.453	0.071	1.486	0.762
1983	-1.494	0.483	1.488	0.479	-0.423	0.039
1984	1.096	0.492	-0.297	0.036	1.073	0.472
1985	-1.568	0.984	-0.153	0.009	0.126	0.006
1986	0.948	0.479	-0.584	0.182	-0.797	0.339
1987	-0.568	0.182	1.199	0.814	-0.083	0.004
1988	-0.871	0.707	0.558	0.291	0.049	0.002
1989	-1.309	0.771	0.690	0.215	-0.178	0.014
1990	0.911	0.887	-0.324	0.112	0.019	0.000
1991	-0.477	0.151	1.113	0.824	-0.190	0.024
1992	0.753	0.063	2.879	0.917	-0.426	0.020
1993	-1.021	0.664	0.571	0.208	-0.448	0.128
1994	2.682	0.714	-1.528	0.232	-0.738	0.054
1995	-0.320	0.049	-1.166	0.649	-0.797	0.302
1996	1.055	0.294	0.296	0.023	1.610	0.683
1997	1.420	0.800	0.335	0.045	0.625	0.155
1998	-1.218	0.685	-0.129	0.008	0.817	0.308
1999	-0.684	0.216	-1.097	0.555	-0.704	0.229
2000	-1.027	0.592	-0.811	0.369	-0.264	0.039
2001	-1.103	0.602	-0.840	0.349	-0.315	0.049
2002	2.255	0.740	1.092	0.174	0.773	0.087
2003	1.680	0.879	-0.559	0.098	0.274	0.023
2004	2.330	0.829	-0.988	0.149	-0.379	0.022
2005	-0.723	0.414	-0.122	0.012	0.852	0.574
2006	-0.293	0.208	-0.437	0.462	0.370	0.330
2007	-1.542	0.871	-0.591	0.128	0.061	0.001
2008	0.347	0.256	-0.547	0.635	0.227	0.109
2009	0.480	0.331	-0.314	0.142	0.606	0.527
2010	-1.780	0.895	-0.609	0.105	0.021	0.000
2011	2.398	0.760	-1.171	0.181	-0.666	0.059
2012	-1.901	0.915	-0.578	0.084	0.065	0.001
2013	-1.320	0.853	-0.506	0.126	0.208	0.021
2014	-1.620	0.736	-0.881	0.218	-0.405	0.046

- Colonne 1 : Coordonnées des individus sur les axes principaux
- Colonne 2 : Cosinus carrés (qualité de représentation)

### 6.4.2 Interprétation dans l'espace des individus

La visualisation dans l'espace des individus, permet de situer les différentes années mais une par rapport aux autres et leurs corrélations avec les composantes principales.

La présence d'une bonne corrélation au cours d'une période indique que les deux stations sont bien touchées par la même perturbation atmosphérique (averse). Au contraire l'absence de corrélation décrit que l'averse a touché qu'une station

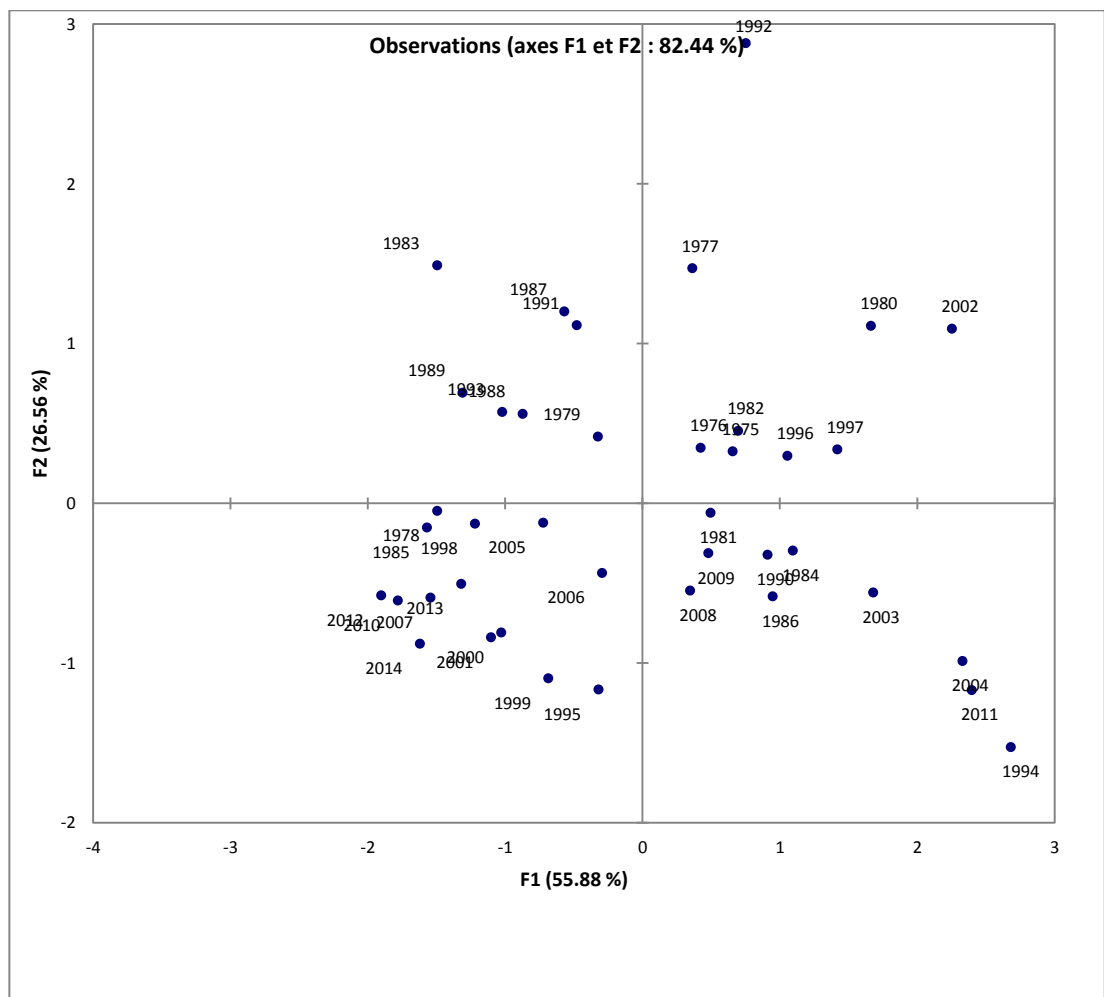


Figure 6.3 Projection des années sur les composantes principales

## 6.5 Conclusion

La méthode d'analyse en composantes principales appliquées aux valeurs des précipitations maximales journalières sur les stations étudiées, nous a permis de conclure que la corrélation entre les stations est contrôlée par plusieurs facteurs, notamment la distance entre les stations...., où la proximité des stations témoigne d'une bonne corrélation entre les variables. D'autre part, les stations qui ont une grande distance entre elles, la corrélation entre elles est faible.

L'analyse multidimensionnelle (ACP) a montré que la proximité des quantités des précipitations entre les années, témoigne d'une moyenne corrélation, ce peut s'expliquer par la nature des perturbations météorologiques.

Cela peut s'expliquer par la nature des perturbations météorologiques dans laquelle les zones sont touchées. Lorsque, dans le cas d'une bonne corrélation, ceci est la preuve que le même niveau de frappe aérienne dans ces années.

# **Conclusion générale**



L'objectif principal de cette étude était le recensement des précipitations sur la zone de Sahara septentrional et de chercher s'il y a une possibilité de corrélation quel que soit entre les trois stations sur la période d'observation 1975- 2014.

En effet, nous avons appuyés sur un ensemble de méthodes statistiques. Dans la première étape, nous avons utilisé les méthodes d'ajustement, dans la deuxième étape, nous avons utilisé l'analyse dans les principaux composants.

L'étude de l'analyse pluviométrique des stations Ghardaïa Hassi R'mel Touggourt a permis de dégager les points suivants :

- Les distributions des séries pluviométriques montrent une dissymétrie très nette, ce qui implique un ajustement aux lois de Galton et loi de Gumbel
- Les précipitations sont faibles
- Il apparaît, que très longues périodes d'observation sont nécessaires pour connaître la pluviosité avec une approximation suffisante
- Pour les stations étudiées, on remarque, qu'elles s'ajustent correctement avec la loi de Galton, cela explique l'alignement des points expérimentaux.
- D'après les résultats du test de Khi2 appliqué sur la série des Pluies journalières maximales observées à Ghardaïa (1975-2014) dans la loi de Galton (log normal) et dans la loi de Gumbel, on remarque que : l'ajustement est accepté avec la loi de Galton et accepté avec la loi de Gumbel
- D'après les résultats du test de Khi2 appliqué sur la série des Pluies journalières maximales observées à Hassi R'mel (1975-2014) dans la loi de Galton (log normal) et dans la loi de Gumbel, on remarque que : l'ajustement est accepté avec la loi de Galton et rejeté avec la loi de Gumbel
- D'après les résultats du test de Khi2 appliqué sur la série des Pluies journalières maximales observées à Touggourt (1975-2014) dans la loi de Galton (log normal) et dans la loi de Gumbel, on remarque que : l'ajustement est rejeté avec la loi de Galton et rejeté avec la loi de Gumbel
- La méthode d'analyse en composantes principales appliquées aux valeurs des précipitations maximales journalières sur les stations étudiées, nous a permis de conclure que la corrélation entre les stations est contrôlée par plusieurs facteurs, notamment la distance entre les stations...., où la proximité des stations témoigne

d'une bien corrélation entre les variables. D'autre part, les stations qui ont une grande distance entre eux, la corrélation entre eux est faible.

- La présence d'une bonne corrélation au cours d'une période indique que les stations sont bien touchées par la même perturbation atmosphérique (averse). Au contraire l'absence de corrélation décrit que l'averse a touché qu'une station

# **Annexes**

**Tableau 01 les valeurs d'Ajustement de la série pluviométrique à la loi de log Normal (Ghardaia)**

Pluie	Pluie classe	Ordre	FND	V red	V-Theo	B-inf.	B-sup
24,7	2,9	1	0,0125	-2,2414	3,3479	-16,5998	23,2956
17,2	3,5	2	0,0375	-1,7805	4,5845	-10,0984	19,2674
17,9	6,7	3	0,0625	-1,5341	5,4232	-6,9407	17,7872
11,3	6,8	4	0,0875	-1,3563	6,1224	-4,7819	17,0268
18,6	7	5	0,1125	-1,2133	6,7495	-3,1115	16,6104
46,5	7,11	6	0,1375	-1,0916	7,3337	-1,7305	16,3979
13	8,6	7	0,1625	-0,9842	7,8909	-0,5402	16,3220
9,4	8,8	8	0,1875	-0,8871	8,4311	0,5160	16,3462
2,9	8,9	9	0,2125	-0,7978	8,9610	1,4735	16,4484
20,2	9,4	10	0,2375	-0,7144	9,4855	2,3562	16,6148
6,8	10	11	0,2625	-0,6357	10,0086	3,1811	16,8361
36,6	10	12	0,2875	-0,5607	10,5335	3,9607	17,1063
8,8	10,9	13	0,3125	-0,4888	11,0631	4,7047	17,4216
8,9	11,3	14	0,3375	-0,4193	11,6000	5,4207	17,7792
6,7	11,9	15	0,3625	-0,3518	12,1465	6,1151	18,1780
27,9	12	16	0,3875	-0,2858	12,7053	6,7933	18,6173
10,9	12,6	17	0,4125	-0,2211	13,2787	7,4598	19,0975
12,6	13	18	0,4375	-0,1573	13,8693	8,1190	19,6195
11,9	14,99	19	0,4625	-0,0941	14,4799	8,7747	20,1850
54,8	16,2	20	0,4875	-0,0313	15,1135	9,4308	20,7963
30	17,2	21	0,5125	0,0313	15,7735	10,0908	21,4563
12	17,9	22	0,5375	0,0941	16,4638	10,7586	22,1689
23	18	23	0,5625	0,1573	17,1886	11,4384	22,9389
3,5	18	24	0,5875	0,2211	17,9531	12,1343	23,7720
26	18,6	25	0,6125	0,2858	18,7633	12,8513	24,6754
18	19,5	26	0,6375	0,3518	19,6265	13,5950	25,6579
18	20,2	27	0,6625	0,4193	20,5512	14,3720	26,7305
24	23	28	0,6875	0,4888	21,5485	15,1901	27,9070
33	23	29	0,7125	0,5607	22,6319	16,0591	29,2047
46	24	30	0,7375	0,6357	23,8189	16,9914	30,6464
7	24,7	31	0,7625	0,7144	25,1325	18,0032	32,2617
16,2	26	32	0,7875	0,7978	26,6036	19,1161	34,0910
10	27,9	33	0,8125	0,8871	28,2755	20,3604	36,1906
23	30	34	0,8375	0,9842	30,2111	21,7800	38,6422
19,5	33	35	0,8625	1,0916	32,5067	23,4425	41,5709
8,6	36,6	36	0,8875	1,2133	35,3203	25,4594	45,1813
50	46	37	0,9125	1,3563	38,9378	28,0335	49,8422
7,11	46,5	38	0,9375	1,5341	43,9579	31,5939	56,3218
10	50	39	0,9625	1,7805	51,9997	37,3168	66,6826
14,99	54,8	40	0,9875	2,2414	71,2075	51,2598	91,1552

**Tableau 02 les valeurs d'Ajustement de la série pluviométrique à la loi de Gumble (Ghardaia)**

Valeurs de départ	Valeurs classées	Ordre de Classement	Fréquence expérimente	Variable réduite	Valeur expérimente	Valeur théorique	Borne inférieure	Borne Supérieure
24.7	2.9	1	0.013	-1.478	2.90	-1.59	-6.17	1.65
17.2	3.5	2	0.038	-1.189	3.50	1.33	-2.61	4.21
17.9	6.7	3	0.063	-1.020	6.70	3.04	-0.55	5.73
11.3	6.8	4	0.088	-0.890	6.80	4.35	1.01	6.91
18.6	7	5	0.113	-0.782	7.00	5.45	2.31	7.92
46.5	7.11	6	0.138	-0.685	7.11	6.42	3.44	8.82
13	8.6	7	0.163	-0.597	8.60	7.31	4.47	9.66
9.4	8.8	8	0.188	-0.515	8.80	8.14	5.42	10.45
2.9	8.9	9	0.213	-0.437	8.90	8.93	6.30	11.21
20.2	9.4	10	0.238	-0.363	9.40	9.68	7.14	11.95
6.8	10	11	0.263	-0.291	10.00	10.41	7.95	12.68
36.6	10	12	0.288	-0.220	10.00	11.12	8.72	13.40
8.8	10.9	13	0.313	-0.151	10.90	11.82	9.47	14.12
8.9	11.3	14	0.338	-0.083	11.30	12.52	10.20	14.84
6.7	11.9	15	0.363	-0.015	11.90	13.20	10.91	15.56
27.9	12	16	0.388	0.053	12.00	13.89	11.62	16.30
10.9	12.6	17	0.413	0.122	12.60	14.58	12.32	17.04
12.6	13	18	0.438	0.190	13.00	15.28	13.01	17.81
11.9	14.99	19	0.463	0.260	14.99	15.98	13.70	18.59
54.8	16.2	20	0.488	0.331	16.20	16.70	14.40	19.39
30	17.2	21	0.513	0.403	17.20	17.43	15.10	20.22
12	17.9	22	0.538	0.477	17.90	18.17	15.81	21.08
23	18	23	0.563	0.553	18.00	18.94	16.53	21.97
3.5	18	24	0.588	0.631	18.00	19.74	17.26	22.90
26	18.6	25	0.613	0.713	18.60	20.56	18.02	23.87
18	19.5	26	0.638	0.798	19.50	21.42	18.80	24.90
18	20.2	27	0.663	0.887	20.20	22.33	19.61	25.98
24	23	28	0.688	0.982	23.00	23.28	20.46	27.12
33	23	29	0.713	1.082	23.00	24.29	21.35	28.35
46	24	30	0.738	1.189	24.00	25.38	22.31	29.67
7	24.7	31	0.763	1.305	24.70	26.55	23.33	31.11
16.2	26	32	0.788	1.432	26.00	27.83	24.44	32.68
10	27.9	33	0.813	1.572	27.90	29.25	25.66	34.43
23	30	34	0.838	1.730	30.00	30.85	27.02	36.41
19.5	33	35	0.863	1.911	33.00	32.68	28.58	38.69
8.6	36.6	36	0.888	2.126	36.60	34.85	30.42	41.40
50	46	37	0.913	2.391	46.00	37.53	32.68	44.75
7.11	46.5	38	0.938	2.740	46.50	41.07	35.65	49.19
10	50	39	0.963	3.264	50.00	46.37	40.08	55.86
14.99	54.8	40	0.988	4.376	54.80	57.61	49.44	70.06

**Tableau 03 les valeurs d'Ajustement de la série pluviométrique à la loi de log Normal (Hassi R'mel)**

Pluie	Pluie clas	ordre	FND	V red	V theo	Binf	Bsup
48.25	4.83	1	0.0125	-2.241	9.336	-29.005	47.677
56.30	14.55	2	0.0375	-1.780	12.424	-15.797	40.646
32.70	17.00	3	0.0625	-1.534	14.474	-9.290	38.239
14.55	17.30	4	0.0875	-1.356	16.161	-4.797	37.120
28.15	17.65	5	0.1125	-1.213	17.659	-1.294	36.613
20.05	19.95	6	0.1375	-1.092	19.044	1.622	36.466
74.80	20.05	7	0.1625	-0.984	20.355	4.150	36.560
79.64	21.21	8	0.1875	-0.887	21.618	6.404	36.831
4.83	23.63	9	0.2125	-0.798	22.849	8.458	37.241
82.94	25.53	10	0.2375	-0.714	24.062	10.359	37.765
23.63	27.05	11	0.2625	-0.636	25.265	12.142	38.388
48.55	27.10	12	0.2875	-0.561	26.467	13.834	39.101
27.10	27.10	13	0.3125	-0.489	27.674	15.453	39.896
28.70	28.07	14	0.3375	-0.419	28.892	17.015	40.769
17.00	28.15	15	0.3625	-0.352	30.127	18.535	41.720
61.15	28.70	16	0.3875	-0.286	31.385	20.021	42.748
27.10	28.96	17	0.4125	-0.221	32.670	21.485	43.854
31.15	31.11	18	0.4375	-0.157	33.988	22.936	45.040
17.30	31.15	19	0.4625	-0.094	35.346	24.380	46.311
82.25	32.45	20	0.4875	-0.031	36.749	25.826	47.672
32.45	32.70	21	0.5125	0.031	38.205	27.282	49.128
88.40	41.41	22	0.5375	0.094	39.722	28.756	50.688
76.00	48.25	23	0.5625	0.157	41.309	30.256	52.361
41.41	48.55	24	0.5875	0.221	42.976	31.791	54.160
28.07	49.53	25	0.6125	0.286	44.735	33.372	56.099
28.96	49.67	26	0.6375	0.352	46.602	35.009	58.195
27.05	56.30	27	0.6625	0.419	48.594	36.717	60.471
87.00	57.54	28	0.6875	0.489	50.733	38.512	62.955
79.13	61.15	29	0.7125	0.561	53.047	40.413	65.680
80.01	65.03	30	0.7375	0.636	55.570	42.447	68.693
49.67	74.80	31	0.7625	0.714	58.349	44.646	72.052
49.53	76.00	32	0.7875	0.798	61.446	47.055	75.838
25.53	76.97	33	0.8125	0.887	64.947	49.734	80.160
57.54	79.13	34	0.8375	0.984	68.977	52.771	85.182
65.03	79.64	35	0.8625	1.092	73.725	56.303	91.147
21.21	80.01	36	0.8875	1.213	79.504	60.551	98.458
76.97	82.25	37	0.9125	1.356	86.874	65.915	107.832
19.95	82.94	38	0.9375	1.534	96.998	73.234	120.763
31.11	87.00	39	0.9625	1.780	113.005	84.783	141.226
17.65	88.40	40	0.9875	2.241	150.387	112.046	188.728

**Tableau 04 les valeurs d'Ajustement de la série pluviométrique à la loi de Gumble (Hassi R'mel )**

Valeurs de départ	Valeurs classées	Ordre de classement	Fréquence expérimentale	Variable réduite	Valeur expérimentale	Valeur théorique	Borne inférieure	Borne supérieure
48	5	1	0.013	-1.478	4.83	4.80	-3.99	11.03
56	15	2	0.038	-1.189	14.55	10.41	2.84	15.94
33	17	3	0.063	-1.020	17.00	13.69	6.80	18.87
15	17	4	0.088	-0.890	17.30	16.21	9.80	21.14
28	18	5	0.113	-0.782	17.65	18.32	12.29	23.07
20	20	6	0.138	-0.685	19.95	20.19	14.48	24.81
75	20	7	0.163	-0.597	20.05	21.90	16.45	26.42
80	21	8	0.188	-0.515	21.21	23.50	18.27	27.94
5	24	9	0.213	-0.437	23.63	25.01	19.97	29.40
83	26	10	0.238	-0.363	25.53	26.46	21.58	30.82
24	27	11	0.263	-0.291	27.05	27.86	23.12	32.21
49	27	12	0.288	-0.220	27.10	29.23	24.61	33.60
27	27	13	0.313	-0.151	27.10	30.57	26.05	34.98
29	28	14	0.338	-0.083	28.07	31.90	27.45	36.36
17	28	15	0.363	-0.015	28.15	33.22	28.83	37.75
61	29	16	0.388	0.053	28.70	34.54	30.18	39.16
27	29	17	0.413	0.122	28.96	35.87	31.52	40.60
31	31	18	0.438	0.190	31.11	37.21	32.85	42.07
17	31	19	0.463	0.260	31.15	38.56	34.18	43.57
82	32	20	0.488	0.331	32.45	39.93	35.52	45.11
32	33	21	0.513	0.403	32.70	41.33	36.86	46.70
88	41	22	0.538	0.477	41.41	42.77	38.22	48.35
76	48	23	0.563	0.553	48.25	44.25	39.60	50.07
41	49	24	0.588	0.631	48.55	45.77	41.02	51.85
28	50	25	0.613	0.713	49.53	47.36	42.47	53.72
29	50	26	0.638	0.798	49.67	49.01	43.97	55.69
27	56	27	0.663	0.887	56.30	50.75	45.53	57.76
87	58	28	0.688	0.982	57.54	52.58	47.16	59.97
79	61	29	0.713	1.082	61.15	54.53	48.88	62.32
80	65	30	0.738	1.189	65.03	56.61	50.71	64.86
50	75	31	0.763	1.305	74.80	58.86	52.67	67.62
50	76	32	0.788	1.432	76.00	61.33	54.80	70.64
26	77	33	0.813	1.572	76.97	64.05	57.15	74.01
58	79	34	0.838	1.730	79.13	67.12	59.77	77.80
65	80	35	0.863	1.911	79.64	70.64	62.77	82.19
21	80	36	0.888	2.126	80.01	74.81	66.30	87.39
77	82	37	0.913	2.391	82.25	79.96	70.65	93.83
20	83	38	0.938	2.740	82.94	86.76	76.35	102.36
31	87	39	0.963	3.264	87.00	96.93	84.86	115.17
18	88	40	0.988	4.376	88.40	118.53	102.83	142.44

**Tableau 05 les valeurs d'Ajustement de la série pluviométrique à la loi de log Normal  
(Touggourt)**

Pluie	Pluie clas	Ordre	FND	V red	V theo	B-inf	B-sup
15,10	2,94	1	0,0125	-2,241	2,9	-5,997	11,796
14,10	3,32	2	0,0375	-1,780	3,739	-2,809	10,288
20,10	3,53	3	0,0625	-1,534	4,284	-1,230	9,798
7,40	4,25	4	0,0875	-1,356	4,726	-0,138	9,589
13,00	4,51	5	0,1125	-1,213	5,113	0,716	9,511
23,30	4,68	6	0,1375	-1,092	5,469	1,426	9,511
11,70	4,83	7	0,1625	-0,984	5,803	2,042	9,563
14,60	5,18	8	0,1875	-0,887	6,122	2,592	9,652
15,00	5,41	9	0,2125	-0,798	6,431	3,092	9,771
12,30	6,30	10	0,2375	-0,714	6,734	3,555	9,914
6,30	6,90	11	0,2625	-0,636	7,033	3,988	10,078
11,80	7,40	12	0,2875	-0,561	7,330	4,399	10,262
15,90	8,26	13	0,3125	-0,489	7,627	4,791	10,463
11,80	8,28	14	0,3375	-0,419	7,925	5,169	10,681
11,40	9,59	15	0,3625	-0,352	8,226	5,536	10,916
12,40	10,84	16	0,3875	-0,286	8,531	5,894	11,167
15,80	11,40	17	0,4125	-0,221	8,841	6,246	11,436
28,10	11,70	18	0,4375	-0,157	9,158	6,593	11,722
11,80	11,80	19	0,4625	-0,094	9,483	6,938	12,027
11,90	11,80	20	0,4875	-0,031	9,817	7,283	12,351
5,41	11,80	21	0,5125	0,031	10,162	7,628	12,697
14,73	11,90	22	0,5375	0,094	10,521	7,976	13,065
16,64	12,30	23	0,5625	0,157	10,894	8,329	13,458
6,90	12,40	24	0,5875	0,221	11,284	8,689	13,879
4,68	12,83	25	0,6125	0,286	11,695	9,058	14,331
4,83	13,00	26	0,6375	0,352	12,128	9,438	14,818
4,51	13,19	27	0,6625	0,419	12,588	9,832	15,344
22,59	13,33	28	0,6875	0,489	13,080	10,245	15,916
13,19	14,10	29	0,7125	0,561	13,610	10,679	16,541
13,33	14,60	30	0,7375	0,636	14,185	11,140	17,230
8,28	14,73	31	0,7625	0,714	14,814	11,635	17,994
8,26	15,00	32	0,7875	0,798	15,512	12,173	18,851
4,25	15,10	33	0,8125	0,887	16,296	12,766	19,826
9,59	15,80	34	0,8375	0,984	17,193	13,433	20,953
10,84	15,90	35	0,8625	1,092	18,243	14,200	22,285
3,53	16,64	36	0,8875	1,213	19,510	15,112	23,908
12,83	20,10	37	0,9125	1,356	21,111	16,248	25,975
3,32	22,59	38	0,9375	1,534	23,288	17,774	28,802
5,18	23,30	39	0,9625	1,780	26,679	20,130	33,227
2,94	28,10	40	0,9875	2,241	34,405	25,509	43,302



**Tableau 06 les valeurs d'Ajustement de la série pluviométrique à la loi de Gumble (Touggourt)**

Valeurs de départ	Valeurs classées	Ordre de classent	Fréquence expérimente	Variable réduite	Valeur expérimente	Valeur théorique	Borne inférieure	Borne Supérieure
15.10	2.94	1	0.013	-1.478	2.94	2.23	0.19	3.67
14.10	3.32	2	0.038	-1.189	3.32	3.53	1.77	4.81
20.10	3.53	3	0.063	-1.020	3.53	4.29	2.69	5.49
7.40	4.25	4	0.088	-0.890	4.25	4.87	3.39	6.02
13.00	4.51	5	0.113	-0.782	4.51	5.37	3.97	6.47
23.30	4.68	6	0.138	-0.685	4.68	5.80	4.47	6.87
11.70	4.83	7	0.163	-0.597	4.83	6.20	4.93	7.24
14.60	5.18	8	0.188	-0.515	5.18	6.57	5.35	7.60
15.00	5.41	9	0.213	-0.437	5.41	6.92	5.75	7.94
12.30	6.30	10	0.238	-0.363	6.30	7.25	6.12	8.27
6.30	6.90	11	0.263	-0.291	6.90	7.58	6.48	8.59
11.80	7.40	12	0.288	-0.220	7.40	7.90	6.82	8.91
15.90	8.26	13	0.313	-0.151	8.26	8.21	7.16	9.23
11.80	8.28	14	0.338	-0.083	8.28	8.52	7.48	9.55
11.40	9.59	15	0.363	-0.015	9.59	8.82	7.80	9.87
12.40	10.84	16	0.388	0.053	10.84	9.13	8.12	10.20
15.80	11.40	17	0.413	0.122	11.40	9.44	8.43	10.53
28.10	11.70	18	0.438	0.190	11.70	9.75	8.74	10.88
11.80	11.80	19	0.463	0.260	11.80	10.06	9.05	11.22
11.90	11.80	20	0.488	0.331	11.80	10.38	9.36	11.58
5.41	11.80	21	0.513	0.403	11.80	10.71	9.67	11.95
14.73	11.90	22	0.538	0.477	11.90	11.04	9.98	12.33
16.64	12.30	23	0.563	0.553	12.30	11.38	10.30	12.73
6.90	12.40	24	0.588	0.631	12.40	11.74	10.63	13.15
4.68	12.83	25	0.613	0.713	12.83	12.10	10.97	13.58
4.83	13.00	26	0.638	0.798	13.00	12.49	11.32	14.04
4.51	13.19	27	0.663	0.887	13.19	12.89	11.68	14.52
22.59	13.33	28	0.688	0.982	13.33	13.32	12.06	15.03
13.19	14.10	29	0.713	1.082	14.10	13.77	12.46	15.58
13.33	14.60	30	0.738	1.189	14.60	14.25	12.88	16.16
8.28	14.73	31	0.763	1.305	14.73	14.77	13.34	16.80
8.26	15.00	32	0.788	1.432	15.00	15.34	13.83	17.51
4.25	15.10	33	0.813	1.572	15.10	15.98	14.37	18.29
9.59	15.80	34	0.838	1.730	15.80	16.69	14.98	19.17
10.84	15.90	35	0.863	1.911	15.90	17.51	15.68	20.18
3.53	16.64	36	0.888	2.126	16.64	18.47	16.50	21.39
12.83	20.10	37	0.913	2.391	20.10	19.67	17.51	22.89
3.32	22.59	38	0.938	2.740	22.59	21.25	18.83	24.87
5.18	23.30	39	0.963	3.264	23.30	23.61	20.81	27.84
2.94	28.10	40	0.988	4.376	28.10	28.62	24.98	34.17

**Tableau 07 Application du test de  $\text{Khi}^2$  a une loi log-normal (Ghardaia)**

ordre	X min	X max	ln Xmin	ln X max	Z X min	Z X max	FND x min	FND x max	foi	fti	Xi <sup>2</sup>
1	$-\infty$	8	0	2.079	-4.015	-0.957	0.000	0.171	6	6.84	0.1031578
2	8	11	2.079	2.398	-0.957	-0.488	0.171	0.315	7	5.76	0.2669444
3	11	17	2.397	2.833	-0.488	0.152	0.315	0.559	7	9.76	0.7804918
4	17	20	2.833	2.996	0.152	0.391	0.559	0.651	6	3.68	1.4626087
5	20	30	2.995	3.401	0.391	0.987	0.651	0.836	7	7.4	0.0216216
6	30	$+\infty$	3.401	$+\infty$	0.987	$+\infty$	0.836	1.00	7	6.56	0.0295122
										som Xi <sup>2</sup>	2.66433665

**Tableau 08 Application du test de  $\text{Khi}^2$  a une loi de Gumbel (Ghardaia)**

ordre	X min	X max	Z X min	Z X max	FND x min	FND x max	foi	fti	Xi <sup>2</sup>	
1	$-\infty$	8	$-\infty$	-0.535	0.000	0.181	6	7.253	0.21650011	
2	8	11	-0.535	-0.235	0.181	0.282	7	4.037	2.17375189	
3	11	14	-0.235	0.065	0.282	0.392	5	4.380	0.08764133	
4	14	19	0.065	0.565	0.392	0.566	7	6.987	2.36646E-	
5	19	24.5	0.565	1.115	0.566	0.720	5	6.159	0.21803878	
6	24.5	35	1.115	2.165	0.720	0.892	5	6.847	0.49804779	
7	35	$+\infty$	2.165	$+\infty$	0.892	1.000	5	4.336	0.10153858	
									som Xi <sup>2</sup>	3.29554217

**Tableau 09 Application du test de  $\text{Khi}^2$  a une loi log-normal (H-rmel)**

ordre	X min	X max	ln Xmin	ln X max	Z X min	Z X max	FND x min	FND x max	foi	fti	Xi <sup>2</sup>
1	$-\infty$	20	0.000	2.996	-5.656	-0.975	0.00	0.166	6	6.64	0.06168
2	20	28	2.996	3.332	-0.975	-0.450	0.166	0.329	7	6.52	0.03533
3	28	32	3.332	3.466	-0.450	-0.241	0.329	0.405	6	3.04	2.88210
4	32	50	3.466	3.912	-0.241	0.456	0.405	0.673	7	10.72	1.29089
5	50	76	3.912	4.331	0.456	1.111	0.673	0.864	6	7.64	0.35204
6	76	$+\infty$	4.331	$+\infty$	1.111	$+\infty$	0.864	1.00	8	5.44	1.20470
										som Xi <sup>2</sup>	5.8267

**Tableau 10 Application du test de  $\text{Khi}^2$  a une loi de Gumbel (H-rmel)**

ordre	X min	X max	Z X min	Z X max	FND x min	FND x max	foi	fti	Xi <sup>2</sup>
1	$-\infty$	20	$-\infty$	-0.7025	0.000	0.133	6	5.312	0.088995249
2	20	28	-0.7025	-0.2865	0.133	0.264	7	5.248	0.585034256
3	28	32	-0.2865	-0.0785	0.264	0.339	6	3.001	2.997822648
4	32	49	-0.0785	0.8054	0.339	0.640	5	12.024	4.103211449
5	49	65	0.8054	1.6374	0.640	0.823	5	7.346	0.74913539
6	65	80	1.6374	2.4174	0.823	0.915	6	3.658	1.500069991
7	80	$+\infty$	2.41748	$+\infty$	0.915	1.000	5	3.412	0.739633273
								som Xi <sup>2</sup>	10.76390226

**Tableau 11 Application du test de  $\text{Khi}^2$  a une loi log-normal (Touggourt)**

ord re	X min	X max	ln Xmin	ln X max	Z X min	Z X max	FND x min	FND x max	fo i	fti	Xi <sup>2</sup>
1	$-\infty$	5	0.000	1.609	-4.10	-1.23	0.00	0.109	7	4.36	1.5985321
2	5	9	1.609	2.197	-1.23	-0.18	0.109	0.428	7	12.7 6	2.6001253
3	9	12	2.197	2.485	-0.18	0.33	0.428	0.629	8	8.04	0.000199
4	12	14	2.485	2.639	0.33	0.60	0.629	0.725	6	3.84	1.215
5	14	16	2.639	2.773	0.60	0.84	0.725	0.799	7	2.96	5.5140540
6	16	$+\infty$	2.773	$+\infty$	0.844	$+\infty$	0.799	1.00	5	8.04	1.1494527
										som Xi <sup>2</sup>	12.077363

**Tableau 12 Application du test de  $\text{Khi}^2$  a une loi de Gumbel (Touggourt)**

ordre	X min	X max	Z X min	Z X max	FND x min	FND x max	foi	fti	Xi <sup>2</sup>
1	$-\infty$	5	$-\infty$	-0.8558	0.000	0.095	7	3.802	2.68908174
2	5	8	-0.8558	-0.1958	0.095	0.296	5	8.051	1.156088131
3	8	11.5	-0.1958	0.5742	0.296	0.569	5	10.923	3.211913674
4	11.5	12	0.5742	0.6842	0.569	0.604	5	1.376	9.543817812
5	12	14	0.6842	1.1242	0.604	0.723	6	4.751	0.328245137
6	14	16	1.1242	1.5642	0.723	0.811	7	3.544	3.370493236
7	16	$+\infty$	1.5642	$+\infty$	0.811	1.000	5	7.552	0.862649348
								som Xi <sup>2</sup>	21.16228908

*Références  
bibliographiques*

---

## Références bibliographiques

- [1] ACHOUR. M. Magister., thème sur Vulnérabilité et protection des eaux souterraine en zone aride : cas de la vallée de M'Zab. Université D'ORAN
- [2] BENSAHA H et ACHOUR.M., 2010, Réflexions pour préserver l'environnement, cas de la valle du M'zab .Communication présentée au colloque international, Univ. Provence Marseille, France 2011
- [3] CHIBANE .T., 2015 (Diagnostic et réhabilitation du réseau d'alimentation en eau potable de la commune de Hassi R'mel), mémoire fin d'étude d'ingénieur, école nationale polytechnique.
- [4] COOLEY W.W et LOHNES P.R., 1971, multivariate data analysis. New York
- [5] David C. Howell., 2004. Méthodes statistiques en sciences humaines cour 12 corrélation et régression Université Paris Descartes Institut de Psychologie
- [6] DAVIS J.C., 1986.statistique and data analysis in geology, Paper back – International Edition
- [7] DOUADI., (1996). Evaluation de la variabilité intra et inter cultivars du palmier dattier dans la région de Ouargla, Oued Souf et Oued Righ. Mém. d'Ing. d'Agro. Ouargla Pp : 14, 15.
- [8] EZECKIEL (M.), Fox (K. A.), 1959 ,Methods of correlation and regression analysis.3e edition, John Wiley and sons, New York.
- [9] Florentina Moatar., 2011. Problématiques hydrologiques et méthodes d'analyse, Université François Rabelais de Tours
- [10] Greenacre M.G., 1984. Theory and applications of correspondence analysis. Academic Press, 364p. University of South Africa

- [11] HARKOUKEN YACIN., 1999. Mémoire de fin d'étude en vue de l'obtention du diplôme d'état en génie rural, analyse pluviométrique du bas SAHRA ALGERIEN
- [12] J.P. LABORDE., 2000, Eléments d'hydrologie de surface, p33, l'Université de Nice
- [13] J.P. LABORDE ., 1982, Eléments d'hydrologie de surface, instit national polytechnique de lorraine. France
- [14] LEBART L, MORINEAU A.et WARWIK K.M., 1984 Multivariate descriptive statistical analysis. New York, 231p.
- [15] LEBDI N., (2001). Dynamique interne du milieu agricole saharien: déclin ou renouveau des systèmes de production? (Cas de cinq zones de la vallée d'oued Righ). Mém. d'Ing d'agro. ITAS Ouargla, pp : 22, 26, 27, 54.
- [16] LEGENDRE.L., 1984, Ecologie numérique-Tome2-La structure des données écologiques Collection d'écologie, Université Laval. France
- [17] MASSON (J. M.) ., 1971 , Thèse de Docteur-Ingénieur. Université des Sciences et Techniques du Languedoc, Montpellier.
- [18] MORLAT. G., 1952. Les méthodes statistiques. Electricité de France, Direction des Etudes et Recherches. Paris.
- [19] Office nationale de météo, données climatiques de la période 2011-2015 station de Hassi R'mel.
- [20] P.DUBREWL., 1974. initiation à l'analyse hydrologie Masson Cie éd. Paris
- [21] RAHALI OUSSAMA. ,2013. Mémoire de master en sciences de l'eau et l'environnement, l'impact du changement climatique sur les régimes pluviométriques de la ville de Ghardaïa période (1970-2014)

- [22] RICCO (R). ,2015. Analyse de corrélation, université lumière Lyon 2
- [23] ROCHE. M., 1961 Contribution au problème de l'abattement des pluies journalières.13 p, Paris.
- [24] SARI AHMED ABDELWAHED., 2002. L'hydraulique de surface, cours et exercices, Université de Bab Ezzouar. Alger

## Résumé

L'étude de la variabilité climatique a une grande importance pour la gestion et la planification des ressources en eau. Ce travail traite la variation du régime pluviométrique sur les trois stations: Ghardaïa et Hassi R'mel et Touggourt, ainsi que d'étudier l'existence d'une corrélation simple ou multiple entre les trois stations. D'une autre façon d'étudier l'étendue des perturbations qui touchent la région. Les séries chronologiques des pluies sur les trois stations utilisées à cet effet sont ceux de la période 1975-2014. L'application des méthodes d'ajustement et les tests statistiques de tendances ont permis de mettre en évidence une variabilité climatique caractérisée par des longues périodes de sècheresse, alternées a des rares périodes pluvieuses. L'application de l'ACP sur données pluviométriques a déciller qu'il ya une absence de corrélation (ou trop faible) entre les trois stations, à l'exception de quelques saisons là où on observe que les perturbations sont suffisamment étendues pour toucher toutes les stations. Ce travail constitue une base indispensable pour une meilleure connaissance du régime pluviométrique de la région.

**Mots-clés :** Changement climatique, pluies journalières, tendances, corrélation multiple, tests statistiques, ACP, ajustement, Ghardaïa, Hassi R'mel, Touggourt

### ملخص

تعتبر دراسة تقلبات المناخ ذات أهمية كبيرة لإدارة وتخطيط الموارد المائية. يتعامل هذا العمل مع اختلاف نظام سقوط الأمطار في المحطات الثلاث: غرداية و وحاسي رمل وتقرت ، وكذلك لدراسة وجود علاقة بسيطة أو متعددة بين المحطات الثلاث. وكذلك طريقة أخرى لدراسة مدى الاضطرابات التي تؤثر على المنطقة.

السلسلة الزمنية للأمطار على المحطات الثلاث المستخدمة لهذا الغرض هي تلك الفترة من 1975-2014. وقد أتاح تطبيق طرق التعديل والاختبارات الإحصائية للاتجاهات إبراز تغير مناخي يتميز بفترات طويلة من الجفاف، يتناوب مع فترات ممطرة نادرة، وقد أظهر تطبيق ACP على بيانات هطول الأمطار أنه لا يوجد ارتباط (أو ضعيف للغاية) بين المحطات الثلاث ، باستثناء عدد قليل من المواسم التي لوحظ فيها أن الاضطرابات واسعة بما يكفي للتأثير على جميع المحطات. هذا العمل هو أساس لا غنى عنه للحصول على معرفة أفضل لنظام هطول الأمطار في المنطقة.

**الكلمات المفتاحية:** تغير المناخ ، أمطار اليومية ، اتجاهات ، ارتباط متعدد ، اختبارات الإحصائية ، ACP ، التعديل ، غرداية ، حاسي الرمل ، تقرت

### Summary

The study of climate variability is of great importance for the management and planning of water resources. This work deals with the variation of the rainfall regime on the three stations: Ghardaïa and Hassi R'mel and Touggourt, as well as to study the existence of a simple or multiple correlation between the three stations. Another way to study the extent of the disturbances affecting the region. The time series of the rains on the three stations used for this purpose are those of the period 1975-2014. The application of the methods of adjustment and the statistical tests of tendencies made it possible to highlight a climatic variability characterized by long periods of drought, alternated with rare rainy periods. The application of PCA on rainfall data has revealed that there is no correlation (or too weak) between the three stations, with the exception of a few seasons where it is observed that the disturbances are wide enough to affect all the stations. This work is an indispensable basis for a better knowledge of the rainfall regime of the region.

**Keywords:** Climate change, daily rainfall, trends, multiple correlations, statistical tests, ACP, adjustment, Ghardaïa, Hassi R'mel, Touggourt